Large-scale simulation of population dynamics for socio-demographic analysis

Cristina Montañola Sales
Computer Scientist

A Thesis presented for the degree of
Doctor in Philosophy

Directed by Dr. Josep Casanovas Garcia
Co-Directed by Dr. Adriana Kaplan Marcusán

Universitat Politècnica de Catalunya - BarcelonaTech

2014

This research was funded by Barcelona Supercomputing Center and The UK Royal Society under the 2009/R2 International Joint Project Program number JP090402. Additional support was provided by Computing Laboratory of Barcelona School of Informatics (inLab FIB) at Universitat Politècnica de Catalunya - BarcelonaTech. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect those of the sponsors.

Summary

Computer modelling and complex systems simulation have dominated the scientific debate over the last decade, providing important outcomes in biology, geology and life sciences. In the social sciences, the number of research groups currently developing research programs in this direction is increasing. The results are extremely promising since simulation technologies have the potential to become an essential tool in the field.

Agent-based modelling is widely recognised as one of the techniques with more potential to develop useful simulations of social interacting systems. The approach allows to specify complex behavioural and cognitive rules at the individual level; and through aggregation, the output at the macro level can be derived. Increasingly, the output of micro-level simulation models is used as an input to policy models. Policy models not only requires detailed micro-level data, but also significant compute power since the number of agents and interactions can be extremely large in some cases.

High performance computing offers a massive supercomputing power which allows us to simulate a large artificial society. In that context, parallel simulation could provide an alternative to speed up the execution of such compute-intensive socio-demographic models. It deals with techniques that allow the use of multiple processors to run a single simulation. Although research in parallel simulation has been around for more than two decades, the number of applications in the social sciences is scarce.

In this thesis, we present a methodology for simulating population dynamics at a large-scale. Especifically, we developed a parallel simulation framework to run demographic models. It simulates the interactions of individuals in a society so the population projection can be obtained. Two of the main obstacles hindering the use of agent-based simulation in practice are (a) its scalability when the analysis requires large-scale models, and (b) its ease-of-use, especially for users with no programming experience. Our approach proposes a solution for both

challenges. On one hand, we give a solution in to simulate large social systems in a parallel environment. We show its potential by studying the peformance of our approach with measures such speedup and execution time. We identify the factors from a demographic model that affect the simulation execution performance. The application is sensible to the architecture where it is run in terms of memory consumption when the population size per node is high. However, our solution scales well when independent of cache issues. Our findings show migrations have a deep effect on performance due to their impact on network communications. Moreover, we investigate the effect of three computer architecture configuration such as unbalanced workload, heterogeneous processing speed and heterogeneous communication latency. Since the application of parallel simulation in demography is new, it is useful to quantify the effect of these factors on performance.

On the other hand, we provide a graphical user interface which allows modellers with no programming background to specify agent-based demographic models and transparently run them in parallel. We believe this will help to remove a major barrier on using simulation although we are aware technical knowledge is necessary to execute scenarios in High Performance Computing facilities.

Two cases studies are presented to support the feasibility of the approach for the social sciences. The first case under study carries out an analysis of the evolution of the emigrated population of The Gambia between 2001 and 2011, a relevant period for immigrations in Spain. The second case study simulates the sociodemographic changes of South Korean during one hundred years. The objective is to rate the feasibility of our methodology for forecasting individual demographic processes. Our results show that agent-based modelling can be very useful in the study of demography. Furthermore, the use of a parallel environment enables the use of larger scale demographic models.

Resumen

En las últimas décadas la modelización computacional y la simulación de sistemas complejos han dominado el debate científico, dando lugar a resultados importantes en áreas como la biología, la geología o las ciencias de la vida. En ciencias sociales el número de grupos de investigación que desarrollan programas en esta dirección no para de crecer. Los resultados son extremadamente prometedores ya que la simulación tiene el potencial para llegar a ser una herramienta esencial en el ámbito de los estudios sociales.

La modelización basada en agentes es ampliamente reconocida como una de las técnicas con mayor potencial para desarrollar simulaciones de sistemas sociales. Este enfoque nos permite especificar reglas de comportamiento y cognición complejas a nivel del individuo y, a través de su agregación, se obtienen resultados a nivel macroscópico. Cada vez más, los resultados individualizados de los modelos de simulación son usados como entrada de modelos de políticas de planificación familiares. Estos modelos no sólo requieren una gran cantidad de datos a nivel microscópico, sino que además precisan de una capacidad de cálculo significativa ya que el nombre de agentes y sus interacciones puede llegar a ser muy grande.

La computación de altas prestaciones ofrece una capacidad de cálculo masiva que nos permite simular a gran escala sociedades artificiales, proveyendo una alternativa para acelerar la ejecución de estos modelos socio-demográficos tan intensivos en sus cálculos. Aunque la investigación en simulación paralela tiene más de veinte años, el número de aplicaciones en las ciencias sociales es escaso.

En esta tesis presentamos una metodología para simular dinámicas poblaciones a gran escala. Concretamente, hemos desarrollado un entorno de simulación paralela que permite emular modelos demográficos, simulando las interacciones de individuos en una sociedad con el fin de obtener la proyección de la población. Dos de los principales obstáculos para el uso de la simulación basada en agentes en su práctica son (a) la escalabilidad cuando el análisis requiere modelos muy grandes, y (b) su facilidad de uso particularmente en usuarios que no tienen experiencia

en programación. Nuestro enfoque propone una solución para ambos retos. Por una parte, proponemos una solución a la simulación a gran escala de sistemas sociales en un entorno paralelo. Mostramos su potencial estudiando su rendimiento, identificando los factores que la hacen más sensible al tiempo de ejecución e investigando el impacto de tres configuraciones conocidas de la arquitectura. Puesto que la aplicación de la simulación paralela en demografía es nueva, es útil calcular el efecto de estos factores en el rendimiento de la metodología que proponemos.

Por otra parte, el entorno desarrollado incluye una interface gráfica de usuario que permite a usuarios sin experiencia en programación especificar modelos demográficos basados en agentes y ejecutarlos de forma transparente en un entorno paralelo. Con este enfoque, ayudamos a superar una gran barrera en el uso de la simulación aunque un conocimiento técnico para ejecutar escenarios en una arquitectura de altas prestaciones sea necesario. Finalmente, presentamos dos casos de estudio que ponen de manifiesto el alcance de nuestro enfoque para las ciencias sociales. El primero realiza un análisis de la evolución de la población gambiana emigrada a España entre 2001 y 2011, un periodo especialmente relevante para las inmigraciones en España. El segundo caso simula los cambios sociodemográficos de Corea del Sud durante cien años. El objetivo es mostrar lo que nuestra metodología puede aportar para el estudio en profundidad de procesos demográficos. Nuestros resultados muestran que la modelización basada en agentes puede ser de gran utilidad en demografía. Así mismo, el uso de un entorno de computación paralelo nos permite simular la demografía poblacional a gran escala.

Contents

List of Figures

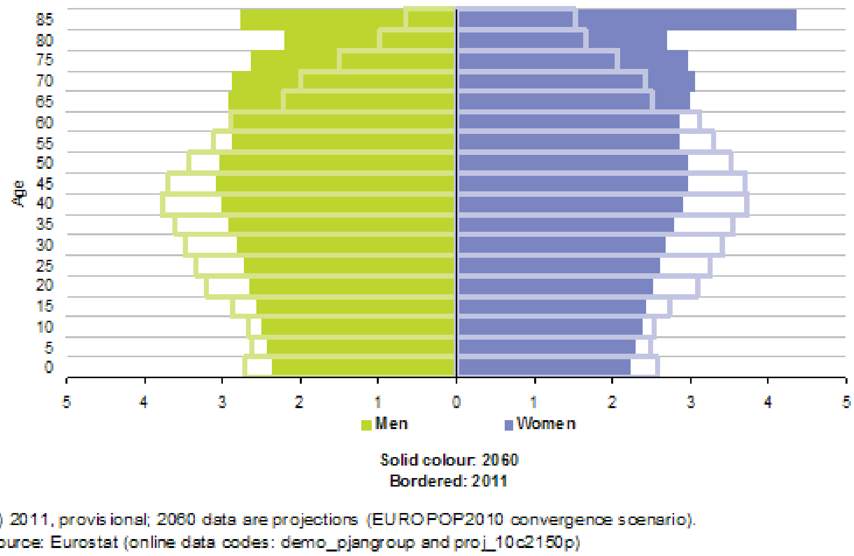
List of Tables

Chapter 1

Introduction

Population dynamics influence every aspect of human, social and economic development. Today, our world faces the challenge to manage an unprecedented population growth. We are already 7 billions of people (up from 2.5 billion in 1950). In the European Union, governments confront a particular scenario. According to the European Commission, by 2025 more than 20% of Europeans will be 65 or over. Moreover, it is expected a particular rapid increase of people over 80. By 2060, the population of Europe is projected to reach 517 million (while today is 501 million according to Eurostat) with 30% aged 65 or over (see Figure 1.1). Today, one out of four working people is above the age of 65. According to these forecasts, in 2060 we will be facing the disproportionate ratio of one person above 65 out of every two working people.

Given these circumstances population trends poses significant challenges for European economy and welfare system. According to Wilmoth (2013), Director of the Population Division at the United Nations Department of Economic and Social Affairs, most policy experts agree the major demands are population growth, population ageing and migrations (both immigrations and emigrations). Population growth and ageing are shaping with increasing influence the social world. Therefore, the need - and to a certain extent, even the responsibility - arises for governments to take the appropriate decisions to respond situations emerged from this trend. Migrations are often perceived as disturbing. However, they can represent a profit both for the migrants and for the countries involved: new labour resources, new individual opportunities, strengthening of the economic and social structures (especially in developed countries, that commonly suffer from a population ageing), and encouragement of social and economic linkages between origin and destination countries.



(a) Population pyramids between 1991 and 2011.

Figure 1.1: Population pyramids of EU-27 countries, between 2011 and 2060 in % of the total population

Demography is a science field that involves the statistical study of human populations. It is often used as the basis for government policies in areas such as labour market, education, healthcare, social welfare and taxation. For example, a fall in the number of births may lead to school closures. Similarly, a rise in the number of elderly people may not only lead to an increase on pensions and social care, but also to a demand rise for healthcare services. In 2011 Europe had an expenditure on pensions of 13% of GDP, ranging from a high of 16.1% in Italy to a low of just over 7% in Ireland. The average expenditure on care for the elderly in hte same year was 0.5% of GDP. Demographers have long studied population trends trying to project them onto the future. However, they have historically been unsuccessful on preventing some of them. For instance, they were not able to forecast the baby-boom that took place after the World War 2 due to the industrialization of European countries. Neither they could foresee the rapid fertility decline in Europe and the attained level of human longevity we experience today. Fertility low rates are now far beyond of what demographers or policy-makers could have expected 50 years ago. Moreover, it seems there is no limit for medical advances that lengthen population life expectancy. Technological and medical science advancements make demographic forecasting a complex problem to handle. Since 1970 elderly death rates in many countries have fallen in an unprecedented way, mostly due to the reduction of heart related and coronary diseases. This situ-

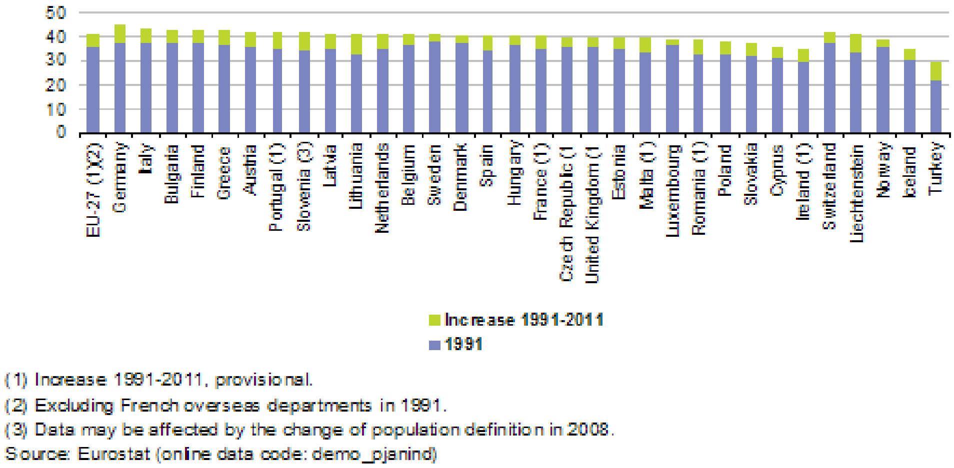
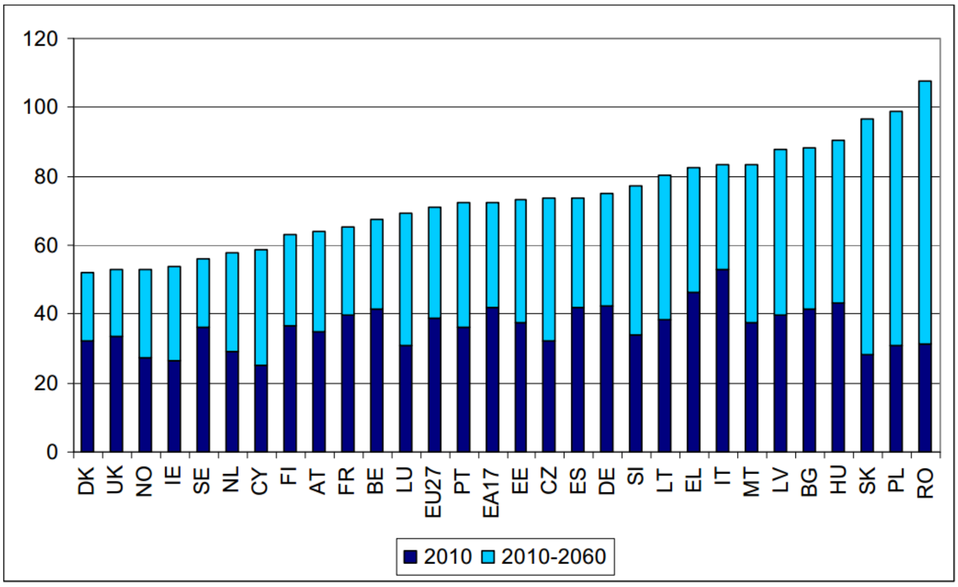


Figure 1.2: Median age of the population in EU-27 countries, 1991-2011.

ation has naturally led to an increase on Europeans' life expectancy and therefore to the increase of the population's age at a global level. To illustrate this, Figure 1.2 shows an increase on median age of European population between 1991 and 2011.

In this framework, Europe is experiencing a slow population growth. With a fertility rate below two (even in traditionally welfare states such as Scandinavian countries) some populations have consequently start to shrink in size. This situation has thereafter led to an acceleration of the population ageing process. In Figure 1.3, we can see the current effective economic old-age dependency in European countries and the expected situation by 2060. Furthermore, the situation has a big implication for government budgets due to the cost of old age pensions, social and medical care.

To confront these challenges, governments need to gather information about population dynamics to understand and analyse population trends. But even when doing so, policy intervention can fail. We have observed such unfortunate cases throughout human history. To mention a particular case which happened recently, about 16 million girls under age 18 give birth each year (UNW, 2013). England has one of the highest teenage pregnancy rates of Western Europe (see Figure 1.4). UK government is investing millions in its prevention due to the risk it supposes on increasing health inequalities, child poverty and the derived economic impact. Only during the period between 2009 and 2010 the amount invested reached £26 million on income support to teenage mothers. But let us analyse how such extraordinary figures were reached and what were the causes that led to that. In 2002, UK government made the morning-after pill more widely available to youngsters on an attempt to improve the situation. However, by 2006, in a



Source: Commission services, EPC.
Note: Inactive population aged 65 and above as a percentage of the employed population aged 15 to 64.

Figure 1.3: Effective economic old-age dependency ratio

blunt demonstration teen pregnancies rose 12% showing that the measure did not have a notable effect on rates of pregnancy and abortion (Glasier, 2006). Later in 2004 the Young People's Development Programme started a 3 years project which consisted of offering education and support for sexual and drug-related topics for 13 to 15 year-old who were deemed at risk of pregnancy, drug abuse or exclusion from school. The policy had been applied previously in New York successfully. However, the same kind of policy in England showed no evidence that the intervention delayed heterosexual experience or reduced the numbers for pregnancy, drunkenness or cannabis use (Wiggins, Bonell, Sawtell, Austerberry, Burchett, Allen, and Strange, 2009).

The main reason behind policy intervention failure is that social problems are most of the time of such complexity and non-linearity that analytic solutions are difficult to outline (Sterman, 2000). In that case we can state social systems are complex non-linear systems. A social system is a collection of individuals who interact among them and their environment, motivated by a set of personal goals. Its behaviour cannot be expressed as the sum of the behaviour of its parts (synergy propriety). That is why dynamics of social systems frequently difficult their analytical understanding (Gilbert and Troitzsch, 2005), and its intricacies create great complexity in their proper evaluation and extracted results.

As an answer to deal with complex non-linear systems, modelling and simu-

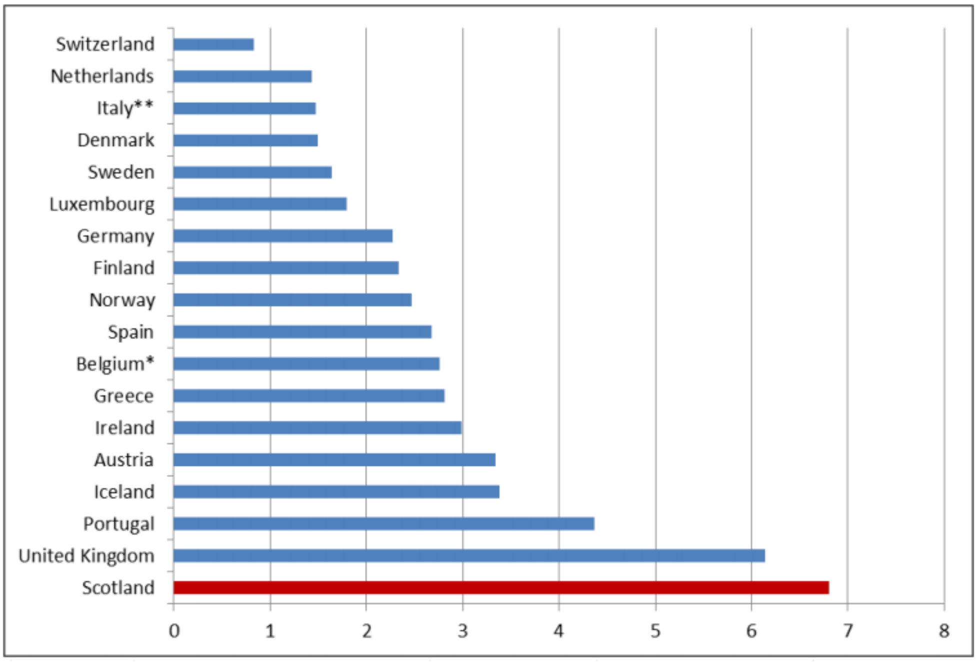


Figure 1.4: Percentage of all live births to mothers under 20 years, Scotland and Western EU 2009 *2007 data, **2008 data, from (Rep, 2013)
Source: The Scotland and European Health for All Database, Scottish Public Health Observatory, 2012

lation techniques were introduced on the second half of the XX century. These techniques have the ability to gain insight into particularly complex problems, hence being a good choice to study social processes. Figure 1.5 shows how both analytical and simulation techniques serve us to find an optimized solution to a real problem. Through observation of the real world, we build models that help us confront problems. To solve them, we either build a mathematical model or a simulation model. With a model, we can then find an optimal solution to the problem under study that can be afterwards applied to reality. The purpose is to avoid doing real experiments, which are often expensive or even infeasible. The main difference between an analytical and simulation solution is that simulation models allow the understanding of the phenomena's insights. However, some social problems can simply not have an analytical solution due to its complexity.

Because of its great potential, scientific areas such as transport, bioinformatics or computational mechanics have adopted simulation as their daily tool to explore complex phenomena. For the social sciences, simulation methodologies are very promising by offering an opportunity to develop a virtual laboratory for exploring and validating current and new approaches. Furthermore, they are not

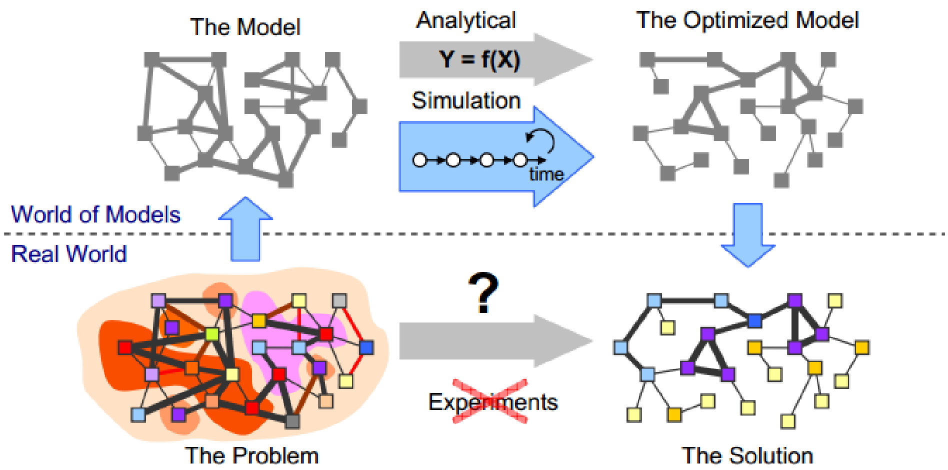


Figure 1.5: Analytical solutions and simulation techniques help us to find an optimized solution to real world problems (after (Borshchev and Filippov, 2004))

exclusively focused on the provision of forecasts or foreseeing the consequences of a hypothetical policy. The aim also expands to the research of past phenomena and providing a better understanding of the present reality. For instance, simulation has been lately used to understand the decay of the Maya empire (Heckbert, 2013). These type of studies are important to deep into the dynamics of a society, understanding the reasons that cause the collapse and guide future course of action decisions.

In policy analysis, the projected population serves as an input to strategy models. Traditionally, policy models allocate people in every population group (usually based on age and gender) to a set of states (such as in full-time employment, in full-time education, unemployed and retired). As computer technology advances, and the data at micro level have become available, multi-state simulation model at the level of individuals has gained popularity for policy analysis. In this model, the state transition of each individual is simulated (such as the transition from being in full-time education to full-time employment). The multi-state model has now become the standard methodology in demography (Willekens, 2005). A complex multi-state model that includes biological factors (such as health-related factors), cognitive factors (such as learning) and social factors (such as social network) is also known as an Agent-based Model in computer science (or a behavioural micro-simulation model in economics).

Simulation is useful for policy evaluation (Arthur and McNicoll, 1975). Quantitatively, simulation models provide the outcomes to justify particular strategies. For example, in evaluating family planning strategies such as giving subsidy aid to families per child or implementing measures to make more flexible work market for women. Options can be studied and compared based on their numerical output in a series of variables. Qualitatively, simulation can show the sequence of events that lead to implications of policy choices that could not have been foreseen before. It can provide information for debate and discussion on particular social issues. Therefore, models not only serve for numerical assessment but also for debate on policy implications.

In this thesis, we present a methodology for simulating large-scale population dynamics. Particularly, we will show how a particular type of simulation model (agent-based) can be used for studying the demographic evolution of a population. Our interest lies on the simulation of large populations, which are often the target of many policies. Our work shows how high-performance computing facilities allow us to simulate large social systems. To this aim we present a solution for parallel simulation of agent-based demographic models.

In the remainder of this chapter, we describe the context of our research in 1.1, addressing the current use of large-scale simulations and presenting how agent-based models can help on simulating population dynamics. Later, we motivate our choice for agent-based modelling in demographic simulation (section 1.2). We show the potential of these techniques and the opportunities it represents for high performace computing and simulation communities. In section 1.3, we introduce our main research questions and hypothesis. Finally, in section 1.4 we explain how this thesis is organized.

1.1 Context

Over the last years technological advances have paved the way for advances in many areas of research such as physics, biology, chemistry and astrophysics. Today, it is difficult to find a science sector which has not been influenced by development on computers. Among the methodologies computers have introduced in the scientific method, simulation has gained a lot of success. Computer simulation started to be used as a scientific tool in meteorology and nuclear physics after de World War II. It provides a virtual laboratory to experiment with models capable of varying numerous assumptions in a set of experiments. Today, simulation has proven to be extremely useful in mathematical modelling in disciplines such as biology, environmental and life sciences, and resulting in the birth of entirely new disciplines (e.g. bioinformatics, geoinformatics, health informatics, etc.).

In the social sciences and humanities, however, simulation has just recently been used in the study of social processes. The use of simulation methodologies to explain population dynamics started in the 1970s. However, it was not until the 1990s that it sparked high interest. Today, social simulation is considered one of the greatest challenges for the following decades as illustrated by initiatives like the European FuturICT project (`http://www.futurict.eu/`). Notwithstanding the interest for simulation in the social sciences, tools and methodologies still need to be further developed. On one hand, computational methods need to be adapted and improve to suit social studies demands. On the other hand, social experts should make an effort to understand simulation capabilities and learn how to build social simulation models.

Social scientists and policy makers have been always confronting the challenge to develop models to understand humans and societies. Now, they can serve from a scientific method that allows them to test their hypothesis, reinforce theories and even, in some cases, make predictions. Nevertheless, social scientists are not alone in this challenge. Computer scientists likewise need to point their energies to understand the needs of social experts on facing simulation and to provide the instruments to help them delve into social processes and their evolution.

Furthermore, today we see latest developments in technology have provide computer infrastructures faster and more powerful than ever. Research in computer architecture makes us witness of big encouraging advances that had lead to high-speed connected systems, cloud computing and supercomputers. Through parallel computation, many research areas (including social simulation) can take profit of these technological improvements and assimilate them to boost their fields. Parallel simulation deals with techniques that allow the use of multiple processors to run a single simulation. In that sense, parallel simulation could provide an alternative to speed up the execution of compute-intensive social models. Although research in parallel simulation has been around for more than two decades the number of applications of parallel simulation in the social sciences is scarce.

Nevertheless, there are reasons to think parallel simulation techniques could become necessary in the future of social simulation. In the real-world, social dynamics are very complex, containing billions of interacting individuals and an important amount of data (both spatial and social). That is the reason a desktop computer or a small cluster might not have enough capacity to manage realistic models. Moreover, the number of simulations to validate social models against evidence records might be substantial (Rubio and Cela, 2010). One solution is to run the simulation model on high-performance environments such as big computer clusters, supercomputers, clouds or grids. However, the problem encounters difficulties on the distribution of computational workload among machines (Fujimoto, 2000), especially when the system is as dynamic as human populations.

There are multiple types of simulation which can be useful for social studies (Gilbert and Troitzsch, 2005). Among them, agent-based modelling and simulation have gain a lot of popularity with application in many areas, from the physical sciences to the social sciences Macal and North (2007). It is useful in exploring problems which entities are capable of executing decision-making processes. These entities, called agents, interact both with other entities and the virtual world where they live (the environment). They execute interaction and decision-making processes over time providing a means to check the evolution of a population. Simulation outputs will produce results which can be compared with empirical data. As a result, we will not only see the most probable situation that explain the phenomena but also we will be able to understand the hidden mechanisms that have driven the population to evolve in that direction.

An agent-based model is a model that is formed by a set of autonomous agents that interact with their environment (including other agents) through a set of internal rules to achieve their objectives. For this reason, agent-based modelling is particularly well suited with the concept of entities with individual decision-making processes interacting within a common environment which can show emergent behaviour. The agent-based model, just like other types of model, is used to represent a real world system to help us understand the system and make decisions. An agent-based model is commonly implemented as a piece of computer code and run using a simulator. Agent-based simulation (ABS) is the computer implementation of an agent-based model.

In social sciences, agent-based modelling is one of the most widely used approaches for simulations of complex social interactions. The main reason is that the object of study in these disciplines, human society present or past, is difficult to analyse through classical analytical techniques due to the unpredictable and changing (dynamic) nature. In this context, agent-based modelling is encouraging the introduction of computer simulations to examine behavioural patterns in complex systems. Simulation opens the door to virtual artificial experimentation of social phenomena. For this reason, agent-based simulation allows the implementation of experiments and studies that would not be feasible otherwise (Pavon, Arroyo, Hassan, and Sansores, 2008). Agent-based simulation is recognized as one of the techniques which could contribute more to develop useful simulations of complex social interactions (Gilbert, 2008).

The purpose of this thesis is to apply agent-based modelling to demography for simulating population dynamics. One of the main applications of demography is the projection of the structure of a population. Traditional methodologies in population projection have been applied at the macro-level. However, supported by the advances in computer technology and the availability of data at micro level (individuals), the use of micro-level simulation models in population projection have become more popular in the last years (Van Imhoff and Post, 1998). The main advantage of this approach is that individual-specific explanatory variables can be included in the model. For example, we may include factors such as age, education level, salary group and ethnicity to model the number of children that an individual female will have. This approach allows us to specify complex behavioural and cognitive rules at the individual level; and through aggregation, the output at the macro level can be derived.

Agent-based modelling is commonly used for small scenarios because the number of agents and interactions between them can be extremely large in some of case studies, thus forcing the scientist to limit its number in order to execute the simulation in a standard computer. However, in the case of policy models, both the amount of compute power required and detailed micro-level data are significant. Those reasons lead us to think parallel simulation techniques might pay an important role in the future of social simulation.

1.2 Motivation

This work started with the interest of Barcelona Supercomputing Center (BSC) to apply high-performance computing to speed up compute-intensive social models. The pursuit came after seen successful applications of parallel simulation in earth science and life science. As a result of that concern, in 2008 a small group was set to explore artificial societies and applications of parallel simulation on humanities and social sciences. As a starting point, BSC decided to profit the deep knowledge of Social and Cultural Anthropology Department of Autonomous University of Barcelona in the demography of The Gambia, and its migration flows to Spain. The purpose of this new line of research is to contribute to policy analysis studies with simulation.

Policy work is mainly based on decision support models that tend to serve large-scale applications to answer real-world questions. New policies may arise processes too dynamic and complex to be foreseen with a good degree of confidence by unaided minds (Arthur and McNicoll, 1975). For example, subsidies can be used to create incentive effects or wealth effects. In many occasions, we want to create one but not the other. Hence, appropriate policies such as allocation formulas and eligibility conditions must be made carefully. Simulation models are able to capture relevant socio-economic phenomena and give the planners a scientific tool to trace out and examine the consequences of applying a certain policy. Therefore, the main goal is to provide a means to test alternative policy choices and to gain insight into the future impact of new plans, which are specially sensible for developing countries. Moreover, the results would free from political influences or biases. These models commonly use real, demographical data and need to be validated to prove their credibility.

After a revision of the literature on parallel simulation, we found little had been done in the social sciences. We saw that as an opportunity and decided to choose agent-based modelling approach to explore large-scale demographic simulations. Currently, there are several tools to support the development and execution of generic agent-based models. However, typically desktop agent-based modelling tools do not scale well to what is required for extremely large applications. Therefore, high-performance computing represents a chance to perform simulations of big scenarios, which are of interest of government and agencies decision problems.

Social simulation community is formed by scientists with very different backgrounds. The reason is the interdisciplinary of population dynamics studies. These backgrounds can be grouped in three different profiles:

1. social scientists who use simulation modelling techniques to test their hypothesis, reinforce social theories, forecast population dynamics or even explain emergent phenomena,

2. scientists between the Artificial Intelligence world and social studies, such as sociology or psychology, who are interested in understanding and reproducing intelligent social behaviour,

3. computer scientists with an interest in social dynamics who can contribute in the design and development of tools and instruments to give a boost in the field.

The work presented in this thesis is framed in the context of this last category. The goal is to help social scientists perform large-scale simulation of social phenomena taking profit of high-speed infrastructures like supercomputers or clusters. In that way, they can perform simulations that would be very costly or just not feasible to perform.

High-Performance Computer (HPC) offers a massive supercomputing power which allows us to simulate a large artificial society. Scientists have long researched on human behaviour. The results can be included in a population model in order to mimic a more realistic human population. This research is exciting because we could find some non-intuitive results due to the interactions among individuals (emergent behaviour). This research is also important because we could compare the effect of different policies on different groups in a population (including minority groups). Relative differences between policies can be analysed before they are put in place. Hence, we could avoid potentially costly and irreversible mistakes. Successful research in this area could improve the quality of life by improving the quality of public policies through the exploitation of computer technologies.

1.3 Objectives

The main objective of this thesis is to investigate how large-scale demographics through computational models could help on understanding population dynamics. The goal of this work is to develop a tool to develop demographic simulation models for supporting decision making processes of policy makers. From the methodological point of view, this work is aimed to answer the following questions:

1. Which are the main processes that describe population dynamics in terms of demographic evolution?

2. Could simulation methodologies be helpful on the study of population dynamics at a large-scale?

3. Which are the determinants that impact the parallel simulation of agent-based models?

4. Could social studies such as demographics, anthropology o sociology take advantage of new technological capacities to boost their research and perform studies to gain deeper knowledge on social dynamics?

5. Could parallel computing, which has successfully been applied to research areas such as earth sciences or life sciences, contribute to simulate large-scale social phenomena?

6. Which are the factors of demographic simulation that affect its execution performance?

7. Is it possible to design a simulation tool able to harness the available compute power efficiently?

To answer these questions our objectives are:

- To research on individual-based approaches to representing population dynamics for simulating them

- To develop a computational framework that can model the demographic evolution of a population to assist in decision making process for policy analysis

- To develop strategies to profit from High Performance Computing infrastructures to improve the simulation of large populations.

- To apply this methodology to a real-case study to simulate the demographic evolution of Gambians in Spain.

- To apply this methodology to a second case study to simulate the population dynamics in South Korea.

In order to achieve these objectives we design and develop a demographic simulator which is able to take advantage of parallel computing to simulate the demographic evolution of large-scale populations. We choose to use agent-based models so we can model the population at an individual-level. Human social behaviours are thereafter simulated through modelling individual agent's behaviour and its interaction with other agents. Since it is not possible to completely model human behaviour, we only model the most common processes of demographics. The main emphasis has been to develop a framework that is flexible enough to allow new social processes to be dynamically integrated with the system.

Moreover, the choice of two complex problems in the Demography and Anthropology domains as cases studies, such as the demographic evolution of Gambians in Spain and the population dynamics in South Korea, was considered essential to support the feasibility of the approach for the social sciences. The problems are modelled following the proposed methodology and using the given agent framework. The agent-based model should provide an insight of the social phenomena studied. The first case under study carries out an analysis of the evolution of the population of The Gambia between 2001 and 2011. The Gambia is a small country in the Sub-Saharan Africa. Despite its small size, its migration rate has been quite high last 30 years. The period of our study is interesting for research because migration flows from The Gambia to Spain increased significantly. In that decade, Spanish economy flourished probably increasing the expectancies of Gambian migrants and increasing migration movements.

The second case study simulates the socio-demographic changes of South Korean from 1990 to 2090. The objective is to rate the feasibility of our methodology for forecasting individual demographic processes. The motivation is the current collaboration with the Big Data Software Platform research department at the Electronics and Telecommunications Research Institute (ETRI) in Daejeon, South Korea. ETRI is currently engaged in a project to explore new simulation concepts for developing an open full-scale individual based modelling simulation platform to analyse and predict socio-economic behaviours based on demographic changes. As a part of their ongoing project, we are collaborating with them to test our methodology for simulating South Korean population dynamics.

Both proposed case studies suit well the high-degree complexity of common agent-based models because the evolution of the population is affected by a large number of interrelated and dynamics factors: gender, age, economic level, family, matchmaking and reproduction patterns, life cycles and so on. Moreover, our agent-based approach is interesting for studying both phenomena since it provides a way to gain insight of the population at an individual-level. For example, it could help to understanding ageing processes on Western societies and the dynamics that can be leading to the current trend. One of the challenges we may encounter is that the structure of households in Gambian society is significantly different from the western society. This case study will help us improve the design and the applicability of the parallel simulation software. Moreover, with the South Korean study we are interested to see if our methodology could be useful on the forecast of socio-demographic dynamics.

In order to apply agent-based models to demographic simulation, our approaches are to:

1. Integrate the individual dynamics' modelling to regional areas where the population is distributed;

2. Develop a set of methods to present the simulation output in ways that are suitable for later analysis. Simulation results will be able to show the population projection at a macro and micro scale at any time of the simulation;

3. Provide a set of different instruments to specify the behaviour of individuals. Users should be capable of represent human behaviour according to the statistical information available for them (quantitative data) or as rule-based knowledge (qualitative data);

4. Develop a graphical user interface to provide a way to model users' own simulation scenarios. Since most policy modellers are not familiar with programming this user interface should be easy to use. The code should be

automatically generated, offering a procedure to profit High Performance Computing resources to run large simulations. This system will approach final users to current parallel computing techniques.

As a result of the objectives' achievement, this research will boost research on several fields making the following contributions:

1. On the social sciences side, this work contributes to demographics applied to policy analysis with a new framework to improve population dynamics simulation studies.

2. On the operations research side, it advances discrete simulation field by employing agent-based modelling to simulate demographics, with an event-driven approach.

3. On the computer science side, it contributes to the research done in High-Performance Computing, by dealing with scalability and performance issues of social agent-based simulations.

1.4 Structure of this thesis

This thesis is structured as follows:

Chapter 2 presents the state-of-art by giving an overview on social modelling. It describes the modelling process for social systems, discusses some approaches to social modelling and introduces agent-based modelling. It also presents the most commonly used tools for social simulation and discusses further research needs.

Chapter 3 introduces the second state-of-art on parallel simulation and High Performance Computing Systems. It highlights the advantages and challenges of parallel simulation. Then it describes the application of parallel simulation to study large-scale social systems. It reviews some of the tools that have been developed in this area of research so far. Moreover, it presents some applications.

The core of our work is presented in chapters 4, 5, and 6. Chapter 4 describes our approach for large-scale demographic simulation. It details our model for population dynamics, shows the design of tee proposed framework, and describes the implementation of agents and processes. The tool is able to harness the available compute power efficiently. Its design requires a good understanding of factors in the policy model that may affect the run-time performance of the simulator. In this chapter, we identify factors from a demographic model that affect the simulation performance.

The tool should be easy to use by model developers (non-computer experts). Therefore, we present a graphical user interface for our tool in chapter 4. We explain in detail its components and the different approaches to using both qualitative and quantitative data to specify demographic processes.

Chapter 5 addresses the performance of simulating large-scale population dynamics. It presents the effect of varying population size and migrations both on the execution time and the scalability. It also studies the response of the tool to different conditions such as a homogeneous environment, an unbalanced population size, dealing with heterogeneous processing elements or heterogeneous communication latency.

In chapter 6, we review the literature on validation and verification of simulation models and use white-box validation to give more credibility to our tool.

Chapter 7 shows our two case studies. First, we simulated the demographic evolution of Gambian population in Spain from 2001 to 2011, and its migration movements. We are particularly interested in understanding migration patterns of migration and seeing how household structure changes. Second, we use our approach to model South Korean population dynamics and project them into the future. Although our methodology is not meant for forecasting, we are interested in analysing the results to see whether we are able to reproduce South Korean patterns of population dynamics.

Finally, in chapter 8 we provide a summary of conclusions, our contributions and some ideas for future research.

Chapter 2

Modeling and simulation for population dynamics

Computer modelling and complex systems simulation have dominated the scientific debate over the last decade, providing important outcomes in biology, geology and life sciences, and resulting in the birth of entirely new disciplines (e.g. bioinformatics, geoinformatics, health informatics, etc.). In the social sciences, the number of groups currently developing research programs in this direction is increasing. The results are extremely promising since simulation technologies have the potential to become an essential tool in the field (Gilbert, 2008).

However, some social scientists are sceptical about the idea of reproducing *in silico* population dynamics, because of the perceived complexity of social structures. This scepticism is understandable given the low number of projects that use this approach and the lack of experience with methodologies of this kind. Nevertheless, the research done in complexity science during recent years shows how computer simulation can be applied to this field. Indeed computer models provide an artificial environment to test hypothesis, to reproduce theories and to analyse scenarios. Furthermore, they provide a way to study problems where the analytical solution can not be found. Thus, they can be seen as a natural complement to classical research methods (Helbing, 2012).

In this chapter, we show the nature of social simulation and why it is becoming so popular. In particular, we first describe the need of modelling in social science and humanities. Later, we revise some approaches to social modelling and compare their limitations with respect to agent-based modelling. Moreover, we discuss the advantages and drawbacks of agent-based modelling. Then, we revise some applications of agent-based simulation in the social sciences and analyse its acceptance in policy analysis studies. Finally, we present some tools for agent-based modelling and simulation and see their further needs. Part of this chapter in the book *Formal Languages for Computer Simulation: Transdisciplinary Models and Applications* authored by Montañola Sales, Rubio-Campillo, Casanovas-Garcia, Cela-Espín, and Kaplan-Marcusan (2014e) Copyright 2014, IGI Global, www.igi-global.com. Posted by permission of the publisher.

2.1 Modeling social systems

Social sciences and humanities are concern with the study of the human being and their world. Using methods of empirical data collection and scientific analysis, the social sciences study human behaviour and society in a variety of fields such as sociology, psychology, political science, economics or anthropology. The reason for having many specialised areas of research around human beings is that societal modelling is complex and can be studied from different approaches (Gilbert and Troitzsch, 2005). According to Rossiter, Noble, and Bell (2010) social systems are complex in three different ways:

- they are composed by many entities which interact between them with a high degree of interconnection that can introduce internal feedbacks,
- their structure and rules may vary over time, so they have limited accuracy which in turn makes validation difficult,
- they have limited available historic data to work.

So how could social researchers tackle their questions about change in social systems? As pointed by Kohler and van der Leeuw (2007) the fieldwork could be a place to start but it is not enough to answer all questions. Therefore, social scientists need to build models as possible explanations to contrast their theories and their data.

Computational modelling and simulation can handle systems with complex, dynamic, and interrelated parts, such as epidemics spread or the extinction of an ethnic group, which occur within a context constrained by many socio-economic factors. This methodology can also handle the emergence of social patterns from individual interactions. In that way, a part of computational modelling is to be capable of model a person and his/her social relationships as networks. In this section, we show the nature of social simulation and why it is becoming so popular. We will also discuss the modelling procedure and why simulation looks suitable for applications in social science and humanities.

2.1.1 A brief history of social science simulation

The study of the real world with simulation technologies in the social sciences started in the earlies 1960s with the advances and developments on computer (Troitzsch, 1997). In the beginning, the research on simulation focused mainly in

Discrete-Event Simulation and System Dynamics. These methodologies are characterized by taking advantage of the big calculus capacity of computers. System Dynamics approach, later described in 2.2.2, uses big systems of differential equations to plot variables trajectories over time (Hanneman, 1988). Sterman (2000) describes it as

> *"a powerful method to gain useful insight into situations of dynamics complexity and policy resistance"*

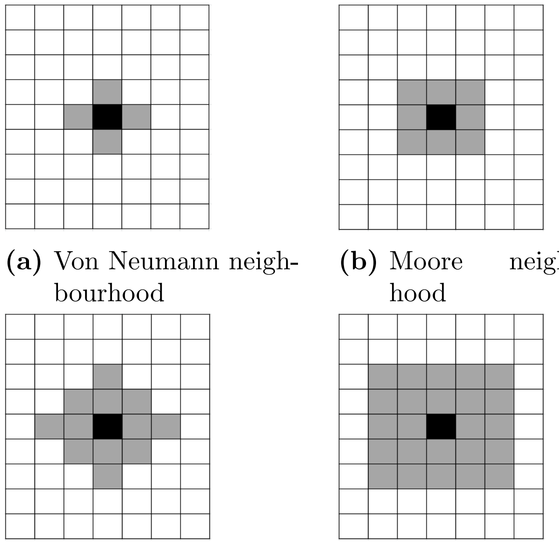
However, System Dynamics is restricted to the social analysis at macro-level. Moreover, it can only allow models that can be translated into equations.

In the early stages, simulation focused more on the prediction of social systems than on their understanding (Gilbert and Troitzsch, 2005). In the same period, another approach to model social behaviour appeared, known as microsimulation (Orcutt, 1957). Microsimulation, later described in 2.2.1, aims to model the evolution of population dynamics over time through specifying a random sampling process for each individual at every simulation time point. Although microsimulation has no pretensions to explanation but to predict as system dynamics, it is interesting to note that the unit of simulation is the individual. However, there is no attempt to model interactions between them. During many years, microsimulation was the only form of simulation that was widespread recognized by social scientists. Today it is still used in many countries for policy issues.

In the 1980s, advances came from mathematics and physics, especially those working in the artificial intelligence field. In these years, cellular automata (Von Neumann, 1966) started to be used to understand social interaction. Cellular automata are mathematical models that simulate dynamic systems that evolve in discreet steps. They consist of a grid where every square is known as *cell*. Each cell has concrete state in each moment of time and also a set of neighbours. In each step of time, the new state of a cell is calculated by a homogeneous transition function that considers the cell state and its neighbours states. In Figure 2.1, we can see different approaches to calculate the cell state. In some real-life applications, there have been some variations of the cellular automata definition such as non-uniform grid, asynchronous update of the cell states, and extension of the cell neighbourhood (Hwang, Garbey, Berceli, and Tran-Son-Tay, 2009).

Later, new advances in computing allowed to combine cellular automata with game theory in the social sciences. One of the most well known examples is the Life Game of Conway (Gardner, 1970) where with four simple rules a cellular automata is build to simulate the life of a complex organism society, including their interactions. At the same time, other types of simulations coming from stochastic processes arose, such as queue theories (Kreutzer, 1986) or multi-level simulation (Helbing, 1994).

In the 1980s, when the computer capacity increased and the Internet appeared, Artificial Intelligence researchers develop the distributed computing in the form of autonomous, independent entities able to interact, called *agents*. This event



(a) Von Neumann neighbourhood

(b) Moore neighbourhood

(c) Extended Von Neumann neighbourhood

(d) Extended Moore neighbourhood

Figure 2.1: Type of neighbours in a cellular automaton

later gave birth to a new technique which allows autonomous objective-driven movements in grids: multi-agent-based systems (Weiss, 1999). In the same period, artificial intelligence community developed *”machine learning”* techniques, systems with the ability to learn from experience, adding the new information to their knowledge and procedural skills (Michalski, Carbonell, and Mitchell, 1985).

A summary on the historical development of contemporary approaches to social simulation can be seen in Figure 2.2.

2.1.2 Why model?

The concept of modelling is widely extended. It comes from the natural observation of the world and the curiosity or need to reproduce it. As Epstein (2008) says

> “Anyone who ventures a projection, or imagines how a social dynamic—an epidemic, war, or migration—would unfold is running some model”

The challenge is to write it down, to turn it from *implicit*, where assumptions or data are hidden, to *explicit*. It does not matter if the model implies a mathematical formulation or any graphical representation. Models are approximations to reality for an intended used (Pidd, 2010). According to Pidd, a model can be graphically represented as a box with inputs and outputs. The box will be black or grey depending on the purpose of the model. If we know little about the

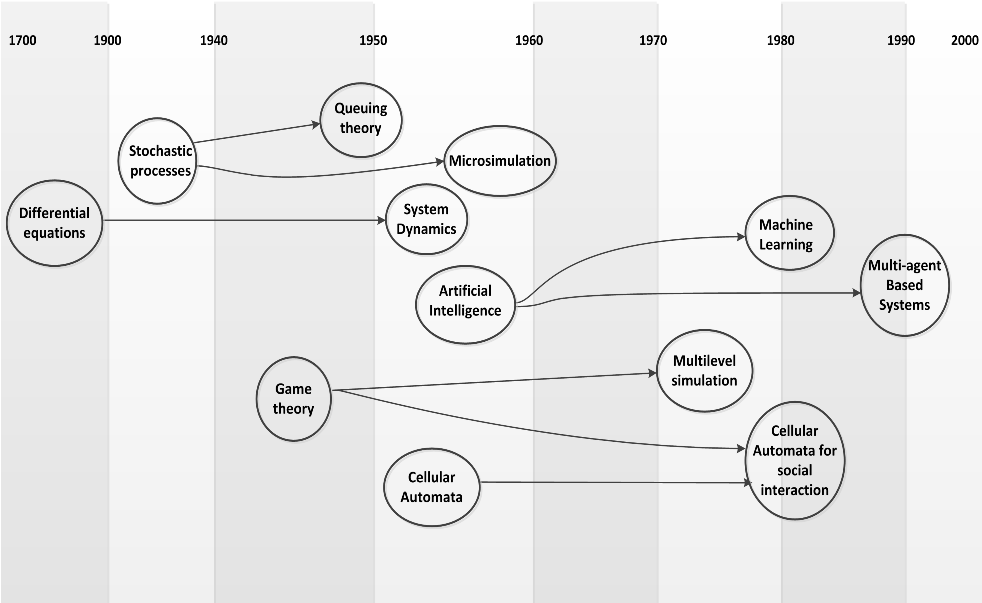


Figure 2.2: Historical approaches to social simulation (based on Gilbert and Troitzsch, 2005)

phenomena under study, the box will be black since the model will be analysed through its outputs under defined inputs. For example, when performing controlled experimentation in some areas of physical science. On the contrary, if the box is grey that means we have some knowledge of the model's interior processes. That is most often the case when investigating "what if" scenarios. Then, processes should be analysed when detecting some unexpected emergent behaviour, as a consequence of the internal dynamic interactions between the variables in the system.

In social sciences and humanities, common mathematical models are based on differential equations which are very useful to describe continuous systems or central systems, where physical laws drives the system dynamics rather than information processing (Parunak, Savit, and Riolo, 1998). However, equation-based models experience more difficulties in systems where the interest is on the interactions between discrete entities, as human being can be. Moreover, to model individuals as discrete autonomous entities looks more natural (Rubio, 2009).

Given the difficulties to apply differential equations, statistical techniques have commonly been used as a powerful analysis tool (Stewart, 1989). They are able to extract general patterns from a set of data that does not appear to have a regular behaviour. However, statistical analysis provides models that indicate tendencies in a sample of variables' values. Therefore the information we can generate is very limited, particularly in forecasting and hypothesis testing.

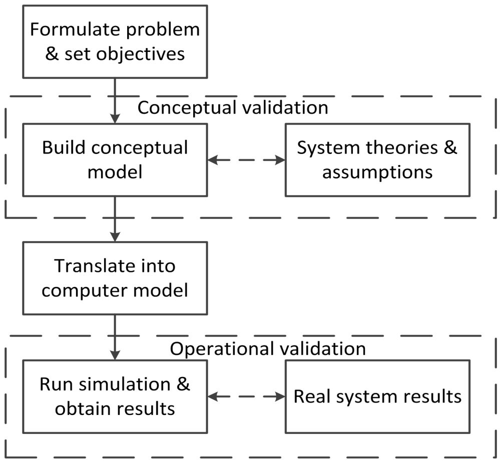


Figure 2.3: Conceptual modelling in a simplified simulation development process by Heath et al. (2009)

Simulation is especially suitable when we want to conduct experiments with a model in order to understand the behaviour of the system under study (Shannon, 1976). Representation of reality through simulation models are often closer to reality processes than other mathematical models (Lozares Colina, 2004). Simulation not only includes the construction of a model to study the system dynamics of interest, but also can generate new knowledge with an impact on the model formulation itself. For instance, we may need to refine the model when new hypothesis arise. The ultimate objective is to get closer to possible answers of the initially formulated questions about the real world. According to (Shannon, 1976), a simulation process should start with the definition of a problem, analysing the important entities which play a role and the relations between variables, followed by a model formulation. At this stage, it is important to decide the number of variables to take into account since simplicity and complexity should be balanced. The model should have as many variables as needed to answer the initial questions. However, it is possible to refine it in later stages.

2.1.3 How to model

It is not ventured to say that modelling is one of the key research processes. A model has impact in all aspects of the simulation study. But which is the process of modelling? Law (2007) describes in the following way. When trying to simulate the real world we talk about *systems* of interest, which are a

> "collection of entities that act and interact together toward the accomplishment of some logical end"

To study them, modellers start from a set of thoughts in the stake-holder's mind around a problem or theory. This set of ideas refer to the structure of the problem:

its objectives, the input and outputs of the system and its content (Robinson, 2008). In this process, some assumptions and simplifications are made. Robinson states that these assumptions help dealing with uncertainties or beliefs about the real world (the scope of the model), while simplifications help reducing the complexity of the model (its level of detail). All this components form the *conceptual model*, which gives us an insight of the behaviour of the system. The process of building a model from a real or proposed system is called *conceptual modelling*. In Figure 2.3, Heath et al. (2009) show how conceptual modelling is embedded in a simplified simulation development process.

Despite the importance of the conceptual model in a simulation study, there is no agreement in the definition of what a conceptual model is. Onggo (2010) suggests this lack might be due to the wide variety of conceptual model representations which have been proposed in the literature. What seems clear is that conceptual modelling is very close to the notion of *abstraction*, which is related to computer science and has originated many specification languages (Roussopoulos and Karagiannis, 2009). To make this abstraction process effective, an appropriate simplification of reality is needed (Pidd, 2003). That is, we need to set the boundaries of the real world portion we want to model in an appropriate way that can give an answer to the question we make. Inspite of this process of simplification, a model should be complex enough to answer the question raised (Banks, 1998). For example, a model that emulates a vehicle routing problem can answer questions on how a company should distribute its products in a given network. However, this model can not say how this distribution will impact on the current traffic of the network. If we were interested in that, we should enlarge our model to include the traffic flow to calculate how a given distribution of vehicles will affect it.

Thus, we need to find a balance between the real world and the conceptual model. If we directly consider the most complex system to do a studywe will encounter several problems, being the most important its credibility. How could we be certain that the added components in our model are not affecting the results? As Robinson (2008) states simple models have many advantages, such as they are faster, require fewer data, are more flexible and more importantly, if we better understand them we can better interpret their results. In fact, a good model design enhances the probability of simulation study success. Nevertheless, not in all research areas simplicity is seen as a positive value. For instance in the social sciences, van der Leeuw (2004) states archaeologists can not assume a simple behaviour until there is some evidence of it. Indeed, one should be careful to simplify certain natural processes since it presumes some assumptions about how they operate with the danger of missing important facets to explain it (Davies, Roderick, and Raftery, 2003). The tendency however is to build KISS (*Keep it simple, stupid*) models. This idea stems from Occam's razor: things should be kept as simple as possible and made as little more complex as explanation purposes demand (Axelrod, 1997). Applied to social simulation, KISS principle ideally seeks simple and abstract models that are general enough to be explanatory for multiple specific cases.

