ITM Web of Conferences **59**, 02002 (2024) https://doi.org/10.1051/itmconf/20245902002
*HMMOCS-II 2023*

# Application of mathematical methods to solving problems of digitization of population movement

*Karolina* Ketova, and *Daiana* Vavilova*

Kalashnikov Izhevsk State Technical University, 7, Studencheskaya street, Izhevsk, 426069, Russian Federation

**Abstract.** This article is devoted to the development of algorithms and mathematical methods for digitalization of population movement. An algorithm for digitalization of demographic flows is proposed. It becomes possible at any time to obtain a complete description of both a specific person and a general characteristic of the state of the economic system in a given context (for example, age, gender, place of residence, type of settlement, level of education, level of health, level of culture). Within the framework of the problem, four tasks are identified, which the research is aimed at solving. The first task is constructing a scheme of a person's digital trace. The second task is aggregating digital traces and structuring demographic flows and related flows of human capital using Big Data technology. The next task is studying the characteristics, properties and qualities of the said flows using Data-analysis technology. The final task is analyzing and forecasting demographic and human capital flows using Data Science technology. When implementing Data Science technology, the use of mathematical methods of statistical data processing, methods of correlation and regression analysis, mathematical models, forecasting methods, artificial intelligence algorithms, including neural network models, is proposed to solve the task.

## 1 Introduction

In the modern world, digitalization is an important tool for the effective development of territories. Digital technologies make it possible to meet the high rates of development in many sectors of human activity and the high rates of development of science and technology.

The information and computer revolution, according to one of the authors of the concept of post-industrial society, American philosopher E. Toffler [1], will make information as important a resource as production and financial resources. According to D. Bell [2], the digitalization process, which entails the information development of society, will bring it to a new stage of social development in all sectors of life.

Digital technologies provide strategically important directions for the activities of both the state as a whole and the individual. Today, the development of digital technologies is an urgent task, since they contribute to strengthening the national security of the state and its

* Corresponding author: vavilova_dd@mail.ru

population. The importance of information technology for the digital economy of the Russian Federation is analyzed in [3].

The use of digital technologies allows to systematic approach for completing tasks and implementing monitoring in the field of national projects of the Russian Federation. The state's compliance with its social guarantees is directly related to ensuring the quality of human life. Particularly significant among them are projects aimed at socio-economic development: "Demography", "Health". The development of information infrastructure in these areas is prescribed in the national project "Digital Economy". The development of algorithms, as well as a systematic approach to the formation of flows of demographic and related information will contribute to the rapid development of infrastructure.

An analysis of studies fully or partially devoted to digitalization processes allows us to conclude that scientific work on systemic modeling of flows of demographic information and associated flows of qualitative characteristics (human capital) is currently absent. Attempts are being made to use data from digital traces of people left on the Internet for research into the quality of life and well-being of the population, a number of other aspects of sociology [4] and migration [5]. The digitalization process is most intensively implemented in the educational environment due to objective reasons related to the constant presence of electronic educational resources in the student's work. In particular, methodological assistance is already being used in mastering the algorithm for working with digital tools and digital traces, as well as the subsequent interpretation of the digital trace and the formation of a student's digital profile. A digital profile as a key element of the infrastructure of the digital economy is considered in [6].

The article [7] pays attention to the creation of an integrated data model from heterogeneous sources containing digital traces. Works are devoted to the use of Big Data in the development of the digital economy [8, 9].

An up-to-date forecast of demographic information flows and associated human capital flows is of great importance. Until recently, the direction associated specifically with the digitalization of population flows did not receive due support and attention in terms of constructing mathematical tools for the implementation of this scientific direction. The impetus for the progress of work was the 2020 pandemic. From that moment on, some works devoted to this topic began to appear in foreign literature. For example, Chinese researchers [10] present a framework for a new digital model of mobility and physical movement of people to monitor virus transmission and the effectiveness of social distancing measures during the ongoing pandemic. The model, based on Hong Kong data, makes it possible to track the effective number of Covid-infected people in almost real time, which is important for quickly assessing the effectiveness of ongoing activities.

A scheme for monitoring population movements using mobile devices generated by Big Data during the coronavirus crisis is presented in the work of Hungarian researchers [11]. They show that accurate current and virus epidemic forecasts can be obtained by integrating up-to-date digital information about people's physical movements into epidemic models.

