



## 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 переменной 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 No-

vosibirsk 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.

## ВВЕДЕНИЕ

Сельское хозяйство 4.0, как четвертый этап его эволюции, выдвигает требования к созданию и освоению современных информационных систем поддержки принятия решений с использованием комплекса цифровых технологий [1]. Предиктивная аналитика в этих системах должна занять ведущее положение, поскольку без прогнозирования трансформации условий, объектов и процессов, происходящих в сельском хозяйстве, трудно или почти невозможно принять правильное решение по его управлению [2, 3].

Как и во многих других сферах человеческой деятельности, объем информации по сельскому хозяйству постоянно увеличивается. Для адекватной оценки информации, поступающей из разнообразных источников, используют различные алгоритмы машинного обучения, позволяющие анализировать данные и помогать специалистам сельского хозяйства в решении частных задач, повышая при этом их эффективность. Применение машинного обучения в сельском хо-

зяйстве в настоящее время сопровождается массовым интересом со стороны мирового научного сообщества. В машинном обучении используются различные модели: регрессия, кластеризация, байесовские и нейронные сети, машины опорных векторов, деревья решений и др. Целесообразность применения тех или иных моделей машинного обучения определяется различными типами сельскохозяйственных данных и решаемыми задачами [4–6].

Одним из инструментов искусственного интеллекта и возможности применения в машинном обучении считаются Байесовские сети доверия (БСД). В материалах обзорной статьи [7] отмечено, что метод БСД подходит для исследований в сельском хозяйстве, так как БСД способны рассуждать с неполной информацией и включать новую информацию, а также решать задачи в условиях неопределенности с учетом причинно-следственных связей [8]. Примеры применения аппарата БСД в русскоязычных публикациях рассмотрены в медицине, экологии, риск-анализе, социологии и в других предметных

областях [9, 10], в приложении к сельскому хозяйству БСД имеют очень ограниченное применение [11].

Наиболее популярным статистическим методом, используемым для прогностического моделирования в сельском хозяйстве, является регрессионный анализ. Этот метод считается наиболее легким в использовании и понимании, позволяет исследовать взаимосвязь между зависимой (целевой) и независимой (предиктор) переменной, выявлять существенные закономерности в общем виде, определять тесноту связей исследуемых факторов [12].

Цель исследований – обучить различные модели анализа эмпирических данных, осуществить прогноз содержания нитратного азота в почве перед посевом, оценить точность прогнозных моделей.

## МАТЕРИАЛ И МЕТОДЫ

При обучении моделей использовали данные длительного многофакторного стационарного полевого опыта СибНИИЗиХ СФНЦА РАН. Эксперимент заложен в 1981 г. на территории ОС «Элитная» – филиала СФНЦА РАН Новосибирской области в центрально-лесостепной подзоне. Данные включают результаты исследований (2013–2018 гг.) четырехпольного зернопарового севооборота (пар – пшеница – пшеница – пшеница). Опыт проведен с разными вариантами основной обработки почвы:

- вспашка (под 1-ю и 3-ю культуры на глубину 20–22 см, под 2-ю – 25–27 см);
- безотвальная обработка (безотвальное рыхление стойками СибИМЭ под 1-ю и 3-ю культуры на глубину 20–22 см, под 2-ю – 25–27 см);
- минимальная обработка (плоскорезная обработка на глубину 10–12 см под все культуры ежегодно);
- нулевая обработка (без зяблевой обработки).

Почва стационара – чернозем выщелоченный среднесуглинистого гранулометрического состава. Моделирование проводили на данных по содержанию нитратного азота в слое почвы 0–40 см перед посевом.

Обучение БСД проведено в программе Netica версия «6.07», модель логистической регрессии осуществлена в пакете модулей SPSS версия «26». При моделировании 80% данных исходной выборки использовали для обучения моделей, 20% – в качестве тестирования (прогнозирования).

Размерность уровня содержания нитратного азота в почве устанавливали в соответствии со шкалой А.Е. Кочергина.

