https://doi.org/10.26898/0370-8799-2021-5-11 Тип статьи: оригинальная
УДК: 631.3:004.04 Type of article: original

ПРОГНОЗИРОВАНИЕ СОДЕРЖАНИЯ НИТРАТНОГО АЗОТА В ПОЧВЕ С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ

Каличкин В.К., (✉) Лужных Т.А., Риксен В.С., Васильева Н.В., Шпак В.А.
Сибирский федеральный научный центр агробиотехнологий Российской академии наук
Новосибирская область, р.п. Краснообск, Россия
(✉) e-mail: tanya.luzhnykh@mail.ru

Исследованы возможности и целесообразность применения Байесовской сети доверия и логистической регрессии для прогнозирования содержания нитратного азота в слое почвы 0-40 см перед посевом. Для обучения моделей использованы данные длительного многофакторного полевого опыта Сибирского научно-исследовательского института земледелия и химизации сельского хозяйства СФНЦА РАН за 2013–2018 гг. Опыт заложен на черноземе выщелоченном на территории центрально-лесостепной подзоны в 1981 г. в Новосибирской области. Учитывая особенности статистической выборки (данных наблюдений и анализов), определены основные предикторы моделей, влияющие на содержание нитратного азота в почве. Байесовская сеть доверия построена в виде ациклического графа, в котором обозначаются главные (основные) узлы и их взаимоотношения. Узлы сети представлены качественными и количественными параметрами рабочего участка (подтип почвы, предшественник, обработка почвы, погодные условия) с соответствующими градациями (событиями). В результате заполнения экспертами таблицы условных вероятностей с учетом анализа эмпирических данных сеть присваивает апостериорную вероятность наступления событий для целевого узла (содержание нитратного азота в слое почвы 0–40 см). Для проверки устойчивости работы сети проанализированы два сценария развития событий, получены удовлетворительные показатели. В результате построения логистической регрессии получены коэффициенты, характеризующие тесноту связи между зависимой переменой и предикторами. Коэффициент детерминации логистической регрессии равен 0,7. Это свидетельствует о том, что качество модели можно считать допустимым для прогнозирования. Дана сравнительная оценка прогностических возможностей обученных моделей. Общая доля правильных прогнозов для Байесовской сети доверия составляет 84%, для логистической регрессии – 87%.

Ключевые слова: Байесовская сеть, регрессионный анализ, нитратный азот, почва

PREDICTION OF NITRATE NITROGEN CONTENT IN SOIL USING MACHINE LEARNING

Kalichkin V.K., (✉) Luzhnykh T.A., Riksen V.S., Vasilyeva N.V., Shpak V.A.
Siberian Federal Scientific Centre of Agro–BioTechnologies of the Russian Academy of Sciences
Krasnoobsk, Novosibirsk Region, Russia
(✉) e-mail: tanya.luzhnykh@mail.ru

The possibilities and feasibility of using the Bayesian network of trust and logistic regression to predict the content of nitrate nitrogen in the 0-40 cm soil layer before sowing have been investigated. Data from long-term multifactor field experience at the Siberian Research Institute of Farming and Agricultural Chemization of SFSCA RAS for 2013-2018 were used to train the models. The experiment was established on leached chernozem in the central forest-steppe subzone in 1981 in the Novosibirsk region. Considering the characteristics of the statistical sample (observation and analysis data), the main predictors of the models affecting nitrate nitrogen content in soil were identified. The Bayesian trust network is constructed as an acyclic graph, in which the main (basic) nodes and their relationships are denoted. Network nodes are represented by qualitative and quantitative plot parameters (soil subtype, forecrop, tillage, weather conditions) with corresponding gradations (events). The network assigns a posteriori probability of events for the target node (nitrate-nitrogen content in the 0-40 cm soil layer) as a result of experts completing the conditional probability table, taking into account the analysis of empirical data. Two scenarios were analyzed to test the sustainability of the network and satisfactory results were obtained. The result of the logistic regression is the coefficients characterizing the closeness of the relationship between the dependent variable and the predictors. The coefficient of determination of the logistic regression is 0.7. This indicates that the quality of the model can be considered acceptable for forecasting. A comparative assessment of the predictive capabilities of the trained models is given. The overall proportion of correct predictions for the Bayesian confidence network is 84%, for logistic regression it is 87%.

Keywords: Bayesian network, regression analysis, nitrate nitrogen, soil

Для цитирования: *Каличкин В.К., Лужных Т.А., Риксен В.С., Васильева Н.В., Шпак В.А.* Прогнозирование содержания нитратного азота в почве с использованием машинного обучения // Сибирский вестник сельскохозяйственной науки. 2021. Т. 51. № 5. С. 91–100. https://doi.org/10.26898/0370-8799-2021-5-11

For citation: Kalichkin V.K., Luzhnykh T.A., Riksen V.S., Vasilyeva N.V., Shpak V.A. Prediction of nitrate nitrogen content in soil using machine learning. *Sibirskii vestnik sel'skokhozyaistvennoi nauki = Siberian Herald of Agricultural Science,* 2021, vol. 51, no. 5, pp. 91–100. https://doi.org/10.26898/0370-8799-2021-5-11

Конфликт интересов
Авторы заявляют об отсутствии конфликта интересов.
Conflict of interest
The authors declare no conflict of interest.

