National Research University Higher School of Economics

as a manuscript

Pokrovskaia Olga Dmitrievna

Algorithms of Computer Research of Random Processes and Structures in Natural Sciences

Dissertation Summary

for the purpose of obtaining academic degree
Doctor of Philosophy in Applied Mathematics

Academic Supervisor:
Doctor of physical and mathematical sciences
professor Shchur L.N.

Moscow – 2023

Research relevance

Computer modeling is one of the most important modern method used in natural and exact sciences, such as mathematics, physics, biology, astronomy. Computer modeling allows you to obtain reliable results when it is not possible to use analytical formulas, to study physical phenomena without having a laboratory, to develop theories that can then be tested on real objects.

The dissertation is devoted to the application of computer modeling to solve various problems. Three chapters of the dissertation are devoted to the following tasks: development of a numerical algorithm of a stochastic process for solving a two-dimensional Laplace equation with a boundary condition, development of a numerical algorithm for generating random structures and numerical verification of the selected statistical criterion.

The random growth structures are studied in the first chapter of the dissertation. This is a special class of objects that appear as a result of various physical phenomena [1]. For example, the growth of native copper in a rock, the formation of snowflakes - ice crystals [2], breakdown in the dielectric [3]. Mathematical modeling of growth processes is one of the main tool for the study of such objects. To study the properties of such structures, it is necessary to simulate hundreds of millions particles [4]. When modeling a large number of particles, several problems arise. Firstly, the process of modeling takes much more time. In addition, other parameters are affected, such as the step size of the wandering particle and the number of particles in the numerical experiment. Thus, the development of effective methods for modeling the wandering of particles is an urgent task.

Another important practical task is the development of elements for storing hydrogen atoms in hydrogen energy. Currently, the most promising for solving this problem are foamed materials of palladium and platinum [5]. Palladium foam stands out among other materials by the quality of the pumping cycle and the release of hydrogen, which may be due to the presence of experimentally detected ferromagnetic properties of palladium foams. Palladium itself is not a ferromagnet, but there are signs that its low-dimensional forms may exhibit ferromagnetic behavior. In the second chapter of the dissertation research, mathematical modeling of foam-like structures, such as palladium foam, and the study of the magnetic properties of the resulting model are carried out. The development of algorithms for generating foam-like structures for further super-computer modeling is an urgent task for hydrogen energy.

Computer modeling is also widely used in genetics. For example, studying the effects of mixing different populations is an important task. This can help to understand how changes occur in the genetic material of people after mixing populations, and how this affects adaptation to the environment. The detection of signs of positive selection in populations around the world helps to reveal the evolutionary history of humans, as well as the genetic basis of various diseases [6]. Most human evolutionary genomic studies have been conducted for European, African and Asian populations. However, the population with a native American origin has not been sufficiently studied [7]. Of particular interest is the study of the Chilean population after the Columbian exchange, because it is after the discovery of America by Columbus that the African and European components appear in the genes of Chileans, which can give new evolutionary advantages. In the work, the presence of natural selection for some haplotypes after the Colombian exchange in the indigenous population of Chile was verified using computer modeling.

Problem statement

With the help of computer modeling, it is possible to solve various classes of problems in many areas. In the dissertation research we solve three problems:

1. For a two-dimensional Laplace equation with a boundary condition, estimate the first correction for the probability of the first intersection. Based on the conducted research, we propose more efficient random walk algorithms on the plane, working many times faster, without loss of accuracy.

2. To study foam-like structures, we propose an algorithm for generating foam on a plane and in 3-D space, as well as to study the magnetic properties of the obtained structures at low temperatures using the Ising model.

3. By using computer modeling and real genotype data of Chileans, to verify that some haplotypes associated with adaptation to climatic conditions and protection from diseases appeared in Chileans as a result of natural selection after mixing populations.

Purpose of research

1. To develop effective methods of computer simulation of random structures growth that represent the solution of the two-dimensional Laplace equation and to evaluate the effect of the finiteness of the step on the simulation result.

2. To develop effective methods of computer modeling of random foam-like structures on the plane and in 3-D space. To investigate the magnetic properties of such structures at low temperatures using the Ising model.

3. To confirm with the help of computer modeling the presence of signs of natural selection in the Chilean population after mixing.

