

## ПРОГНОЗИРОВАНИЕ ЗАСОРЕННОСТИ ПОСЕВОВ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

**В.К. Каличкин<sup>1</sup>**, доктор сельскохозяйственных наук,

**О.К. Альсова<sup>2</sup>**, кандидат технических наук,

**К.Ю. Максимович<sup>1</sup>**, аспирант,

**Н.В. Васильева<sup>1</sup>**, кандидат биологических наук

<sup>1</sup>Сибирский федеральный научный центр агробиотехнологий РАН, 630501, Новосибирская область, пос. Краснообск  
E-mail: vk.kalichkin@mail.com

<sup>2</sup>Новосибирский государственный технический университет,  
630073, Новосибирск, просп. Карла Маркса, 20

*Представлены результаты исследования по оценке влияния качественных факторов и метеорологических параметров на засоренность посевов. Выполнено построение прогнозной модели засоренности посевов на основе дерева решений в условиях ограниченной выборки. В ходе исследования использованы материалы длительных полевых опытов, проведенных в лесостепи Новосибирской области, и сведения о метеорологических параметрах Новосибирского поста метеонаблюдений за 1996–2018 гг. Для решения задач исследования применяли разные классы методов интеллектуального анализа данных, а именно непараметрические вероятностно-статистические методы, методы визуализации и метод деревьев решений. Выбор методов исследования обусловлен особенностями структуры и характера распределения исходных данных: несоответствие модели закону нормального распределения; сравнительно небольшой объем выборки; наличие как качественных, так и количественных предикторов; сложные нелинейные корреляционные связи между степенью засоренности посевов и метеорологическими параметрами. Выявлены качественные факторы, определяющие степень засоренности посевов: фон химизации, размещение культуры после пара, система обработки почвы и метеорологические параметры (среднедекадные температуры воздуха и осадки за период с 3-й декады апреля по конец мая). Выполнена оценка их вклада и статистической значимости. С использованием алгоритма CART построения дерева решений разработана модель прогноза засоренности посевов и сформулированы логические правила развития сорняков в зависимости от управляющих воздействий и агрометеорологических условий. Точность модели характеризуется следующими показателями: MAE (средняя абсолютная ошибка) = 3,75; RMSE (среднеквадратичная ошибка) = 5,70 и R<sup>2</sup> (коэффициент детерминации) = 0,80. Построенный набор логических правил описывает структуру причинно-следственных связей и может быть использован в дальнейшем при создании экспертной системы поддержки принятия решений в растениеводстве.*

## APPLICATION OF MACHINE LEARNING TO FORECAST THE CONTAMINATION OF CROPS

Kalichkin V.K. <sup>1</sup>, Alsova O.K. <sup>2</sup>, Maksimovich K.Yu. <sup>1</sup>, Vasilyeva N. V. <sup>1</sup>

<sup>1</sup>Siberian Federal Research Center of Agricultural Biotechnology, Russian Academy of Sciences,  
630501, Novosibirskaya obl., pos. Krasnoobsk  
E-mail: vk.kalichkin@gmail.com

<sup>2</sup>Novosibirsk State Technical University,  
630073, Novosibirsk, prosp. Karla Marksa, 20

*The paper provides the results on evaluating the influence of qualitative factors and meteorological parameters on the crop infestation index. The prediction model of crops infestation has been built on the application of the decision tree method given limited samples of data. In research work using the materials of long-term field experiments of Siberian Research Institute of Farming and Chemistry of Agriculture – structural subdivision of Siberian Branch of the Russian Academy of Sciences, carried out in the forest-steppe of Novosibirsk region and data on meteorological parameters of Novosibirsk weather observation station for 1996-2018. In order to meet the objectives of the study, different classes of data mining methods were used, such as nonparametric probabilistic-statistical methods, visualization methods, and decision tree methods. Selection of research methods was determined by the peculiarities of the structure and nature of the initial data distribution: nonconformity of the law of normal distribution model; sufficiently small sample size; multiple qualitative and quantitative predictors; complex nonlinear correlations between the crop infestation index and meteorological parameters. Qualitative factors determining the crop infestation index (use of chemicalization, crop placement after fallow, tillage system) and meteorological parameters (average ten-day air temperatures and precipitation during the period from the 3rd ten-day period of April to the end of May) have been identified. Their contribution and statistical significance were evaluated. Using the CART algorithm for building a decision tree, developed the model for crop infestation forecasting and formulated the logical rules of weed development depending on the controlling influences and agro-meteorological conditions. The accuracy of the model is characterized by the following indicators: MAE (Mean Absolute Error) = 3.75; RMSE (Root Mean Square Error) = 5.7 and R<sup>2</sup>(Coefficient of Determination/R Squared)= 0.8. The set of logical rules describes the structure of causal relationships in the data and can be used in the future to create a decision support system in crop production.*

**Ключевые слова:** прогнозирование, непараметрические вероятностно-статистические методы, деревья решений, засоренность посевов, язык R.