Furthermore, in social sciences and humanities there is a particular difficulty when doing conceptual modelling: sometimes data is not available or non-existent. In those cases, stakeholders need to proceed in other ways by filling out the abstract social process or mechanism sufficiently to create a working implementation. Therefore, they will focus more on the instantiation of the desired mechanism than in being faithful to the observation of the real world process (Yang and Gilbert, 2008).

One important property of the conceptual model is that it is not oriented to any software, so stakeholders free from implementation concerns. Thus being able to represent the behaviour of a problem which can be later solved in the preferred computational methodology. Separating the modelling and coding process allow modellers to focus on developing a more appropriate ("right") model to perform the study of interest (Robinson, 2008). This fact has the advantage of allowing all stakeholders to integrate in a simulation project (Roussopoulos and Karagiannis, 2009). That is the reason why it is possible to communicate the model between them, to discuss different points of views and to set common objectives. Therefore, a collaborative effort is needed since they might probably come from different domain knowledge and expertise (Chen, Theodoropoulos, Turner, Cai, Minson, and Zhang, 2008).

2.2 Approaches to social modelling

In previous sections, we have described the modelling process, and we have seen what it consists of. We have also outlined the simplification and abstraction processes of modelling, along with the difficulties we might find in the way. There are some alternatives in the style of methodology to model social systems. As pointed in Siebers, Macal, Garnett, Buxton, and Pidd (2010), people do not always consider different approaches to addressing problem requirements, but tend to use the ones that are more familiar to them. Sometimes, however, the techniques turn out to be inadequate, and they need to switch to another one. Therefore, it is important to know in advance the capabilities and drawbacks of different methodologies to simulate a social system. In this section we present and discuss four methodologies that are most usually employed for social simulation, particularly for demographics.

2.2.1 Microsimulation

One of the commonly used paradigms in demographic simulation is *microsimulation*. Initial studies in microsimulation date back to the work of Orcutt (1957). In this simulation paradigm, modellers have to specify a random sampling process for each individual at each simulation time point, to determine the state of each individual at the next simulation time. Microsimulation uses an initial population set, usually obtained through surveys, with detailed characteristics of the population under study such as household members' age, income, employment and so on. These parameters are intended to represent individual preferences and tendencies. Within the model each entity is defined by a unique identifier and a set of attributes, such as age, gender, economic status, marital status and so on. Then a set of rules is defined so each individual entity moves from one state to another (e.g. moving from single to married). Rules can be stochastic (subject to a probability, for example, the probability of getting married) or deterministic (for example paying taxes or going to primary school, in most western societies). With this initial population setting microsimulation makes predictions on how the sample will look like in the future.

At one extreme, the sampling process requires a simple random sampling. At another extreme, it may require a complex regression model. However, most microsimulation tools have been built to assist public policy analysis. Examples include LABORsim for policies related to labour supply in Italy (Leombruni and Richiardi, 2006) and Pensim2 for the British pension system (O'Donoghue and Redway, 2009). SOCSIM (Hammel, Mason, and Wachter, 1990) is one among the few generic micro-simulation tools for demography. Other works include Dahlen (2009), who developed of an open-source micro-simulation tool called LaMPsim that has been designed for labour market policy in Sweden. Also, Zinn, Gampe, Himmelspach, and Uhrmacher (2009) developed another generic micro-simulation tool called MIC-CORE. Bourguignon and Spadaro (2006) discusses the use of microsimulation for welfare analysis.

The main advantage of microsimulation is that initial population settings come from real data, not being based on randomness or hypothesis. This fact makes microsimulation robust against another kind of modelling techniques, since it is easier to understand or interpret the results of the simulation and make predictions of the future population stake (Gilbert, 2008). However, the ageing process of agents needs detailed transition matrices which are often difficult to obtain since it requires a large amount of data. Furthermore, neither there is interaction between agents, nor often space is taken into account. This fact is an important drawback in some case scenarios, such as modelling people behaviour in self-organizations or not taking into account the space when mobility has a role in the system.

2.2.2 System Dynamics

System dynamics is another commonly used modeling paradigm in developing demographic simulation models. Unlike microsimulation, system dynamics does not keep track changes in the state of each individual but focuses more on the population of individuals and the rates of individuals moving from one state to another. Due to its approach, it is difficult to capture the individual characteristics in system dynamics models since agents are not represented directly. It is commonly used to analyse the complex feedback systems and the mutual interactions in the system over time. It evaluates a set of differential equations that calculate the value of a variable at the next time step from the value of other variables.

System dynamics important works in the area of population dynamics include the World Dynamics (Forrester, 1971) and World3 population model (Meadows, Meadows, Randers, and William, 1972; Meadows, Meadows, and Randers, 2004). A demographic model in system dynamics is often used as a component in a policy model. For example, Ahmad and Billimek (2005) developed a system dynamics model to analyse policies to reduce the harmful effect of tobacco on population health. Key demographic components such as fertility rate, mortality rate and net migration are included in the model. Saysel, Barlas, and Yenigun (2002) developed a system dynamics model to analyse policies on various environmental issues such as water distribution management and agricultural pollution. The model took into account the mutual interactions between the environmental issues and the demography in the region.

2.2.3 Discrete event simulation

Discrete event simulation pictures a system in terms of states and events, basing the simulations on the events that take place in the system under study and looking at the effects these events have on the state of the system. Similarly to microsimulation, discrete event simulation keeps tracking the individuals from their arrival in the system (through births and migrations) until they leave the system (through deaths and migrations). However, discrete event simulation does not inspect each individual at each simulation time point. It inspects an individual only when the state of the individual changes.

Most discrete-event demographic simulation models are used in applications such as healthcare and epidemiology. For example, Rauner, Brailsford, and Flessa (2005) developed a discrete event simulation model to study the effectiveness of intervention programs to reduce the vertical HIV transmission. The model used demographic data to initialize and to project the population. The model took into account the demographic information for activities such as being tested for HIV and receiving treatment. Roderick, Davies, Jones, Feest, Smith, and Farrington (2004) built a discrete event simulation model to estimate the future demand of renal replacement therapy in England, that took into account the demographic population changes. In the area of epidemiology, a number of researchers have attempted to build large-scale simulation models. The main objective is to understand the spread of global epidemics that may include analysis of a large number of individuals. Eubank (2002) and Rao and Chernyakhovsky (2008) showed that parallel discrete event simulation was needed for large-scale epidemiological models. They developed specialized parallel simulation tools for epidemiological models. Recently, there has been a number of works that develop tools for large-scale artificial society simulation (Guo et al. 2012, Bin and Guo 2012, Zia et al. 2012).

2.2.4 Agent-based models

Agent-based models (ABM) also called individual-based models come from the area of Artificial Intelligence, which tries to create artificial systems that think and act like human beings (Russell and Norvig, 2010). In those terms, the unit of study is the *agent*. Although there is no standard definition of agent, it is commonly accepted the one from Huhns and Singh (1997) which defines an agent as

> *"a self-contained program that can control its actions, based on its perceptions of the environment"*

The reason for this discrepancy is the differences found on applications within disciplines, leading to distinct visions of which characteristics an agent should have (Crooks and Castle, 2006). Despite this disagreement, most authors consider agents should satisfy the following properties:

- *Autonomy*, understand as the capability to operate without the direct control of humans or other agents
- *Interactive*, as the social capacity to interact with their environment and/or with other agents besides perceiving their closest local environment objectives
- Having the potential to add characteristics as *adaptation*, *learning*, *complex planning* and *language*

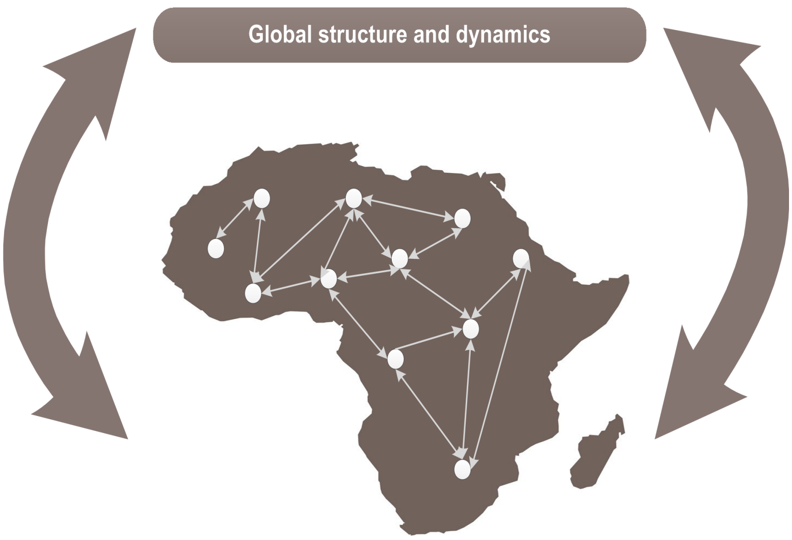


Figure 2.4: The structure of a typical agent-based model

Agent-based models are formed by a set of autonomous agents that interact with their environment through a set of internal rules to achieve their objectives. Interactions with other agents or entities within the system are specified in the model. Therefore, agent-based simulation is the process of designing an agent-based model of a real system and performing experiments with it, either to understand the behaviour of the system or to evaluate a set of strategies for its operation (Shannon, 1976). The environment is a space in which agents take action and constraints potential interactions (see Figure 2.4). As Crooks and Castle (2006) state, agents may be spatially explicit or spatially implicit, wether they have a location in geometrical space or it is irrelevant. In spatially explicit models landscape can be represented as

- a passive space where agents interact or
- spatial landscape represents an agent itself, for example on a land with naturally growing resources.

When the systems evolves through time, interactions of agents occur in certain order. Since agents behave individually in parallel the order of interaction should not be sequential but randomized.

2.2.5 Difference between agent-based simulation and other approaches

Previously we described different approaches commonly used to model and simulate social systems. Now, we want to point at the differences between them and highlight how agent-based modelling differs from other methodologies. However, our purpose is not to demonstrate agent-based modelling is better than the other approaches but to revise the kind of questions it is willing to answer in different application domains with respect to other methodologies. Table 2.1 shows a summary of differences between the previous modelling approaches described. We want to mark out the context where agent-based simulation is meant to. The modeller's challenge is to pick the most appropriate approach to the problem. Thus facilitating the modelling process plus obtaining the expected type of results.

After the description of microsimulation in 2.2.1, we can see this technique has two main limitations (Davidsson, 2001). On one hand, it models individuals' behaviour in terms of probabilities which are supposed to match with preferences, plans and decisions. On the other hand, it simulates people at the individual level, not taking into account interactions with others. Moreover, microsimulation requires realistic micro-data so it can be difficult to apply when appropriate data are not available (Wu and Birkin, 2012). On the contrary, agent-based simulation can perfectly cope with those limitations and gain better results.

Agent-based modelling is able to express the complexity of a real-world system since agents can reproduce individuals' autonomous behaviours and the interactions between them, responding to the stimulus of the environment. Therefore, it

	Microsimulation	**System Dynamics**	**Discrete event simulation**	**Agent-based simulation**
model perspective	microscopic	macroscopic	microscopic	microscopic
modelling approach	bottom-up	top-down	top-down	bottom-up
what is modelled	individual entities	flows	processes modelling the system in detail, not the entities	individual entities and the interactions between them
interval between time steps	constant	constant	when an event occurs	not typically constant, driven by interactions or internal changes
works with flows	models the flow of entities throughout a system	models stocks and flows	models the flow of entities throughout a system	the flow of entities is not modelled, models the individuals which shape macro behaviour
how the simulation deals with randomness	stochastic, need to run multiple times to gain full understanding	deterministic, produces the same results run after run	stochastic, need to run multiple times to gain full understanding	stochastic, need to run multiple times to gain full understanding
statistical detail obtained	general statistics and aggregate statistics of micro behaviour	general statistics and effective rates, cummulative account	general stats and event tracking	general stats and individual tracking

Table 2.1: Comparing different approaches of social modelling

seems more natural than traditional modelling approaches such as discrete-event simulation or system dynamics. According to Bonabeau (2002):

> *"Whether one is attempting to describe a traffic jam, the stock market, voters, or how an organization works, agent-based modelling makes the model seem closer to reality"*

Siebers et al. (2010) illustrate this with an example. During many years, in epidemics system dynamics have been used to understand the spread of infectious diseases, especially using models know as Susceptible-Infected Recovered (SIR). SIR models have been very useful to guide policy decision making. However, they are unable to include human behavioural aspects while we know now affect epidemics dynamics. Thus, agent-based simulation seems a more suitable approach, since it is prepared to include human behaviour on the core of the methodology.

Furthermore, both Bonabeau (2002) and Siebers et al. (2010) state a good candidate problem for the use of agent-based models should be placed in any of these situations:

- when we want to model the behaviour of heterogeneous constituent units' activities of the system, especially when aggregate transition rates can not clearly define it
- when there is interaction between individual entities, particulary when there is some complex structure such as social networks
- when we want to consider a space and how individuals move across it
- when we want to include in the model how individuals adapt or learn
- when we want to reproduce strategic behaviour and anticipate future agents decisions
- when cooperation or organization is included in the model
- when there is structural change as a result of the model
- when we can obtain future conditions that can not be predict from the past

In summary, agent-based modelling looks appropriate for modelling social systems when describing them with equations becomes intractable and when it is very important to validate and calibrate them to ensure they are reproducing what we want to. Systems that can be modelled centrally, where dynamics can be described by physical laws rather than information processing, are most naturally modelled with equation-based approaches (Parunak et al., 1998).

Moreover, compared to other methodologies agent-based simulation allows extensibility, mainly because that in terms of software it follows the object-oriented software paradigm. As an example, Siebers et al. (2010) mention a case where agent-based modelling was used when system dynamics became too complex to deal with. They also showed how agent-based simulation provided a way to understand the system as it grew in complexity. In fact, an agent can be considered a self-directed object with the capability of autonomously choosing actions based on the agent's situation. Therefore classes are naturally used as methods to represent agents and agent behaviours. This fact implies that models can be extended to add more agents, behaviours and components of real-world applications.

Instead, discrete-event simulation is useful when we want to describe a system in terms of its entities and processes, rather than activities, and to represent uncertainty through stochastic distributions (Siebers et al., 2010). For example, discrete event simulation is suitable for queueing systems, very common on manufacturing and service industries. Seldom entities change their state or learn on external conditions, thus having a low level of autonomy. The main difference between discrete event and agent-based simulation lies on the nature of entities driving the system (process-centric or individual-driven), which according to Siebers et al. is not often clear in Operations Research. In both approaches, the simulation engine jumps from one event to another. However, agent-based modelling has a disadvantage compared to discrete-event simulation: the approach typically uses more resources in terms of computation and communications thus resulting in slower simulations (Davidsson, 2001).

2.3 Agent-based modelling in social sciences

Often in the community of social simulation the term agent-based simulation and *multi-agent systems* are used equivalently. However, they refer to different approaches of research. The term multi-agent systems comes from the Artificial Intelligence area and focus more on technical issues related to agent systems than epistemological ones. In that way, multi-agent system methodologies usually tackle issues related to the formalisation of modelling process; its implementation and some issues concerning software engineering methodologies (architecture, design, efficiency and so on). In contrast, agent-based models tend to use more social theory than multi-agent systems, where it is usually not so critical. As Conte, Gilbert, and Sichman (1998) states

> *"If the multi-agent systems field can be characterised as the study of societies of artificial autonomous agents, agent-based social simulation can be defined as the study of artificial societies of autonomous agents"*

When building agent-based models, one of the first decisions modellers need to take is to choose between modelling at the individual level or at the aggregate. The election should depend on the purpose of the model. For example, if we want to understand the dynamics of crowds we might need to look at the factors that drive the behaviour of every individual from a psychological perspective. In that case, an agent-based model at the individual level looks more promising, such Madhavan, Papelis, Kady, and Moya (2009) propose. However, if the object of interest is to deep on the density evolution of a crowd, either a continuous-field model or aggregates of individuals could be possible (Goodchild, 2005). The size of the crowd could change our decision since the more variables intervene in a model, the more difficulties researchers could encounter to understand them. Setting constraints is essential for agent-based modelling, since we can not produce an unlimited number of models to generate an unlimited number of scenarios for social dynamics (Moss, 2001).

Simple observation of real world leads us to the idea of patterns. They are usually the result of the interactions of small pieces that somehow combine in not so expected ways to create a large-scale pattern (Wilensky, 2002). These

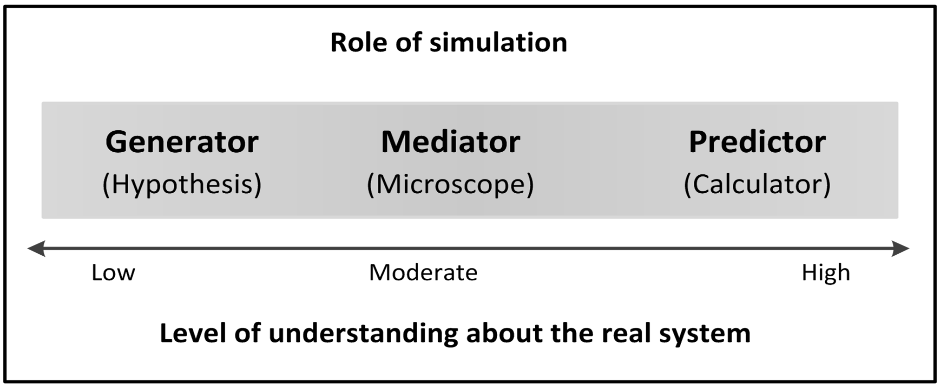


Figure 2.5: Roles of simulation use (from Heath et al., 2009)

patterns or *phenomena emergence* arise out of the interaction of numerous individuals (agents) at the micro-level (Bonabeau, 2002). An example of emergent behaviour could be a flock of birds winging acting complete synchronized while there is no leader to follow. Agent-based simulation can be used to detect patterns of behaviour under the hypothesis that those can emerge from the addition of individual decisions in a social model. Moreover, this property of emergence is especially interesting in situations where the modeller suspects dependent and independent variables alternate their role intermittently, which is very common in social sciences (Menéndez and Collado, 2007).

Which are the ways were agent-based simulation is useful for studying population dynamics? Based on the modeller level of understanding over the system under study, Heath et al. (2009) distinguish three types of purposes for simulation, shown in Figure 2.5. When the level of understanding is high the simulation model can be used as a *predictor*, producing precise forecasts of future behaviour under well defined conditions. When that level of understanding is in a medium stage, the simulator turns to a *mediator* providing insight into the system without offering a complete representation of its behaviour. Mediator simulation models can be used to test theories, and their results can be used to improve the simulation. Finally when low information about the model is available the use of simulation is as a *generator*, producing hypotheses and theories about the system behaviour. In any of those cases, social scientists agree researchers should define the concrete purpose of a proposed model.

Concretely in social sciences, Menéndez and Collado (2007) state there are two goals of simulation. First, to verify if a social model is coherent and adjusts adequately. In fact, when data is non-existent the parameter space is explored to find the optimal adjust for unknown variables. Once the simulation model behaves as expected new hypothesis can be introduced, observing if the model reflects the real behaviour. This property is especially useful when we have qualitative data and we do not know how to quantify it. With trial and error it is possible to modify parameters until observing the behaviour of interest. Through those parameters we can find out their relationships and mutual influence. Second, to virtually defy big limitations on experimenting in these areas, since some hypothesis can not be tested otherwise by recreating empirical situations. Therefore simulation can improve explanatory or even predictable capacities of models. However, if the knowledge in models is poor so the results of the simulation will be, resulting simulation not useful at all.

2.3.1 Advantages and limitations of agent-based models

Moss (2001) points out that the agent-based simulation can be useful in social sciences for several purposes:

- to restate and assess existing theories, which are often sociological and anthropological
- to take advantage of sociological theories and concepts to inform simulation models
- as a formalised description of a system
- to analyse different scenarios
- to help in policy analysis and formation

Beyond these interests, he states the foundational purpose agent-based social simulation is to develop a general social theory. Axtell, Axelrod, Epstein, and Cohen (1996) tried to figure out how general individual models might be by aligning some of them. However, in Moss opinion, this kind of models are far from being general. Simulation can provide insights and hypotheses on population dynamics, but theories on social behaviour are more relevant in the field (Bankes, 2002). That is the reason Moss suggests there is no universally accepted theory of the social process. Moreover, this fact explains why there has been very little use of agent-based models to recommend public policy.

Another reason for the importance of agent-based simulation in the social sciences and humanities is its naturalness as representational formalism (Bankes, 2002). Agent-based methodology provides a way to express the vast amount of data and knowledge about social agents characteristics, including their behaviour, motivations, and relationships with other individuals or institutions. Behaviours and interactions can be formalized by equations or through decision rules (if-then kind or logical operations), giving more flexibility to the modelling approach (Helbing, 2012). Moreover, agent-based simulation makes the construction of heterogeneous agents and their rules flexible enough, which turns out to be particularly helpful when there is a knowledge gap or data are unavailable (Axtell, 2000; Epstein, 1999). However, there are some areas in the field that need further attention such as the calibration of models to data or which methodology should be used to answer specific questions.

Finally, agent-based simulation is flexible since agents can be defined within any given system environment, and move on it freely (Castle and Crooks, 2006). In that way, it is possible to define geospatial reference of agents through variables and parameters of the simulation. Moreover, agent interactions can be governed by space and networks, or a combination of the two, which will be more complicated to explain by mathematical formulation (Axtell, 2000).

Besides agent specification, a key advantage of agent-based modelling is the ability to combine it with another kind of models (Helbing, 2012). For example, modellers may represent the environment by a continuous field and couple it with an agent-based model of the population behaviour. Example can be found in scenarios of people evacuation when a poisonous gas spreads (Epstein, 2009; Parker and Epstein, 2011).

Despite those advantages, agent-based approach also has some limitations. In sociology, some researchers find difficulties on applying agent-based models due to the lack of knowledge in certain social phenomena. This lack might be not only due to empirical evidence, but also to limitations on sociological theory. An example can be found in Menéndez and Collado (2007), where the authors try to study religious evolution in Spanish society in the end of last century. As a conclusion of their study, they point out at the need of finding more formal and documented models able to explain systematically social processes.

Another disadvantage of using agent-based models is the lack of full access to data relating to the phenomenon of interest, because the target of interest is not easy to access, or simply data is non-existent. There is also a limit on the size of parameter space that can be checked for robustness when conducting agent-based simulation. If we increase the number variables we not only increase computational needs, but also we compromise the credibility of the model which is very difficult to validate against real data (Castle and Crooks, 2006).

2.3.2 Some applications in social sciences

Among numerous applications of agent-based simulation to the social science, Schelling (1971) is credited to develop the first social agent-based model. In his work, he reproduced population dynamics in terms of segregation patterns. His work showed these patterns can emerge from migratory movements of two different culture types of households, which were quite tolerant. Later some other social models using the agent-based approached arose, such as Sugarscape (Epstein and Axtell, 1996). Despite the simple behaviour of its agents, Sugarscape results illustrated a variety of features of societies, including the emergence of social networks, trade and markets, cultural differentiation, and evolution. Another icon model of the agent-based modelling community is the Artificial Anasazi model (Dean, Gumerman, Epstein, Axtell, Swedlund, Parker, and McCarroll, 2000; Epstein and Gang, 2006) which describes the population dynamics in the Long House Valley in

Arizona between 800 and 1350. The model helped to prove that simulation could reproduce settlements archaeological records on the occupation of the Anasazi in Long House Valley with simple household rules on choosing locations for farms. Moreover, the model showed that the abandonment of the valley around 1300 can not be explained only by environmental variations.

In this section, we do not pretend to revise the full range of agent-based simulation applications but to point out some examples of interest. We can find examples in very different areas.

Archaeology and anthropology

Archaeologists and anthropologists use agent-based simulation of ancient civilizations to help explain their growth and decline, based on archaeological data. An example of that is Villatoro and Sabater-Mir (2008) who studied the Yamana indigenous of Tierra del Fuego (Patagonia) to deep in the factors that lead this hunter-gatherer society to extinction. In their study, they found that despite living in very hostile geographical conditions Yamana had a strong organisation and a set of norms that made possible a high interaction between different groups. For that, they used agent-based simulation where agents had a set of simple logical rules. More recently, Heckbert (2013) used an agent-based model to represent the ancient Maya socio-ecological system. In their study, the authors identify the variables of Maya society resilience and vulnerability such as climate variability, soil degradation, deforestation, demographic pressure, and the physical configuration of trade networks.

Economy

In economics, we can find numerous applications of agent-based simulation. An early example on evolutionary trade network formation among strategically interacting buyers, sellers, and dealers networks is from Epstein and Axtell (1996). Another example to reproduce dynamics in queues, such those of costumers in check-in desks, banks or airports can be found in Gilbert and Troitzsch (2005). Also, economical factors are considered in the agent-based model of Balbi and Giupponi (2010) of adaptation to climate change. Tesfatsion (2002, 2006) gives some more example of applications of agent-based modelling and simulation to economic systems.

Sociology

Sociologists are also working on agent-based modelling. Cognitive science is starting to extend the idea of artificial agents to social settings (Bedau, 2003). In that sense, Gratch and Marsella (2001) use agent-based simulation to study the influence of emotion and cognition on social behaviour. Menéndez and Collado (2007) studied religious change in Spain by the end of the last century and how it which might be tied to values change.

Applied intelligent agents

If we we talk about more intelligent agents, for instance orientating them rationally based on belief, desires and intentions aligned with psychological theory, we can find applications in traffic simulation where drivers take their decisions based on their perceptions of traffic flow and characteristics (El Hadouaj, Drogoul, and Espié, 2001). In the area of emergencies, we find the study on fire emergency in an airport where agents need to decide their way to escape by (Burmeister, Haddadi, and Mattilys, 1997). However, as Menéndez and Collado (2007) states sometimes it is not possible to precisely define desires and intentions which move human beings in the society to simulate them.

Policy analysis

There are not so many applications of agent-based modelling which actually have been used to support policy decision making. Most of the works done so far correspond to industrial applications that have contributed to guiding policy decisions. For example in the electrical sector, Chappin, Dijkema, van Dam, and Lukszo (2007) built an agent-based model to express the behaviour of electricity producers. The work has contributed significantly to the still ongoing debate to fix the EU emissions trading scheme. Veselka, Boyd, Conzelmann, Koritarov, Macal, North, Schoepfle, and Thimmapuram (2002) also developed an agent-based modelling simulator to model electric power deregulation for the State of in Illinois in the US. Its results provided an objective assessment of the need for real-time electricity pricing information to consumers and transparent market monitors in the absence of regulating body. The tool is currently being used for electric power applications for the US Department of Energy. Zhao, Mazhari, Celik, and Son (2011) proposed an hybrid model with agent-based simulation and system dynamics to study solar power in residential areas at two different regions in the US. They developed a decision support tool to analyse the effectiveness of various policies of establishing distributed photovoltaic systems. In the mining sector, Andriamasinoro and Angel (2012) applied agent-based modelling to forecast industrial mining in Burkina Faso. The approach allowed the authors to deep in the relations between migration and particular types of environments and governance scenarios. Their work is being considered by authorities of Burkinabe Users' Committee. In the energy market, Smajgl and Bohensky (2013) performed a study with agent-based modelling for the Indonesian Government showing the impact of energy price changes on poverty, deforestation and various other natural resources in East Kalimantan, Indonesia.

Other applications to policy include the work done by Christensson, Woodcock, Hitchins, and Cobb (2004) who used a strategy development system using agent-based models for military actions and society response. Specifically, the work focussed on command and control and crisis management procedures involving effective collaborative environments. The project was proposed by Swedish

Armed Forces and later used by US Joint Forces Command experimentation at the MultiNational Exercise version 4. Another example on emergency management is shown in Dugdale, Pavard, and Soubie (2000). The study used an agent-based model to help redesigning an emergency call center, making it more efficient in terms of use of resources. In the area of park and visitors management is interesting to see the work of Itami, Raulings, MacLaren, Hirst, Gimblett, Zanon, and Chladek (2003) who used graphical interface systems and agent-based simulation to simulate human interaction in Port Campbell National Park and Bay of Islands Coastal Reserve, Australia. The tool help on managing increased number of visitors in these parks to improve their experience's satisfaction. With the same idea, although in the private sector, Cheng, Lin, Du, Lau, and Varakantham (2013) presented an agent-based simulation of visitors behaviour in an amusement park in Singapore.

In land-use problems, the work of Brady, Sahrbacher, Kellermann, and Happe (2012) helped to evaluate the potential impacts of three alternative frameworks of reform on agricultural landscapes for the European Union Common Agricultural Policy. The model was calibrated with real data of the counties of Jönköping and Västerbotten in southern and northern Sweden. Agent-based modelling was useful for capturing the heterogeneity of farms and the landscape in regions, as well as modelling farm-agent behaviour and interactions in an optimization framework. In water management and use, another application of agent-based simulation can be found in the study of Gurung, Bousquet, and Trébuil (2006), who used agent-based model to facilitate water management negotiations in Bhutan. The methodology showed its usefulness for mediation purpose, helping to solve a conflict on sharing water resources by creating an institution for collective watershed management. Barthelemy, Moss, Downing, and Rouchier (2001) also used agent-based simulation to improve water demand management in the south of England during a period of climate change. In the project, stakeholders were involved in the model specification and validation process. The model captured the weather system, the hydrological system, the behaviour of households and the behaviour of policy agencies.

2.3.3 Acceptance of agent-based models in policy studies

Up to now we have seen the potentials of agent-based modelling in the social sciences. The agent-based methodological approach is not meant to substitute other mathematical techniques such as equation-based modelling or system dynamics, neither is going to supply them. It just gives a complementary view to study a problem, and it is especially suitable for obtaining generative explanations of complex interactions of human systems through the unforeseen interactions of multiple agents (Epstein, 2006). These characteristics point at the usefulness to guide decision making process in policy studies.

However, not all social research areas use agent-based modelling for proving theories, testing hypothesis or performing case studies. The reason behind this lack of use is the acceptance this type of models have in the social studies community, especially among policy modellers. But what are the reasons behind the resistance of adopting agent-based methodology? Despite its limitations, a pure mathematical-based approach has today more acceptance for policy makers than agent-based models (Andriamasinoro and Angel, 2012). The reason might be that equation-based models are perceived as more feasible than agent-based models.

Moreover, there are also some difficulties that might contribute to the low acceptance of agent-based modelling in policy analysis:

- agent-based modelling is complicated for those who are not skilled in programming (Hamill, 2010)
- agent-based community is quite close, thus having difficulties to reach some application areas and to convince people of its merits (Weyns, Helleboogh, and Holvoet, 2009)
- the need to justify an agent-based approach in sociology, which is not generally accepted (Squazzoni, 2010)
- the need of finding more formal and documented models which should be able to explain systematically social processes (Menéndez and Collado, 2007)
- in some research areas such as archaeology, modelling is understood as including all the details of the system under study. However, this leads to a complex simulation and represents big difficulties when attempting validation (Farmer and Foley, 2009)
- the time that takes to design an application scenario and to justify the scale of time for an agent to behave (days, weeks, months...) (Andriamasinoro and Angel, 2012).

Although it might appear logical that science should inform, affect policy, in the United States today policy-makers do not necessarily want this to happen (Baird, 2013). Actually, the bridge between academia and policy is narrow, and it is often more difficult than it seems to translate theories and findings to policy.

Given this resistance, the challenge is then to find a way to overcome it. Hamill (2010) made interesting suggestions in this respect. First, agent-based models could be more appealing if the benefit in terms of outputs were clear. In that way, the method should provide the kind of outputs both politicians and analysts expect, which differ in their level of detail. Second, there is a need to convince analysts about the benefits of this method compare to others. And finally, the value of modelling needs to be demonstrated by offering training projects (a good example could be the mentioned example in Section 2.3.2 by Veselka et al. (2002)). Beyond these suggestions, Buchanan (2009) points out at the systematic collaboration between multidisciplinary scientists, from computer science and thematic entities. Thematic entities might be for instance sociologists, anthropologists or human geographers in the context of an study about urban segregation and inclusion in cities, or economists, socio-economists (Feitosa, Le, and Vlek, 2011), and geologists in Andriamasinoro and Angel (2012) work. Moreover, a coupled approach combining agent-based modelling and mathematics with the idea of getting closer to analysers might convince stakeholders of agent-based benefits. In their work, Andriamasinoro and Angel also saw that integrating agent-based models with spatial data in the form of maps will increase their confidence in the method.

2.4 Tools for modelling in social simulation

With an interest in developing agent-based modelling as a general technique to be applied to the study of societies, a numerous number of platforms have been designed to model social phenomena. In this section, we will briefly describe some of them and compare them in terms of software capacity, architecture and the type of applications they have been tested. The tools presented in this section are sequentially-based. We have chosen the software platforms we currently believe to be of particular relevance to scientific modelling and simulation of societies. We do not intend to review them fully but just to revise their most important characteristics to help the reader picture the world of agent-based simulation tools. A more completed list on software platforms for agent-based modelling and simulation can be found in Nikolai and Madey (2011) and `http://www2.econ.iastate.edu/tesfatsi/acecode.htm`.

Agent-based systems are very complex applications to program, implement and optimise, especially when the size of the population to simulate is large. Actually, we can find two different situations when modelling and simulating social behaviour. On one hand, there is a computer scientist or someone with advanced skills in mathematical programming that faces the challenge to realize some research in a social domain. Unless he/she has an adequate knowledge in the social model, he/she will experience difficulties to build a valid model to later explain and communicate it to the social community. On the other hand, we can find a social scientist who wants to take advantage of simulation techniques to experiment with his/her theories and knowledge. Unless he/she has the programming experience to develop and implement his/her ideas, he/she will need to find someone to do that. The ideal case, therefore, would be the situation where the social scientist could autonomously use a tool to define his/her models and afterwards run them. Similarly, when social scientists perform a multivariate analysis through a statistical tool without the need of knowing in deep the mathematical algorithms involved, they should be able to test their hypothesis or theories. Thus, the need to give a boost in developing tools to help research community advance in the social simulation field is justified.

As computer technology advanced, the scale and sophistication of the software available for users has increased. Software tool-kits might be difficult to handle, especially when they use object-oriented languages. In that sense, the use of object-oriented languages is very extended: 42% use Java as their primary programming language, 17% uses C++, 11% uses C, 8% uses a variant of Logo and the rest use a platform specific language that was designed to facilitate the modelling and simulation design (Nikolai and Madey, 2009). In fact, there is little consensus about the best general purpose programming language to use on simulation social behaviour (Gilbert and Bankes, 2002). Given the number of different programming languages that can be used, agent-based modelling packages tend to be hard to understand for users who may have little or no programming experience, as social scientists generally are the case.

Apart from that, each tool has its own non-intuitive terminology, so users should learn how to draw or write their models in each particular platform (Allan, 2010). For example, the term *"multi-agent system"*, which properly refers to a small system with heterogeneous agents that have artificial intelligence capabilities, it is in some cases used to refer to a large system of homogeneous agents (an agent-based system) (Nikolai and Madey, 2009). These slight differences on the concepts employment can mislead the user or create difficulties in work teams. This situation is due to the multidisciplinary nature of social agent-based modelling field, so easily there might be a conflict in the use of terms.

Depending on the user's background, the characteristics of the platform should be different. For instance, social scientist might be more concerned about how easily is the interaction with the interface to manage simulations and the degree of programming skills required, while computer scientists may consider if the tool is open source and its capability to be modified or extended for their own purposes. But no matter which type of user should the tool be meant for, they should have a good documentation. However, basic documentation is incomplete in general, with some exceptions (Allan, 2010). It mainly depends on how extended is the community using the tool and the community that supports its development.

Beyond that, there is no standard on how to specify agent-based models not only in the social science and humanities community, but also in the computer science, and specifically in artificial intelligence. That fact makes difficult to address the issue on how the agents' characteristics, the interactions between them and with the environment should be defined. As a consequence, current platforms address that issue in their proper way, according to the type of applications they are meant for, their characteristics, and what the tool designers think can be more convenient for their users. Consequently, it is not surprising to see that several tools have their own language which is used specifically for that tool-kit. Some of them, probably concerned about becoming simpler to learn and use, have support for visual programming (Nikolai and Madey, 2009).

In the following subsections, we will briefly describe the characteristics of four different tools which are most commonly used from the social scientist perspective. Those are Swarm, Mason, Netlogo and Repast. Although they were originally designed as general purpose tools, as an educational tool and as a specific tool for social scientific use respectively, all of them have been used for diverse kind of applications. Their differences on the primary domain of application might not only affect the user interaction with the tool-kit, but also the fact some of them have become more popular among some specific areas of social studies and humanities.

2.4.1 Swarm

Swarm is one of the oldest agent-based modelling tool-kits developed in Santa Fe Institute (Minar, Burkhart, Langton, and Askenazi, 1996). It is intended for general purpose applications, especially those related to artificial intelligence, to develop multi-agent models to simulate complex adaptive systems. In Swarm the design of models follows a schema of a hierarchy of *swarms*. A swarm is a group of agents with a schedule of actions that the agents execute. It is possible to design hierarchies of swarms whereby an agent can be composed of swarms of other agents in nested structures. When this happens, the agent behaviour at the higher level is defined by the emergent phenomena of the agents inside its swarm. Swarm separates the model from its observation. This property enables the model itself to remain unchanged if the observation code is modified.

Swarm simulations can be written mainly in Objective-C and some in Java, both object-oriented languages. Therefore, a knowledge on object-oriented programming is desirable. Swarm has a free source code form under the license GNU General Public License (GPL), which implies to make the source code for their entire model available to anyone who obtains a legitimate copy of the model's binary code. It runs in Windows, Linux, Mac and OS X operating systems. Swarm has no support for Geographical Information Systems although there is an extension of Swarm simulation libraries named *Kenge* that provides functionalities to create cellular automata similar to Geographical Information Systems (Box, 2002).

Swarm has been applied to a variety of domains. There are works in organisation management, for instance, the work of Lin and Pai (2000) to simulate changing business processes to adapt to new business environments. Also there are several contributions to economics (Luna and Stefansson, 2000) and supply chain (Strader, Lin, and Shaw, 1998). In more technological areas, we can find applications in mobile technology (Lingnau and Drobnik, 1999) and in social networks of open source software development (Madey, Freeh, and Tynan, 2002).

Traditionally, Swarm has been the most powerful and flexible simulation platform since it allows to implement very intricate and complicated social mechanisms (Allan, 2010). However, since the modeller needs to have some experience in Objective-C and possible Java, Swarm has a steep learning curve. That is the reason it has remained technically challenging for most social scientist to use. Consequently, it has not generated a broad-based community of practitioners in social science. However, given its history, Swarm contributed to making agent-

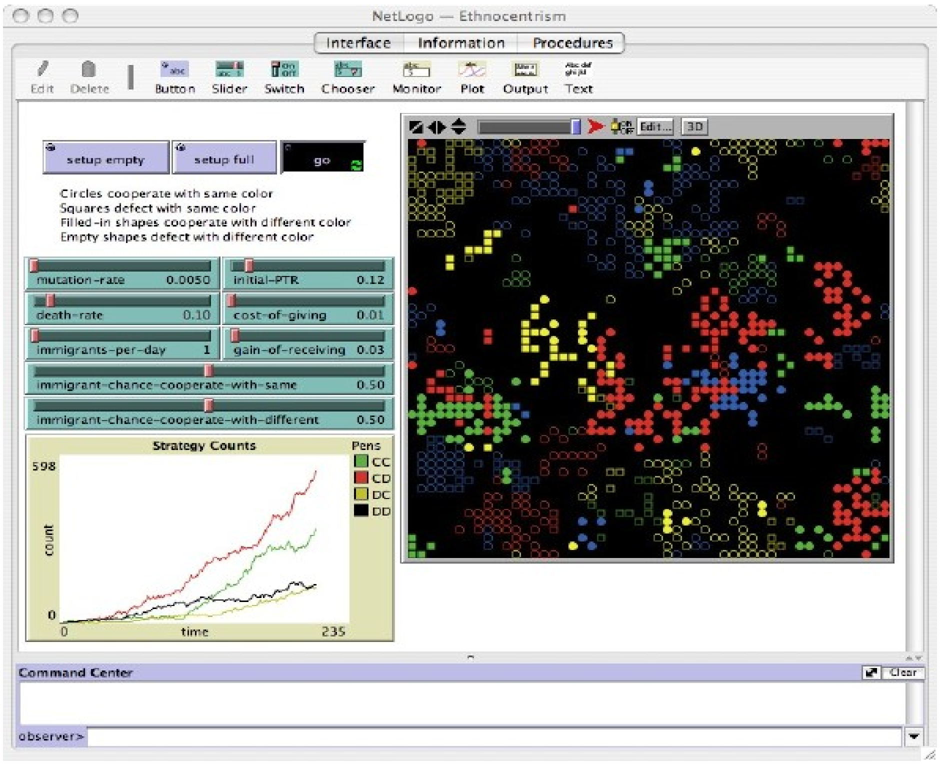


Figure 2.6: Graphical user interface of Netlogo of Axelrod and Hammond, 2003 model of the evolution of ethnocentrism (after Wilensky and Rand, 2007)

based modelling more visible to a large number of scientists (Janssen, Alessa, Barton, Bergin, and Lee, 2008).

2.4.2 Netlogo

Netlogo is an agent-based programming language and modelling environment for simulating complex phenomena. Netlogo derives from StarLogo (Resnick, 1994) and StarLogoT (`http://ccl.northwestern.edu/cm/starlogot/`), an environment for experiment complex dynamics in parallel environments in Macintosh operating system. Netlogo was designed to provide a basic laboratory for teaching complexity concepts. Therefore, it provides a graphical user interface to create models that control graphic agents that reside in a world in the form of a grid of cells, which can be monitored. Since the tool was designed for teaching, the environment and the language in Netlogo is meant to be simple enough ”*to have a low threshold for beginners*“ (Tisue and Wilensky, 2004). Therefore, Netlogo includes a large number of examples. In Tisue and Wilensky (2004) the reader can find some useful information about Netlogo, including the history of its origins, a tour of its interface, an introduction to the Netlogo language, and the acceptance of Netlogo in the research community.

Netlogo uses a modified version of the Logo programming language (Harvey, 1997). The use of this language is a different approach from other tool-kits such as Swarm and Repast use a general-purpose programming language, such as Java. Despite being free, Netlogo is not open source. Netlogo is written in Java so it can be run on all major platforms, requiring Java version 1.4 to run current version 2.0. According to Tisue and Wilensky (2004), the majority of users find Netlogo fast enough for most purposes, especially when running simple code on a large numbers of agents.

There are many applications of agent-based social simulation in Netlogo. Damaceanu (2008) used an agent-based computational model to show how global economy must focus on using renewable resources because this approach may increase the global wealth. Zhao and Li (2008) conducted a study on reputation evaluation mechanisms using Netlogo. Albiero, Fitzek, and Katz (2007) tested a power saving technique for mobile devices in a cooperative framework for the wireless domain. In Koper (2005) a learning network was modelled to study how social interactions might affect acquiring new learning competences. In Millington, Romero-Calcerrada, Wainwright, and Perry (2008) Netlogo was used to run an agent-based model of agricultural land-use decision-making to evaluate potential changes in wildfire risk for a Mediterranean landscape. Barcelo, Cuesta, Del Castillo, Galan, Mameli, Quesada, Santos, and Vila (2010) simulated the emergence of ethnicity and cultural differentiation in prehistoric hunter-gatherer groups in Patagonia (Argentina). Furthermore, Netlogo can be used to replicate models as Janssen (2009) did with the Artificial Anasazi model.

Netlogo is probably the simulation platform most widely used in Social Sciences and Humanities. The main reason is the smooth learning curve of the application and its extensive documentation, particularly if it is compared to other software packages. Unfortunately, Netlogo forces the user to create a model following the concepts and constraints defined by the program itself, and for this reason the researcher will run in trouble in case of creating a model that differs from the Netlogo approach. Moreover, the code of this software is closed, so the possibilities of adapting the platform to other uses are weak.

Actually, Netlogo is known as being by far the most professional platform in its appearance and documentation (Allan, 2010). It is the perfect platform for prototyping the first versions of a model and exploring toy models, and an excellent tool to improve the understanding of social scientists regarding the process of modelling.

2.4.3 Repast

The Recursive Porous Agent Simulation Toolkit (Repast) is a free, open source toolkit that was developed by a collaboration between the University of Chicago and the Argonne National Laboratory (Collier, 2001). It is under constant development and extension. Repast is a set of Java libraries that allow to build

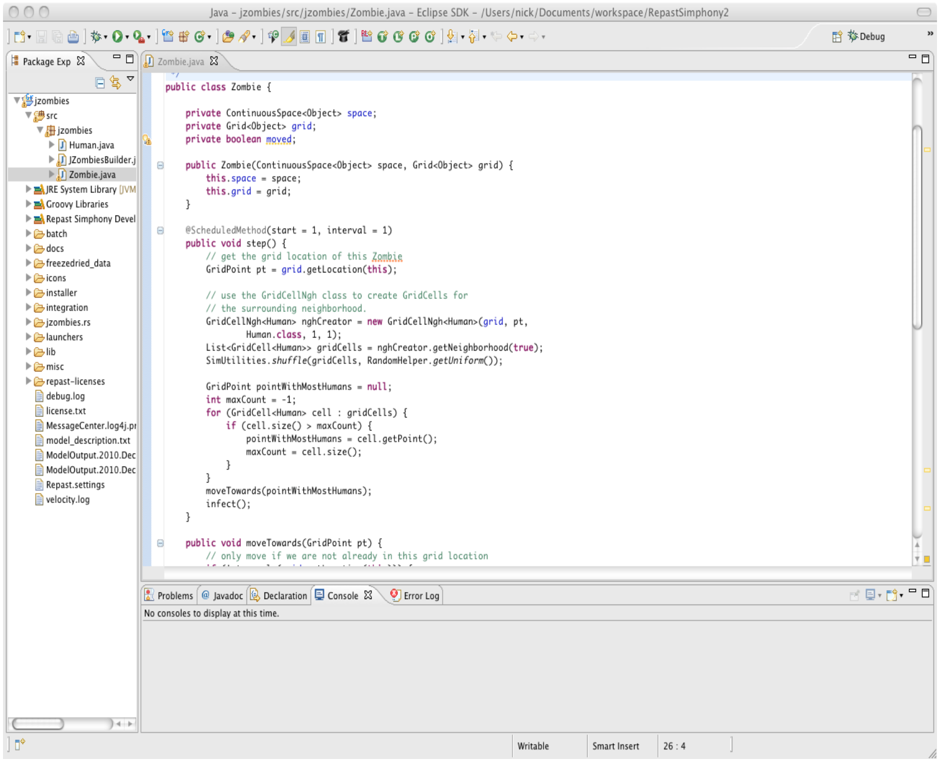


Figure 2.7: Graphical user interface of Repast Suite for Java Modeling (source `http://repast.sourceforge.net/`)

simulation environments, create agents in social networks, collect data automatically and build user interfaces easily. Despite being designed for modelling social behaviour, it is not limited to that (North, Collier, and Vos, 2006). It has a wide variety of applications that range from social systems to evolutionary systems, market modelling, and industrial analysis.

According to Allan (2010), Repast has many similarities with Swarm, both in philosophy and appearance. Similarly, it provides a library of code for designing, running, visualizing, and collecting data from simulations. Perhaps those likenesses are because Repast was initially a Java re-coding of Swarm. However, Repast does not implement swarms. Since Repast was aimed to support social science domain, it includes specific tools for that field and was designed to be easy for inexperienced users. The current version is called Repast Symphony and it is limited to Java. Repast Symphony provides hierarchical and organisation support through *contexts* with special spaces for defining agents' relationships, called *projections* (North and Macal, 2005).

The Repast user community is large and active.It is released under the Berkeley Software Distribution (BSD) license and therefore the source code is freely

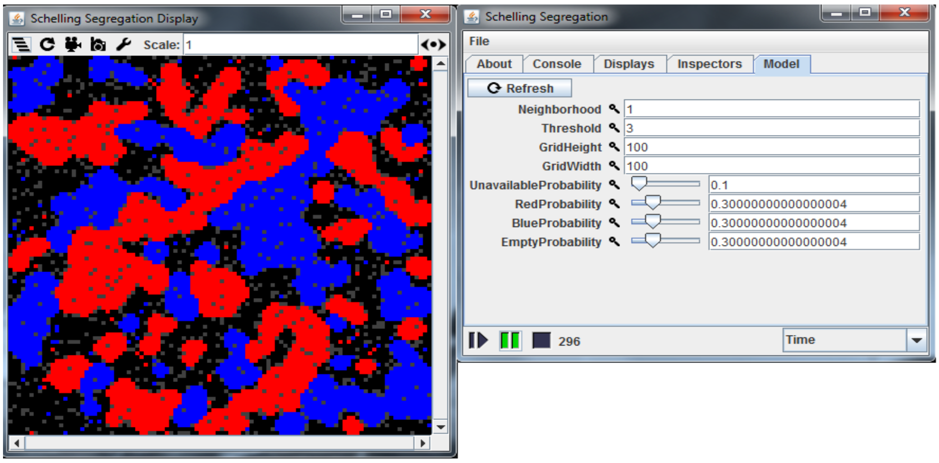


Figure 2.8: Graphical user interface of Mason for Schelling segregation model (source `http://cs.gmu.edu/~eclab/projects/mason/`

available for download. In terms of architecture and computer capabilities, Repast Symphony has a concurrent and multi-threaded discrete event scheduler. Moreover, it has available various numerical libraries such as random number generators and has distributed computing support using the Terracotta Enterprise Suite for Java. Point and click modelling in 2D and 3D is supported. In terms of documentation, although Repast provides some demonstration simulation models such as SugarScape, Swarm's Heatbugs and MouseTrap models, there are very few other simulation models generally available. However, a mailing list of Repast can provide general support and discussion.

Tobias and Hofmann (2004) evaluated free Java-libraries for social-scientific agent-based simulation and found Repast to be the clear winner. Also, Allan (2010) declares Repast as being the agent-based modelling and simulation package with the greatest functionality. Repast has many users involved in a variety of social domains. For example, we find applications in species explorations of landscape (Vidgen and Padget, 2009), reputation systems (Schlosser, Voss, and Brückner, 2005) (Wierzbicki and Nielek, 2011), dynamics of insurgencies (Bennett, 2008), social influence and decision-making (Altaweel, Alessa, and Kliskey, 2010) or evolutionary simulation (Edmonds, 2006). For other Repast application areas, such as evolution and ecosystems, artificial societies, and artificial biological systems see North and Macal (2005).

2.4.4 Mason

Mason (Luke, Cioffi-Revilla, Panait, Sullivan, and Balan, 2005) is the newest entrant into the field of agent-based simulation tool-kits from our list. It was developed by George Mason University's Computer Science Department and the

George Mason University Center for Social Complexity. The limitations of Repast in terms of computational demand inspired Mason design as a smaller and faster alternative. Therefore, Mason was developed as a new platform with emphasis on efficient execution of the code, which is also in Java language (Allan, 2010). It contains a model library and a suite of visualization tools in 2D and 3D, both running independently (see figure 2.8). As one of the newest software tools, it has migration options from other softwares which is interesting for developers. The system is open source and free. For more information about the Mason system and its basic architectural design see Luke et al. (2005).

There are many applications of Mason in social science. An example is the work in Dunham (2005) on epidemiological simulation. Cioffi-Revilla (2010) used Mason to study the emergence and evolution of policies in Inner Asia. Bigbee, Cioffi-Revilla, and Luke (2007) also showed how the Sugarspace model could be replicated with Mason. Luke and Ziparo (2010) used Mason to simulate virtual learning of automata. Other areas of application of Mason include climate change (Hailegiorgis, Kennedy, Rouleau, Bassett, Coletti, Balan, and Gulden, 2010), conflict (Kennedy, Hailegiorgis, Rouleau, Bassett, Coletti, Balan, and Gulden, 2010; Rouleau, Coletti, Bassett, Hailegiorgis, Gulden, and Kennedy, 2009) or nomad societies (Cioffi-Revilla, Rogers, and Latek, 2010). For a complete list of Mason applications see their website (`http://cs.gmu.edu/~eclab/projects/mason/`).

2.5 Discussion and conclusions

In this chapter, we revised the process of modelling in social sciences. We have seen why modelling is necessary for science, particularly in social studies, and the different steps researchers need to be aware of. After briefly introducing the background of social simulation, we saw four different kind of approaches to simulate population dynamics, including microsimulation, System Dynamics, Discrete-Event Simulation and Agent-Based Modelling.

In this thesis, we chose agent-based modelling and showed how it differs from other approaches. Furthermore, we revised the advantages this methodology has and explored the literature to see some applications in the social sciences and humanities. We saw there is a lot of work done in agent-based simulation in social sciences there is much more to do. Particularly in policy, since the political interest in monitoring and global governance of population dynamics is growing up. We found there are not so many studies that have actually used agent-based models to guide governmental decisions. We found this fact surprising, giving the potential of the agent-based approach to gain insight into population dynamics. Therefore, we showed some plausible reasons for its lack of acceptance in policy.

In addition, we looked over four widespread tool-kits for agent-based simulation. We showed their needs, characteristics and some of their shortages. We also revised some social simulation studies that use them. In summary, social scientists experience some specific difficulties when managing agent-based platforms:

- On managing the interaction with the tool, including the need for a good detailed documentation and usability.
- On designing agent-based models users generally have non-existing or basic programming skills
- On defining the kind of features specific to the social sciences and humanities (learning or reasoning capabilities of agents, emergent behaviour, interactions, social networks among others)

Moreover, most current platforms do not provide support to represent and detect emergent structures as explicit entities in the model. The difficulties on following population dynamics during the simulation course should be taken into account not only when designing agent-based platforms with general broad purposes, but also when improving tools that are nowadays in the market (Taillandier, Vo, Amouroux, and Drogoul, 2012). Furthermore, as social simulation field advances new features, new needs will appear demanding for a reasonably fast adaptation of the current tools.

We found there are differences in terminology among software tools. Given that social simulation is a recognised multidisciplinary field, more attention should be paid to this issue. Probably this situation comes from the same background distinction of those who developed the software tool. Therefore, people working in social simulation should discuss and define each of these terms to avoid future misunderstandings.

Agent-based models are being actively applied in many practical areas. The applications range from modelling adaptive behaviours in hunter-gathered prehistoric societies, to understanding consumer behaviours in stock markets. The scope of the applications also varies from minimalist academic models meant to capture the most important features of a system, to decision support systems aim to answer world policy questions, with the difficulty of including real data and later validating the model. According to Allan (2010), in some areas of application is not clear that minimalist applications can be sufficient, although they might contribute to detecting difficulties and shortage in the design of models. What is clear is that there is still a lot of work to do in social simulation field, not only in the technical aspect, but also in terms of definition and agreement on models that emulate population dynamics. In that sense, Allan points out the need of a new generation of agent-based models including more advanced aspects on communication networks, conditional neighbour interaction rules and a protocol for knowledge exchange.

