Comparative Analysis of Modern Simulation Computer Modeling Technologies Using Big Data Technologies

Alexey Tkachenko[a], Maksim Denisenko[b] and Denis Lavrentev[c]
Kaluga State University named after K.E.Tsiolkovski, 26, Stepan Razin St., 248023,Kaluga, Russia
TkachenkoAL@tksu.ru

Keywords: Simulation modeling, Data analysis, Big data, AnyLogic, Alina GPS, Maxima, Simpas, Virus strain, Drives, Cloud storage.

Abstract: This article presents criteria for evaluating simulation systems using Big Data technologies. For modeling, it is important to have an infinitely large data set for analysis, since rapidly changing information can critically affect the final results of modeling. To date, there are many solutions to the problems of design and development of simulation modeling on the technology market. However, computer modeling systems are not standardized, so some of these are either highly narrowly focused on a specific industry. The rest, a smaller part, are professional systems with a huge computing and technological resource. The scope of application of simulation modeling is limitless, but the most striking example of the productive use of this field is medicine: it is possible to test hypotheses on the use of certain drugs in specific, previously unexplored cases, predicting the course of the disease both in a particular patient in particular and in people in a certain area in general. The relevance and relevance of modeling systems is evaluated. A comparative analysis of modeling systems according to a number of the most important criteria is carried out. A product suitable for most of these requirements has been identified. The article builds a simulation model with the connection of big data technologies. As a model, the modeling of the incidence of influenza in the territory of the Russian Federation was carried out. The calculations use data from various official sources, the model takes into account all types of influenza, without dividing into strains and groups.

1 INTRODUCTION

In modern realities, modeling plays a crucial role in a wide variety of areas of human life and activity. The use of this technology allows you to calculate risks in advance and predict future trends in the field under study based on available data.

The foundation in this area is the topic of simulation modeling. Imitation here is just an opportunity to test theoretical knowledge, avoiding using them on real objects, people and other living beings.

The scope of simulation modeling is limitless, but the most striking examples of the productive use of this field are medicine and economics.

From the point of view of medicine, it is possible to test hypotheses on the use of certain medicines in specific, previously unexplored cases, predicting the course of the disease both in a particular patient in particular and in people in a certain area in general.

In the economic sphere, this model can be used to predict the dependencies of supply and demand on various factors, optimize the processes of personnel recruitment and other situations in the business environment.

An important aspect of simulation modeling is the use of specialized software designed to maximize the satisfaction of most of the most common requests for this type of application. There is a relatively wide range of software products on the software market, and the choice of a specific instance should be determined by the correspondence of the modeling goal and the financial capabilities of the developers. Thus, the purpose of this study is to determine the most suitable simulation modeling from the software

[a] https://orcid.org/0000-0003-4276-3017
[b] https://orcid.org/0000-0002-2064-8627
[c] https://orcid.org/0000-0001-6601-1325

available on the market for most of the tasks of building models of the application.

For modeling, it is important to have an infinitely large data set for analysis, since rapidly changing information can critically affect the final results of modeling. The so-called "Butterfly Effect", formulated by Henri Poincare in the postulates of chaos theory, works for each model. The designation of this phenomenon states that the smallest change in the input data leads to critical changes in the output results. Thus, the more information is provided as training and analyzed blocks of the model, the more accurate and fair it will be.

The relevance of this work is to determine the most optimal software for simulation modeling using Big Data technologies. Since there are many different systems for the formation of simple models, without a permanent connection to the database, which greatly impairs the performance and relevance of the information received as output data.

In the course of the study, it is required to determine a software tool for creating computer modeling projects. This need is described by the following factors: firstly, that in the current economic and political situation in the Russian Federation and in the world there is a need for import substitution of foreign applications in circulation; secondly, as mentioned earlier, there is a fairly wide range of various simulation software, and there is a public demand to determine the most functionally complete and convenient simulation application.

2 RESEARCH METHODOLOGY

To date, there are many solutions to the problems of design and development of simulation modeling on the technology market. However, computer modeling systems are not standardized, so some of them are either highly narrowly focused on a specific industry, or simply have a huge base of common tools that are ineffective when going deeper into calculations and graphical construction. The rest, a smaller part, are professional systems with a huge computational and technological resource: the tools accurately count all occurrences, the data output is processed and statistically described.