**Key words:** forecasting, nonparametric probability-statistical methods, decision trees, crop infestation, R.

С распространением инновационных технологических решений в сельском хозяйстве возникает потребность в разработке прогностических инструментов, созданных совместно разработчиками программного обеспечения и исследователями из прикладных областей:

агрономии, защиты растений, агрохимии, экологии, которые могут быть использованы для оценки возможных рисков, обоснования экономической эффективности используемых агроприемов и принятия решений в области сельского хозяйства [1, 2, 3].

Способность предвидеть развитие вредных организмов, оценить возможные риски потери урожайности и оперативно принять взвешенное решение обосновывает актуальность поиска и применения эффективных методов прогнозирования в сельском хозяйстве [4, 5, 6]. Создание предиктивных моделей агросистем осложняется наличием нелинейных взаимосвязей между входными факторами и результирующими показателями, большим количеством предикторов при сравнительно малом количестве наблюдений, практической сложностью проведения многофакторных экспериментов [7, 8, 9]. Оптимизационные и биофизические модели имеют ограниченное применение, поскольку требуют большого объема входных данных [10, 11, 12]. Необходим поиск качественно новых подходов к оценке рисков и управлению посевами, которые основаны на современных разработках в области информационных технологий [13].

Одним из таких подходов может стать метод дерева решений – графическое представление логических правил развития процесса в зависимости от складывающихся условий (сочетание действующих факторов), демонстрирующий альтернативные варианты развития событий [14, 15, 16].

Визуально представленная логическая модель, позволяет объяснять варианты развития событий и исследовать вероятные последствия выбора их сочетаний [17]. Все это открывает возможности для анализа рисков в сочетании с каждым вариантом складывающихся условий. Формализация предметной области на логическом уровне демонстрирует высокие показатели достоверности при ограниченном количестве влияющих факторов и без «погружения» в биофизические свойства исследуемого объекта [13]. Дерево решений – это инновационный инструмент в сельском хозяйстве, который уже применяют по таким направлениям, как моделирование эрозийных процессов [18], оценка засоленности почвы [19], определение болезней [20], прогнозирование урожайности сельскохозяйственных культур [21, 22, 23].

Сельскохозяйственной наукой и практикой накоплено большое количество информации, которая содержит в себе сформированные и устоявшиеся знания, но их практическая реализация на основе инженерии знаний – актуальная проблема методического характера. Разработка моделей, приближенных к логическим рассуждениям человека, и их использование в построении систем поддержки принятия решений – одно из современных направлений цифровизации сельского хозяйства. Становится возможным использование метода деревьев решений для частных задач в области растениеводства [24]. Так, прогнозирование засоренности посевов потенциально позволит помочь в принятии обоснованных решений по выбору агротехнологий и эффективному применению средств защиты растений, что выступает одним из ключевых вопросов управления агрофитоценозом.

Цель исследования – построение прогнозной модели засоренности посевов на основе применения метода дерева решений в условиях ограниченной выборки.

**Методика.** В работе использованы данные по засоренности посевов яровой пшеницы длительных полевых опытов СибНИИЗиХ – структурного подразделения СФНЦА РАН, расположенных в Новосибирской области, р.п. Краснообск, за 1996–2018 гг. В качестве факторов, влияющих на формирование уровня засоренности посевов, рассматривали фон применения средств химизации – экстенсивный (Э), интенсивный (И); систему обработки почвы – ежегодная вспашка на глубину 20...22 см (В), безотвальная глубокая обработка на глубину 20...22 см (БГ), минимальная плоскорезная обработка на глубину 10...12 см (М), без основной обработки почвы (БО); размещение

после пара – первая культура (1), вторая (2), третья (3). На интенсивном фоне азотные удобрения вносили ежегодно в дозе  $N_{90}$  перед посевом, фосфорные – в дозе  $P_{120}$  один раз за ротацию севооборота в паровом поле. По вегетации применяли баковую смесь гербицидов против однодольных и двудольных сорных растений (в фазе кушения), фунгициды против болезней (в фазе флаг-листа) и инсектициды по ситуации. На экстенсивном фоне удобрения и пестициды не использовали.