Research objectives

1. Development of an effective algorithm for modeling the first intersection of a circle of a fixed radius with a random walk of a particle on a plane.

2. Estimation of the finiteness effect of a step in a random walk in computer modeling.

3. Development of an effective algorithm for generating a foam-like structure on a plane and in 3-D space.

4. Study of the magnetic properties of the modeling foam-like structure on the plane at low temperatures.

5. The study using computer modeling of the genes features of native Chileans after mixing European and African genes as a result of the Colombian exchange.

The degree of the problem development

Computer modeling is widely used to study the growth of self-similar structures. One of the methods of constructing fractals is to perform a special recursive procedure. In the article [1], an aggregation algorithm of DLA (Diffusion Limited Aggregation) was proposed. The main idea of the algorithm is that the growth of a cluster on a square lattice occurs as a result of sticking randomly moving particles on a growing structure.

At work [4] an algorithm was proposed for wandering a particle on a plane without a lattice, in which a particle, when removed some distance from a growing cluster, is not destroyed, but with a certain probability returns to the radius of birth and continues to wander. The probability of returning to the radius of birth is derived from the Laplace equation, the relationship between the probability of the first intersection of the absorbing circle during a random walk and the distribution of the electric potential described by the Laplace equation is known [4]. The fractal dimension of the resulting object was also clarified in the article. Avoiding modeling on the grid allows you to get more accurate modeling, but at the same time the modeling speed decreases. As a result, in order to get a more accurate value of the angle at which the particle intersects the circle of a given radius, the particle needs to move with a very small step, so the simulation time increases rapidly.

Computer modeling is often used to study the properties and features of new materials. In the modern world, the question of alternative energy sources is increasingly being raised. Electricity generated by hydrogen installations is considered as an alternative energy source. Fuel cells are used to store hydrogen, and in them hydrogen is stored not in a gaseous or liquid state, but in the form of chemical compounds with other elements. One of the promising materials for fuel cells is foamed palladium. At work [5] the authors, as a result of the experiment, obtain a foamed structure of palladium wires, with a controlled density. The authors also measure the mechanical strength properties of such a foam structure.

At work [8] measurements of the differential conductivity of palladium nanowires have been carried out in order to investigate the nature of Fano resonances detected on nano-contacts with ferromagnets such as, for example, nickel. The article [9] provides an overview of the modern understanding of the mechanics of random fiber networks.

It was not possible to find in the scientific literature works on computer modeling of foam-like structures and the study of the magnetic properties of such structures. Computer research is limited to the study of the magnetic properties of atomic chains.

Now the topic of human genome research has become especially relevant. Often in such tasks, different mathematical models are used together with computer modeling. The detection of a positive natural selection signal in populations around the world helps to reveal the evolutionary history of humans, as well as the genetic basis of various diseases and other phenotypic traits [6]. Most human evolutionary genomic studies have been conducted for European, African and Asian populations. However, the population with Native American ancestry has not been sufficiently studied [7]. This is partly due to: 1) the lack of publicly available genomes of several Native American populations and their mixed Latin American descendants; 2) the technical difficulty of identifying genetic factors specific to the origin in mixed populations with different continental ancestors and a complex demographic history. Unlike Peruvians, Colombians, Mexicans and Puerto Ricans, which have publicly available whole genome sequencing [10], there is no such data for Chileans. The results of this study may help to better understand how changes occur in human genetic material after mixing populations and how this affects adaptation to the environment and protection from diseases.

Scientific novelty

1. For the first time, an effective algorithm is proposed for modeling the random walk of a particle on a plane and its adhesion to a circle of a fixed radius, with a variable particle pitch.

2. Based on the numerical calculations carried out, an analytical expression for the dependence of the modeling accuracy on the particle step is proposed for the first time.

3. An algorithm for generating two-dimensional and three-dimensional foam-like structures with typical properties of experimentally observed structures has been developed for the first time.

4. A model of the magnetic properties of a foam-like structure is proposed and investigated for the first time.

5. The features of the recently sequenced genes of indigenous Chileans for the first time after mixing European and African genes as a result of the Colombian exchange have been studied.

Summary of the work

Introduction. The introduction shows the scientific novelty of the research, the goals and objectives are formulated, the provisions submitted for defense are presented. The relevance and practical significance of scientific work is substantiated.

