

A Micro-Level Data-Calibrated Agent-Based Model: The Synergy between Microsimulation and Agent-Based Modeling

Karandeep Singh^{***†}
Chang-Won Ahn^{***†}
Korea University of Science &
Technology (UST)
Electronics and Telecommunication
Research Institute (ETRI)

Euihyun Paik^{*†}
Jang Won Bae[†]
Chun-Hee Lee[†]
Electronics and Telecommunications
Research Institute (ETRI)

Abstract Artificial life (ALife) examines systems related to natural life, its processes, and its evolution, using simulations with computer models, robotics, and biochemistry. In this article, we focus on the computer modeling, or “soft,” aspects of ALife and prepare a framework for scientists and modelers to be able to support such experiments. The framework is designed and built to be a parallel as well as distributed agent-based modeling environment, and does not require end users to have expertise in parallel or distributed computing. Furthermore, we use this framework to implement a hybrid model using microsimulation and agent-based modeling techniques to generate an artificial society. We leverage this artificial society to simulate and analyze population dynamics using Korean population census data. The agents in this model derive their decisional behaviors from real data (microsimulation feature) and interact among themselves (agent-based modeling feature) to proceed in the simulation. The behaviors, interactions, and social scenarios of the agents are varied to perform an analysis of population dynamics. We also estimate the future cost of pension policies based on the future population structure of the artificial society. The proposed framework and model demonstrates how ALife techniques can be used by researchers in relation to social issues and policies.

Keywords

Artificial life, artificial society, microsimulation, agent-based modeling, social simulation, data-driven simulation, policy evaluation, virtual society

1 Introduction

Artificial life (ALife) is the study of the living world by creating an artificial world in a machine using a simulation. Modeling and simulation are extensively used by scientists and researchers to build artificial populations and study them from their desired perspectives. These populations can include living cells and organisms, including humans. Virtual human populations can be created in a machine to study human behaviors and behavioral byproducts such as human societies. Traditional approaches have not been very effective in tackling artificial human society simulations, especially

* Contact author.

*** Department of Computer Software, Korea University of Science & Technology (UST). karandeep.singh@etri.re.kr (KS); ahn@etri.re.kr (C-WA)

† Smart Data Research Group, SW-Content Research Laboratory, Electronics & Telecommunications Research Institute (ETRI). jwbae@etri.re.kr (JWB); ehpaik@etri.re.kr (EP); ch.lee@etri.re.kr (C-HL)

when dealing with underlying requirements such as nonlinearity, self-organization, and autonomy. ALife can offer some new developments and breakthroughs in this regard. In fact, ALife now includes the techniques and philosophy needed to handle the requirements mentioned above [25].

The study of artificial human societies involves the creation of a complex system composed of various interlinked components. This complex system is modeled either by simple aggregation systems, which treat the constituent units (humans in this case) uniformly and ignore their individual properties, or by assigning heterogeneity to these constituent units, which are provided with at least some degree of autonomy. Therefore, people in an artificially created society can be identical or heterogeneous. The property of heterogeneity is closer to the real world, but involves the execution and management of more complex model systems. In ALife, such complexity is often implemented by agent-based systems. The approach of ALife to the creation and study of artificial human societies is often called *social simulation*. ALife, at its very core, is concerned with producing the bottom-up behavior of the constituent units, which is also a property of agent-based systems [4, 31].

Agent-based social simulation, or *social modeling*, is a wider term that includes all the aspects of ALife. Social simulation simplifies the ALife approach to artificial human societies by leveraging assumptions and a limited implementation of various aspects related to population dynamics. The idea is to implement only the essential properties of the target system. Agent-based models include the basic idea of complex systems: The constituent units of the agent-based systems (the agents) interact to form the global properties of the system [18, 32]. Such artificially created societies can be used by scientists and modelers to analyze and predict the dynamics of human societies. Experimentation on an actual human population by varying control factors, especially if the target population is not aware of it, is a long, expensive, and unethical process. A harmful result may permanently harm the life of an individual.

Advances in techniques and the processing power of computer machines, such as high-performance computing, are making it possible to perform simulations at a larger-than-ever scale. Until recently, agent-based systems were limited by the processing power of machines. However, it is now becoming possible to process millions of agents in simulations. Simulations involving interactions among intelligent agents are computationally rather heavy. Large-scale simulations help us to reduce unexpected errors in the simulation and aid in discovering the hidden trends from the simulation results. For large-scale simulations, it is imperative to use technologies like distributed and parallel computing. However, the use of these technologies presents practical challenges concerning the software and hardware needed for distributed and parallelized computing of the agents. Keeping these key points in mind, we designed a framework that can fulfill all these requirements and is easy to use even for scientists and modelers from fields other than computer science. There are many simulation frameworks available that support agent-based simulations, but are lacking in one or more areas. Primarily, these tools do not permit the efficient large-scale processing of agents. Large-scale processing of agents is becoming important not only in social simulation, but also in other fields such as biology, economics, and humanities. The large numbers of agents, as well as their complexity, are outscaling the existing single-machine architectures. Therefore, we implemented and developed a framework catering to these requirements.

Microsimulation modeling (MSM) and agent-based modeling (ABM) are two popular techniques used for simulation and modeling of artificially created societies in a computer simulation. MSM is a classical technique, while ABM is comparatively new. MSM models individuals with microdata from a real population data set and delivers the final output by adding up simulated individual outputs. ABM also considers and models each individual, but places more emphasis on the interactions and feedback among the individuals in the virtual population. These two techniques have similar features, such as using a bottom-up approach and having heterogeneous entities in the virtual population. However, they have some differences in defining the individual behaviors of the constituent entities. MSMs describe behaviors for different events based on transition probabilities derived from population microdata; ABMs mostly have rule-based behavioral definitions. Considering the features of these two techniques, a hybrid system based on the strong aspects of both of them can be built, which will utilize the microdata as well as incorporate lifelike rule-based decision making to yield

better simulation results. Working along these lines, we have prepared a model to simulate population dynamics using a newly implemented framework. We considered the Korean population as a use case for our current study. The Korean population census data has been used to initialize the simulation as well as to arrive at the required transition probabilities and decision rules.

Experimenting with artificial societies using ALife techniques is becoming more important in view of problems like low birth rates, aging societies, and increasing dependent populations. We discuss these problems in detail in Section 5. Extensive research has been carried out, and many different models have been proposed to analyze human population dynamics. Some well-known examples of these models are discussed in the related research section, Section 2. However, each of these models has certain drawbacks. They are either microsimulation-based or agent-based, each kind possessing certain limitations. For example, the existing models do not take into account microdata for the initialization of the individual agents, and do not use microdata or interactions among agents for decision making; they also use small sample sizes, which can lead to unexpected outputs in the simulation. Although these models provide a general understanding of population dynamics, such as the influence of family welfare policies, and predict future population structures, they do not provide the methods to arrive at desired numbers or structures of the population.

Based on the discussion above, we implemented a new *agent-based computational demography* (ABCD) framework, capable of simulating up to ten million agents, as well as a novel simulation architecture. For this research work, we will refer to this framework as *microsimulation and agent-based modeling for demography* (MAMD). Then, using ALife concepts, we adopt this simulation architecture to implement the new agent- and microsimulation-based hybrid model and generate an artificial human population and society. Further, using this artificial society, we present an analysis of the population dynamics, focusing on control, analysis, and prediction from the micro-level perspective. For the current study, the Korean census data and social setting are used for the model initialization. The results obtained from this analysis show that education and divorce play a significant role in the future population size and structure. The proposed model is unique in its use of a combination of the following characteristics: MAMD framework, microdata for initialization and decision making, the Korean context to arrive at decision rules, the hybrid technique based on MSM and ABM, and the large data set (2% of the Korean population, or about 892,000 individuals), as well as control, analysis, and prediction from the individual-level perspective. In general, MAMD can be used to generate artificial populations with the desired attributes of agents and can be used by scientists to pursue their research and experimentation using “soft” ALife techniques. One such example is a virus infection model developed using MAMD containing 5,000,000 virus agents, which can influence each other, recover, and move from one node to another.

The rest of this article is organized into six sections. Section 2 includes the details of the proposed MAMD framework and constituent agents of the developed artificial society and population. Section 3 describes related studies on models containing artificial populations developed for the analysis of human population dynamics. Section 4 explains how MSM and ABM differ as well as their common properties, giving opportunity for their hybrid use. Section 5 includes an explanation of the structure of the case study. Section 6 contains the results and analysis of the case study, whereby we analyze the outputs of the simulation, predict the future population structures, and estimate the expenditure on pensions as a percentage of GDP, all depending upon the future population structure of the artificial society. Section 7 concludes the article.