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

*Построение БСД.* БСД – это ориентированный ациклический граф, каждой вершине которого соответствует случайная переменная, дуги графа кодируют отношения условной независимости между этими переменными. Вершины могут представлять переменные любых типов, быть взвешенными параметрами, скрытыми переменными или гипотезами [13]. БСД являются вероятностными, потому что построены на основе распределения вероятностей и используют теорию вероятностей для прогнозирования. Для обучения и работы БСД используют какие-либо данные или знания экспертов (эвристики). В основе прогнозной модели лежит формула Байеса, которая позволяет определить вероятность наступления какого-либо события при условии, что уже произошло другое взаимозависимое с ним событие.

Построение БСД начинается с определения структуры графа, в которой обозначаются главные (основные) узлы и их параметры.

Исходя из структуры данных, экспертно допускаем, что количество нитратного азота в почве находится в зависимости от погодных условий и агротехнических приемов. Тогда основные узлы сети представляются качественными и количественными параметрами рабочего участка (подтип почвы, предшественник, обработка почвы, погодные условия) с соответствующими грациями (событиями):

- подтип почвы – дискретная переменная с одной грацией – Leached chernozem (чернозем выщелоченный);
- обработка почвы с четырьмя грациями – Plowing (вспашка), No-tillage (безот-

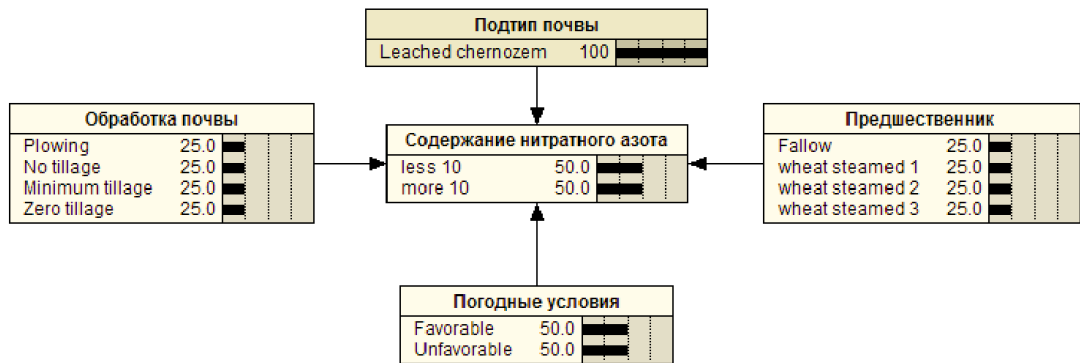
вальная), Minimum tillage (минимальная), Zero tillage (нулевая);

- предшественник с четырьмя градациями – Fallow (пар), wheat steamed 1 (1-я пшеница по пару), wheat steamed 2 (2-я пшеница по пару), wheat steamed 3 (3-я пшеница по пару);
- погодные условия с двумя градациями – Favorable (благоприятные), Unfavorable (неблагоприятные);
- содержание нитратного азота в слое почвы 0–40 см с двумя градациями – Less 10 (меньше 10 мг/кг почвы) и More 10 (больше 10 мг/кг почвы). Данный узел является целевым.

Следующий этап работы заключался в построении ациклического графа и расста-

новках причинно-следственных связей к целевому узлу, что является необходимым условием для дальнейшего заполнения таблицы условных вероятностей (ТУВ) с учетом анализа полученных данных. Построенный граф состоит из четырех основных узлов (подтип почвы, погодные условия, предшественник, обработка почвы) и одного целевого (содержание нитратного азота). На данной стадии система задает по всем узлам равномерное распределение вероятностей (см. рис. 1).

После составления структуры графа и определения причинно-следственных связей в виде стрелок от основных узлов к целевому согласно логике модели осуществляется построение и заполнение ТУВ (см. рис. 2).



