

Neural network method as means of processing experimental data on grain crop yields

*Aleksandr Nikiforov*¹, *Aleksei Kuchumov*¹, *Sergei Terentev*¹, *Inessa Karamulina*¹, *Iraida Romanova*¹, and *Sergei Glushakov*^{1,*}

¹Smolensk State Agricultural Academy, 214000, Smolensk, Russia

Abstract. In the work based on agroecological and technological testing of varieties of grain crops of domestic and foreign breeding, winter triticale in particular, conducted on the experimental field of the Smolensk State Agricultural Academy between 2015 and 2019, we present the methodology and results of processing the experimental data used for constructing the neural network model. Neural networks are applicable for solving tasks that are difficult for computers of traditional design and humans alike. Those are processing large volumes of experimental data, automation of image recognition, approximation of functions and prognosis. Neural networks include analyzing subject areas and weight coefficients of neurons, detecting conflict samples and outliers, normalizing data, determining the number of samples required for teaching a neural network and increasing the learning quality when their number is insufficient, as well as selecting the neural network type and decomposition based on the number of input neurons. We consider the technology of initial data processing and selecting the optimal neural network structure that allows to significantly reduce modeling errors in comparison with neural networks created with unprepared source data. Our accumulated experience of working with neural networks has demonstrated encouraging results, which indicates the prospects of this area, especially when describing processes with large amounts of variables. In order to verify the resulting neural network model, we have carried out a computational experiment, which showed the possibility of applying scientific results in practice.

1 Introduction

Grain production solves the food and strategic problem of the country, as well as ensures its environmental security [1, 2].

Winter crops, including triticale, are the most suitable for the central non-black soil region of the Russian Federation. The advantage of this crop over others manifests itself both on heavy and light sod-podzolic soils [2, 3, 4].

Yield is one of the main indicators of crop evaluation, its level depends on many factors such as: fertilizers, variety (hybrid), timing of seeding, seeding depth, seed quality, soil fertility, meteorological conditions during the years, etc. [5, 6, 7, 8].

The multifactorial nature, non-linearity of dependencies and the randomness of processes affecting cereal yields make it very difficult to make statistical generalizations using previously developed methods. Meanwhile, the data obtained during scientific experiments are random statistical in nature. Moreover, we need to take into account that during the experiment and in the processing of the experimental data obtained, various kinds of errors might appear.

These errors arise due to using approximate methods of processing the physical parameters studied, the accuracy of experimental equipment, the influence of

anthropogenic factors, etc. Such random errors are statistically analyzed. Here the main tool is statistical methods based on the provisions of mathematical statistics, which include regression, dispersion, correlation analysis, etc. Applying them, we have produced a statistical model, which studies the influence of random factors on the functioning of the system.

In order to conduct experiments competently, and then to analyze and process the experimental data obtained, the specialist must have fundamental knowledge in the field of mathematical statistics, have the skills of organizing and planning experiments, as well as be able to apply certain statistical processing methods in practice. Currently, the effective method of processing experimental data is a neural network method or, so-called, neural networks [9].

Active development of neural networks began in the 1950s and 1960s. Their feature lies in the use of a mathematical description of phenomena similar to those occurring in neurons (nervous system) of living organisms for data processing.

Neural networks are applicable to solving tasks that are difficult for traditional computers and humans alike. These include processing large amounts of experimental data, automating image recognition processes, approximating functions, predicting, etc. Artificial neural network similar to the human nervous system is made up

* Corresponding author: sng12@list.ru

of many simple elements operating in parallel. Neural network functions are largely determined by the connections between the elements.

Unlike the classic forecast-based analytics of dependencies, neural networks are capable of:

- incremental accumulation of raw data;
- rapidly providing results;
- refinement of results as new data become available.

Neural networks act as a universal approximator, that is, allow to build generalized models based on processing large amounts of data. In other words, the neural network approximates:

$$Y = \varphi(X) \quad (1)$$

where X is an input vector, Y is a simulated function (output), and φ is a conversion performed by the neural network.

Considering this, the generalization of the results of experimental data on the study of the level of yields of varieties of winter triticale of domestic and foreign breeding using the neural network method is relevant.