In today's environment, there is a need to implement mathematical models to predict various characteristics of the population. For example, the work [12] presents a number of social policy measures based on Data analysis of the healthcare and social systems in Cyprus and Greece. Using Bayesian hierarchical models, the results of demographic forecasts are presented, which serve as the basis for family policy measures to increase fertility and migration in order to improve demographic structure and economic activity. National systems promote preventative strategies in the primary healthcare sector and promote the creation of assistive technologies for healthy aging in older people, taking into account innovative technologies and digitalization in healthcare.

The systematization of problems in the medical field, the formulation of tasks and the search for solutions in the field of digital medicine are outlined in the research [13]. Applied systems analysis, software engineering and artificial intelligence methods are used as research methods. The need for organizational changes in medicine based on information technology is substantiated, which is the essence of the digital transformation of healthcare.

We can also cite a number of modern foreign works that propose using modern technical devices to track a person's location. But these works are in a different plane of research. These developments do not allow solving global system problems of modeling population flows in order to optimize government regulators to improve the quality of life in the complex sense of this concept.

The purpose of this study is to develop a consistent scheme of population flows for the purpose of its further digitalization, as well as the use of data mining methods for modeling and forecasting the characteristics of human capital. Practical implementation consists in constructing tools for storing and processing information about our population using modern technologies for working with Big Data, Data Analysis, Data Science technology, and machine learning methods. To achieve the goal of the study, it is necessary to have statistical information that is structured.

A sequential research plan involves solving 4 problems:

1) to construct a diagram of a person's digital trace over time, starting from birth and throughout life;
2) to structure of demographic flows and a number of socio-economic processes (education, healthcare, culture) using available dynamic data on the digital traces of the population; Big Data technology is used to solve the problem;
3) to study of the characteristics, properties, qualities of demographic flows and a number of socio-economic processes accompanying the movement of demographic flows; Data analysis technology is used to solve the problem;
4) to analyze and forecast of quantitative and qualitative characteristics of population flows; Data Science technology is used to solve the problem.

## 2 The task of constructing a diagram of a person's digital trace in dynamics

Modern people, as a rule, use information technologies, the Internet, in everyday life, and leave digital data (Data) or a digital (information) trace. This term is understood in the literature in precisely this sense.

If you make a request for "digital trace" on the Internet, then the "world wide web" will give out such concepts as "security", "cybercrime", etc. Indeed, you can get a fairly complete picture of how a person behaves in a virtual environment about him. But this information needs to be systematized and directed not to harm, but for the benefit of the individual and society. That is why we will further give a different meaning to the concept of "digital trace". What is important here is information that will be in one place, stored dynamically, and allow one to analyze human development in all areas of life.

A scheme for constructing a person's digital trace is proposed in Figure 1. All parameters that are important for describing the state and movement of a person in time can be divided into constant, conditionally constant and variable. Constant parameters include date and place of birth, gender, and biometric characteristics. Conditionally constant parameters: full name; places of work; educational institutions, medical and other institutions with which he dealt. Variable parameters include the qualitative characteristics of a person's life, namely, level of education, culture, qualifications, health status, marital status, etc.

When such a set of information is formed, it becomes possible to obtain a complete description of a person at any time. This is important in many cases, for example, when identifying economic crimes, to implement the ability to automatically provide citizens with social charges and payments, receive tax deductions, etc. This saves time for the person himself and government services (for example, we need to know that a child. After graduating from preschool, he was registered at the school). The formation of an electronic patient record is an important aspect, since tracking health status over time allows you to respond to emerging problems in a timely manner and thereby improve the quality and length of life.