Chapter 1. Optimal random walk algorithm on the plane to determine the first intersection.

The chapter proposes a new efficient random walk algorithm on the plane and an estimate of the contribution of the step size to the error in computer simulation of the walk. In science , there is a known relationship between the probability of the first intersection of the absorbing circle in the process of a random walk and distribution of the electric potential described by the Laplace equation [1]. A random walk on the plane starts from a point with coordinates (R_b, 0). The radius of the sticking circle is R, $R_b > R$. Then the probability of the first intersection as a result of a random walk at an angle of ϕ is expressed by the formula [4]:

$$P(\phi) = \frac{1}{2\pi} \frac{x^2 - 1}{x^2 - 2x\cos\phi + 1}, \tag{1}$$

where $x = R_b/R > 1$. From the point of view of analytical calculations, the particle wanders along the plane with an infinitesimal step, and always with one hundred percent probability will cross the circle of adhesion. For computer simulation, the step of a particle will always be finite and it is impossible to simulate an infinite process of wandering, waiting for the particle to cross the circle. To speed up the numerical simulation, an additional circle with a radius of R_r is introduced - the return radius. When moving beyond this circle, the particle forcibly returns to the radius of birth R_b, according to the formula [4]:

$$\phi = f(u) = 2\arctan\left(\frac{x-1}{x+1}\tan u\frac{\pi}{2}\right), \tag{2}$$

where $x = R_r/R_b > 1$, u is a random variable uniformly distributed over the interval $[-1, 1]$.

The simulation was carried out with the parameters: $R = 10$, $R_b = 20$, $R_r = 200$. When modeling a walk with different step sizes, it became clear that the dependence of the modeling accuracy on the step size is not linear and with a further decrease in the step, the accuracy ceases to grow proportionally [11]. Based on the analysis of the dependence of the deviation of the numerical result from the exact (1), a formula was proposed for the dependence of the modeling accuracy on the step, where δ is the step size:

$$P_{exp}(\phi) \approx P(\phi) \left(1 - \left(\frac{\delta}{R}\right)^{\alpha} \cos\phi\right), \tag{3}$$

where $\alpha < 1$. The second term can be considered an amendment to the exact expression due to the discretization of the walk. An analytical solution has not yet been found.

To speed up the modeling process, we have proposed two algorithms for changing the step of a random walk. Three algorithms were compared:

1. The particle pitch is unchanged over the entire interval from the birth circle to the adhesion circle.

2. The step of the particle changes when crossing a certain radius of step size change chosen by us.

3. The particle pitch depends linearly on the distance between the particle and the origin.

For comparison, the time for modeling is taken by the usual algorithm (Algorithm 1) with a step of 0.1. Algorithm 3 turned out to be the most efficient, about 27,000 times faster than Algorithm 1.

It was found that the contribution to the statistical error is given by the number of particles in the simulation, the contribution to the systematic error is given by the particle step, that is, the accuracy of the simulation result depends on the step. The best algorithm proposed is an algorithm with a step change according to a linear law. This algorithm is the fastest and most accurate of those considered, and allows you to refine the solution with a given accuracy in a reasonable time.

Chapter 2. An algorithm for generating a foam structure on a plane and in space.

The chapter discusses an algorithm for modeling a foam-like structure on a plane and in space, and also studies the magnetic Ising model for the resulting structure on a plane. A two-dimensional foam structure can be represented as randomly scattered segments of fixed length. An object constructed in this way has some analytically derived properties [12] [13].

At work [14] the properties for a set of randomly located infinite lines on a plane intersecting with each other are considered. As a result of such intersections, polygons are obtained - closed polygons. The article presents analytical calculations for the average perimeter and the average area of such polygons, depending on the density of the lines. Average Perimeter $E(Perimeter) = \frac{2\pi}{\tau}$, the average area $E(Area) = \frac{\pi}{\tau^2}$, where τ is the analog of density. The article [15] shows how to express the parameter τ through the new parameter k, and how to calculate the parameter k. The parameter k determines the density of random lines, this is the average number of lines crossing any straight segment of unit length. Thus, $1/k$ is the «average free length» of the line, and $\tau = \frac{\pi k}{2}$. For our model, we counted k and polygon area values and checked with what accuracy we would get the number π.