Исследования проводили в севообороте пар – озимая рожь – яровая пшеница – яровая пшеница. После 2007 г. озимую рожь заменили яровой пшеницей. Сорная растительность на опытном поле была представлена 30 видами, относящимися преимущественно к типичным для лесостепной зоны. Большую часть сорного фитоценоза составляли мягликовые (злаковые) виды – просо куриное и посевное (*Echinochloa crusgalli* (L.), *Panicum miliaceum* L.), щетинники сизый и зеленый (*Setaria viridis* L., *S.pumila* (Po: ret) Schultes). В разные годы доля мягликовых от общей фитомассы сорных растений достигала 73...84 %. Двудольные сорные растения представлены большей частью однолетними ранними и поздними яровыми. Наиболее распространены были щирица запрокинутая (*Amaranthus retroflexus* L.), пикульник двунадрезанный (жабрей) (*Galeopsis bifida* Boenp.), паслен черный (*Solanum nigrum* L.), липучка щетинистая (*Lappula squarrosa* (Retz.) Dum.), подмаренник цепкий (*Galium aparine* L.) и различные виды семейства гречишные (*Fagopirum tataricum* (L.), *Persicaria lapathifolia* (L.) S.F.Gray, *P. hydropiper* (L.), *Fallopia convolvulus* (L.). Многолетние виды выюнок полевой (*Convolvulus arvensis* L.), бодяк щетинистый (*Cirsium setosum* (Willd.) Bess.) и осот полевой (*Sonchus arvensis* L.) встречались в основном на экстенсивном фоне и на их долю приходилось не более 0,3 % сорного фитоценоза.

В исследовании использовали данные по метеорологическим показателям Новосибирского поста метеонаблюдений, расположенного вблизи г. Новосибирска, а именно, среднесуточные температуры воздуха за декаду (°C), сумма осадков (в мм) за 3-ю декаду апреля и 1...3 декады мая, а также суммы среднесуточных температур воздуха и осадков за эти же периоды (<http://www.pogodaiklimat.ru>).

В ходе работы применяли вероятностно-статистические методы и критерии для исследования законов распределения данных (критерий согласия Шапиро-Уилка [14]), выявления факторов, влияющих на показатель засоренности посевов и оценки их вклада (непараметрический критерий Краскела-Уоллиса [15], ранговый коэффициент корреляции Спирмена). Для прогнозирования степени засоренности посевов была построена модель дерева решений с использованием алгоритма CART [16]. Оценку точности и оптимизацию модели выполняли с использованием 10-кратной 10-блочной кросс-проверки. Для графического представления структуры исходных данных и результатов анализа применяли методы визуализации (построение диаграмм размаха, гистограмм, деревьев решений).

Всего в выборке было представлено 528 значений засоренности посевов (в % от фитомассы надземной части посевов) яровой пшеницы и обуславливающих ее факторов. Для качественной характеристики засоренности посевов вводили показатель уровня относительной засоренности (полученный в эксперименте), принимающий значение одной из 4 категорий соотношения с экономическим порогом вредности (ЭПВ = 10 % от фитомассы надземной части посевов): 1 – существенно ниже ЭПВ (менее 10 %), обработка посевов гербицидами не требуется; 2 – в пределах ЭПВ (10 % ± 1 %), обработка гербицидами по ситуации (учитывая погодные условия); 3 – существенно выше ЭПВ (более 10 %), гербицидная обработка обязательна, 4 – эк-

Табл. 1. Статистические характеристики показателя засоренности посевов

Фактор	Значение фактора	Статистические характеристики				
		число наблюдений	среднее	среднеквадратическое отклонение	минимальное значение	максимальное значение
Фон применения средств химизации	И	264	2,0	2,5	0,0	17,9
	Э	264	16,8	14,3	0,1	67,0
Культура после пара	1	176	4,0	5,5	0,0	31,2
	2	176	8,6	10,8	0,0	50,9
	3	176	15,6	16,4	0,1	67,0
Система обработки почвы	БГ	132	8,6	12,0	0,0	58,1
	БО	132	12,1	14,8	0,1	67,0
	В	132	6,8	9,9	0,0	42,7
	М	132	10,1	13,0	0,0	54,4

тремальный уровень превышения ЭПВ, при котором возможны значительные потери урожая (более 25 %).

Все расчеты выполняли средствами языка R в среде статистического анализа данных R-Studio.

**Результаты и обсуждение.** Для изучения структуры данных и обоснованного выбора методов исследования был проведен анализ характера распределения показателя засоренности посевов по всей выборке и по подвыборкам, сформированным при конкретном значении качественного фактора (система обработки почвы, фон химизации, культура после пара).

Распределение показателя засоренности посевов значительно отличалось для экстенсивного и интенсивного фонов (табл. 1). На экстенсивном фоне соответствующие статистические характеристики примерно в 8 раз выше, чем на интенсивном.

