Application of machine learning methods for crop rotation selection in organic farming system

Kirill Yu. Maksimovich[1,2*], *Vladimir* K. Kalichkin[1], *Dmitry* S. Fedorov[2], and *Vitaly* V. Aleschenko[1]

[1]Novosibirsk State Agrarian University, 160, Dobrolyubov str., Novosibirsk, 630039, Russia
[2]Siberian Federal Scientific Centre of Agro-BioTechnologies of the Russian Academy of Sciences, P.O. box 267, Krasnoobsk, 630501, Russia

Abstract. This study demonstrates the possibility of crop rotation selection based on the assessment of productivity and sustainability of crop production under different atmospheric moisture conditions. The study considers 8 crop rotations oriented to grain production. The data obtained in long-term field experiments in the forest-steppe of the Novosibirsk region were used. As a result of the implementation of the decision tree (CART) and the use of ensemble algorithm (Random Forest) the construction of a model characterized by a fairly high predictive ability was performed. Standardized Precipitation Index was chosen as the main predictor characterizing atmospheric moistening in different periods of vegetation. The most stable from the point of view of stability of crop yield – grain-fallow with winter rye, grain-fallow with legumes (vetch-oats), in conditions of manifestation of atmospheric drought of different severity were selected. The possibility of using machine learning methods (CART, Random Forest) as effective tools in the selection of crop rotation for sustainable grain production without the use of chemicalization in soil and climatic conditions of Siberia, as well as the assessment of possible risks in the transition of crop production to organic technologies were scientifically substantiated.

1 Introduction

Crop rotations play a key role in the rational organization of land use and design of farming systems. A well-designed and consistently implemented crop rotation not only prevents soil degradation but also reduces crop contamination levels and the spread of pests. This typically affects crop yields and the quality of products [1, 2, 3].

Economic-mathematical modeling has been and remains the primary tool for selecting cultivated crops and designing crop rotations [4, 5, 6]. However, traditional approaches are based on static concepts, employing single-criterion optimization procedures instead of multi-criteria evaluations. Uncertainty in information is typically classified as random factors or probabilities, remaining static irrespective of the dynamic evolution of various constraints.

* Corresponding author: kiri-maksimovi@mail.ru

With advancements in computational technology, information-reference and information-computational systems have gained widespread use, applied in tasks like land quality assessment, fertilizer dosage calculations, and field technical passport management [7].

In recent years, attempts have been made to incorporate artificial intelligence methods in solving crop rotation design problems. For instance, [8] conceptualized the subject area of "crop selection and placement" and developed a model with ontological entities represented in UML format. The model describes five classes – "Crop", "Crop Biology", "Growth and Development Conditions", "Management Actions", and "Placement" – with characteristics that influence target indicators. An algorithm based on a control matrix, where experts assign scores from 0 to 1 to seven "Management Actions" factors, assuming that these actions affect 22 factors in the other three classes: "Crop Biology", "Growth and Development Conditions", and "Placement".

Various approaches to crop rotation planning are discussed in the international literature. Or instance, models like FruchtFolge [9] and the Farm Planning Model [10] utilize linear programming techniques. These models establish constraints and goals as formulas that are maximized or minimized using a mixed-integer linear programming solver (MILP). In [11], a knowledge-based and intelligent systems approach is proposed, using five different evolutionary algorithms to create a multi-objective crop rotation plan. Researchers also highlight that traditional data processing methods cannot meet the ever-growing needs of intelligent agriculture, posing a significant obstacle to extracting valuable information from field data [12]. To address this, machine learning (ML), a subset of artificial intelligence, is employed, leveraging the exponential growth in computational power [13].

One widely used ML algorithm is Decision Trees (DT), which are classification or regression models formulated in a tree-like architecture [14]. DT organizes datasets into progressively smaller homogeneous subsets, creating a connected tree-like graph. Each internal node represents a different pairwise comparison on a chosen attribute, while each branch represents the result of that comparison. The terminal nodes hold the final decisions or predictions made following the path from the root to the leaf (expressed as rules). The most common learning algorithm in this category is Classification and Regression Tree (CART) [15, 16].