INTRODUCTION

Agriculture 4.0, as the fourth stage of its evolution, puts forward the requirements for creating and mastering modern information systems for decision support using a set of digital technologies [1]. Predictive analytics in these systems should take a leading position, because without predicting the transformation of conditions, objects and processes occurring in agriculture, it is difficult or almost impossible to make the right decision on its management [2, 3].

As in many other areas of human activity, the amount of information on agriculture is constantly increasing. Various machine learning algorithms are being used to evaluate information from a variety of sources in order to analyze data and assist agricultural professionals in solving specific problems, while improving efficiency. The application of machine learning in agriculture is currently accompanied by massive interest from the global scientific community. Various models are used in machine learning: regression, clustering, Bayesian and neural networks, support vector machines, decision trees, etc. The appropriateness of applying one or another machine learning model is determined by the different types of agricultural data and the problems to be solved [4-6].

One of the tools of artificial intelligence and the possibility of application in machine learning are considered to be Bayesian Belief networks (BBN). A review article [7] noted that the BBN method is suitable for research in agriculture, as BBN are able to reason with incomplete information and include new information, as well as solve problems under uncertainty, taking into account cause-effect relationships [8]. Examples of the application of the BBN apparatus in Russian-language publications are considered in medicine, ecology, risk analysis, sociology and other subject areas [9, 10], in the application to agriculture BBN have very limited application [11].

The most popular statistical method used for predictive modelling in agriculture is regression analysis. This method is considered the easiest to use and understand, it allows to investigate the relationship between the dependent (target) and independent (predictor) variables, to identify significant patterns in general form, to determine the closeness of the relationship of the studied factors [12].

The purpose of the research is to train different models for the analysis of empirical data, to carry out a prediction of the nitrate nitrogen content in the soil before sowing, and to evaluate the accuracy of the predictive models.

MATERIAL AND METHODS

In model training, data from a long-term multi-factor stationary field experiment of the Siberian Research Institute of Soil Management and Chemicalization of Agriculture of SFSCA RAS were used. The experiment was conducted in 1981 on the territory of the farm "Elitnaya" - a branch of SFSCA RAS in the Novosibirsk region in the central forest-steppe subzone. The data include the results of research (2013-2018) of a four-field grain and fallow crop rotation (fallow - wheat - wheat - wheat). The experiments were carried out with different variants of the main tillage:

– ploughing (for the 1st and 3rd crops to a depth of 20-22 cm, for the 2nd - 25-27 cm);

– non-moldboard tillage (non-moldboard loosening with SibIME tines for the 1st and 3rd culture to a depth of 20-22 cm, under the 2nd - 25-27 cm);

– minimum tillage (stubble-mulch to a depth of 10-12 cm for all crops annually);

– zero tillage (no under-winter plowing).

The stationary soil is leached chernozem of medium-loam granulometric composition. Modelling was carried out using data on nitrate nitrogen content in the 0-40 cm soil layer before sowing.

The training of the BBN was carried out in the Netica software version "6.07", the logistic regression model was implemented in the SPSS module package version "26". In modelling, 80% of the original sample data were used for model training and 20% for testing (forecasting).

The dimensioning of the nitrate nitrogen level in the soil was set according to A.E. Kochergin's scale.

RESULTS AND DISCUSSION

BBN construction. BBN is an oriented acyclic graph, each vertex of which corresponds to a random variable, the arcs of the graph encode the conditional independence relations between these variables. The vertices can represent variables of any type, be weighted parameters, latent variables or hypotheses [13]. BBN are probabilistic because they are based on probability distributions and use probability theory to make predictions. Some data or expert knowledge (heuristics) are used to train and run a BBN. A prediction model is based on Bayes formula, which determines the probability of an event occurring, assuming that another event that is interdependent with it has already occurred.

The construction of a BBN begins with the definition of the graph structure, in which the main (main) nodes and their parameters are identified.

Based on the data structure, it is expertly assumed that the amount of nitrate nitrogen in the soil is dependent on weather conditions and farming practices.

The main nodes of the network are then represented by the qualitative and quantitative parameters of the working area (soil subtype, forecrop, tillage, weather conditions) with the corresponding gradations (events):

– soil sub-type - discrete variable with one gradation – Leached chernozem;

– soil tillage with four gradations – Plowing, Non-mouldboard tillage, Minimum tillage, Zero tillage;

– forecrop with four gradations – Fallow, fallow wheat 1, fallow wheat 2, fallow wheat 3;

– weather conditions with two gradations – Favorable, Unfavorable;

– nitrate nitrogen content in the 0-40 cm soil layer with two gradations - Less than 10 (less than 10 mg/kg soil) and More than 10 (more than 10 mg/kg soil). This node is the target node.

The next step was to construct an acyclic graph and arrange causal relationships to the target node, which is a prerequisite for further completion of the conditional probability table (CPT), taking into account the analysis of the data obtained. The constructed graph consists of four main nodes (soil subtype, weather conditions, forecrop, tillage) and one target node (nitrate nitrogen content). At this stage, the system sets up a uniform probability distribution for all the nodes (see Figure 1).

