MODELING AND FORECAST OF SOCIO-ECONOMIC PROCESSES

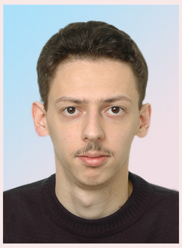
DOI: 10.15838/esc.2019.6.66.4
UDC 332.144; 314.9; 519.876.5; 004.94, LBC 65.04; 60.7; 65.050

Agent-Based Supercomputer Demographic Model of Russia: Approbation Analysis*

Valerii L. Makarov
Central Economic Mathematical Institute of RAS
Moscow, Russian Federation, 47, Nakhimovsky Avenue, 117418
E-mail: makarov@cemi.rssi.ru
ORCID: 0000-0002-2802-2100; ResearcherID: I-9022-2016

Al'bert R. Bakhtizin
Central Economic Mathematical Institute of RAS
Moscow, Russian Federation, 47, Nakhimovsky Avenue, 117418
E-mail: albert.bakhtizin@gmail.com
ORCID: 0000-0002-9649-0168; ResearcherID: S-6203-2016

Elena D. Sushko
Central Economic Mathematical Institute of RAS
Moscow, Russian Federation, 47, Nakhimovsky Avenue, 117418
E-mail: sushko_e@mail.ru
ORCID: 0000-0003-3565-5210; ResearcherID: E-4911-2015



Gennadii B. Sushko
Central Economic Mathematical Institute of RAS
Moscow, Russian Federation, 47, Nakhimovsky Avenue, 117418
E-mail: gennady.sushko@gmail.com
ORCID: 0000-0003-3591-7276; ResearcherID: D-9438-2015

* The study was performed with financial support from the Russian Science Foundation in the framework of research project no. 19-18-00240 "Supercomputer technologies in social sciences" with the use of computing resources of the Joint Supercomputer Center of the Russian Academy of Sciences (JSCC RAS) and the Tianhe-2 supercomputer at the Supercomputer Center of the National University of Defense Technology of China in Guangzhou.

For citation: Makarov V.L., Bakhtizin A.R., Sushko E.D., Sushko G.B. Agent-based supercomputer demographic model of Russia: approbation analysis. *Economic and Social Changes: Facts, Trends, Forecast,* 2019, vol. 12, no. 6, pp. 74–90. DOI: 10.15838/esc.2019.6.66.4

Abstract. The article presents an agent-based demographic model of Russia designed to run on supercomputers. The technologies used in the model allow researchers to create an artificial society with the number of agents up to 109 and effectively parallelize the work of the simulator. The software package designed to implement the model combines separate subsystems written in programming languages of different levels. On the one hand, this provides effective load balancing between computing processes and messaging between agents (implemented in C++), and on the other hand, this simplifies the development of model blocks that implement the simulation of demographic processes (implemented in C#). The demographic processes in the model are simulated based on the actions of individual agents, taking into account their family ties, which they maintain by exchanging messages. Key features of the demographic agent-based models are the following: a) dynamic change in the size and composition of populations of agents – removal of part of the agents (their "death") and the emergence of new ones ("birth"); and b) separation of actions performed at the simulation step in stages, each of which can cause the revision of the general settings that are specific to regions or groups of agents, and/or exchange of messages between agents. In the course of computer experiments, the model has been tested on real data and has shown good results at testing for the following parameters: a) the quality of recreating the age-sex structure of the population for the country as a whole and in the regions with the use of the population of agents; b) the stability of the model and a low margin of error of the results of forecasting the main demographic indicators in comparison with the variants of Rosstat's official forecast; c) efficiency of parallelization of the program code when running on supercomputers. The model is the basic one for an integrated regional simulation model that is currently being developed; however, the model can be useful as an independent forecasting tool.

Key words: agent-based modeling, simulation of demographic processes, supercomputer technologies, application of METIS graph decomposition, demographic forecast for Russia.

Introduction

This article is devoted to the development of an agent-based demographic model for creation of artificial society with many agents-people. It is the continuation of the study presented in [1]. Agent-based models (ABM) are the class of models based on behavior imitation of separate agents, which can act independently in accordance with their interests and possibilities given to them by an environment (including other agents). Demographic processes, such as mortality, fertility rate, and migration, are classic examples of bottom-up processes, because decisions are made by certain individuals, and overall indicators are formed as a result of these individuals' aggregated actions. Therefore, demography is one of the areas of application of the agent-based approach, which is widely represented in the literature.

It is necessary to mention the most peculiar and recently published works.

Agent-based models of social interactions and demographic behavior [2–3], which represent different components of demographic system, such as creation of marriages, changes of fertility, etc. These works also investigate differences of people's behavior, related to their identification with different types of culture, and corresponding differences of reproductive behavior.

Artificial population models [4–5] review ABMs with agents of complex structure and higher number of states, which allow forecasting demographic dynamics on different levels – from households to Great Britain's population.

The most complete description of ABMs usage for simulating demographic processes, from the formation of married couples and the impact of social norms on fertility to people's decisions to change a place of residence, is presented in a book ***"Agent-Based Computational Demography: Using Simulation to Improve Our Understanding of Demographic Behavior"*** [6].

It is obvious that complete simulation of demographic processes on the country-scale requires creation of large population of agents in the model. It transforms ABM into much-agent (multi-agent) system [7] and requires usage of high-performance equipment and specialized software for calculations.

The article [1] included specific features of multi-agent models, used for imitation of dynamics of large socio-economic systems, which are important due to the organization of their work's parallelization:

• population agents are linked to a particular region (jurisdiction) and coordinates on the map of a region and a country (have a place of residence);

• the number and spatial distribution of the agent population changes during the simulation period, because, in the course of a computer experiment, agents can migrate independently from region to region, be destroyed, and create new agents;

• agents have "social connections" with other agents, with whom they exchange messages with varying intensity, and these connections can be dynamically formed by agents from different jurisdictions during the simulation period;

• agents' place of residence and their connection with other agents significantly affect agents' behavior in the simulation of different processes (selection of residence place, family formation, birth of children, etc.).

