Entomology and Applied Science Letters
Volume 7, Issue 2, Page No: 70-76

Available Online at: www.easletters.com

ISSN No: 2349-2864

Forecasting the Growth of Wheat Shoots based on Neural Networks

Svetlana Mustafina[1]*, Natalya Uspenskaya[1], Denis Smirnov[2], Denis Yashin[2] Sofia Mustafina[1], Oleg Larin[3]

[1] Dept. Mathematics and Information Technology, Bashkir State University, Ufa, Russia.
[2]Dept. Information Technology and Management Systems, K.G. Razumovsky Moscow State University of Technologies and Management (the First Cossack University), Moscow, Russia.
[3]Electric Power Department, South-West State University, Kursk, Russia.

ABSTRACT

The article is devoted to the search for optimal growth parameters of wheat shoots using neural networks. The main parameters affecting the growth of wheat shoots are determined. A training sample has been compiled for working with a neural network, and a neural network has been built. Using the forecasting method, the optimal parameters for the growth of wheat shoots were determined depending on the concentration of cadmium in the soil, soil moisture, ambient temperature, and the concentration of added bacteria *Bacillus subtilis.*

Keywords: *wheat shoots, concentration Cd, Bacillus subtilis 26Д, ambient temperature, soil moisture, neural network.*

HOW TO CITE THIS ARTICLE: Svetlana Mustafina, Natalya Uspenskaya, Denis Smirnov, Denis Yashin, Sofia Mustafina, Oleg Larin; Forecasting the Growth of Wheat Shoots based on Neural Networks, Entomol Appl Sci Lett, 2020, 7 (2): 70-76.

Corresponding author: Svetlana Mustafina
E-mail ✉ Mustafina_SA @ mail.ru
Received: 03/04/2020
Accepted: 22/07/2020

INTRODUCTION

Currently, the soil contains a large number of various heavy metals that interfere with plant growth. Heavy metals are considered the most common environmental pollutants [1-3] because of the high utilization in agriculture and industries [4]. In this regard, various methods are being developed to improve the growth of cultivated plants, such as corn, wheat and rye, and others. The most popular way of improving plant growth is to add various pesticides and microorganisms. In this case, the most productive and safe way is the use of *Bacillus subtilis*.

Biological statistics is a branch of knowledge that allows the application of statistical methods in biology to analyze features with a continuous and discrete nature of distribution, to find existing patterns, and make decisions.

Finding optimally selected factors to obtain the best result is a difficult task for an analytical solution. It is necessary to take into account a different number of parameters that may affect the final result. To solve such problems, artificial neural networks are increasingly being used, which help build the dependence of input parameters on the output parameter and find the optimal solution to the problem.

COMPILING TRAINING SAMPLE FOR TRAINING A NEURAL NETWORK

To train the neural network in scientific research, the Statistica software package developed by StatSoft was used. Statistica software implements the functions of data analysis, data management, data mining, and data visualization using statistical methods [5, 6].

When conducting experiments on the effect of the dosage of *Bacillus subtilis* 26D on plant growth in contaminated soil with different metals, an array of data was obtained [7–9]. From the data set, the main parameters were selected that have a significant effect on the growth of shoots in plants. After selecting the main parameters, a data table was created for the neural network model in the Statistica program.

Table 1 presents the input and output data for creating a neural network model. The input parameters are the concentration of cadmium in the soil (mg/kg), soil moisture (%), ambient temperature (°C), and concentration of *Bacillus Subtilis* 26D (cell/ml). The output parameter is wheat shoots (cm). Table 1 shows part of the sample [7-9].

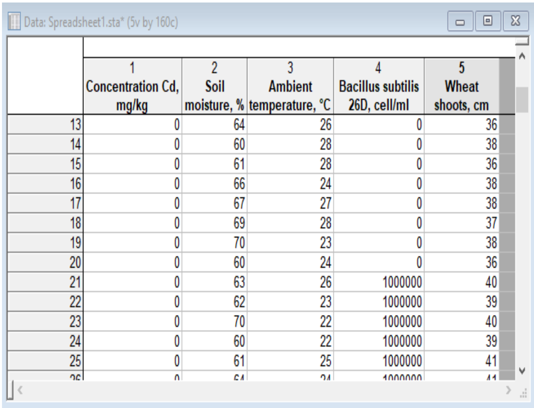
Table 1. Table with data sampling for a neural network