Once the structure of the graph has been drawn up and the cause-and-effect relationships have been defined in the form of arrows from the main nodes to the target node, the CPT is constructed and completed according to the logic of the model (see Figure 2).

In CPT, the program automatically builds possible combinations of random events of the major nodes affecting the two events of the target node (nitrate nitrogen content - Less 10, More 10). In doing so, experts assign a priori probability of occurrence of each of the two events of the target node to each combination of random events of the main nodes, thereby training the model. The higher the percentage, the greater the probability that a given event will occur. In completing the CPT, the experts were guided not only by heuristics (knowledge of the problem) but also by empirical data from field experience. Once the CPT is fully constructed and populated, the BBN is compiled and the network is ready to use.

Before giving a forecast for 2021 for nitrate nitrogen content in the 0-40 cm soil layer, the

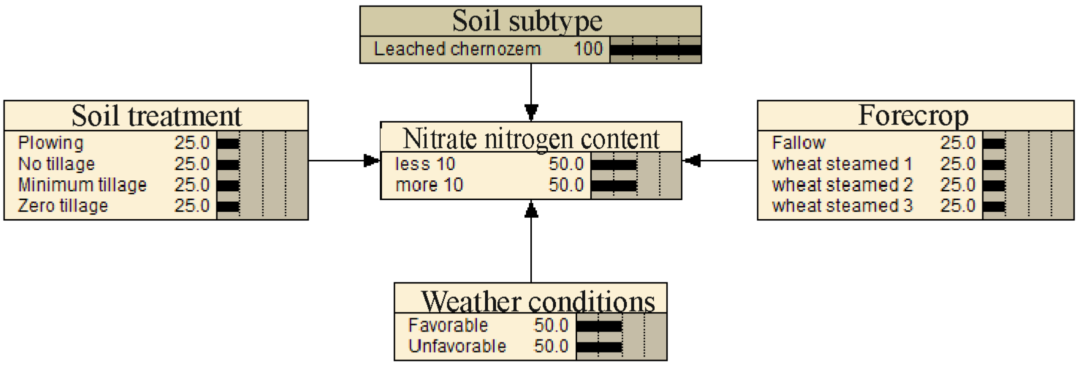
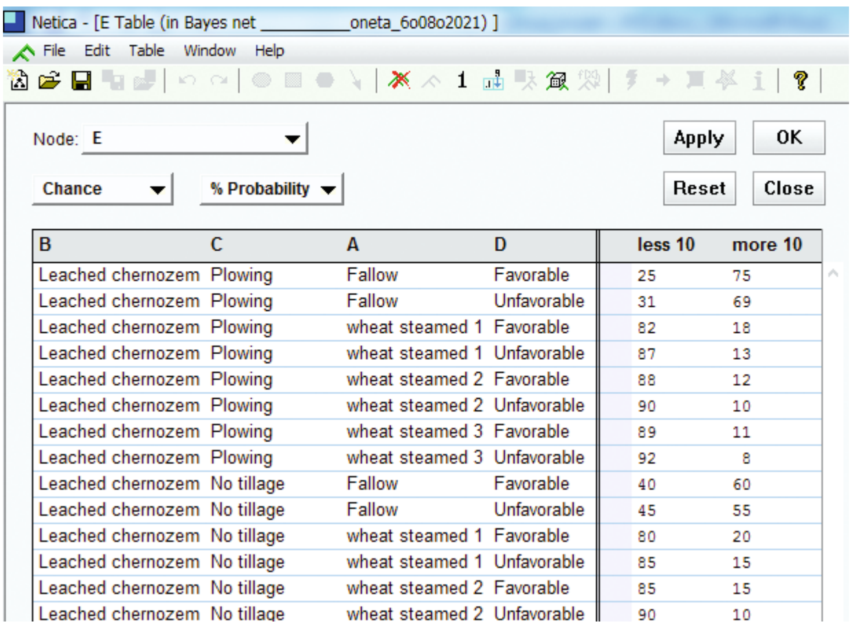


Рис. 1. Структура ациклического графа.
Fig. 1. The structure of an acyclic graph.



B	C	A	D	less 10	more 10
Leached chernozem	Plowing	Fallow	Favorable	25	75
Leached chernozem	Plowing	Fallow	Unfavorable	31	69
Leached chernozem	Plowing	wheat steamed 1	Favorable	82	18
Leached chernozem	Plowing	wheat steamed 1	Unfavorable	87	13
Leached chernozem	Plowing	wheat steamed 2	Favorable	88	12
Leached chernozem	Plowing	wheat steamed 2	Unfavorable	90	10
Leached chernozem	Plowing	wheat steamed 3	Favorable	89	11
Leached chernozem	Plowing	wheat steamed 3	Unfavorable	92	8
Leached chernozem	No tillage	Fallow	Favorable	40	60
Leached chernozem	No tillage	Fallow	Unfavorable	45	55
Leached chernozem	No tillage	wheat steamed 1	Favorable	80	20
Leached chernozem	No tillage	wheat steamed 1	Unfavorable	85	15
Leached chernozem	No tillage	wheat steamed 2	Favorable	85	15
Leached chernozem	No tillage	wheat steamed 2	Unfavorable	90	10

Рис. 2. Фрагмент таблицы условных вероятностей
Fig. 2. A fragment of the conditional probability table

network was trained to determine its behavior during changes in the events of the main nodes and to obtain the posterior probability of the target node. For this purpose, a situation was modelled (first scenario) in which the conditions on the working plot were as follows: forecrop - Fallow, tillage – Mouldboard Plowing, weather conditions - Unfavorable (see Figure 3).