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

Netica - [E Table (in Bayes net oneta\_6o08o2021)]

File Edit Table Window Help

Node: E

Chance % Probability

Apply OK Reset Close

| 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



В ТУВ программа автоматически выстраивает возможные комбинации случайных событий основных узлов, влияющих на два события целевого узла (содержание нитратного азота – Less 10, More 10). При этом эксперты каждой комбинации случайных событий основных узлов в процентном соотношении присваивают априорную вероятность наступления каждого из двух событий целевого узла, тем самым обучают модель. Чем выше процент, тем больше доля вероятности, что произойдет данное событие. При заполнении ТУВ эксперты руководствовались не только эвристиками (знаниями по данной проблеме), но использовали эмпирические данные полевого опыта. Как только ТУВ полностью построена и заполнена, проводится компиляция БСД, после чего сеть готова к использованию.

Перед тем как дать прогноз на 2021 г. по содержанию нитратного азота в слое почвы 0–40 см, проведена тренировка сети для определения ее поведения в процессе изменения событий основных узлов и получения апостериорной вероятности целевого узла. Для этого смоделировали ситуацию (первый сценарий), при которой условия на рабочем участке складываются следующим образом: предшественник – Fallow (пар), обработка почвы – Plowing (отвальная), погодные условия – Unfavorable (неблагоприятные) (см. рис. 3).

В первом сценарии БСД прогнозирует, что с долей вероятности 69% содержание

нитратного азота в слое почвы 0–40 см будет более 10 мг/кг.

Для проверки устойчивости работы сети проведен анализ развития событий по второму сценарию (изменены события в узлах): предшественник – wheat steamed 3 (3-я пшеница по пару), обработка почвы – Zero tillage (нулевая), погодные условия – Favorable (благоприятные) (см. рис. 4).

По второму сценарию БСД прогнозирует, что с вероятностью 95% содержание нитратного азота в слое почвы 0–40 см будет менее 10 мг/кг.

Для БСД также проведен анализ чувствительности. Функция анализа чувствительности используется для определения величины и степени воздействия основных узлов на события целевого узла в порядке убывания (см. табл. 1).

Взаимная информация между двумя узлами указывает на зависимость этих узлов друг от друга. Если она существует, то насколько близки их отношения. Самый большой показатель взаимной информации получен для узла предшественник. Это значит, что узел оказывает наибольшее влияние на целевой показатель (содержание нитратного азота в слое почвы 0–40 см). Погодные условия и обработка почвы оказывают значительно меньшее влияние на целевой показатель.

*Построение модели логистической регрессии.* Для прогноза целевого показателя (содержание нитратного азота в слое почвы

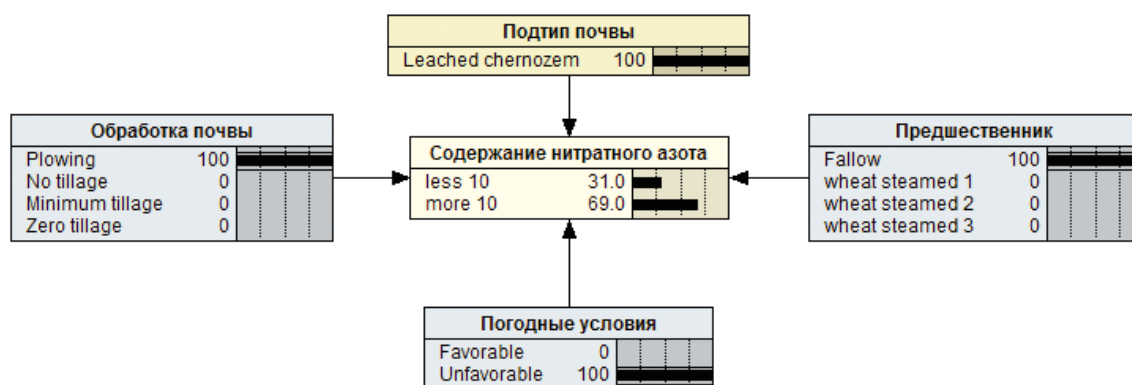
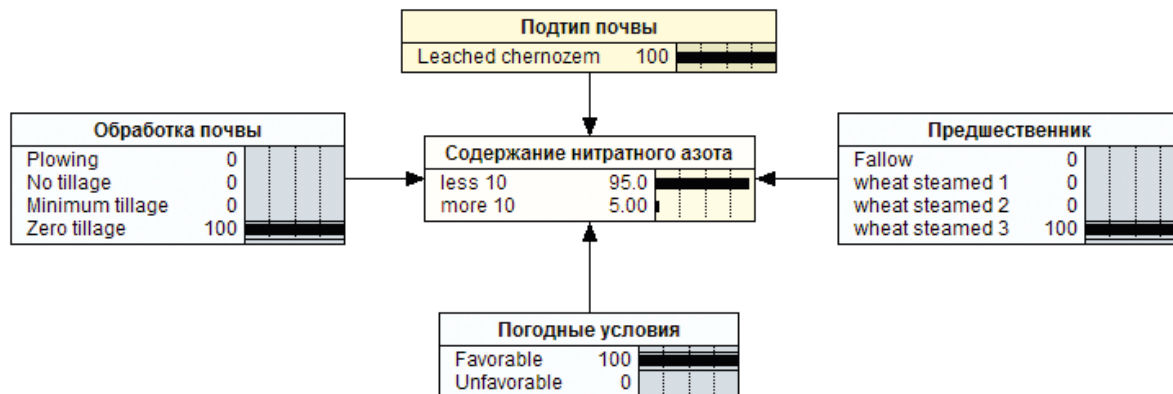