2 Methods

To this end, data from field and laboratory studies conducted between 2015 and 2019 in the six-field crop rotation of the Smolensk State Agricultural Academy were used. Experimental data were processed using neural networks that showed the high effectiveness of this approach in engineering studies [9, 10, 11].

During the first stage, we have used the results of the study of the effect of mineral fertilizers on the yield of winter triticale grains. The table shows the range of nitrogen, phosphorus and potassium fertilizer doses studied, which was broken down in experiments into 10 values, which ultimately gave a data sample consisting of 97 vectors.

Table 1. Maximum and minimum values of winter triticale parameters in the sample described.

Parameter	Unit	Minimum value	Maximum value
N	kg/ha a.i.	0	120
P	kg/ha a.i.	0	120
K	kg/ha a.i.	0	160
t _{avg.1}	°C	14.5	18.7
t _{avg.2}	°C	13.5	20.2
t _{avg.3}	°C	16.0	21.7
t _{avg.4}	°C	18.4	21.9
φ _{avg.1}	mm	0.8	7.6
φ _{avg.2}	mm	2.6	6.2
φ _{avg.3}	mm	2.1	2.7
φ _{avg.4}	mm	2.8	3.6
Y	c/ha	15.4	60.8

The distribution density of fertilizer dose combinations is uniform, making the sample representative. The experiments were conducted over a period of four years, which increased the number of samples to build a neural model to 388.

An important stage in building a model is selecting the minimum sample size required for modeling. The study's four-year-long period allowed an assessment of the impact of weather conditions and moisture on the yields of the crop under study.

3 Results and Discussion

The level of yield is influenced by a significant number of different factors, however, as the nature of this work is exploratory and all the available experimental data were used in it, at this stage of the study it was decided to limit the factors to the ones specified in the table.

To do this, arguments describing temperature and precipitation during different periods of plant growth and development have been introduced in the model.

Thus, the sample size of the initial data for the training of neural models of crop yield was 388 training samples. The parameters forming the input vector included: the amount of fertilizer applied (N – nitrogen, P – phosphorus, K – potash); t_{avg}φ_{avg} – average temperatures and average rainfall at different periods of plant growth and development: 1 – sowing-seedling; 2 – seedling-tillering; 3 – tillering-earring; 4 – earing-full ripeness (harvesting). The output value of the modeled function Y is the yield of the studied culture.

In general, to process a sample of raw data, a researcher can use the following sequence of steps, which the authors formed empirically, using previous experience with neural networks.

1. Choosing input parameters.
 - a) logic and analysis of the subject area;
 - b) analysis of the weights of input neurons;
 - c) perturbation of input parameters and analysis of network response to these perturbations;
 - d) alternate exclusion of input neurons and observation of the network generalization error.
2. Identification of conflict samples.
3. Determination of the number of necessary samples.
4. Improving the quality of neural network training with insufficient sample size (multiple cross-checking, multiple sampling repetition and changing the way training samples are applied).
5. Identification of outliers.
6. Normalization of data.
7. Adding noise to learning samples.
8. Choosing the types of neural networks and activation functions.
9. Decomposition of the network by the number of output neurons.

In the first step, a frequency analysis of each of the parameters was carried out to analyze the sample.

For parameters with multiple repetitive parameters, the frequency of sampling for each of the values was calculated.

For parameters with different values, the frequency of values was calculated in ranges. Frequency analysis made it possible to determine which ranges of parameters were covered with the most values and which areas were less represented in the sample.

In the current task, the ranges of fertilizer dose change are presented tightly and evenly, temperatures and precipitation have low distribution density, but cover a fairly wide range of changes.

Based on this analysis, the generalized expression of the sought model looks like:

$$Y = \varphi(N, P, K, t_{avg.1}, t_{avg.2}, t_{avg.3}, t_{avg.4}, \varphi_{avg.1}, \varphi_{avg.2}, \varphi_{avg.3}, \varphi_{avg.4}) \quad (2)$$

After identifying and excluding insignificant parameters from the training sample, the quality and accuracy of the neural network model tends to improve due to a decrease in its dimension and complexity. But we should remember that excessive reduction in the number of inputs and simplification of the neural network can prevent the identification of patterns in a particular task. It can also lead to *conflicting (contradictory) samples*, i.e. ones that have identical input vectors and different output vectors.

