Chapter 17
Generating Synthetic Population for the Agent-Based Model of the Russian Federation Spatial Development

Aleksandra L. Mashkova, Ekaterina V. Novikova, Olga A. Savina, and Evgenii A. Mashkov

Abstract In our research we generate synthetic population for the agent model of the Russian Federation spatial development according to sample-based approach. The initial data source is All-Russian population census 2010. Within procedure of generation we fill the model database with objects of the model (agents and households) and set interconnections among them. Simulation results show variation of the generated synthetic population from the original Census data.

Keywords Synthetic population · Sample-based approach · Agent model · Census data

Introduction

In the socio-economic system of Russia in the spatial aspect there is a bright disproportion: overpopulation of some regions, causing environmental, housing and transport problems in them; and, conversely, outflow of population from others. The goal of the Strategy of the Russian Federation spatial development is to overcome these problems by smoothing socio-economic differences between regions and improving interregional infrastructure, which would improve resettlement of population and centers of economic activity.

A. L. Mashkova (✉) · E. V. Novikova · O. A. Savina · E. A. Mashkov
Orel State University Named After I.S. Turgenev, Komsomolskaja St. 95, 302026 Orel, Russian Federation
e-mail: aleks.savina@gmail.com

O. A. Savina
e-mail: o.a.savina@gmail.com

E. A. Mashkov
e-mail: ppsnnt@gmail.com

A. L. Mashkova
Central Economics and Mathematics Institute, Russian Academy of Sciences, Nakhimovsky Av. 47, 117418 Moscow, Russian Federation

P. Ahrweiler and M. Neumann (eds.), *Advances in Social Simulation*,
Springer Proceedings in Complexity,
https://doi.org/10.1007/978-3-030-61503-1_17

Our research is aimed at developing a tool for evaluating alternative actions in tax, monetary and investment policies within aims of the Strategy. The model of the Russian Federation spatial development, which is being developed in our research, would reflect sex-age structure, composition of households and spatial distribution of Russian population; infrastructure, production capacities, educational and administrative institutions in different regions. The structure of the model is presented in [6].

In this article we present initial modeling data structures and sources and algorithms of population synthesis for the model. We have selected sample-based approach for this task, because large arrays of detailed information about population and composition of households in different regions of the Russian Federation are available from results of the All-Russian Population Census 2010 [1]. The population is reconstructed as records of individuals in a household in the model database so that attributes of the resulting synthetic population match as closely as possible the distributions in the census data, as it was proposed by Beckman for U.S. census data structure [3] and further implemented for the UK, Switzerland, Australia, Canada, Belgium [2, 7] and other countries. Since the Strategy of the Russian Federation spatial development considers regions as a geographic scale, there is no need to use sample-free techniques to simulate population of small geographic objects [4, 5] for this task.

Algorithm for Generating Synthetic Population

The algorithm is based on a number of tables from the collections of All-Russian Population Census 2010 [2], which is the most recent source with very detailed information. The main initial data tables used for generating population of the Russian Federation are Age-sex structure of population in different regions (1), Marriage status of residents (2), Number of private households (3), Number of residents in collective and homeless households (4), Age-sex composition of private households (5), Age structure of the married couples (6), Number of children in intact and single-parent families (7), Distribution of private households by number of residents (8).

The first step is to set geographical structure of the Russian Federation (Fig. 17.1). After that the original generation of agents is created, in accordance with the sex-age structure of population in each region (1). One agent in the model corresponds to 100 people of the same sex-age group. The created agents are assigned with status "single", "married", "divorced" or "widower" according to the table (2). The required number of private households is created (3), one model household corresponding to 100 real-world households; and also one collective household that represents prisons, shelters, monasteries, nursing homes, and one household of homeless people in each region.

The most complex task in population synthesis is distribution of agents by households, which we do in the following sequence:

