

# ПРИМЕНЕНИЕ БАЙЕСОВСКОЙ СЕТИ ДОВЕРИЯ И МУЛЬТИНОМИАЛЬНОЙ ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ ДЛЯ ПРОГНОЗИРОВАНИЯ СТЕПЕНИ ЗАСОРЕННОСТИ СЕЛЬСКОХОЗЯЙСТВЕННЫХ ЗЕМЕЛЬ

**В.К. Каличкин<sup>1</sup>, К.Ю. Максимович<sup>1</sup>, В.А. Шпак<sup>1</sup>, Р.Р. Галимов<sup>1</sup>, А.Л. Пакуль<sup>2</sup>**

<sup>1</sup> Сибирский федеральный научный центр агробиотехнологий Российской академии наук (СФНЦА РАН), Новосибирская область, р.п. Краснообск, Россия

<sup>2</sup> Кемеровский научно-исследовательский институт сельского хозяйства – филиал СФНЦА РАН, Кемеровская обл., п. Новостройка, Россия

Исследованы возможности применения Байесовской сети доверия (БСД) и мультиномиальной логистической регрессии (МНЛР) для прогнозирования степени засоренности земель сельскохозяйственного назначения. Рассчитана вероятность превышения экономического порога вредоносности (ЭПВ) при участии обеих моделей. Выполнено моделирование влияния природных и антропогенных факторов с использованием БСД, а также осуществлен прогноз превышения ЭПВ по категориям с помощью МНЛР. Для обучения моделей использованы данные длительного многофакторного полевого опыта Кемеровского НИИСХ – филиала СФНЦА РАН. Учитывая особенности статистической выборки, определены основные предикторы моделей, влияющие на засоренность земель. Выбранными предикторами были агротехнические приемы (системы обработки почвы, предшественники) и агрометеорологические ресурсы (суммы активных температур воздуха, осадки). Объясненная часть дисперсии по мере Нэйджелкерка, составляет 80,9 %, что говорит о высоких прогностических возможностях применения МНЛР. Прогнозные результаты обеих моделей совпали в 79 % случаев, что указывает на достижение высоких показателей меры близости прогнозов по БСД и МНЛР. Обе модели показали достаточно высокую достоверность при верификации на эмпирических данных за прошлые годы и могут быть использованы в качестве инструмента для прогноза. Следующим этапом работы станет совместное использование БСД и МЛНР, которое может способствовать усилению достоинств обоих подходов и устранению недостатков отдельных из них.

*Ключевые слова:* Байесовские сети доверия, мультиномиальная логистическая регрессия, прогнозирование, биологические системы, сорная растительность, экономический порог вредоносности

## ВВЕДЕНИЕ

По данным ФАО потери растениеводческой продукции от сорняков, болезней и вредителей достигают 30 % [1], по другим оценкам – от 20 до 80 % [2]. Использование пестицидов – стандартная практика по защите растений от вредных организмов. Несмотря на изменение химического состава пестицидов и современные технологии их применения, общая глобальная тенденция использования химических средств нарастает, причем гербициды составляют не менее две трети от их общего количества [3]. Растут экологические проблемы, связанные с массовым применением пестицидов на посевах сельскохозяйственных культур. Например, исследования показали, что глифосат (распространенный гербицид, используемый с 1974 г.) обладает канцерогенными свойствами для человека [4]. Это требует сбалансированного подхода при применении гербицидов на посевах сельскохозяйственных культур.

Увеличение пестицидной нагрузки на агроценоз не является на самом деле приемлемой альтернативой известным научно-обоснованным агротехническим

приемам регулирования вредных организмов. Однако специалисту занятому непосредственно в производстве оказывается весьма затруднительным, при потенциальном наличии знаний без соответствующей систематизации, полностью оценить их и, следовательно, выработать приемлемую стратегию «хозяйственного поведения» в зависимости от прогнозируемой и складывающейся обстановки.

Для решения этой проблемы целесообразно применение методов искусственного интеллекта и машинного обучения с целью прогнозирования изменения различных условий и создания систем поддержки принятия решений. Способность предвидеть развитие вредных организмов, оценить возможные риски потери урожайности и оперативно принять взвешенное решение закладывает основу эффективного управления в растениеводстве. Возможности прогнозирования при управлении объектами сельскохозяйственного назначения в условиях неопределенности стратегически важны. В настоящее время разработаны и применяются различные подходы к прогнозированию урожайности, развитию сорняков и болезней сельскохозяйственных культур [5-9]. Подавляющее большинство

традиционных подходов (регрессионные, автокорреляционные, спектральные и др.) могут использоваться только для анализа данных, которые подчиняются закону нормального распределения (распределение Гаусса-Лапласа). Для таких данных методы параметрической статистики достаточно эффективны и прогноз имеет приемлемую достоверность. Однако большинство данных исследуемых процессов в сельском хозяйстве не соответствуют закону нормального распределения и не являются независимыми, к тому же не имеют достаточно представительную выборку.