In the first scenario, the BBN predicts with a 69% probability that the nitrate nitrogen content in the 0-40 cm soil layer will be more than 10 mg/kg.

In order to check the stability of the network, the second scenario was analyzed (node events changed): forecrop – fallow wheat 3, tillage - Zero tillage, weather conditions - Favorable (favorable) (see Figure 4).

Under the second scenario, the BBN predicts that there is a 95% probability that the nitrate nitrogen content in the 0-40 cm soil layer will be less than 10 mg/kg.

A sensitivity analysis has also been conducted for the BBN. A sensitivity analysis function is used to determine the magnitude and extent of the impact of the main nodes on the target node events in a descending order (see Table 1).

The mutual information between the two nodes indicates how dependent these nodes are on each other. If it exists, it shows how close their relationship is. The highest rate of mutual information is obtained for the forecrop node. This means that the node has the greatest influence on the target indicator (nitrate nitrogen content in the 0-40 cm soil layer). Weather conditions and soil tillage have much less influence on the target indicator.

Construction of a logistic regression model. To predict the target indicator (nitrate nitrogen content in the 0-40 cm soil layer), the relationship of this indicator (dependent variable) to the following independent variables (factors) was investigated: forecrop, tillage method, weather conditions (precipitation in April-May and September, average monthly air temperature for the same months).

The dependent variable can be categorized as a categorical variable when the values of nitrate nitrogen in the 0-40 cm soil layer take the form - Less 10 and More 10. In this case, the desired relationship can be obtained using a logistic regression model [14]. In this case, one of the categories of the dependent variable becomes the reference variable and the other is compared with it. The independent variables can be categorical or quantitative. The logistic regression equation predicts the probability of belonging to each category of dependent variable by the values of independent variables. The final selection of the predicted category for the dependent variable is made according to the rule of most likely membership.

Data generated in the form of a table of 72 observations (rows) and 9 factors (columns), including the dependent variable, were used to obtain the parameters (coefficients) of the logistic regression (see Figure 5).

The number of observations is distributed by the nitrate nitrogen content in the 0-40 cm soil

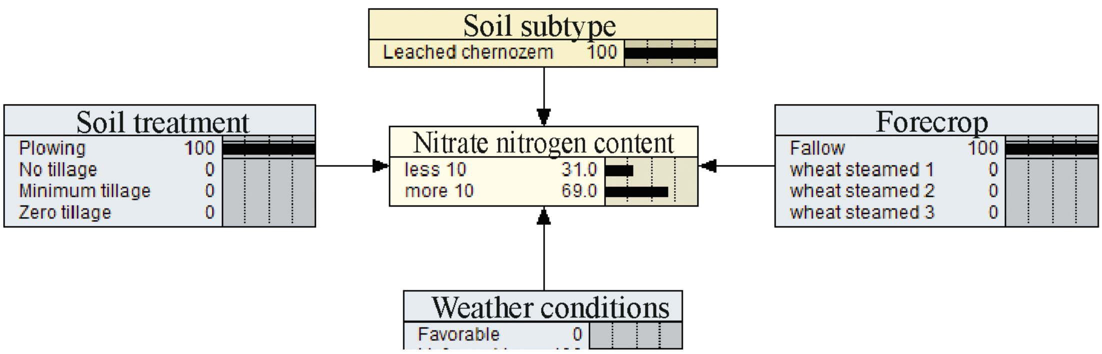


Рис. 3. Первый сценарий БСД

Fig. 3. The first DAG scenario

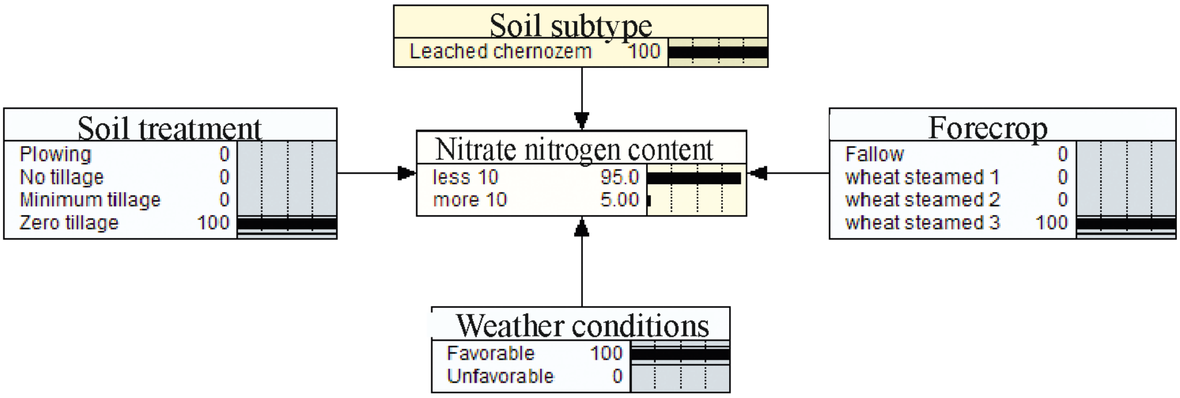


***Рис.* 4.** Второй сценарий БСД

***Fig.* 4.** The second DAG scenario

Табл. 1. Результаты анализа чувствительности узла содержания нитратного азота в слое почвы 0–40 см

