

Применение искусственной нейронной сети для прогнозирования урожайности пшеницы

¹*АУБАКИРОВА Гауля Фаритовна, старший преподаватель, geli8778@mail.ru,

¹МОЛДАХМЕТОВ Саят Сайранович, PhD, доцент, sayatmoldakhmetov@gmail.com,

¹ЗЫКОВА Наталья Владимировна, старший преподаватель, zykovanata1978@gmail.com,

¹ГАГОЛИНА Ольга Сергеевна, старший преподаватель, olenka-russ@mail.ru,

¹ДАРИЙ Елена Михайловна, старший преподаватель, dariyelena@mail.ru,

¹НАО «Северо-Казахстанский университет имени Манаша Козыбаева», Казахстан, Петропавловск, ул. Пушкина, 86,

*автор-корреспондент.

Аннотация. Приведены результаты прогнозирования урожайности пшеницы на основе искусственной нейронной сети (ИНС). Применение ИНС обусловлено использованием в качестве предикторов таких нелинейных факторов, как температура воздуха и количество осадков. Также в исследовании были использованы метеорологические, агрохимические и фитосанитарные факторы, оказывающие наибольшее влияние на урожайность пшеницы. Эти факторы использовались в качестве входных переменных нейронной сети обратного распространения ошибки. После обучения ИНС и получения прогнозных данных были достигнуты хорошие результаты для прогнозирования урожайности пшеницы ($r = 0,52$, $MAPE = 12,02\%$, $RMSE = 3,368$). Разработанная модель может быть легко адаптирована для других культур и мест, а также может быть улучшена путем использования новых источников данных.

Ключевые слова: прогнозирование урожайности, искусственная нейронная сеть, урожайность пшеницы, BPNN, прогноз сезонной урожайности сельскохозяйственных культур, независимые переменные, модель нейронной сети, предикторы урожайности, нейронная сеть обратного распространения.

Введение

Производство зерна занимает приоритетное место как в агропромышленном комплексе, так и в экономике Республики Казахстан. Львиная часть (около 80%) валового сбора зерна принадлежит пшенице. Именно поэтому качеству зерна пшеницы и ее урожайности уделяется огромное значение.

Прогнозирование урожайности – важная, но сложная проблема, необходимая для устойчивой интенсификации и эффективного использования природных ресурсов [1]. Заблаговременное и точное прогнозирование урожайности было и остаётся актуальной проблемой для любого государства, потому что эффективность длинной агропродовольственной цепочки зависит от точности прогноза.

Применяют различные подходы при прогнозировании урожайности, основные из них – это экспертные оценки (например, интервью и полевые исследования), статистические модели и модели, основанные на процессах. Интервью с фермерами, как правило, предоставляют собой сугубо субъективные ожидания урожайности к концу сезона [2]. Полевые исследования с обрезкой урожая дают объективную оценку урожай-

ности до уборки урожая. Статистические модели применяют различные методы (регрессия, байесовские подходы, методы машинного обучения), для построения регрессионных зависимостей между различными статистическими данными, полученными при помощи дистанционных и метеорологических наблюдений [3]. Одним из самых распространенных методов прогнозирования урожайности является статистическая модель на основе агрометеорологических данных. Однако одним из главных недостатков этого метода является то, что многочисленные факторы окружающей среды (температура воздуха, количество осадков), оказывающие наибольшее влияние на формирование урожайности пшеницы нелинейны, т.е. могут иметь большие отклонения от средних значений и поэтому необходимо отойти от традиционных методов в пользу более точных методов прогнозирования. Наиболее подходящей альтернативой являются модели, основанные на искусственных нейронных сетях.

Преимуществом моделей с использованием нейронных сетей является высокая точность прогноза и возможность повышения урожайности. Алгоритмы построения и обучения нейронной сети базируются на функциях, которые опреде-

ляют зависимость признаков и предикторов с выходными данными, в данном случае это урожайность. ИНС имеют некоторые явные преимущества перед традиционными моделями. Так, они могут моделировать нелинейные отношения между несколькими источниками данных, их производительность обычно улучшается при большом количестве обучающих данных.

В данном исследовании представлен метод прогнозирования урожайности пшеницы при помощи ИНС на основе метеорологических, агрохимических и фитосанитарных данных. Целью данного исследования была разработка модели прогнозирования урожайности с использованием ИНС, связывающей урожай пшеницы с количеством продуктивной влаги в почве, плодородием почвы, погодой и факторами наличия вредителей, болезней и сорняков для Северо-Казахстанской области (далее СКО). Данная модель может быть легко адаптирована для других культур и мест, а также может быть улучшена путем использования новых источников данных и более продвинутых функций.