The article, on the example of the Russia's demographic agent-based model, justifies selection of methods for programming multi-agent ABM, which allow implementing author's approach to creating quite realistic simulations of demographic processes, well-scaled for agents' population of the size equal to our country's population.

In the last few years, specialized software environments for building agent models with the function of automatic, or semi-automatic, parallelization of program code with subsequent launch on supercomputers have appeared.

Repast for High Performance Computing (RepastHPC) is the most popular computer software (CS), made for ABM design with the aim of the implementation in high-performance environments [8].

High-performance CS for building large-scale ABM – Pandora – developed in Barcelona Supercomputing Centre and gives full support of geographic information systems (GIS), which is important in cases, when geographical attachment of agents is required for models' functioning [9].

Environment of building agent-based models ABM++, the first version of which appeared in 2009 as a result of the tool's modernization, was developed in 1990–2005 at the Los Alamos National Lab in the process of building large-scale ABM [10].

SWAGES – extensible distributed environment for large-scale agent-based modeling – developed by scientists from Tufts University (Medford, Massachusetts, USA), provides automatic parallelization of program code, supports multiple programming languages and plug-ins for visualization, statistical analysis, and automatic error handling [11].

CyberGIS Toolkit – a set of freely distributed open source software components for spatial analysis and modeling [12].

HPABM – hierarchical parallel modeling environment designed for developing complex agent models for investigating large-scale problems related to geospatial modeling [13].

D-MASON package allows implementing ABM in a distributed environment, increasing their performance, while ensuring backward compatibility with basic MASON environment. D-MASON operation is based on master/slave paradigm: the main application divides the simulated space into parts and distributes the workload across slave processes, each of which uses one or more Logical Processors (LP), according to their computing capabilities. Main tasks, solved with D-MASON, are: distribution of performed work, load balancing, and communication between processes, synchronization, and reproducibility [14].

POLARIS – CS for building ABM for simulations of transport flows was developed in the Argonne National Laboratory (national research center of the United States Department of Energy). Main utilities of the developed package are: (1) a module responsible for parallel processing of events; (2) a module that implements inter-process exchange; (3) a library for visualization; (4) a library for data input and output, etc. By its nature, POLARIS is a set of low-level libraries that provide a user-friendly programming interface and a runtime environment that makes it easier to write code. Transport simulators, developed with POLARIS, use a large amount of data, which is often processed simultaneously by different software modules. In this regard, directives for programming multithreaded applications with shared memory are used in parallel mode [15].

Researchers from The Autonomous University of Barcelona developed an instrument for ABM parallelization – ***Care HPS*** (High Performance Simulation), which allows automatic solution of several tasks: the distribution of executed code, balance of the computational load, communication and synchronization.

Care HPS supports MPI message transfer interface, OpenMP technology and contains several components implemented in C++. Users solve the task of designing the model (including usage of ready-made functional controls), and all the work on distributing agents across processors, synchronizing processes is performed by the framework. Currently, Care HPS developers use this framework in a project aimed at predicting the spread of dengue fever [16].

The number of specialized software products for the implementation of agent-based models on supercomputers grows. In addition to the described ones, we should also mention MUSE [17], LUNES [18], MASS [19], and others.

However, all of these tools were initially developed for solving specialized problems. Although they have expanded significantly to the level of broad profile frameworks, they have nevertheless retained a certain “legacy” in the form of ineffective algorithms for solving other problems.

Practical experience of their usage has shown that, in the end, a less time-consuming way to build a specific agent model for subsequent launch on a supercomputer is its initial implementation from “scratches”.

Thus, it is possible to significantly increase the efficiency of program code parallelization and, in addition, create a reserve for further complication with minimal performance losses.

Software implementation of the supercomputer demographic ABM of Russia

Taking into account conclusions from the analysis of existing software tools for parallelizing software code, we have developed CS that provides:

1. Scalability of ABM operation on a set of nodes in a computing cluster.
2. Efficiency of multiprocessor operation of ABM due to the use of MPI system library, which is usually installed on each cluster, configured on its network system, and provides maximum throughput and minimum data transfer delays. Another feature of this library is also important: it is available for usage on regular personal computers, which allows using one technology without special modification while developing and performing calculations on a personal computer and supercomputers.
3. Simulation of the evolution of the internal state of agents, the formation of permanent and temporary links between them, the exchange of messages, and the appearance or disappearance of agents in the system.
4. Dynamic load balancing mechanism between computing nodes during the model operation, i.e. transferring agents from more loaded processors to another ones while preserving their state and relationships (an imbalance occurs during the model operation due to uneven disappearance of agents tied to different processors, as well as the appearance of new agents).
5. Simplification of ABM engineering and its further development by separating blocks that implement a system of parallelization with load balancing between computing processes and messaging between agents, built in C++, from thematic blocks that implement the simulation of demographic processes, which are written in the high-level programming language C#.

The work [1] showed how a rectangular grid, which binds each cell to a specific pixel in the image, is built in the model, constructed on the basis of source data on the number of regional population, received from a CSV-table, and the information on the geometry of regions, obtained from the map of Russia in raster format. Since each region is painted a different color, while analyzing this image, the connection of grid's cells with numbers of regions and other characteristics was installed. After that, the random distribution of several agent-residents among grid cells with specified total number of agents in the system and region was performed. Besides, the calculation of the decomposition of grid cells by processors was completed.