Information resources obtained using the Internet are used as input materials for analysis and modeling. These include data sets from research groups and teams that are published as "raw" data in tabular form intended for analysis by special software systems. The analysis of this kind of data involves the search for relative and absolute patterns in the proposed information, finding structural features and obtaining data as close to reality as possible.

In our work, universal and qualitative systems will be analyzed, which include the following:

• AnyLogic is a platform application designed for simulation modeling of various structures used in the business sphere. The main advantage of the product is the built-in multimedia animation libraries that allow you to visually reflect the simulated aspects. In addition, the advantages of AnyLogic include cross-platform (there are versions for various operating systems on the market), the availability of free use for educational purposes (the Personal Learning Edition version of the program). The widely used and studied Java language is used for writing programs, which can also be considered an advantage of this platform. The disadvantages of this program are the following: this product is not included in the register of Russian software, thus, in case of sanctions restrictions, access to it will be difficult; for full use, it is necessary to purchase paid versions.

• ALINA GPSS is a software product whose main task is to create simulation models of public services. The advantages include the presence of a graphical user interface and a visual representation of models, the possibility of direct input and output of information in the process of its operation, a free training version, a Russian development (based on a foreign product). The disadvantages include the GPSS language used, which is not studied in sufficient volume at Russian universities, and the difficulty for intuitive understanding.

• Maxima is a program for working with computer algebra, i.e. for using it on mathematical expressions in symbolic form. The advantages include free of charge, open code and the ability to work conveniently with very large quantities. The disadvantages, in turn, are the use of the Common Lisp language, to a certain extent outdated and not studied, as well as the lack of extensive visual animation capabilities.

• Simpas is a domestic development for simulation computer modeling. The main advantage is the country of development – in case of sanctions pressure, this product will remain available. The disadvantages include the narrow specialization of the program – the main purpose is the modeling of information processes, computers of complex architecture and computer networks. At the same time, the use of the Pascal language can be called both a disadvantage and an advantage: on the one hand, the language is quite widespread and is being studied, on the other hand, it is difficult to use a general–purpose language in a fairly narrow field of modeling.

In order to objectively evaluate computer simulation technologies, the following criteria and their corresponding performance indicators were selected.

Criteria (what is it, how we evaluate it):

• Programming language – is a key structure for any software products. Choosing the right language can improve the performance of any software and system. Each language organizes memory and code work in its own way. However, not all languages are suitable for solving specific tasks, and sometimes even degrade performance. For system modeling, it is important how the programming language processes the code and allocates memory, but also a concise and understandable syntax is necessary for the programmer, which minimizes errors in the simulation model. The following evaluation parameters were determined for the analysis of the programming language:

• 1 – poorly readable and inconvenient syntax

• 2 – the language is easy to understand, but a small pool of possibilities

• 3 – the language is simple and productive.

Let's evaluate modeling systems by the criterion – programming language syntax:

- The AnyLogic simulation system is written in the Java programming language. The syntax of the programming language gets a score of 3. Java belongs to the category of high-level languages. One of the main advantages of Java is the Garbage Collector technology, which allows you to control the memory used by the model, thereby improving performance and performance parameters. Also, the memory manipulated by the JVM allows you to logically and adequately set and configure data structures of any complexity, use algorithms of varying complexity, create a weak connection between objects. Java has a huge number of frameworks and libraries, which makes it easier to write and use algorithms and anonymous functions (predict, consumer, etc.).

AnyLogic is constantly adapting to new versions of the language that have the LTS label.

- The Maxima simulation system is written in the Common Lisp programming language. The syntax of the programming language gets a score of 2. Common Lisp programs focus on code support. The language is built in such a way that it allows programming the programming language itself, which leads to the creation of convenient and effective tools. To achieve systematic changes, it is necessary to invest only small changes in the very structure of the source code. Thus, Lisp allows you to bypass the restrictions imposed by the language. Dynamic typing, which allows for more efficient use of code compilation time. The condition processing system allows you to interactively develop code that will handle errors during execution. Also, a huge advantage of the language in computer system modeling tasks is rich and accurate arithmetic. Lisp works very well with large numbers, with their processing and calculation, which allows you to get more accurate solutions for building a model.