Table 1. Results of the sensitivity analysis of the nitrate nitrogen content node in the 0–40 cm soil layer

Node	Mutual information	Percentage	Belief dispersion
Forecrop	0,308	33,9	0,08366
Weather conditions	0,00174	0,204	0,00048
Soil tillage	0,00084	0,0987	0,0002322
Soil subtype	0	0	0

layer as follows: Less 10 - 53, More 10 - 19. More 10 category was selected as a reference category.

Table 2 presents the coefficients obtained by means of logistic regression. Based on the logistic regression calculations, we can conclude that the strongest influence on the resulting variable is exerted by fallow as a forecrop. If the significance of the coefficient is $P < 0.05$, the relationship is statistically significant. A result of $P>0.05$ indicates that the relationship between the variables is weak or not detected. Third fallow wheat and zero tillage do not affect the nitrate nitrogen content in the 0-40 cm soil layer.

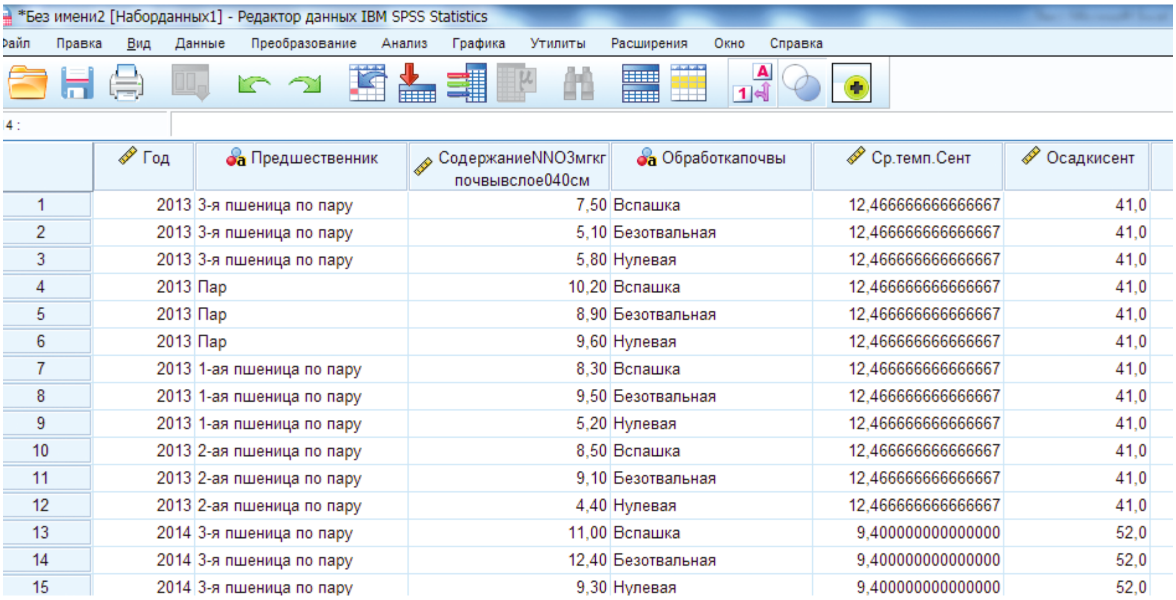
A measure of the adequacy (quality) of the constructed logistic regression model is the pseudo R-squared coefficient. In this case, the resulting variance on the Nagelkerke measure (usually the most used) is 70%, indicating the satisfactory predictive power of the model.

Unlike conventional regression, a logistic regression model does not predict the value of a numerical variable based on a sample of initial values. Instead, the value of the function is the probability that a given initial value belongs to a particular category.

Thus, in the course of the research, BBN and logistic regression models were built and trained, predicting nitrate nitrogen content in 0-40 cm layer of leached chernozem before sowing. The forecast for 2021 was carried out for the experiments presented in Table 3. The weather conditions this year were unfavorable.

The prediction made with the help of BBN shows accumulation of nitrate nitrogen in soil layer 0-40 cm for fallow precedence in all variants is more than 10 mg/kg (with 69% probability for ploughing, 55% for non-mouldboard, 65% for zero tillage). For all other variants, nitrate nitrogen content of less than 10 mg/kg is predicted. In contrast to BBN, logistic regression prediction shows nitrate nitrogen content in 0-40 cm layer in all variants not more than 10 mg/kg.