To calculate the grid decomposition, we used METIS graph algorithm [20] with weights (metis_partgraphrecursive variant). METIS algorithm takes a graph, specified through a matrix of links in the CSR format [21], an array of weights of the graph nodes, and returns the optimal distribution of the graph for a given number of parts with minimizing links between them. With this algorithm, the distribution of the source system by processors was obtained. It is important that this distribution must be made before creating agents, since the latter should initially be correctly distributed among cluster nodes. At the same time, the number of agent populations and the number of processors used are model parameters values of which are set by a user during computer experiments.

Algorithms, which implement these processes in the model, are shown in the article [1], and, in this paper, it is necessary to elaborate on the methods for ensuring the credibility of created artificial analogue of the simulated system.

In this regard, the first necessary thing is to set the starting conditions according to available official statistics for the base year. The second step is a fairly realistic simulation of population reproduction processes. In this paper, we review the simulation of processes of population's fertility and mortality. Simulation of population migration will require inclusion of additional source information and addition of appropriate procedures, the ideology of which was presented in the article [22]: this, however, will not affect the parallelization of ABM functioning.

The following Rosstat data for the base year were used as the initial information for the model.

- On the country's level:

– population's distribution by gender and age (age-sex pyramid), thousand people;

– mortality rates (per 1.000 people) differentiated by gender and age;

– retirement age for men and women by years of transition period, according to the 2018 pension reform.

- On the regional level:

– population, thousand people;

– share of population younger than working age, %;

– share of able-bodied population, %;

– share of population older than working age, %;

– total fertility coefficient;

– distribution of births by mothers' age (share of births by mothers from cohorts in five-year age intervals within reproductive age range: 15–19; 20–24; 25–29; 30–34; 35–39; 40–44 and 45–49), in %.

Setup of starting conditions corresponding to the available official statistics

At the beginning of the model's work, after reading the initial data, scaling the specified number of agents by regions, and creating the calculated number of agents in cells, it is necessary to determine the values of its individual features, associated with the simulated population reproduction processes, for each agent. These features, according to used simulation algorithms, are: agent's age, gender, maximum desired number of children in the family, and the number of children born. In addition, the agent "remembers" its family ties, because they are provided by its individual collections (lists): a collection of parents; children; siblings; other relatives.

The distribution of agents' age and gender values is implemented for accurate reproduction of the population's age and gender structure, specified in the initial data for the country and individual regions. In order to do this, obtained values of the agents' number in each region are further scaled:

a) by shares of population's main age groups in each region: younger than able-bodied; able-bodied and older population (considering the set values for the base year of women and men's retirement age), and then

b) by shares of each age cohort in its age group.

Obtained values of the shares from the total number of agents in the region are used as the probability of falling out of a particular age for an agent belonging to this region. In order to perform this scaling and get the age value for each agent, a specialized auxiliary module was developed. The gender of the agent is also determined in a probabilistic way, considering the sex ratio for the acquired age cohort.

In the model, the maximum desired number of children in a family is a random variable that takes a value from one to seven with the specified beta distribution shifted to the left (the maximum is for two children). The specialized auxiliary module was also developed to determine the specific value of the desired number of children for each agent.

After the distribution of gender and age properties, the establishment of kinship relations between agents begins. First of all, for each agent from the agents' collection of the same region, a "mother" is selected – a female agent with a randomly determined age and with the number of children below the maximum desired number. The selection of the mother-agent's age is based on the usage of the births' distribution according to the mothers' age specified in the source data. After it, the number of children of the selected mother-agent increases, and the mutual entry of new relatives into the corresponding collections, conducted by the child-agent, the mother-agent, and the mother's relative-agents, happens.

Simulation of population reproduction processes at each step of the model operation

Example, given in the work [1], was developed for assessing the influence of the number of used processors on the efficiency of the model's work parallelization. Another parameter of the test model, impact of which on the model's performance was evaluated during the experiments, was the number of agents' connections, with whom they exchanged messages at each step. These connections were established randomly. In addition, the composition of the agent population was constant.

In the full version of the demographic ABM of Russia, presented in this paper, connections between agents are established on the basis of kinship, and there are significantly less of them than in the test example: it reduces the number of messages sent at the step. On the other hand, in a simulation of processes of population reproduction, two things are considered: a) dynamic change of agents' population composition: deletion of agents (their "death") and the emergence of new ("birth" of agents); b) division of step actions into stages: there might be a need for overviewing common parameters, related to regions or groups of agents, and/or messages' exchange between agents. The presence of such stages means appearance of step synchronization points. It implies work suspension of processors which had already finished processing of agents, placed on them, and expect all other processors to finish work.

To implement dynamic addition and removal of agents in the system, it was necessary to switch from a single (end-to-end) index of agents to a double one. Now each agent is characterized by the number of the cell, in which it is located, and its number in this cell. The process of adding agents leads to the situation, when the counter of the agents' number grows in a cell, and its value is used for calculating agents' index. In the process of the agent's deletion, there is a recalculation of indices of other agents within its cell. A deleted agent should be excluded from collections of all agents with which it was bounded by family ties: this is why it sends messages to them.

As a result, it was necessary to stop using regional collections of age cohorts for organizing the simulation of step actions, since it was very time-consuming to keep them updated while deleting agents and switching to a double index. The rejection of collections led to a change in the fertility rate simulation algorithms.

For example, instead of selecting a female agent, who should give birth to a child, from a collection of women of a given age (which was done in order to reproduce the observed distribution of births by the mother's age), the corresponding age probabilities of child births were calculated. These probabilities were calculated for each region on the basis of the total number of women of reproductive age, the regional total fertility rate, the distribution of births by mother's age, and the number of women from each age group for whom the number of children born has not reached the desired maximum.

As the result, actions at a step of work were divided into stages, implementing phases of demographic processes simulation. Each stage may relate to one of the following types of objects: the entire country, a separate region, a separate agent. The model implements the following stages related to different stages:

1. ***Stage and preparation phase:***

• Transition to the next year. Reset of counters, associated with the simulation of the fertility, on the level of the whole country and regions.

