Migration Processes Modeling
with Cellular Automation

Shmidt Yu. D., Ivashina N.V., Ozerova G.P., Lobodin P.N.

Far Eastern Federal University, Vladivostok, Russia
{`syd, ivashina.nv, ozerova.gp,lobodin_pn`}@dvfu.ru

Abstract. The problem of interregional migration flows modeling is being studied. The conjecture that modeling household migration behavior at local level in view of community interactions and using cellular automata results in adequate forecasting of interregional migration flows has been numerically confirmed. A computer program was developed in order to implement the cellular automaton suggested in this study, which models interregional migration flows. Cellular automaton program was tested on Primorsky krai migration flows statistical data; a short-term forecast of the region migration flows was obtained.

Keywords: Modelling, migration, cellular automaton, regions, forecasting, migration flow, software package.

1 Introduction

Migration processes are complex dynamical phenomena influencing the population reproduction and its distribution across the territories. They are affected by many socioeconomic factors having stochastic and probabilistic character. Migration makes a substantial impact on the demographic structure, regional and local labor markets. Therefore, forecasting quantitative and qualitative migration indices and their consequences is a very important and critical task. In order to carry it out modern mathematical tools are needed because these processes have a dynamic character at different time intervals and nonlinear character in terms of interrelation of factors.

There is a relatively successful experience of Russian and foreign scholars in using econometric models to describe migration processes [1], [2], [3], [4], [5], [6], [7]. Non-linear dynamical systems are used for population migration modeling along with traditional econometric techniques [8], [9], [10].

In order to model population movement a theory of stochastic processes is used quite successfully. In the work [11] and the majority of succeeding papers on the subject Markov chains were used as a model for population movement. These were Markov processes in discrete time. Migration processes are dynamic; they change dramatically in time reacting to the changes of external environment and population living conditions. This fact narrows down considerably the potential of using Markov models

for their modeling on long time intervals because transition probabilities in almost all Markov models are assumed to be constant.

In the present work it is proposed to use cellular automata for modeling interregional migration flows. The theory of cellular automata has a relatively short but reasonably productive history, the main stages of which are outlined in [12], [13], [14], [15].

In cellular automaton model each cell changes its state as a result of interaction with a limited number of cells, as a rule with their immediate surroundings. However, a simultaneous change of the state of all cells, i.e. of the entire lattice, is possible on the basis of cellular automaton general rule. This characteristic enables to relate the processes happening at a micro-level with the processes happening at the macro-level while modeling. This is a crucially important characteristic of cellular automata, which enables their successful use for modeling systems where three-dimensional interaction between the elements plays a significant role. They also proved useful in modeling liquid and gas dynamics in various environments, in modeling systems containing a great number of particles non-linearly interacting with each other, in describing the emergence of collective phenomena such as turbulence, order, chaos, symmetry and fractality violations, etc [16].

In recent years cellular automata found applications in modeling social phenomena both as purely theoretical tools for qualitative analysis and for numerical forecasting. Some examples of cellular automata used in sociological problems and urban migration have been introduced in [17], [18], [19], [20]. The results of research in the field of urban migration carried out using cellular automata are presented in [21], [22], where the possibility of using cellular automata for analyzing non-linear dynamic interaction of households was emphasized. The possibility results from simplicity of implementation of household interaction models on micro-level as well as from the ability of such models to take account of the influence of macroeconomic conditions on the modeled processes. In particular, cellular automaton models developed in [22] include the rules, which control the decisions of households to move to a different area of the city. Each household possesses a social structure represented by a summary indicator reflecting some common characteristics of a household: average age of residents, average income, employment status, etc. The decision to move is made when there is a substantial difference between the social status of a household and an average social status of its neighborhood.

In general, cellular automata models have many positive qualities, which undoubtedly influence the intensity of their use in different branches of science. However, there are certain negative features restraining the development of this area of mathematic modeling, among those it is important to mention a weak general theoretical basis of cellular automata, insufficient convergence results for simulations using some types of cellular automata, the problems of robustness of numerical solutions.