№	Input value				Output value
	Concentration Cd, mg/kg	Soil moisture, %	Ambient temperature, °C	Bacillus subtilis 26D, cell/ml	Wheat shoots, cm
1	0	60	28	0	38
2	0	65	25	0	37
3	0	61	24	0	36
4	0	61	26	0	37
5	0	62	27	0	38
6	0	60	27	0	37
7	0	69	23	0	38
153	500	70	26	1000000	44
154	500	65	25	1000000	45
155	500	62	23	1000000	43
156	500	61	28	1000000	43
157	500	65	23	1000000	44
158	500	69	27	1000000	44
159	500	68	28	1000000	43
160	500	69	27	1000000	44

After creating a table with input and output parameters for training, testing, and validation of a neural network, we will start training the neural network in the Statistica software package.

CREATION, TRAINING, AND TESTING OF A NEURAL NETWORK IN THE STATISTICA SOFTWARE PACKAGE

To create a table with the source data, the following steps were taken. Statistica software was launched. The New tab was selected on the main panel, after which a window appeared in which Spreadsheet was selected to create a table in Statistica. In the Create New Document tab, in the Number of Variables field, the number 5 was entered, which corresponds to the number of variables. In the Number of Cases field, number 160 was entered. Double Data was selected as the Default Data Type. The location was selected in a separate window. The image format was common. Next, the OK button was pressed. The program created an empty table with five columns Var1, Var 2, Var 3, Var 4, Var 5, and 160 rows. The completed table is shown in Fig. 1.



Data: Spreadsheet1.sta* (5v by 160c)

	1 Concentration Cd, mg/kg	2 Soil moisture, %	3 Ambient temperature, °C	4 Bacillus subtilis 26D, cell/ml	5 Wheat shoots, cm
13	0	64	26	0	36
14	0	60	28	0	38
15	0	61	28	0	36
16	0	66	24	0	38
17	0	67	27	0	38
18	0	69	28	0	37
19	0	70	23	0	38
20	0	60	24	0	36
21	0	63	26	1000000	40
22	0	62	23	1000000	39
23	0	70	22	1000000	40
24	0	60	22	1000000	39
25	0	61	25	1000000	41

Fig. 1. Filled Source Data Table.

After creating and filling the table with data, it is necessary to create, train, and test neural networks. In the toolbar in the Learning section, Neural Networks was selected. In the New Analysis field, an analysis type was selected from the submitted list to create a new Regression analysis and the OK button was pressed (Fig. 2).

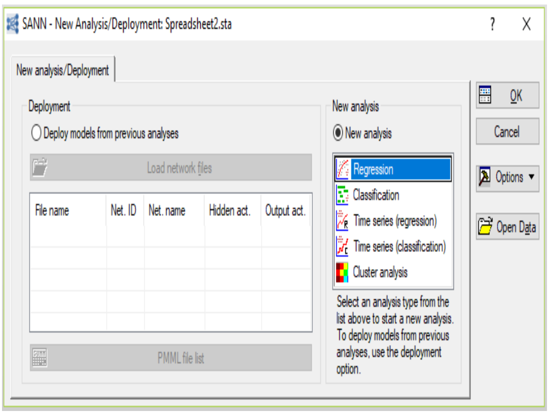


Fig. 2. Neural Network creation.

Next, a new window for configuring the neural network opened, where three tabs were presented: Quick, Sampling (CNN and ANC), and Subsampling. Quick tab was selected. In the field for Strategy for Creating Predictive Models, Automated Network Search was marked (ANS). Then, the Variables button was clicked. Next, the variable window was selected for data analysis. In the first window, the output parameter (wheat shoots, cm) was selected, in the second one the input parameters (concentration of Cd, mg/kg; soil moisture,%; ambient temperature, °C; *Bacillus subtilis* 26D, cell/ml) were selected. Below in the line "Continuous Targets", the number 5 automatically appeared which is responsible for column number of the variable, and in the line "Continuous Inputs" 1-4 next to the inscription "Show Appropriate Variables Only" is unchecked and the OK button was pressed. After that, the selected values were appeared in the Variables field in the neural network settings window (Fig. 3.).