Выполнена проверка гипотезы о соответствии показателя засоренности модели закона нормального распределения (ЗНР). Визуально распределения не согласуются с моделью ЗНР, имеют ярко выраженную асимметрию (рис. 1). Гипотеза о ЗНР показателя засоренности посевов отвергается при использовании критерия согласия Шапиро-Уилка ( $p$ -значение  $< 0,05$ ). Аналогичные результаты были получены по всем сформированным подвыборкам исследуемого показателя.

Выявлены значительные отличия в статистических характеристиках засоренности посевов, вычисленных при разных значениях качественных факторов (фон химизации, культура после пара). Разница в статистических характеристиках засоренности посевов в зависимости от

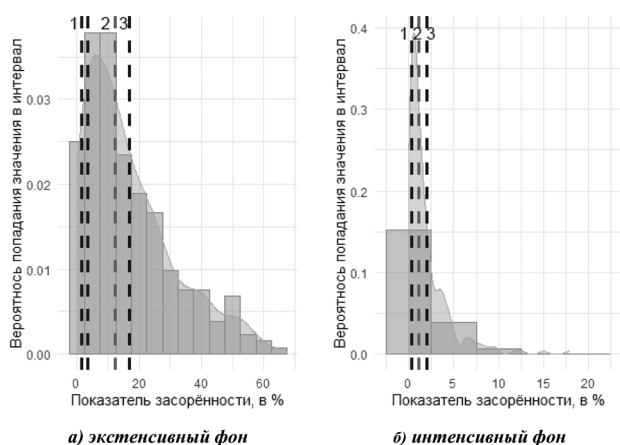


Рис. 1. Гистограмма показателя засоренности посевов с наложением аппроксимирующей функции плотности распределения вероятностей: 1 – мода, 2 – медиана, 3 – среднее.

года проведения исследований и используемой системы обработки почвы сравнительно меньше, чем от фона химизации и удаленности культуры от пара (рис. 2). Для подтверждения предварительных выводов о наличии и степени влияния качественных факторов на формирование уровня засоренности посевов был использован непараметрический критерий Краскела-Уоллиса. Выбор критерия обусловлен характером распределения показателя засоренности посевов (отклонена гипотеза о нормальном распределении) и сравнительно небольшим объемом исходных данных. В результате применения критерия Краскела-Уоллиса выявлена статистически значимая связь между засоренностью посевов и фоном химизации ( $p < 0,00001$ ), размещением

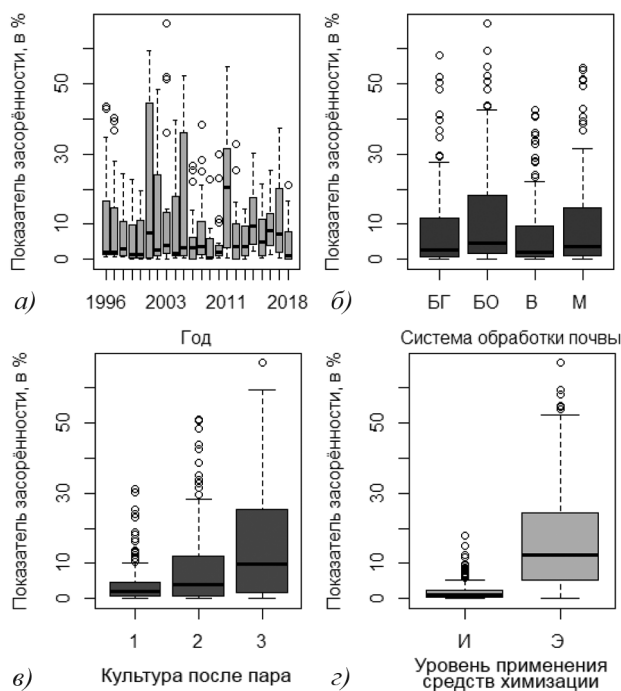


Рис. 2. Диаграммы размаха показателя засоренности посевов в зависимости от значений качественного фактора: а) год; б) система обработки почвы; в) культура после пара; г) фон химизации (положение центральной линии определяет медиану, границы прямоугольника соответствуют нижнему и верхнему квартилям, высота прямоугольника – значение интерквартильного размаха (ИКР), положение верхнего (нижнего) «уса» диаграммы определяется как сумма верхнего (нижнего) квартиля и величины  $1,5 \times \text{ИКР}$ , либо соответствует максимальному (минимальному) значению. Значения за пределами «усов» могут быть аномальными выбросами).

культуры после пара ( $p < 0,00001$ ), системой обработки почвы ( $p < 0,0005$ ), а также особенностями года исследований ( $p < 0,0001$ ).