Chapter 3

Parallel simulation for population dynamics

To understand the complex nature of social systems social researchers have a range of methodologies available, simulation among them. The intrinsic dynamic nature of real-world social phenomena may easily lead to simulations too slow to provide the needed insights for researchers (Allen, 2011). That is why parallel and distributed computing can offer an alternative to manage realistic models. However, the distribution of social models among computers in a network of nodes is not an easy task. Scalability (the capacity of the system to handle a higher amount of work as hardware grows) needs to be addressed, although there is no consensus on dealing with the difficulties it encounters on agent-based models (Tesfatsion, 2002; Hybinette, Kraemer, Xiong, Matthews, and Ahmed, 2006).

In the previous chapter, we revised different methodologies for social modelling and simulation. With this chapter, we want to address some of the technological challenges social researchers could expect in the future of social simulation. Concretely, we will explain the most relevant aspects on dealing with large-scale agent-based simulation in the social sciences and humanities.

The rest of this chapter is organized as follows. In sections 3.1 and 3.2, we will provide a general description of parallel and distributed simulation and the current computer architectures to run simulations. Then, we will also analyse the advantages and drawbacks parallel simulation encounters in section 3.3. Later, we will point out some reasons why we think parallel simulation can be very useful to the study of population dynamics (section 3.4). Section 3.5 will follow with the discussion of some open problems on dealing with large agent-based simulation. Moreover, it will revise some tools designed for running them in parallel and distributed environments. Finally, we will present our conclusions on this subject. Part of this chapter in the book *Interdisciplinary Applications of Agent-Based*

3.1 Background on parallel simulation

There is general interest in using individual-based simulation (such as micro-simulation and agent-based simulation) with larger sample sizes in population dynamics simulation models (including demographics). This fact raises a performance issue because individual-based simulation is relatively slow if the model includes a large number of individuals and the simulation time unit is very small. The execution time will be even worse when the sampling process in each individual requires a complex regression model. It should be noted that the simulation execution time depends not only on the processor speed, but also on the memory capacity. If the agent-based simulation model includes a large number of individuals, it will probably require a bigger memory space than the memory capacity of a single-processor computer. This fact will cause numerous operations on memory swapping. Hence, it is possible that sequential simulation might be too slow. Parallel or distributed simulation may offer an alternative.

Parallel and distributed simulation deals with the distribution of simulation tasks and processes using multiple processors. While serial processing runs a simulation in sequence, parallel and distributed simulation takes advantage of coupled computer systems (e.g. a supercomputer or a shared-memory multiprocessor) to reduce execution time or to execute large simulations. Figure 3.1 illustrates the difference between sequential simulation and parallel and distributed simulation. In our example, the simulation is composed by five tasks ($T_1 ... T_5$) that are performed during simulation time. Tasks have dependencies between them. The figure shows a possible distribution of those tasks across three processors. By executing a simulation program in a set of processors, it is hope that the simulation time can be reduced substantially, up to a factor equal to the number of processors (Law, 2007). In that way, the simulation is portioned somehow, and its resulting partitions are mapped to processors. Through the combination of their results, the output of single simulation model is obtained.

Parallel and distributed simulations distinguish mainly by the computer architecture they use. In parallel simulation, computers are mostly connected with a high speed network and share the same physical space. Moreover, infrastructures are commonly composed by many homogeneous machines. Both characteristics make the delay in transmitting a message from one computer to other (known as communication *latency*) relatively low. Latency is a key element on performance studies since it determines how much time computers will spend on waiting for messages to be delivered. On the contrary, distribute architecture is composed by heterogeneous machines that are located in a broad geographic area. Therefore, communication latency tends to be higher than in parallel architectures because

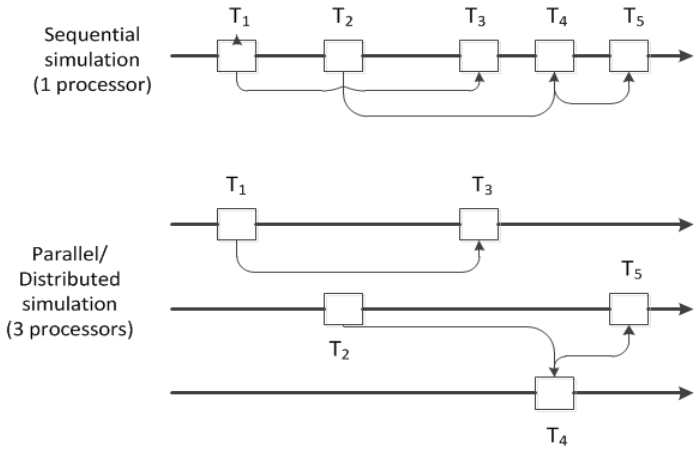


Figure 3.1: Sequential versus parallel and distributed simulation example

it is affected by the distance between the two machines involved in the communication. Moreover, software protocols used to connect both computers tend to be complex and can also have an impact in the communication (Fujimoto, 2000).

Traditionally, parallel and distributed simulation has been applied in military and network simulations. Today, we have seen an increase in the number of papers reporting on parallel simulation applications outside the traditional areas. Tang, Perumalla, Fujimoto, Karimabadi, Driscoll, and Omelchenko (2005) conducted an initial study in applying parallel simulation to a plasma physics application. In the realm of biological science, Lobb, Chao, Fujimoto, and Potter (2005) applied parallel simulation to a neuron model. Parker (2007) used distributed simulation to study an epidemic model. Lan and Pidd (2005) applied parallel simulation to simulate a quasi-continuous manufacturing process. Yoginath and Perumalla (2008) applied parallel simulation to a traffic simulation model. Rubio-Campillo, Cela, and Hernandez-Cardona (2012) explored battlefield dynamics in archaeology using parallel simulation. The situation is encouraging, although it is still far from ideal since the number of applications of parallel and distributed techniques in the social sciences is scarce.

There are two main schemes to divide the simulation: temporal or spatial. Temporal partitioning consists on dividing the simulation along its time dimension. Therefore, each simulation time interval will be executed in a different processor. However, this scheme is atypical in social sciences applications since commonly past events shapes individual or societal behaviour. In a spatial partitioning scheme, the simulation is divided into different sections so each of them contains its corresponding part of the environment and associated entities (agents in agent-based modelling). Furthermore, it is important to analyse the relation between time and space resolutions. Some problems need a specific time step (days, months, years) because entities' dynamics do not have any sense at an-

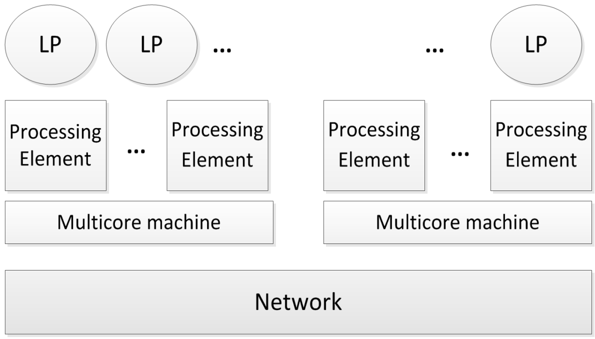


Figure 3.2: Abstraction of a parallel execution architecture

other time scale. For example, if we try to simulate a migration it is difficult to define behaviour at a daily time step rate, so we will use months, or even years. As we will later see in chapter 4, we used month scale to simulate demographic changes in our agent-based demographic model.

To achieve parallelization, commonly a simulation model is partitioned into a set of smaller components called *logical processes* (LP). Logical processes or groups of them are assigned to different processors and run concurrently. Each logical process runs a part of the simulation model and communicates with others logical processes through messages or events. For example, in Figure 3.1 each time line of parallel/distributed simulation would be a logical process (therefore we will have LP_1, LP_2 and LP_3). Each logical process can be viewed a sequential simulation model, having its set of local variables, event list and simulation clock. Figure 3.2 shows the typical execution architecture of a parallel simulation. In order to ensure the execution of the simulation model is correct, events need to be processed at their corresponding time sequence, regardless of their logical processes. This property is named *local causality constraint*. It ensures correct simulation results that will match the equivalent sequential simulation. In that way, if one job is supposed to be done before another job there must be synchronization mechanisms to ensure that sequence of events.

As we saw in Section 2.2.3, simulation methodologies can be classified according to time advancement mechanisms: time-stepped and event-driven. In order to store the events that will take place, discrete event simulation uses an event list. The simulation mechanism consist in taking the first element of the event list, moving the time step to the time of that event, and then simulate the effects of that event. In a time-stepped simulation all events are scheduled with a fixed time increment dt, and entities exchange their status updates via messages while simulation time advances from one time step to the next. Time-stepped simulations are efficient when events are frequent or dense. However, when events are less frequent (when compared to the size of time-steps) the performance of time-stepped simulations drop noticeably (Fujimoto, 2000). Event-driven simulations

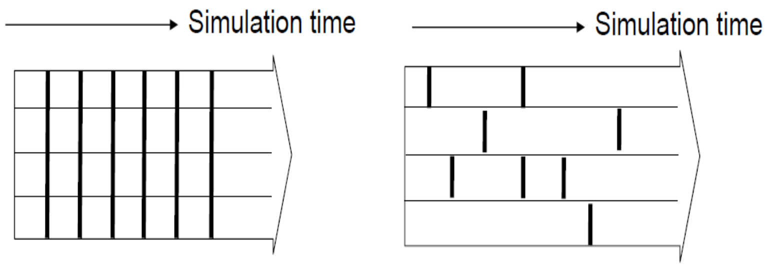


Figure 3.3: Difference between time-step simulation (left) and event-driven simulation (right). Inspired by Fujimoto (2000)

proceed in discrete time steps in which state updates are scheduled. In that way, new state of entities actually drive changes in the simulation. The difference of this approach is that entities exchange events (and not messages) for state updates. Those events are executed in time stamp order. Event-driven simulations have gained significant popularity because they can be effectively applied to a broad spectrum of systems (Fujimoto, 1990). However, time driven simulation could be more suitable when there is human interaction or in real time simulation (Davidsson, 2001). Figure 3.3 illustrates the difference between time-step and event-driven simulation.

3.2 Computer architectures for simulation studies

Simulation models can be run using the capacity of Core Processing Units (CPUs or cores), which are the atom of any computing system. CPU is responsible for calculating arithmetical and logistical operations. With the advances of computer technology, desktop computers went from one CPU to several cores with shared memory. The aggregation of computer nodes through the network is known as a High Performance Computing (HPC) infrastructure. There are four types of HPC systems that can be used to explore simulation problems:

1. A set of computer nodes interconnected with high-speed networks offering HPC at a relatively low-cost (computer clusters),

2. Large sets of computers interconnected with high-speed networks and faster processing speeds, thanks to their memory subsystems and interconnections (dedicated supercomputers),

3. Computers connected through a real-time communication network, typically the Internet, which provides dynamic and scalable resources as services to the end-user (clouds),

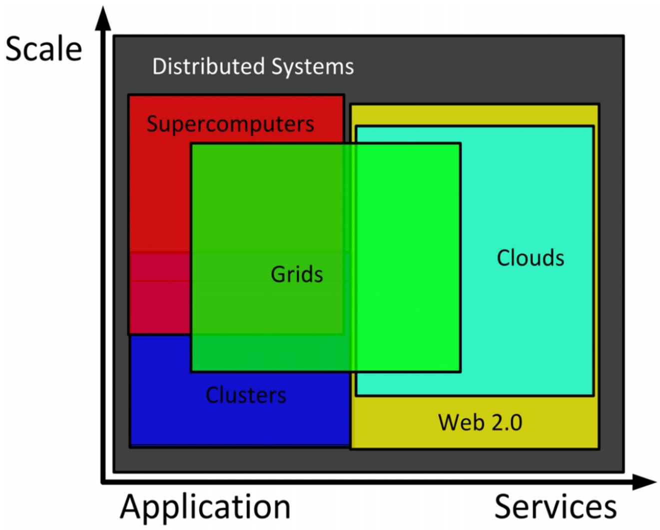


Figure 3.4: Overview of distributed systems according to their scalability and the domain they are applied (source Foster et al. (2008))

4. Computers connected through a real-time communication network as clouds, which require more control by the end user (grids).

An overview of distributed systems can be seen in Figure 3.4. Clusters and supercomputers are oriented to the applications domain, whereas clouds are used in services-oriented applications. While supercomputers and clouds are used for large-scale studies, clusters and web systems are used at small scale. Grid systems overlaps both with application and service-oriented systems but is generally used at a lower scale than supercomputers and clouds.

The main advantage of HPC infrastructures is that they alleviate part of the computational demand by distributing the code among different computer nodes. Each desktop computer will typically have multi-core processors so user can test their parallel applications in their desktop machines and later deploy them in a large system with little or no modification (Parry, 2012). A general architecture of an HPC computer cluster is shown in Figure 3.5. New increases in computer power are coming from embedding more cores in one node.

Whereas desktop machines and small computer cluster are affordable to many research groups, high-performance computing clusters and supercomputers are expensive resources. Internet is an available network to parallelise and distribute simulation work through grids and clouds. Grids are heterogeneous, non-dedicated public systems useful to execute large-scale simulations (Zhang, Theodoropoulos, Minson, Turner, Cai, Xie, and Logan, 2005). They require the distribution of pieces of programs to a large set of computers. This fact has several advantages since work can be divided up among computer nodes, with the corresponding access to resources and work with geographically distributed datasets. However,

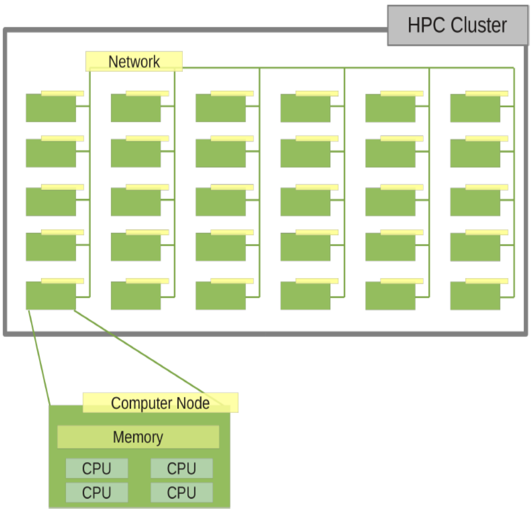


Figure 3.5: General architecture of an HPC computer cluster

they present a disadvantage because if one piece of the software on a computer node fails, other pieces of the software on other nodes may fail. Dependency between node tasks is the weakness point of Grids. Cloud computing evolved out of grid computing and uses grid technologies as its infrastructure support (Foster et al., 2008). It provides access for users to services that can utilize multiple servers anywhere on the globe without knowing which ones they are using or where they are located. Users free from use an application from the PC, or purchase a specific version that's configured for their devices.

Grid and Cloud have similar architecture and, according to Foster et al. (2008), provide services at different layers (see Figure 3.6). Application layer comprises all applications that are available to users (in the case of Cloud with a usage-based pricing mode). Middleware captures collection of resources that can be used build, test and deploy custom applications. Resources layer provides the abstracted/encapsulated access to individual resources of the network, which are in raw hardware resources layer. Grid differentiate from clouds in providing a layer to perform secure network transactions, through authentication protocols and defining the communication channel.

An alternative to CPU-based machines are graphics processing units (GPU). Although GPU have traditionally been designed to handle intensive graphic tasks, recently they have been used to perform more general purpose calculations, including numerical computing (Owens, Luebke, Govindaraju, Harris, Krüger, Lefohn, and Purcell, 2007). GPUs can quickly perform mathematical calculations because of their high bandwidth (rate of data transfer) (Parry, 2012). In agent-based models, they can be useful when the number of communications between agents is high or systems have high mobility. However implementing applications on GPUs

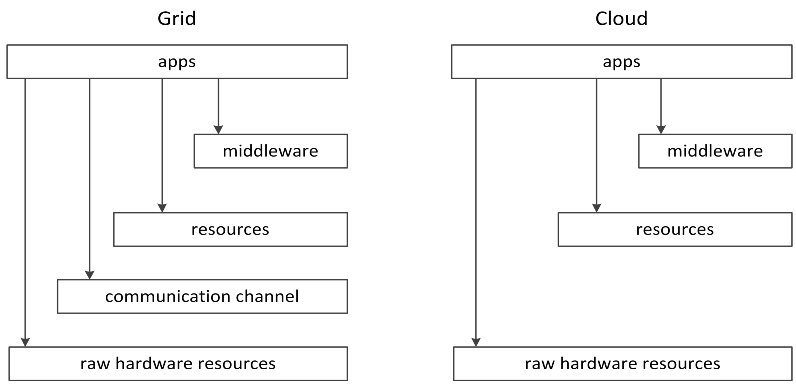


Figure 3.6: Grid vs Cloud architecture (inspired by Foster et al. (2008))

clusters requires high-level programming skills and advance knowledge of the constraints imposed by the architecture. As a result, model implementation needs to be adjusted to GPU particularities. Despite this limitation, GPU can be used efficiently to implement some specific functions (for example to gather information of the environment or to calculate the shortest path). As an example, Wittek and Rubio-Campillo (2013) used GPUs for accelerating the update of knowledge of agents distributed in a graphical information systems. They showed the impact of running the simulation in a GPU cluster had on the implementation of the model. GPUs also have limitations on memory use. Therefore, large-scale scenarios might require distributing simulations between independent GPU cores and subsequently incorporating message passing techniques. In that case, multi-core graphic systems limitations on high latency will need to be taken into account.

3.3 Advantages and challenges of parallel simulation

Parallel and distributed simulation is useful in many ways (Fujimoto, 2000). First, parallel and distribution techniques can help to complete the simulation faster, using a number of processors to distribute the work. Reduction of computer time is critical in applications where simulation is used as a decision aid tool, for example in emergencies that affect air traffic management. It is also interesting in virtual environments for training where participants are located in distant geographical areas. Moreover, it can help handling large or complex systems where the execution is very time-consuming. Second, distributing the simulation workload in different processors makes the system more tolerant to failures. Therefore if a processor goes down others can pick up the work and the simulation can proceed as normal. Third, parallel and distributed simulation makes possible to combine several proprietary simulation tools in a single simulation. In that way, the cost of porting the programs to a common system is avoided. And fourth, they can combine geographically distributed computers and integrate different participants in the simulation (e.g. in flight simulators for air battle). Overall these advantages, parallel and distributed techniques can make possible to simulate larger scenarios, which commonly spend a lot of resources and memory such as large nodes in computer networks or agent-based models with a large number of individuals (Perumalla, 2006).

Nevertheless, parallel and distributed simulation encounters some difficulties on its application. They are mainly due to the architecture of computer HPC systems that are being used. According to (Timm and Pawlaszczyk, 2005), they include

- load balancing between computer cores
- synchronization of events to ensure local causality constraint
- management of communications between nodes
- monitoring the distributed simulation time
- dynamic resource allocation

Among them, good load balancing and inter-node communications with synchronization of events are central to parallel simulation (Parry, 2012). In order to ensure good performance of the simulation, it is important to divide equitable the initial load of the simulation model among nodes. In that way, we will avoid situations of having a computer idle as others are working due to the initial workload unbalance. Pacheco (1997) gives some examples to solve it. However, if there is mobility in the system and computational demands change, techniques of dynamic load balancing will need to be considered. Although there are some solutions to deal with that (Jang and Agha, 2005, 2006), they also represent an additional overhead. Therefore, it will only be worth in cases where either some agents communicate with others with more intensity than others or communication patterns are constantly changing (Parry, 2012). Communications between nodes can slow down substantially the simulation execution time. Therefore, the way the simulation is split between nodes is an important element to obtaining a good performance. However, this highly depends on the simulation model we want to parallelise.

In distributed simulation, new challenges have arisen in recent years. With the increase of the number or processors included in large-scale simulations, the probability of having a fail during the simulation increases. As a result, mechanisms of fault tolerance need to be addressed. Moreover, secure exchange of information through networks should be taking into account in some application areas, especially when data navigates through the Internet (e.g. the case of grid and cloud computing). Therefore, new challenges such as dealing with overheads of encrypted communication need to be studied (Perumalla, 2006).

3.4 Why can parallel simulation be useful to simulate social systems?

There are many reasons why agent-based simulation may require a considerable computational power. Here we point out some of them, partly based on Helbing (2012):

- Agent-based models are non-linear, dynamic systems with high uncertainty and notable degree of stochasticity. Therefore, we will need several runs to obtain results. With parametric sweep, we can run multiple instances of the same program on different sets of input data. HPC infrastructures can minimize the time needed to achieve this task.
- Moreover, parametric sweep can solve the exploration of the model's parameter space. However, determining the number of runs a given case scenario should be executed is not an easy task and the problem needs to be further explored.
- Large-scale scenarios of agent-based modelling might involve a very large number of agents. This situation might require access to a big infrastructure of computing nodes. Although most of the agent-based models in the literature are barely large-scale, advances in the field could make the simulation of these scenarios more necessary.
- Realistic scenarios often include realistic decision-making processes. Enriched models for agents, where a rational behaviour or cognitive and psychological processes take place, require higher computing demands.
- Emergence is a property of many agent-based models. In most studies, it can be explored with small scale simulations. However, there are cases where that is not possible to achieve (Mithen and Reed, 2002). For example Rubio-Campillo et al. (2012) showed how the typical linear warfare can not be correctly modelled while trying to simulate few individuals. They proved it is not possible to simplify the simulation using fewer agents, as the behaviour to explore is linked to the number of them that interact at a given time step.

3.5 Large-scale social simulation

In the last years, many disciplines have profited computer advances to deal with large-scale simulations of agent-based models such as molecular physics, telecommunications, ecology or military research. In the case of social sciences and humanities the number of works is increasing. However, the distribution of an agent-based simulation using HPC resources is not a straightforward task. Agent-based models at a small or medium scale do not often have to care much about their performance. The performance measures the increased computational time needed to simulate a number of N agents in the system. In the ideal case, as we increase the number of agents the scalability of the simulation depends linearly on N. Nevertheless, in many cases time scales worse than linear, as a polynomial or exponential function. That is the characteristic which allow us to distinguish efficient implementations from inefficient ones (Helbing, 2012).

Although there have been some attempts to define protocols for describing agent-based models (Gilbert, 2008; Grimm, Berger, Bastiansen, Eliassen, Ginot, Giske, Goss-Custard, Grand, Heinz, Huse, et al., 2006), there is no common methodology for developing them. Neither it is defined how users should handle large-scale scenarios of agent-based models. Parry (2012) states there are several ways to enlarge large-scale agent simulations: increase computer hardware (using vectorization techniques (Chandak and Browne, 1983) or processors dedicated to specific simulation functions (Comfort, 1984)); reduce the number of agents or revert to simpler modelling; work with aggregates of agents sharing similar characteristics; and reprogram the model in parallel. Parry (2012) describe the potential solutions to implement large-scale agent-based models. The one that has fewer shortcomings is to modify the model to a parallel environment. This solution has been applied successfully to simulate particles on massive parallel computer systems in the physical sciences of fluid dynamics, materials sciences and meteorology. However, it might be monetary expensive and require advanced computational skills.

For some models or case studies, scalability is not crucial and can be solved just by increasing computer power, using vectorization techniques (Chandak and Browne, 1983) or processors dedicated to specific simulation functions (Comfort, 1984). However, there are some cases where scalability is key, so the model needs to be reprogrammed for a parallel environment. In this section, we will describe the elements that affect agent-based simulation performance, and we will briefly see some agent-based simulations platforms that are suited for large-scale scenarios.

3.5.1 Dealing with simulation of large-scale social systems

There are several aspects that particularly affect the scalability of agent-based simulations. First, the complexity of agent-based models has an impact on simulation performance. The overall complexity of an agent-based model depends on the number of agents and the behaviour complexity of each of them. Artificial intelligence has long explored agent's complexity, with approaches like the Belief, Desires, Intentions (BDI) scheme to model agents (Bratman, 1999), which has its roots in cognitive science and comes from a model of human reasoning. It has been used to simulate the behaviour of drivers (El Hadouaj et al., 2001) and to fire emergency situations (Burmeister et al., 1997). Advanced approaches to human modelling of this kind require a high demand in computer power. Therefore, large-scale simulations with complex agents will most of the time require the resources of an HPC system.

Second, complexity and topology of communications between agents have a high effect on simulation speed. The number of communications in the system depends mainly on the distribution of space in the implementation of an agent-based model and the number of movements of agents across the environment scheme. Commonly large agent-based models use spatial partitioning, where each computer node owns a section of the entire simulated scenario. This spatial partitioning is decided due to the interacting topology of agents, which can be either static or dynamic. However, there might be agent-based where space is not important, such studying social mechanisms like evolution of paternal care (Salgado, 2013). In the case it does, it is important to understand the logic of the model to efficiently divide the environment across computer processors. For example, if we think about a case study like migration flows from one region to another it could be interesting to adapt our space partitioning to these population movements and think of the balance between regions and the 'bottlenecks' migrations (communications) can cause.

Third, environment can be as simple as dividing the model in different parts (regions) or being as complex as a graphical interface system. A graphical interface system consists of a set of tools that allow users to interact and understand spatial information. A graphical interface system can not only deal with data at a geographical scale but also with alternative data such as culture, political ideology or religion. A good revision of graphical interface system techniques and capabilities can be found in Castle and Crooks (2006). In agent-based systems, each agent needs to gather knowledge from the environment, as well as from other agents in order to execute its decision making processes. Moreover, agents may modify the environment so environmental data and agents will be shared. The design of the solution for this kind of systems is model dependent. If the environment is rather simple, we can have the same information in all nodes and distribute agents across processors. For example, Parry (2012) used this scheme with an agent-based model of aphid population dynamics in agricultural landscapes of the UK. However, if the agents are complex and have numerous interactions between them it will be costly to have them residing in different processors. Although there have been some initiatives to automatize the parallelisation of agent-based simulations (Coakley, Gheorghe, Holcombe, Chin, Worth, and Greenough, 2012; Kurowski, de Back, Dubitzky, Gulyás, Kampis, Mamonski, Szemes, and Swain, 2009), overall, the nature of the problem and the properties of the computer platform will often guide the method to split the simulation execution.

Finally, the computational consideration of resolving time advancement mechanisms should be taken into account. The synchronously or asynchronously update of time determines how agents change during the simulation. Most social simulations commonly use a time-stepped execution, in which time advances at a fixed interval and agents are updated at each time step. However, there might be cases where either the nature of agent updates is inherently asynchronous or asynchronous update of time might be an alternative to increasing performance (Perumalla, 2010). The reason is asynchronous (discrete-event) implementation can increase simulation speed since agents will only be updated when it is needed to. However, in the parallel execution context it adds additional complexity since the mapping of asynchronous simulations is not so obvious. Therefore, discrete-event simulation can be faster than time-stepped simulations. If there is communication between nodes, an asynchronous update is a problem since some nodes will have to wait for others to finish processes before communication can take place. To ensure simulation results are correct, research in parallel and distributed simulation has produced a number of synchronization protocols that can be classified into two main categories: conservative and optimistic. While conservative protocols do not allow any causality constraint violation throughout the duration of the simulation, optimistic protocols allow its violation but provide mechanisms to rectify it.

One of the frequently used conservative simulation protocols is the CMB protocol Chandy and Misra (1979) Bryant (1984). This protocol uses a dummy message called a null message. When a logical process sends a null message with a timestamp “t”, it tells the receiving logical process that it will not send any message with a timestamp earlier than “t”. The time-stamp “t” depends on what is called the look-ahead. It represents the minimal amount of physical time that is required to complete the process in the real world. The logical processes use this information to advance their local simulation time. Hence, the performance of the CMB protocol depends on the look-ahead quality.

The Time Warp protocol Jefferson (1985) is the most widely used optimistic protocol. In this protocol, once a logical process detects a local causality constraint violation, it rolls back to the last correct state and restarts the simulation from that point. Traditionally, the rollback process is made possible by using a state-saving mechanism. A more recent method includes the *reverse computation* method where the correct state is recovered by reversing the computation process Carothers, Perumalla, and Fujimoto (1999). Fujimoto (2000); Perumalla (2006) provide a good summary on these techniques.

Furthermore, the duration of an agent-based simulation is affected by technical characteristic of computer architecture. Network speed of different platforms (cluster, grids, clouds, supercomputers) may vary and cause differences on latency of message transmission. These differences will greatly affect communications of agents and thus simulation performance (Wang, Turner, and Wang, 2005)). Moreover, the computer power of each processor is a key element for reducing the execution time. Users should be aware that running agent-based simulation in heterogeneous computer infrastructure (such as grid or clouds) can result in unexpected unbalances and delay the execution time. There exists some mechanisms to profit shared-memory capacities of processors with libraries such as OpenMP, which allow to parallelize the execution of agents’ actions within a single computer node (Massaioli, Castiglione, and Bernaschi, 2005). However, there might be contexts where this is not possible to do due to conflicts on accessing and modifying the same data at the same time (for instance, when agents modify the same portion of space). Finally, output requirements or data storage can cause trouble in the implementation of agent-based simulations. For example, imagine that we need to get data at the micro-scale of individual stories. Reprogramming the model might affect to the output handling and decrease simulation performance.

3.5.2 Tools for large-scale agent-based models in social sciences

There is a lot of interest in developing agent-based models as a general technique to be applied to the study of societies. A number of platforms exist to provide the means to study social phenomena. However, they are essentially discrete-event tools designed to execute serially on CPUs and have limited scalability (Lysenko and D'Souza, 2008).

To date, it appears there is not any parallel or distributed implementation of Swarm. Although the Netlogo engine has no fixed limits on the size, it is single-threaded, single processor based thus being problematic to run large-scale models. To fix that, there is an extension of Netlogo called *BehaviorSpace* that allows the user to run the simulation in parallel, one per core in a multicore or a cluster of processors. BehaviourSpace was specially design to explore the parameter space of models, exploring possible behaviours to determine which combinations of settings cause the behaviours of interest. BehaviourSpace is also free and open source.

In the context of parallel and distributed simulation, several frameworks have been recently developed. They range from parallel, distributed approach to the GPU/many-cores or tools to automatize parallel agent-based simulation code. In this section, we will briefly describe some of them and compare them in terms of software capacity and architecture. All of them take advantage in different ways of HPC architecture to deal with either big or very complex simulations of social phenomena. The purpose of this chapter is not to fully revise these tools but introduce some of them to the reader.

Repast HPC

Repast-HPC is the most popular tool for tightly-coupled, large computing clusters and supercomputers (Collier and North, 2012). It is based on the tool Repast for sequentially execute agent-based models. Repast-HPC is intended for users with basic C++ expertise and access to high-performance computers. Repast-HPC implements a dynamic discrete-event scheduler with conservative synchronization. It offers useful features such as data collection, specifying agent interactions by space and networks or automatic management of agent interactions across processes. Repast-HPC allocates a region of space to multiple processors and manages the boundaries by copying them (and their agents) in the adjoining region. In Collier and North (2011) the authors show an example of applying Repast-HPC to

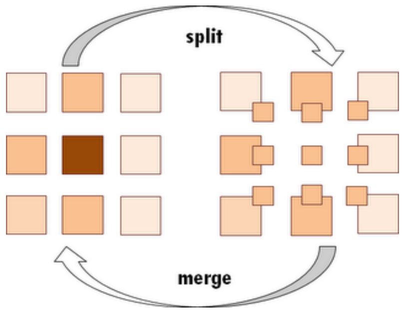


Figure 3.7: Load balance policy in DMASON (source Cordasco et al. (2011))

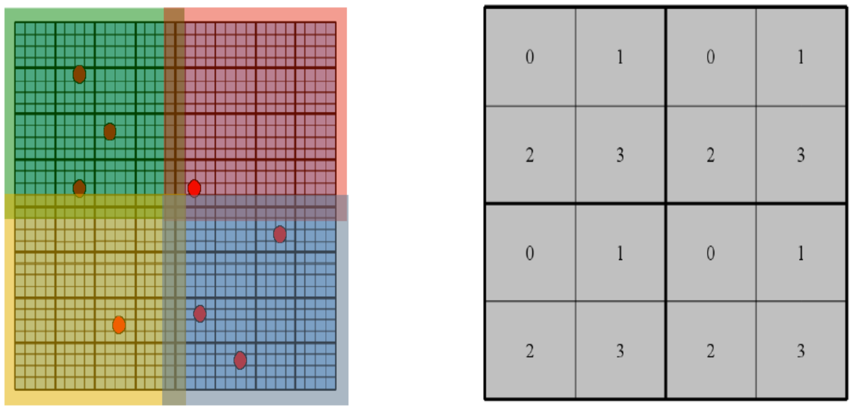
the spread of rumours through a networked population.

D-Mason

DMASON is the Distributed Multi-agents simulation tool-kit, based on Mason. As in Repast-HPC, users can create an agent-based model in MASON and then use the framework to easily distribute it over many machines. It provides a partitioning functionality that self-balances regions and requires and all-to-all communication. As opposite to Pandora and Repast-HPC, partitioning of the field is decided by the user in advance and, in our opinion, may cause an additional unbalance. Agents can migrate from one region to other and, therefore, it does not guarantee load balancing. However, it implements a simple balancing mechanism: when a region is overloaded it decides to split itself in smaller regions, dividing consequently the amount of agents in each of these regions (see Figure 3.7). To deal with boundaries problems, neighbouring regions communicate before each simulation step. DMASON is developed in Java and uses Java Message Service (JMS) for communication between workers, so it is not meant for Java beginners.

Pandora

Pandora is a framework designed to implement agent-based models and to execute them in HPC environments (Rubio-Campillo, 2013). Pandora is currently being used to simulate ancient societies and their relationship with environmental transformations in the project "SimulPast" (`http://www.simulpast.es`). It has been explicitly programmed to allow the execution of large-scale agent-based simulations, and it is capable of dealing with thousands of agents developing complex actions. Pandora implements a time-stepped scheduler and defines the environment where agents live as a set of layers containing raster map structures, following graphical interface systems standards. As in Repast-HPC, in Pandora the user is responsible for writing the agent-based code (in C++ or Python) and the world of the simulation is evenly divided among computer nodes. This layout is depicted in Figure 3.8, where simulation landscape is divided into four main sections which are executed in four different processors. Data and agents located in the border between adjacent nodes will be copied and sent to neighbours every time step, in order to keep data up-to-date in the entire scenario. The range of this border will be defined as the maximum interaction range of any agent, because this is the horizon of actions of any component of the simulation.



(a) space schema **(b)** space distribution

Figure 3.8: Space schema and distribution to solve collisions in Pandora

This solution for space partitioning (adopted as well by Repast-HPC) is highly scalable, given the fact that every computer node will need to communicate, at most, with eight neighbour nodes, independently of the total size of the simulation. To solve collision between agent’s actions, Pandora splits the section of every computer node in 4 equal parts, numbering 0 to 3, as it can be seen in **??**. The agents contained in all 0 sections will be executed simultaneously without the possibility of conflicts. When they finish, modified border data is sent to the neighbours, and a new section will begin its execution (1, 2 and finally 3). Once all of them are done, the entire state of the simulation is serialized, and a new time step can be evaluated.

Other tools

Using grid technology, the open source project GridABM (Gulyás, Szabó, Legéndi, Máhr, Bocsi, and Kampis, 2010) provides a set of templates to enable researchers to run agent-based models in computing clusters and computational grids. It is based on Repast and depending on the topology of communication of agents it allows user to choose the appropriate schema and run their simulations in different parallel and distributed simulation platforms. The programming is very similar to developing Repast sequential applications.

There are also proposed solutions that take advantage of GPU technology to perform large agent-based simulations. Examples of that can be seen in Lysenko and D'Souza (2008) and Perumalla and Aaby (2008). In their solutions, all agent data is stored in large multi-dimensional arrays where each element holds the complete state of an agent. In that way, GPU computer architecture is profited to faster several orders of magnitude agent-based modelling tool-kits. However, the programming of GPUs is counter-intuitive. Moreover, these perspectives do not display well complex behaviours at interactive rates. There have been other approaches to simulate complex agent-based models (Pelechano, Allbeck, and Badler, 2007; Treuille, Cooper, and Popović, 2006) but they are far from scaling the number of agents. Vigueras, Lozano, Perez, and Ordua (2008) proposed a scalable architecture to manage large crowds of agents. They later apply it to a real case scenario (Vigueras, Orduńa, and Lozano, 2010).

There exist tools that automatically generate parallel agent-based simulation code. An example is FLAME (Richmond, Walker, Coakley, and Romano, 2010). In FLAME agents are modelled as finite state machines which consist on a set of finite states, transition functions, and input and output messages. The state of each agent is determined by its internal memory. Agents communicate through messages that can be produced by themselves or can receive from others, similarly to Repast. FLAME takes the model definition of the agent-based model and the C-code to implement agent actions and state changes to automatize the generation of the user application. The resulting code can run either in a serial or in a parallel environment. It is available for both CPU-based and GPU-based systems. Although FLAME is meant to free the user from parallelism complexity, agent model must contain that parallelism at a task level. The advantage of this platform is that it distributes the computational load over the processing elements to achieve computational load balance.

Previously described tools have focused on the distribution of one simulation over several nodes and require high-programming computational skills. However, it would also be interesting to explore the distribution of a set of experiments over an HPC infrastructure. In that case, a single simulation is executed on one node and several simulations are launched in parallel. An example of this approach are the tools OpenMole (Reuillon, Leclaire, and Rey, 2013) and EPIS (Blanchart, Cambier, Canape, Gaudou, Ho, Ho, Lang, Michel, Marilleau, and Philippe, 2011). OpenMole is a cloud approach for large-scale model exploration of complex systems, which distributes experiments on an HPC environment transparently. It uses a Domain Specific Language to design experiments for simulation models to perform their sensitivity analysis. EPIS also allows the deployment of lots of simulation runs over a cluster of nodes to explore the solution space of agent-based simulations. EPIS uses grid computing to distribute the experiments on a cluster. The advantage of both frameworks is the possibility to profit HPC architecture to explore the parameter space without having HPC skills.

3.6 Summary and conclusions

In the last years many disciplines have profited computer advances to deal with large-scale simulations of agent-based models such as molecular physics, telecommunications, ecology or military research. The work done on the social field is not very substantial although the community of social simulation is expanding. Despite being quite new compared to other interdisciplinary fields like geoinformatics or bioinformatics, social simulation can profit from the technological advances to boost its results and examine large-scale social phenomena going beyond organizations, institutions or small groups without losing its micro-scale.

In this chapter, we have revised the technical challenges social researchers need to face when dealing with simulations of large-scale agent-based models. A summary of those challenges is summarized in Table 3.1. One of the motivations was to show the reader that large simulations are feasible. We presented the most common computer architectures that encompass HPC infrastructures and gave several arguments to prove parallel and distributed simulation is interesting for social models. Although most common tools for agent-based simulation are not meant to deal with large-scale scenarios, we presented some that can actually help to overcome this problem. In order to parallelize large simulation models, it is interesting for modellers to know the difference between CPU and GPU or between parallel simulation in multicore clusters and grids or clouds. Technical characteristics of such systems can not only affect simulation performance but also may require significant changes on program restructuring. The debate about which system is better is far from the aim of this chapter, but it is important to note that they can serve for different purposes within a social perspective. On one hand, cloud and grid computing can be an affordable solution to run large social simulation when no high-speed cluster computing is available. On the other hand, the best solution for models with intensive communication is supercomputing (Wittek and Rubio-Campillo, 2013).

Scalability of social agent-based models is not a trivial matter and requires an interdisciplinary effort. On one hand, computer scientists need to know certain aspects of the application domain to participate successfully in the model design. On the other hand, social researchers should be aware of the computational challenges the model generates at different layers (availability of computational resources, experiment design, model scale, and so on). In this sense, this chapter explores the problem of launching large parallel and distributed simulations in the context of an interdisciplinary research. One of the main aspects of using HPC infrastructures to launch large simulations is to ensure we obtain the same results as the equivalent sequential execution. Since the process of transitioning from a sequential to a parallel and distributed version can be costly, it is worth to explore different options and study simulation needs at first. Moreover, there might be cases where parallel simulation is less efficient than a single processor implementation. Parry (2012) showed that, sometimes, a population size should be large

level	characteristic	impact	highly affects	depends on
model	agen'ts complexity	computational time	simulation performance	schema to model the agent
	complexity and topology of communications	communication overheads	scalability and performance	distribution of space, communications' frequency (static or dynamic interactions)
	environment's complexity	computational time and communications	distribution of workload across processors	model needs
computer architecture	computer power	computational time	simulation performance	computer platform (heterogeneous/homogeneus infrastructure)
	time advancement mechanism and synchronization	computational time and communications	simulation performance	frequency of events
	latency of communications	communication overheads	simulation performance	computer platform

Table 3.1: Summary of challenges on large-scale social simulation from the model and architecture perspectives

enough to compensate overheads due to message passing and extra calculations required. Beyond that, we discussed the most important aspects that affect social agent-based simulations. We also described some of the currently available tools to simulate large-scale social agent-based models. With this chapter, we hope to guide social modellers through the possibilities parallel and distributed architectures can offer.

Chapter 4

Yades: a parallel demographic simulation tool

Today, our society is inexorably shifting in form and nature, influenced by challenging transitions driven by social, economic, environmental and technological changes. The European Union is configuring its research agenda in the Horizon 2020 program around societal challenges like ageing, energy saving, smart transport, secure internet, inclusion and preservation of cultural heritage. Most of these challenges can be overcome only by using tools to provide projections of future population which help deepening on the understanding of population dynamics. Simulation is one of these tools which has a lot of potential. However, it has been little used in demographic studies to help explaining dynamics (Billari, Ongaro, and Prskawetz, 2003). Parallel techniques can provide a support to manage large simulations, not only in terms of dealing with complex models but also when working with large data sets.

In the past two decades, research into parallel simulation has been dominated by the development of better synchronization algorithms and their performance evaluations. From 2005, efforts have been made to promote the use of parallel simulation in wider areas of application. As we saw in chapter 3, parallel simulation is widely used across many different application areas including military and network simulations, and shy efforts have been done in areas around the social sciences.

Among social studies, demography is an area of research which has greatly contributed to project population to guide societal planning (Rees, 2009). Demography is the study of human population in relation to changes brought about by the interplay of births, deaths and migration (Pressat, 1985; Posada, Hernández, and López-Paredes, 2007). Population projection is one of the main applications of demography. Although the population projection is a simplification and un-certain representation of the modelled population, it is often used as an input to models used for planning and policy making. Traditionally, policy models allocate people in every population group (usually based on age and gender) to a set of states (for example, in full-time employment, in full-time education, unemployed and retired). One of the traditional methodologies in population projection uses a macro-level model describing the flows of cohorts (people in same age and gender group) from one group to another. People may enter and leave the community through births, deaths and migrations. Hence a good estimation of flow rates entering into and leaving from every group in the community is critical.

As computer technology advances and the data at a micro-level have become available, the multi-state simulation model at the level of individuals has gained popularity for policy analysis. In this model, the state transition of each individual is simulated (such as the transition from being in full-time education to full-time employment). The multi-state model has now become the standard methodology in demography (Willekens, 2005). Complex policy models that include biological factors (such as health-related factors), cognitive factors (such as learning) or social factors (such as social network) may require a significant amount of computing power. As shown later in this chapter, even the demographic model alone requires a significant amount of computing power, let alone if a complex policy model is added on top of it. Parallel simulation may offer an alternative solution.

There are several tools that support the development and execution of generic agent-based models (see Chapter 2). However, desktop agent-based modelling tools may not scale well for large-scale simulations. Recently, some efforts have been made in this direction by using High Performance Computers (see chapter 3). High performance computing enables the execution of large scale simulations by distributing the workload on a number of processors. The work of this thesis contributes to the current body of knowledge addressing specifically large demographic simulations under an agent-based approach.

The use of an agent-based model is justified for its ability of generating individual event histories and produce estimates of the full distribution outcome. Agent-based modelling is particularly useful to project a population answering "what if" questions such as the effect of a policy on a demographic characteristic. The final goal of this work is to model the impact of individual decision units from the changes in strategic planning or government policies. Despite agent-based modelling useful features, there has been a limited use in the area of demography. However, agent-based approaches can improve our understanding of population dynamics, particularly demographic behaviour (Billari et al., 2003; Billari, Fent, Prskawetz, and Scheffran, 2006). Agent-based modelling pre-suppose (realistic) rules of behaviour and try to validate them by showing if they can explain macroscopic regularities.

This chapter presents the design and development of a parallel demographic simulation framework which is able to simulate population dynamics of a geographical areas using agent-based simulation. The framework is called Yades (*Yet*

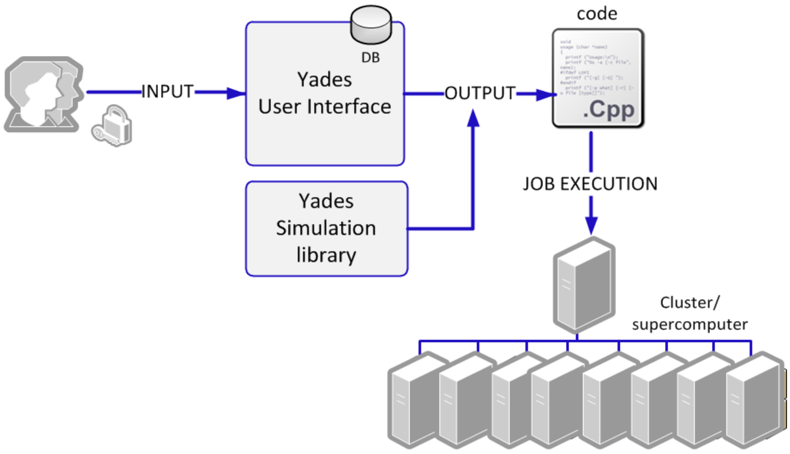


Figure 4.1: Yades framework for paralell demographic simulation

Another Demographic Simulation) and takes advantage of parallel computation to run large-scale demographic models. The work departs from the prototype presented in Onggo (2008). The framework presented was developed in collaboration with the author through an International Joint Program of the UK Royal Society and a grant from the Research and Innovation Ministry from Spanish Government.

Yades has three components: a web user interface, a demographic simulation library and the simulation code generator. The web user interface allows demographic modellers to specify demographic model components in a number of representations familiar to demographers such as regression and statistical distribution function. The simulation code generator can produce the corresponding C++ code that is linked to the demographic simulation library which uses a scalable parallel discrete-event simulation engine. The event-driven approach and a time step of 1 month was improves the efficiency of our simulation model since events are less frequent (when compared to the size of time-steps), as seen in Section 3.1. The generated code is ready for compilation using a target C++ compiler. The demographic simulation library supports both sequential and parallel execution of the simulation model. The framework is picture in Figure 4.1.

The chapter is organized as follows. In section 4.1 we introduce the most general methods for population projection in demography. In section 4.2 an overview of related work in demographic simulation is provided, including some previous work on agent-based simulation in demographics. An explanation on the design and development of the simulation tool then follows in section 4.3. In section 4.4 we present the user interface and justify its need. Finally, in section 4.5 we revise our findings and conclusions.

4.1 Components of a demographic model

The main core of a demographic simulation model is the life course of individuals. In general, a demographic simulation model is formed by a number of demographic components such as birth (fertility), change in economic status and marital status, migration and death (mortality). Hinde (1998); Rowland (2003) provide a good explanation of various demographic components and their modelling methodology. A short summary is given below.

Fertility

Common models for the birth component are related to the fertility model that describes the propensity of the women in a population to bear children. The models include age-specific fertility, parity-specific fertility, birth spacing and their combinations. Age-specific fertility uses age to determine the probability of having a child. Parity-specific fertility takes into account the number of children a female individual has already had. The birth spacing model focuses more on the time between each birth. It is rather complicated to model because it includes factors such as birth control, abstinence period, economic status, etc.

Economic status

The ability to track an individual's economic status is included in the model because there is a plan to build a policy model on top of it (for public policy planning and analysis in such areas as taxation, pensions and benefits). This ability is essential because many policies are linked to economic status. Most models use state-transition diagrams to represent the possible changes in status. In the simplest case, the transitions are made based on a series of simple random samplings. More complex models will include explanatory variables such as an individual's characteristics, the characteristics of the individual's family and external socio-economic factors. Figure 4.2 shows an example of a state-transition diagram. An individual starts from the state "unborn" and moves to state "dependent" until the person reaches the end of the dependent-age. At this stage, the individual may move into one of the three states: "in employment", "in higher education", or "unemployed". As time progresses, the individual may switch between state "in employment" and state "unemployed", and eventually the person will reach the state "pension". An individual in any state can move to state "dead". The state-diagram given in Figure 4.2 assumes that an individual may never move to state "in higher education" in later life. If this assumption is not acceptable, a modeller can change the assumption (and hence the state-transition diagram).

Marital status

Demographers consider the pattern analysis of marriage and cohabitation to be important. One of the reasons is that after age, the most important factor that

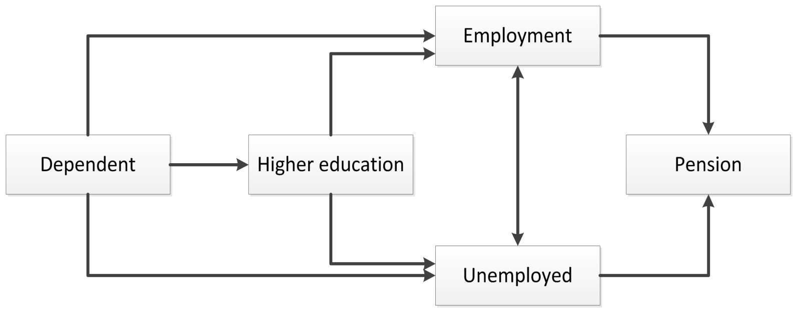


Figure 4.2: An example of a state-transition diagram for economic status

determines the fertility rate of female individuals is marital status. Furthermore, the ability to track an individual's marital status is needed in the model for the same reason as in economic status, i.e., many policies are linked to marital status. Moreover, the difference between marriage and cohabitation has shown to be significant in terms of understanding households dynamics (de Vaus, 2004). Similar to economic status, most models use state-transition diagrams to represent the possible changes in status. In the simplest case, the transitions are made based on a series of simple random samplings. More complex models will include explanatory variables such as an individual's characteristics, the characteristics of the individual's family and external socio-economic factors. Figure 4.3 shows an example of a state-transition diagram for marital status.

In demographic simulation, the marriage model is rather special. Two marriage models are commonly used: open and closed. An open marriage model is simpler because it simply changes the marital status of the individual in the model. A closed marriage model, which is used in our simulation tool, is more complicated. In this model, the simulation can only schedule the start of a 'partner search' rather than a marriage per se. If a suitable individual is found from the list of prospective partners, then the marriage event will be scheduled for both individuals. Otherwise, the individual will be added to the list of prospective partners for a specific duration. The same model applies for cohabiting individuals and civil partnerships.

Migrations

The effect of migration has become more significant in recent years. Migration processes involves the formation and dissolution of households during the simulation. Unlike births and deaths, an individual may migrate (either domestically or internationally) more than once. A simple model uses a simple random sampling to decide whether an individual is going to migrate and to determine the new place. A more complex model employs a combination of individual-specific factors, family-specific factors, region-specific factors and other external factors to explain the individual's decision to migrate to a new place. Furthermore, immi-

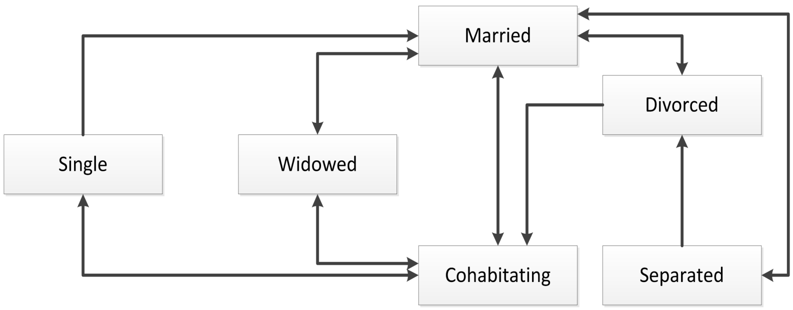


Figure 4.3: An example of a state-transition diagram for marital status

Interim Life Tables, United Kingdom Back to contents

Period expectation of life Office for National Statistics

Based on data for the years 2005-2007

Age	Males					Females				
x	m_x	q_x	l_x	d_x	e_x	m_x	q_x	l_x	d_x	e_x
0	0.005504	0.005489	100000.0	548.9	77.16	0.004439	0.004429	100000.0	442.9	81.47
1	0.000402	0.000402	99451.1	40.0	76.59	0.000371	0.000371	99557.1	37.0	80.83
2	0.000259	0.000259	99411.1	25.7	75.62	0.000181	0.000181	99520.2	18.0	79.86
3	0.000180	0.000180	99385.4	17.9	74.64	0.000147	0.000147	99502.2	14.6	78.87
4	0.000131	0.000131	99367.5	13.0	73.65	0.000105	0.000105	99487.6	10.4	77.89
5	0.000120	0.000120	99354.5	11.9	72.66	0.000093	0.000093	99477.1	9.2	76.89
6	0.000121	0.000121	99342.6	12.1	71.67	0.000103	0.000103	99467.9	10.3	75.90
7	0.000093	0.000093	99330.5	9.3	70.68	0.000084	0.000084	99457.6	8.4	74.91

Figure 4.4: Life table example

grant populations are likely to have different demographic characteristics to the populations they join (Haug, Compton, and Courbage, 2002).

Mortality

The most commonly used method in mortality analysis is the life table (and its variants). The life table summarizes the variation of mortality by age and gender. Survival analysis is another commonly used method where a distribution function of lifetime was used to sample an individual's lifetime at birth. Figure 4.4 and Figure 4.5 show an example of the life table published by the Office of National Statistics and an example of a distribution function of lifetime, respectively.