Одним из инструментов машинного обучения в этой ситуации можно использовать Байесовские сети доверия (БСД). Они применяются для исследований биологических объектов и процессов, в том числе в сельском хозяйстве, благодаря возможности «рассуждать» с неполной информацией и включать новую. Главным достоинством аппарата БСД является возможность визуализации между элементами модели. Кроме того, для обучения БСД применяются не сложные комплекты данных, позволяющие исследовать модель постепенно, не вникая глубоко в «технические» детали процесса. БСД применяются для решения частных задач в области сельского хозяйства: прогнозирование урожайности [10-12], защита растений [13,14], инвестиционная политика [15], оценка рисков [16] и др.

В настоящее время существует множество модификаций и «производных» регрессионных моделей, которые постепенно складываются в самостоятельные направления. Авторами работы с целью разрешения указанного выше противоречия относительно традиционных подходов по обработке данных рассмотрена возможность применения мультиномиальной логистической регрессии (МНЛР) при статистической обработке ограниченной выборки данных. МНЛР широко используется благодаря минимальным требованиям к исходным данным, высокой скорости сходимости, эффективности и наглядности представления результатов [17,18]. Требуется лишь, чтобы целевая переменная была категориальной, а независимые переменные (факторы) были метрическими или дихотомическими. При этом не делается никаких предположений о нормальности и однородности дисперсии для независимых переменных. Кроме того, для применения МНЛР не требуется больших вычислительных ресурсов и какой-либо особой настройки при реализации алгоритма. Известны удачные примеры применения МНЛР в различных приложениях: построение цифровой почвенной карты (на примере Вятско-Камской провинции дерново-подзолистых почв южной тайги) [19], анализ смертности от коронавируса в регионах РФ [20],

исследование составов и динамики групп, занятых на рынке труда РФ [21].

Целью исследований является сравнительная оценка степени достоверности прогноза засоренности рабочего участка сельскохозяйственных земель при использовании моделей БСД и МНЛР.

## МАТЕРИАЛЫ И МЕТОДЫ

В исследовании использованы материалы наблюдений за развитием злаковых сорняков в посевах сельскохозяйственных культур в длительных полевых опытах Кемеровского НИИСХ – филиала СФНЦА РАН за 2013-2019 гг. В ходе анализа учитывались как качественные факторы: предшественники и системы обработки почвы, так и количественные – суммы активных температур воздуха и осадки за период с третьей декады апреля по конец мая. Набор погодных условий весны обусловлен их взаимосвязью с интенсивностью всходов сорной растительности и биологической спелостью почвы. В исследованиях использовали сведения Кемеровского поста метеонаблюдений в Кемеровской области. Источником данных был веб-ресурс «погода и климат» (<http://www.pogodaiklimat.ru/>).

В основу метода БСД заложена формула Байеса, которая вытекает из определения условной вероятности:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}, \quad (1)$$

где  $P(A)$  – априорная вероятность гипотезы  $A$  (смысл такой терминологии см. ниже);

$P(A|B)$  – вероятность гипотезы  $A$  при наступлении события  $B$  (апостериорная вероятность);

$P(B|A)$  – вероятность наступления события  $B$  при истинности гипотезы  $A$ ;

$P(B)$  – полная вероятность наступления события  $B$ .

Реализация БСД осуществлялась с помощью программного пакета Netica версии «6.07». Для определения превышения ЭПВ злаковыми сорняками применена формула условной вероятности. Введены два возможных исхода, по которым степень засоренности посевов сорняками превысила ЭПВ, либо превышения не было. Также заложена априорная вероятность того, что агротехнические приемы влияют на уровень интенсивности всходов злаковых сорняков, а погода весны обуславливает этот уровень и темпы развития. Для расчётов вероятностного события «Превышение ЭПВ» были применены следующие вероятностные отношения событий, такие как: «Обработка почвы», «Предшественник», «Сценарий погоды весны», «Потенциал роста», «Банк семян». При этом переменные «Банк семян», «Потенциал роста» принимают лишь одно из двух возможных значений: Low (низкий) и High (высокий).

Для характеристики засоренности рабочего участка вводили показатель количественного состава сорняков, принимающий значения одной из 4 категорий превышения экономического порога вредоносности (ЭПВ): Н – нет превышения; НП – низкое превышение (12 шт./м<sup>2</sup>); СП – среднее превышение (15 шт./м<sup>2</sup>); ЗП – значительное превышение (более чем в 6-7 раз) [22].