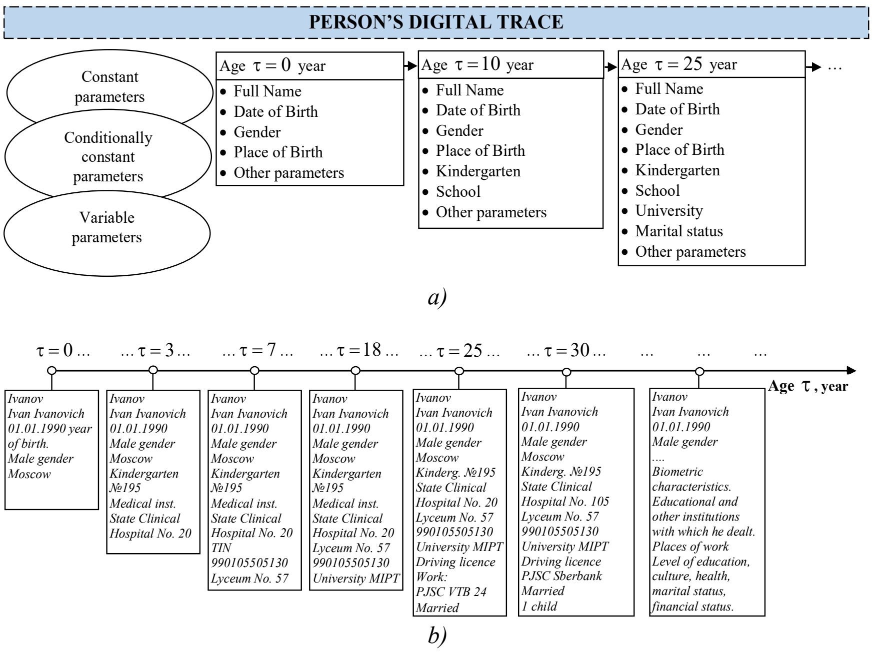



**Fig. 1.** Digital trace of a person throughout life: set of parameters *(a)* and their condition *(b)*.

Obviously, the qualitative characteristics of the human condition play no less a role than the quantitative ones. Such characteristics as the level of education, health and culture of a person are presented in Figure 2.

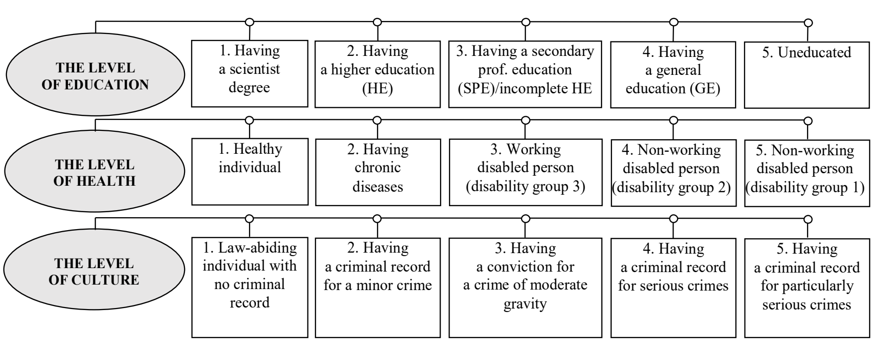
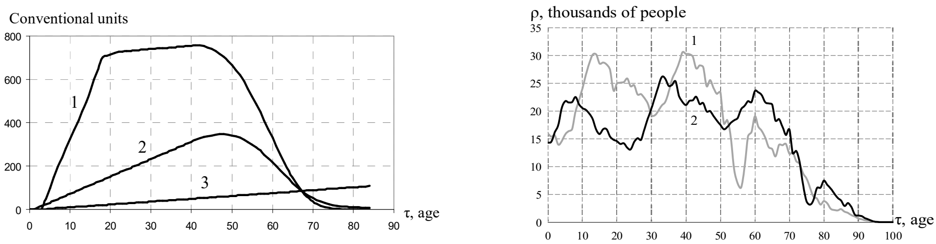



**Fig. 2.** Gradations of characteristics of demographic elements of society.

The rate of change in human performance characteristics varies over time; their qualitative dynamics are presented in Figure 3, *a* [14]. The level of culture is cumulative in nature, while education and health are subject to attrition from a certain point.

To construct demographic flows in the context of a certain territory, it is necessary to aggregate all statistics on demographic elements for anyone at any given time. So, for example, the distribution of population (density) in the Udmurt Republic (UR) region by age in 2000 and 2020 [15] is presented in Figure 3, *b*.




*a)* *b*)

**Fig. 3.** Qualitative dynamics of human characteristics: 1 – level of education, 2 – level of health, 3 – level of culture (*a*); density of population distribution in the UR by ages in 2000 and 2020 (*b*).

Aggregating information about people's digital traces makes it possible to construct information flows about the population in dynamics.