4.2 A demographic model for population dynamics

Traditionally, there are two methods commonly used in population projections: mathematical model and cohort component (Hinde, 1998). In the first method, demographers use a set of mathematical equations to project the size of a future population. The cohort component method is used to project the structure of

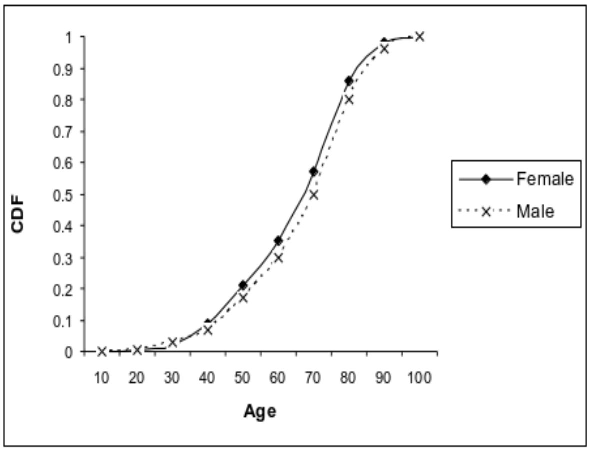


Figure 4.5: Distribution function of lifetime example

$$P_t = \frac{r}{C.r.e.^{-rt}+k}$$

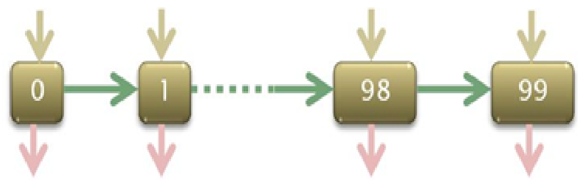


Figure 4.6: Examples of mathematical model and cohort component in demography

a population, i.e., the size of different groups within the population (such as different age groups). This is an iterative method in which each iteration projects the structure of a population in the following year using a set of flow models. Hence, the latter method is able to produce a more detailed projection.

For many purposes, the population structure is more useful than the total population size. The equation on the left in Figure 4.6 is an example of a well known equation in demography. It represents a logistic growth model where P_t is the size of population at time t, C is a constant associated with the initial population size, r is the annual rate of population growth, and k is a parameter to be estimated ((r/k) is the projected maximum population size). The flow model on the right in Figure 4.6 is an example of a model used in cohort component method. Individuals in the population are grouped based on the age of their last birthday. They will progress to the next group until they leave the system due to mortality or migration. New individuals may enter into the system through birth or migration.

The use of simulation for population projection has increased in recent years. Particularly, supported by the advances in computer technology and the availability of data at a micro level (individuals), the use of micro-level simulation models in population projection has become more widespread. The main advantage of this approach is that individual-specific explanatory variables can be included in the model. For example, we may include factors such as age, education level, salary group and ethnicity to model the number of children that an individual female will have. To take one example, Liddle (2002) developed a simulation model to assess the effect of demographic dynamics on sustainable development at three levels: economic, social and environmental. Apart from their application in population projection, demographic models are often used as the basis for policy modelling and analysis. To take two cases, Walker, Percival, and Harding (2000) used simulation to analyze the effects of demographic changes on government expenditure on pharmaceutical benefits in Australia. Similarly, Bonnet and Mahieu (2000) analyzed pension policy in France using simulation. Their simulation model comprised three main components: demographic, labour market and income.

Agent-based models have been used in demographic simulation. Read (1998) used them to explore the interplay between the demographic system and the cultural system in an artificial society of hunter gatherers. Among the recent works in this area, we find a work that proposes agent-based extensions on a spatial microsimulation model of demographic change by Wu and Birkin (2012). Basically they took a microsimulation model and improved it by modeling the heterogeneity, autonomy, explicit inclusion of space and bounded rationality of agents. They modelled ageing, mortality, fertility, health change, marriage and migration demographic components. Their approach helped them to understand the impact of interactions, behaviour and influence of personal history in projecting the student migration and mortality in Leeds (UK). Geard, McCaw, Dorin, Korb, and McVernon (2013) showed an example of how to use agent-based modelling (they called it *"Individual-based model"*) to create a synthetic population which is able to describe basic demographic processes of ageing, mortality, partnership and kinship at a individual and household level. They explored the interaction between these processes and patterns of infection and immunity.

Another example is the work of Kniveton, Smith, and Wood (2011) who proposed an agent-based approach to understand environmental migration (named *"Agent Migration Adaptation to Rainfall Change"*). Following this conceptual model they built an agent-based model for Burkina Faso demographics (including birth, ageing, mortality, marriage, and international migrations processes.) with the objective to assist policy makers. Silverman, Bijak, Hilton, Cao, and Noble (2013) provided an example of combining statistical demographic methods with agent-based modelling to understand how changing family structures in the UK population may influence the provision of health care. To do that, they used an extension of the Wedding Ring, a model of agent-based partnership formation based on Billari, Prskawetz, Diaz, and Fent (2007). With their approach, they state to overcome the limitation of demographic models over-reliance on purely statistical information at the expense of theories and mechanisms.

As in previous studies, our proposed simulation tool implements a set of agent-based demographic models (birth, ageing, mortality, marriage, economic status, and migration processes) to explain population dynamics. However, we use parallel environment to take advantage of High Performance Computing capabilities to run large-scale simulations. Although in agent-based simulation some approaches have already been proposed to explore demographics, none of them is addressing large scenarios. Despite some tools deal with large-scale agent-based models, their interface is not specifically designed to model demographic human behaviour neither they target demographers and policy makers.

4.3 Yades design and implementation

There are a number of environments which have been developed to provide numerous services to building parallel/distributed simulation systems by supporting optimistic, conservative, or hybrid synchronization strategies (Preiss, 1989; Steinman, 1992; Radhakrishnan, Martin, Chetlur, Rao, and Wilsey, 1998; HLA, 2000; Steinman, 2005). Among them, there is one named μsik that supports multiple synchronization algorithms such as: lookahead-based conservative protocol and rollback-based optimistic protocol (state-saving and reverse-computation) (Perumalla, 2005). μsik has been reported to be scalable to run a synthetic benchmark called PHOLD on a large number of processors (Perumalla, 2007). This library adopts the process interaction world-view; hence a simulation model is formed by a set of interacting (logical) processes (LP). These logical processes are implemented as μsik processes. Logical processes communicate through events with the standardized communication protocol Message Passing Interface (MPI) (Pacheco, 1997). Logical processes communicate through events. Multiple logical processes are mapped onto a physical process (μsik kernel) that is run on top of a processing element (PE). A machine can have more than one processing element (e.g., in multi-core architecture). It should be noted that in μsik documentation, the physical process is often referred to as the federate. To avoid confusion with the federate in High Level Architecture (HLA, 2000), in this dissertation the term physical process is used instead of federate.

Yades was implemented using μsik parallel simulation library following a Single Program/Multiple Data (SPMD) paradigm. In SPMD each processor runs that same program, in contrast to a master/slave approach. We used logical process to implement agents in our demographic agent-based model. Figure 4.7 illustrates the layers of software needed to implement the parallel simulation tool using μsik. To implement a simulation model in μsik, we must specify three main components: a physical process (in our case `PopulationSimulator`), a set of logical processes (`Region` and `FamilyUnit`), and a set of events (of type `PopulationEvent`). Figure 4.8 shows the Unified Modeling Language (UML) class diagram of the tool. We use it here because UML is a well-known standardized modelling language for software engineering (Fowler and Scott, 2000). In this figure we can see how the different physical and logical processes described bellow inherit from μsik classes. A detailed explanation on the structure of a simulation model written in μsik can

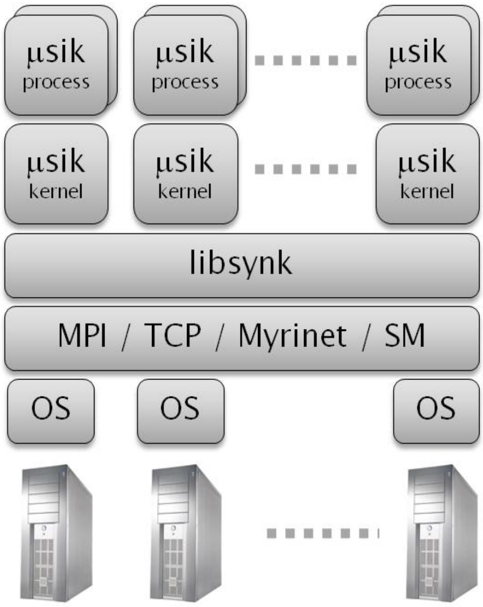


Figure 4.7: Parallel simulation using µsik

be found in Perumalla (2005). The following sub-sections explain how the three main components are implemented for the demographic simulation software.

4.3.1 Physical process

The physical process (PP), implemented as class `PopulationSimulator`, is defined as a subclass of class `Simulator`. The main tasks of this class are to: establish the simulation parameters, generate initial population, manage logical processes and generate simulation reports.

Figure 4.9 shows that the simulator will initialize a number of physical processes, each of which will run on a processing element. Then, the initialization of two types of logical processes: `Region` and `FamilyUnit (FU)` will follow. A logical process `Region` represents an administrative area where a number of families live. Each logical process `FamilyUnit` represents a family in the community. In this paper, we use a strict definition of family which will be discussed later (see section 4.3.2). Hence, for each physical process, there will only be one logical process `Region` and a number of logical processes for `FamilyUnit`. Communication between two logical processes occurs when an logical process (in this case a family) from the region in one physical process wants to migrate to another area on a different physical process. We used logical processes of μsik to define the agents in our agent-based model. Therefore, we have two types of agents: regions and family units.

The detailed structure of class `PopulationSimulator` is shown in Figure 4.10.

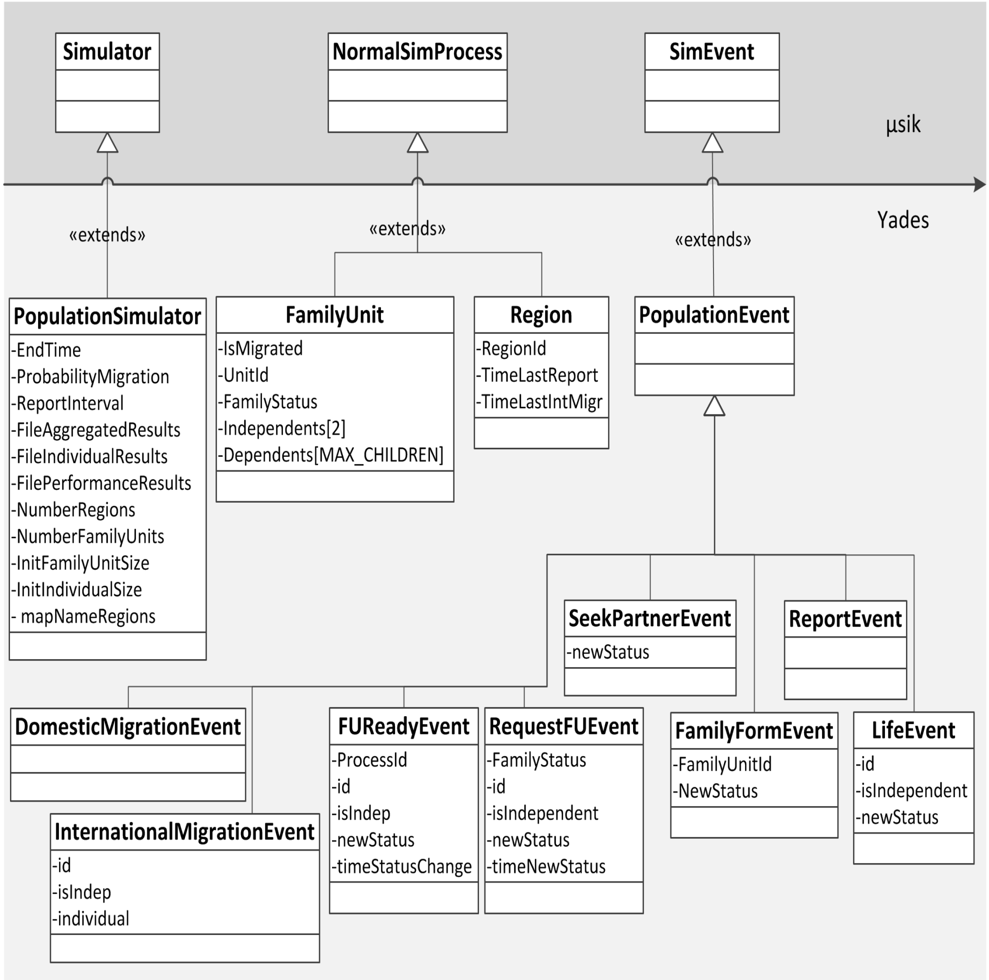


Figure 4.8: UML Class Diagram of Yades

In this class, we use a variable called `endtime` to specify the simulation stopping condition. The number of regions mapped onto a physical process is stored in `nRegions`. The variables for statistical counters, such as average age and average time spent in marriage, are not shown. Any class that inherits µsik class `Simulator` must implement methods `init` and `run`. They are used to initialize the simulation (such as setting up simulation parameters and distributing logical processes across processing elements) and to run the simulation, respectively. The method `Write` is used to produce a detailed report (in CSV format). This report is very useful for model validation or post-analysis using other software (such as Excel and SAS). Finally, the method `GetNewActiveIndividual` is used to add a new individual to a physical process (due to events such as: migration and birth).

One of the modellers' main tasks is to provide data for the initial population

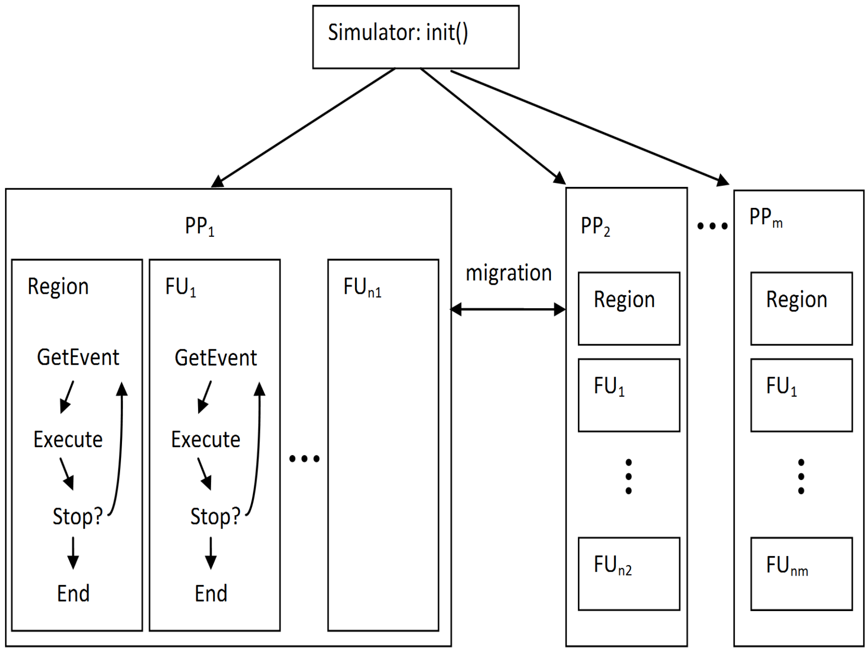


Figure 4.9: Logical process structure and communication

```
class PopulationSimulator : public Simulator
{
public:
  SimTime endtime;
  int nRegions;

  // other variables are not shown

  PopulationSimulator(void);
  virtual ~PopulationSimulator(void);

  virtual void init(int ac, char *av[]);
  virtual void run(void);
  void Write(const PersonalData& p_data,
       const PersonalStat& p_stat);
  long GetNewActiveIndividual(void);
};
```

Figure 4.10: Detailed structure of class PopulationSimulator

in each region. The data includes the proportion of different age groups in the community, the proportion of different types of families by age group, the proportion of different economic status by age group, the proportion of different marital status by age group and the proportion of the number of children in a family.

4.3.2 Logical process: family unit

One of the key design decisions concerns the types of logical processes that are going to be used in the model. We departed from the work presented in Onggo (2008) where each individual was represented as a logical process. This turns out to be problematic. The reason is that public policies may apply to individuals as well as groups of related individuals, such as households and single parents. For example, the UK Department for Work and Pensions and HM Revenue & Customs manage a number of public funds that may apply to individuals (including jobseeker's allowance and incapacity benefit) or groups of related individuals (which could include child benefit and housing benefit). Therefore, it is important that the model recognizes different types of 'policy unit'. Policy unit is often referred to as 'family unit'. A family unit is formed by either a single independent individual or two independent individuals living together (as married, in civil-partnership, or in cohabitation) and any dependent individuals (children). Hence, in this definition, a family unit may represent an independent individual, a single parent, a childless couple or a nuclear family. For completeness, the definition is extended to include orphans, that is, a family unit of dependent children without any parents.

The decision to represent a family unit as a logical process has another benefit. When there is a change in the marital status that affects couples (such as from married to divorced or from married to widowed), only one message needs to be sent to the affected couple. In the earlier work, two messages had to be sent, one for each affected individual. Hence, it reduces the number of sent messages in the simulation.

A family unit may receive events which are related to five demographic components that may change individuals' states. Modellers need to specify models for five demographic components: fertility, a change in economic status, a change in marital status, migration and mortality (described previously in section 4.1). The fertility component determines whether a female individual will give birth, based on the characteristic of the female individual and the current calendar time. The model returns the time when the baby is due. Similarly, the characteristic of an individual and the current calendar time determines a new economic status that will be effective a certain `dt` time from now. The same will happen for a new marital status but this component will take into account if there is a couple. If the new status is either married or cohabitating, modellers need to define the criteria that will be used to match the individual to another individual from the list of prospective partners (i.e. closed marriage model). If a suitable partner is found, then a 'family formation' event will be scheduled for both individuals at time. Otherwise, the individual will be added to the list for a fixed duration. If a partner still cannot be found at the end of the duration, an event will be sent to remove the individual from the list.

Modellers need to specify the model that is used to determine whether a family unit is going to migrate. If the destination is in another country (emigration), the family unit will simply be removed from the simulation and sent to another region. There also new populations of immigrants coming to the current region, which can have different characteristics from the local population. Finally, in the mortality component, modellers need to model the time when an individual will die based on the characteristics of the individual. Commonly used methods, such as life table and survival function can be used for the mortality component. The detailed implementation of logical process family unit is shown in Figure 4.11.

4.3.3 Logical process: region

The second type of logical process represents an administrative area where a number of families live. This logical process will handle domestic migrations, immigration, changes in simulation parameters and periodic reports. Yades allows users to have heterogeneous regions, with different population characteristics. The main limitation of the current version is that it only allows one processing element to run one region. The detailed implementation of class `Region` is shown in Figure 4.12.

A region may receive four types of events. The first event is used when a family unit is going to migrate to a new area. The family unit will send an event to request a place at the destination area (`RequestFUEvent`). The destination area will prepare an empty family unit and send the identification number to the migrating family. Subsequently, all members of the family will be sent to the new location. The second type of event is used to simulate the immigration events, i.e., the number of family units entering the country every month (in batches). This allows modellers to implement different models for immigration policies because it considers heterogeneous numbers and demographic characteristics of the immigrants. Finally, the report event can be used to produce periodical reports, for example, a report on the population structure (by gender, age group, marital status and economic status).

4.3.4 Events

Besides the events already mentioned, in this simulation model, we use the following type of `LifeEvent` to implement demographic processes:

- `ChangeWorkStatusEvent`: an event that changes an individual's work status
- `ChangeMaritalStatusEvent`: an event that changes an individual's marital status

```
class FamilyUnit : public NormalSimProcess
{
private:
  PersonalData data;
  PersonalStat stat;

  // other variables are not shown

    void LifeEvent( LifeEvent* event);
    void FamilyForm( FamilyFormEvent* event);
    void DomesticMigration( DomesticMigrationEvent* event);
    void FUReadyEvent( FUReadyEvent* event);

    void UndoLifeEvent( LifeEvent* event);
    void UndoFamilyForm( FamilyFormEvent* event);
    void UndoDomesticMigration( DomesticMigrationEvent* event);
    void UndoFUReadyEvent( FUReadyEvent* event);

  // supporting methods are not shown

protected:
  virtual void undo_event(SimEventBase *e);
  PopulationSimulator *psim();

public:
  FamilyUnit(const long p_id,
     const int p_auth,
     const bool optimistic);
  virtual ~FamilyUnit(void);

  virtual void init(void) {}
  virtual void wrapup(void);
  virtual void execute(SimEvent *event);
};
```

Figure 4.11: Detailed structure of class FamilyUnit

```
class Region:public NormalSimProcess
{
private:
  vector <PartnerData> male;  // partner list
  vector <PartnerData> female;// partner list

  // other variables are not shown

  void SeekPartner( SeekPartnerEvent* event);
  void RequestFU( RequestFUEvent* event);
  void ReportEvent( ReportEvent* event);
  void InternationalMigrationEvent( InternationalMigrationEvent* event);

  void UndoSeekPartner( SeekPartnerEvent* event);
  void UndoReportEvent( ReportEvent* event);
  void UndoInternationalMigrationEvent( InternationalMigrationEvent* event);

  // supporting methods are not shown

protected:
  virtual void undo_event(SimEventBase *e);
  PopulationSimulator *psim();

public:
  Region (const int p_id,
     const bool optimistic);
  virtual ~Region (void);

  virtual void init(void);
  virtual void wrapup(void);
  virtual void execute(SimEvent *event);
};
```

Figure 4.12: Detailed structure of class Region

- `CancelSeekPartnerEvent`: an event that is used to remove an individual from the list of prospective partners when s/he dies or migrates to another local authority
- `BirthEvent`: an event that creates a new individual
- `DeathEvent`: the death of an individual
- `RequestPIDEvent`, `TransferPIDEvent`, `MigrDataEvent`, and `MigrStatEvent`: are used to implement migration in terms of moving a family unit

We use the competing risk model (see Hosmer, Lemeshow, and May (2008) chapter 9) to ensure that every family unit will have exactly one future event. This approach will sample time-to-event for a number of competing events such as death, giving birth, change in marital status, change in economic status and emigrations. The event with the shortest time-to-event will be chosen and executed. This process is repeated whenever a life event occurs (except for death and emigration). As a consequence of this approach, the model uses a continuous time where future events can happen almost immediately. Hence, the look-ahead is relatively small that makes a conservative protocol less efficient. For this reason, the optimistic protocol is used.

4.4 A Graphical Interface Unit for Yades

Modelling and Simulation platforms are designed to support simulation modelling processes and help modellers to perform challenging models. Platforms are aimed to free users from the unnecessary part of model development processes that can be automated, accelerate the model development process and give a chance to reuse models and analyse results (Kokalj, 2003; Li, Yi, Sun, and Gong, 2012). In the case of agent-based simulation, models have to be specified in computer programs/codes, usually using an object-oriented programming language. For social scientists, this approach is not ideal. To solve this problem, we present a web user interface as a platform for Yades' M&S that is available at `http://yades.fib.upc.edu`. The main advantage of using a web user interface is that it is easily accessible using any supported web browsers.

The user interface is designed so that modellers can define the set of variables and components that will be used in the simulation model. It will also generate the simulation code so that users can download the code to be compiled and run on the target execution platform such as a supercomputer, a cluster of PCs or even a local machine, as shown in Figure 4.1 presented at the beginning of this chapter. Therefore, modellers do not have to worry about the detail on how to harness the parallel computer power.

We will now briefly introduce the state of the art of user interfaces for simulation, justify the need of a graphical user interface for agent-based demographic simulation and describe the design of the proposed interface.

4.4.1 Background on graphical user interfaces design

From the point of view of users, the interface of a simulation tool is the tool itself (Hix and Hartson, 1993), since it is the part of the system with which they interact. In complex systems research, easy-to-use interfaces are specially important since models are difficult enough for non- experts to understand (Saw and Butler, 2008). Human computing interfaces simplify the development of models and relieve social scientists of self-development simulation features, such as simulation input-output procedures (Tobias and Hofmann, 2004).

In Kuljis (1996), issues that influence usability of simulation systems are examined. The author remarks the importance of simulation software design that addresses the specific needs of a certain domain and supports user model development. Therefore, characteristics such as user degree of computer literacy and knowledge or domain knowledge should be taken into account in the interface design (Shneiderman and Plaisant, 2010). Myers, Hudson, and Pausch (2000) state there is always the feeling that a powerful solution to a problem justifies a steep learning curve on graphical user interfaces. However this learning curve can determine the tool's adoption. They point out numerous examples of promising approaches that did not caught on. Among the reasons which determine a tool's success, they point at ease of operation, quality and dealing with diverse dexterity of users. Thus, a good interface should be easy to use, efficient, supportive and satisfy user needs (Smith, 1997). Actually, the development of good graphical interfaces and supportive manuals can be fundamental for software adoption, enabling its diffusion among population both experts and non-skilled users (Bonaccorsi and Rossi, 2003; Murphy and Perera, 2001, 2002; Viorres, Xenofon, Stavrakis, Vlachogiannis, Koutsabasis, and Darzentas, 2007).

Among different kind of user interfaces, web-based approaches for simulation take advantage of a browser to support the interaction of simulation graphical interfaces with users. Actually web-based solutions not only serve as an operating system but also as a distribution channel for applications (Kuljis and Paul, 2003; Macal and North, 2007). Byrne, Heavey, and Byrne (2010) describe several advantages of integrating the Web with the field of simulation, such as the ease-of-use, wide availability, cross-platform capability, collaboration features and control access. Web-based simulation encounters also some disadvantages to consider, including loss in speed, graphical interface limitations, security vulnerability, application stability and license restrictions. Despite these drawbacks, the current tendency is to increase the development of web user interfaces since Web technologies provide means as rich as local PC-based applications (Myers et al., 2000; O'Reilly, 2007). However, areas such as user interface ease-of-use should be further research (Murugesan, Deshpande, Hansen, and Ginige, 2001), particularly in the area of simulation where the number of works in this direction is scarce.

4.4.2 The need of a specialized user interface

To confront the difficulty that social scientists and policy modellers are not familiar with computer programs/codes, we need a specialized GUI which could allow users to easily model different behaviours using familiar methods such as regression or intuitive diagrams such as a state-transition diagram. In that way, we ensure the software design answers the needs and capabilities of the users for whom they are intended Stone and Stone (2005).

Currently, there are some tools that allow the design of generic ABM to be run on parallel environments (seen in chapter 3). However, they required some skills in language programming (Java or C++ mainly). To embrace different simulation models, they provide a desktop user interface where users can define agents' settings and an environment (a grid or a Graphical Information System) where agents can move and a set of simulation displays. Agent behaviour is specified with the corresponding programming language. While this approach is interesting for users who want to use social simulation as a virtual laboratory for their experiments, it is not ideal when users have non-existing or basic programming skills neither when they are not familiar with agent-based models. Moreover, graphical design support attempts (such the ones Repast suite provides) tend to get too complex as soon as models get a particular size.

Social modellers often use statistical estimation techniques to derive into agent decision rules and behaviours. These rules are simple models that relate certain situations with some actions. Key variables that govern agent behaviours have to be identified and statistical relationships have to be obtained (Macal and North, 2005). For example, Yang and Gilbert (2008) give some advices on how to use qualitative data on the design and validation of agent-based models in social simulation. They also point out that in some cases the modellers will have to give some educated guesses to the model based on existing theories in social science, and decide how to quantify some variables.

To summarize, the requirements for the design of the GUI are as follows.

- To design a simple user interface for users who are not familiar with programming and parallel computer environment
- To exploit the potential power offered by parallel computers transparently
- To provide different types of input data: theoretical distributions, multiple regression, logical rules and state-transition diagram
- To produce/show outputs at micro and macro levels

The user interface presented in this research follows what is referred to as a user-centred design approach which involves users throughout the design and development process Stone and Stone (2005). To achieve this, we include the user during the requirement analysis and design. The crucial part, however, is to also involve the user in the evaluation of the system during the whole design and development process.

4.4.3 Graphical user interface design

The demographic agent-based model is made up of five agent behaviours that shape the population dynamics: fertility, mortality, marital status, economic status and migrations. In this subsection, we will explain the GUI design for each part. Before running the simulation, modellers will also need to specify the initial population and simulation settings. The features of the initial population in one area could be different from other regions. Hence, the user interface would also support this requirement. Figure 4.13 shows the internal structure and components of the GUI.

Figure 4.14 shows the sequence of the model implementation wizard that is designed to make the modelling specification process as easy as possible (the demographic modellers, of course, need to prepare the model specification off-line before they enter the specification to the web user interface). The sequence starts with the simulation configuration setting such as the simulation duration. Then, a number of geographical regions will be added. Next, modellers need to specify the initial group settings such as the gender proportion by age groups in each area. Then, the demographic components will be specified: fertility, mortality, marital status, economic status and migrations. To help defining qualitative and quantitative data, users can create distributions, regressions and logical rules in the system, which can be later use to define migration, fertility, economic or partnership changes. A detailed list of what the modellers need to specify for the model components is given in Table 4.1.

Initial population specification

The user interface allows users to specify typical family structures in the region. It includes information on the population distribution by gender and age groups, distribution of different types of family units (parents with/without children, single parents, single individuals, etc.), the age distribution of children in the family, the distribution of different economic statuses, and the distribution of different marital statuses. The GUI allows the information to be entered manually or to import it from a CSV file.

Fertility

Common fertility models to describe the propensity of the women to bear children include age-specific fertility, parity-specific fertility, birth spacing and their combinations. Age-specific fertility uses age to determine the probability of having a child. Parity-specific fertility takes into account the number of children a female person has already had. The birth spacing model focuses more on the time

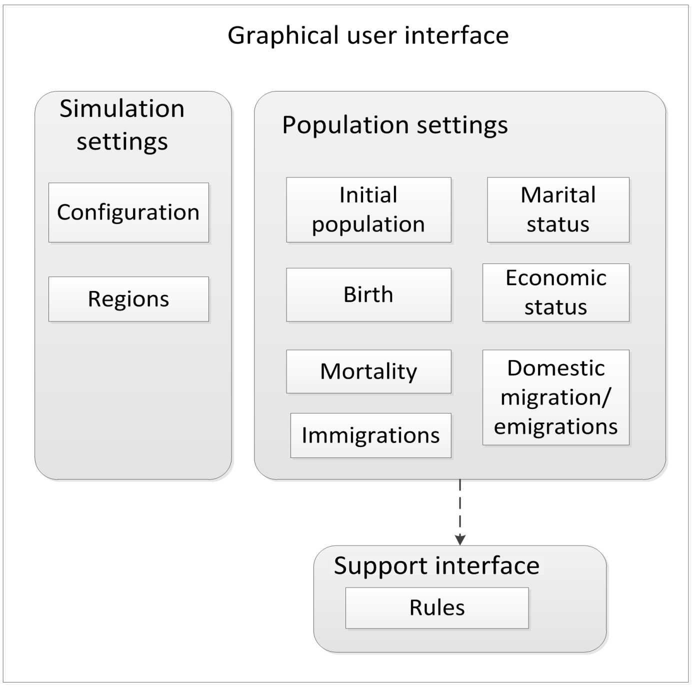


Figure 4.13: Web user interface – modules of the interface

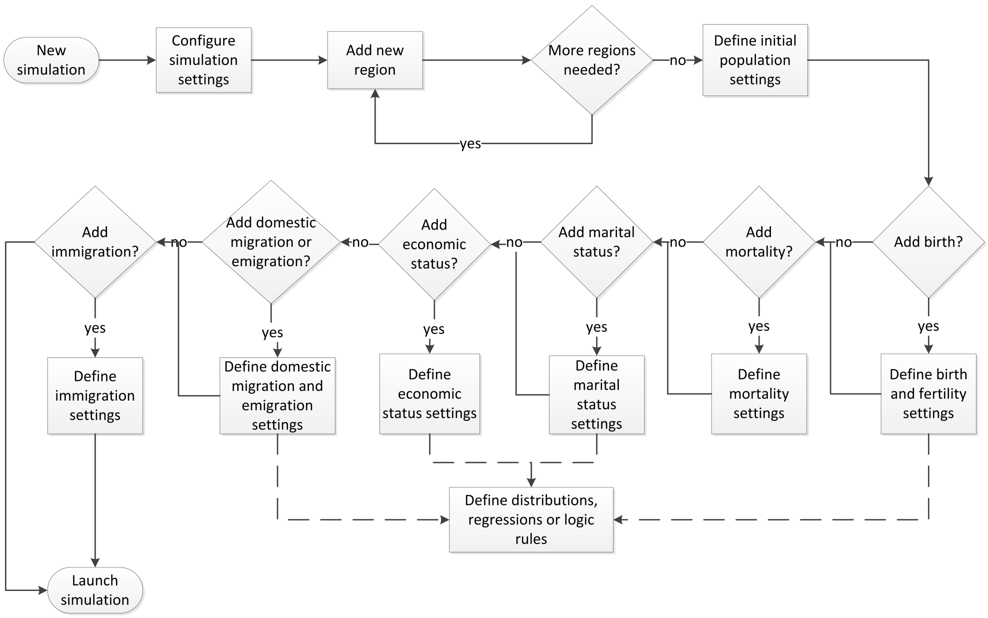


Figure 4.14: Web user interface – flow chart describing the process of creating a simulation model

Configuration	Defines the simulation configuration	Simulation name, number of years, report interval, performance report, individual report, and age groups
Regions	Defines the regions in the model and their settings	Number of regions, homogeneous or heterogeneous setting, region name, family units per region, type of configuration
Initial population	Defines the initial population settings for each region	Population proportions by age group, types of family units by age group, economic status, marital status, children distribution, birthspacing distribution
Birth	Defines birth and fertility settings for each region	Fertility age interval, time to birth
Mortality	Defines mortality settings for each region	Life expectancy at birth or survival function or life table approach
Economic status	Defines economic status settings for each region	Duration in status, transition name, transition origin, transition destination, probability or rule for transition
Marital status	Defines marital status settings for each region	Duration in status, transition name, transition origin, transition destination, probability or rule for transition
Domestic migration/emigration	Defines the settings for domestic migration and emigrations	Logic function to decide whether the family unit is going to migrate, logic function to decide whether it is a domestic migration or emigration, migrations to a domestic destination, migrations to international destination
Immigration	Defines the settings of immigrant population	Number of monthly arrivals of immigrant family units, initial settings for immigrants
Rules	Defines distributions, regressions and logic rules that can be used in the simulation	Distribution name, distribution type, distribution parameters, regression name, regression parameters, rule name, type of result, rule sentence

Table 4.1: Web user interface - definition of modules and parameters

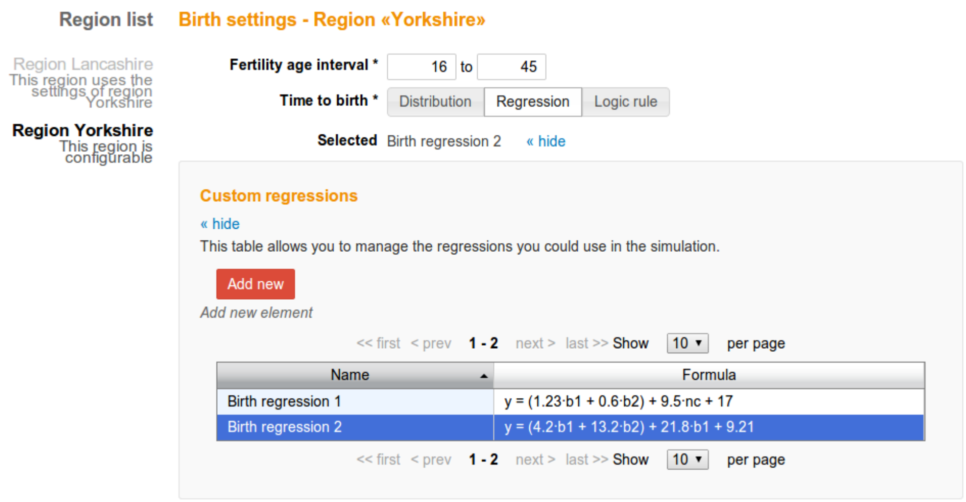


Figure 4.15: Web user interface – view of birth settings

between each birth. This model is rather complicated because it includes factors such as birth control, abstinence period, economic status, etc. The user interface is designed to allow users to specify these models. They will need to indicate the fertility age interval and a model to generate the time to birth. Most of these models can be specified using a theoretical distribution function, a multiple regression function or logical rules based on individual characteristics (see Figure 4.15).

Mortality

The most commonly used mortality models are the life table (and its variants) and survival analysis. The life table summarizes the variation of mortality by age and gender. Survival analysis uses a distribution function to sample an individual's lifetime at birth. The user interface allows users to specify these two commonly used models through entering the life table manually or importing it from a CSV file.

Marital status

In demography, we recognize a number of marital statuses such as single, married and divorced. The GUI is designed to allow users to specify the transition in a state-transition diagram, a graphical tool which helps on defining the multi-state modelling, a standard methodology in demography (Willekens, 2005). Allowing multi-state models enable users to represent different transition models, previously inferred. The difference between marriage and cohabitation has shown to be significant in terms of understanding households dynamics (de Vaus, 2004).

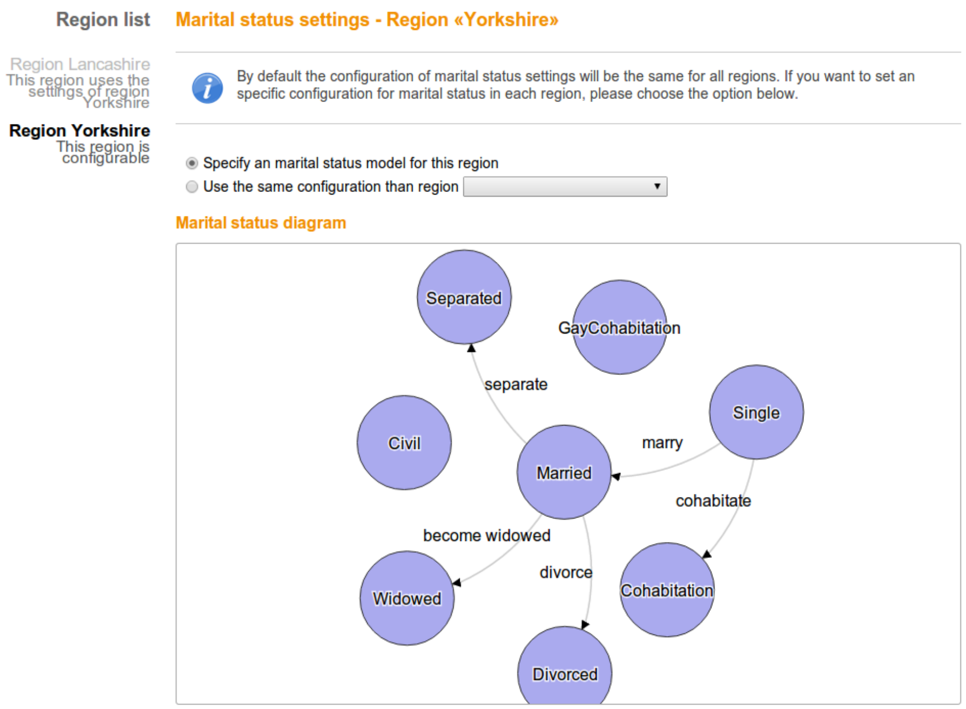


Figure 4.16: Web user interface – view of marital status model

Users will then need to specify each transition and the time to the next state using distribution function, multiple regression or logical rules. Transitions that involve two singles to form a new family unit are treated differently. Modellers will need to specify a "match maker" function, which is based in logic rules according to age, gender, economic status. Figure 4.16 shows how an example of a marital status' state diagram is set in the GUI.

Economic status

The ability to track an individual's economic status is included in the model because there is a plan to build a policy model on top of it (for public policy planning and analysis in such areas as taxation, pensions and benefits). As well as marital status, economic status component is essential for this purpose. The GUI for the change in economic status is very similar to marital status model but in this case we do not need the match maker function.

Migrations

Modellers need to specify two types of migration: national migration and international migration (either emigration and immigration). The user interface allows

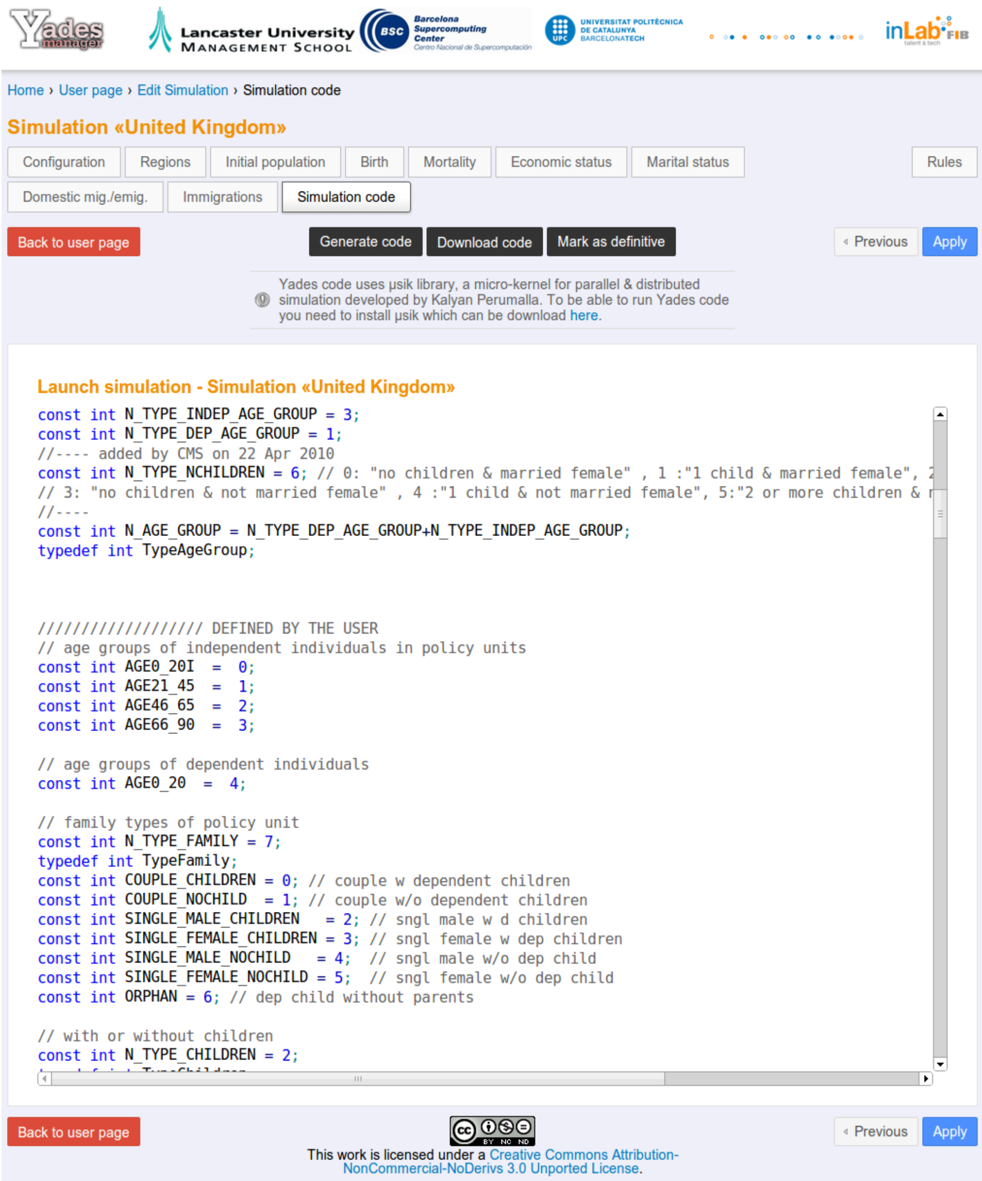


Figure 4.17: Web user interface – screen-shot of simulation launcher and code generation

modellers to specify domestic migration model that determines whether a family unit is going to migrate. Mobility is specified using a constant probability, logit regression or logical rules. For those who migrate, the destination area is determined by a migration matrix. The same user interface is also used for the emigration. Immigrant populations are likely to have different demographic characteristics to the populations they join (Haug et al., 2002). For this reason, users need to specify the number of immigrants per month and their characteristics.

4.4.4 Yades interface code generator

Figure 4.17 shows the interface for code generation. The tool is meant to be run either in a sequential or parallel environment. However, access to a parallel environment is usually restrictive. Moreover, we cannot produce an executable of the simulation because it depends on the architecture is going to be used. Therefore, we provide users the simulation code resulting from their specification of the model. In that way, users with access to a cluster or supercomputer can compile and execute the code. While we free users from the coding process of the model we cannot release them from managing a job in a High Performance Computing environment. Other ABS tools have similar functioning in leaving the management of job's execution to users, though there have been attempts to automatically generate parallel agent-based simulation code (Richmond et al., 2010).

To be able to run the code they need to install libsynk and µsik libraries available from `http://kalper.net/kp/people/kalyan/index.php`.

4.5 Summary and conclusions

We have presented a tool for agent-based demographic modelling and simulation. The main objective of the tool is to tackle the scalability issue in large-scale and complex agent-based models by running the models on top of a parallel discrete-event simulation engine. This tool allows modellers to specify individual behaviour with an agent-based paradigm.

The agent-based model presented in this chapter is able to mimic the human population naturally in the way that individuals can move around and interact with each other and/or the environment where they live. The main reason for an agent-based approach is that it allows to capture the internal structure of the interactions and movements of the population to be studied. In the community of social simulation it is argued that complex systems emerge from the bottom-up, are highly decentralized, and are composed by a multitude of heterogeneous objects (Troitzsch, 2009). This complexity cannot fully be reproduced with a microscopic model so agent-based models seems a better option since it reflects more naturally the typically human capabilities, on the contrary of microsimulation models where actors define their actions based on stochasticity only or macro simulation models which cannot content most of the finding in social research. Moreover, the aggregation of people with the same characteristics would decrease the capacity of the social scientist to understand the processes behind the individual/family/household/organization dynamics.

The agent-based approach can subtly model population changes and the heterogeneous behaviour of the individuals among a large population, characteristics that are important to reproduce demographic processes (such as marriage or migration) where the interactions and behaviours are playing an important role. Since our goal is to use this methodology to experiment with social policies, it is highly important for us to be able to reproduce the complexity at a macro-level from an agent-based level. Moreover, agent-based technology is able to cope not only with quantitative data but also with qualitative information coming from ethnographic records, a characteristic that is especially relevant for demographics. Although the framework is oriented to study demographic dynamics, we expect to add new social models on top of it where the agent-based perspective can be key.

Yades uses an agent-based model to model fertility and birth, mortality, economic status, marital status, and migration demographic processes which permits individuals to flexibly move and interact in a geographical environment:

1. It uses a rich set of attributes which come not only from census or surveys data sources, but also with behavioural rules that help to overcome some data-related limitations of over-reliance on purely statistical information. Individuals attributes change over time due to their demographic evolution.

2. There is no central unit that controls interactions or behaviours of the population

3. Agents behave autonomously according to their own rules

4. Agents do not have global information, they take actions according to simple rules that are based on local knowledge

As a demographic planning tool, Yades is able to monitor the evolution of population structures and demographic changes of a set of geographical regions over time. The model is useful for providing information for demographic planning or policy making, particularly for location-based policies. For instance, it can give us information about migrations patterns which can help calculate their impact in the supply of health or public transport services. Therefore, the tool is useful in assisting decision making,exploration of "what if" situations and testing different hypothesis.

We believe the two of the main issues in the wider adoption of parallel simulation are scalability and ease-of-use. This tool confront both issues by proving a graphical user interface that allows modellers to specify individual behaviour such as fertility and change in marital status using agent-based simulation modelling paradigm and transparently generated codes that enable the model to run on top of a parallel simulation engine hosted in our supercomputer infrastructure. The user interface is designed for social scientists and modellers who are not trained in parallel programming. Hence, it will provide them with familiar modelling tools such as regression editor, an easy-to-use state-transition diagram, and a simplified version of logical rule editor. We believe that this will allow users to concentrate on understanding the modelling process rather than simulation library it is being used. We believe this will help to remove a major barrier on using simulation. However, we are aware technical knowledge is necessary to execute scenarios in High Performance Computing facilities but this limitation is something that all parallel social simulation tools must confront.

Chapter 5

Yades performance analysis

In this chapter, we present the results of some experiments done to study Yades performance under varying conditions. The purpose is to understand how the tool behaves from the technical point of view and how sensible it is to different computer architecture parameters. In chapter 3, we saw agent-based parallel simulation scalability is feasible but depends on some factors: the size of the model, the agents' complexity, and the representation of the environment. At a computer architecture level, the simulation execution time depends not only on the computation time but also on the communication time. Therefore, synchronization is key to scalability because it requires communication in the network. The speed of these communications has to be also taken into account.

To analyse Yades performance, we carried out two experiments. First, we studied the effect of population size and migration activities on performance and scalability. Second, we tested the performance of the tool on different execution settings: homogeneous environment, heterogeneous population size, heterogeneous processing speed and heterogeneous communication latencies. Heterogeneous environments are interesting for us since it is common for data centres to be often upgraded by replacing the part of the infrastructure with the latests processors and memories. Thus, the response of the framework to heterogeneous processing speed and heterogeneous communications can affect not only computation time but overheads.

All experiments were run using μsik settings that gave a roll-back based optimistic parallel simulation execution with a state-saving mechanism and a time window of 12 months. The model uses a continuous time where future events can happen almost immediately. The lookahead is relatively small that makes a conservative protocol less efficient. For this reason, the optimistic protocol is used. Fujimoto (2000, chap. 4,5) provides a good overview of various techniques in optimistic parallel simulation.

The second experiment was initially done in HEC cluster at Lancaster University, UK and presented in (Onggo, Montañola Sales, and Casanovas-Garcia, 2010). Later we studied the scalability of the tool on Marenostrum 2 and 3 supercomputers in Barcelona (Spain) and repeated the second experiments to compare the results in the same architecture. Those experiments were published in Montañola Sales, Casanovas-Garcia, Kaplan-Marcusán, and Cela-Espín (2014b) and are presented in this chapter.

The program was compiled using gcc version 3.3.5 with the optimization flag O3 turned on and mpich version 1.2.7 was used. The experiments were run on Marenostrum 2 and 3 supercomputers, with high-speed Myrinet and Infiniband interconnections respectively. Each node in Marenostrum 2 has two dual-core PowerPC 970 CPUs with a frequency of 2.3GHz and 8GB of memory. Each node in Marenostrum 3 has two 8-core Intel SandyBridge-EP E5-2670 with a frequency of 2.6GHz and 32GB of memory. In the experiments, we used up to 256 processors. All performance results presented in this section are based on the average of five replications. Because the standard deviations are very low, we did not need more than five replications for each experiment. The results of the parallel simulation have been checked against the sequential execution for correctness.

The rest of this chapter is organized as follows. In sections 5.1, we will provide some definition of concepts that are used in this chapter and necessary to understand our work. Then, section 5.2 and 5.3 present the first experiment to measure the effect of population size and migrations in execution time and scalability. Afterwards, the second experiment follows showing the performance of the tool in a homogeneous environment (section 5.4), a heterogeneous population size (section 5.5), a heterogeneous processing speed (section 5.6), and a heterogeneous communications (section 5.7). Finally, we will remark some conclusions of our job.

5.1 Some definitions

The following list consist of a syllabus of definitions that can help the reader to understand the work presented in this chapter.

- **Time window** is the amount of time a simulation entity can advance ahead of others.
- **Elapsed time** is the time for an event to occur.
- **Speedup** is a metric that indicates how much a parallel algorithm is faster than a corresponding sequential algorithm.
- **Simulation bandwith** is the number of events that a simulation executes per second.

- **Latency** is the communication time that is spent when a message is send in a network, from its origin to its final destination.

- **Scalability** is the ability of a system to handle a growing amount of work and accommodate it. When a system is said to scale it means it is efficient when simulating it at a large-scale. So, a simulation model scalable should be able to profit a growing amount of computers to run faster.

- **IPC** is the number of instructions a processor is able to run per cycle.

- **Cache** is the memory with faster access. The processor uses it to reduce the average time to access data from the main memory.

- **Cache miss** is an attempt to read or write a piece of data in the cache that failed. As a result, the processor goes to search the data in the main memory with much longer latency.

- **Throughput** is the amount of data delivered over a communication channel.

5.2 Effect of population size on execution time and scalability

Perumalla (2005, 2007) has carried out a number of experiments to evaluate the performance of μsik simulation library. Hence, we do not repeat it in this dissertation. The focus of the following experiments is to understand the effect of population size on the overall simulation performance. We disable the migrations to measure the effect of the number of family units on computation time. The simulation for a period of 30 years was run with different initial population sizes of 80,000, 160,000 and 320,000 family units. The number of individuals is approximately twice the number of family units. Since the average fertility rate is set to be around two with no immigration, the numbers of individuals at the end of the simulation are approximately the same as their initial size. The result ran in Marenostrum 2 is shown in Figure 5.1. As we can see, it shows super linear speedup, meaning the double of processors increases the speedup more than twice. For example, a simulation of 320,000 families in two nodes has a speedup of 4.66 while in four nodes it is 22.63, showing an improvement higher than two which would be expected in an ideal linear scalability.