The objective of this study is to construct decision trees for crop rotation selection in organic agriculture, based on the usage of long-term field experiment data.

2 Materials and methods

The study was conducted using data from long-term field experiments carried out in the forest-steppe region of Priobye (Ob River region), Novosibirsk Oblast, by the Siberian Research Institute of Agriculture and Agrochemistry from 1999 to 2019. Eight crop rotations oriented towards grain production were examined: Grain (0 fallow), Grain-fallow (25% fallow), Grain-fallow (25% fallow) with winter rye, Grain-grass with winter rye and melilot (for herbage or cover crop), Grain with legumes (vetch-oats), Grain with oilseeds (rapeseed), Grain-grass (clover), and Grain-grass (vetch-oats (herbage)).

All crop rotations were implemented in both time and space, repeated three times, on plots measuring 475 m2. In total, there were 104 experimental fields for the crop rotations. Time series of crop rotation productivity expressed in dt g.e./ha of crop rotation area were used in the calculations. Data were obtained under conditions without the application of intensification measures (fertilizers, pesticides) to better assess the variability of crop rotation productivity in organic farming. The actual crop yield of the studied crops was converted to grain equivalents (g.e.) using crop production conversion coefficients.

Given the characteristics and distribution nature of the original data, including non-compliance with the normal distribution model, a relatively small sample size, and the presence of discrete and continuous factors with complex nonlinear correlations, appropriate statistical methods were chosen. The normality of data was assessed using the Shapiro-Wilk and Lilliefors tests. Due to the deviation from normal distribution (Shapiro-Wilk and Lilliefors tests, $p < 0.05$), the Poisson log model was utilized [17, 18].

The agrometeorological resources during the experiment years were considered using the Standardized Precipitation Index (SPI) [19]. SPI was calculated based on time series data from meteorological observations in Novosibirsk city, sourced from the website "http://www.pogodaiklimat.ru". The SPI calculations were performed using the open-source software Drought Indices Calculator (DrinC) (https://drought-software.com).

The relationship between SPI and crop rotation productivity was assessed using the Spearman rank correlation coefficient, known for its robustness against outliers and non-normal data distribution. The association between crop rotation type and productivity was evaluated using the Kruskal-Wallis test. Nonlinear relationships between variables were modeled using polynomial regression (Equation 1):

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + ... + \beta_n x^n + \varepsilon, \quad (1)$$

where y is the dependent variable, x is the independent variable, β_0, β_1, β_2, ..., β_n are the coefficients of the model, n is the degree of the polynomial, and ε is the random error.

Variation in crop rotation productivity and yield stability was evaluated by calculating the coefficient of variation for each crop rotation, providing a measure of relative data variation. A lower coefficient of variation indicates greater crop rotation stability. Graphical analysis with time series plots for each crop rotation type, visual assessment of productivity stability, and trend analysis were also employed.

To create productivity models for crop rotations, the CART algorithm was utilized with the author's software Crop Yield Analysis & Forecast (CYAF) [20, 21]. Mathematical models using decision trees were built, where each node represents data division based on specific predictors. The hierarchical structure of the obtained decision tree was visualized, presenting the decision-making branches and division conditions clearly. Additionally, the Random Forest (RF) model, an ensemble of multiple decision trees, was applied. Tree complexity calculations, accuracy evaluation, and tree optimization were performed using cross-validation. Additional computations and graphical visualizations were conducted using the R programming language and R-Studio integrated development environment for statistical data processing and graphics.

3 Results and discussion

The obtained results confirm the hypothesis that the maximum productivity under different atmospheric moisture conditions is characteristic of grain-legume and various types of grain-grass crop rotations. The average multi-year productivity of such crop rotations ranged from 20 to 30 dt g.e./ha. For grain-grass rotations (grain-grass with clover, grain-grass with legumes), it was slightly higher, ranging from 23 to 30 dt g.e./ha (Figure 1).

As a result of data analysis, a statistically significant relationship between crop rotation productivity and its type was revealed (Kruskal-Wallis test, $p < 0.0001$).