To test our model, we increased the length of the segments in the box so that it was larger than the size of the box and investigated the asymptotic dependence with an increase in the size of the segments and the density of the segments in the box. It was verified that for our model, the area and perimeter of polygons in asymptotics coincide with those calculated in the article, which means that the model can be used for further research.

An algorithm for generating a foam-like structure on a plane was proposed:

1. Let's set the size of the box L and the size of the segments l.

2. Generate segments on the plane: to do this, randomly generate the coordinates of the origin point (x,y) inside the box, and the angle $\phi \in [0, 2\pi]$. Next, from this point, at the selected angle, we measure the length of l. The segment may extend beyond the box, in which case it will be cut off.

3. When segments intersect with each other, clusters are formed. The generation of segments stops when a cluster appears that has intersections with all sides of the box.

4. All clusters, except the only one that has intersections with all sides of the box, are deleted with all segments included in them.

5. Segments that have only one intersection with the remaining cluster are deleted. This «washing» is repeated until all such segments are removed.

The chapter describes in detail the data structures used for effective cluster modeling. We used the proposed algorithm to study the properties of foam structures.

Next, the magnetic properties of the model obtained earlier are studied. For this purpose, a mathematical model of statistical physics is used, which is used to describe a ferromagnetic material - the Ising model.

On each segment in the cluster, at a distance of 1 from each other, we place the variables σ taking the values -1, +1. Energy of the system:

$$E = -J \sum_{\{i,j|distance(i,j) \leq R\}} \sigma_i \sigma_j, \tag{4}$$

where J is the energy of the exchange interaction, R is the radius of interaction. Summation in this expression is performed for all pairs of spins σ_i and σ_j, the distance between which does not exceed the value of R.

During the simulation, we will calculate the following values of the thermodynamic system:

- Magnetization:
$$M(t) = N_{up} - N_{down};$$
- Heat capacity:
$$C = \frac{\langle E^2 \rangle - \langle E \rangle^2}{(kT)^2};$$
- Magnetic susceptibility:
$$\chi = \frac{\langle M^2 \rangle - \langle M \rangle^2}{kT}.$$

Here, the angle brackets denote the time averaging. The simulation steps are taken as the time.

Two algorithms were used for modeling.

The Local Metropolis Algorithm [16]:

1. We count the energy of the entire system.
2. Choosing a random spin.
3. We flip the selected spin and recalculate the energy of the system in an optimal way. If the energy of the system has decreased or has not changed, then we take

a step (change the sign of the selected spin). If the energy of the system has increased, we take a step with probability $e^{-(E_{new}-E_{old})/T}$.

Wolf's single-cluster Algorithm [17]:

1. Choosing a random spin.

2. Select all the neighbors of the selected spin at a distance of R having the same spin direction. Communication with such neighbors will exist with probability $p = 1 - e^{-2K}$, where $K = \frac{J}{kT}$.

3. If there is a connection with a neighbor, select this neighbor, and look at its neighbors (at a distance of R, having the same direction of spin), build connections with probability $p = 1 - e^{-2K}$. We do not re-consider those spins that have already fallen into our cluster.

4. Repeat the previous step until no new links are created.

5. Changing the direction of spins in the entire cluster to the opposite.

6. Starting the algorithm from the beginning.

Simulations were carried out for different values of R (interaction radius) and T (temperature). It is found that at low temperatures the behavior of magnetization is qualitatively similar to the experimental results for three-dimensional foam [18]

An algorithm for generating a three-dimensional foam-like structure is also proposed.

The experimental study of the magnetic properties of foam-like structures takes place in three-dimensional space on real materials. Therefore, an algorithm for generating a structure in space is of particular interest. The difficulty lies in the fact that if you just randomly generate segments in space, they will intersect with each other with negligible probability. Therefore, we need an algorithm that will allow us to obtain clusters of segments that intersect with each other.

Let's set the size of the box LxLxL and the size of the segment l.

1. We generate three coordinates inside the box (x, y, z) - these are the coordinates of the beginning of the segment. We generate a random direction of the segment using the Box-Muller method [19].

2. We generate a random plane passing through the segment, for this the angle α is generated - the angle between the xOy plane and the plane containing the segment.