Рис. 3. Первый сценарий БСД

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

| Узел             | Взаимная информация | Процент | Дисперсия убеждений |
|------------------|---------------------|---------|---------------------|
| Предшественник   | 0,308               | 33,9    | 0,08366             |
| Погодные условия | 0,00174             | 0,204   | 0,00048             |
| Обработка почвы  | 0,00084             | 0,0987  | 0,0002322           |
| Подтип почвы     | 0                   | 0       | 0                   |

0–40 см) исследовали связь этого показателя (зависимая переменная) от следующих независимых переменных (факторов): предшественник, способ обработки почвы, погодные условия (осадки за апрель – май и сентябрь, среднемесячная температура воздуха за эти же месяцы).

Зависимую переменную можно классифицировать в качестве категориальной, когда значения содержания нитратного азота в слое почвы 0–40 см принимают вид – Less 10 и More 10. В этом случае требуемую зависимость можно получить с использованием модели логистической регрессии [14]. При этом одна из категорий зависимой переменной становится опорной (эталонной), а другая сравнивается с ней. Независимые переменные могут быть категориальными или количественными. Уравнение логистической регрессии прогнозирует вероятность принадлежности к каждой категории за-

висимой переменной по значениям независимых переменных. Окончательный выбор предсказываемой категории для зависимой переменной производится по правилу наибольшей вероятности принадлежности.

Для получения параметров (коэффициентов) логистической регрессии были использованы данные, сформированные в виде таблицы из 72 наблюдений (строк) и 9 факторов (столбцов), включая зависимую переменную (см. рис. 5).

Число наблюдений распределено по показателю содержания нитратного азота в слое почвы 0–40 см следующим образом: Less 10 – 53, More 10 – 19. В качестве опорной выбрана категория More 10.

В табл. 2 представлены коэффициенты, полученные с помощью логистической регрессии.

На основании расчетов, проведенных с помощью логистической регрессии, можно сделать заключение о том, что наиболее сильное влияние на результирующую переменную оказывает пар в качестве предшественника. Если значимость коэффициента  $P < 0,05$ , то связь статистически значима. Результат  $P > 0,05$  свидетельствует о том, что связь между переменными слабая или не обнаружена. Третья пшеница по пару и нулевая обработка не влияют на содержание нитратного азота в слое почвы 0–40 см.