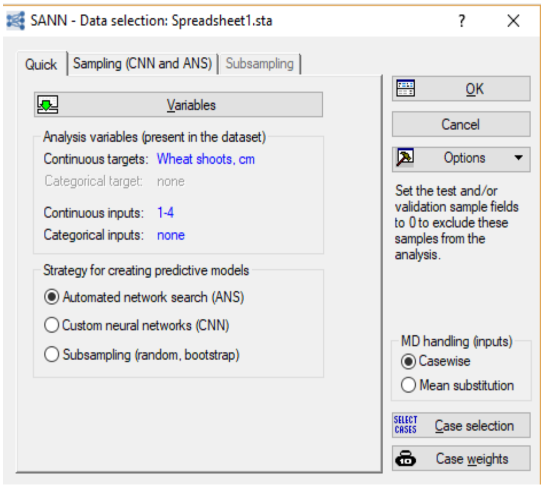


Fig. 3. The neural network architecture settings window.

Selecting the Sampling tab (CNN and ANS), in the Sampling Method group in the Train (%) field, we entered 70, Test (%) - 15, Validation (%) - 15, and Seed for Sampling - 1000 (Fig. 4.). After setting up the neural network, button OK was pressed.

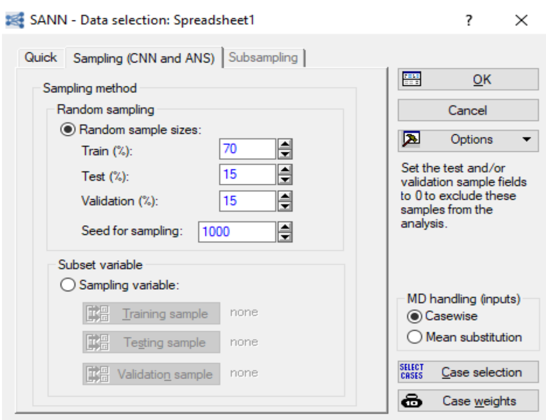


Fig. 4. Sampling (CNN and ANS) Tab.

In the Automated Network Search (ANS) tab (Fig 5.), in the Network Types parameter group, a checkmark was placed next to the inscription MLP (Multilayer Repeater).

We determined the required number of neurons in the hidden layers of the perceptron according to the formula, which is a consequence of the theorems of Arnold-Kolmogorov-Hecht-Nielsen [10, 11]:

$$\frac{N_y \cdot Q}{1+\log_2(Q)} \le N_w \le N_y \cdot \left(\frac{Q}{N_x}+1\right)\cdot(N_x+N_y+1)+N_y \quad (1)$$

$$\frac{1\cdot 112}{1+\log_2(112)} \le N_w \le 1\cdot\left(\frac{112}{4}+1\right)\cdot(4+1+1)+1,$$

$$14{,}34 \le N_w \le 175,$$

where N_y – output signal dimension ($N_y = 1$);
N_w – the required number of synaptic connections;
N_x – input signal dimension ($N_x = 4$);
Q – the number of elements in the set of training examples ($Q = \frac{160\cdot 70}{100} = 112$).

By evaluating the required number of synaptic connections Nw using this formula, it is possible to calculate the required number of neurons in the hidden layers. For example, the number of neurons in the hidden layers of a two-layer perceptron will be equal to:

$$N = \frac{N_w}{N_x + N_y} \quad (2)$$

We calculated by formula (2) the maximum and minimum number of neurons.

The minimum number of neurons is:

$$N = \frac{14{,}34}{4+1} = 2{,}87$$

The maximum number of neurons is:

$$N = \frac{175}{4+1} = 35$$

According to the received data in the Min. Hidden Units was introduced 3, Max. Hidden Units - 35. In the Train/Retain Network group, in the

Networks to Train field, we entered 20, Network to retain– 5, and select Train.

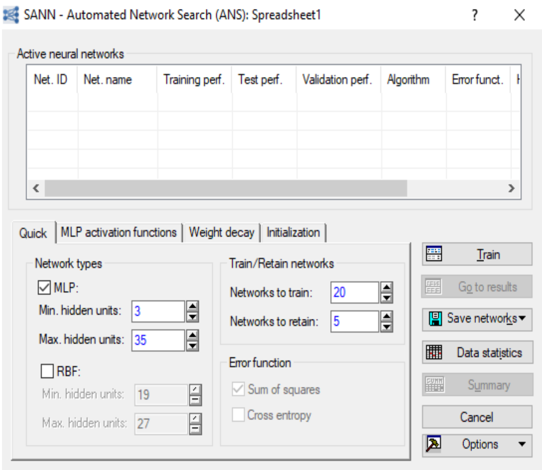


Fig. 4. Automated Network Search (ANS) Tab

After setting up the neural network, training began (Fig 5).

