

Creating a neural network system for forecasting and managing agricultural production using autocorrelation functions of time series

Aleksey Rogachev^{1,2*}, *Elena Melikhova*²

¹ Volgograd State Agrarian University, 26 University Avenue, Volgograd, 400002, Russia

²All-Russian research Institute of irrigated agriculture, 9 Timiryazeva, Volgograd, 400002, Russia

Abstract. The article considers the features of creating an artificial neural network (ANN) for modelling and forecasting the dynamics of long-term time series (TS) levels of grain yield in arid conditions on the example of the Lower Volga region of the Russian Federation. In order to increase the validity of the choice of architecture and macroparameters developed by ANN, statistical characteristics of the simulated TS were analysed. The autocorrelation function of distribution of levels of long-term series of grain yields is constructed. It is proposed to take into account the characteristics of time lags of autocorrelation functions when selecting ins macroparameters for predicting BP yield. On the basis of preliminary statistical analysis, "peaks" corresponding to the time lags of the autocorrelation function, whose values are determined for different groups of grain crops, are identified. The obtained values are recommended to be taken into account when selecting the value of the time window parameter when constructing neural network models of productivity. This is the basis of the proposed information technology for building ins for predicting crop yields. The results of neural network modelling and yield forecasting can be successfully used for managing agricultural production, including in the arid conditions of the Lower Volga region of the Russian Federation.

1 Introduction

Agricultural production is characterized by a significant dependence on hydrothermal conditions, especially in cases of insufficient moisture, typical, for example, for the Lower Volga region of Russia. Accurate crop yield forecasts are important for managing agricultural production. There are various methods for predicting TS yields [1, **Error! Reference source not found.**, 5, 6, 7, 10], but their accuracy is insufficient.

Cultivation of agricultural crops in conditions of insufficient water supply characterized by a high value of the coefficient of variation of time series (TS) of yield levels, reaching values of 0.3 and more for the main grain crops, which limits the use of factor trend-seasonal models. The noted feature of the yield TS causes a methodological error in

* Corresponding author: rafr@mail.ru

assessing the risks associated with planning and forecasting agricultural production based on various methods of nonlinear dynamics [1-**Error! Reference source not found.**]. The use of modern methods of nonlinear dynamics, in particular neural networks, allows us to increase the adequacy of the resulting models.

Methods for modeling complex socio-economic systems based on neural network technologies are becoming more common due to the versatility and possibility of computer implementation using specialized computer programs, among which one of the most convenient is STATISTICA Neural Networks (SNN). An important advantage of SNN is its close integration with the statistical analysis computer program STATISTICA, which allows developing hybrid technologies that include preliminary statistical analysis and neural network data modeling, using a wide range of built-in tools [**Error! Reference source not found.**, 3].

Neural network modeling technologies are based on the construction of artificial neural networks that allow describing the dynamics of simulated nonlinear systems and modelling the processes of interannual yield fluctuations. Modern approaches include artificial intelligence (AI) methods, including deep machine learning [**Error! Reference source not found.**], cognitive modelling (construction and parameterization of fuzzy cognitive maps) [**Error! Reference source not found.**, **Error! Reference source not found.**], and modelling based on artificial neural networks (ANN) [3, 8, 13].

The main objective of the study is to increase the reliability of neural network modelling and, accordingly, to predict the levels of BP yields by taking into account internal patterns of dynamics of changes in yield in previous years, including the use of autocorrelation analysis of long-term BP yields.

2 Research methods

For research was used by TS as grain in General, and individual cereal crops, selected on the data of Federal service of state statistics of Russia for the Volgograd region (Volgogradstat) for the period 1950-2018 years, a fragment of which is for TS "cereal crops" are presented in table.1.