## 3 Big data: the task of structuring demographic flows and related human capital flows

The problem of structuring demographic flows, as well as the flows of socio-economic indicators accompanying this movement, is solved to organize information in the field of demographic and socio-economic processes. Aggregation of information occurs using available dynamic data on the digital traces of the population. To solve the problem, Big Data technology is used.

It is necessary to structure information on the total population, as well as in various sections, important of which are distributions by age and territory. A separate important element of information structuring is the processes of fertility, mortality, and migration. Globally, on the part of society as a whole, it becomes possible to assess changes in the population situation in a specific context of place and time, thereby implementing operational management of socio-economic systems.

The result of this stage is a structured statistical database on population movements, demographic flows, as well as the movement of human capital, which is a consequence of the dynamics of demographic flows. A database filled with statistical information is necessary to develop a digital map of demographic flows and analyze their quantitative and qualitative characteristics. Digital information must be structured in such a way that it is possible to algorithmically implement intelligent methods for processing Big Data.

The quality of the solution to the problem of constructing demographic flows and the flows of human capital accompanying this movement directly depends on the reliability of the information and the degree to which the digital information database is filled. At the initial stage, we will use the reporting data of official statistics, information from the Governments of the relevant constituent entities of the Russian Federation, ministries and departments. In some cases, proprietary verification and addition technologies can be used to restore missing data [16]. A database of demographic and socio-economic information for the Udmurt Republic is integrated into the information and analytical system "Modeling and forecasting indicators of socio-economic processes in the region" developed by the authors [17]. Which could be the basis for further work in this direction.

# 4 Mathematical models and data analysis: the task of studying demographic flows and related flows of human capital

The problem of studying demographic flows, as well as the flows of socio-economic indicators accompanying this movement, is solved to organize information in the field of demographic and socio-economic processes using Data Analysis technology.

The main characteristic of population is its numerical reproduction. What is important here is balance in demographic processes. Analysis and forecast of modern demographic processes is necessary for the implementation of effective demographic policy. When studying demographic phenomena using modern mathematical tools, territorial and socio-economic characteristics are taken into account when analyzing the duration and quality of life, fertility, mortality, population migration, etc.

To model the age dynamics of demographic elements, the following equation is used:

$$\frac{\partial\rho(t,\tau)}{\partial t}+\frac{\partial\rho(t,\tau)}{\partial\tau}=-\mu(t,\tau)\rho(t,\tau)+l(t,\tau)\rho(t,\tau). \quad (1)$$

where $\rho(t,\tau)$ is the population density by age $\tau$ per year $t$ ; $\mu(t,\tau)$– coefficient of distribution of population mortality by age, which specifies the proportion of deaths in an age group $\tau$ per year $t$ ; $l(t,\tau)$– migration interaction coefficient, which specifies the share of migrants in an age group $\tau$ per year $t$ .

To solve equation (1), it is necessary to determine the initial and boundary conditions based on the initial statistical data.

Initial condition is at $t=t_0$ :

$$\rho(t_0,\tau)=\rho_0(\tau), \quad \tau>0 , \quad (2)$$

where $\rho_0(\tau)$ is the population distribution density at the initial time (known function).

Boundary condition is at $\tau=0$ :

$$\rho(t,0)=\int_{\tau_{1f}}^{\tau_{2f}}\beta(t,\tau)\rho(t,\tau)d\tau, \quad t>t_0 , \quad (3)$$

where $\beta(t,\tau)$ is the coefficient of distribution of births by age, which specifies the proportion of births in each age group $\tau$ per year $t$ (density of distribution of births from the range of women's fertility $[\tau_{1f};\tau_{2f}]$).

The dynamics of some quantitative demographic characteristics of the Udmurt Republic for the period 2000-2020 are shown in Figure 4. Figure 4, *a* – total population, Figure 4, *b* – indicators of population reproduction in the region.