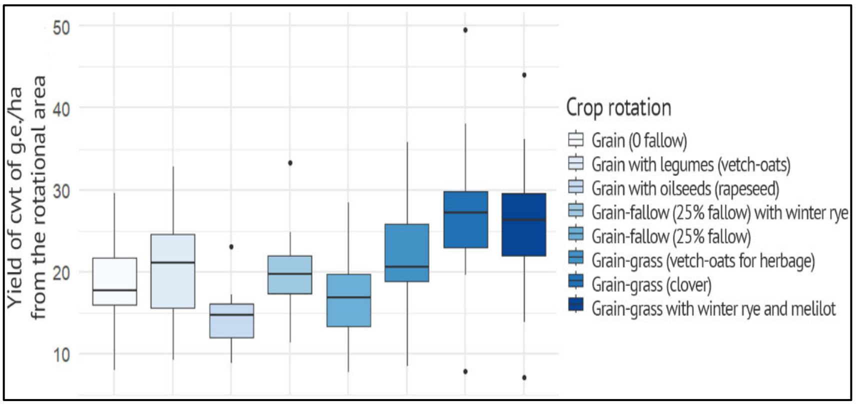

Fig. 1. Boxplot diagram of crop rotation productivity data.

The least variation in crop rotation productivity over time was observed for grain (0 fallow), grain-fallow (25% fallow), grain-fallow (25% fallow) with winter rye, and grain with legumes (Figure 1). In a significantly drought year (1999; SPI = -1.9), the productivity of these crop rotations ranged from 18.0 to 20.4 dt g.e./ha of crop rotation area, which was, on average, 2.4 dt g.e./ha higher than in other crop rotations. In an extremely drought year (2012; SPI = -2.9), the productivity in these rotations ranged from 7.7 to 11.3 dt g.e./ha of crop rotation area, which was, on average, 1.4 dt g.e./ha higher than in other types of crop rotations. The highest productivity values were observed in years with moderate moisture conditions (2018, SPI = 1.01) and within normal parameters in 2017 (SPI = 0.8) for grain-grass rotations with clover (30.2-32.1 dt g.e./ha) and grain-grass rotations for green mass or herbage (31-31.9 dt g.e./ha).

The greatest variation in multi-year crop rotation productivity was observed for all grain-grass crop rotations (Table 1).

Table 1. Assessment of variation in crop rotation productivity.

Crop rotation	Variation coefficient	SD
Grain (0 fallow)	29.0	5.27
Grain with legumes (vetch-oats)	30.5	6.16
Grain with oilseeds (rapeseed)	31.5	4.62
Grain-fallow (25% fallow) with winter rye	22.5	4.52
Grain-fallow (25% fallow)	30.3	5.22
Grain-grass (vetch-oats for herbage)	33.6	7.28
Grain-grass (clover)	34.6	9.42
Grain-grass with winter rye and melilot	30.9	7.95

The relatively low variation in crop rotation productivity was common for grain (0 fallow), grain-fallow (25% fallow), grain-fallow (25% fallow) with winter rye, and grain with legumes. A lower standard deviation (SD) indicates a smaller difference between productivity values in different years and, therefore, a more stable crop rotation productivity. In our case, the aforementioned crop rotations also exhibited relatively low standard deviation values.

Based on the results of the graphical analysis of time series (Figure 2), the least variability and a more stable pattern of productivity data representation were characteristic of grain (0 fallow), grain-fallow (25% fallow), grain-fallow (25% fallow) with winter rye, and grain with legumes crop rotations, indicating their relative stability in terms of actual productivity indicators.

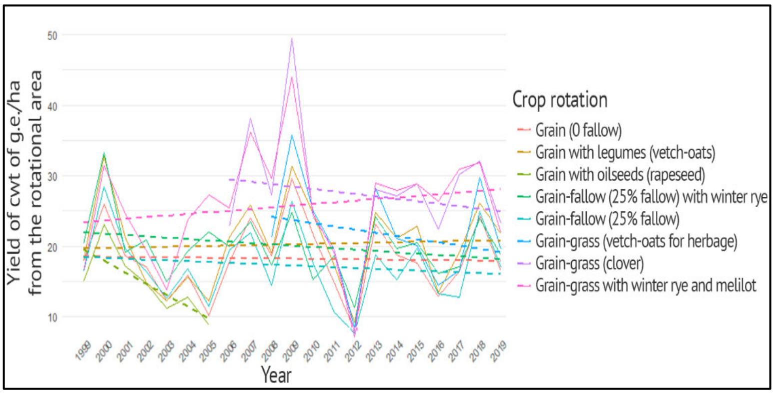

Fig. 2. Graphical analysis of crop rotation productivity variation.