2 Framework

The MAMD model includes large-scale agents and their intricate interactions; therefore, the framework that supports MAMD simulations is required to deal with a large amount of computational complexity. Hence, the proposed framework has been designed and built to be an inherently parallel ABM framework. Also, considering the fact that the end users of the framework might not be familiar with parallel and distributed computing, it is required that the framework provide support

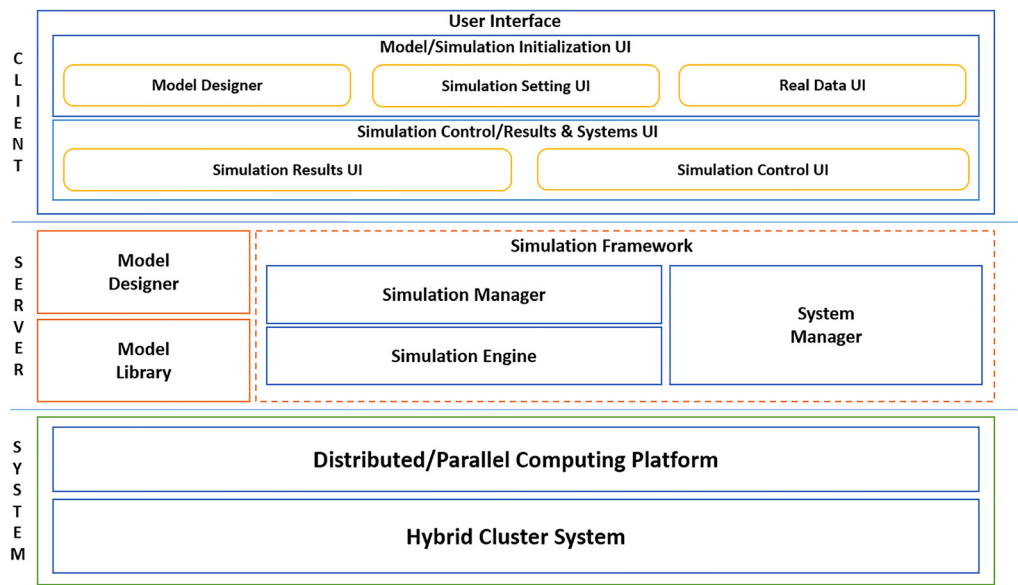


Figure 1. MAMD architecture.

for such users. Ease of use has been considered while designing the user interface (UI) of MAMD. In this way, users are only required to focus on the modeling, and MAMD takes care of the computing needs of the simulation.

The high-level architecture of the proposed framework is shown in Figure 1. The architecture has four main components: user interface (UI), simulation framework (SF), model designer (MD), and computing platform (CP).

The UI block consists of a modeling and simulation UI, which deals with the model, simulation control, real data, and the visualization of results. The model designer UI supports graphical model creation for end users to develop simulation models easily and also provides a tool for the visualization of results obtained from the simulation. Specifically, the model's structure is generated in the .xml files by this UI component, which further generates the simulation code from the model designer component. The MD block consists of the model designer and library module. The MD block is closely connected with the model designer UI. The model code is generated in the C language in .c and .h files and is stored in the model library (ML) component. The model's architecture is generated in the .xml files, whereas the detailed model code is generated in the C language. Further, the executable code generated and stored in the ML is linked to the simulation framework component (SF), where this code is invoked for the simulation to start and proceed.

The SF block holds three components required to execute agent-based simulations: the simulation manager, simulation engine, and system manager. The simulation manager module controls the simulation execution, including agent scheduling, job management, I/O control, and behavior execution. The system manager module monitors and distributes system resources during simulation execution and maintains the hardware resource information. The simulation engine is the core module of the simulation framework (SF), which provides the execution environment for simulation and cooperates with the other modules to perform and control the simulation. It supports parallel and distributed simulation execution, with features including a distributed cache. It manages agent behavior and interactions, and provides MPI-based communication between distributed nodes and caches. Lastly, the computing platform (CP) block supports parallel and distributed simulation execution of MAMD models with the hybrid cluster system consisting of a distributed multi-core execution environment. It is also worth noting that the individual component blocks in the proposed

framework are designed to implement the purposes mentioned above: to support large-scale agent-based simulations (in fact, the architecture applies a hybrid cluster system that supports the processing of about ten million agents in parallel), as well as to facilitate parallel and distributed simulation execution. The framework offers user interfaces that provide support for end users not only to develop MAMD models but also to set up the computing environment settings. The characteristics of the simulation framework are summarized as follows:

- *Simulation framework*: The simulation framework executes several simulation models, such as population dynamics and welfare simulation models.
- *Number of agents*: The proposed simulation framework can execute up to 10 million agents in parallel.
- *Agent types*: The simulation model can have agents of type person, family, region, and so on.
- *Large-scale parallel and distributed simulation engine*: Modeling and simulation functions based on several methods for population dynamics, and the management of various models and data.
- *Distributed simulation parallelization*: CPU-based multi-core distributed system technology, distributed parallel technology, cloud interlocking architecture.

The multi-node simulation system consists of two nodes having two five-core CPUs running on an Intel Xeon E5-2650, V2, 2.6-GHz, 64-GB DDR3 FBDIMM. Using five cores, the average time per step of simulation execution is 3.575 seconds. The technical details of the implementation of each of the framework components and the software development life cycle of MAMD are out of the scope of this article and will be covered in a future one.

Agents in the simulation are created by the framework when the simulation starts, according to the initial model setting, and can also be created dynamically as the simulation proceeds. This dynamic addition of agents can be done by the agents already in the system, such as the offspring of an agent couple. The dynamic attributes of agents are updated at every simulation time step according to their status, the status of other agents related to them, and the characteristics of the environment. The agents can also be dynamically removed from the simulation (during, for instance, a mortality event) and in such a case, every other functionality of the agent is terminated and it no longer participates in the simulation.

Stochastic processes in the simulation are MSM-based, and deterministic ones are ABM-based. The model setting configured by the modeler determines whether the interactions will be part of the simulation. This configuration can be handled at the UI itself, and the user does not need to worry about handling or optimizing interactions among the agents. Thus, code generation by the framework incorporates the stochastic as well as the deterministic processes. As already stated, communication between the agents is handled by MPI-based communication, and the interacting agents can be present on different computing nodes.

Agents in the artificial society undergo simulated life events as depicted in Figure 2. An agent's life covers matrimony, fertility, education, migration, economy, and mortality modules. Only domestic migration is considered in the present model. All these events are regularly updated on each of the simulation ticks, which represent one year in this case. The matrimony event requires message passing between the agents, causing interactions between them. The differently colored functions in Figure 2 handle these interactions. Message passing enables the agents to find a suitable match for themselves by exchanging information such as age and education. Message passing is further based on the rule set we have defined for the agents, keeping the Korean social context in mind; for example, the educational difference between agents must be no more than two levels.

Figure 3 describes the agents' matrimony behavior in more detail. If an agent is not married and satisfies sufficient conditions for marriage by the associated transition probability, it proceeds to the marriage process by looking for other relevant agents. As described above, this interaction is governed by the rule set defined for the agents, which the agents can use to exchange messages and information

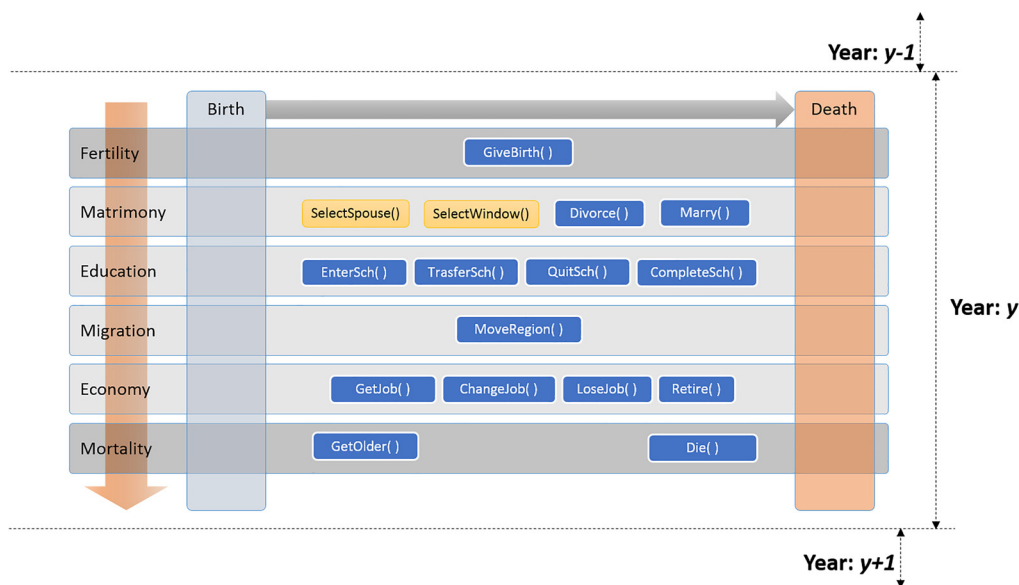


Figure 2. Life events of an agent in the MAMD model.

and look for a potential spouse. Functions Marry(), Divorce(), SelectRegion(), MoveRegion() implement the stochastic behaviors of the agent. Functions SelectWindow() and SelectSpouse() are the deterministic processes of interaction based on a rule set. Married(), MakeCouple(), UpdateCouple(), UpdateFamily(), UpdateHousehold() are state transitions of the agents.