Previously, a statistical test of the null hypothesis about the correspondence of the empirical law of yield distribution to the normal one was performed using two different statistical criteria (Pearson's Chi-square, and Kolmogorov-Smirnov). Artificial neural networks were built using the STATISTICA Neural Networks (SNN) application package, the advantage of which is integration with the statistical analysis computer program STATISTICA.

This approach simplifies the development of hybrid technologies that include preliminary statistical analysis and neural network modeling of the studied time series.

3 Results and discussion

3.1 Pre-forecast statistical analysis of time series

For a sample of the yield of "grain crops as a whole", the calculated value of the Pearson criterion exceeded 51, which is significantly higher than the critical value of 9.5, which was determined from the table at ($\alpha = 0.05$; $\sigma = 4$).

Table 1. Time series of productivity of grain agricultural crops, centners/ha

Years	Productivity	Years	Productivity	Years	Productivity	Productivity	Years
1950	4.6	1968	12.9	1986	10.6	2004	19.5

1951	4.8	1969	6.4	1987	13.2	2005	18.5
1952	7.5	1970	17.5	1988	18.5	2006	17.1
1953	4.8	1971	10.8	1989	18.4	2007	13.5
1954	3.6	1972	4.5	1990	20.6	2008	24.6
1955	8.4	1973	17.2	1991	14.4	2009	19.2
1956	4.8	1974	16.0	1992	15.4	2010	12.0
1957	3.9	1975	3.8	1993	20.1	2011	17.2
1958	13.7	1976	19.2	1994	12.0	2012	16.4
1959	5.5	1977	10.2	1995	5.9	2013	17.5
1960	10.4	1978	19.7	1996	8.8	2014	20.4
1961	10.0	1979	6.8	1997	14.8	2015	17.3
1962	14.3	1980	11.9	1998	4.7	2016	24.0
1963	6.1	1981	8.2	1999	7.0	2017	26.9
1964	14.0	1982	10.2	2000	12.2	2018	19.3
1965	9.2	1983	13.1	2001	17.0		
1966	12.5	1984	4.1	2002	16.1		
1967	11.2	1985	14.1	2003	15.2		

Analysis of the data presented in the table.1, showed that the empirical law of distribution of long-term values of yield levels for grain crops is statistically significantly different from the normal distribution for both criteria used. During the research, the regularities of the distribution of the levels of time cycles of long-term yield levels for other crops from the grain group were revealed (Fig. 1).

In order to justify the architecture and macroparameters of mathematical models of grain yield in arid conditions, which must be taken into account when constructing ins, and to increase the reliability of the obtained forecasts, an autocorrelation analysis of the studied BP was performed.

The results of autocorrelation analysis for the group "grain crops in General" are shown in Fig. 2.

The analysis of the obtained autocorrelation functions of the yield TS performed for different groups of grain crops shows both the presence of significant cyclic components and a significant difference in their statistical characteristics. In accordance with this, in the subsequent simulation of TS yields of each of these crops using ANN, it is necessary to perform taking into account the obtained statistical characteristics.

The autocorrelation function diagrams showed statistically significant peaks at one-, two-, three -, four -, and twelve-year lags (Fig. 2). Note that the most pronounced "peaks" are observed at lags of three and twelve years. A possible hypothesis for explaining the 12th cycle may be a superposition of two-and three-year cycles. Thus, the cyclical time series of grain yields, confirmed in the studies for the above years, as an attribute of the endogenous dynamics of the considered economic TS on the example of yield, should be used in neural network forecasting. This result should also be taken into account when planning agricultural production.

Thus, for predicting TS yields using ins, we can recommend the use of pre-forecast methods of statistical analysis, the algorithms of which are sufficiently developed and implemented as built-in functions in many computer programs. At the same time, the main tasks of the preliminary analysis are to justify the numerical values of the ANN macroparameters in the formation of its architecture even before the start of the training procedure.