This superscalar behavior can be explained due to cache memory problems. To illustrate that, we observe the behavior of cache miss ratio and IPC (instructions per cycle) in our application using use Paraver performance analysis tool (Pillet, Labarta, Cortes, and Girona, 1995). Besides the studied population sizes, we run lower population sizes: 40,000, 20,000 and 10,000 family units. All the executions were made in one node. First, we observe the number of miss ratio produced in

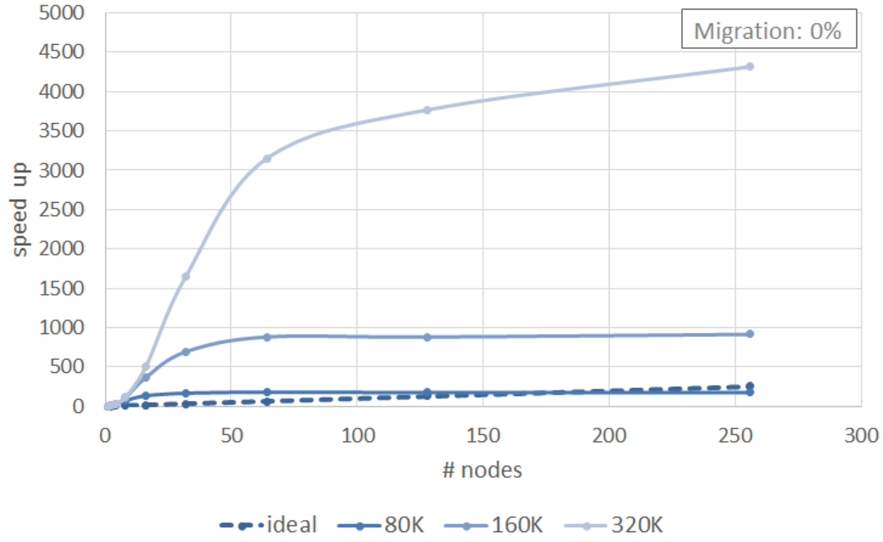


Figure 5.1: Effect of population size on speedup

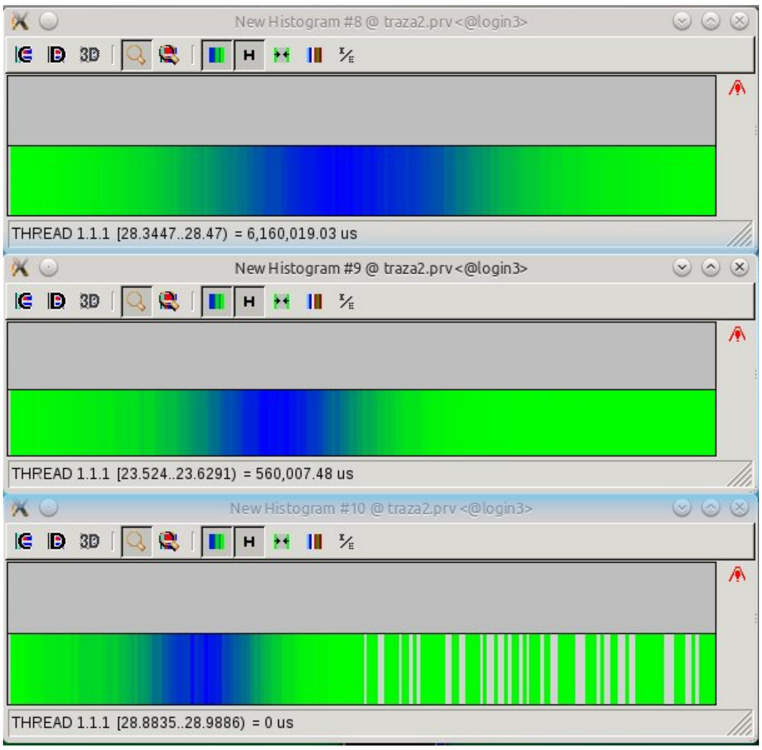


Figure 5.2: Paraver miss ratio histogram in L2 cache for different population sizes scenarios (10,000-20,000-40,000 from bottom to top) in one node

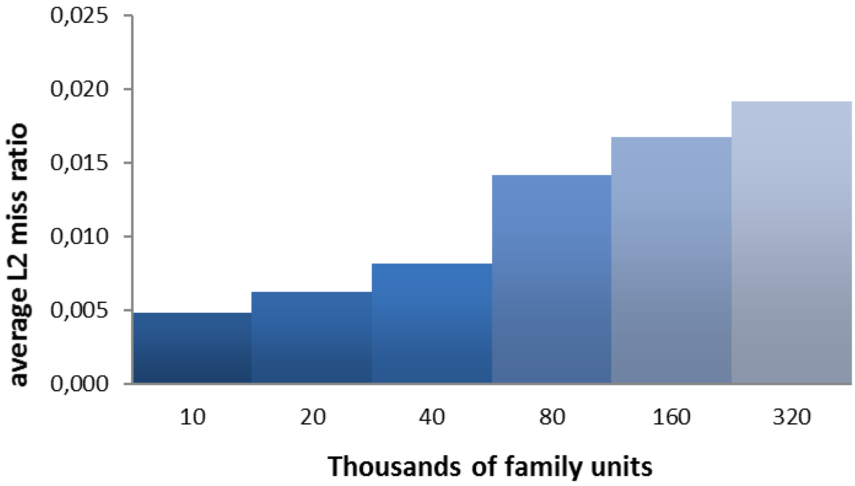


Figure 5.3: Miss ratio observed in L2 cache for different population sizes scenarios in one node

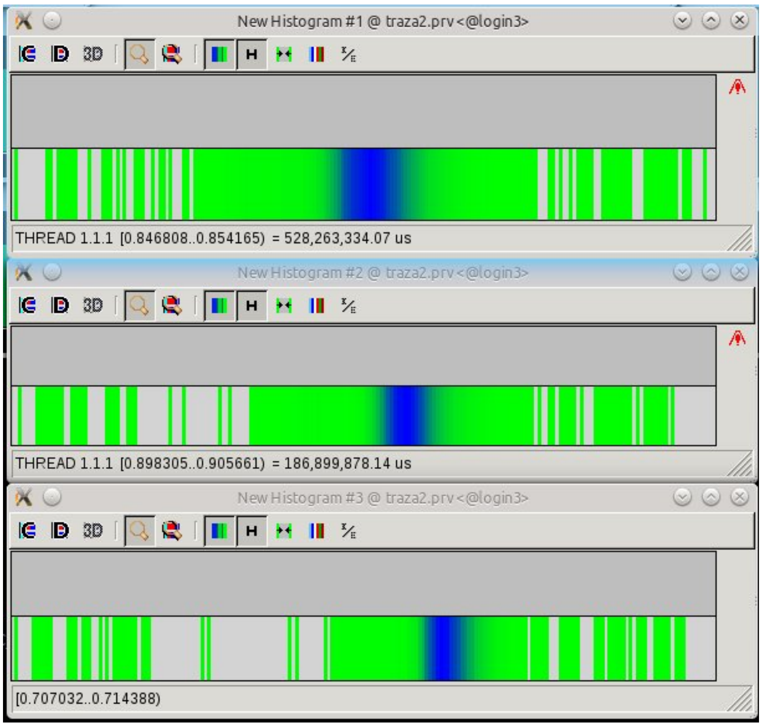


Figure 5.4: Paraver histogram of IPC obtained varying population size scenarios (10,000-20,000-40,000 from bottom to top) in one node

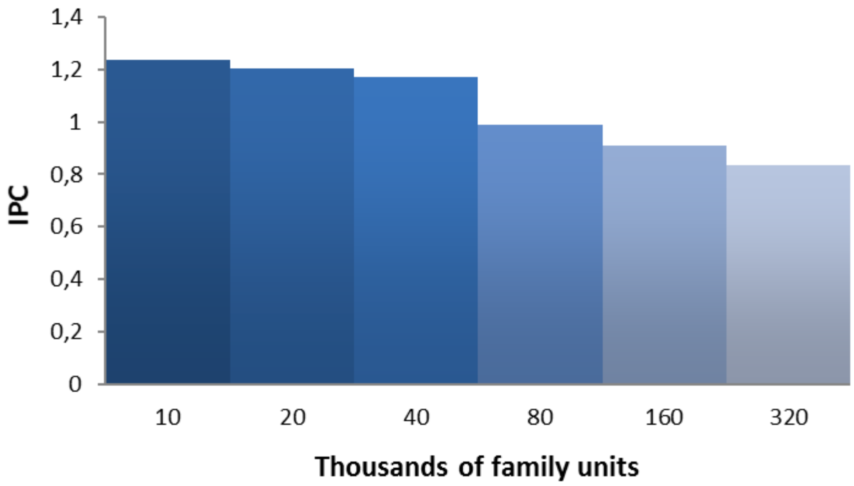


Figure 5.5: IPC obtained for different population size scenarios in one node

different population sizes (see Figure 5.2). The histogram in this figure shows the time the simulation spends in miss ratio at L2 cache level. We choose L2 cache level because it is the most representative in Marenostrum 2. The fact that we see, from bottom to top, that the highest concentration of miss ratio (blue zone) is growing and moving to the right means that for an increasing population size the average miss ratio in L2 cache level is higher. Then, we plot the average miss ratio seen in L2 cache level in Figure 5.3. We confirm that average miss ratio increases with population growth, particularly when we go from 40,000 to 80,000 family units.

Second, we studied IPC behavior. We first plot the histogram of IPC in different population sizes (see Figure 5.4). Here we see how an increase of population sample, affects the IPC of the application. The highest concentration of IPC in our application (blue zone) is moving to the left as we increase the population size (from bottom to top). That means, when we increase the number of families in the simulation, the IPC decreases from 1,2 to 0.9. The IPC in Marenostrum 3 is 4 but most of the applications reach an IPC of 1.5. Then, we plot IPC in Figure 5.5 observing a substantial decrease in IPC when we increase the number of families, especially beyond 40,000. These results prove our application is very sensitive to the size per node, since memory is determinant of performance which explains the super linear speedup found in our experiment.

To deep in the impact of different population sizes in performance, we calculate the effect of increasing the number of nodes and the population size in proportion (weak scaling). We used three different population sizes configurations per node: 40,000, 20,000 and 10,000 family units. Due to Marenostrum 2

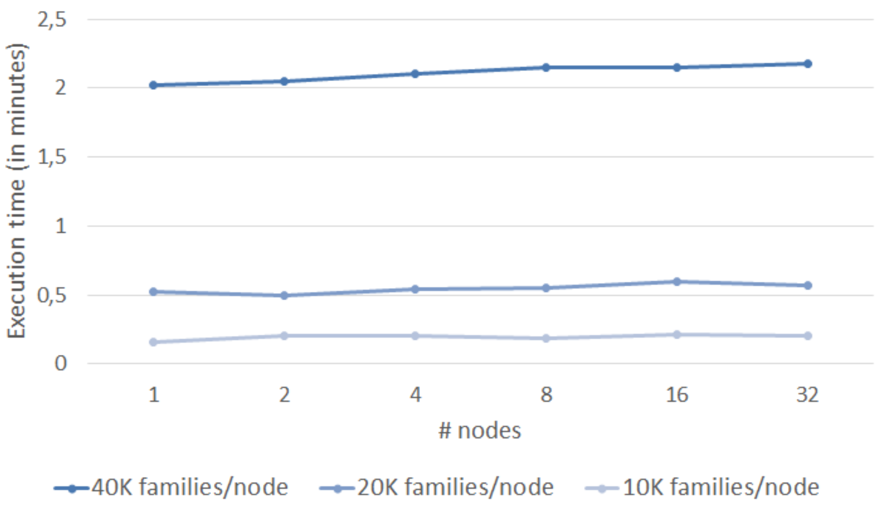


Figure 5.6: Weak scaling

machine decommissioning, we perform the experiment in Marenostrum 3. Despite the difference on the architecture, the results (shown in Figure 5.6) are representative. We can see execution time remains stable when increasing the population and nodes. That means our application is scaling well as we increase the problem size when there are no migrations. Moreover, here we also observe the effect of cache misses on execution time. Looking at the results in one node, going from 20,000 to 40,000 population size the execution takes more than double time (from 31.45 seconds to 2.02 minuts) in comparison to moving from 10,000 to 20,000 families (with a difference of 21.87 seconds).

5.3 Effect of population size and migrations on scalability

The objective of this experiment is to study the effect of varying the number of nodes on execution time and speedup. Both experiments were run in Marenostrum 2.

In the first part, the effect of migrations on execution time is measured. The simulation started with 320,000 family units and was run for a period of 30 years. We varied the number of processors from one to 256 using base-2 logarithmic scale. We considered migrations with probabilities from 0% to 60%. The probability of migrations determines the probability of a family unit to migrate when there is a change in the employment status of one of the parents. The results are shown in 5.7. On one hand, as we can see the execution time decreases as we increase the number of processors. Due to the increase cache misses and the IPC decrease

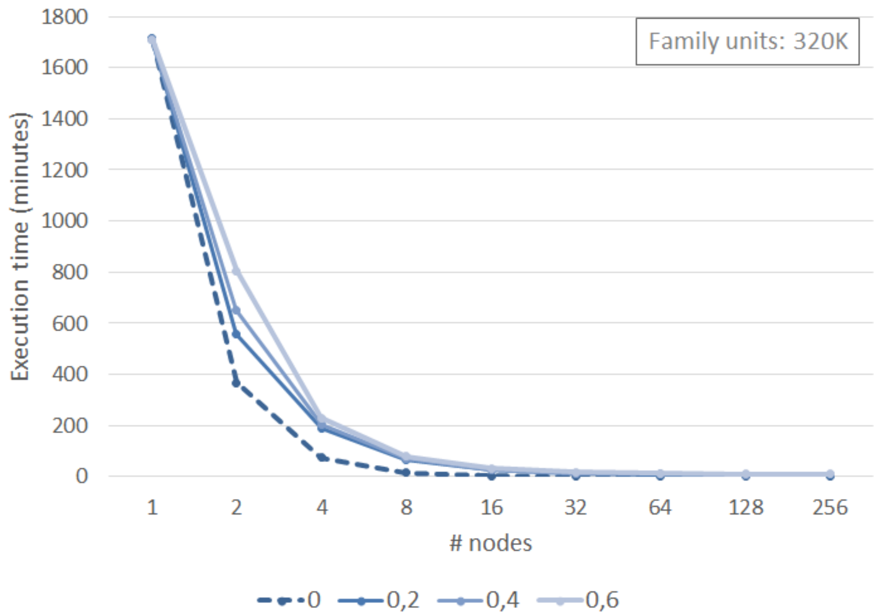


Figure 5.7: Effect of population degree of migration on execution time

shown in Figure 5.3 and Figure 5.5, the execution time is reduced more than half as we double the number of processors. Hence the simulation is slower for big scenarios in few nodes due to the limitations on node memory. On the other hand, the reduction in the execution time becomes less significant as the number of nodes increases. This is because the reduction in the computation cost becomes less significant and at the same time the communication costs becomes more expensive as the number of processors increases, due to migrations and rollbacks.

In the second part, we study the speedup for a fixed population size (strong scaling). Figure 5.8 displays the performance improvement when the simulation is run for 30 years with 320,000 family units running on 1 to 128 processors. The speedups grow when the number of nodes increases. Ideal line shows the linearity. The result shows the super linearity observed in section 6.1 due to cache misses. Moreover, for the same problem size, an increase in the number of processors increases computing power, but at the same time more synchronization overheads are required. This explains the diminishing performance gain as we increase the number of nodes. It can also be seen in Figure that the smaller the proportion of family units who are going to migrate, the larger the speedup obtained. This is because a high proportion of migrations increases the number of inter-processor communications when migrations take place and, consequently, the rollbacks. This also explains why the performance gain diminishes faster for the higher proportion of migrations.

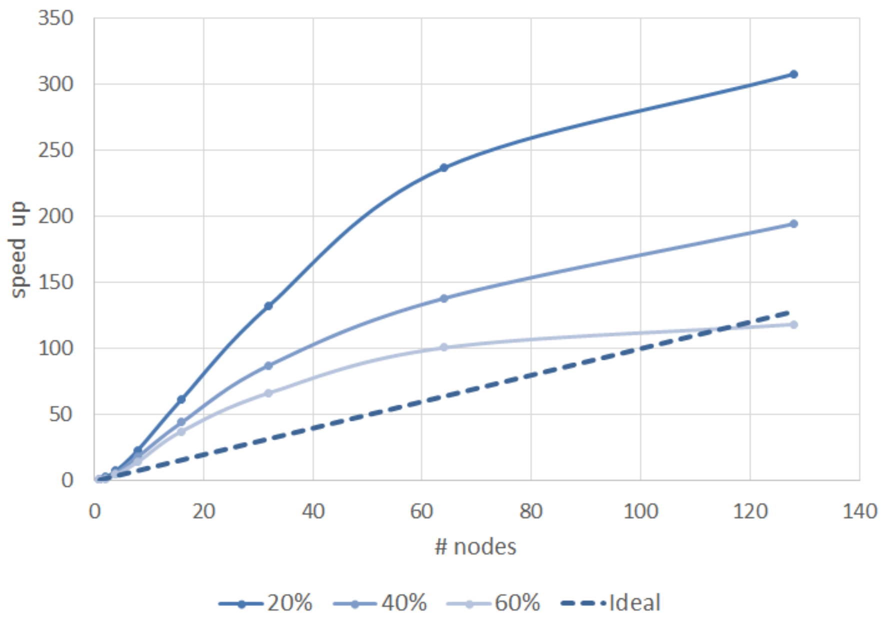


Figure 5.8: Strong scaling with migrations

5.4 Homogeneous environment

The objective of this experiment is to understand the effect of migration activities on the performance of the tool, specifically the execution time and the number of rollbacks, under an ideal execution configuration. In this configuration, we ran the simulation in Marenostrum 2 for a period of 30 years with an initial population size of 320,000 family units (around 650,000 individuals), divided equally among all administrative areas. This would produce a homogeneous workload to all processors. The simulation was run on one compute node containing four processors to minimize the effect of heterogeneous communication latency. The probability of migrations was varied between 0% and 60%. As explained earlier, migrations are responsible for all inter-processor communications in the simulation.

Probability	0%	20%	40%	60%
Average number of migrations (individuals)	0	324,801	642,065	946,524

Table 5.1: Average number of migrations in 30 years

The results are shown in Table 5.1 and Figure 5.9. As expected, the number of migrations is proportional to the migration probability (Table 5.1). Figure 5.9 shows that the increase in the number of migrations increases the execution time. The increase in the number of migrations increases the number of event

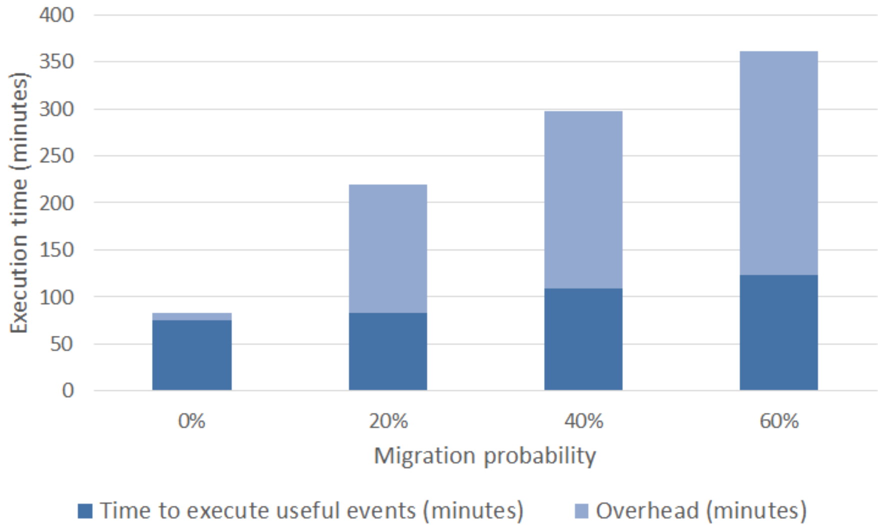


Figure 5.9: Execution Time for Homogeneous Workload

that has to be executed by the simulator. As a result, it requires more time to execute all useful events. In this configuration (homogeneous environment), the average number of rollbacks is close to zero regardless of the migration probability. This indicates that each processor has enough computations and advances its simulation clock slower in such a way that the migrations seldom cause any rollbacks. Consequently, the overhead costs are mainly due to the inter-processor communications and rollbacks.

5.5 Unbalanced population size

In practice, the number of family units may vary across administrative areas. Hence, it is important to measure the effect of an unbalanced distribution of family units on the performance of the tool. As in previous experiments, we set the simulation duration to 30 years but fixed the probability of migrations to 60%. We ran the simulation on four processors in one compute node of Marenostrum 2 with a total of 320,000 family units at the start of the simulation. We varied the distributions of the family units from a balanced distribution of 80,000 family units on each processing element to a rather unbalanced distribution of 215,000 family units on one processing element and 35,000 family units on each of the remaining processing elements (see Table 5.2).

The four configurations are arranged in different columns in Table 5.3 . Row 2 shows total number of migrations. As expected, the total number of migrations is roughly the same regardless of the distribution of the family units. Row 3 onwards shows the total number of rollbacks.

Configuration	Processor 1	Processor 2	Processor 3	Processor 4
Balance	80,000	80,000	80,000	80,000
Low unbalance	125,000	65,000	65,000	65,000
Medium unbalance	170,000	50,000	50,000	50,000
High unbalance	215,000	35,000	35,000	35,000

Table 5.2: Configurations of different population size accross processors

Workload distribution	Balance	Low unbalance	Medium unbalance	High unbalance
Number of migrations (individuals)	946,524	947,359	946,532	948,299
Total rollbacks	0	252,397	348,495	474,626

Table 5.3: Effect of Unbalanced Distribution of Family Units on Performance

Figure 5.10 shows that the equal distribution of family units across processors results in the best execution time. The worst execution time (almost two times slower) was given by the most unbalanced configuration in the experiments (configuration High unbalance). This result is consistent with what has been reported in parallel simulation literature, i.e. an equal distribution of family units will result in an equally distributed workload across the processors. Consequently, the processors can advance their simulation clock at a similar pace, which reduces the number of rollbacks.

A processor with the highest workload in the more unbalanced configuration has to execute more events. This explains the increase in the amount of time spent for executing useful events. In contrast, a processor with the lightest workload in the more unbalanced configuration will execute fewer events. Consequently, it spends less time executing useful events.

In each of the unbalanced configurations, the processor with the higher workload will advance its simulation clock slower than the other processor; hence it will not experience any significant number of rollbacks. Consequently, its overhead can be attributed mainly to the communication costs other than rollback, such as waiting for events from another processor. The busier the processor, the less time is spent on waiting, which explains the decrease in the time spent for overhead. On the other hand, processors with lighter workload execute fewer events so they may advance their simulation time ahead of the busier processor. As a result, they have to rollback more often (see Table 5.3). This explains the increase in the time spent for overhead at the less busy processing elements.

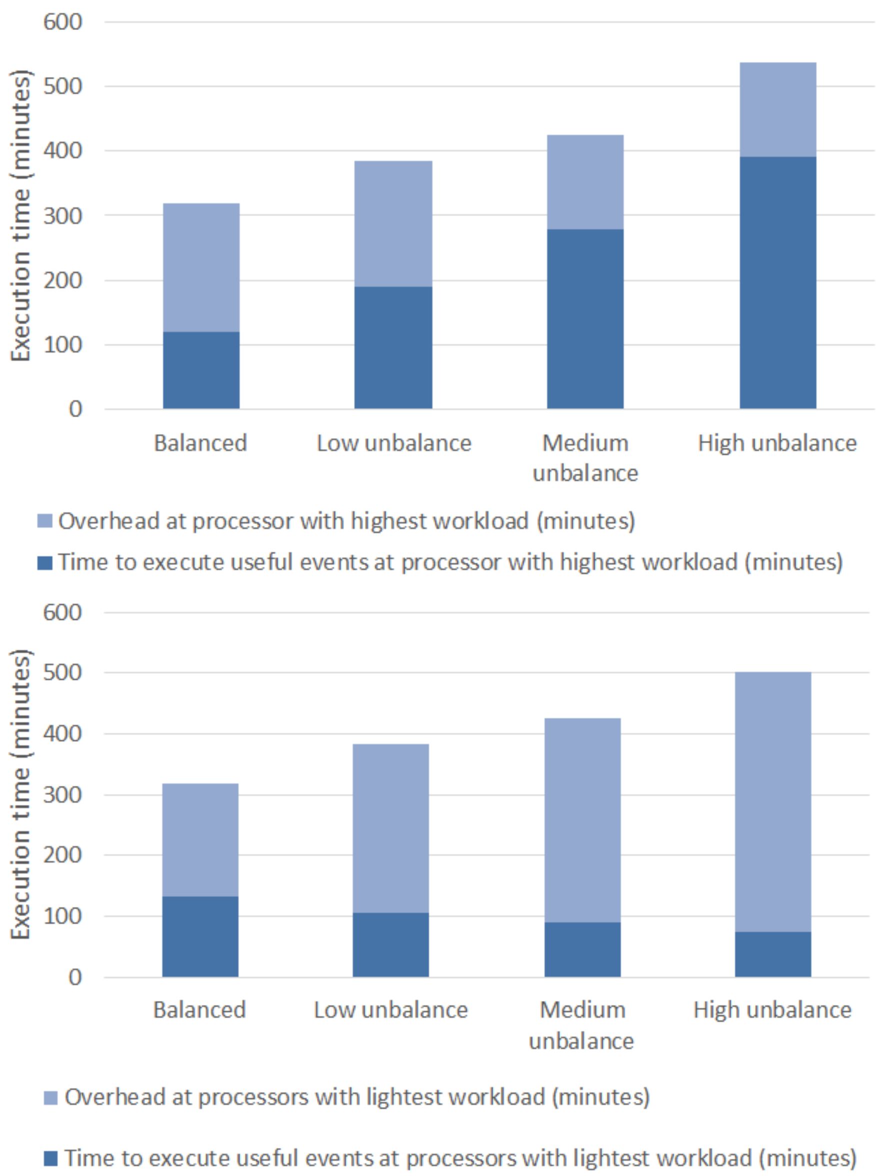


Figure 5.10: Effect of Unbalanced Distribution of Family Units on Performance

5.6 Heterogeneous processing elements

In this section, we measure the effect of using heterogeneous processors on the performance of the tool. In the experiment, we ran the simulation in Marenostrum 2 for a period of 30 years with an initial population size of 320,000 family units, divided equally among all administrative areas. The probability of migrations was fixed at 60%. To emulate the difference in processor speed, we inserted a delay for every simulation year at one of the processors (1 second and 2 seconds for each experiment, respectively). This is done by adding a delay to the event that generates an annual report. The result is shown in Table 5.4. The result is consistent with what has been reported in literature on parallel simulation, i.e., the wider gap in processor speed will result in more rollbacks (see the last row).

Delay (seconds)	0	1	2
Number of migrations (individuals)	946,524	946,532	947,327
Time to complete simulation (minutes)	361.1	429.1	480.2
Total rollbacks	0	252.4	348,495

Table 5.4: Effect of Unbalance in Processor Speed on Performance

5.7 Heterogeneous communication latency

Finally, we are also interested in the effect of heterogeneous latency in the communication between processing elements. The event size used in Yades is 512 bytes. For this event size, we used the Intel MPI Benchmark Suite to measure the inter-node latency and intra-node latency in Marenostrum 2 and found that the inter-node latency was 4 times slower than the intra-node latency. In the experiment, we used the same configuration as in the previous experiments but without any delay. We varied the locations of the four processing elements used in the experiment: using one compute node with four processing elements, using two compute nodes with two processing elements each, and using four compute nodes with one processing element each. The performance result is shown in Table 5.5. As expected, the number of migrations is about the same (row 2). The time spent in executing useful events is roughly the same because we expect similar number of useful events (row 4). The last two rows show that when the latency is homogeneous, the number of rollbacks is zero (row 6). As a result, it incurs some additional overhead cost (row 5). The overall performance (row 3) shows that a configuration with heterogeneous communication latencies (2×2) performs worse than a configuration with higher but more homogeneous communication latencies (4×1) due to rollbacks. However, the difference in performance is not very visible because the inter-node latency and intra-node latency are within the same order of magnitude.

Nodes × Processors	1×4	2×2	4×1
Average number of migrations (individuals)	947,327	946,990	949,679
Average time to complete simulation (minutes)	358	363.4	359.6
Average time to execute useful events (minutes)	124.1	124.2	122.7
Average overhead time (minutes)	233.9	239.2	238.7
Average number of rollbacks	0	6.4	0

Table 5.5: Effect of Unbalance in Communication Latency on Performance

5.8 Conclusions

We have presented the performance evaluation result of Yades. The study was done first in the HEC high performance computing cluster at Lancaster University and later in Marenostrum 2 and 3. The performance measures such as speed-up and scalability of the tool using up to 256 processors showed the potential of parallel simulation for large-scale scenarios. Those experiments in different computer infrastructures showed us the application is sensible to the architecture where it is run in terms of memory consumption when the population size per node is high. However, when analysing weak scaling we saw the application scales well when independent of cache issues. Moreover, migrations have a deep effect on performance due to their impact on network communications.

We also conducted some more fine grained performance measures such as time spent in executing useful events, time spent for overhead and the number of roll-backs. Specifically, we have investigated the effect of three factors: unbalanced workload, heterogeneous processing speed and heterogeneous communication latency. The results are consistent with what has been reported in other application areas where parallel simulation has been used. Since the application of parallel simulation in demography is new, it is useful to quantify the effect of the three factors on performance. The findings are useful because it is likely that the simulation users will run the tool using heterogeneous population configurations.

Chapter 6

Yades validation

Validation and verification are two of the crucial parts in the process of performing a simulation study. The reason is these techniques help on increasing the confidence in the model, since it is not possible to demonstrate its absolute validity in all contexts. However, unlike many other fields such as Physics or Economics, there is no established standard for the validation of social simulation models. The social simulation community is very concern about that since this lack complicates the acceptance of social simulation models by those outside the field.

This chapter presents the validation of our proposed agent-based demographic model. First, we discuss the verification and validation and validation of simulation models in Section 6.1. Particularly we will focus on the factors that should be taken into account when validation agent-based models. Then, we present the validation of our model following a white-box approach (Section 6.2). Although this work is far from a complete validation of the tool, we believe it will contribute to increasing the acceptance of our work to final users.

6.1 On verifying and validating simulation models

Validation and verification (V & V) is a significant element of any simulation study. As pointed by Robinson (1997),

> “”without V & V there are no grounds on which to place confidence in a (simulation) study's results”

In simulation, we often differentiate between verification and validation. Verification is a process to determine whether a conceptual model has been implemented correctly in its computerized form. To borrow the computer programming

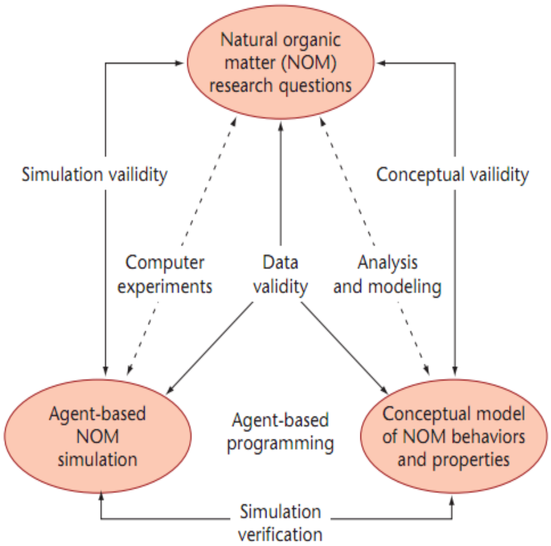


Figure 6.1: Validation and verification processes of the agent-based social model

term, we debug the model. Validation is a process to determine whether the model is an accurate representation of the system being studied for a given set of modelling objectives. A simplified version of agent-based simulation models of Sargent (2005) is shown in Figure 6.1. In this figure, we can see verification confronts the conceptual model with the simulation model, while validation is done at two different levels: operational and conceptual. In fact, we can look at validation as an *a posteriori* constraint while verification as *a priori* constraint (Moss, 2001). The reason is verification limits the specification of the model, establishing some logic or theory limits before generating any output from the simulation. However, validation aims to adjust the model (through simulation outputs) with an observation. A failure on validating a verified model would imply to modify the model itself. As a consequence, either the model should be changed or the theory or formalism should be revised.

Although conceptually simple, verification can be challenging, especially when we are dealing with a relatively complex computer program. Law (2007) and Banks, Carson, Nelson, and Nicol (1999) lists a number of techniques that can be used in a verification process. Validation is neither an easy job. Actually, Robinson (1997) states that it is not possible to prove that a model is valid in all contexts, because a model is only a simplified version of reality. Consequently, the model cannot describe the whole complexity of a real system. Hence, the main objective of validation is to prove that a model is *sufficiently* accurate for parts of the real world under study. Indeed, one of the key aspects of validation is to assess whether the outcomes of a model can explain the real phenomenon that is being studied (Ormerod and Rosewell, 2009). This issue can be fulfilled by performing as many validation methods as possible during a simulation study until we (and users) can gain enough confidence in the model and accept its results. Therefore, validation is a continuous process (Edmonds, 2001). Validation should also take into account the domain of the system under study (Sargent, 2005). Therefore, a validated model may not be valid for a set of different experimental conditions outside its domain.

(Robinson, 1997) identifies four different forms of validation in simulation modelling: conceptual model validation, data validation, white-box validation and black-box validation. Conceptual model validation deals with issues such as the level of detail of the model and determines if the model is enough for the purpose it was developed. Data validation is needed to determine whether the data used in the simulation study is sufficiently accurate. The black-box validation concerns with the relationship between inputs and outputs of the model, ignoring the elements inside. The objective is to determine if the output of the model reflects the real world observation for the same set of inputs. Finally, white-box validation tries to answer the question *does each element of the model, and the structure of the model elements represent the real world with sufficient accuracy?*

In the following subsections, we will point at some specific issues of verification and validation processes in agent-based social models and the role of formal languages in these processes.

6.1.1 Validating agent-based models

In the area of agent-based simulation applied to social sciences and humanities validation is a big issue of concern. Despite the increasing popularity of agent-based simulation in the last two decades, validation techniques are neither as widely used nor as formalised as one would expect. According to a survey conducted by Heath et al. (2009) on the articles related to agent-based models published between 1998 and 2008, 29% of the articles did not even discuss the validation of their models. Moreover, their findings showed only 17% of the articles used solely the conceptual validation, while 19% compared simulation results with real observations (operational validation). 35% used both approaches. Beyond that, they observed qualitative methods dominated the validation of agent-based models, possibly because many models are not conducive for quantitative validation methods. Agent-based models often exhibit behaviour that can be problematic for validation purposes, such as non-linearities and multi-level properties (Klügl, 2008). Furthermore, they also require the finer level of model detail in which data at that level of detail may be difficult to obtain. In general, agent-based models often use significantly more assumptions making the assessment of the validity more difficult.

Kennedy, Xiang, Madey, and Cosimano (2006) suggest the lack of verification and validation for agent-based and social sciences could be attributed to agent-based modelling not being as mature as engineering modelling. Klügl noted that agent-based models often exhibit behaviour that can be problematic for validation purposes, such as non-linearities and multi-level properties (Klügl, 2008). In addition, agent-based models often use significantly more assumptions that make the assessment of the validity of these assumptions more difficult. Agent-based models also require the finer level of model detail in which data at that level of detail may be difficult to obtain.

Duong (2010) also examines this issue and points at three reasons contributing to the difficulties in agent-based validation: the greater uncertainty in social sciences compare to others, the lack of consensus on how to represent social environment and the lack of experimental controls in data collection. There are some methodological problems in the empirical validation of agent-based models (Windrum, Fagiolo, and Moneta, 2007). The problems seem to have arisen due to, among other reasons, the lack of techniques to build and analyse these models and the lack of comparability between the ones that have already been developed. A number of validation techniques have been proposed for agent-based simulation modelling. For instance, Klügl (2008) defined a validation process for agent-based simulation models combining face validation and statistical methods. Another technique comes from Moss, Edmonds, and Wallis (1997) who used a declarative formalism to address the validation and verification of agent-based models with cognitive agents. Despite these works, there seems to be a general concern on the lack of validation framework or methodology in agent-based simulation.

A number of validation techniques have been proposed for agent-based simulation modelling. Klügl (2008) proposes a validation process for agent-based simulation models combining face validation and statistical methods. Arifin, Davis, and Zhou (2011) explain there are three ways to validate an agent-based simulation:

1. Through comparison of simulation output with real phenomena. Although its simplicity, this method has disadvantages when real data is not complete.

2. Constructing mathematical models of the system under study and compare these models with the simulation results. However, is not always possible to build a mathematical model of reality, particularly when formulating complex systems.

3. The third technique is docking (also known as alignment, replication, cross-model validation, or model-to-model comparison), a process of comparing two similar models which address the same question with the objective of not only finding their similarities and differences, but also to gain understanding of the phenomenon under study (Burton, 1998). Performing this comparison between independent simulation tools, docking might find differences of interpretation in the model specification and also in the implementation. Nevertheless, finding a similar behaviour among multiple simulations will increase the validation confidence. Some examples of docking can be found in the literature. For instance, the beer distribution game by North and Macal (2002), the simulation of organisations by Ashworth

and Louie (2002) and the work in collaboration networks by Xu, Gao, and Madey (2003).

Another technique is the validation at two levels, called cross-validation (Moss and Edmonds, 2003). It consists of two steps. First, we do a qualitative validation at the agents' level, checking that the behaviour of the computational agents is similar to the target agents. Second, a statistical validation of patterns of behaviour of the overall system is performed. The concept of cross-validation comes from Physics, specially on looking at systems with high volatility, and the social theories of social embeddedness.

6.2 Yades Validation

Simulation community has recognized a number of validation techniques. Sargent (2005) provides a summary of the techniques, such as animation, comparison to other models, degenerative tests, event validity, extreme condition tests, face validity, historical data validation, historical methods, internal validity, multistage validation, operational graphics, parameter variability or sensitivity analysis, predictive validation, traces and Turing tests. Robinson (1997) identifies four different forms of validation in simulation modelling: conceptual model validation, data validation, white-box validation and black-box validation. Conceptual model validation deals with issues such as the level of detail of the model and determines if it is enough for the purpose it was developed. Data validation is needed to determine whether the data used in the simulation study is sufficiently accurate. The black-box validation concerns with the relationship between inputs to the model and its outputs, ignoring the elements inside a model. The objective is to determine if the output of the model reflects the real world observation for the same set of inputs. Finally, white-box validation tries to answer the question does each element of the model and the structure of the model elements represent the real world with sufficient accuracy?

In demographic modelling, common used techniques appear to be the variants of white-box validation model. The reason is validation is exacerbated by the inaccuracy of some demographic data and the changes in the definition of some data categories. For example, the definition of stillbirth now could be different from 20 years ago. For example, Bohk, Ewald, and Uhrmarcher (2009) and Zinn et al. (2009) use plots of descriptive statistics to show the output results. As a consequence, we used the white-box method to test whether our tool has been correctly implemented. The method used in this project is described in (Pidd, 2004). The demographic simulation model creates an artificial society based on the characteristics of the UK population using Yades. We divide the model into smaller components and test the correctness of each component. It is suggested that we start with the simplest possible behaviour so that the simulation output can easily be understood. Hence, errors can be easily spotted. This is the approach that we have used. We have tested the implementation of a simple model for each demographic component in our model.

In the following subsections, we present the result of our evaluation on each model component using the white-box method. In the evaluation, we use a population of 110,000 family units. In each test, we run the simulation five times and report the average results.

Mortality

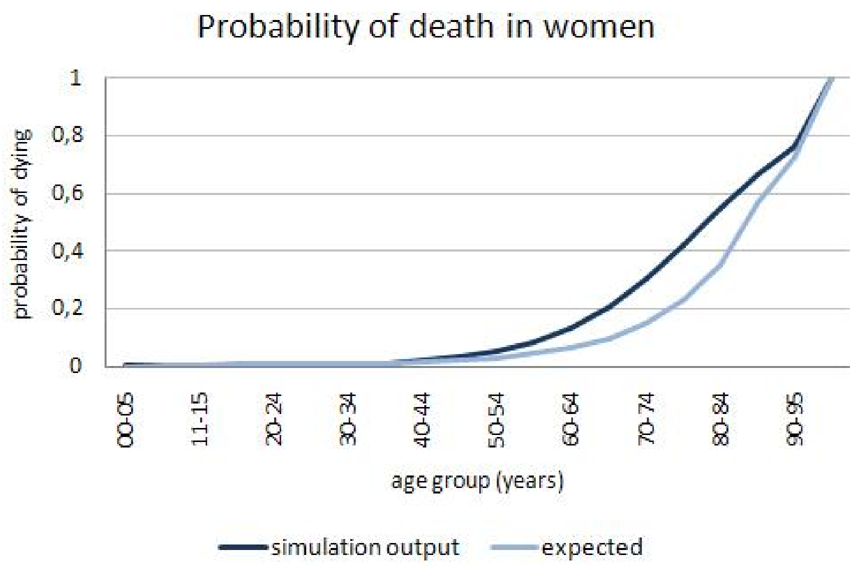
Yades allows modellers to sample the lifetime of individuals using two commonly used methods: life table and survival function. In order to evaluate this component, we disable all other demographic components. This action helps us to detect any error and to isolate the cause of the error easily. We vary the life tables. One of the results is shown in Figure 6.2. In both cases, the Pearson product moment correlation coefficient of the original distribution and the outputs is very high, 0.9870 for women and 0.9783 for men. The same evaluation is repeated using various life tables. They also produce high correlation values. This exercise has increased our confidence that the mortality component can produce the intended behaviour.

Fertility

Modellers can specify a number of fertility models in Yades using age-specific fertility model, parity-specific fertility model, birth spacing model and their combinations. To test the fertility model component, the rest of demographic components are disabled in order to isolate fertility results. To simplify the model, birth function is set to follow a Poisson distribution with parameter $\lambda = 2$ in women from 16 to 49 years old (assumed to be the reproductive age). Birth function is calculated for every woman regardless their marital status, and birth spacing is uniformly distributed. After running the simulation, the accumulated number of births by age group is obtained. In Figure 6.3, the percentages of births are represented according to the number of children's group. As we can see, the simulator is producing the expected number of births in the fertility interval. We repeat this experiment with different parameters, and all of them produced the expected results.

Marital status

Yades recognizes the following marital statuses: single, married, cohabitation, separated, divorced and widowed. Individuals will move from one marital status to another during their lifetime. The transitions from one status to another can be specified based on a simple probability function, a regression function or a set of logical rules. Likewise, time spent in one status can be sampled using a distribution function, a regression function or a set of logical rules. In the formation of a family



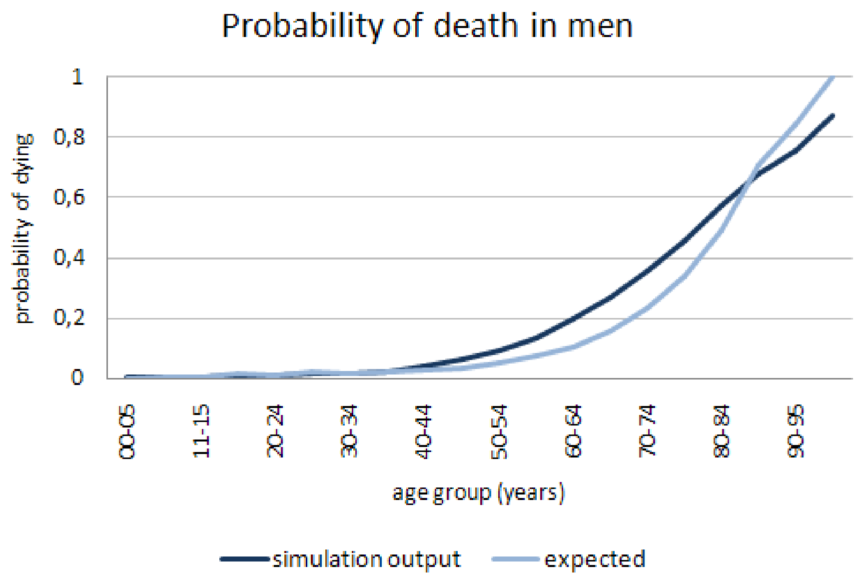


Figure 6.2: Comparison between simulation output and expected output for mortality

unit (e.g., marriage and cohabitations), we need to specify a function that matches a pair of individuals.

In the following evaluation, we use a probability function for the state transition and apply a simple matching criteria where we choose the first person that

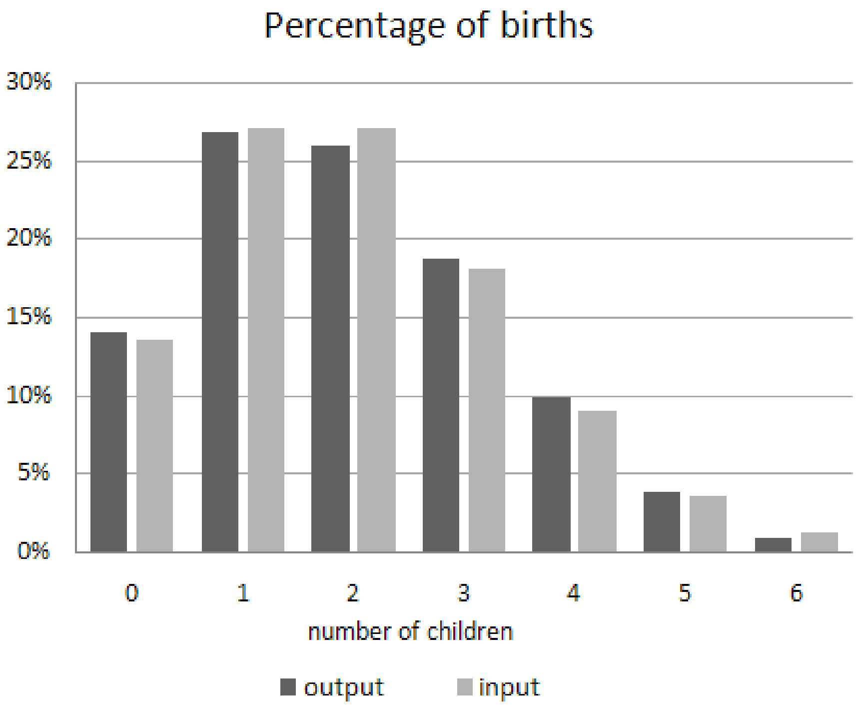


Figure 6.3: Comparison between simulation output and expected output for fertility

we find in the list regardless of his/her characteristics. Transitions are uniformly distributed between 1 and 10 years. First, we want to test the correctness of family unit formations. Hence, we disable all other transitions. We also disable all other life events (as fertility and mortality). The top chart in Figure 6.4 shows the result. As expected, the number of marriages and cohabitations increases with time while the number of singles decreases progressively. The second test has the same settings as before, but in this case we enable the mortality. The result can be seen from the bottom chart in Figure 6.4. The figure shows that the number of widowed increases with time. These results are the behaviour that we expected.

Economic status

As in the marital status, an individual may move from one economic status to another during his/her lifetime. Yades recognizes the following economic statuses: dependent, in employment, unemployed, in full-time higher education, pension and economically inactive. Modellers will need to model the transitions from one status to another and the time spent in any of the status. In the evaluations, the transitions are scheduled using an uniform distribution between 1 and 5 years. The top chart in Figure 6.5 shows the changes in economic status. As expected, without mortality, the number of pensioners increases steadily. At the same time, the number of dependent individuals, individuals in higher education, working individuals and unemployed individuals decreases over time.

In the second test, we use exactly the same setting but in this case we enable the mortality. The scenario should produce a similar behaviour as before, but the proportion of retired individuals will be less because some of them will die. The result can be seen from the bottom chart in Figure 6.5.

Migrations

Yades provides a facility for a modeller to define a model that determines whether a family unit is going to migrate. There are two types of migration: domestic migration and international migration (emigration and immigration). These models can be specified using a constant probability, regression or a set of logical rules. The destination region is determined using a probability matrix where each row represents the originating region, and each column represents the destination. To validate this component we tried two different scenarios using four regions. In the first scenario, we set the probability to migrate to be the same across all regions. As expected, the number of population in each region is relatively the same. In the second scenario we set one of the regions (i.e., region 4) to be the most attractive, such that once people have moved to that region they will never leave the region. In this example, we expect an increase in the population of region four while the rest of the regions experience a decrease in their population. The result is shown in Figure 6.6.

The validation of a complex agent-based model is challenging. The reason is partly due to the quality of the available data that are needed to calibrate and to validate the model. A large number of model parameters makes it even more challenging. In this project, we have conducted the verification and validation of an agent-based demographic simulation model implemented using Yades using white-box method. This method allows us to assess the correctness of the model components and their interactions. The results show that the five components of the simulator are behaving correctly in terms of what the modellers should expect from them for the given scenarios. To increase our confident in the model, we need to conduct more testing using different validation methods.

The graphical user interface explained in Chapter 3 can be useful to help users specify the model more easily without having to write the codes. This interface would help potential users who do not have any programming experience to test their models and provide feedback on the tool.

Simulation output

The chart on the left in Figure 6.7 is the population pyramid of the initial population which is based on the UK Family Resource Survey data in 2008/09. A simple combination of age-specific fertility, parity-specific fertility and birth spacing are used for the fertility model (representing a population with an average fertility rate of 1.3). The UK life table based on the 2005-2008 data is used for the mortality model. A simple random sampling method is used for domestic migrations and changes in marital status and economic status. Finally, there are no international migrations (representing a closed population). The chart on the right in Figure 6.7 shows the population pyramid ten years later. This setting provides an example of how a low birth rate affects the demography of a closed population in 10 years.

6.3 Conclusions

The validation of a complex agent-based model is challenging. This is partly due to the quality of the available data that are needed to calibrate and to validate the model. A large number of model parameters makes it even more challenging. In this paper, we have presented the verification and validation of an agent-based demographic simulation model implemented using Yades using white-box method. This method allows us to assess the correctness of the model components and their interactions. The results obtained in this paper show that the five components of the simulator are behaving correctly in terms of what the modellers should expect from them for the given scenarios. To increase our confident in the model, we need to conduct more testing using different validation methods. At the moment, we are implementing a graphical user interface to help users specify the model more easily without having to write the codes. This would help potential users who do not have any programming experience to test their models and provide feedback on the tool.

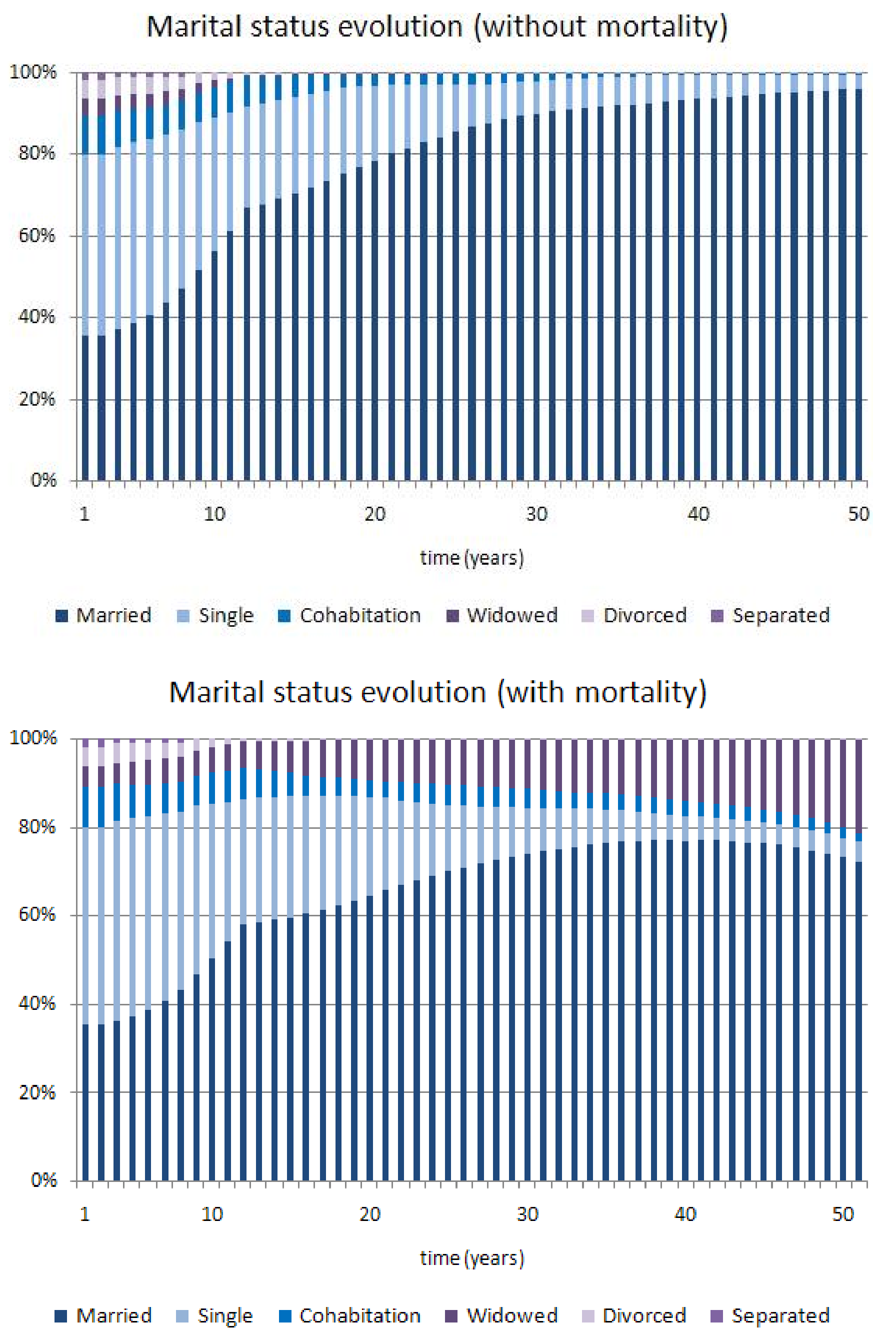


Figure 6.4: Reasonableness check for marital status

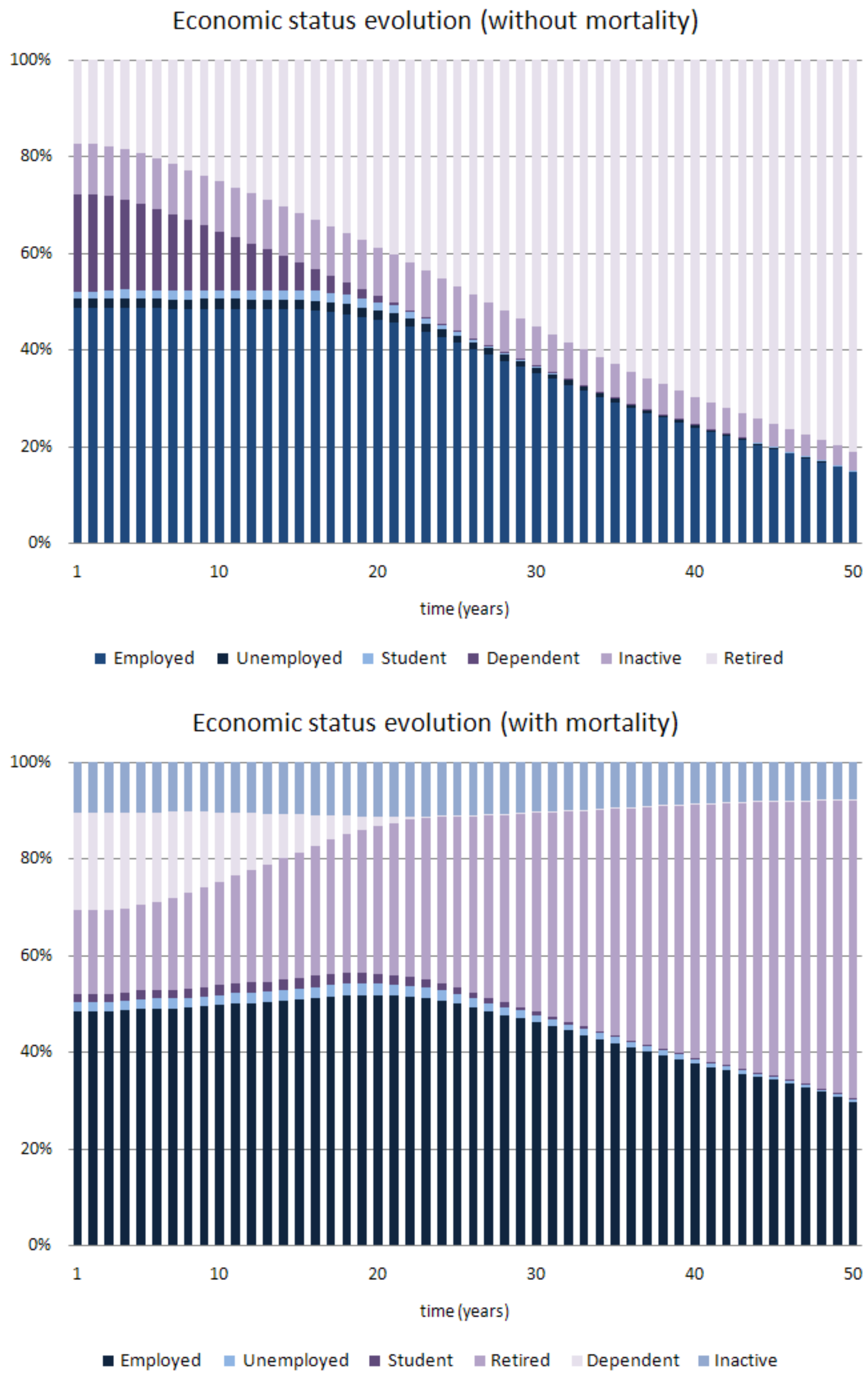


Figure 6.5: Reasonableness check for economic status

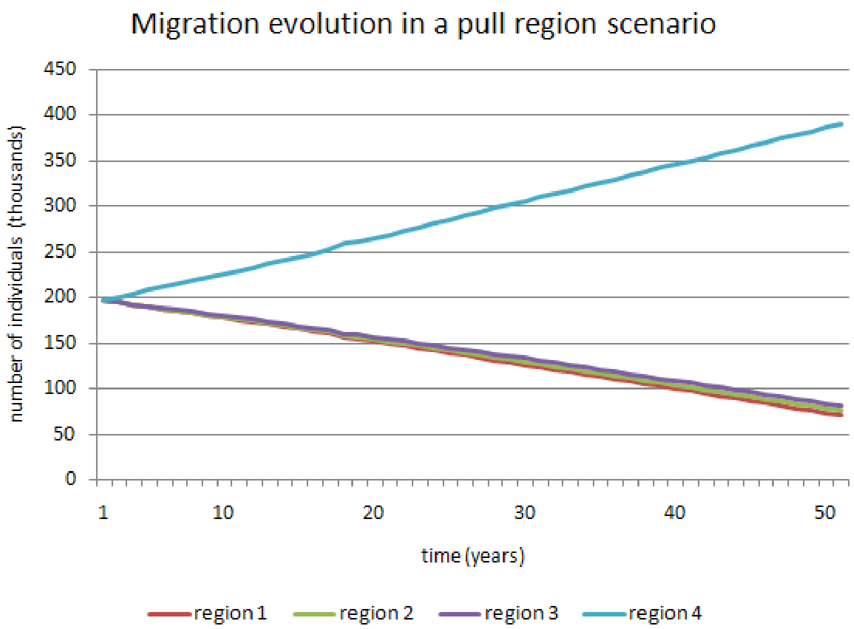


Figure 6.6: Reasonableness check for migrations

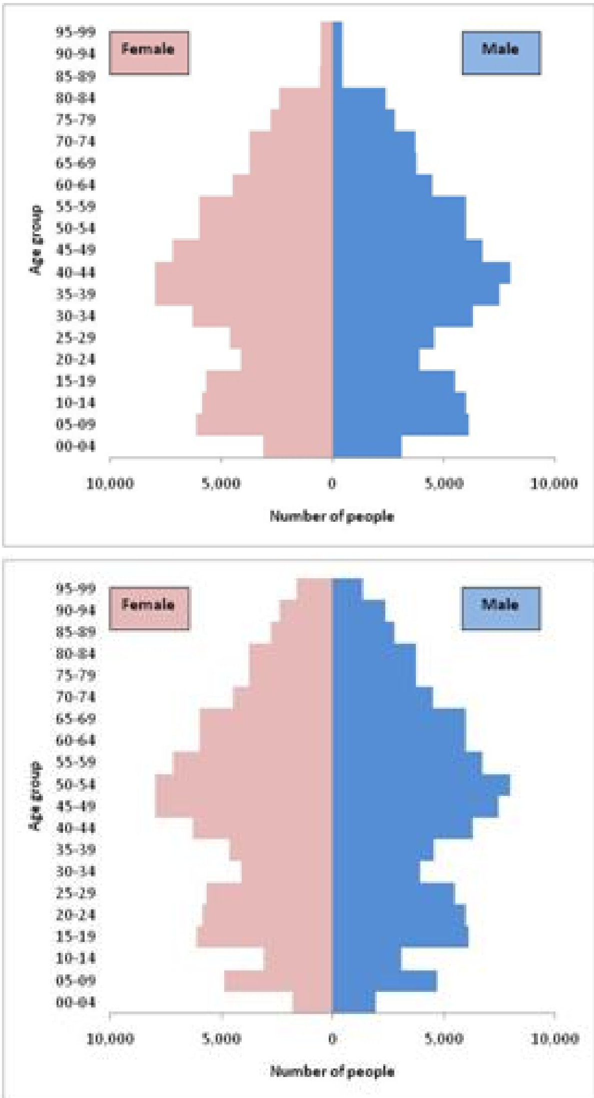


Figure 6.7: Population pyramids (left: initial population, right: 50 years later)

Chapter 7

Two case studies

This chapter explores two case studies showing the use of Yades for a real application in demographic analysis. In the first one, we take advantage of our agent-based simulation approach to project the population of Gambian migrants in Spain during 10 years (Section 7.1). In the second one, we explore how the demographic evolution of South Korean population affects its ageing in 100 years (Section 7.2).