The goal of this study is to develop a cellular automaton model that could be successfully applied for modeling interregional migration flows. The analysis of existing migration models shows that, despite a wide range of available models, most of them use only aggregated data. This leads to a certain discrepancy between the behavior of economic entities at a micro-level corresponding to empirical data and the projected figures of migration at macro- and meso-levels.

The assumption of the present study is that the modeling of household behavior with regard to their migration decisions at local levels performed by cellular automata allows to achieve adequate forecasts for interregional migration flows.

2   The cellular automaton model description

A cellular automaton is a finite collection of objects (cells) which, as a rule, form a regular lattice. In other words, without using strict mathematical terminology, this is a collection of cells joined together in a specific way to form a uniform lattice. The state of the i-th cell at time t is described by a number or a vector which we denote $a(i, t)$. The collection of states of all the cells is called the state of the lattice. In the cellular automaton model, each state of the lattice corresponds to a certain point of time that changes in discrete steps. The state of the lattice changes according to a certain law that is called the rule of the cellular automaton. The rule determines a cells state at the time $t + 1$ based on its states and the states of other cells at time t. This can be expressed mathematically by the following formula:

$$a(j, t+1) = F(a(j, t), \sum_{i \in Q(j)} a(i, t)), \tag{1}$$

where F is a function that describes the corresponding rule, $Q(j)$ is the set of cell numbers that influence the state of the cell j in time period from t to $t + 1$. Theoretically, cell automata may have any dimension. However, it is one- and two-dimensional automata that are most often used. In 2-dimensional cellular automata, the lattice is represented by a 2-dimensional array, and each cell is numbered with an ordered pair of numbers (a, b). In this case, the cells closest neighbors are either those having a common side with the initial cell (the von Neumann neighborhood), which produces 4 neighboring cells for a rectangular cell, or those having common vertices with the initial cell (the Moore neighborhood), which gives 8 neighboring cells.

The classical cellular automaton model has the following features:

– the lattice of a cellular automaton is homogeneous, i.e. the transition rules are the same for all the cells;
– the set of states is finite for each cell;
– a cell can be influenced only by cells from its neighborhood, i.e. its closest neighbors;
– the states of all the cells change simultaneously at end of an iteration.

Various criteria exist for classification of cellular automata. In particular, based on the transition rules, they can be classified as deterministic or probabilistic. In deterministic cellular automata, the state of each cell at time $t + 1$ is determined univocally by the state of this cell and its closest neighbors at time point t. The corresponding rule is represented by formula (1). A cellular automaton that determines its states based on certain transition probabilities is called probabilistic. In this case, the rule for transition from the state $a(i, t)$ of a cell i into the state $a(i, t + 1)$ can be written as follows:

$$P = P(a(j, t+1) \mid a(j, t), \sum_{i \in Q(j)} a(i, t)), \tag{2}$$

where P is the probability of the j-th element transitioning from the state $a(j,t)$ into state $a(j, t + 1)$ given certain states of its neighbors.

In a modern world, the decision of a household to move can be influenced not only by its closest neighbors, but also by randomly located agents. To model this situation one might use a cellular automaton that accounts the states of 8 or 4 randomly chosen cells on the lattice when determining the state of each cell. However, in a more realistic model both the opinions of the closest neighbors and the remote agents contribute to the migration decision. To this end, it is possible to use the combined probabilistic cellular automaton introduced in [16], where the state of each cell is determined by on the states of its 4 closest neighbors (the von Neumann neighborhood) and 4 randomly chosen cells.

When implementing the cellular automaton model, we will use a rectangular lattice for each Federal district of Russia and the neighboring countries. Each rectangular lattice contains the same number of cells as the number of households in a given region, assuming that a household consists of three people. Also, when building a cellular automaton, it is necessary to specify the rules of its behavior on the boundary of the lattice. In this paper we will use an automaton with the toric topology, i.e. the first row viewed as a continuation of the last whereas the last one precedes the first (the same applies to columns as well).