- The ARENA GPSS simulation system is written in the GPSS, Pc, GPSS/PC programming language. The syntax of the programming language gets a score of 1. The general purpose modeling system is designed for modeling discrete systems. This language was created precisely as a solution to modeling problems, which is why simulation programming is more concise. Another advantage is the ability to take into account many different factors bypassing restrictions.

GPSS/PC allows you to simulate continuous and discrete processes. But the Plus language is considered a low–level programming language. It extends the capabilities of GPSS, making it more flexible and open to work with other applications.

- The Simpas simulation system is written in the Simpas programming language, Pascal. The syntax of the programming language gets a score of 2. A feature of the language is the ability to implement a simulation system on different operating systems: MS DOS, Windows 95/98, Windows NT, Linux. Another advantage is a more advanced version of interaction with Pascal, which allows you to use more efficient data types: scalar and multiple (transits, drives, devices). The system also provides random value generation to improve the accuracy of the model and ease of testing.

• Functionality – this criterion is suitable for the ability to create graphs, conditions, drives, program events, the processing speed of model input data, the creation of computational and visual simulations using the system in question. To analyze the functionality, the following estimated parameters were determined:

• 1 – variety of tools.

• 1 – visual and computational graphical representation.

• 1 – interaction with the programming language directly.

The score in Table 1 is set as the sum of the parameters according to the designated criteria.

• Usability is one of the main parameters for the successful development of the system by the user. Thanks to the well-developed usability of the graphical user interface, fewer errors are allowed and the overall functionality of the system is used more successfully.

Table 1: Criterion "Functionality".

Modeling system	Evaluation	Tools
AnyLogic	3	Drives, Databases, all kinds of graphs, Correlation Analysis, Drag&Drop, IDE, GIP, Video simulation result
Maxima	2	There is no possibility of graphical representation
ALINA GPSS	3	There are a lot of tools, a good graphical component, there is interaction with the code
Simpas	1	Exclusively working with code

The following evaluation parameters were determined for Usability analysis:

• 1 – no graphical user interface

• 2 – the graphical user interface is inconvenient, you should study the system documentation to understand how to work with a particular model

• 3 – intuitive user interface

• 4 – intuitive graphical user interface with the ability to customize the windows of the working environment.

The graphical interface of the AnyLogic modeling system receives a score of 4 (Figure 1).

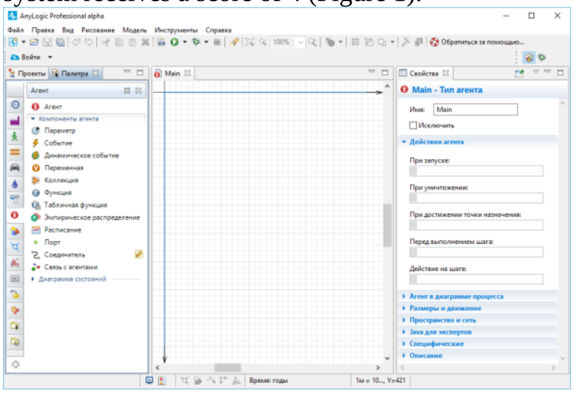

Figure 1: Graphical interface of AnyLogic modeling system.

The graphical interface of the Maxima modeling system receives a score of 2 (Figure 2).

The Simpas simulation system gets a score of 1, since it works without a graphical user interface.

• Variability in the use of the graphical interface of the program – allows you to work with all the resources offered by the simulation system, which leads to a more convenient way of compiling the report and its further information purity. To analyze the variability of the use of the graphical user interface, the following estimated parameters were determined:

• 1 – many different charts

• 1 – the ability to run a simulation of the model in a separate window

• 1 – availability of graphic space

• 1 – Drag&Drop function

• 1 – workspace separation

The score in Table 2 is set as the sum of the parameters according to the designated criteria.

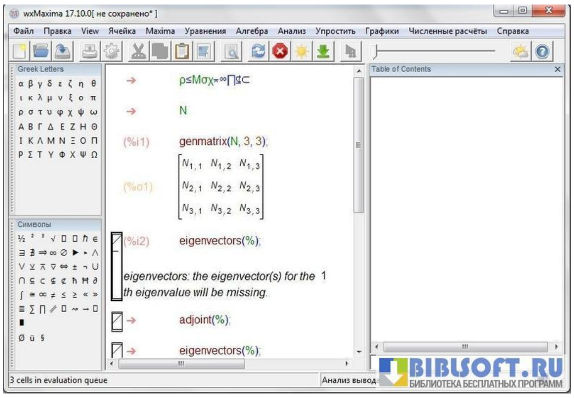

Figure 2: The graphical interface of the Maxima modeling system.