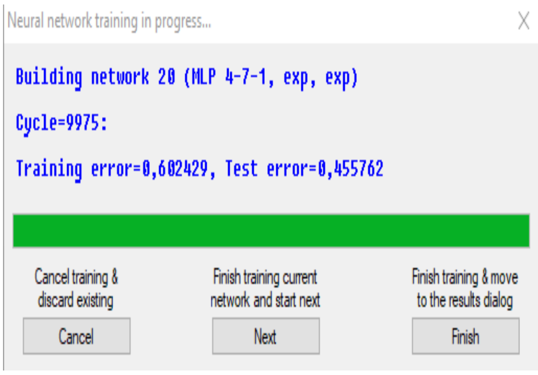


Fig. 5. Neural network training

Ready-made neural networks are shown in Fig. 6. The window presented 5 neural networks that have the smallest mean-square errors for the control and test samples.

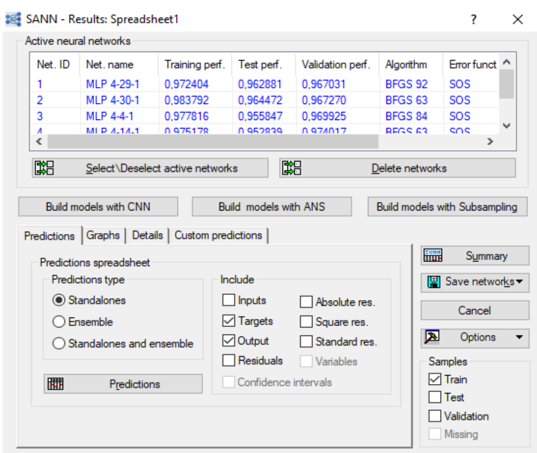
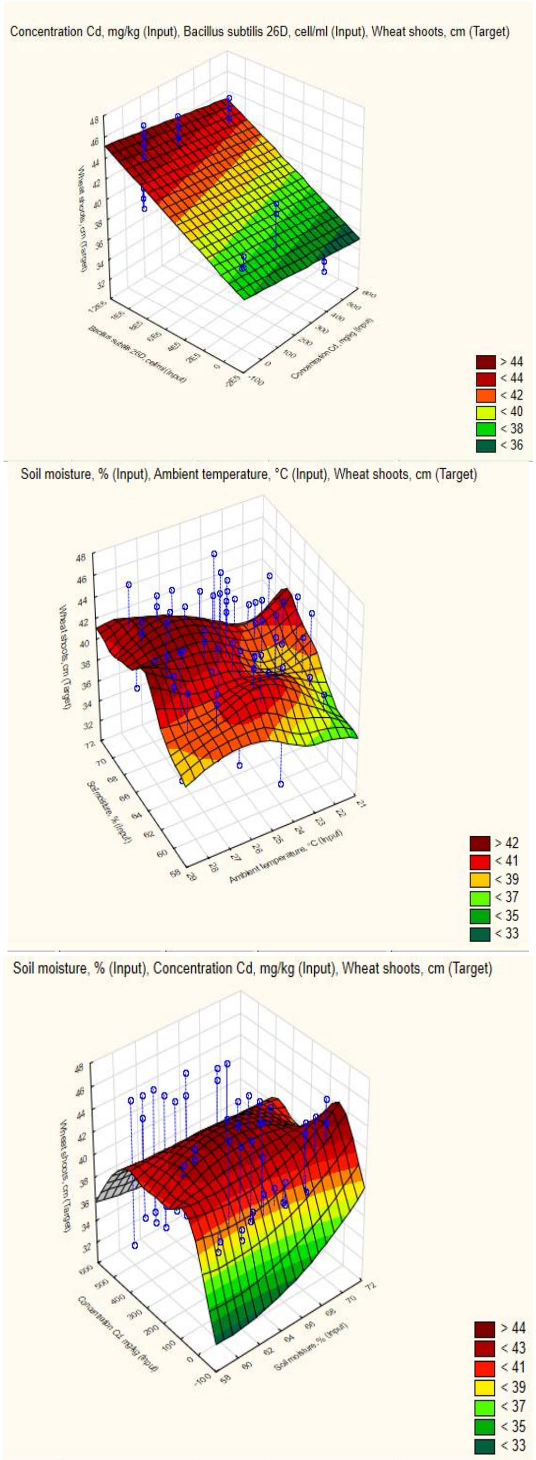


Fig. 6. Ready-made neural networks

RESULTS

Having studied the 3D histograms of the dependencies presents in Fig. 7, we can conclude that the highest growth of wheat shoots is observed at the highest content of *Bacillus Subtilis* 26D equal to 1×10^6 cell/ml, at Ambient Temperature from 25 to 29 °C.