More stable trend patterns were characteristic of grain-fallow (25% fallow), grain-fallow (25% fallow) with winter rye, and grain with legumes crop rotations, which also confirms the aforementioned assumption about the stability of these crop rotations.

In earlier studies [22-24] evaluating the resistance of crop rotations to the most harmful biotic stressors – weeds, it was shown that grain-fallow crop rotations also reacted to increased weed infestation with a lesser decline in crop rotation productivity. Crop rotations with legumes and cabbage were in the second group in terms of resistance, with productivity decline depending on the increase in weed biomass ranging from 0.46 to 0.58 dt g.e./ha. It should also be noted that under high crop infestation, these rotations coped more successfully with the weed component due to increased crop competitiveness. The grain crop rotation (0 fallow) was the least balanced and characterized by the highest productivity losses. Without the application of mineral fertilizers, grain-fallow crop rotations were preferred [22]. Regarding grain yield per hectare of crop rotation area without the application of fertilizers and pesticides, the most productive were grain-fallow crop rotations with winter rye (24.6 dt/ha) and grain with legumes (24.4 dt/ha) [22-24]. The advantages of grain-fallow crop rotations in increasing the productivity of arable land and the technological properties of grain crops in different agricultural systems, through the optimization of black soil fertility elements and the phytosanitary state of agrophytocenosis, have been noted by many authors [25-30].

Using polynomial regression and actual data on crop rotation productivity (Figure 3), the relationship and variation of indicators were studied when different values of atmospheric moisture were alternated, expressed through the Standardized Precipitation Index (SPI) during various vegetation periods. The graphs show that the influence of SPI on the target variable is not linear but rather has a more complex form, confirming the use of polynomial regression. By analyzing polynomials and cross-validation, the most suitable polynomial degrees for each SPI were determined: SPI May – 13, SPI June – 13, SPI July – 12, SPI Vegetation (May-July) – 12. For each polynomial degree, the Root Mean Square Error

(RMSE) was calculated, which is a measure of model accuracy. This allowed the selection of the most accurate models that take into account the nonlinear relationship between SPI and the target variable.

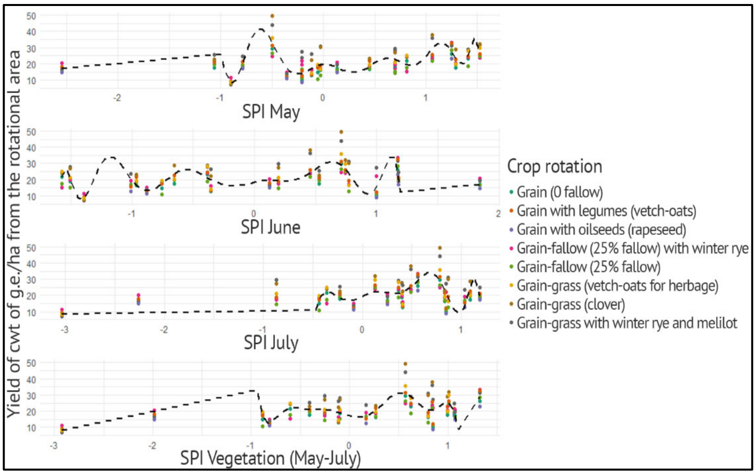

Fig. 3. Scatter plot of variables containing crop rotation productivity values and SPI index over the vegetation period with an overlaid regression spline.

Using the Spearman rank correlation coefficient, a relatively high statistical significance of the influence of SPI during specific vegetation periods on crop rotation productivity was found (SPI May – 0.47, $p < 0.001$, and SPI May-July – 0.43, $p < 0.001$). This relationship was utilized in constructing a decision tree (CART) (Figure 4).