2. ***Stage of agents' extinction:***

• ***Phase of a "black mark" distribution (on the agent level):*** agent's marking in case of a "black mark" (death), given to it in accordance with the age-sex mortality coefficient. Otherwise, an increase of the agent's age by one.

• ***Phase of recalculation of agents' indices (on the country level):*** calculation of new indices for surviving agents in cells, where agent deletions are scheduled.

• ***Phase of indices' replacement (on the agent level):*** sending of an updated agent index to all associated agents; the old index is replaced with a new one in all collections.

• ***Phase of agents' removal (on the country level):*** removal of marked agents from collections.

3. ***Stage of new agents' creation:***

• ***Phase of calculating birth probabilities (on the country level):*** calculation of birth probabilities for women of different ages (by region);

• ***Phase of agents' birth (on the agent level):*** creation of a subsidiary agent in case of a childbirth event. Setup of kinship relationships.

4. ***Stage of simulation step completion:***

• ***Phase of statistics' collection (on the country level):*** collection of regional and country statistics and preservation of results in external files.

At the end of each phase, the message queue is checked. If it is not empty, messages are sent to recipients.

Analysis of the experiments' results using the demographic ABM of Russia

1. Analysis of model's initial state compliance with the original data, and the stability of recreating the specified demographic parameters

The models' parameters are the number of agent populations and the number of processors used. The initial distribution of agents by region and processors, as well as the assignment of age and gender to individual agents, is carried out using probability distributions, the stability of which depends on the number of tests. Therefore, it was necessary to assess the stability of the distribution of agent features depending on the values of the model's parameters

Table 1 presents the comparison of the results of agents' features distribution, gained during model experiments, with assigned distribution of population by gender and age (model data on agents' scaling distribution is changed into the number of population in thousands of people).

Table 1. Analysis of adequacy and stability of the reconstruction in the model of the age-sex pyramid depending on the number of agents

Number of agents, million	Mean-squire deviation from the specified age distribution, thousand people		Deviation from experimental average values by age and gender, %
	Men	Women	
1	11.0	12.0	±2.7
2	8.0	8.4	±1.4
3	7.6	7.5	±1.1
4	6.3	6.6	±0.8
Source: calculated according to the results of computer experiments with the demographic ABM of Russia.			

Rosstat data was used as the initial information[1]. The experiments were performed on a personal computer (for a single processor) in order to estimate the minimum size of the population of agents required for an acceptable match with the age-sex pyramid of the Russian population specified in the initial data. In addition, *Table 1* provides data on the stability of the obtained gender and age distribution of agents, estimated by deviation from the experimental average values.

Data shows that there is a significant improvement of analyzed statistical indicators after increase of agents' number from one to two million. Thus, the standard deviation from the specified age distribution decreased in 1.38 times for men and 1.43 times for women, and the indicator characterizing the stability of the distribution, obtained during the experiments, improved in 1.93 times. With a further increase of agents' population to three and four million, these indicators continued to improve, but not so rapidly. As the result, when the number of agents was four million, the standard deviation decreased in almost two times, in comparison with the basic (1.75 and 1.82 times for men and women, respectively), and the deviation from the average experimental values – in more than three times (in 3.38 times). The spread of the latter indicator's ±1.4% values, which is achieved when the number of agents is two million, may be considered acceptable. It is exactly the number of agents which might be considered minimum for obtaining relevant results of experiments on the country level.

If we estimate the necessary number of agents on the regional level, we should proceed from the number of agents obtained for the most sparsely populated regions of the Russian Federation. Thus, the population of the Magadan Oblast is only 144.1 thousand people, and, with two million agents in the model, this region accounts for less than two thousand agents. This number is clearly not enough to ensure the stability of the distribution of features of this region's agents-residents. For example, the stability of the beta distribution, which determines the desired number of children for agents and establishes the relationship between mothers and children, is achieved when the number of agents is 10 thousand. Based on these considerations, the requirements for the total number of agents should be at least five times higher.

However, all subsequent experiments were conducted to assess the sustainability of indicators on the country level.

2. Stability analysis of models' functioning

The next series of experiments was a simulation of reproduction processes in Russian population for twenty steps (years) ahead

[1] The Demographic Yearbook of Russia. 2017: stat. coll. Rosstat. Moscow, 2017.

Table 2. Analysis of the stability of the resulting model indicators over entire forecast period

Indicator	Deviation from the experimental average values for the forecast period years, %	
	lower limit	upper limit
Population, thousand people	-0.07	0.06
Share of able-bodied population, %	-0.12	0.12
Total fertility coefficient per 1,000 people	-2.19	1.57
Total mortality coefficient per 1,000 people	-1.02	0.88
Source: calculated according to the results of computer experiments with the demographic ABM of Russia.		

with a population of two million agents and constant values of mortality and fertility. For Russia, forecasts of the dynamics of the following indicators were obtained:

- population, thousand people;
- distribution of population by major age groups, %;
- total fertility rate per 1000 people;
- total mortality rate per 1000 people.

The purpose of the experiments was to evaluate the stability of obtained simulation results. *Table 2*, which shows the boundaries of the variances of key output variables from the average values by years of the forecast period, allow assessing stability of these indicators (maximum and minimum values were indicated throughout the forecast period).

The table shows that indicators, such as the population size and the share of able-bodied population, demonstrate remarkable stability with a variation of the values, obtained in the course of experiments, around ±0.1%. The next largest deviation is the total mortality rate (deviation around ±1%), and the largest variation is the total fertility rate (±2%). Thus, even if the number of agents is two million, the stability of the model may be considered satisfactory.