Мерой оценки адекватности (качества) построенной логистической регрессионной модели является коэффициент псевдо  $R$ -квадрат. В данном случае полученная дис-

|    | Год  | Предшественник       | Содержание NNO3 мг/кг почвы в слое 0-40 см | Обработка почвы | 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

| Фактор                                           | Коэффициент переменных | Стандартная ошибка | Значимость коэффициента (P) |
|--------------------------------------------------|------------------------|--------------------|-----------------------------|
| Среднемесячная температура (сентябрь)            | −1,7                   | 0,6                | 0,007                       |
| Среднемесячное количество осадков (сентябрь)     | − 0,16                 | 0,068              | 0,023                       |
| Среднемесячная температура (апрель – май)        | −2,08                  | 0,84               | 0,013                       |
| Среднемесячное количество осадков (апрель – май) | −0,22                  | 0,076              | 0,003                       |
| Предшественник – пар                             | 5,1                    | 1,7                | 0,002                       |
| Предшественник – 1-я пшеница по пару             | −0,613                 | 0,8                | 0,48                        |
| Предшественник – 2-я пшеница по пару             | −1,5                   | 1,03               | 0,167                       |
| Предшественник – 3-я пшеница по пару             | 0                      | 0                  | 0                           |
| Обработка почвы – вспашка                        | 1,03                   | 0,81               | 0,27                        |
| Обработка почвы – безотвальная                   | −0,58                  | 0,87               | 0,607                       |
| Обработка почвы – нулевая                        | 0                      | 0                  | 0                           |
| Константа                                        | 55,1                   | 19,06              | 0,004                       |

персия составляет по мере Нэйджелкерка (как правило, наиболее применяемой) 70%, что свидетельствует об удовлетворительных прогностических возможностях модели.

В отличие от обычной регрессии, в модели логистической регрессии не производится предсказание значения числовой переменной на основании выборки исходных значений. Вместо этого значением функции является вероятность того, что данное исходное значение принадлежит к определенной категории.

Таким образом, в ходе исследований построены и обучены модели БСД и логистической регрессии, прогнозирующие содержание нитратного азота в слое почвы 0–40 см чернозема выщелоченного перед посевом. Прогноз на 2021 г. осуществлен для вариантов опыта, представленных в табл. 3. Погодные условия данного года были неблагоприятные.

Прогнозом, осуществленным с помощью БСД, накопление нитратного азота в слое почвы 0–40 см по паровому предшествен-

нику во всех вариантах показано больше 10 мг/кг (по вспашке с долей вероятности 69%, по безотвальной обработке – 55, по нулевой обработке – 65%). По всем остальным вариантам прогнозируется содержание нитратного азота меньше 10 мг/кг. В отличие от БСД, прогнозом логистической регрессии показано содержание нитратного азота в слое почвы 0–40 см во всех вариантах не более 10 мг/кг.

В качестве критерия оценки прогнозных моделей определена величина отклонения (погрешность) фактического содержания азота от прогнозируемого. В табл. 4, 5 представлены сравнительные прогностические способности моделей, проверенные на исходной выборке. Исходная выборка состоит из 72 наблюдений, из них 53 наблюдения – с содержанием нитратного азота Less 10 и 19 наблюдений – с содержанием нитратного азота More 10.

В БСД правильные прогнозы для категории Less 10 достигают 90%, при этом 48 на-

блюдений спрогнозировано правильно и 5 наблюдений – ошибочно (см. табл. 4).

Для категории More 10 в БСД доля правильного прогнозирования находится на уровне 68%. При этом из 19 наблюдений правильно предсказанными моделью были 13 и ошибочно – 6. По всем наблюдениям общая доля правильных прогнозов составляет 84%. Достоверность прогнозных показателей отмечена ниже, чем в логистической регрессии, но эффективность ее прогностической способности достаточно высока.

При сравнении фактических наблюдений по содержанию нитратного азота в слое почвы 0–40 см с прогнозными оценками с помощью логистической регрессии можно отметить, что полученная доля правильных прогнозов составляет 94,3% для категории Less 10 (см. табл. 5). Данная модель правильно спрогнозировала 50 наблюдений, три наблюдения были спрогнозированы ошибочно, попав в категорию More 10.