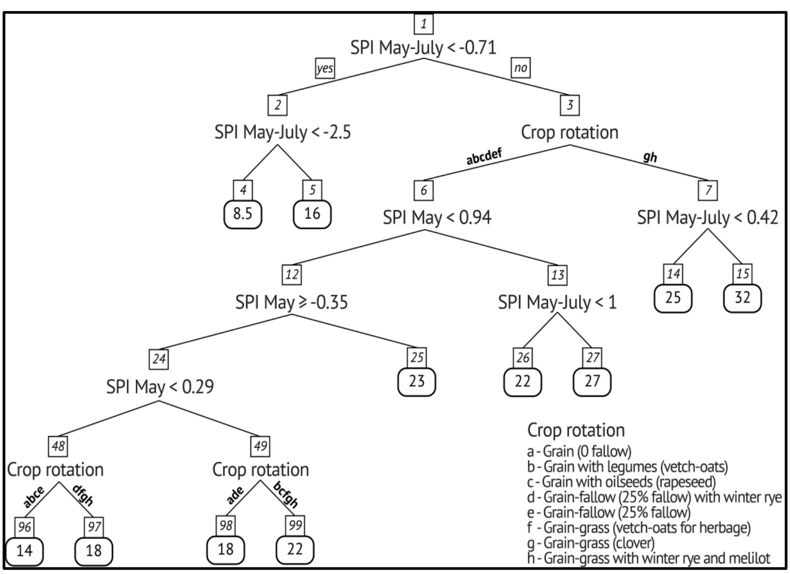

Fig. 4. Decision tree visualization for selecting crop rotation and predicting productivity values (At the top is the tree node number. This node inquires whether the specified partitioning condition expression is true. If yes, the tree descends to the left child node of the root, otherwise it descends to the right child node. Yield of cwt of g.e./ha from the rotational area shown below).

In the gradient of atmospheric moisture during different vegetation periods, a set of logical rules allows identifying crop rotations with optimal potential productivity parameters.

For example:

– SPI May-July < -2.5 (extremely dry conditions) – low productivity for all crop rotations.
– SPI May-July > -0.71(range of average long-term moisture conditions) – high productivity is characteristic of grass-grain (clover) and grass-grain (herbage or cover crop) rotations.
– SPI May-July > -0.71; -0.35 <= SPI May < 0.29 (range of average long-term moisture conditions) – maximum productivity will be characteristic of grain-fallow (25% fallow) with winter wheat and majority types of grass-based crop rotations.

Variations in choosing crop rotations are possible depending on their potential stability under evolving moisture conditions and the level of agricultural practices in specific farms.

The constructed model based on a simple decision tree demonstrated relatively high predictive capability (R2 = 0.56, RMSE = 6.84, MAE = 4.55). However, to increase the model's accuracy, an ensemble approach was applied by building multiple decision trees (Random Forest algorithm) based on the same predictors. This led to an improved model accuracy. The Random Forest algorithm showed (Figure 5) higher accuracy in estimating potential productivity depending on varying conditions during different vegetation periods.

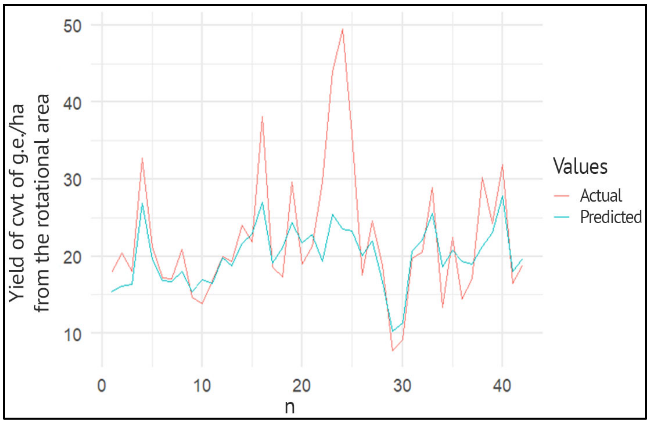

Fig. 5. Results of modeling crop rotation productivity values using the Random Forest method.