3. Analysis of model's parallelization efficiency

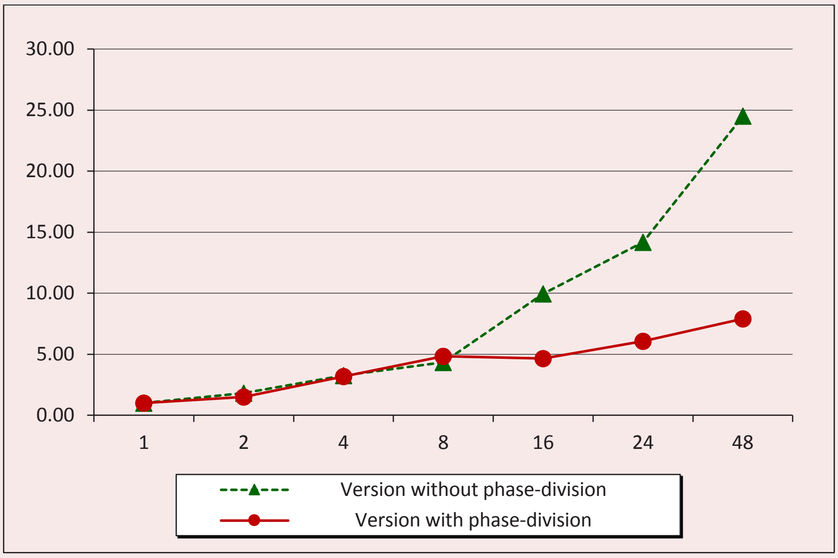
The next stage of testing the demographic ABM was evaluation of the effectiveness of parallelizing of its operation on a super-computer. In the test example, ten steps were calculated, and the average time value for a step was considered. *Table 3* shows the results of running the model on a different number of processors for two and eight million agents. *Figure 1* shows the corresponding acceleration curves of the model depending on the number of processors, where the number of processors on the abscissa axis increases by degrees of two.

It is interesting to compare the obtained dependence of the model's acceleration on the number of processors with the curves of the well-known Amdahl's law [23], which

Table 3. Estimation of acceleration in parallel calculation depending on the number of processors

Number of processors	Calculation time for 2M, s	Acceleration for 2M, times	Calculation time for 8M, s	Acceleration for 8M, times
1	3.029	1.00	17.211	1.00
2	2.016	1.50	11.064	1.55
4	0.948	3.19	7.053	2.44
8	0.625	4.84	5.215	3.30
16	0.653	4.64	2.962	5.81
24	0.499	6.07	2.245	7.66
32	0.476	6.35	1.019	16.88
48	0.383	7.90	1.395	12.33
Source: calculated according to the results of computer experiments with the demographic ABM of Russia.				

Figure 1. Comparison of acceleration in parallel calculation depending on the number of processors with different number of agents



Source: calculated according to the results of computer experiments with the demographic ABM of Russia.

connects maximum achievable acceleration with the number of processors and the share of parallel calculations. Thus, we may assess how well the algorithms, used in the model to simulate the behavior of agents, are configured for parallelization of its work. A comparison with the theoretically achievable level of acceleration shows that, for two million agents, the reduction of the parallelization efficiency is comparable to the situation, when the share of sequential calculations would be about 10%, for eight million agents – as if it would be less than 5%.

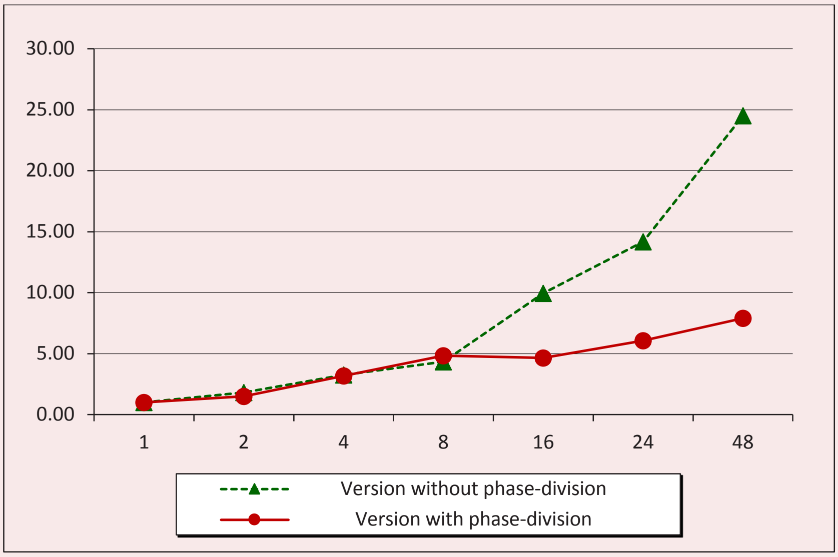
4. Comparison of effectiveness indicators of model parallelization with synchronization points at the simulation step with the test example

The decrease of the parallelization efficiency is largely caused by the presence of intermediate synchronization points at each step of the model's functioning. Therefore, it was of great interest to find out, how much the parallelization performance of the model deteriorated in comparison with the first version [1] due to the appearance of stages, performed at each step. A comparison of two versions is shown in *Figure 2*.

The figure shows that an increase of the number of processors up to eight provides almost the same acceleration of two versions of the model, but a further increase of the number of processors reveals a significant difference of the parallelization effectiveness. So, for 16 processors, the acceleration of the model with additional synchronization points is half as fast as for the first version of the model (2.1 times) and for 48 processors – 3.1 times lower.