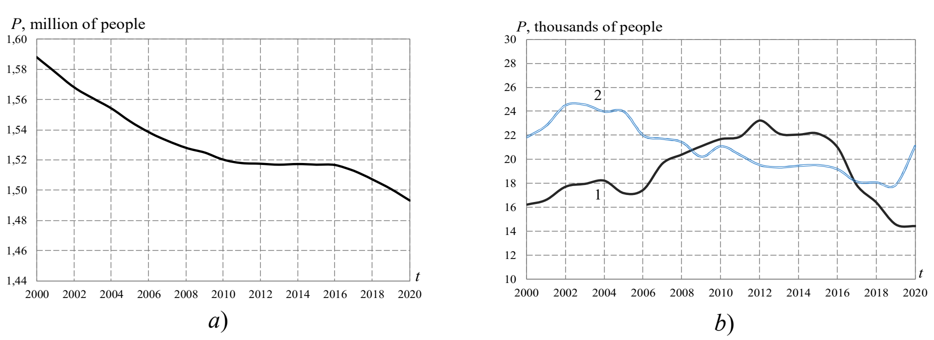



**Fig. 4.** Dynamics of quantitative demographic characteristics of UR for the period 2000-2020: (*a*) population size; (*b*) birth rate (1) and death rate (2).

In addition to taking into account quantitative demographic characteristics in the socio-economic system, it is planned to study social groups formed by demographic elements according to qualitative characteristics. It is proposed to use mathematical models of the dynamics of the population structure by levels of education, health and culture. In particular, to calculate the population with higher education in the socio-economic system, a dynamic equation is proposed in the form:

$$\frac{\partial \rho_{he}(t,\tau)}{\partial t} + \frac{\partial \rho_{he}(t,\tau)}{\partial \tau} = [\chi_{he}(t,\tau) - \mu(t,\tau)]\rho(t,\tau). \quad (4)$$

Here $\rho_{he}(t,\tau)$ is the density of the population with higher education, aged $\tau$ per year $t$; $\chi_{he}(t,\tau)$ is the proportion of the population with higher education at the age $\tau$ per year $t$.

Initial and boundary conditions are:

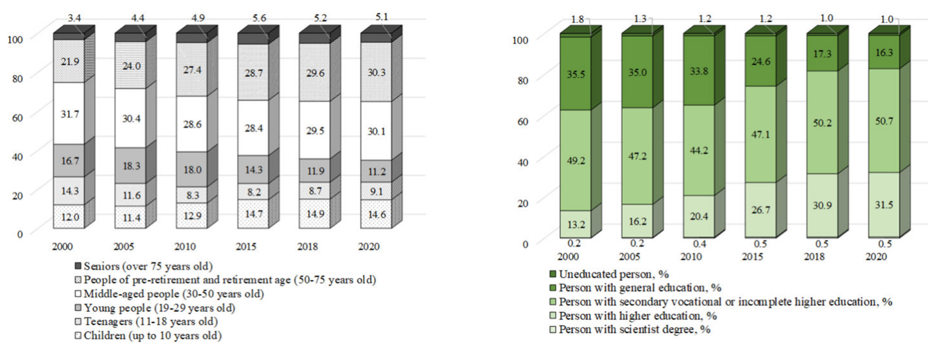
at $t = t_0$:

$$\rho_{he}(t_0,\tau) = \rho_{he0}(\tau), \quad (5)$$

at $\tau = 0$ or $\tau = \infty$:

$$\rho_{he}(t,0) = \rho_{he}(t,\infty) = 0. \quad (6)$$

The change in the structure of the population of the Udmurt Republic by age groups for the period 2000-2020, estimated by using models (1)-(6), is presented in Figure 5.




*a*) *b*)

**Fig. 5.** Dynamics of the population structure of the UR for the period 2000-2020: (*a*) by age groups; (*b*) by level of education.

The use of mathematical models and Data Analysis technology allows you to obtain important results necessary for assessing the current state of the system, its analysis and further planning. So, for example, for the result given here (Figures 4, 5), the following conclusions can be drawn.

During the period under review, the population of the region decreased by 5.9% compared to 2000. The average annual rate of population decline in the UR was 0.3%. Population dynamics are influenced by fertility and mortality. The maximum birth rate for the period was recorded in 2012 (23.2 thousand children), the minimum in 2020 (14.5 thousand children). For the mortality rate, the highest value was recorded in 2003 (24.6 thousand people), the lowest value was in 2019 (17.9 thousand people), while for the period 2003-2019 there is a visible trend towards a decrease in mortality (average annual rate a decrease of 2.0%), but 21.1 thousand people died in 2020, which is 17.9% more than the previous year.