Материалы исследования

Северо-Казахстанская область – одна из ведущих областей сельскохозяйственного производства Республики Казахстан. Область занимает 3,6% территории Республики, при этом дает 25% от общего сбора зерновых культур. Площадь области составляет 98 тыс. км², из них сельхозугодия занимают 58,8 тыс. км², что составляет 60% территории области [4]. Для разработки ИНС использовались данные урожайности 13 районов Северо-Казахстанской области с 2008 по 2020 гг. (таблица 1), которые предоставлены Бюро наци-

ональной статистики Агентства по стратегическому планированию и реформам Республики Казахстан [5] (таблица 2).

В исследовании мы предположили, что урожайность пшеницы варьируется внутри и между районами в зависимости от различных факторов, из которых было выбрано 16 наименований входных данных для каждого поля (таблица 2).

Ежедневные климатические данные наблюдений, включая осадки, а также минимальную и максимальную температуру воздуха, были получены от гидрометеорологических станций и постов Филиала РГП «Казгидромет» по Северо-Казахстанской области. Осадки были разделены на четыре периода и упоминаются как май, июнь, июль и август.

Данными агрохимического фактора использовались результаты замеров, произведенных гидрометеорологическими станциями и постами с целью получения данных о запасе продуктивной влаги в 100 см слое почвы в качестве входных показателей ИНС. Также в ИНС были включены четыре фактора плодородия почвы: pH почвы, фосфор, азот, а также процентное содержание гумуса в почве (таблица 2). Коэффициенты плодородия почвы были получены из анализа образцов почвы, взятых с полей. Данные получены за период с 2008 по 2020 гг. Были рассчитаны средние значения для дальнейшего использования в эксперименте.

Также в исследовании были учтены фитосанитарные факторы, наносящие наибольший ущерб урожаю зерна. Распространение вредителей, болезней и сорняков наблюдалось на всех полях и учитывалось в виде коэффициента от нуля (плотность = 0) до очень высокого (плотность = 4).

Таблица 1 – Урожайность пшеницы яровой по Северо-Казахстанской области

Год	Районы Северо-Казахстанской области												
	Айыртауский	Акжарский	Жумабаева	Есильский	Жамбылский	Кызылжарский	Мамлютский	Шалакына	Аккайынский	Тайыншинский	Тимирязевский	Уалихановский	Мусрепова
2008	13,0	5,9	14,9	13,0	13,7	14,2	13,5	11,4	13,9	12,5	12,7	6,5	13,4
2009	14,1	13,0	17,1	15,7	13,6	16,8	15,7	13,1	17,6	13,9	14,4	13,1	12,7
2010	9,2	12,0	11,8	10,0	7,6	12,6	12,5	8,9	10,6	9,3	7,6	9,6	7,2
2011	23,0	15,8	21,9	23,4	19,4	22,8	21,0	19,1	21,7	19,1	25,5	14,6	22,5
2012	15,5	7,7	13,7	16,1	12,7	16,5	15,9	8,7	14,0	10,1	9,5	7,4	8,5
2013	12,6	9,1	15,2	11,5	11,0	13,7	11,2	9,1	12,8	12,5	9,4	12,3	14,5
2014	13,8	12,9	6,7	13,1	13,6	16,0	13,6	11,9	14,1	13,4	12,3	12,5	13,7
2015	14,8	15,0	18,3	17,7	17,3	18,7	15,9	13,2	18,2	12,9	14,4	15,1	14,5
2016	14,1	13,2	17,4	16,4	17,4	16,6	14,3	14,1	18,2	11,8	13,8	12,7	14,8
2017	16,8	12,6	19,9	18	18,8	17,4	17,4	15,9	19,8	14,6	17,3	11,6	16,2
2018	13,1	14,2	16,7	17	19,8	16,9	14,6	14,3	16,9	13,9	15,6	15,3	14,6
2019	12,7	12,4	13,5	17,3	15,4	18,2	17,7	14,1	16,3	14,2	11,8	14,5	13,2
2020	14,0	13,2	14,0	18,1	14,3	18,3	17,9	15,5	16,1	12,4	14,5	10,2	14,0

Таблица 2 – Слои информации, использованные для построения ИНС

Входной слой				Выходной слой
Тип	Фактор	Размерность	Обозначение	
Агрохимические факторы	влажность почвы в 100 см слое почвы	%	X ₁	урожайность (y)
	содержание азота	мг/кг	X ₂	
	содержание фосфора	мг/кг	X ₃	
	кислотность почвы (Ph)	мг/кг	X ₄	
	гумус	%	X ₅	
Метеорологические факторы	количество осадков за май	мм	X ₆	
	количество осадков за июнь	мм	X ₇	
	количество осадков за июль	мм	X ₈	
	количество осадков за август	мм	X ₉	
	ср. температура воздуха за май	°C	X ₁₀	
	ср. температура воздуха за июнь	°C	X ₁₁	
	ср. температура воздуха за июль	°C	X ₁₂	
	ср. температура воздуха за август	°C	X ₁₃	
Фитосанитарные факторы	сорняки	0-4 у.е.	X ₁₄	
	вредители	0-4 у.е.	X ₁₅	
	болезни	0-4 у.е.	X ₁₆	