При составлении прогноза с помощью МНЛР, необходимо получить связь ЭПВ (зависимая переменная) от независимых факторов (переменных): способа обработки почвы, предшественника, суммы эффективных температур и осадков. Поскольку зависимая переменная ЭПВ является категориальной и принимает значения перечисленных 4 категорий, то требуемую зависимость можно получить с использованием модели мультиномиальной логистической регрессии. При этом одна из категорий зависимой переменной становится опорной (эталонной) и все другие категории сравниваются с ней. Независимые переменные могут быть категориальными или количественными. Процедура мультиномиальной логистической регрессии прогнозирует вероятность принадлежности к каждой категории зависимой переменной по значениям независимых переменных. Окончательный выбор предсказываемой категории для зависимой переменной производится по правилу наибольшей вероятности принадлежности. Предполагается, что имеется серия из  $N$  наблюдений. Каждое наблюдение состоит из набора  $m$  независимых переменных  $X_i$ ,  $i=1, \dots, m$  (также называемых предикторами) и соответствующего категориального значения зависимой переменной  $Y_j$ ,  $j=1, \dots, K$ , которая может принимать одно из  $K$  возможных значений (категорий). Для каждой категории зависимой переменной (за исключением опорной) строится уравнение бинарной логистической регрессии, которое определяет отношение  $p_j$  - вероятности отнесения рассматриваемого наблюдения к этой категории к вероятности для опорной категории -  $p_1$ :

$$\ln\left(\frac{p_j}{p_1}\right) = \alpha_j + \sum_i \beta_j^i X_i, \quad j = 2, \dots, K \quad (2)$$

где  $\alpha_j$  – свободный член,  $\beta_j^i$  - коэффициент регрессии, связанный с  $i$ -й независимой переменной для категории  $j$ .

Здесь 1-категория принята в качестве опорной, вместо нее можно использовать любую другую.

Неизвестные коэффициенты  $\alpha_j$ ,  $\beta_j^i$ ,  $j=2, \dots, K$ ,  $i=1, \dots, m$  совместно оцениваются для всей выборки наблюдений по максимальной апостериорной оценке (MAP), которая является расширением максимального правдоподобия с использованием регуляризирующих весов. Решение находится итеративно путем минимизации пересмотренных наименьших квадратов (IRLS).

Используя (2) и тот факт, что сумма всех  $K$  вероятностей должна составлять единицу, получим искомые вероятности принадлежности к категориям:

$$p_1 = 1 / (1 + \sum_j \exp(\alpha_j + \sum_i \beta_j^i X_i)) \quad (3)$$

и

$$p_j = \exp(\alpha_j + \sum_i \beta_j^i X_i) / (1 + \sum_j \exp(\alpha_j + \sum_i \beta_j^i X_i)), \quad \text{для } j=2, \dots, K \quad (4)$$

При обучении МНЛР были использованы исходные данные в виде таблицы из 288 наблюдений (строк) и 8 факторов (столбцов), включая зависимую переменную ЭПВ. Число наблюдений распределено по показателям ЭПВ следующим образом: Н – 39 шт., НП – 100 шт., СП – 126 шт., ЗП – 23 шт. В качестве опорной для ЭПВ выбрана категория Н. Построение МНЛР и расчеты осуществлены в программном пакете SPSS [23,24]. Массив данных делили в соотношении 80/20, а именно 80 % примеров использовали в качестве обучающих, 20 % – в качестве тестовых для проверки результата тренировки моделей.

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Первой моделью, использованной в исследованиях, была БСД. Математически, теорема Байеса показывает взаимоотношения между вероятностью события «Превышение ЭПВ» и вероятностного отношения событий: «Сценарий погоды весны», «Потенциал роста», «Способ обработки почвы», «Предшественник», «Запас семян». Каждая из переменных («Банк семян» и «Потенциал роста») может принимать лишь одно из двух возможных значений: Low (низкий) и High (высокий) и отражают результаты причинно-следственных связей родительских вершин по отношению к ним. Графически удобно продемонстрированы взаимосвязи между объясняющими переменными и исходным результатом в виде степени превышения ЭПВ. Метод подобран с целью изучения ожидаемых реакций превышения экономического порога вредоносности в различных сценарных условиях. Фактически обучением Байесовской сети стало заполнение таблиц априорных вероятностей для основных вершин, на основе существующих данных, наблюдений и результатов исследований. Следуя вышесказанному, допускаем, что наибольшее влияние на уровень засоренности рабочего участка злаковыми сорняками (априорная вероятность количества сорняков выше ЭПВ) влияют такие факторы, как: «Обработка почвы», «Предшественник», «Сценарий погоды весны», «Потенциал роста», «Банк семян». Шесть вершин простого ациклического графа БСД показаны на рис. 1. В приведенном примере рассмотрены: теплые и влажные условия погоды весны текущего

сценария, конкретные участки, отличающиеся предшественником и обработкой почвы.