In the period 2000-2008, there was a natural population decline for the socio-economic system of the UR (maximum value of 6.8 thousand people in 2005). During 2009-2016, the number of births in the region exceeded the number of deaths, and natural population growth was observed (the maximum value was 3.7 thousand people in 2012). Since 2017, there has been a natural population decline in the UR, which in 2020 amounted to 6.6 thousand people.

The population is aging: the proportion of adolescents and young people in the total population is decreasing. The average age of the population in the UR increased from 32.2 years in 2000 to 39.1 years in 2020. In 2020, compared to 2000, the number of centenarians has increased. The share of the population with an academic degree increased from 0.2% in 2000 to 0.5% in 2020 and the share of the population with higher education increased from 13.2% to 31.5%; at the same time, the proportion of people of working age with only general and secondary basic education decreased significantly, from 35.5% to 16.3%.

We see that the use of a database of digital information on the movement of demographic flows and the use of Data-analysis technology to study their characteristics will generate a large amount of information that needs systematization to organize it and identify internal relationships. This problem is solved using Data Science technology, and has already been partially solved using the proposed mathematical models.

## 5 Mathematical models and data science: the task of analyzing and forecasting demographic flows and related human capital flows

The problem of using Data Science technology is solved to analyze and forecast the quantitative and qualitative characteristics of population flows. At this stage, mathematical methods for processing statistical data are used, for example, such as methods of correlation and regression analysis, constructing mathematical models, forecasting methods, artificial intelligence algorithms, including neural network models.

A diagram of the use of Data Science technology for the digitalization of the country's population movement is presented in Figure 6. When forming such a set of information, it becomes possible at any time to obtain a complete description of both a specific person and a general characteristic of the state of the economic system in a given context (for example, age, gender, place of residence, type of settlement, level of education, level of health, level of culture). For example, let $M$ is the complete dynamic set of the population (a set of people within a given territory on a known time interval), which has the form $M = \{M_{jkl...};\ j,k,l... \in N\}$. The combination of subscripts $M_{jkl...}$ characterizes the demographic flow in a given context. If, for example, the first index denotes gender ( $j=1-$ male, $j=2-$ female), the second index - the type of settlement ( $k=1-$ urban, $k=2-$ rural) and the third index is age ( $l=1-$ age group $\tau \in [0;3)$ years, $l=2-$ $\tau \in [3;7)$, $l=3-$ $\tau \in [7;16)$, $l=4-$ age group $\tau \in [16;65)$, $l=5-$ age group $\tau \geq 65$ ), then the set $M_{213}$ consists of schoolgirl girls living in the city.

When implementing Data Science technology, we will use mathematical models of quantitative and qualitative characteristics to solve the problem. Thus, we will model quantitative demographic flows using the equation of demographic dynamics (Ketova, 2013), which contains dynamic functions of population distribution by age, functions of mortality, fertility and migration interaction. We will evaluate the qualitative characteristics of demographic flows using methods of correlation and regression analysis and constructing their mathematical models [18-20].

The forecast problem is solved by methods of multidimensional analysis of population flows, taking into account changes in the quantitative and qualitative characteristics of human capital (education, health, culture), which accompany the physical flows of movement of demographic elements across the territory. Neural network models occupy an important place among forecasting methods [21-23].