The MAMD (or ABCD) is an ongoing project, and it is still under development as we are in the process of incorporating newer functionality. We are developing it as a proprietary tool that is funded by government and private company partnerships. Due to the project regulations and copyright policies, the tool set cannot be shared as an open source project for the time being. However, we are very open to potential collaborative activities, and researchers can contact the authors for further details. We are planning international collaborations to challenge our techno-socioeconomic problems with various complex social simulations.

3 Related Research

In this section we discuss some of the related research, focusing on the study of population dynamics using artificial populations and societies created in the simulations. The popular techniques for the analysis of population dynamics are statistical, microsimulation-based, and agent-based. The statistical methods use mathematical equations to determine how the model would evolve. This technique suffers from the problem of inflexibility regarding explanation and understanding of demographic situations [8]. By the choice of method, the analysis of the population dynamics can be done from the collective (macro) or individual (micro) perspectives. The collective, or macro-level, population dynamics performs the analysis based on behavioral patterns or aggregation properties. The individual, or micro-level, one performs the analysis based on the individuals' choices in the population. With a combination of both, we can obtain a new vision for understanding the issues pertaining to population dynamics.

The MSM simulation is driven by transition probabilities, derived from the actual data, that determine the changes and progression in the simulation. MSM for the population dynamics, as well as in general, requires a large sample of detailed data, such as age, location, and household income. This

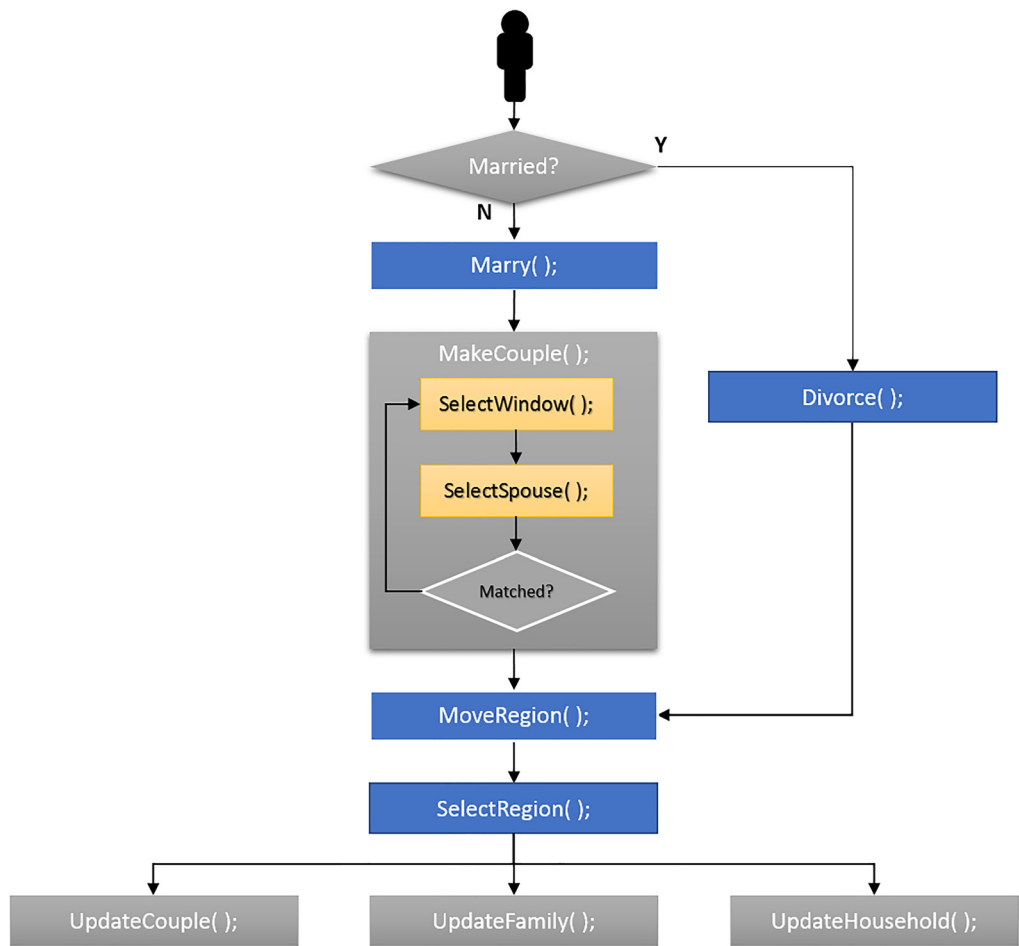


Figure 3. Overview of the matrimony module of the MAMD model.

data is used to assign properties or states to individuals. The idea is to represent the constituent units of the simulation with individual characteristics, which are based on real data rather than the average or distributed values. MSM starts by deploying the artificial population where properties are associated with a particular unit or individual so that the distributions of the virtual world population conform to the real one. There is neither any interaction among the constituent units nor any feedback from the macro to the micro level. When the simulation ends, the simulation results are provided as the aggregation of individual properties. As a technique for socioeconomic policy analysis, the microsimulation model was pioneered by Orcutt [30]. He identified and represented individual entities in the socioeconomic system by temporal changes in their behaviors. MSM describes the simulation system at the individual level, so that the system's dynamics are represented as the accumulation, or adding up, of individualistic values of attributes, that is, it represents a bottom-up approach [19]. The simulation models developed by MSM were utilized as an evaluation tool for public policies [22]. Many MSM-based models have been developed in different countries for different purposes: DYNASIM3 in the USA for retirement and aging issues, DYNACAN in Canada for fiscal and policy analysis, MOSART in Norway for general population dynamics, SESIM in Sweden for the pension system, SAGE in the UK as a tax model, DYNAMOD in Australia for population dynamics and policy changes, and DESTINIE in France for pensions.

The agent-based model consists of autonomous agents with their own set of rules that are used to perform their actions. They take into account their characteristic properties, the environment where they are situated, and most importantly, their interactions with other agents. The interaction is understood as a key factor because it enables an emergence or a formation of trends during the simulation and helps to analyze complex population dynamics [6]. Agents have properties like autonomy, learning, perception, heterogeneity, proactiveness, reactivity, bounded rationality, mobility, and especially interactions with other agents [21]. The origin of ABM can be traced back to the Neumann von machine [27], which, though it was a cellular automaton, can be considered as a starting point. Some famous early examples of ABM are the segregation model by Schelling [40] and John Conway's Game of Life [9]. Mainly during the last two decades, ABM has been improved from the theoretical viewpoint, and its application range has widened in the social sciences [14], biology [37], and economics [15]. The interactions among individuals produce an emergence in the simulated system. The primary assumption of ABM is that interactions can create a system behavior that is more, or less, than the aggregated sum of the individual activities. The emergence generated from ABM is expected to provide newer viewpoints for understanding a target system. There have been a variety of ABMs developed to understand population dynamics. The agent-based social simulation for population dynamics was initiated by Billari and Prskawetz [6, 7]. The majority of the research previously conducted in demographic contexts has been proposed from the perspective of interactions, such as family formation [6], residential mobility [5], households [16], and migration [21].

Policy evaluation is performed to investigate the effectiveness of policies and to gauge their cost, efficiency, and merits. There are many examples of policy evaluation using simulation models. Pensim2, developed by the British Department for Work and Pensions [13], is a micro-level simulation model used for estimating pension policies in the UK. EuroMod is a simulation model for assessing tax policies in the European Union and is utilized for estimating the size of future populations and the tax policies for them [44]. Also, TRIM (Transfer Income Model), developed in the USA, has been applied to analyze potential impacts of changes in tax and health policies [17].

The ABM models take into account the individual or micro-level dynamics, but generally do not consider real data for the initialization of the simulation. The rules and inferences used by agents are also often assumed analytically rather than derived from the actual scenarios of the population under consideration. In contrast, MSM models are completely dependent on high-quality detailed data, and each individual in the virtual population acts in isolation from the others. This lack of interaction means that such models fail to support the phenomenon of emergence from the simulation and certain missing patterns remain undiscovered. Data-fed simulation was proposed in articles like [41], which was an extension of the well-known "Wedding Ring" [7]. Hassan et al. [20] proposed the use of data in an agent-based simulation. Bae et al. [2], Sajjad et al. [39], and Singh et al. [42] proposed general population dynamics models. Our work is an improvement over these studies in enhancing the general structure of the model, in increasing the number of agents in the simulation, and in performing validation, control, analysis, and prediction. We also estimated the future pension expenditure of the Korean government as a percentage of its GDP, depending on initial conditions in the simulation.