Our approach not only enables to simulate the life course of individuals, but also allow us to go further into the the movements, interactions, and behaviours of the target population. Our methodology is able to capture individual characteristics and to overcome some data-related limitations with assumptions on behavioural rules. The purpose of these case studies is not to provide a calibrated model but to show Yades' potential with the study of two real case scenarios. The results of the first study have been presented in Montañola Sales, Casanovas-Garcia, Onggo, and Li (2014c).

7.1 Gambian immigrants in Spain

The Gambia is a small country in the Sub-Saharan Africa, bordering the North Atlantic Ocean and Senegal. It is almost an enclave of Senegal; smallest country on the African continent with 1,88 million people (2003 national census source). It is an ex-colony of the United Kingdom who got its independence in 1965. The country is situated on each side of the Gambia River, which flows through the country's centre and empties into the Atlantic Ocean.

According to the 2003 national census, the population distributes among different ethnics groups: 42% Mandinka, 18% Fula, 16% Wolof/Serer, 10% Jola,

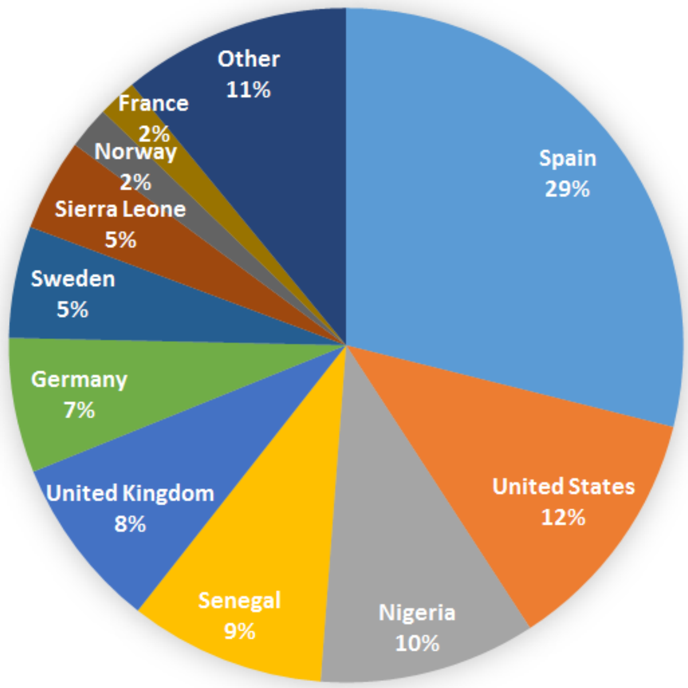


Figure 7.1: Distribution of Gambian Emigrants by Destination Country, 2010. Source: The World Bank. Migration and Remittances Factbook 2011 [http://www.worldbank.org/migration]

9% Serahuli, 4% other Africans, and 1% non-African. Due to the fertile land of the country, the economy is based on farming, fishing, and tourism. About 75% of the population depends on crops and livestock for its livelihood. Small-scale manufacturing activity features the processing of peanuts, fish, and hides. Thus, most of its population has a rural background. Literacy is quite high, 40.1% of total population in 2003. Around a third of the population lives bellow the international poverty line of US$1.25 a day (according to the Human Development Report 2009, United Nations).

Despite its small size, Gambian emigration rate has been quite high last 30 years. In 2010, Spain received the 29% of emigrants from the Gambia, being the leading destination country (see Figure 7.1). In the region of Gambia, movements outside the country are perceived as a familiar strategy and follow two differentiate patterns: surviving and mobility (Kaplan Marcusán, 2005). Mobility movements generally answer to a familiar project, where the household does a significant financial investment. To maximize the success of the project, a selection process in the origin characterizes migrations.

The case study presented in this chapter carries out an analysis of the evolution of the population of The Gambia living in Spain between 2001 and 2011. The

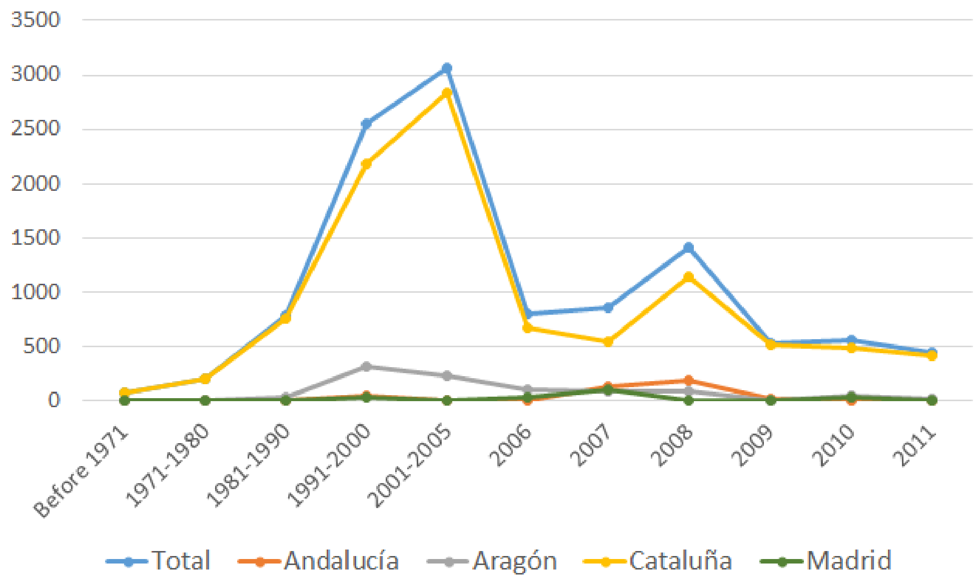


Figure 7.2: Arrivals of Gambian immigrants by destination region in Spain, from before 1971 to 2011. Source: INE. Censo de Población y Viviendas 2011

period is interesting for research because migration flows from Sub-Saharan Africa (including The Gambia) to Spain increased significantly. Spanish economy prospered in that decade probably increasing the expectancies of Gambian migrants and increasing migration movements. Figure 7.2 shows the arrival of Gambians to Spain in different Autonomous Communities. As we can see, Catalonia had the highest concentration of the incoming flows.

7.1.1 Context

The Gambian is the four largest Sub-Saharan African group of emigrants in Spain, with a total of 20,639 people according to the 2011 census of National Institute of Statistics of Spain (INE). It is the group of emigrants with the highest fertility, the higher proportion of children under 16 years old and the lowest proportion of people older than 65 in Spain (Bledsoe, Houle, and Sow, 2007). Thus, we are facing a young population, with a very high fertility rate (Total Fertility Rate in 2013 was 5.43 children per woman according to the National Institute of Statistics of Spain - INE). Moreover, Gambians show an unequal sex and age distribution as we can see in Figure 7.3. The population pyramid illustrates the age and sex structure of a country's population. The population is distributed along the horizontal axis, with males shown on the left and females on the right. The male and female populations are broken down into 5-year age groups represented as horizontal bars along the vertical axis, with the youngest age groups at the bottom

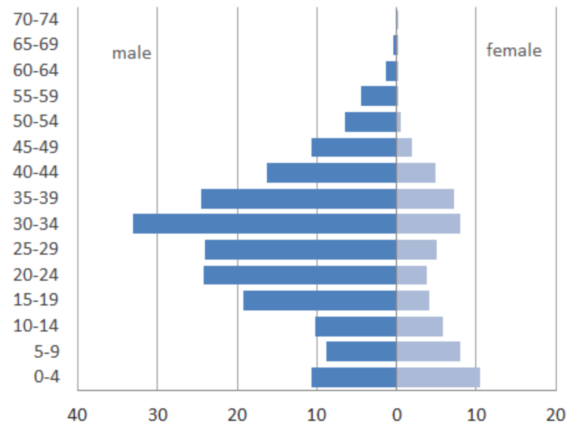


Figure 7.3: 2013 population pyramid of Gambians in Spain showing a high masculin component. Source: INE 2013 continuos register

and the oldest at the top.

The gender disequilibrium shows two different migratory strategies, which is common among Sub-Saharan immigrants in Spain (Sepa Bonaba, 1993; Prieto Sarro and López Trigal, 1993; Domingo, López-Falcón, and del Rey, 2008, 2010). Figure 7.4 shows the magnitude of this phenomena. From one hand, the high masculine proportion, especially when compared to other migrants coming from the rest of Africa, might respond to economic reasons in the origin. On the other hand, low women proportion might be a consequence of family reunification processes.

The particularities of Gambian population make an impact on local population structures in Spain. Gambian immigrants exhibit a distinctive pattern from the rest of migrants, i.e. they tend to reside only in certain areas, dispersed across the various regions, with the largest concentrations in Catalonia from the beginning (76.16% in 2011). Moreover, due to their rural origin in most cases, they tend to work in agricultural sectors, domestic services, construction, or the service industry.

7.1.2 Gambian model with Yades

In this section we will introduce how we built a model for Gambian immigrants in Spain using Yades. In this model, we are including the two types of agents provided by our framework: families and regions. We made the assumption to just model Gambian immigrant demographics in Spain and not include the rest of Spanish population to focus on the target population dynamics.

In Gambia, polygenic marriage arrangements to compose extended families (up to four women per man) are very common (Kaplan Marcusán, 2005; Bledsoe

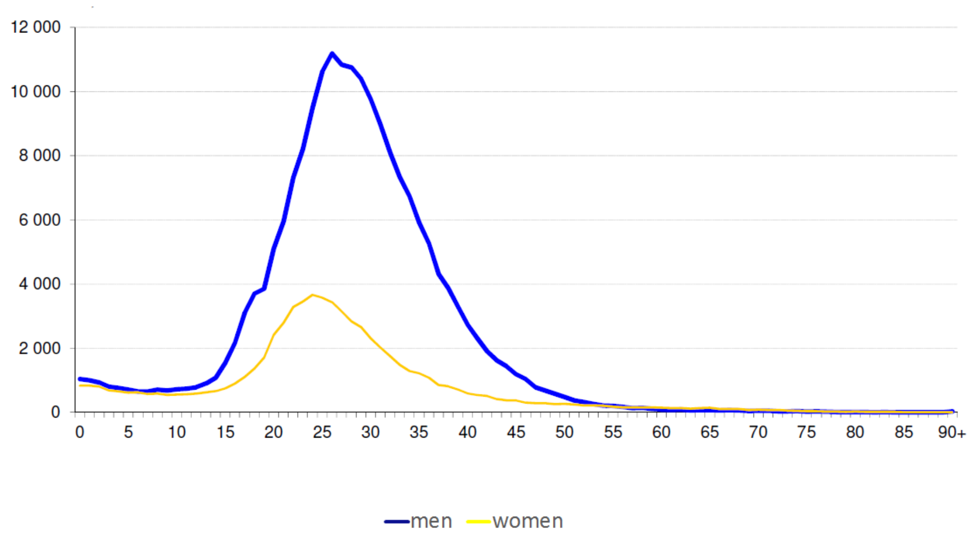


Figure 7.4: Sub-Saharan population flows by sex and age 1988-2009. Source: Domingo i Valls and Vono de Vilhena (2012)

et al., 2007). Kaplan Marcusán (1998); Bledsoe et al. (2007) argue polygamy is practiced in Spain although illegal. In the study made by Kaplan Marcusán (1998) of 121 Senegambian immigrants in Girona, 27% of the married men had two wives. They follow the strategy where the first wife returns to Gambia with her children and then the second wife comes to Spain. Despite the evidences, it is difficult to quantify the number of cases of this kind. Therefore in our model we make the assumption to consider a family unit as one or two adults with (some) children.

In the model we included eight regions covering 97% of Gambians in Spain: Almería, Huesca, Zaragoza, Madrid, Barcelona, Girona, Lleida, and Tarragona. Since most of Gambians live in Catalonia, our results focus on the four provinces of Girona, Lleida, Tarragona and Barcelona. We start the simulation with 7843 families, 15402 individuals.

Table 7.1 summarizes the data, either quantitative or quantitative, that was used in each demographic component of our simulation model and the different sources that were used. Next subsections describe the model in terms of how a population is initialised, how it is updated over time in each demographic component of the model, and which data related issues should be taken into account with current available data.

Initialisation

An initial set-up population is created by randomly generating families with individuals characterized by their age, sex and the family to which they currently

Model component	Source of data and type
Mortality	Life Table for each region (source: INE)
Work/ Education Status	Four states: Employed, Unemployed, Studying and Retired Transitions for the state diagram (source: OPI) Duration ~ Uniform
Marital Status	Five states: Single, Married, Separated, Divorced and Widowed Transitions and durations for the state diagram (source: INE) Closed Model for Mate-Matching (Billari et al., 2007)
Fertility	Assume a maximum fertile age (Kaplan Marcusán, 2005) Age-specific fertility rate for Gambians (Domingo et al., 2010) High fertility pattern (Bledsoe et al., 2007; Kaplan Marcusán, 2007)
Domestic Migrations	Probability based (source: INE) At any work/education status The whole family unit migrates
Immigrations	Estimated arrivals to each region (source: INE) Young component (Kaplan Marcusán, 2005)
Time	Transitions are set at the initial simulation time and do not change with time

Table 7.1: Summary of data sources and types for the Gambian demographic model in Yades

belong. A population consists of the set of current families. The individuals are assigned to the families according to a specified family size distribution. Families can be formed by one or two adults and a number of children. The ages of the initial population are drawn from a specific age distribution. The structure of this initial population will diverge in several ways from a real population since constraints on birth interval and inter-generational age difference are not respected. More complex methods could be used to gain greater realism (Gargiulo, Ternes, Huet, and Deffuant, 2010; Ajelli and Merler, 2009).

Mortality

In demography, age and sex mortality rates are used to determine an individual's probability of death. The most common method for mortality is the life table approach. The Spanish National Statistics Institute (INE) provides life tables by province, so we used those for Almería, Huesca, Zaragoza, Madrid, Barcelona, Girona, Lleida, and Tarragona. Life tables for Gambian population in each province were not available. Therefore, we assumed Gambian population would have a similar mortality than the general population in the region. For example, Figure 7.5 shows the life table for Barcelona. We contemplate the possibility of using as well data from the Gambia. However, in Gambia mortality data is not available either in the 2003 or 2013 Population and Housing Census.

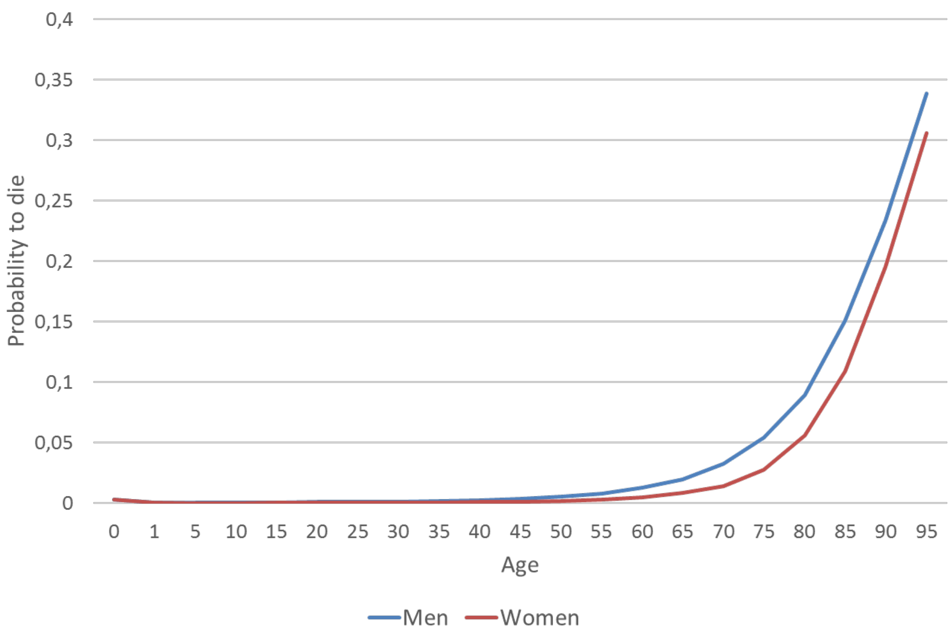


Figure 7.5: Life table for the population in Barcelona in 2001. Source: INE 2001 census

In our model, if a death results in a family containing only children then the family is dissolved. Any adult children (i.e., aged 18 or over) leave home and create new single-person family, while any children under 18 is randomly allocated to other families containing at least one child.

Fertility

In Gambia, the main function of marriage is to have children (Bledsoe, 2002). According to Kaplan Marcusán (2005), motherhood is seen as the *"primary life project"* of women. Therefore, when reunited wives arrive to Spain, they *"start or continue building their family"*. That is one of the reasons why Gambian fertility is quite high compare to other immigration groups in Spain. Although Bledsoe et al. (2007) attribute part of the high fertility rates among Gambian women to the circulation of co-wives and their children between Spain and Gambia to evade Spanish prohibition on polygyny. They argue that the residency permit is exchanged among the co-wives. As a consequence, second and third wives become "invisible" in Spain.

Despite this evidence, in our simulation model we do not consider the circulation of co-wives due to the lack of data available. Therefore we assume a maximum fertility age from 15 to 50 (Kaplan Marcusán, 2005; Domingo et al., 2010) and a fertility component based on the age specific fertility ratio plotted in Figure 7.6.

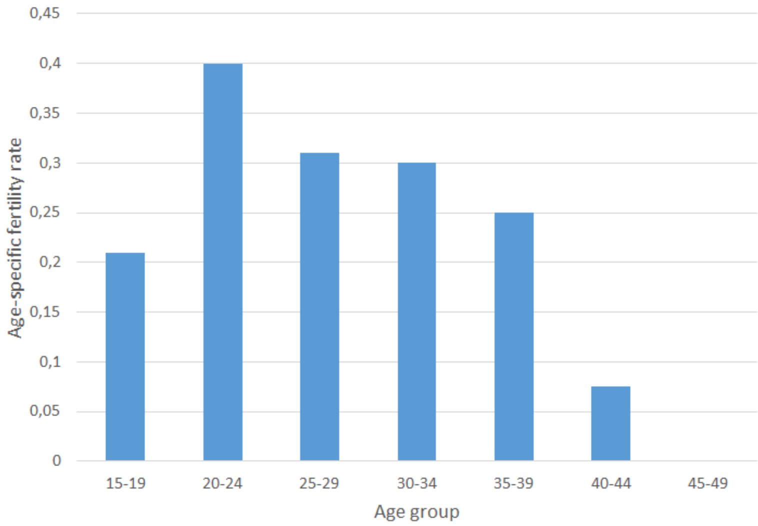


Figure 7.6: Age specific fertility rate for Gambian emigrants in Spain in 2001. Source: INE 2001 census

In our model, upon giving birth a mother is excluded from being a candidate for future births for a random number of months, with a minimum duration of 12 (9 months of pregnancy plus 4 months of maternity leave).

Marital status

We assume five states for the marital status model: single, married, separated, divorced and widowed (see the diagram in Figure 7.7). Transitions and durations among states were based on age, gender and current marital status of a person. For example, an individual within a given range who is currently single has a fixed probability per time unit of forming a new couple. Transitions and durations were quantified with data from demographic indicators from the National Statistics Institute of Spain (INE).

Demographers and sociologists often relied in mathematical and statistical macro-level methods to model partnership formation through age-at-marriage curves. With those methods, they have seen the age pattern of marriage is characterized by constant features over time and space (Coale and McNeil, 1972). However, psychologists and economists have traditionally focussed on modelling the process of partner search. Agent-based models allows to combine both approaches. Thus, we are proposing to add this feature to our Gambian model.

We assume a closed marriage model where we match a person with another from a list of prospective partners. For that, we are using the "wedding-ring" model proposed by Billari et al. (2007). This model decides whether a person will find a partner based on its social interactions. Billari et al. state the chances

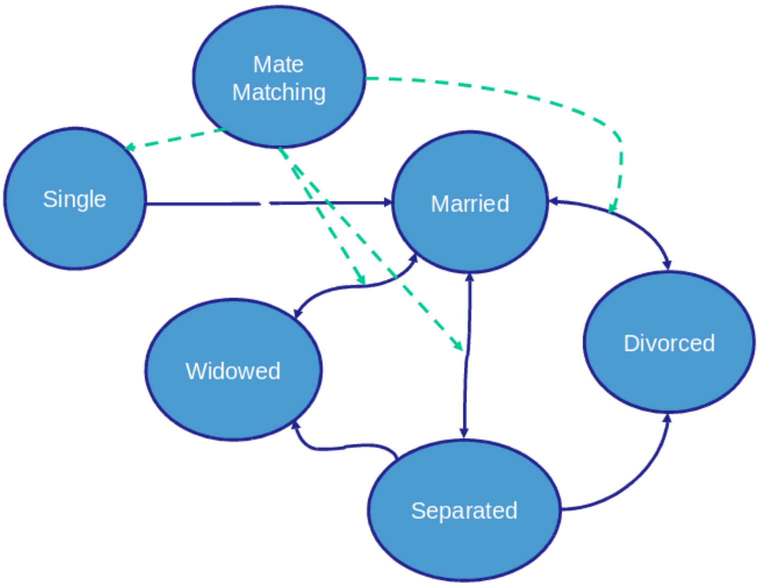


Figure 7.7: Marital status diagram for Gambians immigrants in Spain. The mate matching function is calculated from transitions from single, widowed, separated or divorced to married.

of marrying depend on the availability of partners, and the willingness to marry which is influenced by people already married in the same social network ("social pressure"). The social network is linked to a neighbourhood, and it is located in a theoretical space of two dimensions (age and location). The social pressure is based on Hernes (1972) model which showed already married couple would affect the not yet married population, assuming that members of the same cohort constitute the influential peer group of everyday life. Moreover, Billari et al. define a functional form of age influence since the literature shows the size of the social network varies with age.

In Billari et al. model, it is assumed that a person lives in a theoretical space ϕ. Every year, the group of the individual's relative others is determined by selecting a random set of people in ϕ. This group of relevant others includes individuals of both gender, but only people from opposite gender may marry. The proportion of married people (pom) in the relative others defines the social pressure (sp) for marriage according to the following formula:

$$sp = \frac{exp(\beta(pom - \alpha))}{1 + exp(\beta(pom - \alpha))} \tag{7.1}$$

where α and β are some parameters that determine the inflection point and slope of a social pressure function.

In our model, we make a simplification of Billari et al. and assume that if two unmarried people with opposite sex exist in each one's window for potential partners respectively, then they get married. Our window is defined by the local region where the agents live. A difference with Billari et al. model is that we calculate this function not only for single, but also for divorced, separated, and widowed (see Figure 7.7). Thus, extending the use of the match making model for all kind of partnership formation. When a partner is chosen from the pool of individuals who are currently single and of the opposite gender, the new couple moves into a new family, together with any dependents (e.g., children from previous couples). In the same way, when a couple dissolution takes place one member of the couple moves into a new single person family, while the other remains in the family. Qhen this dissolution happens, we assume children are split evenly between the new two families.

Economic status

For the work status component, we assume four states: employed, unemployed, studying and retired. Transitions among states were quantified with data from the Permanent Observatory of Immigration of Spain (OPI). The lack of data to determine durations among states made as to assume duration follow an uniform distribution $\sim U(1,5)$ in years. The economic component also calculates when an individual above a specified age leaves the current family and forms a new single person family.

Domestic migrations

In our model, we included domestic migrations across regions. Since we are modelling eight provinces of Spain, we can have 56 different movements. For instance, from Barcelona individuals could migrate to Girona, Lleida, Tarragona, Huesca, Zaragoza, Madrid or Almeria. We made the simplification of isolating Gambians from the rest of Spanish citizens. We also assumed a common trend of domestic migrations based on a global probability of 30%, which was extracted from general trend of domestic migrations in 2001 according to the National Statistics Institute of Spain (INE). Since we did not have any data on the destination of those migrations, we assume a random destination. Our model allows to perform a migration at any economic status. The change of region is produced at the level of the family, with all its members migrating.

Immigrations

The model includes new arrivals from Gambia. Based on the data provided by the Spanish National Stastics Institute (INE), we adjusted immigration arrivals for the period 2002-2011 to simulate 89.7% of the total number of arrivals (see Figure 7.8). As a simplification, we do not differentiate between incoming migrations and family reunification processes.

Figure 7.8: Real vs simulated Gambian migrations in Spain, 2002-2011. Source: INE census, Estadistica de Variaciones Residenciales – Residential Variations Statistics

Data considerations

One of the difficulties we confronted on building the model was the lack of data. For instance, in The Gambia we have very few data coming from the census (even 2013 Population and Housing Census has not been fully published yet). Some data coming from Unicef, United Nations and the World Bank is based on estimations, an issue that should be considered. For instance, the World Factbook of the Central Intelligence Agency (CIA) estimates a population pyramid for the Gambia (see Figure 7.9) but this data is not coming from the Gambia Bureau of Statistics who elaborates the census.

Moreover, managing demographic data in the European context requires a careful study. For example, when analysing migration and its consequences, who counts as a migrant is of crucial importance. Yet there is no consensus on a single definition of a 'migrant' across Europe. In Spain, children born in the country are counted as migrants if their parents have a foreign nationality. However, this is not the case in France.

Another issue that should be taken into account on the management of Spanish demographic data is the different sources of available data from the National

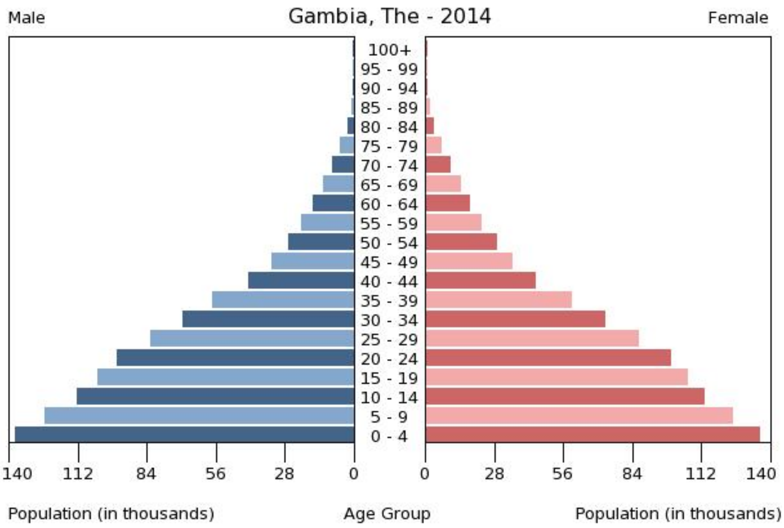
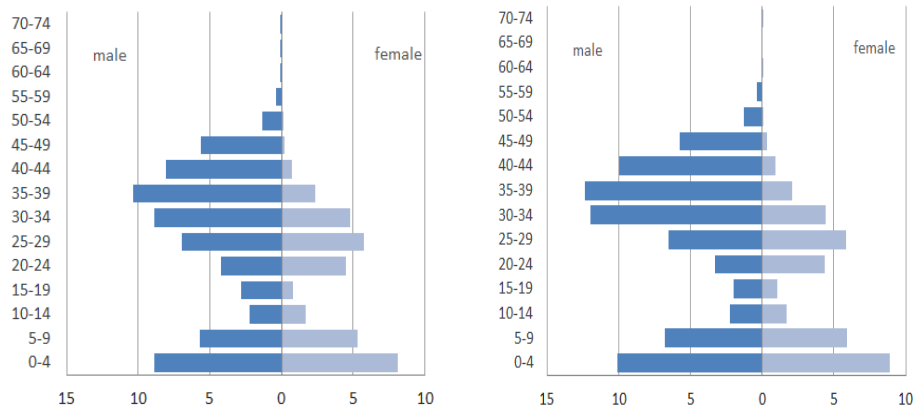


Figure 7.9: Estimated population pyramid of The Gambia in 2014. Source: The World Factbook. Central Intelligence Agency (CIA). Retrieved 17 November 2014.

Statistics Institute of Spain (INE). For instance, Figure 7.10 compares a population pyramid built from the 2001 INE census with another built from the 2001 INE continuous register. As we can see, there are slight differences in data: the continuous register estimates the number of children up to 4, and men between 30 and 44 to be higher than the census. These differences can be explained because of the overestimation of the continuous register, since in Spain it is not mandatory to update one's state when moving abroad. This overestimation is the reason why we did not rely in the INE continuous register for our model. On the downside, the official nature fo the microcensus also limits the validity of the data source since the sample only contains officially registered foreign citizens. Besides this sampling error, the available microcensus is a sample of just 10% of the population which in turn can add some bias to the immigrant population. Issues like that should be considered in order to analyse simulation results properly.

7.1.3 Simulation results

Starting from the 2001 structure by sex, age and marital status, we have propagated the demographic dynamics forward, obtaining the expected picture of an ageing population. Results presented in this section have been obtained from an average of 100 runs of the simulation, in other to smooth the randomness in the underlying patterns. Since the number of outcomes obtained from simulations is very high, we show a selection of what is obtained. In this section, we present some of the results for the simulation year 2011, which we are able to compare



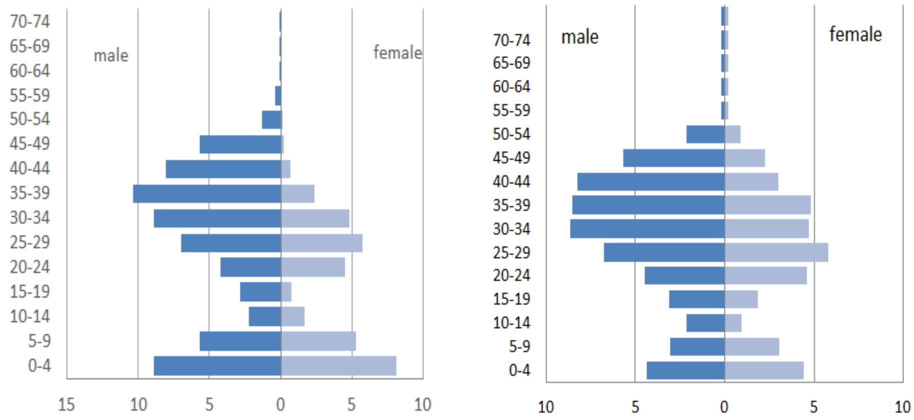
(a) census **(b)** continuos register

Figure 7.10: Comparison of INE sources for the Gambian population in Catalonia in 2001: census vs. continuous register

with the census data provided by the National Statistics Institute of Spain (INE).

First, as described in previous section, we calculate the initial population for the simulation. Figure 7.11 shows the differences between the built set-up with the initial population from the INE 2001 census. As we can see, there is slight overestimation on the size of the initial population (7.15% higher than real data), particularly in women. However, the gender unbalance is respected. Children bellow 5 are under-represented. However, in general terms we can conclude the initial population set-up responds to the needs of the simulation and reflects the current distribution of the target population.

During 10 years, Gambians migrated across regions in Spain. Figure 7.12 presents the intensity of those flows across the territory. Since we assumed a global migration probability with random destiny, regions registered an emigration rate around 30% of their population. By the end of the simulation, 2444 families (5000 individuals, 30,31% of the total population) have change their residence in average. Figure 7.13 compares the total number of individuals obtained by the simulation (in blue) with the 2011 INE census data (in red). It shows the variance of results with information on quartiles 1 and 3 (Q1 and Q3 respectively). As we can see, Barcelona and Girona decrease their size probably because we used a general 30% migration rate when this might not correspond to reality for those specific regions. As a result, Lleida, Tarragona and Madrid have a little more population than the census shows. Differences can also refer to migrations to outsider areas from the ones included in the model, which is a limitation of our simulation and can increment perturbations. Despite the differences, we can see that the maximum value for the variance includes the value observed in the 2011 census. Moreover, we observe a correspondence pattern between our results and



(a) real

(b) simulation set-up

Figure 7.11: Population pyramid for Gambian population in Catalonia in 2011 (source: INE 2001 census) vs. the set-up used as the initial population in the simulation

the 2011 census data.

By the end of the simulation we obtain the evolution of the immigrant population for 2011. Figure 7.14 illustrates the overall population structure in the simulation for Catalonia, compared to the INE 2011 census. As expected, there are some structural differences due to not including emigrations to the origin, especially for groups over 45. These differences are common in both gender, and particularly for women whose population decreases abruptly from Spain over 45 years. This data confirms anthropologist and demographic studies with respect to the particularities of Gambian familiar forming, reunification processes. We can also observe men over 60 are practically non-existent which could confirm the explanation that emigrations are playing a key role in this population. There also visible differences in male children from 0 to 14 years. This issue could be explained by a over estimation of fertility, but the number of births for females is lower than real records. As a consequence, we need to investigate further this issue in future experiments. For groups of men from 15 to 39 the simulation shows results significantly lower than reality. We believe this might be due to not covering the total number of immigrations that happened in this period. Nevertheless, the overall process of population dynamics matches the Gambian emigrant population data quite closely.

Regarding the family structure, it is important to notice that individual births are allocated to parents using age and parity-specific fertility rates, but not to families of a specific size or composition. Thus, family size distributions are not controlled in our approach, but rather result from the interaction between individuals of different families and dissolution arising from individual-level events.

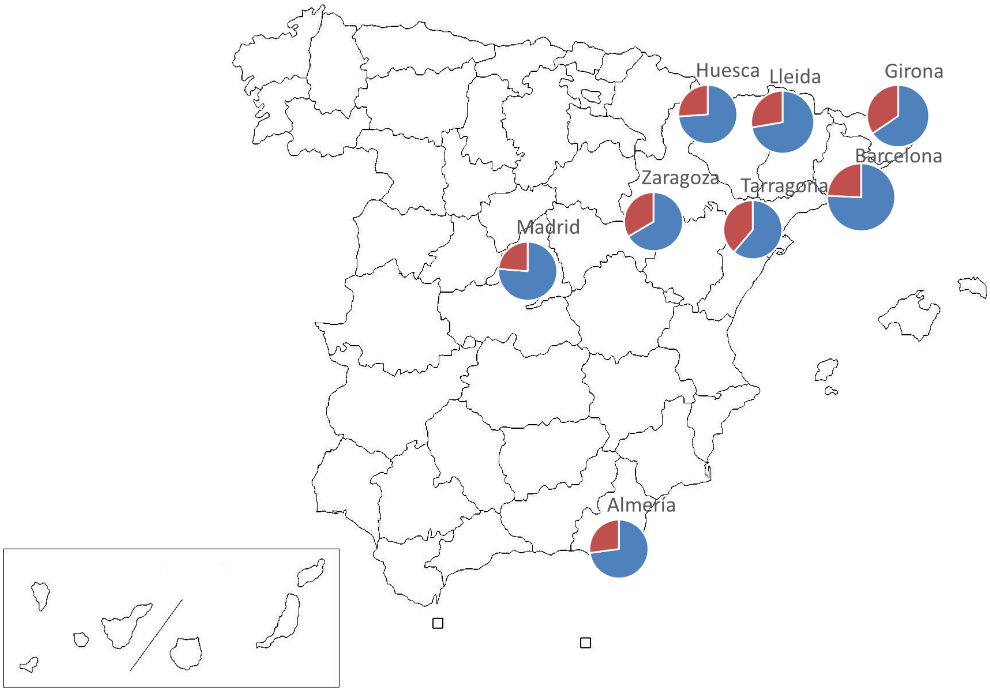


Figure 7.12: Density of migration incoming and outgoing domestic migration flows of Gambians across Spain provinces.

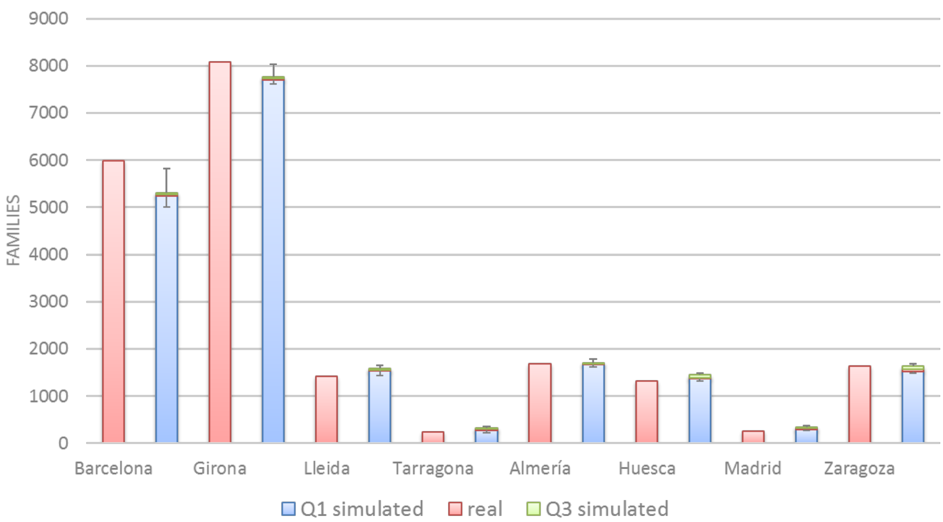
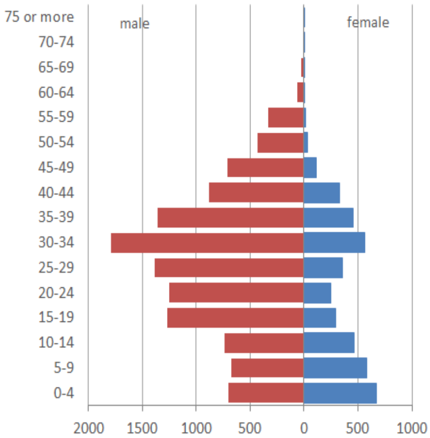
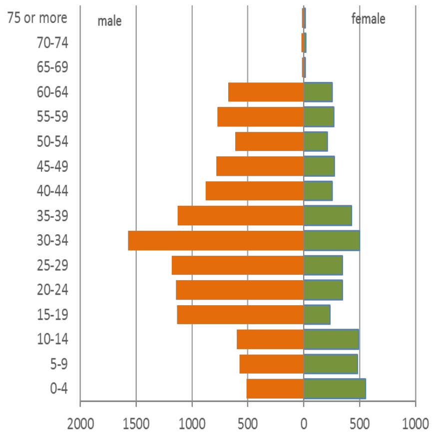


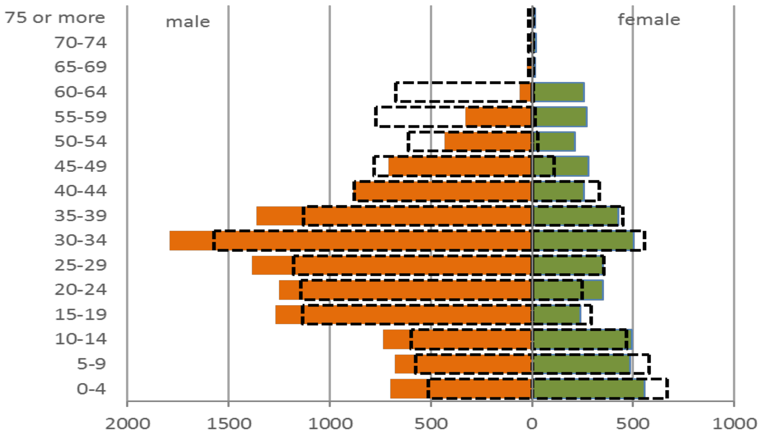
Figure 7.13: Comparison between population sizes per regions obtained by the simulation with the data of the 2011 National Statistics Institute of Spain, INE.



(a) real



(b) simulation results



(c) simulation results

Figure 7.14: Population pyramid for Gambian population in Catalonia in 2011 (a) (source: INE 2011 census) vs. the one at the end of the simulation (b). Figure (c) shows the differences observed when comparing (a) and (b). Black dot lines correspond to the census data.

Further calibration against family-level data would be possible, but it is not the purpose of this work to go beyond evaluating the effectiveness of our simple-based model.

In terms of computer resources, the average time of the simulations was 128.7 sec. (2.14 min), using 8 processors of one node in Marenostrum 3 supercomputer. Although in this case study the population is not very large neither the simulation very long, we increased its performance by 8.015 (in sequential the same simulation takes 1031.6 sec., 17,19 min in average). Therefore, the parallel computing approach is justified.

7.2 South Korean population dynamics

We are currently exploring socio-demographics of South Korea with Yades. The project is a collaboration with the Electronics and Telecommunications Research Institute (hereinafter ETRI) located in Daejeon, South Korea. Big Data Software Platform ETRI research department is currently studying new simulation concepts for developing an open full-scale individual based modelling simulation platform to analyse and predict socio-economic behaviours based on demographic changes.

The objective of this case study is to apply our methodology for population dynamics simulation to rate the feasibility in the South Korean context. The goal of the research is to analyse South Korean demographic patterns with our demographic parallel simulation framework and analyse the goodness of fit in forecasting individual demographic processes.

7.2.1 Korean context

South Korea is a country in East Asia of 50 million people whose economy has grown very quickly in the last 60 years. Today, its citizens enjoy a very high standard of living and an education system ranked among the five highest-performing countries in the world (OECD, 2014). It is also one of the world's most densely populated countries, with half of its population residing in the area around its capital, Seoul. In 2005, Korea encountered 9 percent of its population over 65, far beneath the average of 15 percent for developed countries. However, according to the latest government projections, this situation will dramatically change by 2050 where 38 percent of Korea's population will be elderly (see Figure 7.15).

Therefore, if the forecasts are correct, by 2050 South Korea may be the oldest country on earth. This situation could bring difficult challenges to face for Korean government and population. Despite the rapid economic growth of last decades, Korea is a traditional society. As a result, workers are expected to retire early from employment, women to quit their jobs when marrying and elders to be supported by their extended families when retired. These issues poses many interrogation for the Government in the long run, such as managing the rising expenditures on pensions and health care, facing the deficit of entry-level workers and young consumers or handling a new configuration of families. Moreover, Korea is one of the world's most ethnically homogeneous nations, being this homogeneity part of the national identity (Shin, 2006). Unless the country takes adequate measures to prepare for the demographic ageing trend, it is expected that Korea will face a slower economic growth and living standards stagnation (Howe et al., 2007).

Concerned by this situation, in 2006 the Government launched the program Vision 2020 Plan to raise fertility and prepare for a society with extreme ageing.

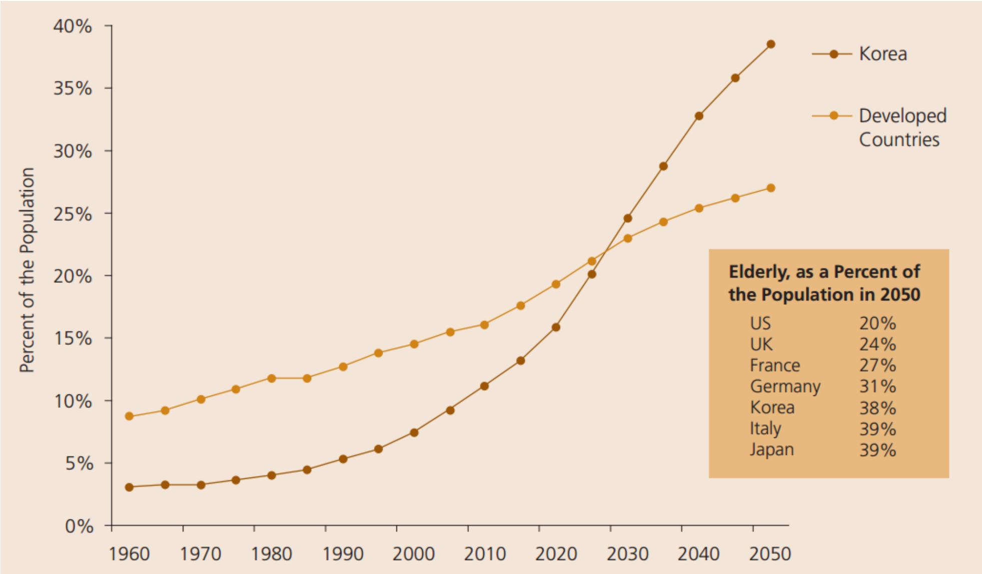


Figure 7.15: Elderly (aged 65 and over), as a Percent of the Population, Korea versus developed-country average, 1960-2050. Source: (Howe et al., 2007) from Korea National Statistical Office (2006) and United Nations (2005)

Age of women	1970	1980	1990	2000	2004
20 ∼ 24	460.3	405.6	426.3	363.6	259.0
25 ∼ 29	362.8	285.4	219.8	254.5	193.4
30 ∼ 34	217.5	120.9	54.78	96.8	99.5

Table 7.2: Change in Marital Fertility Rate(MFR) in Korea. Source: Korea National Statistical Office (2005)

The main objective was to raise Total Fertility Rate (TFR) from current 1.2 to 1.6 children per women by 2020 (the average for OECD countries). In Table 7.2 we can see the change in Marital Fertility Rate in Korea from 1970 to 2004. However the tendency is very difficult to revert in a society that provides job opportunities for women and their increased reluctance to get married compared with the past (Korea Institute for Health and Social Affairs, 2006). For instance, in 2009 252,000 births less were registered, down from 493,200 in 2007. Moreover, the TFR declined from 1.19 in 2008 to 1.15 in 2009 as well as marriages from 343,600 in 2007 to 309,800 in 2009.

This evidence shows there is no single solution to increase low birth rates in South Korea (Haub, 2010) and there is a need to establish programs and policies by government and businesses. As part of Korean commitment to that and supported by Korean government, we had the change to use our approach for parallel demographic simulation with ETRI. The final goal of ETRI's project

Region name	Real population	Simulated families
Seoul	10935230	109352
Busan	3882389	38823
Daegu	2284191	22841
Incheon	2065866	20658
Gwangju	1223784	12237
Daejeon	1134843	11348
Gyeonggi-do	6613094	66130
Gangwon-do	1554106	15541
Chungcheongbuk-do	1403633	14036
Chungcheongnam-do	1870723	18707
Jeollabuk-do	2027454	20274
Jeollanam-do	2283108	22831
Gyeongsangbuk-do	2872170	28721
Gyeongsangnam-do	3846825	38468
Jeju-do	505784	5057
Total	44503200	445032

Table 7.3: List of South Korean regions included in the model and their population in 1990. Source: Korea National Statistics Office

is to have a framework that not only forecasts demographic tendencies, but also helps on decision making processes derived from demographic changes.

In the following section, we introduce the model we built for Korean demographics and the simulation experiments for 100 years.

7.2.2 Korean model with Yades

In this section we will introduce how we built a model for South Korean demographics using Yades. As in the previous case study, we are including the two types of agents provided by our framework: families and regions. We covered fifteen out of a total of seventeen regions of the country, which are listed in Table 7.3. Unfortunately, the Korean National Statistics Office did not had enough information about the population in Ulsan and Sejong so we could not incorporate them in the model. In this model, we scaled the population by 100 as shown in the third column of Table 7.3.

In terms of the simulation scenario, there are three main differences with the previous Gambian case:

- On one hand, in this study we run the simulation model for a period of 100 years (from 1990 to 2090) with the idea to compare the evolution of demographics up to 2013 and see the expected trend far beyond. However,

Model component	**Source of data and type**
Mortality	Unique gender-based life table for all regions (source: KNSO)
Education Status	Five states: Dependent, Elementary school, Middle school, High schoool and College or university Transitions for the state diagram (source: KNSO) Duration ~ Uniform
Marital Status	Six states: Single, First-married, Married, Separated, Divorced and Widowed Transitions ~ Uniform Closed Model for Mate-Matching (Billari et al., 2007)
Fertility	Assume a maximum fertile age (assumption by ETRI) Age-specific fertility rate (source: KNSO) Only married women have children (assumption by ETRI)
Domestic Migrations	Probability matrix (source: KNSO) Only single male without children migrate (assumption by ETRI)
Time	Transitions change with time from 1990 to 2013. From 2013 to the end simulation time they remain constant

Table 7.4: Summary of data sources and types for the Gambian demographic model in Yades

the comparison was not possible since we did not keep the proportion ratio of people across regions.

- On the other hand, we are relying mainly on data from Korean National Statistics Office. Moreover, this data includes changes over time. That means, for instance, that birth function will change from one year to another to adjust to the provided information. Data is available from 1990 to 2013. Continuous data in demographic components was new in Yades, so we had to define additional functionalities to use it.

- Finally, ETRI required to limit migrations to domestic movements across Korean regions, and do not include international migrations (both immigrations and emigrations). In the context of this country, this requirement is understandable since between 1945 and 1985 the urban population in South Korea grew from 14.5% to 65.4% of the total population, specially in Seoul which concentrated 23.8%. That phenomena is mostly attributable to migrations rather than to natural growth (South, 2013).

Table 7.4 summarizes the datathat was used in each demographic component of our simulation model and the different sources that were used. As opposite to the Gambian case study, in Korea in Korean National Statistics Office (KNSO) has quite a large number of quantitative data in their census. Therefore, most of the variables in our model relied on it. Next subsections describe the model

	First Married	Remarried
Without children	0.75	0.25
With children	0.5	0.5

Table 7.5: Probabilities assumed for the proportion of people who were married for the first time or not with or without children

in terms of how a population is initialised, how it is updated over time in each demographic component of the model, and which data related issues should be taken into account with current available data.

Initialisation

Homogeneous initial population settings for the 17 regions where calculated from the Korean micro census from 1990. According to that, the setup of the simulation was defined, including

- the proportion of different age groups in the community,
- the proportion of different types of families by age group,
- the proportion of different education status by age group,
- the proportion of different marital status by age group, and
- the proportion of the number of children in a family.