The graphical interface of the ALINA GPS simulation system receives a score of 3 (Figure 3).

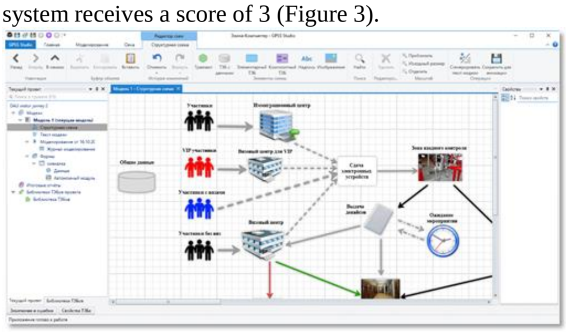

Figure 3: The graphical interface of the ALINA GPS simulation system.

• Reliability. This criterion is responsible for storing data in cloud storage, which allows you to create backups and set access for specific users authorized in the system. For the Reliability analysis, the following estimated parameters were determined:

• 1 – it is not possible to create cloud storage. Everything is stored on local media.

• 2 – it is possible to store data on the cloud, for authorized users

• 3 – it is possible to create both private and public access

The assessment according to the criterion of "Reliability" is presented in Table 3.

Table 2: Criterion "Variability of the use of the graphical user interface".

Modeling system	Evaluation	Opportunities
AnyLogic	5	Excellent graphic structure. The ability to model not only computational data, but also to create animations, support for working with objects
Maxima	2	There is no possibility to run the simulation in a separate window, there is no separation of the workspace, the Drag&Drop function does not work well
ALINA GPSS	4	The separation of the workspace is limited
Simpas	1	Without having a graphical user interface, it still has many graphical representations of the model

Table 3: Criterion "Reliability".

Modeling system	Evaluation	Cloud Characteristics
AnyLogic	3	AnyLogic Cloud is a software solution for cloud storage and the launch of simulation projects
Maxima	1	-
ALINA GPSS	2	If there are several licenses that include corporate servers, it is possible to create a single cloud system to increase the reliability of data storage
Simpas	1	-

• Multi-user mode allows users to work remotely on one model: save overall progress, pull up new changes, resolve merge conflicts, and so on. This option is not as mandatory as it is convenient and increases productivity for working in a team. The score according to the criterion "Multi–user mode" is presented in Table 4 and, in fact, consists of two parameters, where 2 is a multi-user mode, and 1 is not.

Thus, the results of the comparative analysis are presented in Table 5.

Table 4: Criterion "Multi-user mode".

Modeling system	Evaluation	Multiplayer mode
AnyLogic	2	Using AnyLogic Cloud to create, view, and run models together
Maxima	1	
ALINA GPSS	2	If you have several licenses that include corporate servers, it is possible to create a single cloud system for collaboration
Simpas	1	

Table 5: Results of the analysis of simulation modeling systems.

Criteria/System	AnyLogic	Maxima	ALINA GPSS	Simpas
Programming language	3	2	1	2
Functionality	3	2	3	1
Usability	4	2	3	1
Variability in the use of the graphical user interface	5	2	4	1
Availability	3	4	3	1
Reliability	3	1	2	1
Multiplayer mode	2	1	2	1
total	23	14	18	8

Based on Table 5, the following results are summarized:

• the most suitable of the systems under consideration is the AnyLogic software product;

• AnyLogic combines reliability and multi-user access due to the use of cloud storage technology, the possibility of free access for educational purposes;

• the concise and intuitive interface of the product opens up wide opportunities for novice developers.

It is also worth noting the possibility to use AnyLogic to solve various simulation tasks: financial models, simulation of the spread of viral infections, etc. The AnyLogic modeling system has a huge list of effective tools.