Наибольший уровень засоренности посевов выявлен на 3 культуре после пара и без обработки почвы (см. табл. 1). Связь засоренности с размещением культуры в севообороте была прямой – чем дальше культура расположена от пара, тем выше засоренность. Это объясняется накоплением семян однолетних сорных растений и многолетних корнеотпрысковых сорняков в пахотном слое почвы по мере удаления от парового поля. Наиболее явно зависимость наблюдается на экстенсивном фоне без применения гербицидов. На интенсивном фоне тенденция также сохраняется, хотя проявляется в меньшей степени. На заключительной культуре севооборота отмечена максимальная засоренность посевов по всем вариантам обработки почвы.

Отмечено значительное влияние системы обработки почвы на засоренность посевов. Безотвальные обработки (БГ и М) способствовали накоплению на поверхности почвы семян сорных растений, которые весной в массе прорастали в последующей культуре. В варианте со вспашкой засоренность была минимальной, так как оборот пласта способствует попаданию семян сорняков в нижние слои пахотного горизонта, где большая часть их погибает. Кроме того, при вспашке уничтожается значительная часть многолетних сорных растений. Отсутствие обработки почвы (вариант БО) способствует увеличению засоренности.

Связь показателя засоренности посевов и года исследований подтверждает гипотезу о его зависимости от метеорологических условий. Для оценки связи между уровнем засоренности посевов и метеорологическими показателями использовали ранговый коэффициент корреляции Спирмена, который не требует выполнения условия нормального распределения исходных данных и может быть использован для получения корректных выводов о наличии и силе связи между исследуемыми количественными показателями.

Выявлено наличие слабой корреляционной линейной связи между уровнем засоренности посевов и метеорологическими показателями (табл. 2): статистически значимые положительные связи на экстенсивном фоне химизации выявлены только с температурой воздуха в 3-й декаде мая и суммарной температурой воздуха за период с 3-й декады апреля по конец мая, отрицательная статистически значимая связь – с осадками за 3-ю декаду апреля; на интенсивном фоне химизации отмечена статистически значимая связь только с осадками в 3-й декаде апреля. Такой результат может свидетельствовать как об отсутствии сильных корреляций

между исследуемыми факторами, так и о более сложном нелинейном характере их связи, для описания которого необходимо применение методов других классов.

Исходя из полученных результатов анализа характера распределения и взаимосвязей данных, для построения прогнозной модели был выбран метод дерева решений. При его выборе учитывали небольшой объем выборки, наличие как качественных, так и количественных предикторов, значительные отклонения распределения показателя засоренности посевов от закона нормального распределения. Кроме того, дерево решений часто позволяет построить сравнительно точную прогнозную модель при отсутствии сильных линейных корреляций между целевым показателем (степень засоренности посевов) и предикторами, как в нашем случае.

Дерево решений графически представляет собой логическую модель взаимосвязи степени засоренности культуры, качественных факторов и метеорологических показателей (рис. 3). Каждый узел дерева показывает прогнозируемое значение степени засоренности посевов (в %), число наблюдений в узле и долю наблюдений от общего количества (в %).

Система обработки почвы не была выделена в качестве значащего фактора в прогнозной модели, что объясняется высоким комплексным вкладом факторов «фон интенсификации», «культура после пара» и выделенных метеорологических показателей в формирование засоренности посевов. В результате включение в модель данных по засоренности в зависимости от системы обработки почвы не приводит к статистически значимому улучшению показателей ее точности.

Минимальное количество сорного компонента наблюдается на интенсивном фоне (лист дерева 2). Наименьший уровень засоренности посевов формируется по 1-й культуре после пара и равен 6,6 % на экстенсивном фоне (лист дерева 12). По 2-й культуре после пара доля сорняков в общей биомассе посевов увеличивается с 10 % до 35 % (листья 21, 54, 55). Регулирующим фактором в варианте размещения пшеницы 2-й культурой после пара на экстенсивном фоне выступают осадки 1-й декады мая, которые способствуют значительному увеличению биомассы сорняков (лист 55). Наибольшая доля сорной растительности (45%) достигается при среднесуточной температуре воздуха за вторую декаду мая выше 9,5 °С и сумме осадков выше 12 мм (лист дерева 61). Фактором, значительно лимитирующим рост и развитие сорняков на 3-й культуре после пара на экстенсивном фоне, была среднесуточная температура воздуха за 2-ю декаду мая ниже 9,5 °С.