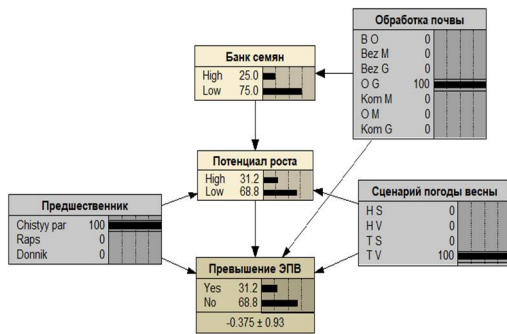


Рис. 1. Пример сети доверия: теплая, влажная погода весны, отвальная глубокая обработка почвы по чистому пару

При отвальной глубокой обработке почвы, в условиях теплой и влажной погоды текущего периода (T\_V), превышение ЭПВ сорняками не будет достигнуто с вероятностью 68,8%. Запас семян сорной растительности, при данной комбинации факторов тоже будет незначителен. Благоприятное соотношение тепла и влаги сказывается на развитии сорной растительности, обуславливая высокую интенсивность всхожести сорняков и уровень засоренности посевов. В то же время вероятность превышения ЭПВ сорняками при соответствующем предшественнике и способе обработки почвы, меняется, что связано с видом технологии возделывания предшествующей культуры, при которой, возможно осуществляется высокая эффективность контроля сорной растительности, и, следовательно, обуславливает низкий банк семян и низкую вероятность превышения ЭПВ [25]. Первичные результаты проведенного эксперимента на обучающей выборке так же позволили определить оптимальные диапазоны значений, которыми можно руководствоваться при принятии решения: верхний – 80 % и нижний – 35 % вероятности превышения ЭПВ. Достоверность результатов тестирования равна 83 %.

Вторым методом исследований предметной области было применение МНЛР. Проведены расчёты коэффициентов регрессии и оценка эффективности МНЛР. Стандартные ошибки и оценка значимости факторов получены с применением статистики Вальда. Самыми значимыми факторами являются (значимость  $\leq 0,05$ ): суммы активных температур воздуха, осадки. Из агротехнических приемов наиболее значимыми были: системы обработки почвы (без обработки, безотвальная, комбинированная минимальная, комбинированная глубокая, отвальная минимальная, отвальная глубокая, отвальная) и предшественники (чистый пар, сидеральные пары донника и рапса). Как известно, в случае обычной линейной регрессии, используется показатель R-квадрат, который описывает ту часть дисперсии,

которую можно объяснить с помощью регрессии. В случае МНЛР такую роль играет показатель псевдо R-квадрат [18]. Наиболее распространенными являются меры, предложенные Нэйджелкерком, Коксом и Снеллом [24]. В нашем случае объясненная часть дисперсии составляет по мере Нэйджелкерка (как правило, наиболее применяемой) 80,9 %, а Кокса и Снелла равна 73 %, что говорит о высоких прогностических возможностях применения МНЛР.

В таблице 1 представлены сравнительные прогностические способности данного метода, проверенные на тестовой выборке. Общая доля правильных прогнозов по всем категориям составляет 79,5%.

Табл. 1. Классификационная таблица.

Наблюдения	Предсказанные				% правильных
	Н	НП	СП	ЗП	
Н	31	8			79,5%
НП	5	75	20		75,0%
СП	0	15	106	5	84,1%
ЗП	0	0	6	17	73,9%
Общая % доля					79,5%

Выполнен тестовый прогноз для сравнения прогнозных и фактических значений на конкретной выборке по некоторым сценариям комбинации факторов с целью верификации обеих моделей на конкретном землепользовании (табл. 2).

Табл. 2. Сравнительный анализ тестового прогноза условной вероятности и степени превышения р (ЭПВ) на 2019 г.

Обработка почвы	Предшест. венник	Превышение ЭПВ по БСД, %		Прогноз превышения по МНЛР (категория превышения)
		Да	Нет	
Без обработки	Чистый пар	62,8	37,2	НП
Без обработки	Рапс	79,0	21,0	НП
Без обработки	Донник	83,7	16,3	НП
Безотвальная минимальная	Чистый пар	60,5	39,5	СП
Безотвальная минимальная	Рапс	78,0	22,0	СП
Безотвальная минимальная	Донник	82,5	17,5	СП
Безотвальная глубокая	Чистый пар	60,5	39,5	НП
Безотвальная глубокая	Рапс	78,0	22,0	НП
Безотвальная глубокая	Донник	82,5	17,5	НП
Отвальная глубокая	Чистый пар	31,2	68,8	НП
Отвальная глубокая	Рапс	65,0	35,0	НП
Отвальная глубокая	Донник	66,2	33,8	НП