3. In the plane from the previous step, centered at the beginning of the segment, a circle with radius l is drawn.

4. Searches for all points of intersection of segments with the plane that fall inside this circle.

5. We find the angles formed by the segment and the rays from the beginning of the segment to these points.

6. Select the minimum angle by the absolute value of ψ. We rotate the segment relative to its beginning by this minimum angle. As a result, we get the intersection of two segments.

7. We cut off the segment if some part of it has gone beyond the box. We remember which side the segment touched.

8. If a segment intersects with another one, it falls into an already existing cluster, if there are no intersections, then it creates a cluster in which it is alone.

9. We return to step 1 until the end condition of the algorithm is met: the presence of a cluster of segments intersecting all six sides of the box.

Examples of structures that are built according to this algorithm are given.

Chapter 3. Search for natural selection in the Chilean population by predicting the local origin of genomes.

The chapter examines the Mapuche Chilean population after the discovery of America by Columbus (the Colombian Exchange). Chileans have been found to have haplotypes associated with pigmentation, thermogenesis, and immune defense against pathogens that have been subjected to natural selection. These haplotypes were acquired by Chileans as a result of mixing with Indians, Europeans and Africans after the Colombian exchange. The genomes reconstruct the population and evolutionary history of the Chilean population. As a result of computer modeling, signs of PAS (The Post-admixture Selection, selection after admixture) were revealed in the data used in the following genome regions: the strongest signal of natural selection refers to the haplotype, including rs12821256, which regulates expression on KITLG. rs12821256 is the cause of lighter hair color in northern Europeans, KITLG regulates the amount of melanin pigments in hair follicles and in the skin, and is also involved in thermogenesis.

Using the SELAM [20] modeling tool, a population of Chileans with different proportions of European and African components was generated and dynamics for 10, 12 and 15 generations was launched. It was shown that the selected sections of the genome could not be obtained during a random process and natural selection took place.

Conclusion. In the conclusion of the dissertation, the general conclusions and the main results of the research are formulated.

The main results submitted for defense:

1. An efficient algorithm with an adaptive step has been developed for modeling a random walk on a plane followed by sticking to a circle of a fixed radius.

2. The error introduced by the step size of a randomly wandering particle is estimated based on the results of a large number of computer experiments. A possible form of the analytical formula of the first-order correction to the exact result of solving the Laplace equation is proposed.

3. An algorithm has been developed for generating foam-like structures on the plane and in space. The correspondence of the geometric characteristics of the simulated structures with analytical calculations is verified.

4. The values of physical quantities obtained using the Ising model on a flat foam-like structure were compared with experimental ones. It is shown that the values are qualitatively similar.

5. The population of indigenous Chileans was studied after mixing European and African genes. Sections of the genome with signs of natural selection have been found.

Personal contribution of the author to the development of the problem

The author personally implemented computational algorithms, participated in discussions on research areas, studied the scientific literature in the field of research. The author was engaged in the preparation, analysis, processing and discussion with the supervisor of the results obtained. Together with the supervisor, ideas and hypotheses were proposed, tasks were set.

Approbation of research results

1. «Dependence of the results of modeling the random walk process on the final step», annual interuniversity Scientific and Technical conference of students, postgraduates and young specialists named after E.V. Armensky, 19 February - 1 March 2018.

2. «Algorithms with variable step size for modeling the random walk process on the plane», annual interuniversity scientific and technical conference students, postgraduates and young specialists named after E.V. Armensky, 18-28 February 2019.

3. «Algorithm for generating segments on a plane for modeling two-dimensional foam», annual interuniversity scientific and Technical conference of students, postgraduates and young specialists named after E.V. Armensky, 25 February - 4 March 2020.

4. «Method for foam generation in plane» at the international Conference International Conference on Computer Simulation in Physics and beyond, October 12-16, 2020, Moscow, Russia

5. «Modeling of magnetic properties of two-dimensional metal foam» at the international conference Supercomputer Days in Russia on September 27-28, 2021

Certificate of state registration of software:

- Certificate of state registration of the computer program No.2018665832 «Random walk with variable step and first border crossing», 2018

Author's publications on the topic of the dissertation

All publications are included in the Scopus international citation system.

1. Klimenkova O., Menshutin A., Shchur L. Variable-step-length algorithms for a random walk: Hitting probability and computation performance // Computer Physics Communications. 2019. Vol. 241. P. 28-32.