The only information that was not available on the micro census was the number of people who were married for the first time, or re-married. Therefore, we had to assume this data with ETRI. The current simulations were done with the probabilities from Table 7.5, depending on if people had children or not.

Mortality

Life tables were provided on 6/26/14 taking into account gender and marital status. We made some adjustments on the code to adapt to that. However, they were only available for 2000, 2005 and 2010. Later on 7/25/14 we were requested to use a mortality table by gender from 1990 to 2012. This required a changed on the mortality model logic. The current version is adjusted to this last mortality. As a result, the maximum age in the simulator is set to 120. When an event of mortality is scheduled, we assumed the time to execute this event is set to up randomly between 1 and 12 months. It is also important to mention than in South Korea people retired at 54.

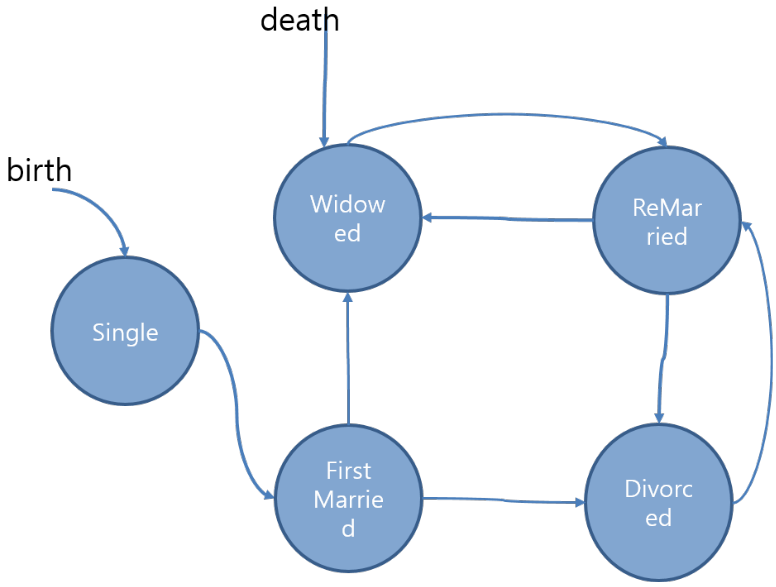


Figure 7.16: Marital status diagram of five states for South Korea.

Fertility

The periodical events related to births where specify with the following data:

- The fertility age interval for women in the population, was set from 16 to 45.
- Birth only takes place on married women, therefore in marital status states First Married or Remarried
- The logical rules guiding birth model, where built from the Age-Specific Fertility Rate provided from 1990 to 2012.
- When the simulator calculates a new birth should take place, we assume the event will take place between 9 months (pregnancy) and 1.9 years.

Marital Status

Marital status was specified with an homogeneous transition diagram for all regions. It is shown in Figure 7.16.

The different between first marriage and remarriage was incorporated in the model due to the data provided. Transitions between status where either provided by ETRI from Korean National Statistics Office or inferred (for instance, the transition from married to widowed was calculated with mortality data). The data included in the model sets transitions from 1990 to 2012. There was a missing value for year 2000 in the transition for remarriage after divorce, so 1995 data was assumed for this transition. The duration in a status was set to be from 1 to 1.5 years, according to the discussion with ETRI.

As in the previous case study, we assume when a family unit splits (for divorce or separation for instance) children will be split evenly in the new family units. Moreover, we assume a closed marriage model with the same simplification of the match making by (Billari et al., 2007), as described in Section 7.1.2.

Education status

For the education component, we assume five states: dependent, elementary school, middle school, high school and college or university. Transitions among states were quantified with data from the Korean National Statistics Office. The lack of data to determine durations among states made as to assume durations follow an uniform distribution $\mathcal{U}(1, 5)$ in years. The education component also calculates when an individual above a specified age leaves the current family and forms a new single person family.

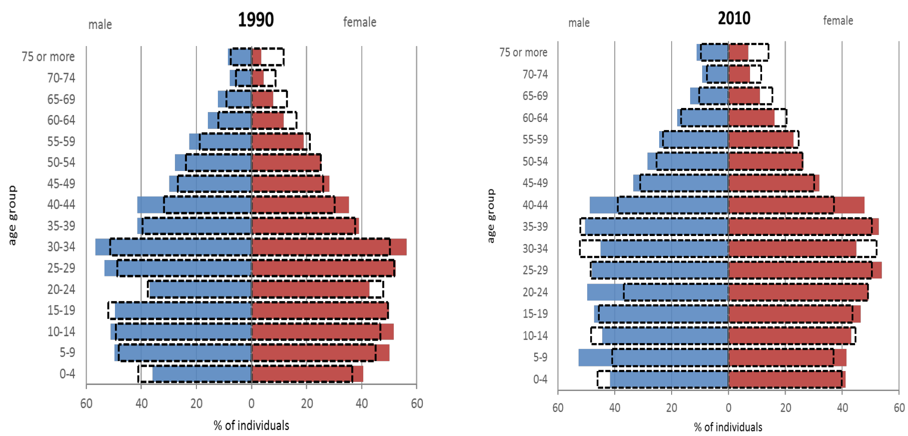
Domestic migrations

A migration probability matrix was built with the data of Korean National Statistics Office provided by ETRI. It contains the probability to migrate from one region to another. With this probability matrix, a cumulative distribution function (CDF) is calculated to decide the region to migrate from the current region. In the literature on Korean migration, single-member migration is seen as a strategy of wealthier households (Mobrand, 2012). Although it can be argued this tendency might have changed over the last years, ETRI requested to explicitly model that only male in divorce, single or widowed marital status could migrate.

7.2.3 Simulation results

All the experiments were done in Marenostrum 3 supercomputer, located in Barcelona Supercomputing Center, whose architecture has been described before. Results presented in this section have been obtained from an average of 100 simulation runs. In this case, we used up to 17 nodes, with 1 processor each (1 per region). In this section, we present some of the results for simulation years 2010, 2030, 2050 and 2090.

We started the simulation building the initial set-up with data from 1990. Since ETRI provided data from the Korean National Statistics Office (KNSO) census, we stopped the simulation after 20 years to check the evolution of the population. The results are displayed in Figure 7.17 which shows the population pyramid for 1990 and 2010 and their differences with real data. We can observe several things. In the initial set-up, men from ages 0 to 4 and 15 to 19 are a bit under-represented, On the contrary, the number of men from 25 to 69 is higher than in real data. For women, we see an under-estimation in age groups 20-24 and from 55 to 75 or more while there then number of females in 0-4, 5-9, 10-14, and 30-35 surpasses the census records. Those results make us think we need to improve the initialisation of the simulation set-up, although the simulation results look quite faithful to what is expected.



(a) initial population **(b)** population after 10 years

Figure 7.17: Population pyramid for the initial population (in 1990) and after 20 years (in 2010). Black dot lines correspond to the real data from the KNSO.

After 20 years, we display the progression of South Korean demographics in Figure 7.17 (b). If we look at males from 20 to 24 and 40 to 44, we see the real population is much lesser than the natural population growth and we therefore fail to project those groups. We can think of several possibilities to explain that. The most probable is that emigrations have caused this high decrease. Since we did not include them in our model, we are unable to adjust the population in those ages. However, other plausible explanation could be that either the mortality of those groups is not well adjusted or the model is not incorporating it adequately. A similar issue happens with males in the groups 30-34 and (in lesser measure) 35-39. The real data shows lower numbers than a natural growth and we observe differences with the simulation results. The arrival of immigrants in those age groups could explain it.

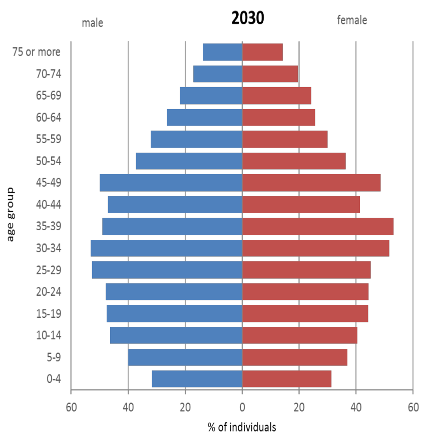
Looking at the progression of females we can also observe several differences:

- it seems that the fertility of the simulation is slightly higher than real records. This is observable either for the initial and the 20-years-after population pyramid.
- Real data for 10-14, 15-19 and 30-34 groups in 2010 is higher than the natural population growth. Again, the hypothesis of immigrant arrivals to

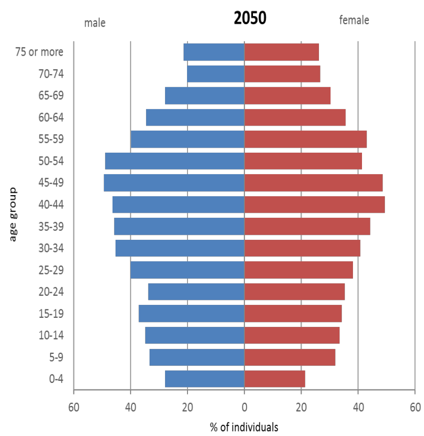
the country (with their families) could explain this phenomena.

- The hypothesis of emigration to explain a not justified decrease in the population might be true for females between 35 and 39, and specially from 40 to 44.
- Simulation results for retired women (over 55) are significantly lower than real data, as a result of the accumulated error in the initial population from 1990.

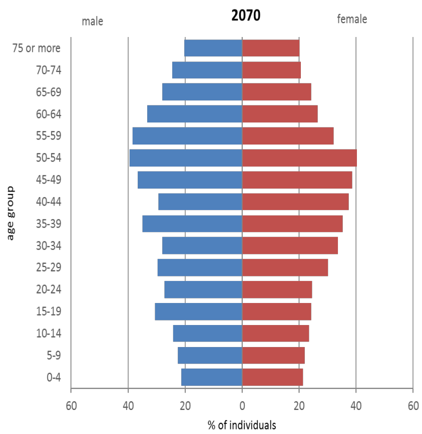
Despite the differences, the projected population in 2010 looks quite similar to census records from the Korean National Statistics Office.



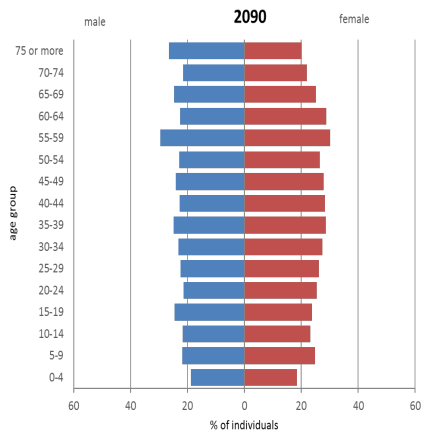
(a) projection for 2030



(b) projection for 2050



(c) projection for 2070



(d) projection for 2090

Figure 7.18: Projected population pyramids for 2030, 2050, 2070, and 2090 obtained from simulation results.

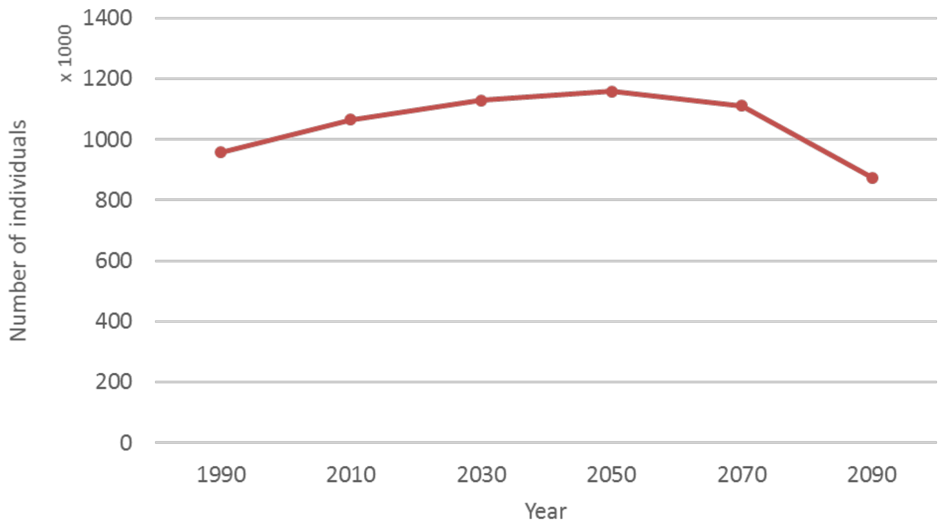
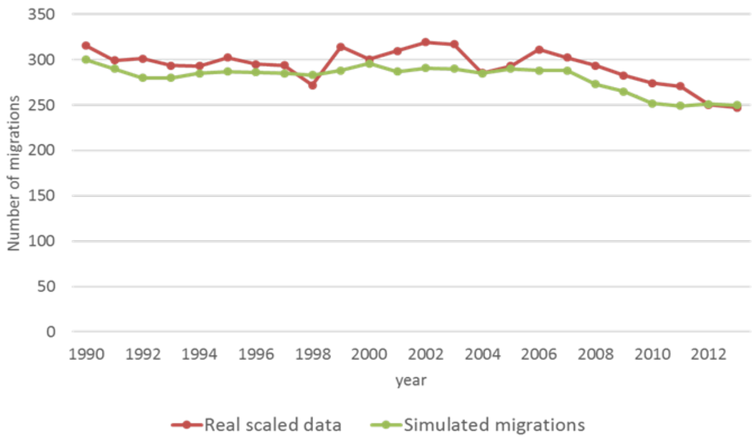


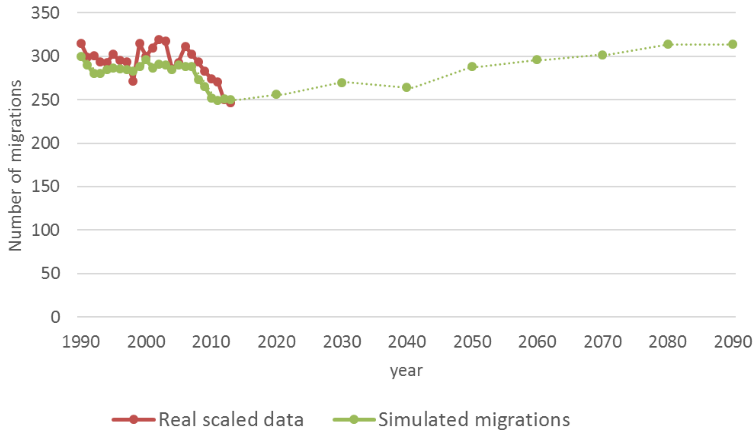
Figure 7.19: Evolution of the total population during 100 years of simulation

From this point, we need to be aware we do not have more data to validate next results of the simulation. Figure 7.18 shows the projections for every 20 years up to the simulation end time (in 2090). As we can see, as time goes on the population pyramid of the population becomes more narrow. This issue correspond with the common idea that South Korea would collapse in the next decades (of Economic, 2001; Coleman, 2002; Howe et al., 2007). In 2030 we can observe the fertility already diminish, which can be understood since the simulation is working with a constant birth rate taken from 2013 data of the Korean National Statistics Office. Moreover, we see the 2010 pyramid suits the natural growth as we expected, including the accumulated error for elder women (over 75 in this case) with respect to the initial population. During 2050 and 2070 the pyramid base is reducing in size with time and there is a progressive increment of elder population. By 2090 the pyramid appears very narrow as a consequence of an accumulated decrease of fertility during the last 80 years and the accumulated number of retired people. Although medical advancements and development tend to reduce mortality rates at elder ages, we did not include any variation in the simulation to decrease mortality rates. However, it would be interesting to enhance the model with information to this respect. Figure 7.19 shows the resulting evolution of the total number of individuals during 100 years of simulation.

If we have a look at the progression of domestic migrations during the simulation, we find a positive trend with time, as shown in Figure 7.20. The figure shows first the evolution of migrations from 1990 to 2010 with respect to real data. As we can see the tendency is correct for those years, although it could be more adjusted. However, we are able to reproduce 95.46% of the domestic migrations in that period. Figure 7.21 plots the distribution of age groups who migrate from 1990 to 2010 and compares it with census data. Since our simulation size is a sample of current population of South Korea, we scaled the results for single migrations to make the comparison. Figure 7.20 also shows the progression provided by simulation results up to 2090. The simulation forecasts that the number of domestic migrations will increase over the next years. However, we need to recall that our model is not taking into account the economic status of individuals and it is based just on probabilities provided by current records. In the future, it would be worth to explore if we can improve the accuracy of these results by including economic variables in the model.



(a) domestic migrations from 1990 to 2010



(b) domestic migrations from 1990 to 2090

Figure 7.20: Number of domestic migrations during the simulation vs. census data from KNSO in the period 1990-2010 (a), and for the rest of simulation time to 2090 (b)

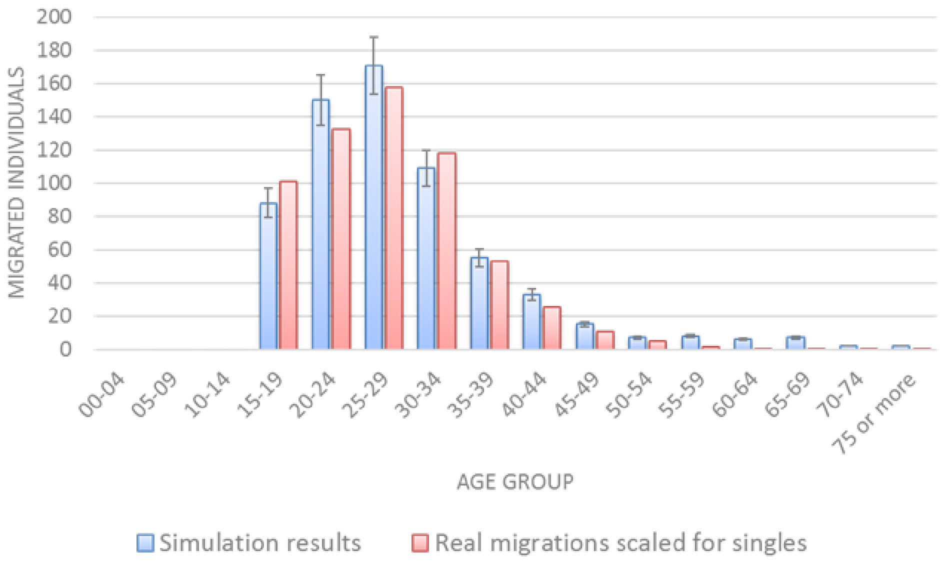


Figure 7.21: Average number of domestic migrations per age group after 10 years of simulation

In terms of computer resources, the average time of the simulations was 1056.0 sec. (17.6 min), using 15 processors of one node in Marenostrum 3 supercomputer. In this case, since the population is very big we are unable to calculate the simulation time in sequential and properly calculate the increased performance ratio. However, we can state in this case study we multiplied the population by 62.12 with respect to the previous case study and the simulation time increased just by 8.21.

7.3 Conclusions

The presented case studies have been of great value to test our demographic simulation methodology. In this exercise, we revised and questioned some of the decisions made on the implementation of the different demographic components included in Yades. Table 7.6 summarizes the comparison between both case studies, both in terms of complexity and performance. Our first model considers the aspects which best characterize the Gambian emigrant population in Spain: high fertility, partnership trends, deaths, labour market, and mobility. Data was obtained from census and country surveys on migration patterns. Behavioural rules were built from ethnographic records of anthropological studies (Kaplan Marcusán, 2005; Bledsoe and Sow, 2006; Bledsoe et al., 2007). The simulation results have been compared to the observed distribution of the Gambian population in 2011. Through the implementation of behavioural simple rules at the individual

Parameter	Value	
	Gambians in Spain	**South Korea**
Simulation time	10 years	100 years
Initial simulation time	2001	1990
Final simulation time	2011	2090
Initial number of families	7843	445032
Initial number of individuals	15402	956819
Domestic migration rate	30% homogeneous	probability-matrix based
Immigration rate	estimated arrivals per region	not included
Number of simulated domestic migrations	2444	6719 (in 1990-2010, the 95.46% of total registered migrations)
Time	transitions do not change with time	transitions change with time (1990-2013). From 2013 they remain constant
Performance		
Number of processors	8	15
Mean simulation time	128.7 sec (2.15 min)	1056.0 seconds (17.6 min)
Increased performance ratio	8.015	62.12 increase on population size increase than case study 1 8.21 increase on simulation time than case study 1
Average memory used	9.4 GB	11.13 GB
Max memory used	12.7 GB	18.32 GB

Table 7.6: Comparison between the cases studies of the Gambian emigrants in Spain and the South Korean demographic models in Yades

level, the heterogeneity of the group and its residential mobility is captured in the model.

As we have seen, demography is a complex phenomena difficult to simulate if we want to incorporate the anthropologist's perspective. We agree that the current work still has some limitations, among which two points seem to be important to understand Gambian immigrants dynamics. First, we did not consider Bledsoe et al. (2007) arguments on reproducing the high fertility with the circulation of co-wives between the Gambia and Spain. The problem here is to quantify the magnitude of this phenomenon, which skips current official statistical records. Second, we did not consider family reunification processes or the possibility to arrange marriages from the origin country. Again, changes in the immigration law and lack of statistical records is an impediment for including them in the model. However, these two limiting points should not totally call into question the validity of the methodology itself. In fact, agent-based approach has shown to be useful for the study of migrations flows and policy analysis already (Andriamasinoro and Angel, 2012; Kniveton et al., 2011).

With the case of South Korean demographics we had the chance to test how Yades is behaving in forecasting demography. The current results from 1990 to 2010 look promising and increase our confidence in the simulation model. The projection for the following years up to 2090 looks also good, particularly if we compare it with other forecast demographics studies made by agencies such as the United Nations (of Economic, 2001). As a result, we believe Yades could be used to complement other techniques commonly used to generate population projections.

As a final conclusion, the proposed case studies suits well the high-degree complexity of general agent-based models because the development of the population is affected by a large number of interrelated and dynamics factors: gender, age, economic level, family, matchmaking and reproduction patterns, life cycles and so on. Moreover, the proposed agent-based approach is interesting for studying these particular phenomena since it provides a way to gain insight of the population at an individual-level. Thus, with our work we show the potential of agent-based simulation on group dynamics with two real case studies. Furthermore, this work will increase Yades credibility for the community of modellers and policy makers, showing the possibilities agent-based modelling tools can provide to assist them in the understanding of demographic processes to guide decision-making or exploration of "what if" situations.

Chapter 8

Conclusions

We conclude offering a summary of the topics discussed in the previous chapters, as well as a summary of the contributions and a discussion of their significance. No research is ever finished and the one discussed in this thesis is no exception. First, Section 8.1 exposes our main research findings. Second, we revise the publications of this research. Finally, some possible directions for future work are proposed in Section 8.3.

In this thesis, we presented a methodology for simulating population socio-demographic dynamics at a large-scale. First, we revised classical modelling and simulation techniques and showed their usefulness for social modelling (see Chapter 2). A particular emphasis was done to introduce the advantage and cons of the agent-based approach for simulating complex socio-dynamics. Then, we exposed the current state of agent-based tools and their lack of support for large-scale systems. Second, Chapter 3 addressed some of the technological challenges social researchers confront in performing large-scale social simulations. We believe social simulation community should be aware of them to boost the research on the field. Third, our framework was introduced in Chapter 4 showing the potential of using agent-based modelling for demographic simulation in a parallel environment. This framework is composed by a parallel simulation library and a user interface meant for modellers such as demographers, anthropologists or policy makers who have little knowledge about computer programming. Our goal is that finals user reap the performance offered by parallel computers transparently. We later test the performance of our tool in two High Performance Computing infrastructures: the High End Cluster at Lancaster University, UK, and in Marenostrum 2 and 3 supercomputers in Barcelona, Spain. The results of those tests are presented in Chapter 5. Fifth, we performed some validation work of our tool following a white-box validation approach and examine the results in Chapter 6. Finally, we tested our framework with two case studies (presented in Chapter 7): the evolution of Gambian immigrants in Spain and South Korean demographics. Both real scenarios demonstrated the applicability of our methodology and help us to evaluate it.

8.1 Thesis contributions

In this thesis, the following list of contributions have been made:

- The proposal of an application of agent-based modelling to demography at a large-scale. Although agent-based models have already been used for simulating demographics (see Section 2.3.2 and (Billari et al., 2003, 2006) work), the application of parallel simulation to this specific research topic is new. On one hand, we decided to prototype social units as agents, with attributes and methods, versus other statistical/quantitative models used already. The designed architecture allows to expand the characteristics and processes of agents (which can be individuals, groups of individuals or collectives) on demand. This approach has been showed to be important for social sciences research, since it permits to add new behaviours and relationships among people. On the other hand, we analysed the performance of our approach and showed the framework has a good scalability, which highly depends on the architecture infrastructure configuration such as memory, latency or heterogeneous processing. We proved that migrations are one of the key elements to determine the limit of scalability, since they require communications through the computer network. The increasing complexity of social models and the management of large data sets makes us think this application would be of high interest in the future of both demographic and simulation research.

- The development of a framework designed to approach parallel agent-based simulation to social scientists and modellers. The tool is useful for demographers, anthropologists and other social practitioner since it allows a flexible modelling to easily incorporate already existing social models. Consequently, we provide an interface to visualize the information in a way that adapts to their needs. Furthermore, our approach allows to take into account not only variables at the individual level, but also at the collective level and to combine both quantitative and qualitative data. Qualitative information is crucial for some social studies such as anthropology or sociology. Furthermore, this methodology has potential in areas such as epidemiology, an area we are starting to work in, since epidemic models rely on a socio-demographic base model to deep on the dynamics that drive to infection and spread of diseases.

- We analysed the sensibility of agent-based models in a High Performance Computing context and review the strategies to better profit parallel architectures. The knowledge and expertise in High Performance Computing, programming models, and computer architecture from Barcelona Supercomputing Center allow us to enrich socio-demographics simulations. Our works highlight there is a need of (a) training social researchers and modellers in the basics to use this new paradigm, and (b) learn about the singularity of social complex systems and the underlying models on population dynamics. For the social sciences, the critical point is not a methodology to visualize simulations (as in traffic simulation or cellular growth), but the capacity to scale and detect the tasks or processes that highly affect the performance. We showed the parallelism of agent-based models depends notably in their size, the inherent agent complexity, the communication among agents, and the management of space where agents move and interact. From the point of view of modellers, those are crucial points that need to be taking into account when choosing a parallel or distributed simulation tool for population dynamics. From the computer science perspective, it poses a major challenge in terms of the methods, programming models and algorithms that need to be developed to confront a efficient large-scale social simulations.

- We performed two case studies to apply our methodology to real scenarios of social interest. First, we modelled the complex dynamics of Gambian immigrants in Spain. Here, the ethnographic information was crucial for building the model and understanding the results of simulations. Our assumptions where based on current research in anthropology and the lack of information to quantify certain social dynamics (for instance on the familiar reunification flows or the co-wives management and distribution between Gambia and Spain). Despite these simplifications, the projected population presents similarities with statistical records and encourages us to further enhance the methodology. Second, we modelled South Korean demographic evolution to test if our approach can be useful for forecasting demography. The complexity of population dynamics of South Koreans was taken into account and added to the model with quantitative and qualitative data. The interest of ETRI research institute is to understand those dynamics at a large scale with agent-based modelling techniques. The impact of our work in Korea will allow Korean government to build a platform to take future decision concerning the social policies and economy.

As a conclusion, our work boost the research in this interdisciplinary field, bringing closer different disciplines such as Parallel Computing, Operations Research simulation and social areas of Demography and Anthropology. From this interdisciplinary perspective, it opens new research lines since a socio-demographic engine is the base for other models from epidemiology, anthropology, sociology or economics.

8.2 Publications

The contributions of this work have led to a number of articles in both international conferences and journals. In this section, they are classified attending to their relevance.

8.2.1 Journals

Manuscripts submitted

- Cristina Montañola Sales, Bhakti S.S. Onggo, Josep Casanovas-Garcia, Jose M. Cela-Espin, and Adriana Kaplan-Marcusán. Yades: A parallel simulation tool for demography. *Parallel Computing*, 2015. (submitted)
  This work presents the design and development of our framework, along with the performance tests done in Marenostrum 2 and 3.

8.2.2 Articles in conference proceedings

- B.S.S. Onggo, C. Montañola Sales, and J. Casanovas-Garcia. Performance Analysis of Parallel Demographic Simulation. In *Proceedings of the 24th European Simulation and Modelling Conference, 25-27 October 2010, Hasselt, Belgium (ESM10)*, pages pp. 142–148, Ostend, Belgium, 2010. Eurosis-ETI
  This paper proposed some performance tests of a first version of our framework in Lancaster High End Computing cluster.

- Cristina Montañola-Sales, Bhakti S.S. Onggo, and Josep Casanovas-Garcia. Agent-based simulation validation: A case study in demographic simulation. In *SIMUL 2011, The Third International Conference on Advances in System Simulation*, pages 101–107, 2011. ISBN 978-1-61208-169-4
  The white-box validation work introduced in Chapter 6 was presented in this paper.

- Cristina Montañola Sales, Josep Casanovas-Garcia, Adriana Kaplan-Marcusán, and Jose M. Cela-Espín. Demographic agent-based simulation of Gambians immigrants in Spain. In Francisco J. Miguel, Frederic Amblard, Joan A. Barcelo, and M. Madella, editors, *Advances in Computational Social Science and Social Simulation, Proceedings of the Social Simulation Conference, Barcelona, 1-5th Sept. 2014*, Barcelona, Spain, 2014b. ESSA. doi: 10.13140/2.1.3052.4166. URL `http://ddd.uab.cat/record/125597`
  In this paper, we presented the results of our methodology applied to Gambian immigrants in Spain from Chapter 7.

Conference posters

The following conference posters describing the framework were presented in three different research communities:

- Cristina Montañola Sales, Josep Casanovas-Garcia, Jose Maria Cela-Espín, B.S.S. Onggo, and Adriana Kaplan-Marcusan. Parallel simulation of large population dynamics. In *ACM SIGSIM Conference on Principles of Advanced Discrete Simulation (PADS)*, Montreal, Canada, May 2013a. ACM SIGSIM

- Cristina Montañola Sales, Josep Casanovas-Garcia, Jose Maria Cela-Espín, B.S.S. Onggo, and Adriana Kaplan-Marcusan. Agent-based simulation of large population dynamics. In *European Conference on Complex Systems, 16-20 September 2013, Barcelona, Spain*, Barcelona, Spain, September 2013c

- Cristina Montañola Sales, Josep Casanovas-Garcia, Jose Maria Cela-Espín, B.S.S. Onggo, and Adriana Kaplan-Marcusan. Parallel simulation of large population dynamics. In *Winter Simulation Conference, 8-12 December 2013, Washington D.C.*, Washington D.C., December 2013b. Informs

The graphical user interface described in Chapter 4, Section 4.4 was presented in the following conferences:

- C. Montañola Sales, J. Casanovas-Garcia, B.S.S. Onggo, J. M. Cela-Espín, and A. Kaplan-Marcusán. Approaching simulation to modelers: A user interface for large-scale demographic simulation. In A. Tolk, S. D. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, editors, *Proceedings of the 2014 Winter Simulation Conference*. ACM, December 2014a

- Cristina Montañola Sales, Josep Casanovas-Garcia, Bhakti S.S. Onggo, and Zengxiang Li. A user interface for large-scale demographic simulation. In Juan Guerrero, editor, *Procceding of the IEEE 6th International Conference on Cloud Computing Technology and Science (CloudCom 2014)*. IEEE, December 2014c

8.2.3 Book chapters

- Cristina Montañola Sales, Xavier Rubio-Campillo, Josep Casanovas-Garcia, Jose Maria Cela-Espín, and Adriana Kaplan-Marcusan. Overview on Agent-Based Social Modelling and the use of formal languages. In Pau Fonseca, editor, *Formal Languages for Computer Simulation: Transdisciplinary Models and Applications*, pages 333–377. IGI Global, Hershey PA, US, 2014e. doi: 10.4018/978-1-4666-4369-7.ch011. ISBN 978-1-4666-4371-0
  This chapter presented the state of the art of social modelling and simulation described in Chapter 2.

- Cristina Montañola Sales, Xavier Rubio-Campillo, Josep Casanovas-Garcia, Jose M. Cela-Espín, and Adriana Kaplan-Marcusán. Large-scale social simulation, dealing with complexity challenges in high performance environments. In Diana Francisca Adamatti, Graçaliz Dimuro, and Helder Coelho, editors, *Interdisciplinary Applications of Agent-Based Social Simulation and Modeling*, chapter 7, pages 106–123. IGI Global, May 2014d. doi: 10.4018/978-1-4666-5954-4.ch007
  The state of the art in parallel social simulation introduced in Chapter 3 was published in this book chapter.

8.3 Future directions

There are two main areas where our work will continue on developing. From the point of view of the parallel simulation, there are three main topics interesting to improve our methodology:

- The exploitation of the multi-core capacity of the node. The design of our architecture is currently using a single processor to simulate each region and each (large) set of families. However, most of computers nowadays have a multi-core architecture. We would like to be able to use several processors to increase the performance of a simulation in one node. To do that, we need to study the possibility of using an application programming interface to support multi-platform shared memory multiprocessing such as the standard Open Multi-Processing (OpenMP) (Dagum and Menon, 1998) or OmpsS (Duran, Ayguadé, Badia, Labarta, Martinell, Martorell, and Planas, 2011), developed by Barcelona Supercomputing Center. Although the mathematical calculus complexity of our framework is not very high, the management of large sets of families and their events lead us to think we could optimize the simulation performance with such approach.

- Dynamic balance of computer workload. As we have seen, there is a high sensibility to heterogeneous size populations in terms of computer performance. Differences in population sizes can come from (a) the initial distribution of agents across regions, (b) movements such as domestic migrations, and (c) arrivals in the system (immigrations). That is why it would be very interesting to develop some dynamic balancing behaviour in the system. In Chapter 3, we saw there are some approaches to do that. However, the overhead of load balancing solutions should be also taken into account, since communications between nodes can slow down substantially the simulation execution time. Furthermore, we have seen the management of a large set of agents and the optimistic synchronization of the simulation encounter some problems in terms of memory demand, especially RAM. Virtual allocation of population across the computer network would be interesting to study to optimise the use of memory.

- Use of more fine-grained data.

From the point of view of the demographic model, we would like to highlight four different ideas of features which could highly improve our results:

- To add household structures. Introducing the concept of household would allow one or more members of the same family unit to live in separate regions. This social structure is particularly interesting for policy analysis since numerous policies are not applied to individuals, but to the household set. We are currently developing this new feature which will be helpful for our next epidemiological study and the enhancement of the South Korean demographic simulation.

- To add the economic framework. In our work, we followed the software engineering principle of extensibility, where the implementation takes future growth into consideration. Therefore, it is possible to incorporate in the model economic effects such as labour market particularities, social trade networks, economic policies and so on. For instance, we could allow modellers to introduce periodical changes in simulation parameters such as life table and fertility rates based on economic effects at the region level. This new feature can be easily achieved by adding a set of new events in regions.

- To increase the credibility of our framework. This topic can be done at to levels. First, go beyond the white-box validation and develop further the comparison of Yades' results to other microsimulation models would increase the credibility of our methodology. Second, we need to review and improve our interface with end users. We have consulted an anthropologist, who uses demographic models, during the design of the user interface to understand how our simulation framework might be used by the end-users. In the future, we will invite a number of end-users to evaluate our tool. Experts reviews can be done late in the design process (Shneiderman and Plaisant, 2010). We also plan to expand our graphical tool to include a results feedback presentation panel, so a visual exploration of simulation output could be performed. Right now, the tool works with initial aggregate date for families and individuals. In the future, we would like to enhance it so data at the micro-level could be also used.

- To enrich the framework with models such as social dynamics, where individuals are influenced by one another's behaviour. This new feature would make possible to define models with more complex social behaviour and incorporate new interactions among people. Moreover, we could enrich the family agent at the individual level to simulate preferences, objectives or ethnics differentiate behaviour, profiting of Artificial Intelligence techniques such as reasoning, learning or autonomy of agents. This would be especially interesting for areas such as sociology or psychology.

We believe the presented future directions for our methodology would provide a way not only to improve the simulation model, but also to increase the scope of applications we can apply it. For instance, we are currently studying migration patterns of Singapore with researchers of the Institute of High Performance Computing in Singapore, extending our framework to add more variables that influence migrations in that country. Moreover, our methodology is currently being expanded for the study of tuberculosis spread in Barcelona relying the demographic simulation model we proposed. We the ongoing work we expect to keep contributing to social and parallel computing research.

Bibliography

IEEE standard for modeling and simulation (M&S) high level architecture (HLA) - std 1516-2000, 1516.1-2000, 1516.2-2000, 2000.

Report on inquiry into teenage pregnancy, 5th report (session 4). Technical Report 355, Health and Sport Committee, Scottish Parliament, 2013.

United Nations World population Day, 2013 Theme: Focus is on Adolescent Pregnancy, July 2013. URL http://www.un.org/en/events/populationday/.

S. Ahmad and J. Billimek. Estimating the health impacts of tobacco harm reduction policies: A simulation modeling approach. *Risk Analysis*, 25(4):801–812, 2005.

Marco Ajelli and Stefano Merler. An individual-based model of hepatitis a transmission. *Journal of Theoretical Biology*, 259(3):478 – 488, 2009. ISSN 0022-5193. doi: http://dx.doi.org/10.1016/j.jtbi.2009.03.038.

F. Albiero, F.H.P. Fitzek, and M. Katz. Cooperative power saving strategies in wireless networks: an agent-based model. In *4th International Symposium on Wireless Communication Systems, ISWCS 2007*, pages 287–291. IEEE, 2007. doi: 10.1109/ISWCS.2007.4392347.

Rob Allan. Survey of agent based modelling and simulation tools. Technical report, Computational Science and Engineering Department, STFC Daresbury Laboratory, Warrington WA4 4AD, 2010.

Theodore T. Allen. *Introduction to Discrete Event Simulation and Agent-based Modeling*. Springer London, 2011. doi: 10.1007/978-0-85729-139-4.

Mark Altaweel, Lilian N. Alessa, and Andrew D. Kliskey. Social influence and decision-making: Evaluating agent networks in village responses to change in freshwater. *Journal of Artificial Societies and Social Simulation*, 13(1):15, 2010. ISSN 1460-7425. URL http://jasss.soc.surrey.ac.uk/13/1/15.html.

Fenintsoa Andriamasinoro and Jean-Michel Angel. Artisanal and small-scale gold mining in Burkina Faso: Suggestion of multi-agent methodology as a complementary support in elaborating a policy. *Resources Policy*, 37(3):385 – 396, 2012. ISSN 0301-4207. doi: 10.1016/j.resourpol.2012.04.004.

SM Arifin, G.J. Davis, and Y. Zhou. A spatial agent-based model of malaria: Model verification and effects of spatial heterogeneity. *International Journal of Agent Technologies and Systems (IJATS)*, 3(3):17–34, 2011. doi: 10.4018/jats.2011070102.

W Brian Arthur and Geoffrey McNicoll. Large-scale simulation models in population and development: What use to planners? *Population and Development Review*, pages 251–265, 1975.

Michael J. Ashworth and Marcus A. Louie. Alignment of the garbage can and NK fitness models: A virtual experiment in the simulation of organizations, 2002.

R. Axelrod. Advancing the art of simulation in the social sciences. *Complexity*, 3(2): 16–22, 1997.

Robert Axelrod and Ross A Hammond. The evolution of ethnocentric behavior. In *Midwest Political Science Convention*, pages 2–30, 2003.

Robert Axtell. Why agents?: on the varied motivations for agent computing in the social sciences. Technical Report 17, Center on Social and Economic Dynamics, The Brookings Institution, 1775 Massachusetts Ave. NW, Washington, DC 20036;, November 2000.

Robert Axtell, Robert Axelrod, Joshua M. Epstein, and Michael D. Cohen. Aligning simulation models: A case study and results. *Computational and Mathematical Organization Theory*, 1(2):123–141, 1996. doi: 10.1007/BF01299065.

Brian Baird. To Improve Public Policy. *Science*, 340(6131):432–433, 2013. doi: 10.1126/science.1237522.

S. Balbi and C. Giupponi. Agent-Based Modelling of Socio-Ecosystems: A Methodology for the Analysis of Adaptation to Climate Change. *International Journal of Agent Technologies and Systems (IJATS)*, 2(4):17–38, 2010. doi: 10.4018/jats.2010100103.

Steven C. Bankes. Agent-based modeling: A revolution? *Proceedings of the National Academy of Sciences of the United States of America*, 99(Suppl 3):7199–7200, 2002. doi: 10.1073/pnas.072081299.

Jerry Banks. *Handbook of simulation: principles, methodology, advances, applications, and practice*. Wiley-Interscience, 1998. doi: 10.1002/9780470172445.

Jerry Banks, John S. Carson, Barry L. Nelson, and David M. Nicol. *Discrete-event system simulation*. Prentice-Hall, 1999. Chapter 10.

Juan A. Barcelo, Jose A. Cuesta, Florencia Del Castillo, Jose M. Galan, Laura Mameli, Miguel Quesada, Jose I. Santos, and Xavier Vila. Patagonian ethnogenesis: towards a computational simulation approach. In *Proceedings of the 3rd World Congress on Social Simulation*, pages 1–9, Kassel, Germany, 2010.

Olivier Barthelemy, Scott Moss, Thomas Downing, and Juliette Rouchier. Policy Modelling with ABSS: The Case of Water Demand Management. Technical Report CPM-02-92, Centre for Policy Modeling, Manchester Metropolitan University, UK, 2001.

Mark A. Bedau. Artificial life: organization, adaptation and complexity from the bottom up. *Trends in Cognitive Sciences*, 7(11):505 – 512, 2003. ISSN 1364-6613. doi: 10.1016/j.tics.2003.09.012.

D. Scott Bennett. Governments, civilians, and the evolution of insurgency: Modeling the early dynamics of insurgencies. *Journal of Artificial Societies and Social Simulation*, 11(4):7, 2008. ISSN 1460-7425. URL `http://jasss.soc.surrey.ac.uk/`

11/4/7.html.

A. Bigbee, C. Cioffi-Revilla, and S. Luke. Replication of Sugarscape using MASON. *Agent-Based Approaches in Economic and Social Complex Systems IV*, pages 183–190, 2007.

F.C. Billari, Th. Fent, A. Prskawetz, and J. Scheffran. *Agent-based computational modelling: applications in demography, social, economic and environmental sciences*, volume X. Springer, 2006.

Francesco C. Billari, F. Ongaro, and Alexia Prskawetz. *Agent-Based Computational Demography: Using Simulation to Improve Our Understanding of Demographic Behaviour*, volume IX of *Physica Verlag Heidelberg*. Elsevier, 2003.

Francesco C Billari, Alexia Prskawetz, Belinda Aparicio Diaz, and Thomas Fent. The "wedding-ring": An agent-based marriage model based on social interaction. *Demographic Research*, 17(3):59–82, 2007. doi: 10.4054/DemRes.2007.17.3.

Eric Blanchart, Cristophe Cambier, Clive Canape, Benoit Gaudou, The Nhan Ho, Tuong Vinh. Ho, Cristophe Lang, Fabien Michel, Nicolas Marilleau, and Laurent Philippe. EPIS: A Grid Platform to Ease and Optimize Multi-agent Simulators Running. In Yves Demazeau, Michal Pechoucek, Juan M. Corchado, and Javier Bajo P'erez, editors, *Advances on Practical Applications of Agents and Multiagent Systems*, volume 88 of *Advances in Intelligent and Soft Computing*, pages 129–134. Springer Berlin Heidelberg, 2011. ISBN 978-3-642-19874-8. doi: 10.1007/978-3-642-19875-5_17.

C. Bledsoe, R. Houle, and P. Sow. High fertility Gambians in low fertility Spain: mutually entailed lives across international space. *Demographic Research*, (16): 375–412, 2007.

Caroline H Bledsoe. *Contingent lives: Fertility, time, and aging in West Africa*. The University of Chicago Press, Chicago, 2002.

Caroline H. Bledsoe and Papa Sow. Emancipation or marginalization: The dilemma of aging gambian women in spain. In *Paper For The European Population Conference, Liverpool, England*, 2006.

C. Bohk, R. Ewald, and A.M. Uhrmarcher. Probabilistic population projection with JAMES II. In *Proceedings of the 2009 Winter Simulation Conference*, Piscataway, NJ, 2009. IEEE Computer Society Press.

E. Bonabeau. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the United States of America*, 99(Suppl 3):7280–7287, 2002.

Andrea Bonaccorsi and Cristina Rossi. Why open source software can succeed. *Research Policy*, 32(7):1243 – 1258, 2003. ISSN 0048-7333. doi: 10.1016/S0048-7333(03)00051-9.

C. Bonnet and R. Mahieu. *Microsimulation modelling for policy analysis: challenges and innovations*, chapter Public pensions in a dynamic microanalytic framework: the case of France, pages 175–199. Cambridge University Press, Cambridge, UK, 2000.

Andrei Borshchev and Alexei Filippov. From system dynamics and discrete event to practical agent based modeling: reasons, techniques, tools. In *Proceedings of the 22nd international conference of the system dynamics society, July 25-29 2004,* *Oxford, England*, number 22, 2004.

François Bourguignon and Amedeo Spadaro. Microsimulation as a tool for evaluating redistribution policies. *The Journal of Economic Inequality*, 4(1):77–106, 2006. ISSN 1569-1721. doi: 10.1007/s10888-005-9012-6.

P. Box. *Integrating Geographic Information Systems and Agent-Based Modeling Techniques for Simulating Social and Ecological Processes*, chapter Spatial units as agents: Making the landscape an equal player in agent-based simulations, pages 59–82. Oxford University Press, USA, 2002. doi: 10.1111/0033-0124.5502028.

Mark Brady, Christoph Sahrbacher, Konrad Kellermann, and Kathrin Happe. An agent-based approach to modeling impacts of agricultural policy on land use, biodiversity and ecosystem services. *Landscape Ecology*, 27(9):1363–1381, 2012. ISSN 0921-2973. doi: 10.1007/s10980-012-9787-3.

Michael E Bratman. Intention, plans, and practical reason. 1999.

R.E. Bryant. A Switch-Level Model and Simulator for MOS Digital Systems. *IEEE Transaction on Computers*, 33(2):160–177, 1984.

M. Buchanan. Meltdown modelling. Could agent-based computer models prevent another financial crisis? *Nature*, 7256:680–682, 2009.

B. Burmeister, A. Haddadi, and G. Mattilys. Application of multi-agent sistems in traffic and transportation. In *Software ingeneering IEEE Proceedings*, volume 114, pages 51–60, 1997. doi: 10.1049/ip-sen:19971023.

R. Burton. *Simulating Organizations: Computational Models of Institutions and Groups*, chapter Validating and docking: An overview, summary and challenge. AAAI Press, Menlo Park, CA, 1998.

James Byrne, Cathal Heavey, and P.J. Byrne. A review of web-based simulation and supporting tools. *Simulation Modelling Practice and Theory*, 18(3):253 – 276, 2010. ISSN 1569-190X. doi: http://dx.doi.org/10.1016/j.simpat.2009.09.013.

C. Carothers, K.S. Perumalla, and R.M. Fujimoto. Efficient optimistic parallel simulations using reverse computation. In *Proceedings of the 13th Workshop on Parallel and Distributed Simulation*, pages 126–135, 1999.

C.J.E. Castle and A.T. Crooks. Principles and concepts of agent-based modelling for developing geospatial simulations. *University College London, Working Paper Series*, 110, 2006. URL `http://discovery.ucl.ac.uk/3342/`.

Avinuh Chandak and JC Browne. Vectorization of discrete event simulation. In *Proceedings of the 1983 International Conference on Parallel Processing: August 23-26*, page 359. IEEE Computer Society Press, 1983.

K.M. Chandy and J. Misra. Distributed Simulation: a Case Study in Design and Verification of Distributed Programs. *IEEE Transaction on Software Engineering*, 5(5):440–452, 1979.

Emile JL Chappin, Gerard PJ Dijkema, Koen Haziël van Dam, and Zofia Lukszo. Modeling strategic and operational decision-making: an agent-based model of electricity producers. In *The 2007 European Simulation and Modelling Conference, St. Julians, Malta, Eurosis*, 2007.

D. Chen, G.K. Theodoropoulos, S.J. Turner, W. Cai, R. Minson, and Y. Zhang. Large scale agent-based simulation on the grid. *Future Generation Computer Systems*, 24 (7):658–671, July 2008. ISSN 0167-739X. doi: 10.1016/j.future.2008.01.004.

Shih-Fen Cheng, Larry Lin, Jiali Du, Hoong Chuin Lau, and Pradeep Varakantham. An Agent-Based Simulation Approach to Experience Management in Theme Parks. In R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, editors, *Proceedings of the 2013 Winter Simulation Conference, Washington DC*, 2013.

S Anders Christensson, Alexander ER Woodcock, Derek K Hitchins, and Loren Cobb. Modeling the governance and stability of political dynamical systems. *The Cornwallis Group VIII: analysis for governance and stability*, page 123, 2004.

Claudio Cioffi-Revilla. A methodology for complex social simulations. *Journal of Artificial Societies and Social Simulation*, 13(1):7, 2010. ISSN 1460-7425. URL `http://jasss.soc.surrey.ac.uk/13/1/7.html`.

Claudio Cioffi-Revilla, J. Daniel Rogers, and Maciej Latek. The mason householdworlds of pastoral nomad societies. In C. Cioffi-Revilla K. Takadama and G. Deffaunt, editors, *Simulating Interacting Agents and Social Phenomena: The Second World Congress in Social Simulation*, Berlin and Tokyo, 2010. Springer.

S. Coakley, M. Gheorghe, M. Holcombe, S. Chin, D. Worth, and C. Greenough. Exploitation of high performance computing in the flame agent-based simulation framework. In *IEEE 14th International Conference on High Performance Computing and Communication 2012 (HPCC-ICESS)*, pages 538–545, 2012. doi: 10.1109/HPCC.2012.79.

Ansley J Coale and Donald R McNeil. The distribution by age of the frequency of first marriage in a female cohort. *Journal of the American Statistical Association*, 67(340):743–749, 1972.

David A Coleman. Replacement migration, or why everyone is going to have to live in korea: a fable for our times from the united nations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 357(1420):583–598, 2002.

N. Collier. Repast: An extensible framework for agent simulation. *Natural Resources and Environmental Issues*, 8(1):4, 2001.

Nicholson Collier and Michael North. Parallel agent-based simulation with repast for high performance computing. *SIMULATION*, 2012. doi: 10.1177/0037549712462620.

John Craig Comfort. The simulation of a master-slave event set processor. *Simulation*, 42(3):117–124, 1984.

Rosaria Conte, Nigel Gilbert, and Jaime Sichman. MAS and Social Simulation: A Suitable Commitment. In Jaime Sichman, Rosaria Conte, and Nigel Gilbert, editors, *Multi-Agent Systems and Agent-Based Simulation*, volume 1534 of *Lecture Notes in Computer Science*, pages 47–53. Springer Berlin / Heidelberg, 1998. ISBN 978-3-540-65476-6. doi: 10.1007/10692956_1.

G. Cordasco, R. De Chiara, V. Scarano, M. Carillo, A. Mancuso, D. Mazzeo, F. Raia, F. Serrapica, C. Spagnuolo, and L. Vicidomini. D-mason: Distributed multi-agent based simulations toolkit. 2011. URL `http://isis.dia.unisa.it/projects/dmason/`.

Andrew Crooks and Christian Castle. Principles and Concepts of Agent-Based Modelling for Developing Geospatial Simulations. Technical report, Centre for Advanced Spatial Analysis, University College London, September 2006.

Leonardo Dagum and Ramesh Menon. Openmp: an industry standard api for shared-memory programming. *Computational Science & Engineering, IEEE*, 5(1):46–55, 1998.

F.J. Dahlen. LaMPsim - an open source microsimulation tool in development. In *Presented at the 2nd General Conference of the International Microsimulation Association, June 8th to 10th, 2009, Ottawa, Ontario, Canada*, 2009.

Romulus-Catalin Damaceanu. An agent-based computational study of wealth distribution in function of resource growth interval using netlogo. *Applied Mathematics and Computation*, 201(1-2):371–377, 2008. ISSN 0096-3003. doi: 10.1016/j.amc.2007.12.042.

Paul Davidsson. Multi Agent Based Simulation: Beyond Social Simulation. In Scott Moss and Paul Davidsson, editors, *Multi-Agent-Based Simulation*, volume 1979 of *Lecture Notes in Computer Science*, pages 97–107. Springer Berlin Heidelberg, 2001. ISBN 978-3-540-41522-0. doi: 10.1007/3-540-44561-7_7.

R. Davies, P. Roderick, and J. Raftery. The evaluation of disease prevention and treatment using simulation models. *European Journal of Operational Research*, 150(1):53–66, 2003. doi: 10.1016/S0377-2217(02)00783-X.

David de Vaus. Diversity and change in Australian families: Statistical Profiles. *Melbourne, Australia. Australian Institute of Family Studies*, 2004.

Jeffrey S. Dean, George J. Gumerman, Joshua M. Epstein, Robert L. Axtell, Alan C. Swedlund, Miles T. Parker, and Steven McCarroll. Understanding Anasazi culture change through agent-based modeling. *Dynamics in human and primate societies: Agent-based modeling of social and spatial processes*, pages 179–205, 2000.

Andreu Domingo, Diana López-Falcón, and Alberto del Rey. Current status and trends of african migration to spain. In *International Seminar Beyond Plurality in the Africa Diaspora, 25-27 September 2008, Staffordshire, UK*, 2008.

Andreu Domingo, Diana López-Falcón, and Alberto del Rey. Current status and trends of African migration to Spain. In *Papers de Demografia - Centre d'Estudis Demogràfics*. Universitat Autònoma de Barcelona, 2010.

Andreu Domingo i Valls and Daniela Vono de Vilhena. Africans in the Southern European countries: Italy, Spain and Portugal. Documentos de Proyectos No.461 pp. 15-59, Development, institutional and policy aspects of international migration between Africa, Europe and Latin America and the Caribbean, 2012. URL `http://hdl.handle.net/11362/3960`.

Julie Dugdale, J Pavard, and B Soubie. A pragmatic development of a computer simulation of an emergency call center. In R. Dieng, A. Giboin, L. Karsenty, and G. de Michelis, editors, *Designing Cooperative Systems: The Use of Theories and Models*, pages 241–256. IOS Press, 2000.

Jill Bigley Dunham. An Agent-Based Spatially Explicit Epidemiological Model in MASON. *Journal of Artificial Societies and Social Simulation*, 9(1):3, 2005. ISSN 1460-7425. URL `http://jasss.soc.surrey.ac.uk/9/1/3.html`.