3 RESULTS OF THE STUDY

To build a model, it is necessary, as described earlier, to form a data set for the study. The described set is necessary as a training and comparative information resource that will be used in the process of the simulation modeling system. For the study below, the data were obtained from open sources on the Internet, namely from the official website of the Federal Service for Supervision of Consumer Rights Protection and Human Welfare of the Russian Federation. Weekly, the specialists of this state organization publish a bulletin containing information on the number of cases of viral infections during the past seven days, broken down into separate types and types of health problems. Unfortunately, these data were not combined into a single tabular structure, therefore, for modeling, work was done to structure the information into a single database for subsequent use.

The model was built in the free version of AnyLogic Learning Edition as a test task to check the usability of the program and evaluate the output data. The graphical representation of the model is shown in Figure 4.

This test model considers the incidence of influenza in the territory of the Russian Federation. The calculations use data from various official sources (for example, the percentage of the vaccinated population). The model takes into account all types of influenza, without dividing into strains and groups.

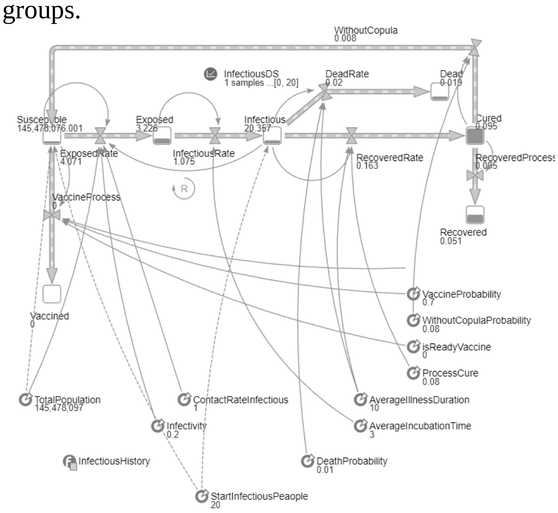

Figure 4: Graphical representation of the influenza incidence model on the territory of the Russian Federation.

A brief description of the key variables used in the model:

• Population (TotalPopulation) - 145,478,097

• Rating of contact between people in one day (ContactRateInfectious) – this parameter is calculated in the range from 0 to 1;

• Probability of transmission of infection (Infectivity) – this parameter takes values from 0.1 to 0.2;

• Incubation period (AverageIncubationTime) – lasts 3 days;

• The average duration of the disease (AverageIllnessDuration) is 10 days;

• The probability of a fatal outcome (DeathProbability) – takes values from 0 to 0.01 - remains

• The probability of recovery from infection WITHOUT the production of antibodies is 0.08;

• The probability of successful vaccination (VaccineProbability) is 0.7;

• Vaccination rating among the population (Vaccination) – takes values from 0 to 0.00028

The development of influenza was analyzed for the period from 34 to 52 weeks of 2021, as the largest increase in the number of cases over the year was observed in this time period.

4 CONCLUSIONS

From the model presented above, one can observe the operability and functionality of the AnyLogic software product. The graphical interface is completely intuitive, which allows even a novice developer to learn how to work in the system without any problems. Anylogic has the ability to build time graphs reflecting the change in a set of parameters over time (for example, the growth and decline in the incidence of infection). Settings of parameters and drives are made as accurately as possible, which contributes to a high level of validity of the model. It was verified that the data obtained as a result of the simulation correlate with the actual number of cases, based on the statistics of the incidence of influenza. It is important to mention that not all existing simulation modeling systems are considered in this paper, since their number does not allow even mentioning each one within the framework of a scientific article. However, from the point of view of the objectives of this study, the available products are sufficient to make an objective and balanced conclusion about the most suitable systems. Based on the purpose outlined in the introduction, we can say that the most suitable program for simulation modeling, which can be used to create layouts of various processes, systems and patterns, is AnyLogic.

REFERENCES

Tkachenko, A., Lavrentev, D., Denisenko, M., Kuznetsova, V., 2021. E3S Web of Conferences, 270, 01003.

Liying, C., Xiaohui, S., Yueling, Zh., Guifen, Ch., 2013. *Mathematical and Computer Modeling,* 58, pp. 507-513.

Tarik, R., 2017. We Create a Neural Network: Trnsl. from English, 272.

Chesalin, A.N., Grodzensky, S.Ya., Van, T.F., Nilov, M.Yu., Agafonov, A.N., 2020. *Russ. Technol. J.*, 8 (6), pp. 167-183.