The deviation (error) of the actual nitrogen content from the predicted one was determined as a criterion for evaluating the predictive models. Table 4, 5 present comparative predictive capabilities of the models tested on the initial sample. The initial sample consists of 72 observations, including 53 observations with nitrate nitrogen content Less than 10 and 19 observa-



*Без имени2 [Наборданных1] - Редактор данных IBM SPSS Statistics

Файл Правка Вид Данные Преобразование Анализ Графика Утилиты Расширения Окно Справка

	Год	Предшественник	СодержаниеNNO3мгкг почвывслое040см	Обработкапочвы	Ср.темп.Сент	Осадкисент
1	2013	3-я пшеница по пару	7,50	Вспашка	12,466666666666667	41,0
2	2013	3-я пшеница по пару	5,10	Безотвальная	12,466666666666667	41,0
3	2013	3-я пшеница по пару	5,80	Нулевая	12,466666666666667	41,0
4	2013	Пар	10,20	Вспашка	12,466666666666667	41,0
5	2013	Пар	8,90	Безотвальная	12,466666666666667	41,0
6	2013	Пар	9,60	Нулевая	12,466666666666667	41,0
7	2013	1-ая пшеница по пару	8,30	Вспашка	12,466666666666667	41,0
8	2013	1-ая пшеница по пару	9,50	Безотвальная	12,466666666666667	41,0
9	2013	1-ая пшеница по пару	5,20	Нулевая	12,466666666666667	41,0
10	2013	2-ая пшеница по пару	8,50	Вспашка	12,466666666666667	41,0
11	2013	2-ая пшеница по пару	9,10	Безотвальная	12,466666666666667	41,0
12	2013	2-ая пшеница по пару	4,40	Нулевая	12,466666666666667	41,0
13	2014	3-я пшеница по пару	11,00	Вспашка	9,400000000000000	52,0
14	2014	3-я пшеница по пару	12,40	Безотвальная	9,400000000000000	52,0
15	2014	3-я пшеница по пару	9,30	Нулевая	9,400000000000000	52,0

Рис. 5. Фрагмент таблицы данных в SPSS для построения логистической регрессии
Fig. 5. A fragment of a data table in SPSS for constructing a logistic regression

Табл. 2. Оценка параметров (коэффициентов) логистической регрессии
Table. 2. Estimation of parameters (coefficients) of logistic regression

Factor	Coefficient of variables	Standard error	Coefficient significance (*P*)
Average monthly temperature (September)	–1,7	0,6	0,007
Average monthly rainfall (September)	– 0,16	0,068	0,023
Average monthly temperature (April to May)	–2,08	0,84	0,013
Average monthly rainfall (April to May)	–0,22	0,076	0,003
Forecrop – fallow	5,1	1,7	0,002
Forecrop – 1-st wheat on fallow	–0,613	0,8	0,48
Forecrop – 2-nd wheat on fallow	–1,5	1,03	0,167
Forecrop – 3-rd wheat on fallow	0	0	0
Tillage - ploughing	1,03	0,81	0,27
Tillage – non-mouldboard	–0,58	0,87	0,607
Tillage – zero	0	0	0
Constant	55,1	19,06	0,004

tions with nitrate nitrogen content More than 10.

In BBN, correct predictions for the Less 10 category are as high as 90%, with 48 observations predicted correctly and 5 incorrectly (see Table 4).

For the More 10 category in the BBN, the correct prediction rate is 68%. However, out of 19 observations, 13 were correctly predicted by the model and 6 were incorrectly predicted. For all observations, the overall proportion of correct predictions is 84%. The reliability of the predictions is lower than in logistic regression, but the efficiency of its predictive ability is quite high.

When comparing actual observations of nitrate nitrogen content in the 0-40 cm layer with predicted estimates using logistic regression, it can be noted that the obtained proportion of correct predictions is 94.3% for the category Less 10 (see Table 5). This model correctly predicted 50 observations, three observations were predicted incorrectly, falling into the category More 10.

For the sample of observations with nitrate nitrogen content More 10, the proportion of correct predictions was 68.4%, of which 13 observations were correctly predicted and 6 predictions were incorrect, falling into the category Less 10. The total proportion of correct predictions in all categories was 87%. Thus, given the small size of the statistical sample and the small number of predictors, the predictive properties of the tested models can be considered satisfactory.

CONCLUSION

Using machine learning techniques, models have been built and trained to predict the pre-sowing nitrate nitrogen content in the 0-40 cm soil layer with acceptable confidence. The difference in methodologies is that the BBN in predictive modelling allows for "what-if" scenario analysis, can combine patterns inferred from statistical data and expert knowledge derived from actual data, and is able to handle incomplete and different types of data.

In the future, it is planned to improve the quality of the models by adding other predictors affecting the resultant trait, to search for machine learning methods that allow analysis of small-scale data, and to verify the models with actual data. The possibility of using the built models to develop an expert system for the selection and maintenance of agricultural technologies, which will allow making competent decisions in relation to the objectives of agricultural production is being considered.

СПИСОК ЛИТЕРАТУРЫ

1. *Zhai Z., Fernán J., Beltran M.V., Martínez N.L.* Decision support systems for agriculture 4.0: Survey and challenges // Computers and Electronics in Agriculture. 2020. Vol. 170. P. 105256. DOI: 10.1016/j.compag.2020.105256.