Табл. 2. Результаты расчета коэффициента корреляции Спирмена

Показатель	Экстенсивный фон		Интенсивный фон	
	значение коэффициента	p-значение	значение коэффициента	p-значение
Среднесуточная температура воздуха за:				
3 декаду апреля	-0,02	>0,05	0,06	>0,05
1 декаду мая	0,10	>0,05	0,03	>0,05
2 декаду мая	0,09	>0,05	-0,05	>0,05
3 декаду мая	0,13	<0,05	0,04	>0,05
Сумма осадков за:				
3 декаду апреля	-0,18	<0,05	-0,16	<0,05
1 декаду мая	-0,06	>0,05	-0,11	>0,05
2 декаду мая	0,06	>0,05	-0,08	>0,05
3 декаду мая	-0,04	>0,05	0,01	>0,05
Сумма среднесуточных температур воздуха за период (с 3-ей декады апреля по конец мая)	0,15	<0,05	-0,02	>0,05
Сумма осадков за период (с 3-ей декады апреля по конец мая)	-0,06	>0,05	0,05	>0,05

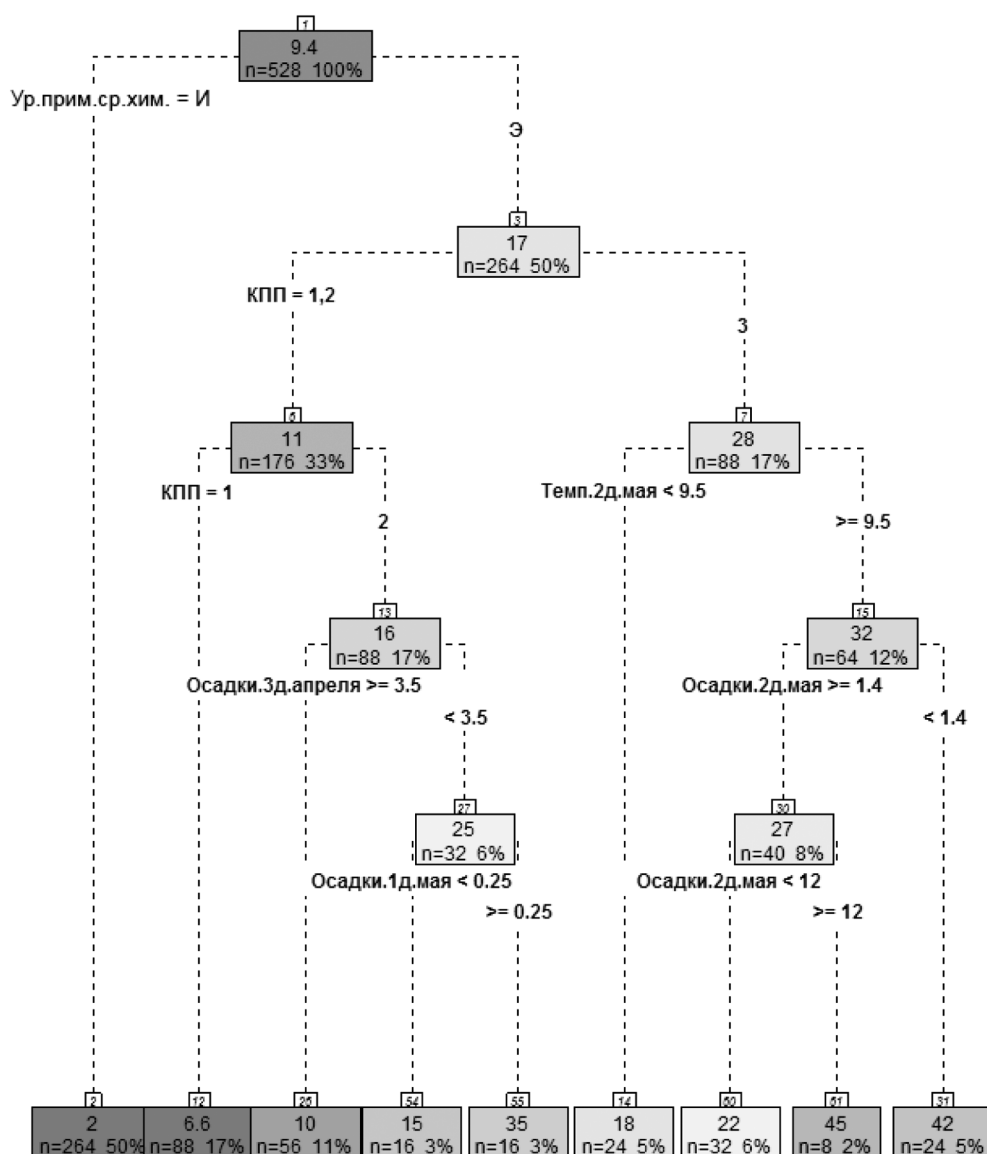


Рис. 3. Дерево решений для прогнозирования показателя засоренности посевов.