We consider probabilities of households leaving a region in 9 directions (Federal districts of Russia, the Commonwealth of Independent States (CIS) and countries outside the CIS). Denote by p_0 the probability that a household stays in a given region and by pi the probability of moving to the i-th region for permanent residence, $i = 1 \ldots 9$. Estimate these probabilities as averages over a number of years.

We will form the cells corresponding to a given region as follows. The number of cells is equal to the number of residents divided by 3, i.e. we assume that an average household totals three. We assign the initial state by sampling from the uniform distribution over integers $\{0, 1, .., 9\}$. If a cell is assigned 0, then the respective household remains to live in the region in question for the following period. If the state is i, then the household leaves for the region $i, i = 1 \ldots 9$.

Each iteration comprises the following steps:

1. Detecting the households that decide to leave the examined region.
2. Detecting the number of households that decide to arrive in the examined regions.
3. Reviewing the cells in the given region and assigning new states to them.

During the first step, browse the cells in the lattice that describes the given region. For each cell, go over its 4 neighbors and 4 randomly chosen cells in the region. Looking at their states do the following:

– For each cell, if there is a non-zero state among its neighbors that is encountered more often than any other state, then such state is assigned to this cell and it will move to the corresponding region.
– If there is no such state, calculate $p = \sum_{j}^{m} p_{i_j}$ using m neighbor cells in non-zero states and their corresponding probabilities $p_{i_1}, p_{i_2}, \ldots p_{i_m}$. Divide the interval $[0, 1]$ into subintervals of length $\frac{p_{i_j}}{p}, j = 1, m$. Generate a random number from the

interval $[0, 1]$ and choose the region of destination according to the subinterval hit by this number.

In view of the states of all the neighboring cells (8 cells) and using Laplaces function, build the transition probabilities $P(n)$:

$$W(n) = P\left(-3\sigma \leq x \leq -3\sigma + \frac{3}{4}\sigma n\right) = \Phi\left(-3 + \frac{3}{4}n\right) + \Phi(3), \qquad (3)$$

$$P(n) = \frac{W(n)}{k}, \ k = \frac{0,9973}{\sum\limits_{i=1}^{9} p_i}, \qquad (4)$$

where n is the number of neighboring cells for the current cell in non-zero state; σ is the mean square deviation; $\Phi(x)$ is Laplace's function.

This pretty much agrees with the central limit theorem of the probability theory. According to it, if a random variable represents a sum of a large number of mutually independent random variables, then it has near-normal distribution. Moreover, the three-sigma rule is valid for a normally distributed variable: if a random variable is normally distributed, then with very high probability the absolute value of its deviation from mathematical expectation does not exceed the tripled mean square deviation.

Generate a random number . If $x < P(n)$, then the household decides to leave for the selected region. Memorize the coordinates of this cell. Otherwise, the cell state will be 0, and the household will not leave.

In a similar way, on step 2 determine the households that decide to move in the region under examination, browsing the lattices for Federal districts and the CIS countries. The probability of leaving for the current region is known for each Federal district and the CIS countries and is denote g_1. All the cells of the respective lattice have the state of 0 or 1 depending on whether the household decides to stay in its region or leave for the examined region, respectively. For each cell, go over 4 neighboring and 4 randomly chosen cells in the region. In this case, the transition probability $P(n)$ is calculated as follows:

$$P(n) = \frac{W(n)}{k}, \ k = \frac{0,9973}{g_1}, \qquad (5)$$

Where n is the number of cells from the neighborhood of the current cell with state 1.

Generate a random number . If $x < P(n)$, then the household decides to leave for the region in question. Otherwise, the cell state is 0, and the household will not leave. For each lattice, calculate the number of newcomers to the region in question and compute their total.