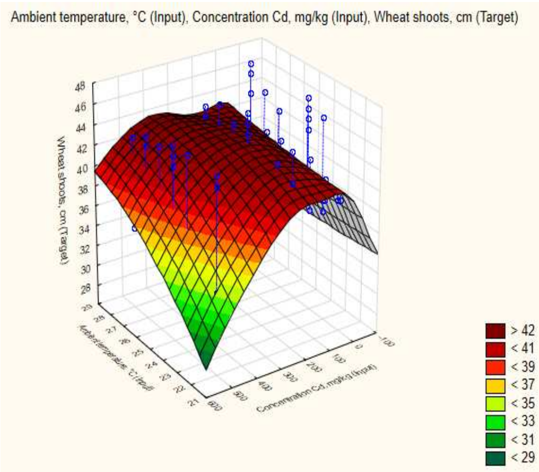


Fig. 7. 3D histograms of the dependencies: x (Input), y (Output), z (Target)

However, the dependence of soil moisture on the concentration of Cd had two options: with a concentration of Cd from 0-100 mg/kg, Soil moisture should be equal to 70-71%, and with a concentration of Cd from 150 to 300 mg/kg, soil moisture should be from 63% to 66 %. There is also a relationship between the concentration of Cd and ambient temperature [12]. The highest shoot growth at ambient temperature from 25 °C to 29 °C was observed when the Cd content in the soil was from 100 mg/kg to 300 mg/kg.
From this we can conclude the input parameters lying in the range were the best:

- *Bacillus Subtilis* 26D – no less 1×10^6 cell/ml;
- Ambient temperature from 25 °C to 29 °C;
- Soil moisture from 63 % to 71 %;
- Concentration of Cd from 0 mg/kg to 300 mg/kg.

To predict the length of wheat shoots, we used the created neural network. To do this, in the Custom Inputs window, we added the selected parameters and examined the results. According to the table in Fig. 8, the neural network at number 5 had the best learning indicators, so when forecasting data, we focused on its indications [13, 14].

Summary of active networks (Spreadsheet1)

Index	Net. name	Training perf.	Test perf.	Validation perf.	Training error	Test error	Validation error
1	MLP 4-29-1	0,972404	0,962881	0,967031	0,456757	0,583606	0,525657
2	MLP 4-30-1	0,983792	0,964472	0,967270	0,290071	0,578607	0,595894
3	MLP 4-4-1	0,977816	0,955847	0,969925	0,365270	0,686458	0,474170
4	MLP 4-14-1	0,975178	0,952839	0,974017	0,409308	0,741867	0,472684
5	MLP 4-22-1	0,973109	0,978199	0,979704	0,481655	0,345809	0,502621

Fig. 8. Window «Summary»

We conducted the first experiment in predicting the output parameter using a neural network. We introduced the following parameters: *Bacillus Subtilis* 26D was 1×10^6 cell/ml, ambient temperature was 25 °C, soil moisture was 63%, Cd concentration was 0 mg/kg. To predict the output parameter, we selected Custom Inputs [15, 16]. Next, in the window that opens, we entered the minimum parameters and selected OK. With these parameters, the length of wheat shoots was 40.49 cm (Fig. 9).