Because of the erroneously prepared data, the error will not fall below this level, whatever methods of teaching the neural network are applied.

Conflict examples may occur in the learning set due to a measurement error.

Search for conflict examples can be done with the aid of software, using special algorithms that search for matching learning samples, or by carefully reviewing the sample in search of repetitive input vectors.

In our case, no conflict examples were found in the processing of the original sample.

As mentioned above, successful neural network modeling requires the *necessary number of training samples*.

The study [12] provides an estimate of the minimum required sample size:

$$Q = 7 * N_x + 15 \quad (3)$$

where N_x is the number of arguments in the neural network model; Q is the number of samples of the learning sample.

In the case under study, the number of arguments N_x is 11, meaning that the minimum required training sample volume should be 92 samples, which is several times less than the 388 samples used.

In the next phase of data preparation, outliers were excluded from the training sample. Outliers in the original sample are the values of parameters that differ significantly from other similar ones in terms of value.

In the simplest cases, outliers can be detected by carefully reviewing the sample, in more complex multi-parametric dependencies and for large amounts of information simplest neural networks with minimal number of neurons in a hidden layer can be used to search for outliers.

As our sample size was large, we needed to use various algorithms to search for outliers [13].

Several examples of outliers were found in the sample. In order not to reduce the sample size, the detected outliers were corrected by approximating the type of dependency. This allowed to preserve the size of the learning sample.

A significant step in preparing data for modeling is to bring it to a dimensionless form and normalize it. Dimensionless values were obtained by splitting the source data by maximizing each parameter within their ranges.

To normalize the data, a range of 0.1 was adopted, in the interval from 0 corresponding to the minimum value of the dimensionless parameter to 1—the maximum one.

After modeling, neural network responses are converted back from normalized to natural range of values. The process of normalization has been detailed in the works of a number of scientists [12, 14, 15, 16].

In this case, the normalization of the data was particularly relevant because of the physical nature of the data.

After the initial sample was prepared for the simulation, the sample size was 388 samples. From this amount, a test set of 26 samples was selected that did not participate in the training. This set was used to determine the error of the constructed model.

In order to train the neural network, based on the accumulated international experience, we chose the perceptron-type networks with two-layer perceptrons [17, 19, 20].

As an activation function, we adopted logical sigmoid for both layers [20, 21].

The selection of suitable network architecture is empirical. Without taking into account a number of existing network architecture guidelines to address dependency approximation, we had to choose the architectures that are most appropriate to each specific task and data.

Two-layer neural networks with 25 neurons in the first (hidden) layer were created to select the learning function. The accuracy of the simulation when changing training functions was tested: with the BFGS quasi-Newtonian method, with the Levenberg-Marquardt method of optimization and with the optimization of Levenberg-Marquardt with Bayesian regularization.

The first layer used logical sigmoid as an activation function. In the second layer we used neural networks with logical sigmoid (logsig) and linear activation function (purelin) for each learning function.

After training, errors were calculated for each type of neural network. In accordance with the minimum error, the training function with the method of optimization of Levenberg-Marquardt with Bayesian regularization was chosen.

Then a computational experiment was conducted in order to optimize the network architecture on the chosen learning function. We have built the following models with the number of neurons:

- two-layer with the number 10, 15, 20, 25, 30 in a hidden layer and one neuron in the output layer;
- three-layer with the number 10, 15, 20, 25, 30 in the first (hidden) layer and 10, 15, 20, 25, 30 in the second (hidden) layer for each variant of the number of neurons in the first layer. The third output layer contained one neuron.

The same activation function, logical sigmoid, was used for all the layers. Additionally, from the theorem of Arnold-Kolmogorov-Hecht-Nielsen it follows that to

build a neural network model of complex function it is enough to use a perceptron with one hidden layer of sigmoid neurons, the number of which is determined by the formulas below.

To determine the required number of synaptic weights of the neural network:

$$\frac{N_y Q}{1 + \log_2(Q)} \leq N_w \leq N_y \left(\frac{Q}{N_x} + 1 \right) (N_x + N_y + 1) + N_y \quad (4)$$

where N_x is the number of neurons in the input layer (number of parameters); N_y is the number of neurons of the output layer (the number of simulated values); Q is the number of elements of many training samples, that is, the number of pairs of input and output vectors X_q and D_q ; N_w is the necessary number of synaptic connections.