During step 3, randomly distribute the households that have arrived to the region under examination among the cells of the households that have left. The states of such cells are determined according to the departure probabilities for the examined region for the current year. Then proceed to step 1 or finish the calculation. During the learning phase, the number of iterations is determined by the difference between the calculated value and the actual number of people who left the region in this year. During the forecasting phase the number of iterations is determined as the average number of the automatons iterations for one-year modeling over the period of 5 years during the learning phase.

3   The result of the numerical experiment

The cellular automaton described above was tested on the statistical data of Primorsky krai. The following data is available for the period from 2004 to 2014, which was used as an input for the computer program that implemented the proposed cellular automaton:

- population number of Primorsky krai, each Federal district, including the Far Eastern (without Primorsky krai) and the CIS countries;
- vector of average probabilities for migration from Primorsky krai to Federal districts and CIS countries and countries outside the CIS by years;
- vector of average probabilities for migration to Primorsky krai from Federal districts and CIS countries by years;
- in- and out- flow data for Primorsky krai by years.

The data for the years from 2005 to 2014 was used for the learning of the cellular automaton. The number of iterations of the cellular automaton required to model one year migration flow varied within the limits between 61 and 79. The forecasting error for the total annual outflow from Primorsky krai amounted from 0,08% to 0,72%, total error taking account of the region of destination – from 4,43% to 7,13%. A considerable increase of error is largely due to the countries outside the CIS. The number of immigrants to these counties was relatively small so that absolute deviation of a few people accounted for a great percentage of the total error for that region. This is reflected in the average error for the migration flow in general.

In order to illustrate the aforementioned we will cite the 2005 data for Primorsky krai migration flows and their calculated values. The results obtained for that year are typical. 61 iterations of cellular automaton were run. The calculated total outflow from Primorsky krai was 12909 people, actual number amounted 13003 people who departed the region, the error is 0,72%.

Table 1 shows the actual and calculated migration outflows from Primorsky krai in 2005. The compound error in forecasting the migration flow by the regions totaled 4,95%.

Table 1. Actual and calculated migration flows from Primorsky krai in 2005, number of individuals

Number	Region	Calculated value	Actual value	Error, %
1.	Central Federal district	2894	2857	1,30
2.	North-Western Federal district	1122	1075	4,37
3.	Southern and North-Caucasus Federal district	1287	1310	1,76
4.	Privolzhsky Federal district	1152	1219	5,50
5.	Ural Federal district	625	638	2,04
6.	Siberian Federal district	1851	1834	0,93
7.	Ear Eastern Federal district (without Primorsky krai)	3507	3631	3,42
8.	Countries outside the CIS	136	150	9,33
9.	CIS countries	335	289	15,92

In the present research a short-term forecast of interregional migration outflow from Primorsky krai for 1 or 2 years has been made. The year 2014 was used as the forecast base that is 2014 population data was entered into the cellular automaton. The results of the forecast are presented in table 2.

Table 2. Projected values of migration flow from Primorsky krai, number of individuals

Number	Region	One-year forecast	Two-year forecast
1.	Central Federal district	2799	5547
2.	North-Western Federal district	1805	3600
3.	Southern and North-Caucasus Federal district	1759	3561
4.	Privolzhsky Federal district	990	2009
5.	Ural Federal district	447	861
6.	Siberian Federal district	1573	3322
7.	Ear Eastern Federal district (without Primorsky krai)	3066	6336
8.	Countries outside the CIS	192	332
9.	CIS countries	215	339
10.	Total	12846	25907
11.	Number of incomers to Primorsky krai	5860	12670

In order to model migration flows by a cellular automaton a cross-platform program using Go language was developed. The calculations were performed on a computer with quad-core processor Intel Core i7 with clock frequency 2.5 GHz and 16 GB of RAM memory. The amount of computation made during the experiments was rather big. On average it took 7 minutes to perform one year calculation. In order to improve the computation speed it is reasonable to consider the possibility of parallel calculations. Conclusion

4   Conclusion

The present study shows the possibility and viability of interregional migration flow modeling by cellular automata. Modeling of migration processes at macro-economic level based on simulations by cellular automata of household behavior at local level appears to be a promising research direction. In order to widen the capabilities of the developed cellular automaton in mid-term and long-term forecasting of interregional migration flows it is necessary to incorporate birth and mortality processes into a model. In other words, it seems reasonable to make adjustments for natural population decrease as well as for migration balance while forecasting the migration flows. The projected birth and mortality rates can be estimated by an application of econometric models.