Прогнозы превышения ЭПВ сорной растительностью, выполненные обеими моделями, совпали в 19 случаях из 21 при проверке на экспериментальной выборке за 2019 г. По двенадцати опытам, приведенным в табл. 2, более вероятен прогноз категории НП – низкое превышение ЭПВ, полученный по МНЛР, который так же подтверждает апостериорную вероятность БСД в 19 из 21 случаях. Максимальное значение превышения ЭПВ сорняками будет при обработке почвы «Без обработки», предшественник «Донник», погоды весны «Теплый, влажный» и составляет 83,7 %. Высокий показатель превышения ЭПВ обусловлен благоприятным соотношением тепло- и влагообеспеченности для роста и развития сорных растений, т.к. их массовые всходы отмечаются при температуре почвы 10-14°C и при высокой влажности почвы.

Обе модели показывают достаточно высокую эффективность при верификации на эмпирических данных за прошлые года. МНЛР позволяет не только уточнить прогноз, полученный при помощи БСД, но и спрогнозировать «категорию» превышения ЭПВ. Учитывая вышесказанное, можно говорить о достижении высоких показателей меры близости прогнозов по БСД и МНЛР и существенной достоверности полученных результатов. Обе модели могут быть использованы в качестве инструмента для прогноза (как самостоятельно, так и в основе гибридного метода) на будущий вегетационный период по рабочим участкам землепользования.

### ЗАКЛЮЧЕНИЕ

Проведен сравнительный анализ и оценена эффективность применения БСД и МНЛР для прогнозирования степени засоренности земель сельскохозяйственного назначения. Используемые модели включали анализ агротехнических факторов (система обработки почвы, предшественник, банк семян) и агрометеорологический ресурс (сумма активных температур воздуха, осадки), а также использование эвристики для построения БСД и вычисления априорных вероятностей. Достоверность результатов тестирования БСД оказалась равной 83 %. Объясненная часть дисперсии по мере Нэйджелкерка, составила 80,9 %. Прогнозные результаты обеих моделей совпали в 79 % случаев. Обе модели показали достаточно высокую эффективность при верификации на эмпирических данных за прошлые годы и могут быть использованы в качестве инструмента для прогноза засоренности земель сельскохозяйственного назначения. Следующим этапом работы станет создание гибридной модели БСД и МНЛР, которая может способствовать усилению достоинств обоих подходов и устранению недостатков отдельных из них.

### СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Долженко, В.И. Повысить фитосанитарную безопасность Российской Федерации // Защита и карантин растений. - 2011. - № 2. - С. 4-7.
2. Oerke E.C., Dehne H.W., Schönbeck F., Weber A., «Crop production and crop protection: estimated losses in major food and cash crops,» Amsterdam: Elsevier Science, 808 p. 2012.
3. Benbrook C.M., «Trends in glyphosate herbicide use in the United States and globally,» Environmental Sciences Europe, vol. 28, no 1, pp. 1-15. 2016.
4. Myers J.P., Antoniou M.N., Blumberg B., «Concerns over use of glyphosate-based herbicides and risks associated with exposures: a consensus statement,» Environmental Health, vol. 15, no 1, pp. 1-13. 2016.
5. Van Wart J., Grassini P., Cassman K. G., «Impact of derived global weather data on simulated crop yields,» Global change biology, vol. 19, no 12, pp. 3822-3834. 2013.
6. Chipanshi A., Zhang Y., Kouadio L., Newlands N., «Evaluation of the Integrated Canadian Crop Yield Forecaster (ICCYF) model for in-season prediction of crop yield across the Canadian agricultural landscape,» Agricultural and Forest Meteorology, vol. 206, pp. 137-150. 2015.
7. Shahhosseini M., Hu G., Huber I., Archontoulis Sv., «Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt,» Scientific reports, vol. 11, no 1, pp. 1-15. 2021.
8. Asad M. H., Bais A., «Weed detection in canola fields using maximum likelihood classification and deep convolutional neural network,» Information Processing in Agriculture, vol. 7, no 4, pp. 535-545. 2019.
9. Bedi P., Gole P., «Plant disease detection using hybrid model based on convolutional autoencoder and convolutional neural network,» Artificial Intelligence in Agriculture, vol. 5, pp. 90-101, 2021.
10. Nathaniel K. N., Zamar D. S., Kouadio L. A., Zhang Y., Chipanshi A., Potgieter A., Toure S. and Hill H. S., «An integrated, probabilistic model for improved seasonal forecasting of agricultural crop yield under environmental uncertainty,» Frontiers in Environmental Science, vol. 2. pp. 17-34. 2014.
11. Gandhi, N., Armstrong, L.J., Petkar, O., «Predicting rice crop yield using Bayesian networks,» 2016 International Conference on Advances in Computing, Communications and Informatics. (ICACCI). IEEE, pp. 795-799, 2016.
12. Chandrababha M., Dhanaraj R.K., «Machine learning based Pedantic Analysis of Predictive Algorithms in Crop Yield Management,» 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA). IEEE, pp. 1340-1345, 2020.
13. Bressan, G.M., Oliveira, V.A., Hruschka, E.R., Nicoletti, M.C., «Using Bayesian networks with rule extraction to infer the risk of weed infestation in a corn-crop,» Engineering Applications of Artificial Intelligence, vol. 22. no 4-5, pp. 579-592, 2009.
14. Pérez-Ariza C.B., Nicholson A.E., Flores M.J., «Prediction of coffee rust disease using bayesian networks,» Proceedings of the Sixth European Workshop on Probabilistic Graphical Models, vol. 6, pp. 259-266, 2012.
15. Yet B., Lamanna C., Shepherd K.D., Rosenstock T.S., «Evidence-based investment selection: Prioritizing agricultural development investments under climatic and socio-political risk using Bayesian networks Plos one,» vol. 15, no. 6, pp. e0234213. 2020.
16. Yet B., Constantinou A., Fenton N., Neil M., Luedeling E., and Shepherd K. A., «Bayesian network framework for project cost, benefit and risk analysis with an agricultural development case study,» Expert Systems with Applications, vol. 60, pp. 141-155, 2016.
17. Бююль А., Цеффель П. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей: пер с нем./Ахим Бююль, Петер Цеффель. – СПб.: ООО «ДиаСофтЮП», 608 с. 2002.
18. Наследов А.Д. IBM SPSS Statistics 20 и AMOS: профессиональный статистический анализ данных. – СПб.: Питер, 2013. 416 с.