Bondarenko, G.G., Andreev, V.V., Stolyarov, A.A., Chukhraev, I.V., Tkachenko, A.L., 2001. Fizika i Khimiya Obrabotki Materialov, (4), pp. 94–99.

Kashirina, I.L., Demchenko, M.V., 2018. p. 10 VSU.

Rashka, S., 2017. Python and machine learning. Translator: Logunov, A.V., Movchan, D.A. (eds.), p. 408.

Esmaeili Gookeh, M., Tarokh, M.J., 2017. J. Ind. Eng. Manage. Stud. Summer Autumn, 24 (2), pp. 85-102.Working paper.

Tkachenko, A., Lavrentev, D., Denisenko, M., Kuznetsova, V., 2022. SHS Web of Conferences 141, 05001.

Shih, Y.Y., Liu, C.Y. J., 2003. Database Mark Cust. *Strategy Manage.*, 11 (2), pp. 159-172.

Abrukov, V.S., Karlovich, E.V., Abrukov, S.V., 2011. 2010 Research Bulletin of the Australian Institute of High Energetic Materials, 2, pp. 129-144.

Bondarenko, G.G., Andreev, V.V., Stolyarov, A.A., Tkachenko, A.L., 2002. *Vacuum,* 67(3-4), pp. 617–621.

Abrukov, V.S., Troeshestova, D.A., Abrukov, S.V., Karlovich, E.V., Polykarpov, A.I., 2014. Conference Paper (electronic format): Tenth Int. Symp. on Special Topics in Chemical Propulsion (10-ISICP) At Ensma-Poitiers (France), p. 21.

Prokopenko, O., Larina, Y., Chetveryk, O., Kravtsov, S., Rozhko, N., Lorvi, I., 2019. *International Journal of Innovative Technology and Exploring Engineering,* 8 (12), pp. 4982-4987.

Cuesta, H., Kumar, S., 2016. Practical Data Analysis. Packt Publishing Ltd, Birmingham.

Tkachenko, A.L., Mazin, A.V., Tyurina, M.M., 2023. Proceedings of SPIE - The International Society for Optical Engineering, 12564, 125640R.

Data Science & Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data. EMC Education Services. Wiley, Indianapolis, 2015.

Gorbatiuk, K., Mantalyuk, O., Proskurovych, O., Valkov, O., 2019. Proceedings of the 6Th International Conference on Strategies, Models and Technologies of Economic Systems Management (SMTESM 2019), Pp. 271–276.

Gruzdev, A., 2018. Prognoznoe modelirovanie v IBM SPSS Statistics i R: metod derev'ev reshenij [Predictive Modeling in IBM SPSS Statistics, R and Python: decision trees and random forest method], 274 p. DMK Press, Moscow.

Katsko, I.A., Paklin, N.B., 2009. Praktikum po analizu dannyh na komp'yutere [Workshop on Data Analysis on a Computer: Training manual for universities] 276 P. Koloss Gorelova, G.V. (ed.), Moscow, In Russian.

Nilsen, E., 2021. Prakticheskij analiz vremennyh ryadov. Prognozirovanie so statistikoj i mashinnoe obuchenie [Practical analysis of time series: forecasting with statistics and machine learning]. Transl. from Eng. Dialektika, St. Petersburg.

Beksultanova, A.I., Tkachenko, A.L., 2023. Proceedings of SPIE - The International Society for Optical Engineering, 12564, 1256410.

Plas, J., 2021. Python dlya slozhnyh zadach, nauka o dannyh: i mashinnoe obuchenie [Python for complex problem solving: Data science and machine learning]. Transl. from Eng. 572 p. Peter, St. Petersburg.

Russell, S., Norvig, P., 2021. Iskusstvennyj intellekt. Sovremennyj podhod [Artificial Intelligence: A Modern Approach], 4th edn. Transl. from Eng. Volume 1. Problem solving: knowledge and reasoning. Dialektika, St. Petersburg.

Wade, R., 2021. Analitika v POWER BI s pomoshch'yu R i Python [Advanced Analytics in Power BI with R and Python: Ingesting, Transforming, Visualizing]. Transl. from Eng. DMK Press, Moscow.

Abdollahi A., Behzadi S., 2022. *Journal of Racial and Ethnic Health Disparities.*