**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.

| Предшественник      | Обработка почвы | Содержание N-NO <sub>3</sub><br>(прогноз по БСД), мг/кг | Содержание N-NO <sub>3</sub><br>(прогноз логистической<br>регрессии), мг/кг |
|---------------------|-----------------|---------------------------------------------------------|-----------------------------------------------------------------------------|
| Пар                 | Вспашка         | More 10                                                 | Less 10                                                                     |
| Пар                 | Безотвальная    | More 10                                                 | Less 10                                                                     |
| Пар                 | Нулевая         | More 10                                                 | Less 10                                                                     |
| 1-я пшеница по пару | Вспашка         | Less 10                                                 | Less 10                                                                     |
| 1-я пшеница по пару | Безотвальная    | Less 10                                                 | Less 10                                                                     |
| 1-я пшеница по пару | Нулевая         | Less 10                                                 | Less 10                                                                     |
| 2-я пшеница по пару | Вспашка         | Less 10                                                 | Less 10                                                                     |
| 2-я пшеница по пару | Безотвальная    | Less 10                                                 | Less 10                                                                     |
| 2-я пшеница по пару | Нулевая         | Less 10                                                 | Less 10                                                                     |
| 3-я пшеница по пару | Вспашка         | Less 10                                                 | Less 10                                                                     |
| 3-я пшеница по пару | Безотвальная    | Less 10                                                 | Less 10                                                                     |
| 3-я пшеница по пару | Нулевая         | Less 10                                                 | Less 10                                                                     |

**Табл. 4.** Классификационная таблица БСД

**Table 4.** DAG classification Table

| Количество наблюдений | Предсказанные |         | Процент правильных прогнозов |
|-----------------------|---------------|---------|------------------------------|
|                       | Less 10       | More 10 |                              |
| Less 10 (53)          | 48            | 5       | 90                           |
| More 10 (19)          | 6             | 13      | 68                           |
| Общая доля            |               |         | 84                           |

**Табл. 5.** Классификационная таблица модели логистической регрессии

**Table 5.** Classification table of the logistic regression model

| Количество наблюдений | Предсказанные |         | Процент правильных прогнозов |
|-----------------------|---------------|---------|------------------------------|
|                       | Less 10       | More 10 |                              |
| Less 10 (53)          | 50            | 3       | 94,3                         |
| More 10 (19)          | 6             | 13      | 68,4                         |
| Общая доля            |               |         | 87                           |



Для выборки наблюдений с содержанием нитратного азота More 10 доля правильных прогнозов составила 68,4%, из них верно спрогнозировано 13 наблюдений, неправильными были 6 прогнозов, попав в категорию Less 10. Общая доля правильных прогнозов по всем категориям составила 87%. Таким образом, учитывая небольшой объем статистической выборки и малое количество предикторов, прогностические свойства проверенных моделей можно признать удовлетворительными.

## ЗАКЛЮЧЕНИЕ

С помощью методов машинного обучения построены и обучены модели, позволяющие с допустимой достоверностью прогнозировать содержание нитратного азота перед посевом в слое почвы 0–40 см. Разница методологий заключается в том, что БСД при прогностическом моделировании позволяет проводить анализ по сценарию «что, если», может совмещать закономерности, выведенные из статистических данных и знания экспертов, полученных фактическим путем, а также способна обрабатывать неполные и разных типов данные.

В дальнейшем планируется улучшать качество работы моделей, добавляя другие предикторы, влияющие на результативный признак, осуществлять поиск методов машинного обучения, допускающих анализ данных небольшого объема, и провести верификацию моделей на фактических данных. Рассматривается возможность использования построенных моделей для разработки экспертной системы по выбору и сопровождению агротехнологий, которая позволит грамотно принимать решения в отношении поставленных задач сельскохозяйственного производства.

## 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.
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., Pylaniadis C., Athanasiadis I. Machine learning for large-scale crop yield forecasting. *Agricultural Systems*, 2021, vol. 187, p. 103016. DOI: 10.1016/j.agry.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.knsys.2016.09.12.
9. Aivazyan S.A. Bayesian approach in econometric analysis. *Applied econometrics. Prikladnaya ekonomika = 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 danih = 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