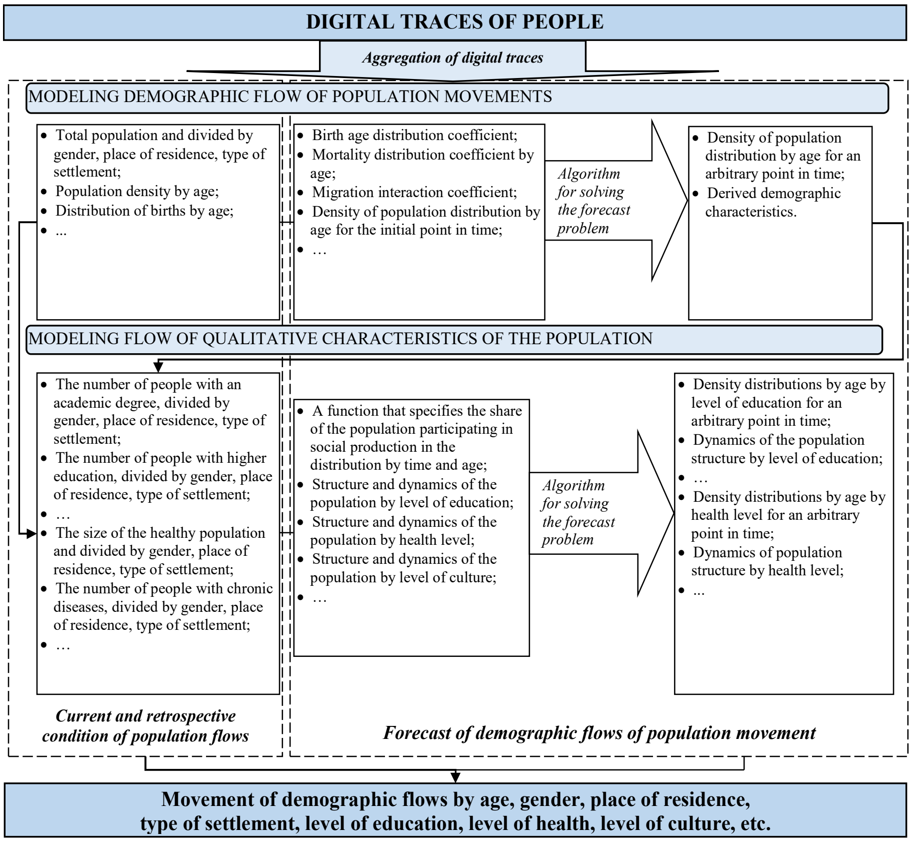



**Fig. 6.** Algorithm for digitalization of population movement.

The authors conducted research in the field of forecasting socio-economic processes using models of regression and autoregressive analysis, the method of principal components, neural networks, as well as their comparison. The authors [23] show that the most effective tool for making forecasts is a neural network. Thus, a neural network model is used to predict the quantitative and qualitative characteristics of population movement. In this model, the input data is a set of socio-economic indicators of the region for a fixed period.

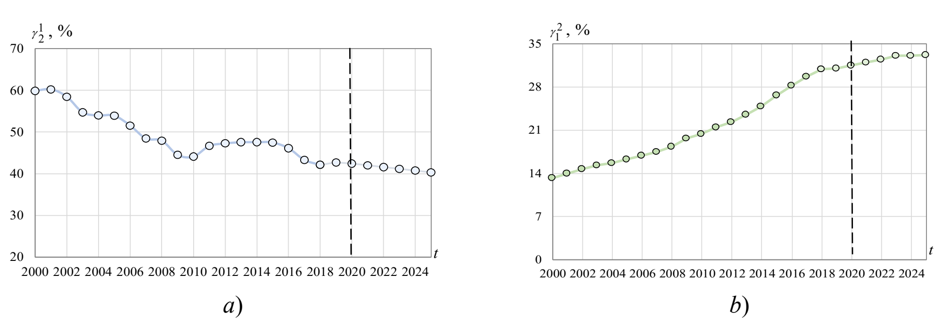



**Fig. 7.** Dynamics and forecast of qualitative characteristics of the population of the UR: (*a*) the share of healthy people; (*b*) the share of people with higher education.

The indicators included in the aggregate may be changed depending on the purposes of the study. In particular, a study was conducted where the output data for forecasting human capital were budgetary and private investments in education, health and culture, indicators of socio-economic development, inflation rate, and qualitative characteristics of the population. The output data of the neural network is the forecast indicators of the flows of quantitative demographic and related qualitative characteristics of population movement.

Some results of forecasting the characteristics of demographic flows using the example of the UR are presented in Figure 7.

Numerical implementation of algorithms for modeling and forecasting the movement of population flows is also possible using application packages in a modern programming language (C#, Python, etc.).

## 6 Conclusion

An algorithm for digitalization of population movement has been compiled, which is configured to take into account the maximum number of characteristics. This is both the movement of quantitative demographic flows and the movement of qualitative flows (human capital). The logic of the algorithm includes a transition from the particular to the general, i.e. from constructing a digital trace of a person at every moment in time to generating information about the movement of the entire population.

The task of constructing a diagram of a person's digital trace is formulated; the task of aggregating digital traces and structuring demographic flows and related flows of human capital using Big Data technology; the task of studying the characteristics, properties and qualities of these streams using Data analysis technology; the task of analyzing and forecasting demographic flows and human capital flows using Data Science technology.