Методы исследования

Существует несколько типов моделей искусственной нейронной сети (ИНС), различаемых по способу соединения узлов, методам расчета весов, количеству узлов в скрытых слоях и типу передаточной функции между слоями. В данном исследовании для прогнозирования урожайности используется нейронная сеть обратного распространения ошибки. Ее основная идея состоит в том, чтобы пересмотреть веса и пороги сети путем обратного распространения, чтобы минимизировать ошибку между фактическим выходным значением и ожидаемым выходным значением. Топология применяемой нейронной сети обратного распространения ошибки с одним скрытым слоем показана на рисунке 1.

Важным элементом прогнозного моделирования является точная оценка правильности функциональности модели. Для этого используются индексы качества ретроспективного прогноза. Одним из наиболее часто используемых индикаторов ошибки прогноза является средняя абсолютная ошибка в процентах (MAPE), которая рассчитывается по формуле [6].

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - y_i'}{y_i} \right| \times 100\%. \quad (1)$$

MAPE измеряет ошибку в процентах и указывает среднее процентное отклонение между прогнозируемым значением и фактической реализацией. Согласно [7], если значение MAPE ниже 10%, степень соответствия модели идеальна; если

он попадает в диапазон 10-20%, качество соответствия модели хорошее. В диапазоне от 20% до 30% уровень ошибки является приемлемым, в то время как ошибка, превышающая 30%, считается плохим результатом, и поэтому модель должна быть отклонена. По утверждению Kim, S.; Kim, H. [8], когда реальное значение близко или равно нулю, MAPE предоставляет бесконечные или неопределенные значения, что считается его существенным недостатком. Поэтому для детальной проверки оценки эффективности модели прогнозирования используется комбинация MAPE и среднеквадратичной ошибки (RMSE). RMSE указывает на абсолютное соответствие наблюдаемой точки данных прогнозируемым значениям. RMSE определяется согласно формуле (2) как корень второго порядка из среднего квадрата всех ошибок. Чем ниже значение MAPE и RMSE, тем выше точность полученной прогнозной модели. Другие меры для определения ошибки включают среднюю абсолютную ошибку (MAE), среднеквадратичную ошибку (MSE), относительную абсолютную ошибку (RAE), относительную квадратную ошибку корня (RRSE) и другие.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - y_i')^2}{n}}. \quad (2)$$

В исследовании использовалась стандартная полностью связанная нейронная сеть с прямой связью и обратным распространением ошибки (BPNN). В качестве входных данных использова-

лась сигмовидная функция. Архитектура BPNN, использованного в этом исследовании, были следующими:

- количество слоев = 3 (входной, скрытый и выходной);
- количество нейронов в скрытом слое = от 3 до 15;
- тип функций активации = сигмоида для скрытого слоя, линейный для выходного слоя;
- количество узлов во входном слое = 16 (таблица 1);
- количество узлов в выходном слое = 1;
- тип сетевой ошибки = среднеквадратичная ошибка.

Для построения и обучения нейронной сети использовался Neural Network Toolbox программной среды MatLab. При этом использовался метод проб и ошибок, чтобы выбрать оптимальный параметр значения (веса соединений), которые дадут наиболее точные результаты. На рисунке 2 изображена модель нейронной сети с прямой связью и обратным распространением (BPNN).

На входе нейросети установлены 16 нейронов, представляющие входные данные. В качестве входных данных использованы 3 группы факторов

(таблица 2). На вход скрытого слоя (Hidden Layer) подаются сигналы, после чего они умножаются на веса «w» (каждый сигнал – на свой вес). К данной сумме прибавляется смещение нейрона «b» и затем поступают на суммирующий блок. Суммирующий блок алгебраически складывает взвешенные входы, создавая выход. Полученный сигнал преобразуется активационной функцией нейрона, которая формирует выходной сигнал. Функцией активации является сигмоида, т.к. диапазон значений наших данных [0;1].

Научные результаты

Для обучения разработанной нейронной сети использовались данные урожайности с 2007 по 2016 годы. Таким образом, использовалась выборка для 10 итераций и 14 эпох. При этом созданная нейронная сеть обучается достаточно быстро. В зависимости от мощности компьютера время обучения составляет от 10 до 25 секунд.