4 Microsimulation plus Agent-Based Simulation

In this section, we will discuss the properties of both MSM and ABM with regards to modeling artificial societies and populations. Analysis of the characteristics of MSM and ABM reveals that these techniques have some common features as well as some that are different. The bottom-up approach—smaller building blocks integrated to give rise to bigger ones—is one feature that is common to both. In both ABM and MSM, the constituent units are the building blocks of the simulation. Also, each constituent unit is parameterized using the actual individual information from the micro (individual) data, which enables heterogeneity in the virtual population. On the other hand,

interactions that are almost absent in MSM are a significant and key feature of ABM. These different features arise from the nature of the techniques. The primary objective of microsimulation is a projection of the future, provided the present circumstances of individuals remain similar, so it uses transition probabilities derived from micro population data. The objectives of ABM are to analyze, understand, and explain the behavior of the target system, and discovering trends from the simulation is one way to accomplish them. Therefore, ABM is more focused on how the interactions are taking place among the agents. The stochastic processes in the simulation are defined using the Monte Carlo method, that is, random number generation checked against some probability. Most interactions of ABMs are set and calibrated against the target system's social setting, social norms, and assumptions according to the social context and previous research results.

Microsimulation defines the behavior of constituent units using stochastic processes, mainly transition probabilities, which implies that it requires high-quality detailed initial data. As the operations of the constituent units in MSM are comparatively simple during the simulation, MSM has lower modeling complexity. On the contrary, each constituent unit in ABM usually acts in accordance with its rule set and inferences, which are assumed to be in line with the general trends or the social setting of a target system, and are therefore less dependent on the real data. As ABM involves interactions among agents, it requires more resources, and the simulation process is more complicated than MSM. The differences in their primary objectives and functioning also produce different requirements for the initial data and different simulation complexities.

To summarize, MSM and ABM have properties that are similar, such as heterogeneous individuals, leveraging a bottom-up approach, and the micro-level modeling of individuals. On the other hand, there are some features that distinguish them, such as interactions that are absent in MSM, but prominent in ABM; the objectives, which are projections in MSM and explanations in ABM; and the state change process for an individual, which is stochastic in MSM and rule-based in ABM [42]. The detailed discussion of the features of both these techniques led to the preparation of a model that possesses common features of both MSM and ABM and where the distinguished features are used to complement each other. In order to achieve this, we needed to identify the relevant characteristics of the model to be handled by the appropriate technique. We segregated these features and identified the behaviors of individuals such as mortality, fertility, and migration as controlled by the stochastic processes. We defined finding a marriage partner, or *matrimony*, as an *interaction*, or deterministic behavior. Stochastic processes are handled by transition probabilities that are derived from real data, whereas deterministic processes are handled by the rule set.

The proposed model is used to simulate the Korean population dynamics by using a large number of agents (2% of the Korean population, or about 892,000 agents). In this model, the individual characteristics of agents such as gender, age, income, region, fertility, and mortality are derived from actual empirical data, and interactions between agents, such as finding a marriage partner, are governed by social norms and setting. The data required by the MSM part is taken from various government sources and studies, which are readily available these days. Governments generally collect this data to analyze the population size and many related matters. As stated above, the stochastic processes are handled by transition probabilities, which are derived from real data, whereas the deterministic processes are handled by a rule set derived from the social norms of the target society, previous research results, and assumptions based on these. The resulting artificial population can be subjected to various social environments, and the analysis of population dynamics of this artificial society can help us predict and understand future population dynamics.

The size and structure of the population are affected by a variety of factors such as fertility, mortality, education, employment, migration, and matrimony. At present, our model takes into account all these factors except international migration. The separation of these modules helps us to better understand and analyze the results obtained from the model. The effect of these factors on population dynamics can be easily monitored by bringing about changes in these factors, running the simulation, and then analyzing the results. In this way, we can identify the factors that have a major influence on the population, and further policy changes can be made in accordance with the observations. The proposed model is general in nature, but at present, the data used and case study is

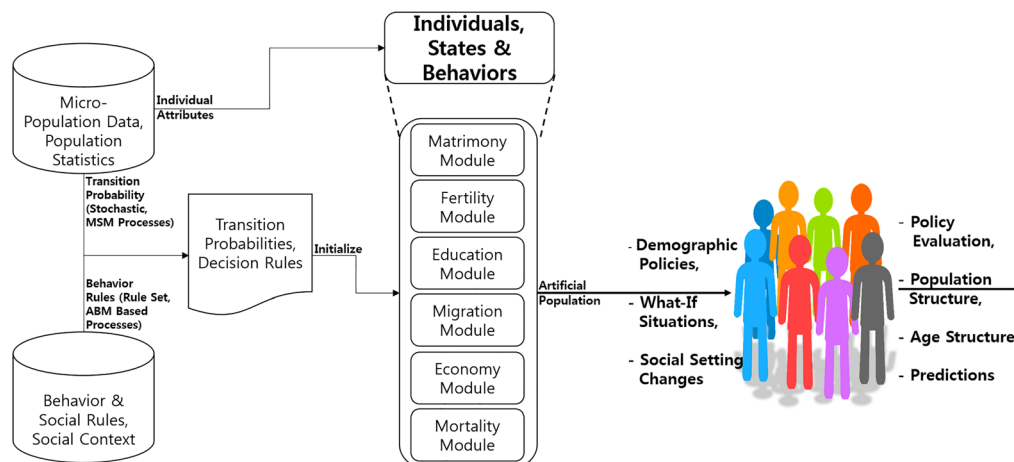


Figure 4. Conceptual diagram of the artificial society of the proposed model.

based on the Korean context. The careful use of the MSM, ABM, and actual data for the model helps us to improve the simulation results and brings the simulation a step closer towards eliminating outliers. The proposed ABM + MSM model is shown in Figure 4.

5 Korean Population Analysis Using the Micro-Level Population Model

In general, excessive population growth is the main focus of population-related studies and analysis. Recently, however, this field has received special attention from the many countries of the world for a different reason, namely the problem of low fertility and inadequate population growth. While countries like Niger have very high fertility rates (6.76 children per woman), others like Singapore (0.81), Korea (1.25), and Japan (1.40) have low fertility rates [47]. The fertility rates are well below the sustainable rate of about 2.1 [10]. The study of population dynamics is carried out to understand and predict the changes that occur in a population over time. It is the scientific study of a population, which focuses on the variation of size, distribution, and age composition of the population over the course of time. The understanding of this phenomenon is rather difficult, because it is not only related to the aggregate behavior of people but also to the individual-level choices that people make according to their personal and social surroundings. The primary focus of this case study is on the size and age structure of the population.

The problem of low fertility mostly occurs in the developed countries of the world. With development come different preferences and lifestyle changes that are considered to be the reason for low fertility [26]. Although a layman might think so, it is incorrect to believe that a smaller population necessarily implies better facilities for fewer people rather than fewer facilities for many people. Many problems are associated with low fertility rate. Low fertility means that fewer children are being born and therefore fewer people will grow up to be available as a part of the workforce. As stated above, low fertility is generally associated with developed countries, where the life expectancy of people is high due to better health care. In fact, the life expectancy in Korea is 82 years. In such scenarios, the size of the dependent population, that is, the population in the age groups 0 to 14 years and above 65 years, increases, primarily due to increase in the aged population above 65 years of age. In countries where a low fertility rate prevails and a large elderly population exists, governments have to invest more money to cater to the needs of the elderly population with pensions and other facilities, using the decreasing available workforce resulting from decreased fertility.

This expenditure can create tremendous pressure on governments and on the working population, which can lead to a variety of social, welfare, and economic problems.

Governments have undertaken many policies to alleviate the problem of low fertility, but the outcomes of these efforts have not met expectations. For instance, despite the changing policies related to maternity and childbirth, the current fertility rates of Korea and Japan are 1.25 and 1.40, respectively; the fertility rate of Korea is 220th among the 224 countries listed in [23, 47]. The only countries with a lower fertility rate than Korea are Singapore (0.82), Macao (0.94), Taiwan (1.12), and Hong Kong (1.19). A large amount of money has been spent and many policy changes have been implemented to promote an increase in the fertility rate. As satisfactory changes have not been seen, there is a need to test and evaluate such policies before they are implemented to optimize the results as well as the resulting expenditure. This is where ALife comes into the picture. As stated in the introduction, ALife now has both the techniques and the philosophy to provide a new viewpoint and promote the development of this field. The optimization of public policies designed by governments is vital for the welfare of people. However, future policy design and evaluation is complicated, as it is closely related to the population size and structure. Therefore, to design an effective policy and evaluate it, the population characteristics should first be understood and predicted efficiently. Subsequently, the effectiveness and correct evaluation of policies can be assured by predicting and experimenting on an artificial society.

The artificial society's population consists of agents, which represent individual people. An agent in the simulation is a self-contained, modular, and uniquely identifiable individual. It is autonomous and has states that can vary over time. It is social in that it interacts with other agents for decision making related to matrimony. The attributes of agents can be static or dynamic. Static attributes are the ones that do not vary during the simulation, viz. identifier, gender, birth year, and birth region, while dynamic attributes are those that evolve or change with the simulation, such as age, employment type, spouse, and region of residence.