2. Klimenkova O., Menshutin A., Shchur L. Influence of the random walk finite step on the first-passage probability // Journal of Physics: Conference Series. 2018. Vol. 955. No. 012009. P. 1-6.

3. Klimenkova O., Shchur L. Algorithm for foam generation in plane // Journal of Physics: Conference Series. 2021. Vol. 1740. No. 012030. P. 1-5.

4. Vicuña, L., Klimenkova, O., Norambuena, T., Martinez, F. I., Fernandez, M. I., Shchur, V., Eyheramendy, S., Postadmixture selection on Chileans targets

haplotype involved in pigmentation, thermogenesis and immune defense against pathogens //Genome Biology and Evolution. - 2020 - 12(8) - p. 1459-1470.

References

1. T. A. Witten, L. M. Sander, Diffusion-Limited Aggregation, a Kinetic Critical Phenomenon, Physical Review Letters 47, 1400 (1981)

2. J. Nittman, H. E. Stanley, Non-deterministic approach to anisotropic growth patterns with continuously tunable morphology: the fractal properties of some real snowflakes, J. Phys.A: Math. Gen. 20, L1185 (1987).

3. L. Niemeyer, L. Pietronero, H. J. Wiesmann, Fractal Dimension of Dielectric Breakdown, Physical Review Letters 52, 1033 (1984).

4. A. Yu. Menshutin, L. N. Shchur, Test of multiscaling in a diffusion-limited-aggregation model using an off-lattice killing-free algorithm, Phys. Rev. E **73**, 011407 (2006).

5. Gilbert D. A. et al. Tunable low density palladium nanowire foams //Chemistry of Materials. – 2017. – T. 29. – №. 22. – C. 9814-9818.

6. Nielsen, R., Akey, J. M., Jakobsson, M., Pritchard, J. K., Tishkoff, S., Willerslev, E. Tracing the peopling of the world through genomics //Nature. – 2017. – T. 541. – №. 7637. – C. 302-310

7. Cheng JY, Racimo F, Nielsen R. 2019. Ohana: detecting selection in multiple populations by modelling ancestral admixture components. BioRxiv.

8. M.S. Islam, et al. Fano profiles in palladium nanoconstrictions // Solid State Communications – 2017. – T. 262. – C. 16.

9. Picu, R. C. Mechanics of random fiber networks—a review. Soft Matter, 7(15), 6768.(2011)

10. 1000 Genomes Project Consortium, A global reference for human genetic variation. //Nature. - 2015 - 526(7571), p.68-74.

11. Klimenkova O., Menshutin A., Shchur L. Influence of the random walk finite step on the first-passage probability // Journal of Physics: Conference Series. 2018. Vol. 955. No. 012009. P. 1-6.

12. S. Goudsmit, Random distribution of lines in a plane //Reviews of Modern Physics. – T. 17. – №. 2-3. – C. 321 (1945)

13. R. E. Miles The various aggregates of random polygons determined by random lines in a plane //Advances in Mathematics. – T. 10. – №. 2. – C. 256-290 (1973)

14. R. E. Miles, Random polygons determined by random lines in a plane, Proceedings of the National Academy of Sciences of the United States of America. T. 52.– No. 4. – C. 901. (1964)

15. Richards P. I., Averages for polygons formed by random lines //Proceedings of the National Academy of Sciences of the United States of America. – 1964. – T. 52. – №. 5. – C. 1160.

16. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E., Equation of state calculations by fast computing machines //The journal of chemical physics, - 1953. - 21(6), pp.1087-1092.

17. Wolff U., Collective Monte Carlo updating for spin systems. //Physical Review Letters - 1989 - 62(4), p.361.

18. Teng, X., Han, W.Q., Ku, W. and Hücker, M., Synthesis of ultrathin palladium and platinum nanowires and a study of their magnetic properties //Angewandte Chemie - 2008 - 120(11), pp.2085-2088.

19. Box G. E. P., Muller M. E. A note on the generation of random normal deviates //The annals of mathematical statistics. – 1958. – T. 29. – №. 2. – p. 610-611.

20. Corbett-Detig, R., Jones, M. SELAM: simulation of epistasis and local adaptation during admixture with mate choice //Bioinformatics. - 2016 - 32(19), 3035-3037.