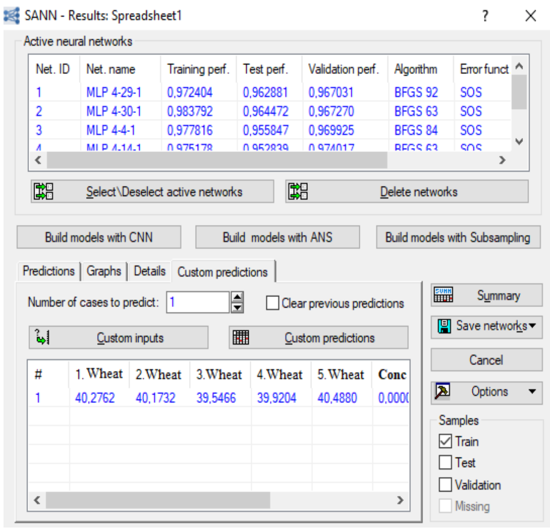


Fig. 9. Prediction of wheat shoots length with the minimum selected parameters

Similarly, we forecasted the maximum parameters in the neural network. We introduced the following parameters: *Bacillus Subtilis* 26D was 1×10^6 cell/ml, the ambient temperature was 29 °C, soil moisture was 71%, Cd concentration was 300 mg/kg. With these parameters, the length of wheat shoots was 43.22 cm (Fig. 10).

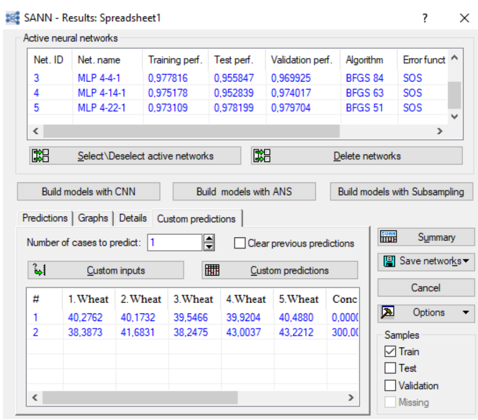


Fig. 10. Prediction of wheat shoots length with the maximum selected parameters

Similarly, we carried out forecasting for the average selected parameters in the neural network. We introduced the following parameters: *Bacillus Subtilis* 26D was 1×10^6 cell/ml, the ambient temperature was 27 °C, soil moisture was 66%, Cd concentration was 150 mg/kg. With these parameters, the length of wheat shoots was 44.67 cm (Fig. 11).

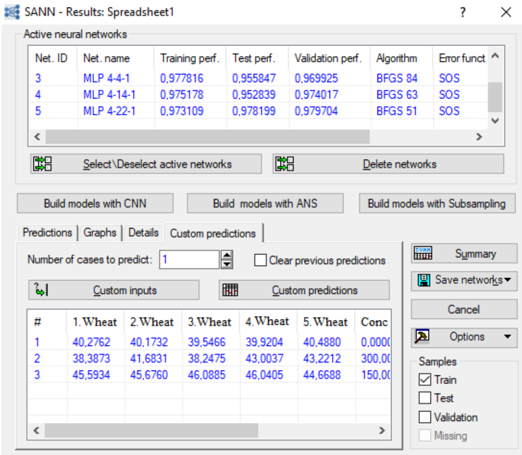


Fig. 11. Prediction of wheat shoots length with the average selected parameters

Similarly, we predicted the shoot length for constant input variables, except for Cd concentration: *Bacillus Subtilis* 26D was 1×10^6 cell/ml, the ambient temperature was 29 °C, soil moisture was 71%, and Cd concentration varied from 0 mg/kg to 500 mg/kg. The results are presented in Fig. 12.

As you can see, the greatest length of wheat shoots is achieved with a small concentration of Cd equal to 100 mg/kg. If we compare the first study with the seventh one, we can see that the obtained dependence between Cd concentration and soil moisture is correct [17]. Therefore, it is necessary to choose this parameter correctly to get the best result.

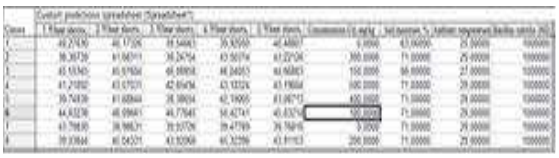


Fig. 12. Prediction of the length of wheat shoots at different concentrations of Cd

CONCLUSIONS

When predicting the optimal parameters for the growth of wheat shoots, the following input parameter ranges were identified:

- *Bacillus Subtilis* 26D – more 10^6 cell/ml;
- Ambient temperature from 25 °C to 29 °C;
- Soil moisture from 63 % to 71 %;
- Cd concentration from 0 mg/kg to 300 mg/kg.

The results showed the applicability of neural networks for predicting plant growth and the optimal set of input parameters [18, 19]. Neural networks can also be used to predict situations that require immediate adjustment. This forecasting method can be used in laboratories, agriculture, as well as for educational purposes.

The reported study was funded by the Ministry of Science and Higher Education of the Russian Federation according to the research project FZWU-2020-0027.