Agents not only have these attributes, but also exhibit behaviors. Agent behaviors are the activities that are related to population dynamics. Behaviors include mortality, health, education, economy, matrimony, fertility, and migration. For an agent, these behaviors are interconnected. For instance, for a particular agent, the level of educational attainment determines the type of job he is doing, whereas the type of job may determine the region of residence of the agent. The transition probabilities and deterministic rules are derived from population micro data. The behaviors of the agents evolve based on the criterion set and through interactions with other agents.

The data used for the analysis in our model is the Korean population census data. The sources of this data are the Korean Statistical Information Service (KOSIS) [23] and Statistics Korea (KOSTAT) [43]. KOSIS provides detailed data regarding the population census, containing a wide variety of fields with specific data regarding individuals. We have filtered out the required data from the available data dump. Data from specific fields (age, region, employment, education, family information, etc.) are chosen and are used for the model initialization. For experimentation purposes, we are currently using about 2% of the micro population data, which is equivalent to the data of about 892,000 individuals. By using this large data set to initialize the heterogeneous agent population in our model, we hope to discover emergent behaviors from the agent population and minimize the outliers in the simulation.

The simulation begins with micro-data-seeded agents wandering in virtual space, looking for suitable partners, and getting involved in matrimony. The simulation is a discrete-time event-based algorithm. The simulation proceeds with a yearly time step, and each year, the dynamic attributes and the behaviors of the agents evolve and are exchanged with the other agents. By measuring the various attributes of the virtual population, such as the number of births, education levels, and employment status, we can analyze population behavior as well as predict the future population structure. We vary the agents' rule set for decisions and change various conditions in the model's artificial society to perform an analysis and make observations.

In this case study, we present the analysis of the Korean population with respect to its size and structure. As already discussed in previous sections, low fertility causes the working population to

Table 1. Summary of various scenarios created in the simulation by varying different parameters.

Activity	Parameters	Scenarios	Significance
Marriage	Difference of age	Male up to 5 years older, female up to 3 years older (2 scenarios)	The age difference between couple when getting married
	Difference of education	Random, 1 level, 2 Levels (3 Scenarios)	Difference of Education Level between the couple
	Rate of divorce	Same, 30% increase, 30% decrease (3 scenarios)	30% change in divorce probability
Education	Rate of education	Same, 15% increase, 15% decrease (3 scenarios)	Changing the probability of having higher education by 15%
Employment	Rate of employment	Same, 15% increase, 15% decrease (3 scenarios)	Changing the probability of getting Employment by 15%
Total scenarios	$2 \times 3 \times 3 \times 3 \times 3 = 162$ scenarios		Each scenario run 10 times

decrease and the dependent population to increase; therefore, we focus on the working population in the age range of 15 to 65 years and the dependent population in the age groups from 0 to 14 years (the young population) and above 65 years (the elderly, or aged, population). The choice of the sizes of these groups is based on the fact that a larger working population experiences less stress on itself in supporting the dependent population. In the long run, the working population is determined by the number of people in the 0-to-14 age group, which means that a larger working population would come from more people in that age group, which in turn would come from increased fertility [38]. In addition, we estimated the pension amount required to support the elderly population, based on the future population structure.

To pursue the desired analysis, a wide variety of test cases were designed and the simulations were run for each one of them. While designing these test cases we not only consider the factors that may affect the size of the population, but also take the Korean social context into consideration so that the test cases are plausible for the Korean context. Four main factors were considered and varied for the analysis: matrimony, employment, education, and divorce rate. We varied the matrimony rules for partner selection, which is mainly handled by the interaction among the agents. Employment and education were considered because they are closely related to the partner selection and economic conditions of the agents, which are further considered to be related to the fertility behavior. The divorce rate was considered for experimentation because it has been on a rise in recent years. It has increased 88% in the last 24 years [23]. Given that in the Korean context cohabitation is negligible, we determined that the divorce rate should be a major factor as far as fertility is concerned. We compared our test observations against the predictions made by not changing any of the above factors, which served as our control. To design these test cases, we varied a number of factors, as summarized in Table 1. These rules determine the artificial society’s behavior.

5.1 Matrimony

The rules for matrimony are utilized for choosing a life partner; male and female agents take these rules into consideration while undertaking this activity. A female agent can marry a male agent who is up to 5 years older or 3 years younger than she is, and vice versa for male agents. Education in the simulation is expressed as six different levels. Couples can get married if the education level

difference is up to two levels. The implication of these settings is to manifest the effect of Korean social context in the simulation. As an example, our test case implies that a Ph.D. graduate can marry a person who has at least a bachelor's degree.

5.2 Education of All Agents in Simulation

Higher education is a serious matter in the Korean societal context. Almost 85% of high school graduates go to college. A bachelor's degree is held by 65% of Korean students between the ages of 25 to 34 years, against the OECD average of 39% [12]. Higher education, particularly for women, is also considered as one of the reasons for lower fertility [24]. We determined that experimenting with the education field could be informative. We experimented with no change in education attainment, and also increased and decreased the education attainment of people in general by 15%. This increase or decrease was distributed over the entire agent population.

5.3 Divorce Rate

South Korea's divorce rate has been surging. It has increased by about 88% in the last 24 years, according to data from 1990 to 2014 [23]. As cohabitation is negligible in Korean society and the number of children born out of wedlock is quite low (about 2% [23]), a change in divorce rate may bring about some changes in fertility and hence in the population size. We varied the divorce rate by increasing and decreasing it by 30% and also kept it unchanged.

5.4 Employment Rate

It is postulated that the employment rate also affects fertility choices, especially through the participation of women in the workforce [24]. Higher participation in the workforce may mean less time for family and consequently less tendency towards having children. We varied the employment rate by increasing and decreasing it by 15%, as well as keeping it unchanged.

By combining the various test scenarios described above, 162 test cases were prepared, and a simulation was run for each of these cases. Each simulation or each test case was run 10 times and for 100 years each. In total, this resulted in 162,000 iterations. Further, in the simulation outputs we grouped the data according to age and years. A typical test case output had 4,040 rows of information. This resulted in the evaluation of about 6.5 million rows of data for our analysis.

6 Results and Analysis

Figure 5 presents the comparison of the population numbers of the simulation results with the statistical data provided by the Korean government available on KOSIS [23]. The government provides the actual census data and prediction for the population numbers, along with upper and lower bounds. The comparison of the simulation data with the actual government-provided statistical data serves as a validation of our model. The figures provided by the government are for the entire population, so we scaled our simulation results to the whole population. By analyzing the graph, we can see that the simulated population numbers not only lie between the upper and lower bounds but are quite similar to the predicted values, which is an argument for the quality of our simulation configuration. Figure 5 provides comparisons for the young, working, and elderly population independently, which further prove the correct calibration of our model.

The design of agents and their behavioral properties are discussed in Section 3 and 4. We have explained how an agent in the simulation has stochastic and deterministic properties and how those properties are derived. In particular, we depend on census data provided by the government to obtain the stochastic processes and social norms existing in the target social context (Korean society). Whereas the use of trustworthy and publicly available data is straightforward, mapping social norms into a simulation can be tricky and should be arrived at by proper agreement and discussions with the appropriate domain experts [1, 35, 36]. The social norms in this experiment were derived from discussions with researchers and social scientists, and it was possible to reach a common viewpoint

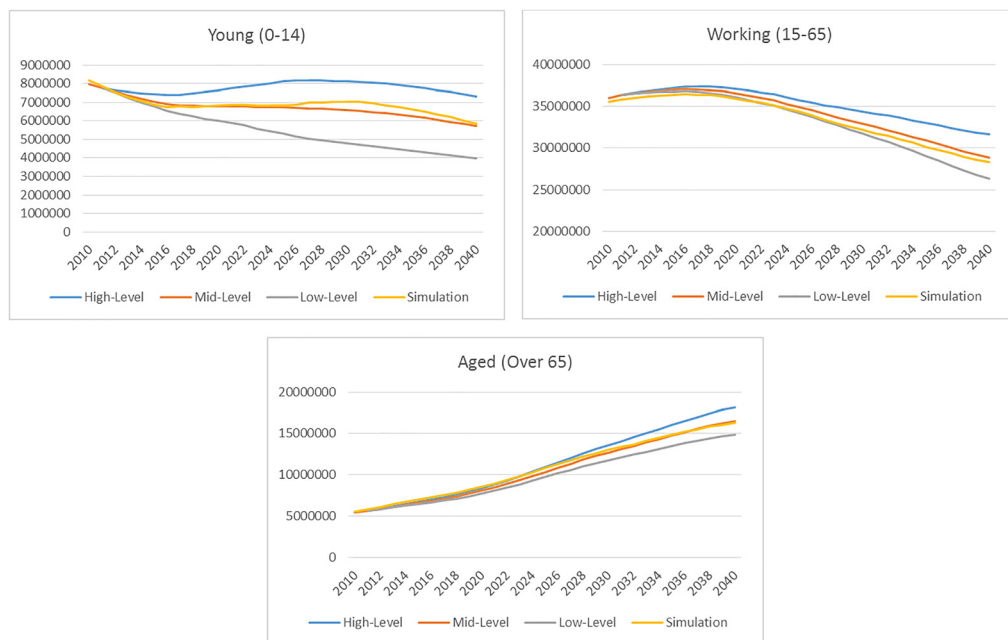


Figure 5. Clockwise from top: comparison of simulation population numbers for young (0–14 years), working (15–65 years), and elderly (over 65 years) population numbers with government-provided population data. In all the graphs, the y axis represents the population numbers and the x axis represents the years.

in this regard. Though these norms cannot be considered precise, it was agreed by the experts to be at a reasonable level of abstraction. Further, independent validation of the population structure for the young, working, and aged populations against the actual census data, from a starting point in the past, yields a solid validation of our model. As stated in [35, 36], the role of *fitness for purpose* is to demonstrate reasonable adequacy, and the fitness of a simulation in the context of an established purpose; the descriptions of the design of components, component interactions, and appropriate expected behavior in a simulation serve as arguments for the fitness for purpose of our simulation system.