19. Dokuchaev P. M., Meshalkina J. L., Yaroslavtsev A. M., «Mathematical models application for mapping soils spatial distribution on the example of the farm from the North of Udmurt Republic of Russia.» IOP Conference Series: Earth and Environmental Science. – IOP Publishing, vol. 107, no. 1, p. 012113. 2018.
20. Иванова, А.А. Применение мультиномиальной логистической регрессии для анализа смертности от коронавируса в регионах РФ / А.А. Иванова, Ю.В. Скородумова, С.А. Кулагина, А.А. Тарасов, Д. Л. Павлов, // Молодой ученый. - 2020. - № 27 (317). - С. 9-12.
21. Гимпельсон, В.Е. Низкооплачиваемые рабочие места на российском рынке труда: есть ли выход и куда он ведет? / В.Е. Гимпельсон, Р.И. Капелюшников, А.В. Шарунина. – М.: Изд. дом Высшей школы экономики. (Серия WP3 «Проблемы рынка труда»), 2018. - 58 с.
22. Алехин В.Т. Экономические пороги вредности вредителей, болезней и сорных растений в посевах сельскохозяйственных культур: справочник / В.Т. Алехин, В.В. Михайликова, Н.Г. Михина – М.: ФГБНУ «Росинформагротех», 2016. - 76 с.
23. Cohen A., Rom M., «A method for hypothesis tests in polychotomous logistic regression.» Computational statistics & data analysis, vol. 17, no 3, pp. 277-288. 1994.
24. SPSS Base 8.0 для Windows. Руководство по применению. – М.: СПСС Русь, Т.658, 1998. – 397 с.
25. Каличкин, В. К. Байесовская сеть доверия как инструмент прогнозирования уровня засоренности овсом пустым (овсюгом) / В. К. Каличкин, К. Ю. Максимович, Р. Р. Галимов // Труды международной научной онлайн-конференции «АгроНаука-2020». - 2020. - С. 49-54.
- Каличкин Владимир Климентьевич, доктор сельскохозяйственных наук, главный научный сотрудник; адрес: СФНЦА РАН, Краснообск, Новосибирская область, 630501, Россия; e-mail: kvk@ngs.ru*
- Максимович Кирилл Юрьевич, аспирант, младший научный сотрудник; адрес: СФНЦА РАН, Краснообск, Новосибирская область, 630501, Россия; e-mail: kiri-maksimovi@mail.ru*
- Шпак В.А. Исследователь; адрес: СФНЦА РАН, Краснообск, Новосибирская область, 630501, Россия; e-mail: shpakva54@gmail.com*
- Галимов Руфан Рамильевич, аспирант, младший научный сотрудник; адрес: СФНЦА РАН, Краснообск, Новосибирская область, 630501, Россия; e-mail: rufangalimov@yandex.ru*
- Пакуль А.Л. Исследователь; адрес: Улица Центральная, 47, Новостройка, Кемеровская область, 650510, Россия; e-mail: lelikpakul@mail.ru*