While evaluating the adequacy and effectiveness of the author's approach to the organization of the model's parallelization, it is necessary to compare it with closest analogues. For example, in [24], a distributed ABM of epidemics is described: within it, it is

Figure 2. Comparison of acceleration of the model functioning with parallel calculation depending on the number of processors for two versions with the number of agents 2 million



Source: calculated according to the results of computer experiments with the demographic ABM of Russia.

possible to create a society with the number of agents up to 6 billion. The model, as well as the one presented by the authors, uses an algorithm for distributing agents between computing nodes, which reduces intergroup interaction. In [25], an approach to parallelization of resource-intensive ABM is reviewed, agents of which exchange information and have spatial attachment. The acceleration, obtained by the authors in comparison with sequential version of the model, was, on average, 20 times. In our case, in the test example, the acceleration was up to 25 times with 48 processors, but addition of synchronization points significantly reduces this number. However, the total calculation time, as shown in *Table 3*, is quite acceptable.

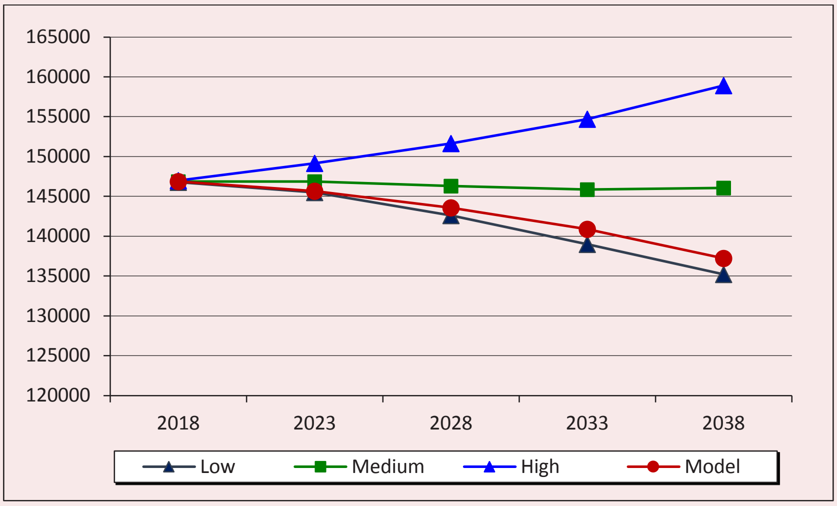
5. Forecast results and comparison of received forecast with the Rosstat prediction

After making sure of the adequacy of the initial state of our ABM and the stability of its operation, we now evaluate the results of the simulation from a meaningful point of view. Thus, in the course of experiments with the model, forecasts of demographic characteristics were obtained in the context of regions and the country. In the same 2017 Demographic Yearbook of Russia[2], the variants of forecast (low, medium, and high) of the following main characteristics of country's population up until 2051 are given:

- population number;
- number of men and women;
- population number by separate age groups;
- births, deaths, and natural population growth;
- total fertility coefficient;
- life expectancy at birth.

[2] The Demographic Yearbook of Russia. 2017: stat. coll. Rosstat. Moscow, 2017.

Figure 3. Comparison of forecast variants. Population, thousand people



Source: calculated according to the results of computer experiments with the demographic ABM of Russia and data from 2017 Demographic Yearbook of Russia: Tab. 8.1.

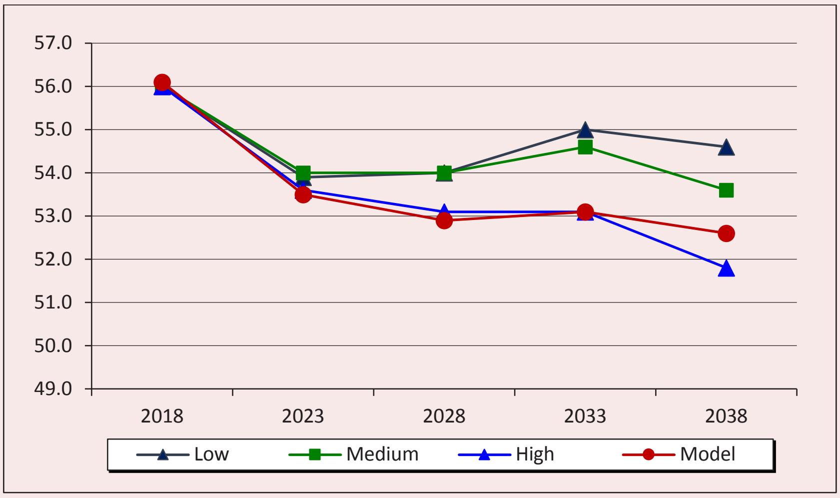
Due to the fact that mortality rates in the experiments were considered unchanged, as well as the total fertility rate, on the basis of which the birth rate is simulated, it is obvious that the forecast of these indicators does not make sense in comparison with forecasts of Rosstat. The same applies to life expectancy at birth indicator, which is calculated on the basis of values of mortality indicators, differentiated by gender and age.

However, it is possible and interesting to compare forecasts on the number of Russian population and its structure within individual age groups obtained through model experiments. Figures 3 and 4 show a comparison of four variants of the forecast up to 2038: the low, medium, and high versions of the Rosstat forecast, as well as the forecast obtained as a result of ABM work (average values of several experiments with the population of two million agents were taken). *Figure 3* shows that the model population forecast for the entire period slightly exceeds the low version of Rosstat (at the end of the period – by 1782.6 thousand people, or 1.3% of Russian population), while the model differs significantly from the average and high versions (by -6.2 and -13.8%, respectively).

Figure 4 shows the comparison of the model forecast of the share of able-bodied population with the corresponding versions of the Rosstat forecast. It can be seen that, during the entire period, the model forecast almost coincides with the high version of Rosstat and, at the end of the period, exceeds it by only 0.8%. The largest deviation of the model forecast was from the low version of Rosstat, but it was only -2%. In other words, even considering the assumption that the death rate and the total birth rate remain unchanged, the model allows us to obtain a fairly realistic forecast of this indicator.