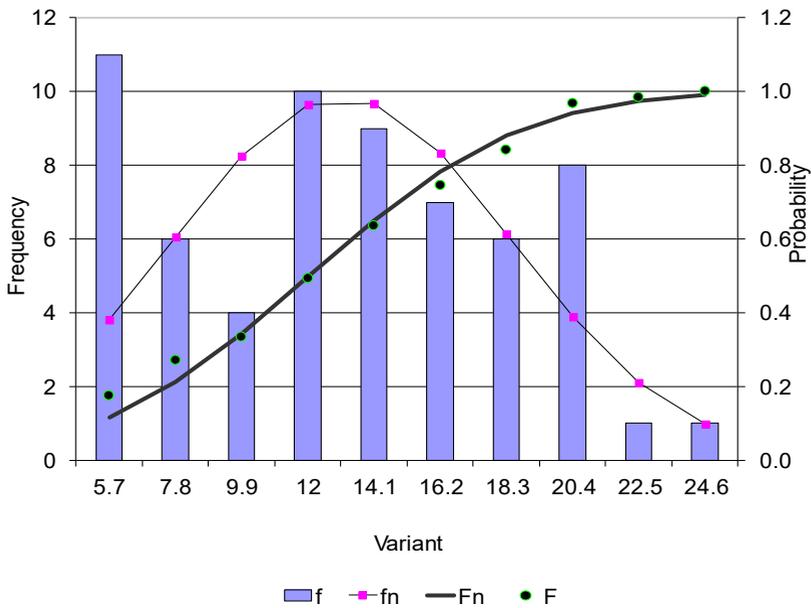


Fig.1. Empirical distribution of grain yield levels in comparison with the normal law

It is also interesting to compare TS yields formed from geographically similar or different regions, according to Russian Application «B. Zagaytov, Yanovsky P. (RU), A method for predicting interannual fluctuations in winter wheat yield. Application no. 2005114095. A. IPC A 01 G7/00 (2006.01). - Announced 11.05.2005; publ. 20.11.2006». At the same time, the data sets for ANN training are significantly expanded, which can have a positive impact on the quality of network training and prediction accuracy.

3.2 Building and training artificial neural networks

The practical implementation of the developed ins models was carried out by means of computer mathematics in the SNN v4.0 environment. – The interface of a neural network system for predicting the yield of grain crops based on SNN v.4 is shown in the Fig. 3.

In the process of selecting neural network parameters, the "Time window" parameter of the Create Network dialog box was assumed to be equal to the number of the most pronounced lag on the correlogram (Fig.2). In the case of the "grain in General" TS simulation, the "Time window" parameter was assumed to be 3.0. The ANN architecture included a three-layer perceptron according to the recommended [Error! Reference source not found., 5] methods (Fig. 3).

The selection of the initial BP from the yield levels and the formation of training and verification samples, as well as the training of the created ANN, were carried out in a partially automated mode [Error! Reference source not found.]. As a result of performing the ins training procedure with built-in SNN tools, a family of preferred ins architectures and macroparameters was obtained.

To analyze the predictive characteristics of the trained network, various time series projections were used in the "Time Series Projection" window of the SNN program. These projections characterize the quantitative and qualitative possibilities for obtaining forecasts with different initial values of BP levels and forecast horizons. The conducted research allows us to recommend the resulting family of neural network models for short-term

forecasting with a horizon of 1-2 years, which can be performed directly in the "Run Single Case" window.