REFERENCES

1. Al-Maathidy A, Alzyoud J A M, Al-Dalaen S, Al-Qtaitat A. Histological alterations in the Thyroid Follicular cells induced by lead acetate toxicity in adult male albino rats. Int. J. Pharm. Phytopharm. Res. 2019; 9(5): 19-26.
2. Zainy F M A, Alotaibi O A. The Quality Control of Eye Shadow, Eyeliner, and Mascara Products that Sold on Saudi Markets. Int. J. Pharm. Phytopharm. Res. 2019; 9(3): 107-118.
3. Nafees S, Nafees H, Rehman S, Rahman S Z, Amin K M. Microbial load, Pesticides residue, Aflatoxin Estimation and Heavy metals Analysis of a Single Unani Drug Badranjboya (Melissa officinalis). Pharmacophores. 2018;9(4): 8-13.
4. Edori O S, Nwineewii J D, Nwoke I B. Metals and Phytochemical Composition of Leaves and Peels of Pawpaw (Carica papaya) sold Within Port Harcourt, Rivers State, Nigeria. Pharmacophores. 2019;10(4): 57-61.
5. Short tour of the system STATISTICA. [Electronic resource] – URL: http://hr-portal.ru/statistica/gl1/gl1.php (date accessed 15.06.2020)
6. STATISTICA. [Electronic resource] – URL: http://soft.mydiv.net/win/download-STATISTICA.html (date accessed 15.06.2020)
7. Kuramshina Z.M., Smirnova Y.V., Khairullin R.M. Increasing Triticum Aestivum Tolerance To Cadmium Stress Through Endo-

phytic Strains Of Bacillus Subtilis. Russian Journal of Plant Physiology, 2016; 63(5): 636-644.

8. Kuramshina Z.M., Smirnova Y.V., Khairullin R.M. Cadmium And Nickel Toxicity For Sinapis Alba Plants Inoculated With Endophytic Strains Of Bacillus Subtilis. Russian Journal of Plant Physiology, 2018; 65(2): 269-277.
9. Kuramshina Z.M., Khairullin R.M. Influence of Endophytic Strains Of The Bacterium Bacillus Subtilis On Cell Number In Monocultures Of Green Algae. Russian Journal of Plant Physiology, 2013; 60(4): 571-575.
10. Artificial Intelligence: A popular introduction for teachers and students. [Electronic resource] – URL: https://inf.1sept.ru/view_article.php?ID=200902304 (date accessed 15.06.2020).
11. Grigoryev I., Miftakhov E., Mustafina S. 2016. Mathematical Modelling Of The Copolymerization Of Styrene With Maleic Anhydride In A Homogeneous Environment. International Journal of Chemical Sciences, 2016; 14(1): 381-386.
12. Mustafina S.A., Valieva Yu.A., Davletshin R.S., Balaev A.V., Spivak S.I. Optimization of catalytic processes and reactors. Kinetics and Catalysis, 2005; 46(5): 705-711.
13. Vaytiev V.A., Mustafina S.A. Searching for uncertainty regions of kinetic parameters in the mathematical models of chemical kinetics based on interval arithmetic. Bulletin of the South Ural State University, Series: Mathematical Modelling, Programming and Computer Software, 2014; 7(2): 99-110
14. Mustafina S.A., Vaytiev V.A., Stepashina E.V. Search the kinetic parameters of the reduced scheme of a-methylstyrene dimerization reaction. ARPN Journal of Engineering and Applied Sciences, 2014; 9(7): 211-217.
15. Sanders, D., Kendrot, E.. CUDA by Example: An Introduction to General-Purpose GPU Programming. - DMK Press, 2011.
16. The particle swarm algorithm - [electronic resource] - Access mode. [Electronic resource] – URL: http: //habrahabr.ru/post/105639/ (date accessed 13.05.2020)
17. Shi, Y., Eberhart, R. A modified particle swarm optimizer». The 1998 IEEE International Conference on Evolutionary Computation Proceedings, 1998.
18. Weise, T. Global. Optimization Algorithms - Theory and Application: Ph.D. thesis. University of Kassel, 2008.
19. Ikramov R., Mustafina S. Numerical study of the belousov-jabotinsky's reaction models on the basis of the two-phase rozenbrock's method with complex coefficients. International Journal of Applied Engineering Research, 2014; 9(22): 12797-12801.