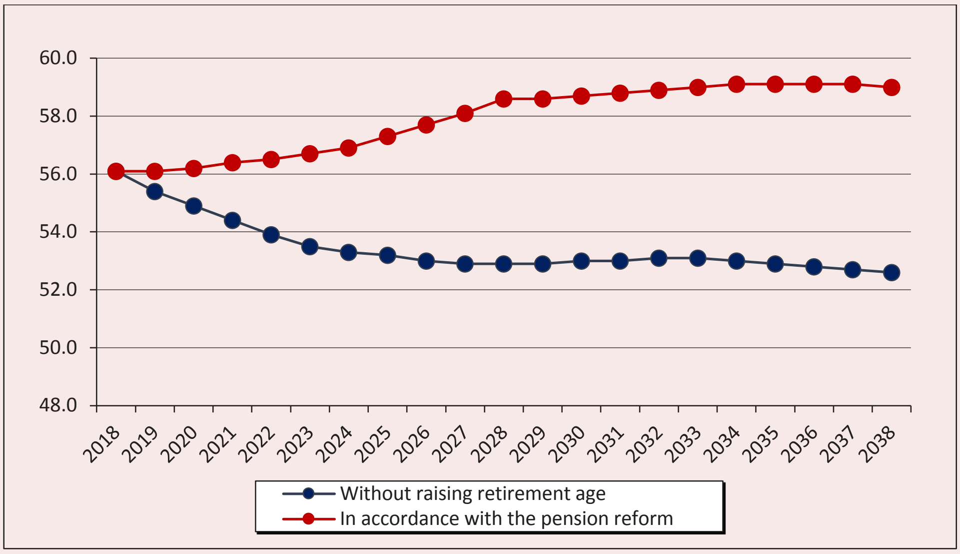
It should be noted that this forecast, as well as Rosstat forecasts, was made under the assumption that the retirement age remains unchanged – 55 years for women and

Figure 4. Comparison of forecast variants. Share of able-bodied population, %



Source: calculated according to the results of computer experiments with the demographic ABM of Russia and data from 2017 Demographic Yearbook of Russia.

Figure 5. Comparison of forecast variants. Share of able-bodied population, %



Source: calculated according to the results of computer experiments with the demographic ABM of Russia.

60 years for men. However, in 2018, the situation changed after the adoption of the pension reform, which was designed to gradually raise the retirement age over the next ten years. The forecast should take this into account. Therefore, in the next series of model experiments, the retirement age of agents was raised in accordance with the pension reform. The obtained results are presented in *Figure 5*, where two variants of model forecasts of the share of able-bodied population are compared: without raised retirement age and under the conditions of pension reform.

The figure shows that the share of able-bodied population, during the ten-year period of the reform, steadily decreases in the first version (from 56.1 to 52.9%) and grows in the second one (to 58.6%). As a result, at 2028 mark, the difference was 5.7% (8184.1 thousand people). At the end of the entire forecast period, the share of able-bodied population decreased to 52.6% in the first version and increased to 59.0% in the second one. The difference between them was 6.4% (9189.2 thousand people).

Conclusions

The analysis of the results of testing the presented demographic Russian ABM allowed us to draw the following conclusions.

The model showed a high degree of stability in the course of a comprehensive test despite the widespread usage of probabilistic mechanisms in setting the initial state of the system and simulating processes of population reproduction.

The technologies used in the model for parallelizing its operation on a set of computing nodes of a supercomputer allow conducting simulations with a large number of agents and achieve acceptable indicators of parallel effectiveness.

Presented demographic model is designed as a basic platform for a comprehensive regional ABM, because the mechanism of dividing the simulation steps into stages allows connecting blocks which simulate any socio-economic processes people participate in. Moreover, the implemented mechanism also allows specifying the type of objects the stage (phase) belongs to. Thus, the model is configured to introduce other types of objects into the artificial environment (for example, enterprises, municipalities, etc.) with the appropriate addition of simulated processes (social, environmental, economic, political), in which these objects participate. However, the demographic model also has an independent value as a tool that allows getting forecasts of demographic characteristics in the course of experiments within regions and the country.

References

1. Makarov V.L., Bakhtizin A.R., Sushko E.D., Sushko G.B. Development of an agent-based demographic model of Russia and its supercomputer implementation. *Vychislitel'nye metody i programmirovanie=Computational methods and programming*, 2018, vol. 19, pp. 368–378. DOI: 10.26089/NumMet.v19r433. (In Russian).
2. Billari F.C., Prskawetz A., Diaz B.A., Fent T. The "Wedding-Ring": an agent-based marriage model based on social interaction. *Demographic Research*, 2007, vol. 17, article 3, pp. 59–82.
3. Diaz B.A. *Agent-Based Models on Social Interaction and Demographic Behaviour (Ph.D. Thesis)*. Wien: Technische Universität, 2010. 93 p.
4. Silverman E., Bijak J., Hilton J., Cao V.D., Noble J. When demography met social simulation: a tale of two modelling approaches. *Journal of Artificial Societies and Social Simulation (JASSS)*, 2013, vol. 16 (4), article 9. Available at: http://jasss.soc.surrey.ac.uk/16/4/9.html.
5. Silverman E., Bijak J., Noble J., Cao V., Hilton J. Semi-artificial models of populations: connecting demography with agent-based modelling. In: Chen S.-H. et al. (Eds.). *Advances in Computational Social Science. Agent-Based Social Systems. Vol. 11.* Tokyo: Springer Japan, 2014. Pp. 177–189. DOI: 10.1007/978-4-431-54847-8_12.
6. Billari F.C., Prskawetz A. (Eds.). *Agent-Based Computational Demography: Using Simulation to Improve Our Understanding of Demographic Behaviour*. Heidelberg: Springer – Verlag, 2003. 210 p.
7. Tarasov V.B. *Ot mnogoagentnykh sistem k intellektual'nym organizatsiyam: filosofiya, psikhologiya, informatika* [From multi-agent systems to intellectual organizations: philosophy, psychology, computer science]. Moscow: Editorial URSS, 2002. 352 p.