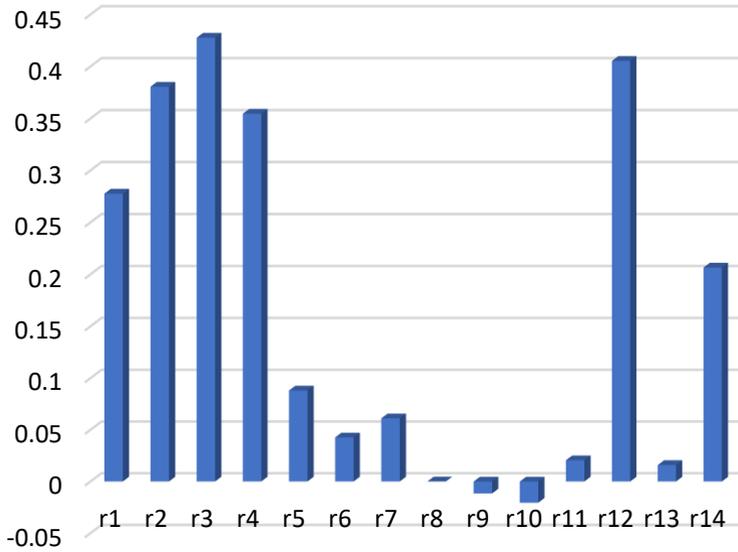


Fig.2. Diagram of the autocorrelation function of the BP yield for the group " grain crops"

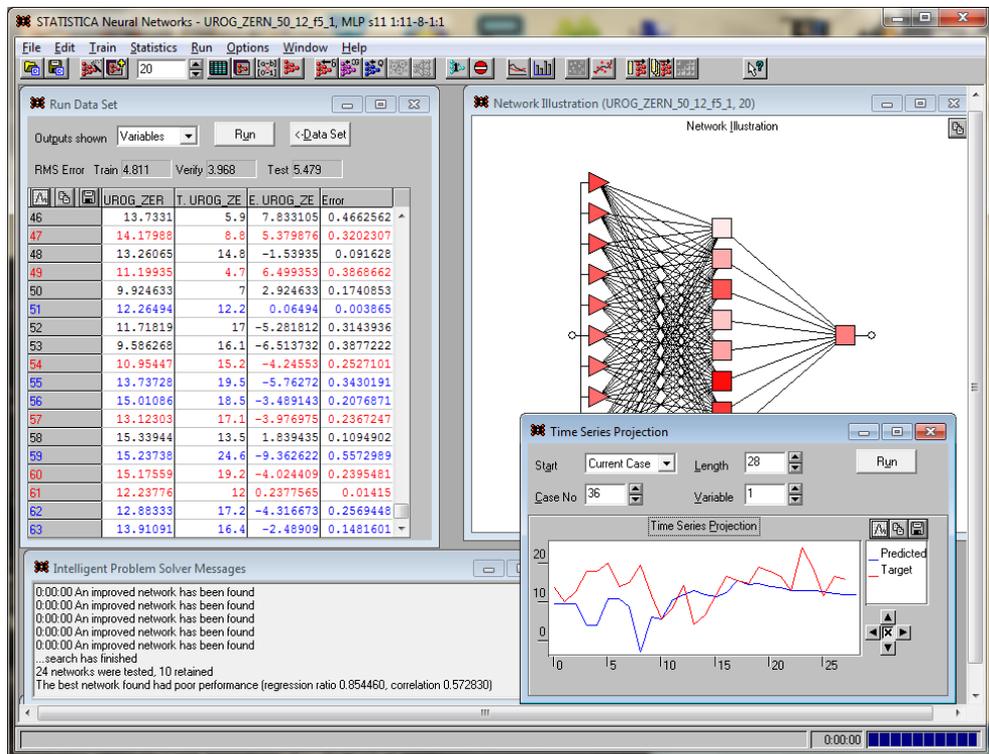


Fig. 3. Working Windows of the interface of the developed system for forecasting crop yields

4 Conclusion

Thus, the proposed information technology for short-term modelling of TS levels of crop yields using ANN is based on a pre-forecast autocorrelation analysis of long-term TS levels of crop yields. This approach makes it possible to more reasonably take the value of the characteristic parameter of the time window "Steps" when selecting the ANN macroparameters, which reduces the error of short - and medium-term forecasting of grain yield in the arid conditions of the Lower Volga region. The achieved error value of 6 ... 13% is quite acceptable for forecasting such TS and planning agricultural production for the next 1-2 economic years.