We evaluated our results based on the population numbers in the age groups 15 to 65 years (the working population), 0 to 14 (the young age group), and above 65 (the elderly, or aged, population). For the best case, with the largest population, the numbers in the artificial population varied from 892,024 to 980,896; and in the worst case, with the smallest population, the numbers went up to 955,132 from 892,024, before the population numbers started decreasing in both cases.

6.1 Best Case

The test conditions in the best case with the largest population numbers are:

- The female was up to 3 years older than the male agent.
- The education of male and female partners for marriage was random.
- The divorce rate was 30% less.
- The education of people in general was 15% less.
- The employment was 15% less.

This test case not only delivered the best results in terms of overall population numbers, but also gave the highest numbers in the working-age population. Further, this case featured the fourth highest number of people in the 0-to-14-years age category. It conforms to the fact that the higher the number of individuals in the 0-to-14-years age group, the higher the number of people who would grow up and would be available in the workforce.

6.2 Worst Case

The conditions in the worst case with the lowest population numbers are:

- The male was up to 5 years older than the female agent.
- The education of male and female partners for marriage was random.
- The divorce rate was 30% more.
- The education of people in general was 15% more.
- The employment was 15% less.

This test case had not only the smallest overall population, but also the smallest working-age population numbers. It also had the smallest number of individuals in the 0-to-14-years age category, which reaffirms the fact that the number of individuals in the 0-to-14-years age group has a direct effect on the number of working people available in the future. The best- and worst-case overall population, working population, and elderly population graphs can be seen in Figure 6.

If we have a closer look at the graph in Figure 6c, one interesting phenomenon can be noticed: It is the point at which the best- and worst-case scenarios start to differ substantially in the number of elderly people. If we observe the trend and compare it with the graphs in Figure 6a and b, we see that it occurs much later over the course of time. This means that although the population numbers in the best and the worst case start to differ at an early point of time in Figure 6a and b, that happens much later for the elderly population numbers. This implies that, in spite of the fact that the overall population in the best case is higher, the best case does not have the corresponding higher numbers of elderly people. It can further be concluded from this that the increase in the population is due to a higher number of births and a higher number of people in the 0-to-14-years age group, which is a good characteristic of population structure setting. For further clarification, we can take a look at the ratio of the dependent to working population in Figure 6d. The best-case setting in this graph initially has a higher dependent ratio than the worst-case setting; this must be due to the higher number of people in the 0-to-14-years age group, as the number of elderly people is higher in the worst-case setting, as shown in Figure 6c. It is safe to conclude from this that when the 0-to-14-years age group grows up and starts participating in fertility-related activities, the dependent ratio decreases due to an increase in the working population.

The median age of a population gives us an idea of the population structure. It means that half the people are above that age and half are below it. A population with a good fertility rate and a large youth population has a lower median age than a population with a lower fertility rate and a larger elderly population. With the test data generated by the simulation, the median age of the population in the artificial society was also calculated. The calculations were done for two different scenarios, the best case and the worst case, and the results are shown in Figure 6e. For analysis purposes, the median age of the population of the twenty best and the twenty worst cases is shown in the graph in Figure 6f. Continuing the trend of favorable population measures, this scenario again yielded better results. In both graphs, it is clear that the median age is less in the best-case setting. In the graph in Figure 6e, the median age in the best-case setting starts decreasing as early as about 2025, and this trend then continues, creating a substantial difference between the two scenarios towards the end of the simulation, the values being 50 and 60 years approximately for the best- and the worst-case setting, respectively. Similarly, for the top best and worst cases, the median age is less for the

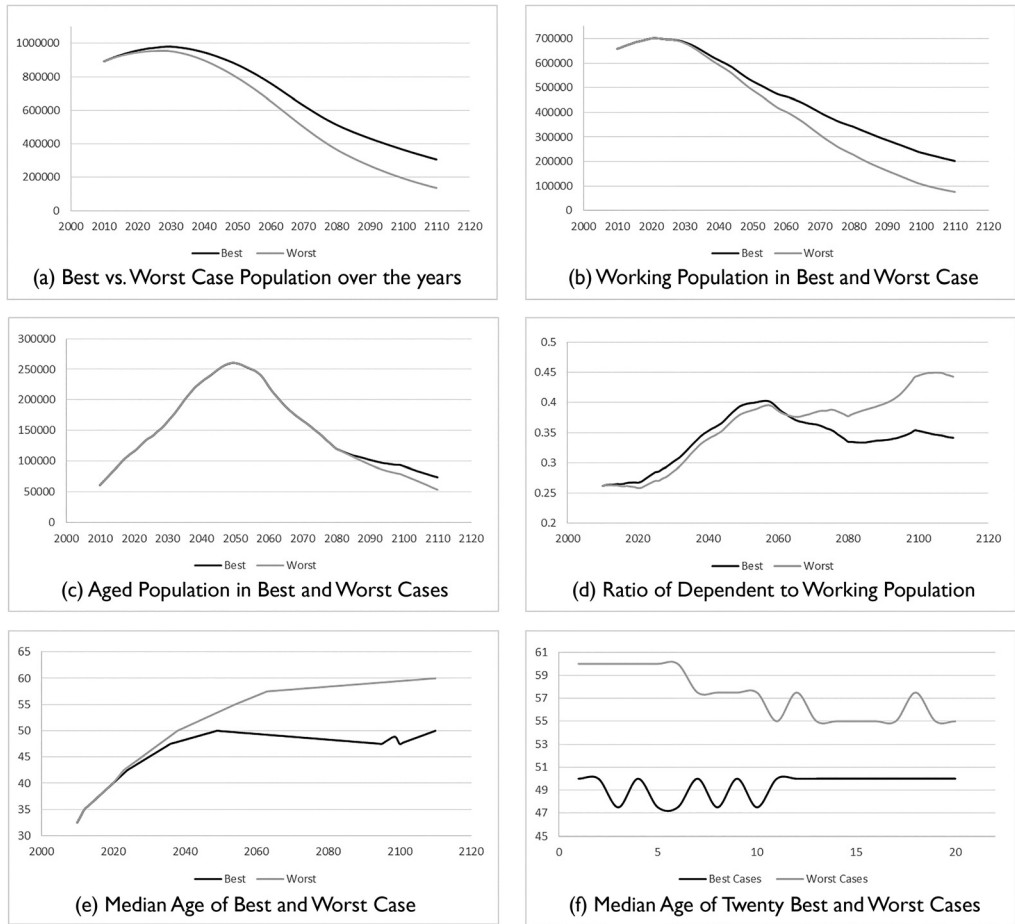


Figure 6. (a) Best- and worst-case total population numbers over the years. (b) Best- and worst-case working-population numbers over the years. (c) Best- and worst-case elderly-population numbers over the years. (For (a), (b), and (c) graphs, the y axis represents population numbers, and the x axis represents years.) (d) Ratio of dependent to working population (y axis is the ratio value, and x axis is the years). (e) Best- and worst-case median age over years (y axis is age in years, x axis is years). (f) Median age for twenty best and twenty worst cases (y axis is age in years, x axis is serial number).

best-case settings. This analysis further solidifies our view of the favorable effect of the best-case setting on the population numbers and structure. The population structure of the best-case setting is shown in Figure 7b.