Табл. 3. Прогноз содержания нитратного азота в слое почвы 0–40 см в 2021 г.
Table 3. Forecast of nitrate nitrogen content in the 0–40 cm soil layer in 2021.

Forecrop	Soil tillage	$N\text{-}NO_3$ content (forecast by DAG), mg/kg	N-NO3 content (logistic regression forecast), mg/kg
Fallow	Ploughing	More 10	Less 10
Fallow	Non-mouldboard	More 10	Less 10
Fallow	Zero tillage	More 10	Less 10
1-st wheat on fallow	Ploughing	Less 10	Less 10
1-st wheat on fallow	Non-mouldboard	Less 10	Less 10
1-st wheat on fallow	Zero tillage	Less 10	Less 10
2-nd wheat on fallow	Ploughing	Less 10	Less 10
2-nd wheat on fallow	Non-mouldboard	Less 10	Less 10
2-nd wheat on fallow	Zero tillage	Less 10	Less 10
3-rd wheat on fallow	Ploughing	Less 10	Less 10
3-rd wheat on fallow	Non-mouldboard	Less 10	Less 10
3-rd wheat on fallow	Zero tillage	Less 10	Less 10

Табл. 4. Классификационная таблица БСД
Table 4. DAG classification Table

Number of observations	Predicted		Percentage of correct predictions
	Less 10	More 10	
Less 10 (53)	48	5	90
More 10 (19)	6	13	68
Total share			84

Табл. 5. Классификационная таблица модели логистической регрессии
Table 5. Classification table of the logistic regression model

Number of observations	Predicted		Percentage of correct predictions
	Less 10	More 10	
Less 10 (53)	50	3	94,3
More 10 (19)	6	13	68,4
Total share			87

2. *Bendre M., Manthalkar R.* Time series decomposition and predictive analytics using MapReduce framework // Expert Systems with Applications. 2019. Vol. 116. P. 108–120. DOI:10.1016/j.eswa.2018.09.017.
3. *Khaki S., Wang L., Archontoulis S.V.* A cnn-rnn framework for crop yield prediction // Frontiers in Plant Science. 2020. Vol. 10. P. 1750. DOI: 10.3389/fpls.2019.01750.
4. *Benos L., Tagarakis A.C., Dolias G., Remigio B.* Machine Learning in Agriculture: A Comprehensive Updated Review // Sensors. 2021. Vol. 21. N 11. P. 3758. DOI: 10.3390/s21113758.
5. *Cravero A., Sepúlveda S.* Use and Adaptations of Machine Learning in Big Data – Applications in Real Cases in Agriculture // Electronics. 2021. Vol. 10. N 5. P. 552. DOI: 10.3390/electronics10050552.
6. *Paudel D., Boogaard H., Wit A. Janssen S., Osinga S., Pylianidis C., Athanasiadis I.* Machine learning for large-scale crop yield forecasting // Agricultural Systems. 2021. Vol. 187. P. 103016. DOI: 10.1016/j.agsy.2020.103016.
7. *Drury B., Valverde-Rebaza J., Moura M.-F., Andrade Lopes, A.* A survey of the applications of Bayesian networks in agriculture // Engineering Applications of Artificial Intelligence. 2017. Vol. 65. P. 29–42. DOI: 10.1016/j.engappai.2017.07.003.
8. *Fenton N., Neil M., Lagnado D., Yet B., Constantinou A.C.* How to model mutually exclusive events based on independent causal pathways in Bayesian network models // Knowledge-Based Systems. 2016. Vol. 113. P. 39–50. DOI: 10.1016/j.knosys.2016.09.12.
9. *Айвазян С.А.* Байесовский подход в эконометрическом анализе // Прикладная эконометрика. 2008. № 1 (9). С. 93–130.
10. *Торопова А.В.* Подходы к диагностике согласованности данных в байесовских сетях доверия // Труды СПИИРАН. 2015. Т. 6. № 43. С. 156–178.
11. *Прищепов А.В., Понькина Е.В., Сун Ж., Мюллер Д.* Выявление детерминант урожайности пшеницы в Западной Сибири с использованием байесовских сетей // Пространственная экономика. 2019. Т. 15. № 1. С. 39–83. DOI: 10.14530/se.2019.1.039-083.
12. *Zvizdojevic J., Vukotic M.* Application of statistical methods in analysis of agriculture-correlation and regression analysis // Poljoprivreda i Sumarstvo. 2015. Vol. 61. N 1. P. 309–323. DOI: 10.17707/AgricultForest.61.1.38.
13. *Литвиненко Н.Г., Литвиненко А.Г., Мамырбаев О.Ж.* Байесовские сети. Теория и практика: монография. Алматы: Институт информационных и вычислительных технологий КН МОН РК, 2020. 197 с.
14. *Наследов А.Д.* IBM SPSS Statistics 20 и AMOS // Профессиональный статистический анализ данных. СПб.: Питер, 2013. 416 с.