References

1. Rogers, A.: A Regression Analysis of Interregional Migration in California. The Review of Economics and Statistics. 49(2), 262–267 (1967).
2. Haurin, D.R.: The regional distribution of population, migration, and climate. The Quarterly Journal of Economics. 95(2), 293–308 (1980).
3. Andrienko, Y., Guriev, S.: Determinants of interregional mobility in Russia. Evidence from panel data// Economics of Transition. 12(1), 1–27 (2004).
4. Gerber, T.: Regional economic performance and net migration rates in Russia, 19932002. International Migration Review. 40(3), 661–697 (2006).
5. Sarra, A.L., Signore, M. A.: Dynamic Origin-constrained Spatial Interaction Model Applied to Polands Inter-provincial Migration. Spatial Economic Analysis. 5(1), 29–41 (2010).
6. Vakulenko E.S., Mkrtchyan N.V., Furmanov K.K.: Modeling of registered migration flows between the regions of the Russian Federation. Applied Econometrics. 1(21), 35–55 (2011).
7. Schmidt, Yu.D.: Prognosis of demand and supply at the regional labor market. Vladivostok (2012).
8. Vasiluev A.M.: Model of labor market self-organization. Economics and mathematical methods. 37 (2), 123–127 (2001).
9. Korovkin A.G.: Dynamics of labor market and employment: problems of macro-economic analysis and prognosticating. Moscow (2001).
10. Havinson, M.Y., Kulakov, M.P., Mishchuk, S.N.: Mathematical model of population dynamics of economically active population and of foreign workers in the region (on the example of the Jewish autonomous region). Informatics and control systems. 1 (31), 95–106 (2012).
11. Blumen, I., Kogan, M., McCarthy, P.: The Industrial Mobility of Labor as a Probability Process. N.-Y.(1955).
12. Toffoli, T., Margolus, N.: Machines cellular automata. Moscow (1991).
13. Wolfram, S.: A New Kind of Science. (2002). URL: http//www.wolframscience.com/nksonline/toc.html.
14. Astafjev, G.B., Koronovskii, A.A., Khramov, A.E.: Cellular automata. Saratov (2003).
15. Lobanov A.I.: Models of cellular automata. Computer studies and modeling. 2 (3), 273–293 (2010).
16. Schmidt, Yu.D., Lobodina, O.N.: On some approaches to modeling the spatial diffusion of innovations// Spatial Economics. 2, 103–115 (2015).
17. Plotinskii, Yu. M.: Models of social processes. Moscow (2001).
18. Batty, M.: Cities and Complexity Understanding Cities with Cellular Automata, Agent-Based Models and Fractals, MIT Press, Cambridge, Massachusetts (2005).
19. Cheng, J., Masser, I.: Cellular Automata Based Temporal Process Understanding of Urban Growth. Lecture Notes in Computer Science. 2493, 325–336 (2002).
20. Ward, D.P., Murray, A.T., Phinn, S.R.: Integrating spatial optimization and cellular automata for evaluating urban change. Regional Science. 37, 131–148 (2003).
21. Benito-Ostolaza, J.M., Hernndez, P., Palacios-Marqus, D., Vila, J.: Modeling local social migrations: A cellular automata approach. Cybernetics and Systems. 46 (3-4), 287–302 (2015).
22. Dabbaghian, V., Jackson, P., Spicer, V., Wuschke, K.: A cellular automata model on residential migration in response to neighborhood social dynamics. Mathematical and Computer Modelling. 52 (9-10), 1752–1762 (2010).