По полученным результатам можно заключить, что количество атмосферных осадков существенно влияет на засоренность посевов яровой пшеницы. При этом максимальное воздействие оказывают осадки и температура в период посева и всходов (май), когда культурные и сорные растения только формируются. Высокая температура воздуха способствует более быстрому росту яровой пшеницы, что дает возможность ей занять площадь питания и подавить прорастающие сорные растения. В период от фазы кушения до уборки (июнь–август) зависимость засоренности от атмосферных осадков значительно слабее, так как сформировавшаяся масса растений пшеницы препятствует росту сорняков и появлению их новых всходов.

Показатели точности модели дерева решений составили: MAE (средняя абсолютная ошибка) = 3,75; RMSE (среднеквадратичная ошибка) = 5,70 и R<sup>2</sup> (коэффициент детерминации) = 0,80.

На основе построенного дерева решений можно сформулировать наиболее значимые логические правила, которые дают возможность прогнозировать степень засоренности посевов яровой пшеницы.

если фон химизации экстенсивный и первая культура после пара, то показатель засоренности посевов равен 6,6%;

если фон химизации экстенсивный, 2 культура после пара и сумма осадков 3 декады апреля выше 3,5 мм, то показатель засоренности посевов равен 10%;

если фон химизации экстенсивный, 3 культура после пара и среднесуточная температура воздуха 2 декады мая ниже 9,5 °С, то показатель засоренности посевов равен 18%;

если фон химизации экстенсивный, 3 культура после пара, среднесуточная температура воздуха 2 декады мая выше 9,5 °С и сумма осадков 2 декады мая выше 12 мм, то показатель засоренности посевов равен 45%.

Таким образом, по результатам изучения разных классов методов интеллектуального анализа данных, как непараметрических вероятностно-статистических, так и построения логической модели (набор правил «если – то»), установлено, что для прогнозирования изменения показателя засоренности посевов лучше всего подойдет метод дерева решений. Такой выбор обусловлен особенностями структуры и характера распределения исходных данных: несоответствие

модели закону нормального распределения; сравнительно небольшой объем выборки; наличие как качественных, так и количественных предикторов; сложные нелинейные корреляционные связи между степенью засоренности посевов и метеорологическими показателями.

К основным факторам, определяющим засоренность посевов яровой пшеницы, относятся: фон интенсификации, удаленность культуры от пара, а также среднедекадные температуры воздуха и осадки за период с 3-ей декады апреля по конец мая.