Approximately 62.77% of the variations in the target variable (crop rotation productivity) can be explained by the Random Forest model. The Root Mean Square Error (RMSE) is 6.44, the Mean Absolute Error (MAE) is 3.99, and the coefficient of determination (R^2) is 0.60.

Based on the developed model, it becomes possible to select crop rotations under conditions of alternating different categories of atmospheric moisture and assess the preliminary productivity of their types. The set of logical rules allows evaluating stable or most productive crop rotation types under different moisture variations (Figure 4). This model will enable more efficient planning of technological resources according to the prevailing moisture conditions during various vegetation periods and can become an integral part of decision support systems [31-33].

4 Conclusion

Using machine learning methods (Random Forest and CART algorithms), an attempt has been made to address one of the pressing challenges in agriculture – the selection of crop rotations for sustainable grain production in the Siberian region. The main predictor chosen for this study is the atmospheric moisture index expressed as the Standardized Precipitation Index (SPI) during various vegetation periods. The effectiveness of employing an ensemble approach was demonstrated, highlighting the model's highest forecasting capabilities. As a result of this work, a model with sufficiently high predictive ability ($R^2 = 0.60$, RMSE = 6.44, MAE = 3.99) has been developed for selecting crop rotation types based on their stable productivity over time. Among the most suitable crop rotations identified are grain-fallow (25% fallow), grain-fallow with winter wheat, and grain-legume rotations.

This research was carried out with financial support from the grant of the President of the Russian Federation for the state support of leading scientific schools, grant number NSH-1129.2022.2.

References

1. V.I. Kiryushin, *Agroecological assessment of land, designing adaptive landscape agricultural systems, and agrotechnologies. Methodological guide* (Moscow, Federal State Budgetary Institution "Rosinformagrotekh", 2005)
2. V.K. Kalichkin, A.I. Pavlova, Siberian Bull. Agric. Sci. **12**, 5-11 (2008)
3. V.A. Tyulin, V.P. Sutyagin, Int. J. Appl. Fund. Res. **10**, 297-301 (2017)
4. V.I. Kiryushin, I.L. Frumin, Proc. Timiryazev Agric. Acad. **2**, 27-36 (2004)
5. I.Y. Gorodetsky, A.E. Sergeev, *Mathematics in land management*, Fundamental and Applied Scientific Research: Current Issues, Collection of Articles of the 20th International Scientific and Practical Conference: MCNS "Science and Enlightenment", Penza, Russia (2019)
6. A.I. Yuzhakov, V.K. Kalichkin, G.M. Zakharov, *Economic-mathematical model of crop rotation selection*, Problems of Development and Scientific Support of the Agro-Industrial Complex of Yakutia, Materials of the Joint Meeting of the Presidium of the Russian Academy of Agricultural Sciences, the Presidium of the Siberian Branch of the Russian Academy of Agricultural Sciences, and the Council of the Far Eastern Research and Methodological Center, August 3-5, Yakutsk, Russia (2005)
7. I.I. Vasenev, G.N. Cherkasov, *Information and reference systems for the optimization of land use in the conditions of central emergency protection* (Kursk, All-Russian Research Institute of Agriculture and Soil Protection from Erosion, 2002)
8. V.K. Kalichkin, R.A. Koryakin, T.A. Luzhnykh et al., IJRTE **8(4)**, 6059-6063 (2019). https://www.doi.org/10.35940/ijrte.D8638.118419
9. C. Pahmeyer, T. Kuhn, W. Britz, Comput. Electron. Agric. **181**, 105948 (2021). https://www.doi.org/10.1016/j.compag.2020.105948
10. J.L.E.K. Fendji, C.T. Kenmogne, D.J. Fotsa-Mbogne et al., Appl. Sci. **11(15)**, 6775 (2021). https://www.doi.org/10.3390/app11156775
11. C. Lücken, A. Acosta, N. Rojas, *Solving a Many-Objective Crop Rotation Problem with Evolutionary Algorithms*, Intelligent Decision Technologies: Proceedings of the 13th KES-IDT 2021 Conference, Singapore (2021). https://www.doi.org/ 10.1007/978-981-16-2765-1_5
12. L. Benos, A.C. Tagarakis, G. Dolias et al., Sensors **21(11)** 3758 (2021). https://www.doi.org/10.3390/s21113758
13. K. Liakos, P. Busato, D. Moshou et al., Sensors **18(8)**, 2674 (2018). https://www.doi.org/10.3390/s18082674