Acknowledgement

The work was carried out with the financial support of the RFBR and the Administration of the Volgograd region under project No. 19-416-340014 "Creation of a neural network management system for programmed agricultural production using retrospective data and remote sensing results for arid conditions in the Volgograd region".

References

1. R. Ilyasov, V. Yakovenko, T. Malyutina, D. Tkachenko, E. Magomadov *European Proceedings of Social and Behavioural Sciences*. **58**, 497-504 (2019). DOI: 10.15405/epsbs.2019.03.02.57
2. K. Tokarev, *et al*, *IOP Conf. Ser.: Earth Environ. Sci.* **421**, 042015 (2020). DOI:10.1088/1755-1315/421/4/042015.
3. E. Popova, A. Komratova, T. Nedogonov, A. Vasilenko, *Modern Economics: Problems and Solutions* **6 (102)**, 20-30 (2018). DOI: 10.17308/meps.2018.06/1907
4. A. Rogachev, *Asian Social Science* **11 (20)**, 178-185 (2015) - DOI: 10.5539/ass.v11n20p178.
5. A. Rogachev, E. Melikhova, A. Shokhnekh, *Espacios* **39 (11)**, 1 (2018).
6. A. Rogachev, *Proceedings of the Lower Volga agricultural University complex: Science and higher professional education* **4 (52)**, 332-339 (2018). DOI: [10.32786/2071-9485-2018-04-47](https://doi.org/10.32786/2071-9485-2018-04-47)
7. I.F. Yurchenko, *Journal of Physics: Conference Series* **1015 (4)**, 042063 (2018).
8. A. Komratova, E. Popova, S. Turley, *et. al.*, *Modern Economics: problems and solutions* **5 (113)**, 8-16 (2019). DOI: [10.17308/meps.2019.5/2100](https://doi.org/10.17308/meps.2019.5/2100)
9. G. Khubaev, *Management of economic systems: electronic scientific journal*, **23**, 243-249 (2010).
10. E.V. Melikhova, A.F. Rogachev, *Studies in Computational Intelligence* **826**, 1193-1201 (2019)
11. J. Gooijr, R. Hyndman, *A Selective Review*. Tinbergen Institute Discussion Paper, **068 (4)** (2005).
12. T. Huang, R. M-T. Kechadi, *International Journal of Computational Intelligence* **3 (1)**, 1304-1386 (2006).
13. Papa Parse - Powerful CSV Parser for JavaScript. Available at: <http://papaparse.com> (25.10.2017).
14. Simple feed-forward neural network in JavaScript. Available at: <https://github.com/BrainJS/brain.js> (25.10.2017).

15. G. Zahng, B. Patuwo, M. Hu., *International Journal of Forecasting* **14**, 35-62 (1998).
16. R. Hraskoa, A. Pacheco, R Krohling, *Procedia Computer Science* **55**, 990 – 999 (2015). DOI: 10.1016/j.procs.2015.07.104.
17. G. Yun, D. Medinsky, *Science and technology* **36 (4)**, 35-41 (2017).
18. Ya. S. Lozovoy, A.I. Sekirin, Solving the forecasting problem using neural networks
access Mode. Available at:
http://www.rusnauka.com/1_NIO_2011/Informatica/78176.doc.htm (19.10.2019).
19. A.M. Komratova, E.V. Popova, S.I. Turley, T.A. Nedogonov, *Modern economy: problems and solutions* **7 (115)**, 33-41 (2019). DOI: [10.17308/meps.2019.7/2155](https://doi.org/10.17308/meps.2019.7/2155).
20. I.E. Egorova, A.G. Gagarin, S.Y. Kuznetsov, et. al., *Advances in Intelligent Systems and Computing* **726**, 1156-1162 (2019). DOI: 10.1007/978-3-319-90835-9_130.