We also estimated the pension requirements for the best- and worst-case scenarios as a percentage of total GDP of the country. For this particular estimation, the low data availability for the pension-only expenditure of the government created hurdles for comparing it with the actual figures. Nevertheless, we can observe the relative increase in the pension expenditure of the government for the simulated time frame. The pension expenditure of the government would depend upon the total elderly population and the pension amount per person. The pension amount in Korea is a function of income and a certain fixed allowance. To calculate the income, we used the long-term OECD forecast of Korean GDP until 2060 [11] and then extrapolated the data. Based on this data, the formula for pension calculation as given by the OECD [28, 33, 34], and the cap (maximum pension per person), we calculated the pension amount for our simulation time period. Further, using this amount, we calculated the total pension expenditure of the government against the total number of aged people in the simulation for the best and worst cases, as a percentage of GDP, scaled to our

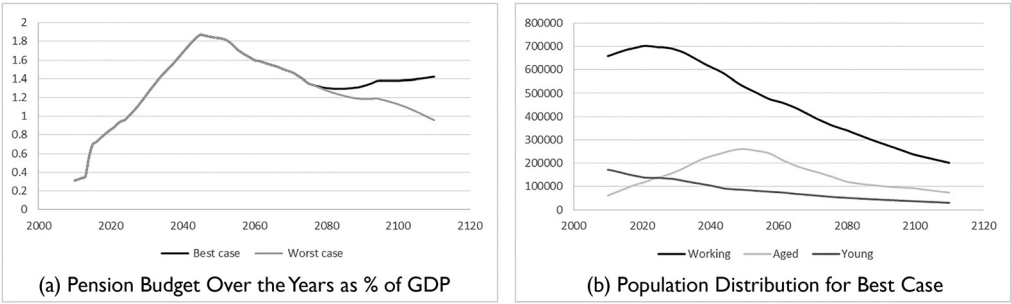


Figure 7. (a) Pension budget as a percentage of GDP over the years (y axis represents the values in percent, and x axis represents year). (b) The population age structure of the best case (y axis is the population number and x axis is the year).

simulation. The resulting graph is shown in Figure 7a. As already discussed, the number of aged people in the best case is almost the same as that in the worst case, implying that the working-population numbers are higher in the best case. The difference between the pension expenditures only starts appearing in the late 2070s, which is in line with Figure 6c, which displays the number of elderly people. The abrupt increase in expenditure at the start, around the year 2013, is due to a change in the pension policy by the government. From the year 2013 onwards, the pension amount per person increased. The flat part of the graph around the year 2045 represents the cap on the maximum pension per person as specified in [33, 34]. Although these numbers cannot be compared directly with the actual expenditure due to lack of data, we can predict the relative increase in the expenditure as compared to the present time. Going by the simulation figures, the current expenditure, for the year 2015, is about 0.70% of GDP, for both the best and the worst case, and that rises to about 1.872% and 1.874%, respectively, in the year 2045 (more for the worst case) before it starts decreasing again. The percentage increase in pension expenditure is thus staggering: 267% and 268%, respectively. Adding to the woes would be the ever shrinking working population to contribute to the economy. This is a worrisome situation.

To assess the contributions of various factors to the population numbers, we prepared a linear regression model of our simulation, which is summarized in Table 2. The independent variables in the regression model are the experimental variables that we varied, viz., age difference, educational

Table 2. Linear regression model with standardized coefficients.

	Total population	Aged population	Working population	Median age of population
Age difference	-0.2071	-0.1175	-0.2201	0.2328
Education difference (random)	-0.3217	-0.2107	-0.3399	0.3825
Education difference (I level)	-0.0121	-0.0117	-0.0118	0.0510
Divorce rate	-0.3047	-0.4317	-0.2864	0.2110
Education rate	-0.7780	-0.8109	-0.7655	0.6608
Employment rate	0.0131	-0.0077	0.0151	-0.0043

Notes. Dependent variables are the total, working, and elderly populations and the median age of the population. Independent variables are age difference, education level difference between the partners (random and I level), divorce rate, education rate, and employment rate. For the dependent variables, the employment rate and I-education-level difference (shown in bold italics) have p -value > 0.05 ; other than these, all the variables have a p -value < 0.05 .

difference, divorce rate, education rate, and employment rate. The dependent variables are the total population, elderly population, working population, and median age of the artificial society's population.

6.3 Discussion

Many interesting conclusions can be drawn from the results obtained through the simulation. The results imply that higher education affects fertility. Korea spends large amounts of money on education—about 8% of its GDP, against the OECD average of 6%. The private expenditure share of education in Korea is also the highest, particularly at the tertiary level [12], which suggests financial stress on parents to raise their children. This expenditure signifies that a lot of attention and resources are invested in the education sector. A study showed that nine out of ten parents feel burdened by the costs of childcare [46], a major portion of which is related to education. Higher education has a positive relation with better jobs and earnings [3, 12], and therefore, the preferences of individuals seem to shift from family towards education and professional life. While advocating a decrease in higher education seems unrealistic, efforts should be made by the government and parents to change the present education policies. Most students seek admission in prestigious institutions of higher education and placement in a large enterprise for handsome salaries and better lifestyles. Students begin working hard in school at an early age, spending long hours in school and utilizing private tutoring. This often spills over to high schools, colleges, universities, and ultimately the workplace.

All this effort and attention towards education leads to less attention towards family life, which includes having children. In a survey, 62.6% of Koreans in their 20s and 30s responded that it is better not to have children if they cannot be raised under financially abundant conditions [46]. 77.4% of people in this survey also responded that children are not necessary for a happy marriage, and 42.3% said that it is better not to have children or that to have children is not a must. The social setting should be such that while pursuing and inculcating the importance of education in children, family life should also be given its due importance. The sheer stress associated with education should be managed—for children as well as parents.

Furthermore, the divorce rate in Korea is growing, and couples increasingly have unhappy marriages, mainly due to stress in their professional life and lack of time for each other [23, 45]. Simulation results show that a decrease in the divorce rate positively impacts the population numbers. A decline in the divorce rate would imply less stress in domestic life and consequently more chances of better attention towards family life. The average annual number of hours worked per worker for Korea is 2,113, which is third highest in the OECD [29], and is bound to have an effect on family life. As discussed above, this job culture might have arisen from the educational culture of the country, so setting things right from the start would bring about a change in overall social setting.

7 Conclusions

In this article, we have demonstrated the implementation of a new agent-based framework called MAMD and discussed its use in the development of an artificial life experiment. The MAMD framework makes use of parallel and distributed computing techniques to process large-scale simulations and can support simulations containing up to ten million agents. The framework design enables modelers and scientists with limited knowledge of parallel and distributed computing to develop simulations without worrying about the underlying computing complexities.

We made use of the MAMD framework and “soft” ALife techniques to prepare an artificial society of humans and run experiments to analyze and predict the important issues of future population structure and future expenditure of the government on the pension policy. We made use of about 892,000 agents, controlled the conditions in the simulation, and analyzed the results. Our aim was to discover the main factors that are affecting the Korean population size and structure.

The MAMD framework implemented a unique population dynamics model, which is a hybrid system of MSM and ABM and possesses the advantages of both these modeling techniques. For the simulation, 162 test scenarios were run to analyze the effect of education, divorce, employment, and matrimony on the population dynamics. We mainly focused on the size and age structure of the population, but also attempted to estimate the future pension budget for the government. The results indicate that educational attainment and divorce rate have a noteworthy effect on the population numbers in the long run. They both are inversely related to the population numbers and a favorable population structure. Both education and divorce rate are of great importance in many developed nations. While the attainment of education cannot be discouraged, the divorce rate is something that could be tackled. Also, how education, work life, and family life should and can go hand in hand is the next major question that should be answered. Providing a problem-free future for the coming generation is the responsibility of today's generation.

This work is a step toward further exploration of ALife techniques in the social simulation field. At the recent conferences on ALife, the contributions of ALife techniques for alleviating social problems, such as low fertility, were minimal. Birth rates are dropping in almost all developed and developing nations, and therefore, this application becomes highly relevant when considering the global population, aging societies, caring for the aging populations, and the limited resources we have for future generations. Our work is a concrete example of the use of ALife techniques in this area. An artificial society generated in this manner could be explored for various types of experiments resembling the ones exhibited in this study. Recognizing the main reasons behind declining fertility rates and consequently skewed population structures could help governments to shape the future population policies in a way that benefits society as a whole. The MAMD approach, modeling method, case study structure, and approach can be implemented by modelers or scientists to gain insight into the desired target systems consisting of artificial populations by controlling the initial conditions and observing the induced changes in the simulation. The fitness-of-purpose arguments should be closely monitored.

As a part of future research, we will involve more of the agent-based component in the simulation and perform a deeper analysis to discover more hidden trends, thereby discovering more of the tertiary and quaternary effects of changes in initial control factors. Our analysis would involve how different factors affect each other, taking into consideration additional details on the micro level, as well as increasing the test scenarios. We are currently exploring the use of machine learning (ML) techniques at appropriate levels in such simulations. ML could be very useful in providing additional insights and in tasks such as the optimization of simulation variables. We also plan to widen the scope of the simulation by including additional modules, such as international migration. In addition, using our model, we plan to analyze the present government policies and, eventually, help to design and reshape social and public policies.

Acknowledgment

This work is supported by the Korea ICT R&D program of MSIP/IITP (10047117, Development of Distributed/Parallel Multi-Dimensional Demographic Micro Simulation Technologies for Population Dynamics and Socio-Economic Experimentation). The authors also extend their thanks to the whole social simulation team for their constant advice during the development of the work.