14. Y.Y. Song, L.U. Ying, Shanghai Arch. Psychiatry **27(2)**, 130-135 (2015). https://www.doi.org/10.11919/j.issn.1002-0829.215044
15. R.J. Lewis, *An Introduction to Classification and Regression Tree (CART) Analysis*, in Proceedings of Annual Meeting of the Society for Academic Emergency Medicine, May 22-25, 2000, San Francisco, CA, USA (2000)
16. A. Priyam, Abhijeeta, R. Gupta et al., Int. J. Curr. Eng. Technol. **3(2)**, 334-337 (2013)
17. F.S. Roberts et al., *Discrete Mathematical Models with Applications to Social, Biological, and Environmental Problems* (Nauka, 1986)
18. I.Ya. Liepa, *Mathematical methods in biological research. Factor and component analysis* (Riga, 1980)
19. K.Y. Maksimovich, V.K. Kalichkin, D.S. Fedorov, V.V. Aleschenko, Bull. Agric. Stavropol Region **2(46)**, 32-38 (2022)
20. D.S. Fedorov, O.K. Alsova, V.K. Kalichkin, K.Y. Maksimovich, *Program for Crop Yield Analysis & Forecast (CYAF),* Certificate of official registration of software for computers No. 2021667894, November 8, 2021 (2021)
21. V.K. Kalichkin, D.S. Fedorov, O.K. Alsova, K.Y. Maksimovich, Achiev. Sci. Technol. AIC **36(1)**, 51-56 (2022). https://www.doi.org/10.53859/02352451_2022_36_0_0
22. K.Y. Maksimovich, V.K. Kalichkin, E.V. Rudoy, V.V. Aleschenko, *Crop Agro-Technologies Adaptation in Organic Farming*, VIII International Conference on Advanced Agritechnologies, Environmental Engineering and Sustainable Development, AGRITECH-VIII 2023, 390 (2023). https://www.doi.org/10.1051/e3sconf/202339005007
23. G.M. Zakharov, V.K. Kalichkin, A.E. Malygin et al., Siberian Bull. Agric. Sci. **1**, 5-13 (2012)
24. A.L. Pakul, N.A. Lapshinov, G.V. Bozhanova et al., Siberian Bull. Agric. Sci. **48(6)**, 21-29 (2019). https://www.doi.org/10.26898/0370-8799-2018-6-3
25. V.V. Chibis, Bull. Krasnoyarsk State Agrar. Univ. **5(182)**, 51-57 (2022)
26. L.V. Yushkevich, V.V. Chibis, Bull. Krasnoyarsk State Agrar. Univ. **9(174)**, 35-43 (2021)
27. I.F. Khramtsov, N.P. Drankovich, *Technological Systems of Cereal and Leguminous Crops Cultivation: Recommendations* (Omsk: LITERA, 2014)
28. I.F. Khramtsov, N.A. Voronkova, 75 Years Geogr. Netw. Experim. Fertilizers, 303-309 (2016)
29. Z.M. Azizov, V.V. Arkhipov, I.G. Imashev, Agrar. Sci. J. **7**, 4-9 (2020)
30. V.I. Kiryushin, A.N. Vlasenko, *Adaptive-Landscape Systems of the Novosibirsk Region* (Novosibirsk, 2002)
31. V.P. Yakushev, V.V. Yakushev, V.L. Badenko, Agric. Biol. **55(3)**, 451-467 (2020). https://www.doi.org/10.15389/agrobiology.2020
32. V.V. Chibis, Living Biocenotic Syst. **20**, 5-5 (2017)
33. V. Riksen, K. Maksimovich, T. Kizimova, R.Galimov, and D. Fedorov, *Elements of the Decision Support System in the Agricultural Production Processes*, Agriculture Digitalization and Organic Production: Proceedings of the Second International Conference, ADOP 2022, June 06–08, 2022, St. Petersburg, Russia (2022)