The final result is a digital spatiotemporal map of demographic and human capital flows. This result is a significant result for the development of the digital economy in our country.

## References

1. Al. Toffler, H. Toffler, *Revolutionary Wealth: How it will be created and how it will change our lives* (Currency Publications Ltd., 2007)
2. D. Bell, *The Coming of Post-Industrial Society. A Venture in Social Forecasting* (New York, Basic Books, 2001)
3. K.V. Pitelinsky, A.O. Tikhomirov, A.A. Shirokov, Defence complex - scientific and technical progress of Russia, **146** (2020)
4. E.V. Shchekotin, Bulletin of Tomsk State University **467** (2021). https://www.doi.org/10.17223/15617793/467/21
5. A.V. Smirnov, Demographic Review **9**, 2 (2022). https://www.doi.org/10.17323/demreview.v9i2.16205
6. M.A. Tobien, M.M. Markhaichuk, O.A. Kosareva, Problems of theory and practice of management **11** (2021). https://www.doi.org/10.46486/0234-4505-2021-11-177-192
7. P.K. Chernov, E.A. Rabchevsky, Bulletin of Perm University. Mathematics. Mechanics. Computer science **57**, 2 (2022). https://www.doi.org/10.17072/1993-0550-2022-2-81-87
8. S.D. Belov, D.P. Zrelova, V.V. Korenkov, System analysis in science and education, 2 (2020). https://www.doi.org/10.37005/2071-9612-2020-2-187-197
9. A.A. Sidorov, P.V. Senchenko, V.F. Tarasenko, Bulletin of Tomsk State University. Management, computing and information science, 51 (2020). https://www.doi.org/10.17223/19988605/51/14

10. K. Leung, J.T. Wu, G.M. Leung, Nature Communication **12**, 1 (2021). https://www.doi.org/10.1038/s41467-021-21776-2
11. M. Szocska, P. Pollner, I. Schiszler, A. Toth, P. Gaal, Scientific Reports **11**, 1 (2021). https://www.doi.org/10.1038/s41598-021-81873-6
12. D. Lamnisos, K. Giannakou, M.M. Jakovljevic (2021). Health Research Policy and Systems **19**, 1 (2021). https://www.doi.org/10.1186/s12961-020-00666-x
13. V.N. Pavlov, A.M. Khanov, A.G. Tyurganov, Public health **2**, 2 (2022). https://www.doi.org/10.21045/2782-1676-2022-2-2-73-76
14. K.V. Ketova, *Mathematical models of economic dynamics* (Izhevsk, IzhSTU Publishing House, 2013)
15. Official statistical information on the population in the Udmurt Republic (2023). URL: https://udmstat.gks.ru/folder/51924
16. A.A. Sorokin, I.M. Borodyansky, A.V. Dagaev, News of the Southern Federal University. Technical science, 4 (2020). https://www.doi.org/10.18522/2311-3103-2020-4-93-107
17. D.D. Vavilova, K.V. Ketova, Certificate of registration of a computer program No. 2022619947, 05/27/2022
18. Yu.M. Romanovsky, K.V. Ketova, I.G. Rusyak, Computer Research and Modeling **11**, 2 (2019). https://www.doi.org/10.20537/2076-7633-2019-11-2-329-342
19. K.V. Ketova, D.D. Vavilova, Economic and Social Changes: Facts, Trends, Forecast **13**, 6 (2020). https://www.doi.org/10.15838/esc.2020.6.72.7
20. D.D. Vavilova, K.V. Ketova, *Demographic problems of modern Udmurtia*. (Izhevsk, IzhSTU named after M.T. Kalashnikov, 2022)
21. L.Yu. Emaletdinova, Z.I. Mukhametzyanov, D.V. Kataseva, A.N. Kabirova, Computer Research and Modeling **12**, 4 (2020). https://www.doi.org/10.20537/2076-7633-2020-12-4-737-756
22. M.A. Reshitko, A.B. Usov, Computer Research and Modeling **14**, 3 (2022). https://www.doi.org/10.20537/2076-7633-2022-14-3-539-557
23. I.G. Rusyak, K.V. Ketova, D.D. Vavilova, Tomsk State University Journal of Control and Computer Science **53** (2020). https://www.doi.org/10.17223/19988605/53/2