References

1. Andrews, P. S., Polack, F., Sampson, A. T., Timmis, J., Scott, L., & Coles, M. (2008). Simulating biology: Towards understanding what the simulation shows. In S. Stepney, F. Polack, & P. Welch (Eds.), *Proceedings of the 2008 Workshop on Complex Systems Modelling and Simulation* (pp. 93–123). Somerset: Luniver Press.
2. Bae, J. W., Paik, E., Kim, K., Singh, K., & Sajjad, M. (2016). Combining microsimulation and agent-based model for micro-level population dynamics. *Procedia Computer Science*, 80, 507–517.

3. Baum, S., Ma, J., & Payea, K. (2013). *Education pays: The benefits of higher education for individuals and society*. New York: The College Board.
4. Bedau, M. A. (1997). Artificial life. In M. Matthen & C. Stephens (Eds.), *Philosophy of biology* (pp. 585–603). Amsterdam: Elsevier.
5. Benenson, I., Orner, I., & Hatna, E. (2003). Agent-based modeling of householders' migration behavior and its consequences. In F. C. Billari & A. Prskawetz (Eds.), *Agent-based computational demography* (pp. 97–115). Heidelberg: Physica-Verlag.
6. Billari, F. C., Ongaro, F., & Prskawetz, A. (2003). Introduction: Agent-based computational demography. In F. C. Billari & A. Prskawetz (Eds.), *Agent-based computational demography* (pp. 1–17). Heidelberg: Physica-Verlag.
7. Billari, F. C., Prskawetz, A., Diaz, B. A., & Fent, T. (2008). The “Wedding-Ring”: An agent-based marriage model based on social interaction. *Demographic Research*, 17(3), 59–82.
8. Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting*, 22(3), 547–581.
9. Conway, J. (1970). The Game of Life. *Scientific American*, 223(4), 4.
10. Craig, J. (1993). Replacement level fertility and future population growth. *Population Trends*, (78), 20–22.
11. Economic outlook no. 95—May 2014—long-term baseline projections. OECD. Available at: https://stats.oecd.org/Index.aspx?DataSetCode=EO95_LTB (accessed March 2017).
12. Education at a glance 2014: Country notes. OECD. Available at: <http://www.oecd.org/edu/education-at-a-glance-2014-country-notes.htm> (accessed March 2017).
13. Emmerson, C., Reed, H., & Shephard, A. (2004). *An assessment of PenSim2* (no. 04/21) (Working paper). London: Institute for Fiscal Studies (IFS).
14. Epstein, J. M. (2006). *Generative social science: Studies in agent-based computational modeling*. Princeton University Press.
15. Farmer, J. D., & Foley, D. (2009). The economy needs agent-based modelling. *Nature*, 460(7256), 685–686.
16. Geard, N., McCaw, J. M., Dorin, A., Korb, K. B., & McVernon, J. (2013). Synthetic population dynamics: A model of household demography. *Journal of Artificial Societies and Social Simulation*, 16(1), 8.
17. Giannarelli, L. (1992). *An analyst's guide to TRIM2: The transfer income model, version 2*. Washington DC: The Urban Institute Press.
18. Gilbert, N., & Bullock, S. (2014). Complexity at the social science interface. *Complexity*, 19(6), 1–4.
19. Gilbert, N., & Troitzsch, K. (2005). *Simulation for the social scientist* (2nd ed.). Berkshire: McGraw-Hill Education.
20. Hassan, S., Pavón, J., Antunes, L., & Gilbert, N. (2010). Injecting data into agent-based simulation. In K. Takadama, C. Cioffi-Revilla, & G. Deffaunt (Eds.), *Simulating Interacting Agents and Social Phenomena: The Second World Congress* (pp. 177–191). Tokyo: Springer.
21. Heiland, F. (2003). The collapse of the Berlin wall: Simulating state-level East to West German migration patterns. In F. C. Billari & A. Prskawetz (Eds.), *Agent-based computational demography* (pp. 73–96). Heidelberg: Physica-Verlag.
22. Imhoff, E. V., & Post, W. (1998). Microsimulation methods for population projection. *Population: An English Selection*, 10(1), 97–138.
23. KOSIS—Korean Statistical Information Service. Available at: <http://kosis.kr/eng/> (accessed March 2017).
24. Kye, B. (2008). *Delay in first marriage and first childbearing in Korea: Trends in educational differentials* (working paper). Los Angeles: California Center for Population Research.
25. Langton, C. G. (1989). *Artificial life*. Massachusetts: Addison-Wesley.
26. Myrskylä, M., Kohler, H. P., & Billari, F. C. (2009). Advances in development reverse fertility declines. *Nature*, 460(7256), 741–743.
27. Neumann von, J., & Burks, A. W. (1966). Theory of self-reproducing automata. *IEEE Transactions on Neural Networks*, 5(1), 3–14.

28. OECD Pensions at a Glance 2005. OECD Publishing. Available at: http://www.oecd-ilibrary.org/finance-and-investment/oecd-pensions-at-a-glance-2005_pension_glance-2005-en (accessed March 2017).
29. OECD Statistics. Available at: <http://stats.oecd.org/> (accessed March 2017).
30. Orcutt, G. H. (1957). A new type of socio-economic system. *The Review of Economics and Statistics*, 39(2), 116–123.
31. Parisi, D. (2009). Artificial life and human societies. In L. D. Kiel (Ed.), *Knowledge management, organizational intelligence and learning, and complexity*, Vol. 1 (pp. 270–287). Oxford, UK: EOLSS Publishers.
32. Penn, A. (2016). Artificial life and society: Philosophies and tools for experiencing, interacting and managing real world complex adaptive systems. In C. Gershenson, T. Froese, J. M. Siqueiros, W. Aguilar, E. Izquierdo, & H. Sayama (Eds.), *Proceedings of the Artificial Life Conference 2016* (pp. 26–27). Cambridge, MA: MIT Press.
33. *Pension markets in focus—OECD 2015*. Available at: <http://www.oecd.org/daf/fin/private-pensions/Pension-Markets-in-Focus-2015.pdf> (accessed March 2017).
34. *Pensions at a glance 2015*. OECD Publishing. Available at: http://www.oecd-ilibrary.org/social-issues-migration-health/pensions-at-a-glance-2015_pension_glance-2015-en (accessed March 2017).
35. Polack, F. (2015). Filling gaps in simulation of complex systems: The background and motivation for CoSMoS. *Natural Computing*, 14(1), 49.
36. Polack, F. A., Andrews, P. S., Ghetiu, T., Read, M., Stepney, S., Timmis, J., & Sampson, A. T. (2010). Reflections on the simulation of complex systems for science. In R. Calinescu, R. Paige, & M. Kwiatkowska (Eds.), *2010 15th IEEE International Conference on Engineering of Complex Computer Systems* (pp. 276–285). Piscataway, NJ: IEEE.
37. Politopoulos, I. (2007). *Review and analysis of agent-based models in biology*. Liverpool: University of Liverpool.
38. Prskawetz, A., Kögel, T., Sanderson, W. C., & Scherbov, S. (2007). The effects of age structure on economic growth: An application of probabilistic forecasting to India. *International Journal of Forecasting*, 23(4), 587–602.
39. Sajjad, M., Singh, K., Paik, E., & Ahn, C. W. (2016). A data-driven approach for agent-based modeling: Simulating the dynamics of family formation. *Journal of Artificial Societies & Social Simulation*, 19(1), 9.
40. Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology*, 1(2), 143–186.
41. Silverman, E., Bijak, J., Hilton, J., Cao, V. D., & Noble, J. (2013). When demography met social simulation: A tale of two modelling approaches. *Journal of Artificial Societies and Social Simulation*, 16(4), 9.
42. Singh, K., Sajjad, M., & Ahn, C. W. (2015). Towards full scale population dynamics modelling with an agent based and micro-simulation based framework. In D. H. Kim & B. N. Yoon (Eds.), *2015 17th International Conference on Advanced Communication Technology* (pp. 495–501). Piscataway, NJ: IEEE.
43. Statistics Korea. Available at: <http://kostat.go.kr/portal/eng/index.action> (accessed March 2017).
44. Sutherland, H. (2001). *EUROMOD: An integrated European benefit-tax model* (no. EM9/01) (working paper). Cambridge, UK: University of Cambridge.
45. *The Korea Observer*. (2015). Marriage Rate Declines in South Korea. Available at: <http://www.koreaobserver.com/number-of-marriages-drops-last-year-28213/> (accessed March 2017).
46. *The Korea Times*. (2017). Korea's birthrate lowest among OECD members. Available at: http://www.koreatimes.co.kr/www/nation/2017/03/119_225996.html (accessed March 2017).
47. *The world factbook*. Central Intelligence Agency. Available at: <https://www.cia.gov/library/publications/the-world-factbook/rankorder/2127rank.html> (accessed March 2017).