According to the modeling data, the synaptic bond range for the model ranges from 58 to 797. Accordingly, $58 \leq N_w \leq 797$.

Then the number of neurons in the hidden layer of the two-layer perceptron will be equal to:

$$N = \frac{N_w}{N_x + N_y} \quad (5)$$

These computed estimates show that the optimal number of neurons in the hidden layer for the two-layer perceptron model ranges from 8 to 114 ($8 \leq N \leq 114$).

There is no strict theory of choosing the optimal number of hidden layers and neurons in hidden layers. In practice, perceptrons with one or two hidden layers are the ones most commonly used, while the number of neurons in hidden layers usually ranges from $N_x/2$ to $3N_x$.

The computational experiment on neural networks showed that the minimum error is observed in a network with a number of neurons in the hidden layer equaling 30. This value fits in the range obtained in theoretical calculations using the above formulas.

The smallest error was obtained for a neural network with the following architecture: two-layer, number of neurons in the hidden layer 30, logical sigmoid activation function for all layers, learning function with the optimization of Levenberg-Marquardt with Bayesian regularization. The average margin of error for all crops studied did not exceed 3.3%.

As an illustration of the results obtained, we demonstrate the dependence of the yield of winter triticale on doses of nitrogen and potassium fertilizers using the spline function in the figure 1.

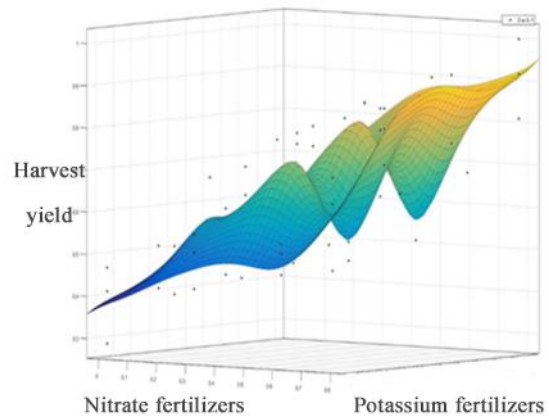


Fig. 1. Effect of single doses of nitrogen and potassium fertilizers on triticale yields.

The choice in favor of two-layer networks was made due to the fact that the difference in the errors of the two-layer and three-layer neural networks is minimal, but obviously, models with the two layers of perceptrons are more economical in terms of the required computing power.

Checking the effects of normalizing when comparing a neural network trained on non-normalized data and a neural network trained on already normalized data was performed in our previous work [22].

Once the original data are normalized, the margin of error decreases by 3% on average compared to the same model trained on non-normalized data.

4 Conclusions and confirmations

Pre-processing the sample for neural network training is practical, as it significantly reduces modeling errors compared to neural networks created on unprepared raw data.

During the work with neural networks, various recommendations for the preparation and processing of raw data have been collected and studied. They include analysis of the subject area and weights of neurons, identification of conflicting examples and outliers, data normalization, determination of the required number of samples for training a neural network and improving the quality of training of neural networks with insufficient sample size, as well as the choice of types of neural networks and decomposition based on the number of output neurons.

The accumulated experience with neural networks has shown encouraging results, which indicates that this approach is prospective, especially when describing processes with a large number of arguments.

The high accuracy of the developed neural network models of winter triticale yield allows to use them in practice during the research of other cereals, as well as in calculating yield forecasts taking into account a different combination of a large number of biotic and abiotic factors.