Deborah. Duong. Verification, Validation, and Accreditation (VV&A) of Social Simulations. In *Spring Simulation Interoperability Workshop, Orlando, April 12-16 2010*, 2010.

Alejandro Duran, Eduard Ayguadé, Rosa M Badia, Jesús Labarta, Luis Martinell, Xavier Martorell, and Judit Planas. Ompss: a proposal for programming heterogeneous multi-core architectures. *Parallel Processing Letters*, 21(02):173–193, 2011.

Bruce Edmonds. The use of models - making mabs more informative. In Scott Moss and Paul Davidsson, editors, *Multi-Agent-Based Simulation*, volume 1979 of *Lecture Notes in Computer Science*, pages 269–282. Springer Berlin / Heidelberg, 2001. ISBN 978-3-540-41522-0. doi: 10.1007/3-540-44561-7_2. URL http://dx.doi.org/10.1007/3-540-44561-7_2.

Bruce Edmonds. The Emergence of Symbiotic Groups Resulting from Skill-Differentiation and Tags. *Journal of Artificial Societies and Social Simulation*, 9(1): 10, 2006. ISSN 1460-7425. URL http://jasss.soc.surrey.ac.uk/9/1/10.html.

S. El Hadouaj, A. Drogoul, and S. Espié. How to combine reactivity and anticipation: the case of conflicts resolution in a simulated road traffic. *Multi-Agent-Based Simulation*, pages 157–167, 2001. doi: 10.1007/3-540-44561-7_6.

G.S. Epstein and I.N. Gang. The Influence of Others on Migration Plans. *Review of Development Economics*, 10(4):652–665, 2006.

J.M. Epstein. Why model? *Journal of Artificial Societies and Social Simulation*, 11 (4):12, 2008. URL http://jasss.soc.surrey.ac.uk/11/4/12.html.

Joshua M Epstein. Agent-based computational models and generative social science. *Complexity*, 4(5):41–60, 1999.

Joshua M Epstein. Remarks on the Foundations of Agent-Based Generative Social Science. *Handbook of Computational Economics*, 2:1585–1604, 2006.

Joshua M Epstein. Modelling to contain pandemics. *Nature*, 460(7256):687–687, 2009.

Joshua M. Epstein and Robert Axtell. *Growing artificial societies: social science from the bottom up*. The MIT Press, 1996. doi: 10.2307/1060800.

S. Eubank. Scalable, efficient epidemiological simulation. In *Proceedings of the 2002 ACM Symposium on Applied Computing*, pages 139–145, New York, NY, 2002. ACM Press.

J Doyne Farmer and Duncan Foley. The economy needs agent-based modelling. *Nature*, 460(7256):685–686, 2009.

Flávia F. Feitosa, Quang Bao Le, and Paul L.G. Vlek. Multi-agent simulator for urban segregation (MASUS): A tool to explore alternatives for promoting inclusive cities. *Computers, Environment and Urban Systems*, 35(2):104 – 115, 2011. ISSN 0198-9715. doi: 10.1016/j.compenvurbsys.2010.06.001.

J. Forrester. *World Dynamics*. Productivity Press, New York, NY, USA, 2nd edition, 1971.

I. Foster, Yong Zhao, I. Raicu, and Shiyong Lu. Cloud computing and grid computing 360-degree compared. In *Grid Computing Environments Workshop, 2008. GCE '08*, pages 1–10, 2008. doi: 10.1109/GCE.2008.4738445.

Martin Fowler and K. Scott. *UML Distilled: A Brief Guide to the Standard Object Modeling Languange*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2 edition, 2000. doi: 10.1109/MS.2005.81.

Richard M Fujimoto. Parallel discrete event simulation. *Communications of the ACM*, 33(10):30–53, 1990.

R.M. Fujimoto. *Parallel and distributed simulation systems*. John Wiley & Sons, Hoboken, New Jersey, 2000.

M. Gardner. Mathematical games: The fantastic combinations of John Conway's new solitaire game 'Life'. *Scientific American*, 223(4):120–123, 1970.

Floriana Gargiulo, Sônia Ternes, Sylvie Huet, and Guillaume Deffuant. An iterative approach for generating statistically realistic populations of households. *PloS one*, 5(1):e8828, 2010. doi: 10.1371/journal.pone.0008828.

Nicholas Geard, James M McCaw, Alan Dorin, Kevin B Korb, and Jodie McVernon. Synthetic population dynamics: A model of household demography. *Journal of Artificial Societies and Social Simulation*, 16(1):8, 2013. ISSN 1460-7425. URL `http://jasss.soc.surrey.ac.uk/16/1/8.html`.

G.N. Gilbert. *Agent-based models*. Sage Publications, Inc, 2008. ISBN 1412949645.

G.N. Gilbert and K.G. Troitzsch. *Simulation for the social scientist*. Milton Keynes: Open University Press, 2nd edition, 2005.

Nigel Gilbert and Steven Bankes. Platforms and methods for agent-based modeling. *Proceedings of the National Academy of Sciences of the United States of America*, 99(Suppl 3):7197–7198, 2002. doi: 10.1073/pnas.072079499.

Anna Glasier. Emergency contraception: Is it worth all the fuss? *British Medical Journal*, 333:560–561, 2006. doi: 10.1136/bmj.38960.672998.80.

M.F. Goodchild. *GIS, spatial analysis, and modeling*, volume 414, chapter GIS and Modeling Overview, pages 1–18. ESRI Press, Redlands, CA, 2005.

J. Gratch and S. Marsella. Tears and fears: Modeling emotions and emotional behaviors in synthetic agents. In *Proceedings of the fifth international conference on Autonomous agents*, pages 278–285. ACM, 2001. doi: 10.1145/375735.376309.

V. Grimm, U. Berger, F. Bastiansen, S. Eliassen, V. Ginot, J. Giske, J. Goss-Custard, T. Grand, S.K. Heinz, G. Huse, et al. A standard protocol for describing individual-based and agent-based models. *Ecological modelling*, 198(1-2):115–126, 2006. doi: 10.1145/375735.376309.

László Gulyás, Attila Szabó, Richárd Legéndi, Tamás Máhr, Rajmund Bocsi, and George Kampis. GridABM - templates for distributed agent based simulation. In *Open Grid Forum 28*, Munich, Germany, 2010.

T. R. Gurung, F. Bousquet, and G. Trébuil. Companion modeling, conflict resolution, and institution building: sharing irrigation water in the Lingmuteychu Watershed, Bhutan. *Ecology and Society*, 11(2):36, 2006.

A.B. Hailegiorgis, W.G. Kennedy, M. Rouleau, J.K. Bassett, M. Coletti, G.C. Balan, and T. Gulden. An agent based model of climate change and conflict among pastoralists in East Africa. In David A. Swayne, Wanhong Yang, A. A. Voinov, A. Rizzoli, and T. Filatova, editors, *Proceedings of the International Congress on Environmental Modeling and Software (IEMSS2010)*, Ottawa, Ontario, Canada, 2010.

Lynne Hamill. Agent-Based Modelling: The Next 15 Years. *Journal of Artificial Societies and Social Simulation*, 13(4):7, 2010. ISSN 1460-7425. URL `http://jasss.soc.surrey.ac.uk/13/4/7.html`.

E. Hammel, C. Mason, and C. Wachter. SOCSIM II, a sociodemographic microsimulation program, rev. 1.0, operating manual. Graduate Group in Demography Working Paper 29, University of California, Berkeley, 1990.

R. Hanneman. *Computer-assisted theory building: Modeling dynamic social systems*. Sage, Newbury Park, CA, 1988. doi: 10.1002/sdr.4260070208.

B. Harvey. *Computer science logo style: Symbolic computing*, volume 1-3. MIT Press, 2 edition, 1997.

Carl Haub. Did South Korea's Population Policy Work Too Well? *Population Reference Bureau*, 2010. URL `http://www.prb.org/Articles/2010/koreafertility.aspx(accessedFebruary2,2015)`.

Werner Haug, Paul A Compton, and Youssef Courbage. *The Demographic Characteristics of Immigrant Populations.* Number 38. Council of Europe Publishing, 2002.

Brian Heath, Raymond Hill, and Frank Ciarallo. A survey of agent-based modeling practices (january 1998 to july 2008). *Journal of Artificial Societies and Social Simulation*, 12(4):9, 2009. ISSN 1460-7425. URL `http://jasss.soc.surrey.ac.uk/12/4/9.html`.

Scott Heckbert. Mayasim: An agent-based model of the ancient maya social-ecological system. *Journal of Artificial Societies and Social Simulation*, 16(4):11, 2013. ISSN 1460-7425. URL `http://jasss.soc.surrey.ac.uk/16/4/11.html`.

D. Helbing. Stochastic methods and models of social interaction processes. Kluwer, Dordrecht, 1994.

D. Helbing. *Social Self-Organization, Understanding Complex Systems*, chapter Agent Based Modelling, pages 25–70. Springer-Verlag, Berlin Heidelberg, 2012. doi: 10.1007/978-3-642-24004-12.

Gudmund Hernes. The process of entry into first marriage. *American Sociological Review*, pages 173–182, 1972.

A. Hinde. *Demographic methods.* Arnold, London, UK, 1998.

Deborah Hix and H Rex Hartson. *Developing user interfaces: ensuring usability through product & process.* John Wiley & Sons, Inc., 1993.

D.W. Hosmer, S. Lemeshow, and S. May. *Applied survival analysis: Regression modeling of time-to-event data.* John Wiley & Sons, Hoboken, New Jersey, 2n edition, 2008.

Neil Howe, Richard Jackson, and Keisuke Nakashima. The aging of korea. demographics and retirement policy in the land of the morning calm. Technical report, Global Aging Initiative, Center for Strategic and International Studies, April 2007.

M.N. Huhns and M.P. Singh. *Readings in agents.* Morgan Kaufmann Publishers Inc., San Francisco, 1997.

M. Hwang, M. Garbey, Scott A. Berceli, and R. Tran-Son-Tay. Rule-Based Simulation of Multi-Cellular Biological Systems - A Review of Modeling Techniques. *Cellullar and Molecular Bioengineering*, 2:285–295, 2009.

M. Hybinette, E. Kraemer, Y. Xiong, G. Matthews, and J. Ahmed. Sassy: A design for a scalable agent- based simulation system using a distributed discrete event infrastructure. In *Proceedings of the 36th Conference on Winter Simulation (Monterey, California)*, pages 926–933, 2006.

Robert Itami, Rob Raulings, Glen MacLaren, Kathleen Hirst, Randy Gimblett, Dino Zanon, and Peter Chladek. Rbsim 2: simulating the complex interactions between human movement and the outdoor recreation environment. *Journal for Nature Conservation*, 11(4):278 – 286, 2003. ISSN 1617-1381. doi: 10.1078/1617-1381-00059.

Myeong-Wuk Jang and Gul Agha. Adaptive agent allocation for massively multi-agent applications. In T. Ishida, L. Gasser, and H. Nakashima, editors, *Massively*

Multi-Agent Systems I: First International Workshop MMAS 2004, pages 25–39. Springer, Berlin, Kyoto, Japan, December 2005.

Myeong-Wuk Jang and Gul Agha. Agent framework services to reduce agent communication overhead in large-scale agent-based simulations. *Simulation Modelling Practice and Theory*, 14(6):679–694, 2006.

Marco A. Janssen. Understanding artificial anasazi. *Journal of Artificial Societies and Social Simulation*, 12(4):13, 2009. ISSN 1460-7425. URL `http://jasss.soc.surrey.ac.uk/11/2/6.html`.

Marco A. Janssen, Lilian Na'ia Alessa, Michael Barton, Sean Bergin, and Allen Lee. Towards a community framework for agent-based modelling. *Journal of Artificial Societies and Social Simulation*, 11(2):6, 2008. ISSN 1460-7425. URL `http://jasss.soc.surrey.ac.uk/11/2/6.html`.

D.A. Jefferson. Virtual Time. *ACM Transaction on Programming Language System*, 7(3):404–425., 1985.

A. Kaplan Marcusán. De Senegambia a Cataluña. Procesos de aculturación e integración social. *Fundación La Caixa, Barcelona, Spain*, 1998.

A. Kaplan Marcusán. From Senegambia to Spain: Migration Process and Social Integration. *Monograph Migration: European Journal of International Migration and Ethnic Relations*, (43/44/45):52–65, 2005. Berlin, EMZ.

A. Kaplan Marcusán. Las migraciones Senegambianas en España: una mirada desde la perspectiva de género. *Ferran Iniesta (Coord.) Africa en diáspora: movimientos de población y políticas estatales*, 153–168, 2007.

RC Kennedy, X. Xiang, GR Madey, and TF Cosimano. Verification and validation of scientific and economic models. In M. North, D.L. Sallach, and C. Macal, editors, *Proceedings of Agent 2005: Generative Social Processes, Models, and Mechanisms*, pages 177–192, Chicago, IL: Argonne National Laboratory, 2006.

W.G. Kennedy, A.B. Hailegiorgis, M. Rouleau, J.K. Bassett, M. Coletti, G.C. Balan, and T. Gulden. An agent-based model of conflict in east africa and the effect of watering holes. In *Proceedings of the 19th Conference on Behavior Representation in Modeling and Simulation (BRiMS). Columbia, South Carolina.*, 2010. doi: 10.1145/1363686.1363696.

Franziska Klügl. A Validation Methodology for Agent-Based Simulations. In *Proceedings of the 23rd Annual ACM Sysmposium on Applied Computing*, pages 39–43, 1515 Broadway, NY 10036-9998 USA, 2008. ACM SIGAC; Univ Fortaleza; Federal Univ Ceara, ACM. 23rd Annual ACM Symposium on Applied Computing, Fortaleza, BRAZIL, MAR 16-20, 2008.

Dominic Kniveton, Christopher Smith, and Sharon Wood. Agent-based model simulations of future changes in migration flows for burkina faso. *Global Environmental Change*, 21, Supplement 1(0):S34 – S40, 2011. ISSN 0959-3780. doi: 10.1016/j.gloenvcha.2011.09.006. Migration and Global Environmental Change - Review of Drivers of Migration.

T.A. Kohler and S.E. van der Leeuw. *The model-based archaeology of socionatural systems*. SAR Press, Santa Fe, New Mexico, 2007.

A. Kokalj. Computer graphics and graphical user interfaces as tools in simulations of matter at the attomic scale. *Computational Material Science*, 28(2):155–168, 2003.

doi: 10.1016/s0927-0256(03)00104-6.

R. Koper. Increasing learner retention in a simulated learning network using indirect social interaction. *Journal of Artificial Societies and Social Simulation*, 8(2), 2005.

Korea Institute for Health and Social Affairs. The First Five-Year Plan for Ageing Society and Population (2006-2010). Technical report, Ministry of Health and Welfare, August 2006.

W. Kreutzer. *System Simulation: Programming styles and languages*, volume 3. Addison-Wesley, Addison-Wesley, Sydney, 1986.

Jasna Kuljis. Hci and simulation packages. In *Proceedings of the 28th conference on Winter simulation*, pages 687–694. IEEE Computer Society, 1996.

Jasna Kuljis and Ray J. Paul. Web-based discrete event simulation models: Current states and possible futures. *Simulation & Gaming*, 34(1):39–53, 2003. doi: 10.1177/1046878102250484.

Krzystof Kurowski, Walter de Back, Werner Dubitzky, Laszlo Gulyás, George Kampis, Mariusz Mamonski, Gabor Szemes, and Martin Swain. Complex system simulations with qoscosgrid. In *Computational Science–ICCS 2009*, pages 387–396. Springer, 2009.

C. Lan and M. Pidd. High performance simulation in quasi-continuous manufacturing plants. In M. E. Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, editors, *Proceedings of the 2005 Winter Simulation Conference*, pages 1367–1372, Picataway, NJ., 2005. IEEE Computer Society Press.

Averill M. Law. *Simulation modeling and analysis*. Mcgraw-Hill series in industrial engineering and management science, 4 edition, 2007.

R. Leombruni and M. Richiardi. Laborsim: An agent-based microsimulation of labour supply–an application to italy. *Computational Economics*, 27(1):63–88, 2006.

N. Li, Wenqing Yi, M. Sun, and G. Gong. Development and application of intelligent system modeling and simulation platform. *Simulation Modelling Practice and Theory*, 29:149–162, 2012. doi: 10.1016/j.simpat.2012.08.001.

B. Liddle. Demographic dynamics and sustainability: insights from an integrated multi-country simulation model. Working Paper WP 2002-039, Max Planck Institute for Demographic Research, Rostock, Germany, 2002.

F. Lin and Y.H. Pai. Using multi-agent simulation and learning to design new business processes. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 30(3):380–384, 2000. doi: 10.1109/3468.844361.

A. Lingnau and O. Drobnik. Simulating mobile agent systems with swarm. In *Agent Systems and Applications, 1999 and Third International Symposium on Mobile Agents. Proceedings. First International Symposium on*, pages 272–273. IEEE, 1999. doi: 10.1109/ASAMA.1999.805417.

C.J. Lobb, Z. Chao, R.M. Fujimoto, and S.M. Potter. Parallel event-driven neural network simulations using the Hodgkin-Huxler neuron model. In *Proceedings of the 19th Workshop on Principles of Advanced and Distributed SImulation*, pages 16–25, New York, NY, 2005. ACM Press.

C. Lozares Colina. La simulación social, una nueva manera de investigar en ciencia social ? *Papers: revista de sociologia*, (72):165–188, 2004.

S. Luke and V.A. Ziparo. Learn to Behave! Rapid Training of Behavior Automata. In

Proceedings of Adaptive and Learning Agents Workshop at AAMAS, pages 61–68. Citeseer, 2010.

S. Luke, C. Cioffi-Revilla, L. Panait, K. Sullivan, and G. Balan. Mason: A multiagent simulation environment. *Simulation*, 81(7):517, 2005.

F. Luna and B. Stefansson. *Economic Simulations in Swarm: Agent-based modelling and object oriented programming*, volume 14. Springer, 2000. doi: 10.1080/17517570701275390.

M. Lysenko and R.M. D'Souza. A framework for megascale agent based model simulations on graphics processing units. *Journal of Artificial Societies and Social Simulation*, 11(4):10, 2008.

Charles M. Macal and Michael J. North. Tutorial on agent-based modeling and simulation. In *Proceedings of the 37th conference on Winter simulation*, WSC '05, pages 2–15. Winter Simulation Conference, 2005. ISBN 0-7803-9519-0.

Charles M. Macal and Michael J. North. Agent-based modeling and simulation: desktop abms. In *Proceedings of the 39th conference on Winter simulation: 40 years! The best is yet to come*, WSC '07, pages 95–106, Piscataway, NJ, USA, 2007. IEEE Press. ISBN 1-4244-1306-0.

G. Madey, V. Freeh, and R. Tynan. Agent-Based Modeling of Open Source using Swarm. In *Americas Conference on Information Systems (AMCIS2002)*, pages 1472–1475, Dallas, TX, USA, 2002.

P. Madhavan, Y. Papelis, R. Kady, and L. Moya. An agent-based model of crowd cognition. In *Proceedings of the 18th Conference on Behavior Representation in Modeling and Simulation*, pages 139–140, 2009.

Federico Massaioli, Filippo Castiglione, and Massimo Bernaschi. Openmp parallelization of agent-based models. *Parallel Computing*, 31(10):1066–1081, 2005.

D.H. Meadows, D.L. Meadows, J. Randers, and W.B. William. *Limits to growth.* Universe Books, New York, NY, USA, 1972.

D.L. Meadows, D.H. Meadows, and J. Randers. *The limits to growth: The 30-year update.* Earthscan, London, UK, 2004.

M.A. Menéndez and S.H. Collado. Simulación de procesos sociales basada en agentes software. *EMPIRIA. Revista de Metodología de Ciencias Sociales*, 14:139–161, 2007.

R.S. Michalski, J.G. Carbonell, and T.M. Mitchell. *Machine learning: An artificial intelligence approach*, volume 1. Morgan Kaufmann, 1985. doi: 10.1002/zamm.19850651113.

James Millington, Raúl Romero-Calcerrada, John Wainwright, and George Perry. An agent-based model of mediterranean agricultural land-use/cover change for examining wildfire risk. *Journal of Artificial Societies and Social Simulation*, 11(4):4, 2008. ISSN 1460-7425. URL `http://jasss.soc.surrey.ac.uk/11/4/4.html`.

N. Minar, R. Burkhart, C. Langton, and M. Askenazi. The swarm simulation system: A toolkit for building multi-agent simulations. Working Papers 96-06-042, Santa Fe Institute, 1996.

Steven Mithen and Melissa Reed. Stepping out: a computer simulation of hominid dispersal from africa. *Journal of Human Evolution*, 43(4):433 – 462, 2002. ISSN 0047-2484. doi: 10.1006/jhev.2002.0584.

Erik Mobrand. Reverse Remittances: Internal Migration and Rural-to Urban Remittances in Industrialising South Korea. *Journal of Ethnic and Migration Studies*, 3 (38):389–411, 2012. doi: 10.1080/1369183X.2012.658544.

C. Montañola Sales, J. Casanovas-Garcia, B.S.S. Onggo, J. M. Cela-Espín, and A. Kaplan-Marcusán. Approaching simulation to modelers: A user interface for large-scale demographic simulation. In A. Tolk, S. D. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, editors, *Proceedings of the 2014 Winter Simulation Conference*. ACM, December 2014a.

Cristina Montañola Sales, Josep Casanovas-Garcia, Jose Maria Cela-Espín, B.S.S. Onggo, and Adriana Kaplan-Marcusan. Parallel simulation of large population dynamics. In *ACM SIGSIM Conference on Principles of Advanced Discrete Simulation (PADS)*, Montreal, Canada, May 2013a. ACM SIGSIM.

Cristina Montañola Sales, Josep Casanovas-Garcia, Jose Maria Cela-Espín, B.S.S. Onggo, and Adriana Kaplan-Marcusan. Parallel simulation of large population dynamics. In *Winter Simulation Conference, 8-12 December 2013, Washington D.C.*, Washington D.C., December 2013b. Informs.

Cristina Montañola Sales, Josep Casanovas-Garcia, Jose Maria Cela-Espín, B.S.S. Onggo, and Adriana Kaplan-Marcusan. Agent-based simulation of large population dynamics. In *European Conference on Complex Systems, 16-20 September 2013, Barcelona, Spain*, Barcelona, Spain, September 2013c.

Cristina Montañola Sales, Josep Casanovas-Garcia, Adriana Kaplan-Marcusán, and Jose M. Cela-Espín. Demographic agent-based simulation of Gambians immigrants in Spain. In Francisco J. Miguel, Frederic Amblard, Joan A. Barcelo, and M. Madella, editors, *Advances in Computational Social Science and Social Simulation, Proceedings of the Social Simulation Conference, Barcelona, 1-5th Sept. 2014*, Barcelona, Spain, 2014b. ESSA. doi: 10.13140/2.1.3052.4166. URL `http://ddd.uab.cat/record/125597`.

Cristina Montañola Sales, Josep Casanovas-Garcia, Bhakti S.S. Onggo, and Zengxiang Li. A user interface for large-scale demographic simulation. In Juan Guerrero, editor, *Procceding of the IEEE 6th International Conference on Cloud Computing Technology and Science (CloudCom 2014)*. IEEE, December 2014c.

Cristina Montañola Sales, Xavier Rubio-Campillo, Josep Casanovas-Garcia, Jose M. Cela-Espín, and Adriana Kaplan-Marcusán. Large-scale social simulation, dealing with complexity challenges in high performance environments. In Diana Francisca Adamatti, Graçaliz Dimuro, and Helder Coelho, editors, *Interdisciplinary Applications of Agent-Based Social Simulation and Modeling*, chapter 7, pages 106–123. IGI Global, May 2014d. doi: 10.4018/978-1-4666-5954-4.ch007.

Cristina Montañola Sales, Xavier Rubio-Campillo, Josep Casanovas-Garcia, Jose Maria Cela-Espín, and Adriana Kaplan-Marcusan. Overview on Agent-Based Social Modelling and the use of formal languages. In Pau Fonseca, editor, *Formal Languages for Computer Simulation: Transdisciplinary Models and Applications*, pages 333–377. IGI Global, Hershey PA, US, 2014e. doi: 10.4018/978-1-4666-4369-7.ch011. ISBN 978-1-4666-4371-0.

Cristina Montañola Sales, Bhakti S.S. Onggo, Josep Casanovas-Garcia, Jose M. Cela-Espin, and Adriana Kaplan-Marcusán. Yades: A parallel simulation tool for demography. *Parallel Computing*, 2015. (submitted).

Cristina Montañola-Sales, Bhakti S.S. Onggo, and Josep Casanovas-Garcia. Agent-based simulation validation: A case study in demographic simulation. In *SIMUL 2011, The Third International Conference on Advances in System Simulation*, pages 101–107, 2011. ISBN 978-1-61208-169-4.

S. Moss and B. Edmonds. Sociology and simulation: Statistical and qualitative cross-validation. Technical Report 03105, 2003.

Scott Moss. Messy Systems - The Target for Multi Agent Based Simulation. In Scott Moss and Paul Davidsson, editors, *Multi-Agent-Based Simulation*, volume 1979 of *Lecture Notes in Computer Science*, pages 1–14. Springer Berlin / Heidelberg, 2001.

Scott Moss, Bruce Edmonds, and Steve Wallis. Validation and Verification of Computational Models with Multiple Cognitive Agents. Discussion Papers 97-25, Manchester Metropolitan University, Centre for Policy Modelling, July 1997.

S.P Murphy and T Perera. Successes and failures in uk/us development of simulation. *Simulation Practice and Theory*, 9(6–8):333 – 348, 2002. ISSN 0928-4869. doi: 10.1016/S0928-4869(01)00048-9.

Stephen P Murphy and Terrence D Perera. Simulation practice: key enablers in the development of simulation. In *Proceedings of the 33nd conference on Winter simulation*, pages 1429–1437. IEEE Computer Society, 2001.

San Murugesan, Yogesh Deshpande, Steve Hansen, and Athula Ginige. Web engineering: A new discipline for development of web-based systems. *Lecture notes in computer science*, pages 3–13, 2001.

B. Myers, S.E. Hudson, and R. Pausch. Past, present, and future of user interface software tools. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 7 (1):3–28, 2000. ISSN 1073-0516.

Cynthia Nikolai and Gregory Madey. Tools of the trade: A survey of various agent based modeling platforms. *Journal of Artificial Societies and Social Simulation*, 12(2):2, 2009. ISSN 1460-7425. URL `http://jasss.soc.surrey.ac.uk/12/2/2.html`.

Cynthia Nikolai and Gregory Madey. Comparison of agent-based modelling software, August 2011. URL `http://en.wikipedia.org/wiki/Comparison_of_agent-based_modeling_software`.

M.J. North and C.M. Macal. The beer dock: Three and a half implementations of the beer distribution game. In *SwarmFest 2002*, 2002.

M.J. North and C.M. Macal. *Artificial life models in software*, chapter Escaping the accidents of history: an overview of artificial life modeling with Repast, pages 115–141. Springer, Heidelberg, Germany, 2005. doi: 10.1007/1-84628-214-4.

M.J. North, N.T. Collier, and J.R. Vos. Experiences creating three implementations of the repast agent modeling toolkit. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 16(1):1–25, 2006. doi: 10.1145/1122012.1122013.

C. O'Donoghue and H. Redway. Modelling migration in Pensim2: a dynamic microsimulation model of the British pension system. In *Presented at the 2nd General Conference of the International Microsimulation Association, June 8th to 10th, 2009, Ottawa, Ontario, Canada*, 2009.

OECD. Pisa 2012 results: What students know and can do - student performance in mathematics, reading and science (volume i, revised edition, february 2014). Technical report, Programme for International Student Assessment. OECD, February 2014.

United Nations. Dept. of Economic. *Replacement migration: is it a solution to declining and ageing populations?*, volume 206. United Nations Publications, 2001.

B.B.S. Onggo. *Conceptual Modelling for Discrete-Event Simulation*, chapter Methods for Conceptual Model Representation, pages 337–354. Taylor & Francis, Abingdon, 2010. doi: 10.1201/9781439810385-c13.

B.S.S. Onggo. Parallel discrete-event simulation of population dynamics. In *Proceedings of the 40th Conference on Winter Simulation*, pages 1047–1054. Winter Simulation Conference, 2008.

B.S.S. Onggo, C. Montañola Sales, and J. Casanovas-Garcia. Performance Analysis of Parallel Demographic Simulation. In *Proceedings of the 24th European Simulation and Modelling Conference, 25-27 October 2010, Hasselt, Belgium (ESM10)*, pages pp. 142–148, Ostend, Belgium, 2010. Eurosis-ETI.

G.H. Orcutt. A new type of socio-economic system. *The Review of Economics and Statistics*, 39(2):116–123, 1957.

Tim O'Reilly. What is web 2.0: Design patterns and business models for the next generation of software. *Communications & strategies*, (1):17, 2007.

Paul Ormerod and Bridget Rosewell. Validation and Verification of Agent-Based Models in the Social Sciences. In Squazzoni, F, editor, *Epistemilogical Aspects of Comuter Simulation in the Social Sciences, Second International Workshop, EPOS 2006*, volume 5466 of *Lecture Notes in Artificial Intelligence*, pages 130–140, Heidelberger Platz 3, D-14197 Berlin, Germany, 2009. Lucchini Fdn; European Social Simulat Assoc, Springer-Verlag. ISBN 978-3-642-01108-5. doi: 10.1007/978-3-642-01109-2_10. 2nd International Workshop on Epistemological Perspectives on Simulation, Brescia, Italy, oct 05-06, 2006.

John D Owens, David Luebke, Naga Govindaraju, Mark Harris, Jens Krüger, Aaron E Lefohn, and Timothy J Purcell. A survey of general-purpose computation on graphics hardware. In *Computer graphics forum*, volume 26, pages 80–113. Wiley Online Library, 2007.

Peter S Pacheco. *Parallel programming with MPI.* Morgan Kaufmann Publishers Inc., San Francisco, 1997.

Jon Parker. A flexible, large-scale, distributed agent based epidemic model. In Shane G. Henderson, Bahar Biller, Ming-Hua Hsieh, John Shortle, Jeffrey D. Tew, and Russell R. Barton, editors, *Proceedings of the 39th conference on Winter simulation, Washington, DC, USA, December 9-12, 2007*, WSC '07, pages 1543–1547, Piscataway, NJ, USA, December 9-12 2007. IEEE Press. ISBN 1-4244-1306-0. URL `http://portal.acm.org/citation.cfm?id=1351542.1351816`.

Jon Parker and Joshua M. Epstein. A distributed platform for global-scale agent-based models of disease transmission. *ACM Trans. Model. Comput. Simul.*, 22(1): 2:1–2:25, December 2011. ISSN 1049-3301. doi: 10.1145/2043635.2043637. URL `http://doi.acm.org/10.1145/2043635.2043637`.

HazelR. Parry. Agent based modeling, large scale simulations. In Robert A. Meyers, editor, *Computational Complexity*, pages 76–87. Springer New York, 2012. ISBN 978-1-4614-1799-6. doi: 10.1007/978-1-4614-1800-9_5. URL `http://dx.doi.org/10.1007/978-1-4614-1800-9_5`.

H Van Dyke Parunak, Robert Savit, and Rick L Riolo. Agent-based modeling vs. equation-based modeling: A case study and users' guide. In *Multi-Agent Systems and Agent-Based Simulation*, pages 10–25. Springer, 1998.

J. Pavon, M. Arroyo, S. Hassan, and C. Sansores. Agent-based modelling and simulation for the analysis of social patterns. *Pattern Recognition Letters*, 29(8):1039–1048, 2008. ISSN 0167-8655.

Nuria Pelechano, Jan M Allbeck, and Norman I Badler. Controlling individual agents in high-density crowd simulation. In *Proceedings of the 2007 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 99–108, Aire-la-Ville, Switzerland, 2007. Eurographics Association.

K. Perumalla. *Modeling Simulation and Optimization - Focus on Applications*, chapter Computational Spectrum of Agent Model Simulation, pages 185–204. InTech, 2010. ISBN 978-953-307-055-1.

K.S. Perumalla. μsik: A Micro-Kernel for Parallel/Distributed Simulation Systems. In *Proceedings of the 19th Workshop on Principles of Advanced and Distributed Simulation*, pages 59–68. IEEE Computer Society Washington, DC, USA, 2005.

K.S. Perumalla. Parallel and distributed simulation: traditional techniques and recent advances. In *Simulation Conference, 2006. WSC 06. Proceedings of the Winter*, pages 84–95. Simulation Conference, 2006. WSC 06., 2006.

K.S. Perumalla. Scaling time warp-based discrete-event execution to 104 processors on a blue gene supercomputer. In *Proceedings of the 4th International Conference on Computing Frontiers*, pages 69–76, New York, NY, 2007. ACM Press.

K.S. Perumalla and B.G. Aaby. Data parallel execution challenges and runtime performance of agent simulations on GPUs. In *Proceedings of the 2008 Spring simulation multiconference*, pages 116–123. Society for Computer Simulation International, 2008.

M. Pidd. *Tools for Thinking: Modelling in Management Science.* Wiley, Chichester, UK, 2 edition, 2003. doi: 10.1057/palgrave.jors.2600969.

M. Pidd. *Computer simulation in management science.* John Wiley & Sons, Inc. Chichester, England, 5th edition edition, 2004.

M. Pidd. Why modelling and model use matter. *Journal of the Operational Research Society*, 61(1):14–24, 2010. doi: 10.1057/jors.2009.141.

Vincent Pillet, Jesús Labarta, Toni Cortes, and Sergi Girona. Paraver: A tool to visualize and analyze parallel code. In *Proceedings of WoTUG-18: Transputer and occam Developments*, volume 44, pages 17–31, 1995. URL `http://www.bsc.es/paraver`.

M. Posada, C. Hernández, and A. López-Paredes. Emissions Permits Auctions: An Agent Based Model Analysis. In *Social Simulation: Technologies, Advances, and New Discoveries*, chapter 15, pages 43–60. IGI Global, Hershey, New York, 1 edition, August 2007.

Bruno R. Preiss. The Yaddes Distributed Event Simulation Specification Language and Execution Environments. In *Proceedings of the SCS Multiconference on Distributed Simulation*, volume 21, pages 139–144. Society for Computer Simulation, 1989.

R. Pressat. *Demographic analysis.* Aldine Atherton, Chicago, 1985.

Ignacio Prieto Sarro and Lorenzo López Trigal. Portugueses y caboverdianos en españa. *Estudios geográficos*, 54(210):75–96, 1993.

Radharamanan Radhakrishnan, Dale E Martin, Malolan Chetlur, Dhananjai Madhava Rao, and Philip A Wilsey. An object-oriented Time Warp simulation kernel. In *Proceedings of the International Symposium on Computing in Object-Oriented Parallel Environments, LNCS 1505*, pages 13–23. Springer, 1998.

D.M. Rao and A. Chernyakhovsky. Parallel simulation of the global epidemiology of avian influenza. In *Proceedings of the 22nd Workshop on Principles of Advanced and Distributed Simulation*, pages 1583–1591, Picataway, N.J., 2008. IEEE Publisher Press.

MS Rauner, SC Brailsford, and S. Flessa. Use of discrete-event simulation to evaluate strategies for the prevention of mother-to-child transmission of HIV in developing countries. *Journal of the Operational Research Society*, 56(2):222–233, 2005.

Dwight Read. Kinship based demographic simulation of societal processes. *Journal of Artificial Societies and Social Simulation*, 1(1), 1998. URL `http://jasss.soc.surrey.ac.uk/1/1/1.html`.

P. Rees. Demography. In Rob Kitchin and Nigel Thrift, editors, *International Encyclopedia of Human Geography*, pages 75 – 90. Elsevier, Oxford, 2009. ISBN 978-0-08-044910-4. doi: http://dx.doi.org/10.1016/B978-008044910-4.00815-4. URL `http://www.sciencedirect.com/science/article/pii/B9780080449104008154`.

M. Resnick. *Turtles, Termites and Traffic Jams: Explorations in Massively Parallel Microworlds.* MIT Press, 1994.

Romain Reuillon, Mathieu Leclaire, and Sebastien Rey. Openmole, a workflow engine specifically tailored for the distributed exploration of simulation models. *Future Generation Computer Systems*, 29(8):1981–1990, 2013.

Paul Richmond, Dawn Walker, Simon Coakley, and Daniela Romano. High performance cellular level agent-based simulation with flame for the gpu. *Briefings in Bioinformatics*, 11(3):334–347, 2010. doi: 10.1093/bib/bbp073. URL `http://bib.oxfordjournals.org/content/11/3/334.abstract`.

S. Robinson. Simulation model verification and validation: increasing the users' confidence. In *Proceedings of the 29th Conference on Winter Simulation*, pages 53–59. IEEE Computer Society, 1997. doi: 10.1145/268437.268448.

S. Robinson. Conceptual modelling for simulation Part I: definition and requirements. *Journal of the Operational Research Society*, 59(3):278–290, 2008. doi: 10.1057/palgrave.jors.2602368.

P. Roderick, R. Davies, C. Jones, T. Feest, S. Smith, and K. Farrington. Simulation model of renal replacement therapy: predicting future demand in england. *Nephrology Dialysis Transplantation*, 19(3):692, 2004.

Stuart Rossiter, Jason Noble, and Keith R.W. Bell. Social simulations: Improving interdisciplinary understanding of scientific positioning and validity. *Journal of Artificial Societies and Social Simulation*, 13(10):1, 2010. ISSN 1460-7425. doi: 10.1145/268437.268448. URL `http://jasss.soc.surrey.ac.uk/13/1/10.html`.

M. Rouleau, M. Coletti, J.K. Bassett, A.B. Hailegiorgis, T. Gulden, and W.G. Kennedy. Conflict in complex socio-natural systems: Using agent-based modeling to understand the behavioral roots of social unrest within the mandera triangle. In

Proceedings of the Human Behavior-Computational Modeling and Interoperability Conference, 2009.

Nick Roussopoulos and Dimitris Karagiannis. Conceptual modeling: Past, present and the continuum of the future. In Alexander Borgida, Vinay Chaudhri, Paolo Giorgini, and Eric Yu, editors, *Conceptual Modeling: Foundations and Applications*, volume 5600 of *Lecture Notes in Computer Science*, pages 139–152. Springer Berlin / Heidelberg, 2009. ISBN 978-3-642-02462-7. doi: 10.1007/978-3-642-02463-4_9.

Donald T Rowland. *Demographic methods and concepts*. Oxford University Press Inc., New York, 2003. ISBN 0-19-875263-6.

X. Rubio. *Modelització i simulació aplicades a la recerca i interpretació de camps de batalla*. PhD thesis, Universitat de Barcelona. Departament de Didáctica de les Ciéncies Socials, Barcelona, Spain, June 2009. URL `http://hdl.handle.net/10803/1339`.

X. Rubio and J.M. Cela. Large-scale agent-based simulation in archaeology: an approach using High-performance computing. In *Computer Applications in Archaeology*, Granada, Spain, April 2010.

X. Rubio-Campillo. Pandora: An hpc agent-based modelling framework. 2013. URL `https://github.com/xrubio/pandora/`.

Xavier Rubio-Campillo, Jose María Cela, and Francesc Xavier Hernandez-Cardona. Simulating archaeologists? Using agent-based modelling to improve battlefield excavations. *Journal of Archaeological Science*, 39(2):347–356, 2012. ISSN 0305-4403. doi: 10.1016/j.jas.2011.09.020. URL `http://www.sciencedirect.com/science/article/pii/S0305440311003475`.

S.J. Russell and P. Norvig. *Artificial intelligence: a modern approach*. Prentice hall, 2010. doi: 10.1016/0004-3702(96)00007-0.

Mauricio Salgado. The evolution of paternal care. In ArielM. Greenberg, WilliamG. Kennedy, and NathanD. Bos, editors, *Social Computing, Behavioral-Cultural Modeling and Prediction*, volume 7812 of *Lecture Notes in Computer Science*, pages 1–10. Springer Berlin Heidelberg, 2013. ISBN 978-3-642-37209-4. doi: 10.1007/978-3-642-37210-0_1. URL `http://dx.doi.org/10.1007/978-3-642-37210-0_1`.

R.G. Sargent. Verification and validation of simulation models. In *Proceedings of the 37th Winter Simulation Conference*, pages 130–143. IEEE, Inc., 2005. doi: 10.1109/WSC.2005.1574246.

J. Saw and M. Butler. Exploring graphical user interfaces and interaction strategies in simulations. In *Hello! Where are you in the landscape of educational technology? Proceedings ascilite Melbourne 2008 http://www.ascilite.org.au/conferences/melbourne08/procs/saw.pdf*, 2008. URL `http://www.ascilite.org.au/conferences/melbourne08/procs/saw.pdf`.

A.K. Saysel, Y. Barlas, and O. Yenigun. Environmental sustainability in an agricultural development project: a system dynamics approach. *Journal of Environmental Management*, 64(3):247–260, 2002.

T.C. Schelling. Dynamic models of segregation. *Journal of mathematical sociology*, 1 (2):143–186, 1971. doi: 10.1080/0022250X.1971.9989794.

Andreas Schlosser, Marco Voss, and Lars Brückner. On the Simulation of Global Reputation Systems. *Journal of Artificial Societies and Social Simulation*, 9(1):4,

2005. ISSN 1460-7425. URL http://jasss.soc.surrey.ac.uk/9/1/4.html.

Edmundo Sepa Bonaba. Els negres catalans. *Barcelona: Fundició Serveis de Cultura Popular, Editorial Alta Fulla*, 1993.

R.E. Shannon. Simulation modeling and methodology. In *Proceedings of the 76 Bicentennial conference on Winter simulation*, pages 9–15. Winter Simulation Conference, 1976. URL http://informs-sim.org/wsc76papers/prog76sim.html.

Gi-Wook Shin. *Ethnic nationalism in Korea: Genealogy, politics, and legacy*. Stanford University Press, 2006.

Ben Shneiderman and Catherine Plaisant. *Designing The User Interface: Strategies for Effective Human-Computer Interaction*. Addison-Wesley Publishing Company, 5th edition, 2010.

PO Siebers, CM Macal, J. Garnett, D. Buxton, and M. Pidd. Discrete-event simulation is dead, long live agent-based simulation! *Journal of Simulation*, 4(3):204–210, 2010.

Eric Silverman, Jakub Bijak, Jason Hilton, Viet Dung Cao, and Jason Noble. When demography met social simulation: A tale of two modelling approaches. *Journal of Artificial Societies and Social Simulation*, 16(4):9, 2013. ISSN 1460-7425. URL http://jasss.soc.surrey.ac.uk/16/4/9.html.

A. Smajgl and E. Bohensky. Behaviour and space in agent-based modelling: Poverty patterns in east kalimantan, indonesia. *Environmental Modelling & Software*, 45(0):8 – 14, 2013. ISSN 1364-8152. doi: http://dx.doi.org/10.1016/j.envsoft.2011.10.014. URL http://www.sciencedirect.com/science/article/pii/S1364815211002258. <ce:title>Thematic Issue on Spatial Agent-Based Models for Socio-Ecological Systems</ce:title>.

Andy Smith. *Human computer factors*. McGraw-Hill, 1997.

Korea South. Korea south: Country study guide. volume 1. strategic information and developments. Technical report, International Business Publications, USA, Washington DC, USA, 2013.

Flaminio Squazzoni. The impact of agent-based models in the social sciences after 15 years of incursions. *History of Economic Ideas*, XVIII(2):197–233, 2010.

Jeff S Steinman. SPEEDES: A unified approach to parallel simulation. In *Proceedings of the 6th Workshop on Parallel and Distributed Simulation*, volume 24, pages 75–84, 1992.

Jeffrey S Steinman. The WarpIV simulation kernel. In *Proceedings of the 2005 Workshop on Principles of Advanced and Distributed Simulation, PADS*, pages 161–170. IEEE, 2005.

J.D. Sterman. *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Mcgraw-Hill Higher Education, 2000. doi: 10.1036/007238915X.

I. Stewart. *Does god play dice? The new mathematics of caos*. Blackwell Publishing, 1989.

D.L. Stone and D. Stone. *User interface design and evaluation*. Morgan Kaufmann, 2005. ISBN 0120884364.

T.J. Strader, F.R. Lin, and M.J. Shaw. Simulation of order fulfillment in divergent assembly supply chains. *Journal of Artificial Societies and Social Simulation*, 1(2): 36–37, 1998. URL http://jasss.soc.surrey.ac.uk/1/2/5.html.

Patrick Taillandier, Duc-An Vo, Edouard Amouroux, and Alexis Drogoul. Gama : a simulation platform that integrates geographical information data, agent-based modeling and multi-scale control. In Nirmit Desai, Alan Liu, and Michael Winikoff, editors, *Principles and Practice of Multi-Agent Systems 13th International Conference, PRIMA 2010,Kolkata, India, November 12-15, 2010, Revised Selected Papers*, volume 7057 of *Lecture Notes in Computer Science*, pages 242–258. Springer, 2012. doi: 10.1007/978-3-642-25920-3.

Y. Tang, K.S. Perumalla, R.M. Fujimoto, H. Karimabadi, J. Driscoll, and Y. Omelchenko. Optimistic parallel discrete event simulations of physical systems using reverse computation. In *Proceedings of the 19th Workshop on Principles of Advanced and Distributed Simulation*, pages 26–35, New York, NY, 2005. ACM Press.

L. Tesfatsion. Agent-based computational economics: Growing economies from the bottom up. *Artificial life*, 8(1):55–82, 2002. doi: 10.1162/106454602753694765.

L. Tesfatsion. *Handbook of Computational Economics: Agent-based Computational Economics*, chapter Agent-based Computational Economics: A Constructive Approach to Economic Theory, pages 831–880. North-Holland, Amsterdam, The Netherlands, 2006. URL `http://www.econ.iastate.edu/research/books-and-chapters/p7004`.

Ingo J. Timm and D. Pawlaszczyk. Large Scale Multiagent Simulation on the Grid. In *Proceedings of the Workshop on Agent-Based Grid Economics (AGE 2005) at the IEEE International Symposium on Cluster Computing and the Grid (CCGRID)*, pages 334–341, Cardiff, 2005. doi: 10.1.1.90.2116.

S. Tisue and U. Wilensky. NetLogo: A simple environment for modeling complexity. In *Proceedings of the International conference on Complex Systems, Boston, May 2004*, pages 16–21. Citeseer, 2004.

R. Tobias and C. Hofmann. Evaluation of free java-libraries for social-scientific agent based simulation. *Journal of Artificial Societies and Social Simulation*, 7(1), 2004. URL `http://jasss.soc.surrey.ac.uk/7/1/6.html`.

Adrien Treuille, Seth Cooper, and Zoran Popović. Continuum crowds. *ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH 2006*, 25(3):1160–1168, 2006.

K.G. Troitzsch. *Simulating Social Phenomena*, volume 456, chapter Social science simulation - origins, prospects, purposes, pages 41–54. Springer- Verlag, Berlin, Germany, 1997.

Klaus G Troitzsch. Perspectives and Challenges of Agent-Based Simulation as a Tool for Economics and Other Social Sciences. In Decker, Sichman, Sierra, and Castelfranchi, editors, *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2009)*, pages 35–42, Budapest, Hungary, May, 10-15 2009. International Foundation for Autonomous Agents and Multiagent Systems.

S.E. van der Leeuw. Why model? *Cybernetics and Systems*, 35(2-3):117–128, 2004. doi: 10.1080/01969720490426803.

Evert Van Imhoff and Wendy Post. Microsimulation methods for population projection. *Population*, 10:97–138, 1998.

TD Veselka, G. Boyd, G. Conzelmann, V. Koritarov, C.M. Macal, M.J. North,

B. Schoepfle, and P Thimmapuram. Simulating the behavior of electricity markets with an agent-based methodology: the electric market complex adaptive systems (EMCAS) model. In *The 22nd International Association for Energy Economics International Conference*, 2002.

Richard Vidgen and Julian Padget. Sendero: An Extended, Agent-Based Implementation of Kauffmanś NKCS Model. *Journal of Artificial Societies and Social Simulation*, 12(4):8, 2009. ISSN 1460-7425. URL `http://jasss.soc.surrey.ac.uk/12/4/8.html`.

Guillermo Vigueras, Miguel Lozano, Carlos Perez, and JM Ordua. A scalable architecture for crowd simulation: Implementing a parallel action server. In *Proceedings of the 37th International Conference on Parallel Processing (ICPP 2008=*, pages 430–437. IEEE, 2008. doi: 10.1109/ICPP.2008.20.

Guillermo Vigueras, JuanM. Orduńa, and Miguel Lozano. A gpu-based multi-agent system for real-time simulations. In Yves Demazeau, Frank Dignum, JuanM. Corchado, and JavierBajo Pérez, editors, *Advances in Practical Applications of Agents and Multiagent Systems*, volume 70 of *Advances in Intelligent and Soft Computing*, pages 15–24. Springer Berlin Heidelberg, 2010. ISBN 978-3-642-12383-2. doi: 10.1007/978-3-642-12384-9_3.

D. Villatoro and J. Sabater-Mir. Mechanisms for social norms support in virtual societies. In *Proceedings of the 5th conference of the European Social Simulation Association (ESSA08), Brescia, Italy*, 2008.

Nikos Viorres, Papadopoulos Xenofon, Modestos Stavrakis, Evangelos Vlachogiannis, Panayiotis Koutsabasis, and John Darzentas. Major hci challenges for open source software adoption and development. In *Online Communities and Social Computing*, pages 455–464. Springer, 2007.

J. Von Neumann. *Theory of self-reproducing automata*. University of Illinois Press, Urbana, IL., 1966. doi: 10.1002/asi.5090180413.

A. Walker, R. Percival, and A. Harding. *Microsimulation modelling for policy analysis: challenges and innovations*, chapter The impact of demographic and other changes on expenditure on pharmaceutical benefits in 2020 in Australia, pages 149–171. Cambridge University Press, Cambridge, UK, 2000.

Fang Wang, StephenJohn Turner, and Lihua Wang. Agent communication in distributed simulations. In Paul Davidsson, Brian Logan, and Keiki Takadama, editors, *Multi-Agent and Multi-Agent-Based Simulation*, volume 3415 of *Lecture Notes in Computer Science*, pages 11–24. Springer Berlin Heidelberg, 2005. ISBN 978-3-540-25262-7. doi: 10.1007/978-3-540-32243-6_2. URL `http://dx.doi.org/10.1007/978-3-540-32243-6_2`.

G. Weiss. *Multiagent systems: a modern approach to distributed artificial intelligence*. The MIT press, 1999.

Danny Weyns, Alexander Helleboogh, and Tom Holvoet. How to Get Multi-Agent Systems Accepted in Industry ? *Int. J. Agent-Oriented Softw. Eng.*, 3(4):383–390, 2009. ISSN 1746-1375. doi: 10.1504/IJAOSE.2009.025316. URL `http://dx.doi.org/10.1504/IJAOSE.2009.025316`.

Adam Wierzbicki and Radoslaw Nielek. Fairness emergence in reputation systems. *Journal of Artificial Societies and Social Simulation*, 14(1):3, 2011. ISSN 1460-7425. URL `http://jasss.soc.surrey.ac.uk/14/1/3.html`.

Meg Wiggins, Chris Bonell, Mary Sawtell, Helen Austerberry, Helen Burchett, Elizabeth Allen, and Vicki Strange. Health outcomes of youth development programme in england: prospective matched comparison study. *British Medical Journal*, 339: b2534, 7 2009. doi: 10.1136/bmj.b2534.

U. Wilensky. Modeling nature's emergent patterns with multi-agent languages. In *Proceedings of EuroLogo 2001 Linz, Austria*, pages 1–6, 2002.

U. Wilensky and W. Rand. Making models match: Replicating an agent-based model. *Journal of Artificial Societies and Social Simulation*, 10(4):2, 2007.

F. Willekens. Biographic forecasting: Bridging the micro-macro gap in population forecasting. *New Zealand Population Review*, 31(1):77–124, 2005.

John R. Wilmoth. New Director highlights importance of population dynamics, January 2013. URL `http://www.un.org/en/development/desa/news/population/population-division-director.html`.

Paul Windrum, Giorgio Fagiolo, and Alessio Moneta. Empirical Validation of Agent-Based Models: Alternatives and Prospects. *Journal of Artificial Societies and Social Simulation*, 10(2):8, 2007. ISSN 1460-7425. URL `http://jasss.soc.surrey.ac.uk/10/2/8.html`.

P. Wittek and X. Rubio-Campillo. Social simulations accelerated: Large-scale agent-based modeling on a gpu cluster. In *GPU Technology Conference*, San Diego, USA, 2013.

Belinda M Wu and Mark H Birkin. Agent-based extensions to a spatial microsimulation model of demographic change. In *Agent-based models of geographical systems*, pages 347–360. Springer, 2012.

J. Xu, Y. Gao, and G. Madey. A docking experiment: Swarm and repast for social network modeling. In *Seventh Annual Swarm Researchers Meeting (Swarm2003)*. Citeseer, 2003.

L. Yang and N. Gilbert. Getting away from numbers: Using qualitative observation for agent-based modeling. *Advances in Complex Systems*, 11(2):175–186, 2008. ISSN 0219-5259.

S.B. Yoginath and K.S. Perumalla. Parallel Vehicular Traffic Simulation using Reverse Computation-based Optimistic Execution. In *Proceedings of the 22nd Workshop on Principles of Advanced and Distributed Simulation*, pages 145–152, Piscataway, NJ, 2008. IEEE Computer Society Press.

Yi Zhang, Georgios Theodoropoulos, Rob Minson, Stephen Turner, Wentong Cai, Yong Xie, and Brian Logan. Grid-Aware Large Scale Distributed Simulation of Agent-Based Systems. In *Proceedings of the 2005 European Simulation Interoperability Workshop*, 2005.

H. Zhao and X. Li. H-Trust: A robust and lightweight group reputation system for peer-to-peer desktop grid. In *The 28th International Conference on Distributed Computing Systems Workshops*, pages 235–240. IEEE, 2008. doi: 10.1109/ICDCS.Workshops.2008.96.

Jiayun Zhao, Esfandyar Mazhari, Nurcin Celik, and Young-Jun Son. Hybrid agent-based simulation for policy evaluation of solar power generation systems. *Simulation Modelling Practice and Theory*, 19(10):2189–2205, 2011. ISSN 1569-190X. doi: DOI:10.1016/j.simpat.2011.07.005. URL `http://www.sciencedirect.com/`

science/article/pii/S1569190X11001341.

S. Zinn, J. Gampe, J. Himmelspach, and AM Uhrmacher. MIC-core: A tool for microsimulation. In *Proceedings of the 2009 Winter Simulation Conference, WSC 2009, Austin, TX, December 13-16, 2009. WSC 2009*, pages 992–1002. IEEE, 2009.