# APPLICATION OF THE BAYESIAN TRUST NETWORK AND MULTINOMIAL LOGISTIC REGRESSION TO PREDICT THE DEGREE OF CONTAMINATION OF AGRICULTURAL LAND

V.K. Kalichkin<sup>1</sup>, K.Yu. Maksimovich<sup>1</sup>, V.A. Shpak<sup>1</sup>, R.R. Galimov<sup>1</sup>, A.L. Pakul<sup>2</sup>

<sup>1</sup>*Siberian Federal Scientific Centre of AgroBioTechnologies of the Russian Academy of Science, Novosibirsk region, Krasnoobsk, Russia*

<sup>2</sup>*Kemerovo Research Institute of Agriculture-branch of the Siberian Federal Scientific Center of Agrobiotechnologies of the Russian Academy of Sciences, Kemerovo region, Novostroika village, Russia*

The possibilities of using the Bayesian Network of Trust (BSD) and multinomial logistic regression (MNLr) to predict the degree of contamination of agricultural land are investigated. The probability of exceeding the economic threshold of harmfulness (ETH) with the participation of both models is calculated. Modeling of the influence of natural and anthropogenic factors using BSD was carried out, and the forecast of the excess of ETH by category was carried out using MNLr. To train the models, data from a long-term multifactorial field experience of the Kemerovo Research Institute of Agricultural Sciences - branch of the SFSCA RAS were used. Taking into account the features of the statistical sample, the main predictors of the models affecting land contamination are determined. The selected predictors were agrotechnical techniques (tillage systems, precursors) and agrometeorological resources (sums of active air temperatures, precipitation). The explained part of the variance with the Nagelkerk measure is 80.9%, which indicates high prognostic possibilities of using MNLr. The forecast results of both models coincided in 79% of cases, which indicates the achievement of high indicators of the measure of proximity of forecasts for BSD and MNLr. Both models have shown sufficiently high reliability when verified on empirical data from previous years and can be used as a tool for forecasting. The next stage of the work will be the joint use of BSD and MDR, which can contribute to strengthening the advantages of both approaches and eliminating the shortcomings of some of them.

Keywords: Bayesian networks of trust, multinomial logistic regression, forecasting, biological systems, weed vegetation, economic threshold of harmfulness.

## REFERENCES

1. Dolzhenko V. I. To increase the phytosanitary safety of the Russian Federation // Protection and quarantine of plants. 2011. No. 2. pp. 4-7.
2. Oerke E.C., Dehne H.W., Schönbeck F., Weber A. Crop production and crop protection: estimated losses in major food and cash crops. – Amsterdam: Elsevier Science, 2012. 808 p.
3. Benbrook C.M. Trends in glyphosate herbicide use in the United States and globally // Environmental Sciences Europe. 2016. Vol. 28. N 1. P. 1-15.
4. Myers J.P., Antoniou M.N., Blumberg B. Concerns over use of glyphosate-based herbicides and risks associated with exposures: a consensus statement // Environmental Health. 2016. Vol. 15. No 1. P. 1-13.
5. Van Wart J., Grassini P., Cassman K. G. Impact of derived global weather data on simulated crop yields // Global change biology. 2013. Vol. 19. N 12. P. 3822-3834.
6. Chipanshi A., Zhang Y., Kouadio L., Newlands N. Evaluation of the Integrated Canadian Crop Yield Forecaster (ICCYF) model for in-season prediction of crop yield across the Canadian agricultural landscape // Agricultural and Forest Meteorology. 2015. Vol. 206. P. 137-150.
7. Shahhosseini M., Hu G., Huber L., Archontoulis S. Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt // Scientific reports. 2021. Vol. 11. N 1. P. 1-15.
8. Asad M. H., Bais A. Weed detection in canola fields using maximum likelihood classification and deep convolutional neural network // Information Processing in Agriculture. 2019. Vol. 7. N 4. P. 535-545.
9. Bedi P., Gole P. Plant disease detection using hybrid model based on convolutional autoencoder and convolutional neural network // Artificial Intelligence in Agriculture. 2021. Vol. 5. P. 90-101.
10. Nathaniel K. N., Zamar D. S., Kouadio L. A., Zhang Y., Chipanshi A., Potgieter A., Toure S. and Hill H. S. An integrated, probabilistic model for improved seasonal forecasting of agricultural crop yield under environmental uncertainty // Frontiers in Environmental Science. 2014. Vol. 2. P. 17-34.
11. Gandhi, N., Armstrong, L.J., Petkar, O. Predicting rice crop yield using Bayesian networks // 2016 International Conference on Advances in Computing, Communications and Informatics. (ICACCI). IEEE, 2016. P. 795-799.
12. Chandraprabha M., Dhanaraj R.K. Machine learning based Pedantic Analysis of Predictive Algorithms in Crop Yield Management // 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA). IEEE, 2020. P. 1340-1345.
13. Bressan, G.M., Oliveira, V.A., Hruschka, E.R., Nicoletti, M.C. Using Bayesian networks with rule extraction to infer the risk of weed infestation in a corn-crop // Engineering Applications of Artificial Intelligence. 2009. Vol. 22. N 4-5. P. 579-592.
14. Pérez-Ariza C.B., Nicholson A.E., Flores M.J. Prediction of coffee rust disease using bayesian networks // Proceedings of the Sixth European Workshop on Probabilistic Graphical Models. 2012. Vol. 6. P. 259-266.
15. Yet B., Lamanna C., Shepherd K.D., Rosenstock T.S. Evidence-based investment selection: Prioritizing agricultural development investments under climatic and socio-political risk using Bayesian networks // Plos one. 2020. Vol. 15. № 6. P. e0234213.
16. Yet B., Constantinou A., Fenton N., Neil M., Luedeling E., and Shepherd, K. A Bayesian network framework for project cost, benefit and risk analysis with an agricultural development case study // Expert Systems with Applications. 2016. Vol. 60. P. 141-155.
17. . Byul A., Cefel P. SPSS: the art of information processing. Statistical data analysis and recovery of hidden patterns: translated from German./Achim Byul, Peter Cefel. - St. Petersburg: LLC "dasOftYUP", 2002. 608 p.