#### Литература.

1. *Agriculture in 2050: recalibrating targets for sustainable intensification* / M. C. Hunter, R. G. Smith, M. E. Schipanski, et al. // *Bioscience*. 2017. Vol. 67. No. 4. P. 386–391. doi: 10.1093/biosci/bix010.
2. *Climate smart agriculture, farm household typologies and food security: An ex-ante assessment from Eastern India* / S. Lopez-Ridaura, R. Frelat, M. T. van Wijk, et al. // *Agricultural systems*. 2018. Vol. 159. P. 57–68. doi: 10.1016/j.agsy.2017.09.007
3. *Decision support systems for agriculture 4.0: Survey and challenges* / Z. Zhai, J. F. Martinez, V. Beltran, et al. // *Computers and Electronics in Agriculture*. 2020. Vol. 170. P. 105256. URL: <https://www.sciencedirect.com/science/article/pii/S0168169919316497?via%3Dihub> (дата обращения: 12.10.2021), doi: 10.1016/j.compag.2020.105256.
4. Köksal Ö., Tekinerdogan B. *Architecture design approach for IoT-based farm management information systems* // *Precision Agriculture*. 2019. Vol. 20. No. 5. P. 926–958. doi: 10.1007/s11119-018-09624-8.
5. *IoT and agriculture data analysis for smart farm* / J. Muangprathub, N. Boonnarn, S. Kajornkasirat, et al. // *Computers and electronics in agriculture*. 2019. Vol. 156. P. 467–474. doi: 10.1016/j.compag.2018.12.011.
6. *A systems approach to forecast agricultural land transformation and soil environmental risk from economic, policy, and cultural scenarios in the north central United States (2012–2062)* / B. L. Turner, M. Wuellner, T. Nichols, et al. // *International Journal of Agricultural Sustainability*. 2017. Vol. 15. No. 2. P. 102–123. doi: 0.1080/14735903.2017.1288029.
7. *Opinion: Smart farming is key to developing sustainable agriculture* / A. Walter, R. Finger, R. Huber, et al. // *Proceedings of the National Academy of Sciences*. 2017. Vol. 114. No. 24. P. 6148–6150. doi: 10.1073/pnas.1707462114.
8. *Big data analytics and precision animal agriculture symposium: Machine learning and data mining advance predictive big data analysis in precision animal agriculture* / G. Morota, R. V. Ventura, F. F. Silva, et al. // *Journal of animal science*. 2018. Vol. 96. No. 4. P. 1540–1550. doi: 10.1093/jas/sky014.
9. Кравченко Ю. А., Лежебоков А. А., Запорожец Д. Ю. *Способы интеллектуального анализа данных в сложных системах* // *Известия Кабардино-Балкарского научного центра РАН*. 2012. № 3. С. 52–57.
10. Кремер Н. Ш. *Теория вероятностей и математическая статистика как фундамент новой комплексной прикладной дисциплины «анализ данных»* // *Современная математика и концепции инновационного математического образования*. 2019. Т. 6. № 1. С. 333–337.
11. Хворова Л. А., Топаж А. Г. *Построение моделей агро-экосистем и их адаптация к конкретным условиям* // *Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. Информатика, телекоммуникации и управление*. 2011. № 1 (115). С. 99–105.
12. *Использование динамической модели агроэкосистемы для оценки влияния климатических изменений на продуктивность посевов (теория и реализация)* / Р. А. Полуэктов, А. Г. Топаж, В. П. Якушев и др. // *Вестник Российской академии сельскохозяйственных наук*. 2012. № 2. С. 7–12.
13. *Имитационная модель агроэкосистемы как инструмент теоретических исследований* / В. Л. Баденко, А. Г. Топаж, В. В. Якушев и др. // *Сельскохозяйственная биология*. 2017. Т. 52. № 3. С. 437–445.
14. *Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход: монография* / Б. Ю. Лемешко, С. Б. Лемешко, С. Н. Постовалов и др. Новосибирск: Изд-во НГТУ, 2011. 888 с.
15. Kruskal W. H., Wallis W. A. *Use of ranks in one-criterion variance analysis*. // *Journal of the American Statistical Association*. 1952. Vol. 47. No. 260. P. 583–621.
16. *Classification and regression trees* / L. Breiman, J. H. Friedman, R. A. Olshen, et al. Monterey, CA: Wadsworth & Brooks / Cole Advanced Books & Software. 1984. 368 p. doi: 10.1201/9781315139470.
17. Орлов А. И. *Новая парадигма прикладной статистики* // *Заводская лаборатория. Диагностика материалов*. 2012. Т. 78. № 1–I. С. 87–93.
18. Gayen A., Pourghasemi H. R. *Spatial modeling of gully erosion: a new ensemble of CART and GLM data-mining algorithms* // *Spatial Modeling in GIS and R for Earth and Environmental Sciences*. Elsevier. 2019. P. 653–669. doi: 10.1016/B978-0-12-815226-3.00030-2.
19. *Засолённые почвы Венгрии: прогноз распространения на основе гидрогеологических карт* / Т. Том, Л. Пастор, Ш. Кабош и др. // *ББК 40.3. Э 40*. 2009. С. 116–158.
20. *Detection and classification of citrus diseases in agriculture based on optimized weighted segmentation and feature selection* / M. Sharif, M. A. Khan, Z. Iqbal, et al. // *Computers and electronics in agriculture*. 2018. Vol. 150. P. 220–234. doi: 10.1016/j.compag.2018.04.023.
21. Zeinvand Lorestani E., Kamkar B., Razavi S. E. *Study on the effect of agricultural management factors on fungal diseases diversity indices and wheat yield in Gorgan using decision tree analysis CART* // *Cereal Research*. 2016. Vol. 6. No. 4. P. 489–505.
22. Monjezi N. *The Application of the CART and CHIAD Algorithms in Sugar Beet Yield Prediction* // *Basrah Journal of Agricultural Sciences*. 2021. Vol. 34. No. 1. URL: <https://bjas.bajas.edu.iq/index.php/bjas/article/view/318> (дата обращения: 17.10.2021). doi: 10.37077/25200860.2021.34.1.01.
23. *Use of classification and regression trees (CART) for analyzing determinants of winter wheat yield variation among fields in Poland* / M. Iwańska, A. Oleksy, M. Dacko, et al. // *Biometrical Letters*. 2018. Vol. 55. No. 2. P. 197–214. doi: 10.2478/bile-2018-0013.
24. Kalichkin V. K., Alsova O. K., Maksimovich K. Y. *Application of the decision tree method for predicting the yield of spring wheat* // *IOP Conference Series: Earth and Environmental Science*. IOP Publishing. 2021. Vol. 839. No. 3. P. 032042. URL: <https://iopscience.iop.org/article/10.1088/1755-1315/839/3/032042> (дата обращения: 15.10.2021). doi: 10.1088/1755-1315/839/3/032042.

Поступила в редакцию 02.12.2021  
После доработки 28.12.2021  
Принята к публикации 20.01.2021