REFERENCES

1. Zhai Z., Fernán J., Beltran M.V., Martínez N.L. Decision support systems for agriculture 4.0: Survey and challenges. *Computers and Electronics in Agriculture,* 2020, vol. 170, p. 105256. DOI: 10.1016/j.compag.2020.105256.
2. Bendre M., Manthalkar R. Time series decomposition and predictive analytics using MapReduce framework. *Expert Systems with Applications,* 2019, vol. 116, pp. 108–120. DOI: 10.1016/j.eswa.2018.09.017.
3. Khaki S., Wang L., Archontoulis S.V. A cnn-rnn framework for crop yield prediction. *Frontiers in Plant Science,* 2020, vol. 10, p. 1750. DOI: 10.3389/fpls.2019.01750.
4. Benos L., Tagarakis A.C., Dolias G., Remigio B. Machine Learning in Agriculture: A Comprehensive Updated Review. *Sensors,* 2021, vol. 21, no. 11, p. 3758. DOI: 10.3390/s21113758.
5. Cravero A., Sepúlveda S. Use and Adaptations of Machine Learning in Big Data – Applications in Real Cases in Agriculture. *Electronics,* 2021, vol. 10, no. 5, p. 552. DOI: 10.3390/electronics10050552.
6. Paudel D., Boogaard H., Wit A. Janssen S., Osinga S., Pylianidis C., Athanasiadis I. Machine learning for large-scale crop yield forecasting. *Agricultural Systems,* 2021, vol. 187, p. 103016. DOI: 10.1016/j.agsy.2020.103016.
7. Drury B., Valverde-Rebaza J., Moura M.-F., Andrade Lopes, A. A survey of the applications of Bayesian networks in agriculture. *Engineering Applications of Artificial Intelligence,* 2017, vol. 65, pp. 29–42. DOI: 10.1016/j.engappai.2017.07.003.
8. Fenton N., Neil M., Lagnado D., Yet B., Constantinou A.C. How to model mutually exclusive events based on independent causal pathways in Bayesian network models. *Knowledge-Based Systems,* 2016, vol. 113, pp. 39–50. DOI: 10.1016/j.knosys.2016.09.12.
9. Aivazyan S.A. Bayesian approach in econometric analysis. Applied econometrics. *Prikladnaya*

ekonometrika = Applied Econometrics, 2008, no. 1 (9), pp. 93–130. (In Russian).

10. Toropova A.V. Approaches to Diagnosing Data Consistency in Bayesian Trust Networks. *Trudy SPIIRAN = SPIIRAS Proceedings*, 2015, vol. 6, no. 43, pp. 156–178. (In Russian).
11. Prishchepov A.V., Pon'kina E.V., Sun Zh., Myuller D. Identification of determinants of wheat yield in Western Siberia using Bayesian networks. *Prostranstvennaya ekonomika = Spatial Economics*, 2019, vol. 15, no. 1, pp. 39–83. (In Russian). DOI: 10.14530/se.2019.1.039-083.
12. Zvizdojevic J., Vukotic M. Application of statistical methods in analysis of agriculture-correlation and regression analysis. *Poljoprivreda i Sumarstvo,* 2015, vol. 61, no. 1, pp. 309–323. DOI: 10.17707/AgricultForest.61.1.38.
13. Litvinenko N.G., Litvinenko A.G., Mamyrbaev O.Zh. Bayesian networks. *Theory and practice.* Almaty: Institute of Information and Computational Technologies, CS MES RK, 2020, p. 197. (In Russian).
14. Nasledov A.D. IBM SPSS Statistics 20 and AMOS. *Professionalny Statistichesky analiz dannikh = Professional statistical data analysis*, St. Petersburg, Piter, 2013, p. 416. (In Russian).

ИНФОРМАЦИЯ ОБ АВТОРАХ

Каличкин В.К., доктор сельскохозяйственных наук, профессор, главный научный сотрудник; e-mail: vk.kalichkin@gmail.com

(✉) **Лужных Т.А.,** младший научный сотрудник**; адрес для переписки:** Россия, 630501, Новосибирская область, р.п. Краснообск; СФНЦА РАН, а/я 463; e-mail: tanya.luzhnykh@mail.ru

Риксен В.С., младший научный сотрудник; e-mail: riclog@mail.ru

Васильева Н.В., кандидат биологических наук, старший научный сотрудник; e-mail:vasileva_nv@prometeus.sbras.ru

Шпак В.А., кандидат физико-математических наук, научный сотрудник; e-mail: shpakva54@gmail.com

AUTHOR INFORMATION

Vladimir K. Kalichkin, Doctor of Science in Agriculture, Professor, Head Researcher; e-mail: vk.kalichkin@gmail.com

(✉) **Tatyana A. Luzhnykh**, Junior Researcher; **address**: PO Box 463, SFSCA RAS, Krasnoobsk, Novosibirsk Region, 630501, Russia; e-mail: tanya.luzhnykh@mail.ru

Vera S. Riksen, Junior Researcher; e-mail: riclog@mail.ru

Nadezhda V. Vasilyeva, Candidate of Science in Biology, Senior Researcher; e-mail:vasileva_nv@prometeus.sbras.ru

Vladimir A. Shpak, Candidate of Science in Physics and Mathematics, Researcher, e-mail: shpakva54@gmail.com

Дата поступления статьи / Received by the editors 21.06.2021
Дата принятия к публикации / Accepted for publication 18.10.2021
Дата публикации / Published 25.11.2021