18. Heritov A.D. IBM SPSS Statistics 20 and AMOS: Professional statistical data analysis. - St. Petersburg: Peter, 2013. 416 p.
19. Dokuchaev P. M., Meshalkina J. L., Yaroslavtsev A. M., «Mathematical models application for mapping soils spatial distribution on the example of the farm from the North of Udmurt Republic of Russia,» IOP Conference Series: Earth and Environmental Science. – IOP Publishing, vol. 107, no. 1, p. 012113. 2018.
20. Ivanova A.A., Skorodumova Yu.V., Kulagina S.A., Tarasov A.A., Pavlov D. L. Application of multinomial logistic regression for the analysis of mortality from coronavirus in the regions of the Russian Federation // Young Scientist. 2020. No. 27 (317). P. 9-12.
21. Gimpelson, V.E., Kapelyushnikov, R.I., Sharunina, A.V. Low-paid jobs in the Russian labor market: is there a way out and where does it lead? - M.: Publishing House of the Higher School of Economics. 2018– - (WP3 series "Labor Market Problems"). 58 p.
22. Alekhin V.T., Mikhailova V.V., Mikhina N.G. Economic thresholds of harmfulness of pests, diseases and weeds in agricultural crops: handbook. - M.: FSBI "rOsinformagrotech". 2016. 76 p.
23. Cohen A., Rom M. A method for hypothesis tests in polychotomous logistic regression // Computational statistics & data analysis. 1994. Vol. 17. N 3. P. 277-288.
24. SPSS Base 8.0 dlja Windows. Rukovodstvo po primeneniju (Application Guide), M.: SPSS Rus', vol. 658, 1998, 397 p.
25. Kalichkin V. K., Maksimovich K. Yu., Galimov R. R. Bayesian network of trust as a tool for predicting the level of contamination with empty oats (oatmeal)//Proceedings of the international scientific online conference "Agronauka 2020", 2020, pp. 49-54.

*Kalichkin Vladimir Klimentyevich, Doctor of Science in Agriculture, Head Researcher; address: PO Box 463, SFSCA RAS, Krasnoobsk, Novosibirsk Region, 630501, Russia; e-mail: kvk@ngs.ru*

*Maksimovich Kirill Yurievich, Graduate student, Junior Researcher; address: PO Box 463, SFSCA RAS, Krasnoobsk, Novosibirsk Region, 630501, Russia; e-mail: kiri-maksimovi@mail.ru*

*Shpak V.A. Researcher; address: PO Box 463, SFSCA RAS, Krasnoobsk, Novosibirsk Region, 630501, Russia; e-mail: shpakva54@gmail.com*

*Galimov Rufan Ramilevich, Graduate student, Junior Researcher; address: PO Box 463, SFSCA RAS, Krasnoobsk, Novosibirsk Region, 630501, Russia; e-mail: rufangalimov@yandex.ru*

*Pakul A.L. Researcher; address: 47 Tsentralnaya Street, Novostroika, Kemerovo region, 650510, Russia; e-mail: lelikpakul@mail.ru*