References

- [1] S.E. Terentyev, et al., Yields of new varieties of cereals depending on the level of mineral nutrition, *Grain Industry of Russia*, **3**, 60-63 (2012).
- [2] I.N. Romanova, et al., Yields of cereals and the level of soil fertility depending on the application of mineral fertilizers, soil type in the crop rotation system, *Grain Industry of Russia*, **2**, 44, 57-60 (2016).
- [3] S.A. Guchanov, O.V. Melnikova, V.E. Torikov, Yield and grain quality triticale winter depending on elements of cultivation technology, *Herald of the Kursk State Agricultural Academy*, **4**, 90-95 (2018).
- [4] I.N. Romanova, et al., Ways to increase the yield and quality of grain with winter triticale of Consul variety, *Herald of Bashkir State Agricultural University*, **1**, 49, 37-41 (2019).
- [5] N.N. Semenenko, I.I. Vaga, Effect of hydrothermal conditions of growth and development on the yield of winter triticale, *Melioration*, **2**, 66, 137-144 (2011).
- [6] I.N. Romanova, et al., Productivity of varieties of grain crops depending on the mineral nutrition backgrounds, *Grain Industry of Russia*, **2**, 37-43 (2012).
- [7] O.S. Tikhonova, I.S. Fatykhov, Effect of the rate of seeding on the quality of grain of winter cereals in the Middle Pre-Ural, *Herald of Bashkir State Agricultural University*, **4**, 24, 14-16 (2012).
- [8] V.E. Torikov, et al., Comparative characteristics of the quality of the grains of winter triticale varieties grown in southwest Russia, *Herald of the Altai State Agricultural University*, **2**, 172, 49-56 (2019).
- [9] A.G. Nikiforov, et al., Using neural network modeling to calculate the energy characteristics of a centrifugal compressor, *Compressor Equipment and Pneumatics*, **3**, 18-21 (2015).
- [10] A. Nikiforov, A. Kuchumov, S. Terentev, E. Petukhov, K. Kabalyk, Simulation of gas-dynamic characteristics of a centrifugal compressor vane diffuser using neural networks, *International Scientific Conference on Energy, Environmental and Construction Engineering, EECE-2019, E3S Web of Conferences*, **140**, 05003 (2019). DOI: 10.1051/e3sconf/201914005003.
- [11] A. Nikiforov, D. Avramenko, A. Kuchumov, S. Terentev, A network application for modeling a centrifugal compressor performance map, 11th International Conference on Compressors and their Systems, London, UK, *IOP Conf. Series: Materials Science and Engineering*, **604**, 012046 (2019). DOI:10.1088/1757-899X/604/1/020.
- [12] L.N. Yasnitsky, *Intelligent Systems* (Knowledge Laboratory, Moscow, 221, 2016).
- [13] V.P. Shkodyrev, et al., Review of methods for detecting anomalies in data streams, *Proc. of the Second Conference on Software Engineering and Information Management*, St.Petersburg, Russia, **1864**, 7 (2017).
- [14] T. Jayalakshmi, Dr.A. Santhakumaran, Statistical normalization and back propagation for classification, *International Journal of Computer Theory and Engineering*, **3**, 1, 89-93 (2011).
- [15] E.S. Ogasawara, L.C. Martinez, D. de Oliveira, G. Zimbrão, G.L. Pappa, M. Mattoso, Adaptive normalization: a novel data normalization approach for non-stationary time series, *The 2010 International Joint Conference on Neural Networks, IJCNN* (July 2010). DOI: 10.1109/IJCNN.2010.5596746.
- [16] S.C. Nayak, B.B. Misra, H.S. Behera, Impact of data normalization on stock index forecasting, *International Journal of Computer Information Systems and Industrial Management Applications*, **6**, 257-269 (2014).
- [17] S. Osovsky, *Neural Networks for Information Processing*, trans. from Polish by I.D. Rudinsky (Finance and Statistics, Moscow, 334, 2002).
- [18] V.V. Kruglov, M.I. Dli, R.J. Golunov, *Fuzzy Logic and Artificial Neural Networks* (Publishing House for Literature on Physics and Mathematics, Moscow, 224, 2001).
- [19] V.V. Borisov, V.V. Kruglov, A.S. Fedulov, *Fuzzy Models and Networks* (Hotline-Telecom, Moscow, 284, 2016).
- [20] V.S. Medvedev, V.G. Potemkin, *Neural Networks. MATLAB 6* (DIALOGUE-MIFI, Moscow, 496, 2002).
- [21] V.V. Borisov, A.S. Fedulov, M.M. Zernov, *Basics Hybridization of Fuzzy Models* (Hotline-Telecom, Moscow, 100, 2017).
- [22] A.G. Nikiforov, D.Y. Avramenko, Preparation of experimental data for neural network modeling of characteristics of centrifugal compressors, *Scientific and Technical Records of SPBPU, Natural and engineering sciences*, **24**, 4, 61-71 (2018). DOI: 10.18721/JEST.24406.