8. Collier N., North M. Parallel agent-based simulation with Repast for High Performance Computing. *Simulation*, 2012, vol. 89, no. 10, pp. 1215–1235. DOI: 10.1177/0037549712462620.
9. Wittek P., Rubio-Campillo X. Scalable agent-based modelling with cloud HPC resources for social simulations. In: *IEEE 4th International Conference on Cloud Computing Technology and Science (CloudCom). December 3-6, 2012, Taipei, Taiwan*. Pp. 355–362.
10. Roberts D.J., Simoni D.A., Eubank S. A National scale microsimulation of disease outbreaks. *Advances in Disease Surveillance*, 2007, vol. 4, no. 15.
11. Scheutz M., Connaughton R., Dingler A., Schermerhorn P. SWAGES – an extendable distributed experimentation system for large-scale agent-based alife simulations. In: *Proceedings of Artificial Life X*, 2006, pp. 412–419. Available at: https://hrilab.tufts.edu/publications/scheutzetal06alifeswages.pdf .
12. Shaowen W., Yan L., Anand P. Open cyberGIS software for geospatial research and education in the big data era. *SoftwareX*, 2015, no. 5. DOI: 10.1016/j.softx.2015.10.003.
13. Tang W., Wang S. HPABM: A hierarchical parallel simulation framework for spatially-explicit agent-based models. *Transactions in GIS*, 2009, no. 13 (3), pp. 315–333.
14. Cordasco G., Scarano V., Spagnuolo C. Distributed MASON: A scalable distributed multi-agent simulation environment. *Simulation Modelling Practice and Theory*, 2018, vol. 89, pp. 15–34. DOI: 10.1016/j.simpat.2018.09.002.
15. Auld J., Hope M., Ley H., Sokolov V., Xua B., Zhang K. POLARIS: Agent-based modeling framework development and implementation for integrated travel demand and network and operations simulations. *Transportation Research Part C: Emerging Technologies*, 2016, vol. 64, pp. 101–116.
16. Borges F., Gutierrez-Milla A., Luque E., Suppi R. Care HPS: A high performance simulation tool for parallel and distributed agent-based modeling. *Future Generation Computer Systems*, 2017, vol. 68, pp. 59–73.
17. Gebre M.R. MUSE: *A parallel agent-based simulation environment (Doctoral Thesis)*. Oxford, Ohio: Miami University, 2009. 99 p.
18. D'Angelo G., Ferretti S. LUNES: Agent-based simulation of P2P systems. In: *Proceedings of 2011 IEEE International Conference on High Performance Computing & Simulation*, Istanbul, Turkey, July 2011. Pp. 593-599. DOI: 10.1109/HPCSim.2011.5999879.
19. Emau J., Chuang T., Fukuda M. A multi-process library for multi-agent and spatial simulation. In: *Proceedings of 2011 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing – PACRIM'11, Victoria, BC, Canada, August 24–26, 2011*. Pp. 369–376.
20. Karypis G., Kumar V. *METIS-unstructured graph partitioning and sparse matrix ordering system, version 2.0*. Available at: http://dm.kaist.ac.kr/kse625/resources/metis.pdf.
21. Tinney W., Walker J. Direct solutions of sparse network equations by optimally ordered triangular factorization. *Proceedings of the IEEE*, 1967, no. 55 (11), pp. 1801–1809.
22. Makarov V.L., Bakhtizin A.R., Sushko E.D., Ageeva A.F. Artificial society and real demographic processes. *Ekonomika i matematicheskie metody=Economics and the Mathematical Methods*, 2017, vol. 53, no. 1, pp. 3–18. (In Russian).
23. Amdahl G.M. Validity of the single processor approach to achieving large scale computing capabilities. In: *AFIPS Conference Proceedings,* 1967*,* vol. 30, pp. 483–485,
24. Parker J. A flexible, large-scale, distributed agent based epidemic model. In: Henderson S.G., Biller B., Hsieh M.-H., Shortle J., Tew J.D., Barton R.R. (Eds.). *Proceedings of the 2007 Winter Simulation Conference*. Washington, D.C. December, 2007. Available at: https://www.brookings.edu/wp-content/uploads/2016/06/12_epidemicmodel_parker.pdf.
25. Gong Z., Tang W., Bennett D.A., Thill J.C. Parallel agent-based simulation of individual-level spatial interactions within a multicore computing environment. *International Journal of Geographical Information Science*, 2013, vol. 27, no. 6, pp. 1152–1170.

Information about the Authors

Valerii L. Makarov – RAS Academician, Doctor of Sciences (Physics and Mathematics), Scientific Director, Central Economic Mathematical Institute of the Russian Academy of Sciences (47, Nakhimovsky Avenue, Moscow, 117418, Russian Federation; e-mail: makarov@cemi.rssi.ru)

Al'bert R. Bakhtizin – RAS Corresponding Member, Doctor of Sciences (Economics), Director, Central Economic Mathematical Institute of the Russian Academy of Sciences (47, Nakhimovsky Avenue, Moscow, 117418, Russian Federation; e-mail: albert.bakhtizin@gmail.com)

Elena D. Sushko – Candidate of Sciences (Economics), Leading Researcher, Central Economic Mathematical Institute of the Russian Academy of Sciences (47, Nakhimovsky Avenue, Moscow, 117418, Russian Federation; e-mail: sushko_e@mail.ru)

Gennadii B. Sushko – Candidate of Sciences (Physics and Mathematics), Senior Researcher, Central Economic Mathematical Institute of the Russian Academy of Sciences (47, Nakhimovsky Avenue, Moscow, 117418, Russian Federation; e-mail: gennady.sushko@gmail.com)

Received November 12, 2019.