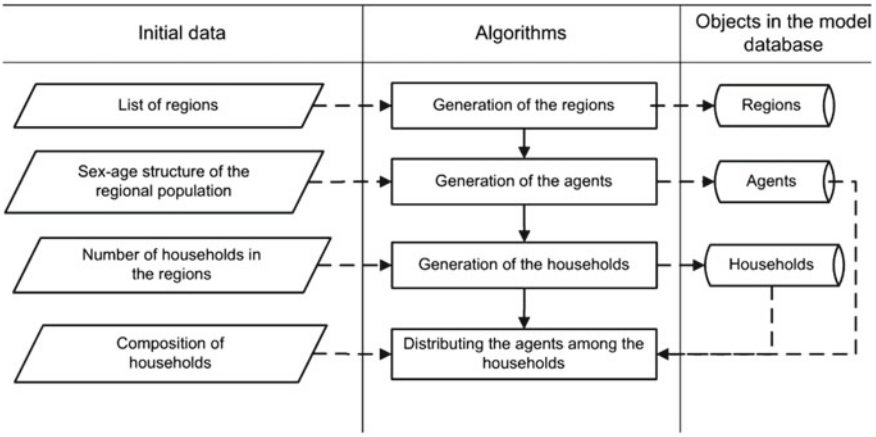

Fig. 17.1 Sequence of population synthesis in the model

1. The specified number of agents is fixed for the collective household and the homeless household (4). Agents associated with collective or homeless households are excluded from further consideration.
2. Adult agents with status "single", "divorced" or "widower" are distributed one by one to a given number of single private households, according to sex-age structure of their residents (5).
3. The specified number of married couples (2) is created, choosing male and female agents with status "married" in accordance with the age structure of the married couples (6). Created couples are assigned one by one to private households.
4. The given number of children is attached to intact families (7), respecting difference between mothers and children in the range of 16 to 45 years.
5. Single-parent families are created and the given number of children (7) is attached to them, sticking to the agreed age difference. Single parents are selected from adult agents with status "single", "divorced" or "widower". Each single-parent family is distributed to a separate private household.
6. Remaining adult agents are randomly distributed by two into empty private households.
7. Number of agents in private households is brought up in accordance with (8) by settling the remaining agents into created private households.

Generated population is stored in a database for later use in scenario calculations on the further stages.

Table 17.1 Simulation results

Parameter	Census data	Modeling data	Variation
Number of residents	142,856,536	1,428,565	$2{,}52 * 10^{-5}\%$
Number of households	54,560,627	545,606	$4{,}95 * 10^{-5}\%$
Married couples	33,206,726	332,067	$7{,}83 * 10^{-5}\%$
Residents of collective households	1,832,386	18,324	$7{,}64 * 10^{-4}\%$
Children in intact families	16,996,690	169,967	$5{,}88 * 10^{-5}\%$

Simulation Results

The model of the Russian Federation spatial development is being programmed on C# in Microsoft Visual Studio 2015. Results of the generation procedure are stored in the model database, access to them is provided by SQL-queries. Table 17.1 presents variation of characteristics of the generated population of the Russian Federation in 2010 from the original Census data. To cope with computational complexity of generating synthetic population of 142 million of residents we have reduced number of agents (100 to 1), the opposite operation was made to modeling data for calculating variation of results.

Conclusions

In this article we presented results of the first stage of development of an agent-based computer model of the Russian Federation spatial development. The model is a tool for assessing control actions aimed at eliminating imbalances in population distribution, economic achievements and quality of life in various regions. At this stage we reconstructed population of different regions and their distribution among households in accordance with information from All-Russian Population Census 2010.

The purpose of the generation procedure was to create objects of the model (population and households) and reflect their interconnections in the model database. In the process of program implementation, agents and organizations were aggregated: one agent in the model corresponds to 100 residents. Simulation results show very small variation from Census data due to used sample-based technique: the number of created objects is pre-determined and only their distribution is stochastic. Meanwhile, the obtained results show validity of the proposed population synthesis algorithms and their program implementation.

Acknowledgements The reported study was funded by RFBR according to the research project № 18-29-03049.

References

1. All-Russian population census 2010 official website. https://www.gks.ru. Last Accessed 22 Mar 2019
2. J. Barthelemy, P. Toint, Synthetic population generation without a sample. Transport. Sci. **47**(2), 266–279 (2013). https://doi.org/10.1287/trsc.1120.0408. https://pubsonline.informs.org/doi/abs/10.1287/trsc.1120.0408
3. R.J. Beckman, K.A. Baggerly, M.D. McKay, Creating synthetic baseline populations. Transport. Res. Part A: Policy Pract. **30**(6), 415–429 (November 1996). https://ideas.repec.org/a/eee/transa/v30y1996i6p415-429.html
4. K. Harland, A. Heppenstall, D. Smith, M. Birkin, Creating realistic synthetic populations at varying spatial scales: a comparative critique of population synthesis techniques. J. Artif. Soc. Social Simul. **15**(1), 1 (2012). https://doi.org/10.18564/jasss.1909. https://jasss.soc.surrey.ac.uk/15/1/1.html
5. N. Huynh, J. Barthelemy, P. Perez, A heuristic combinatorial optimisation approach to synthesising a population for agent based modelling purposes. J. Artif. Soc. Social Simul. **19**(4), 11 (2016). https://doi.org/10.18564/jasss.3198. https://jasss.soc.surrey.ac.uk/19/4/11.html
6. A.L. Mashkova, O.A. Savina, Y.A. Banchuk, E.A. Mashkov, Using open data for information support of simulation model of the Russian Federation spatial development, in *Electronic Governance and Open Society: Challenges in Eurasia*. ed. by A. Chugunov, Y. Misnikov, E. Roshchin, D. Trutnev (Springer International Publishing, Cham, 2019), pp. 401–414
7. K. Mueller, K.W. Axhausen, Hierarchical IPF: Generating a synthetic population for Switzerland. ERSA conference papers ersa11p305, European Regional Science Association (Sep 2011). https://ideas.repec.org/p/wiw/wiwrsa/ersa11p305.html