Тестирование уже обученной нейронной сети заключалось в способности предсказывать урожайность в годы с 2017 по 2020. Анализ полученных результатов прогнозирования осуществлялся путем подсчета средней абсолютной ошибки в

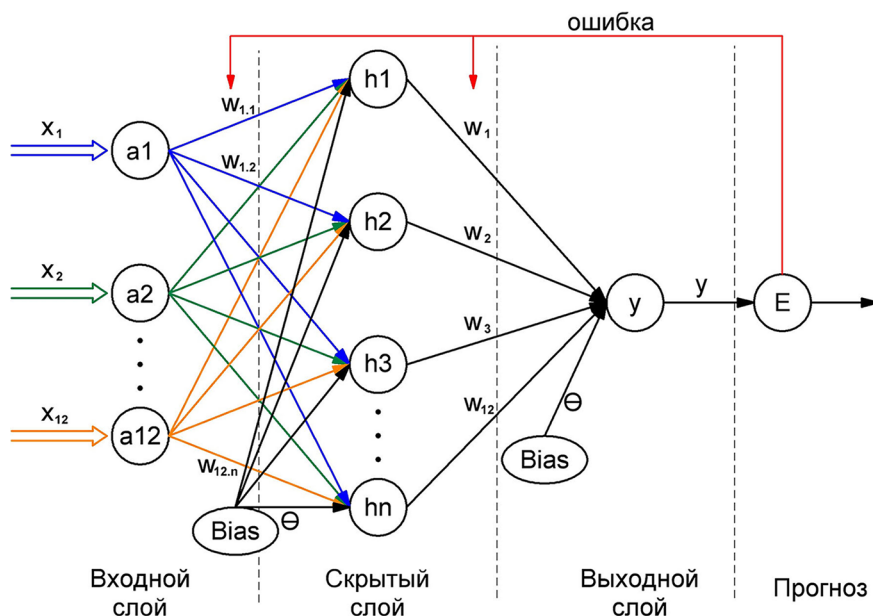


Рисунок 1 – Структура применяемой нейронной сети обратного распространения

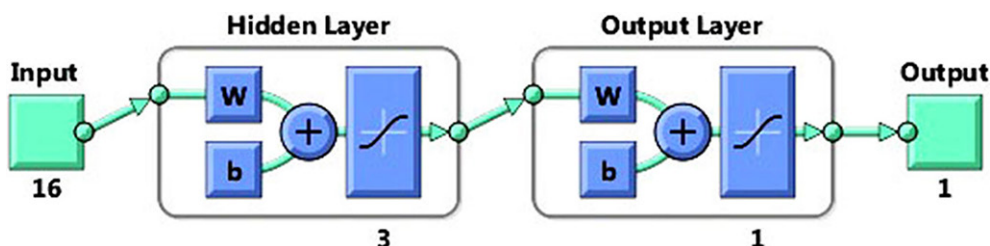


Рисунок 2 – Модель ИНС в среде MatLab

процентах MAPE и среднеквадратической ошибки RMSE. Полученные в результате расчета ошибки и коэффициента корреляции между прогнозируемой нейронной сетью урожайностью и фактическими данными представлены в таблице 3.

Для наглядности также рассчитаны уравнения линейной регрессии для результатов тестирования. На рисунке 3 показаны результаты прогнозируемой урожайности в сравнении с фактической. Данные, использованные для обучения, обозначены треугольником, а тестирования – кружком.

Полученные результаты показывают, что разработанная нейронная сеть успешно освоила основные зависимости урожайности от входных данных. При тестовом прогнозировании на созданной нейронной сети получается прогноз урожайности с ошибкой MAPE не более 12%, что является подтверждением адекватности выдаваемого прогноза.

Большое количество учитываемых факторов несильно перегружает саму систему прогноза, но в то же время позволило добиться высокой точ-

ности прогнозирования. Полученная величина MAPE в сравнении с нейронными сетями, представленными в [9], также позволяет считать результаты достоверными.

Выводы

В настоящем исследовании разработана система прогнозирования урожайности пшеницы в Северо-Казахстанской области на основе искусственной нейронной сети с использованием нескольких источников данных, таких как метеорологические, топографические данные, данные плодородия почвы и наличия вредителей, болезней и сорняков. Данные для системы взяты по Северо-Казахстанской области. В настоящем исследовании достигнуты удовлетворительные результаты для прогнозирования урожайности пшеницы ($p = 0,52$, MAPE = 12,02%, RMSE = 3,368). В дальнейшем исследовательская группа планирует разработку беспроводной системы удаленного мониторинга температуры воздуха и почвы, влажности почвы на разной глубине и кислотности непосредственно на посевном поле.

Таблица 3 – Результаты анализа производительности разработанной НС

Этап	Годы	Итерации	Ошибка		Коэффициент корреляции
			MAPE	RMSE	
Обучение	2007-2016	60	8,76	2,241	0,689
Тестирование	2017-2020	10	12,02	3,368	0,534

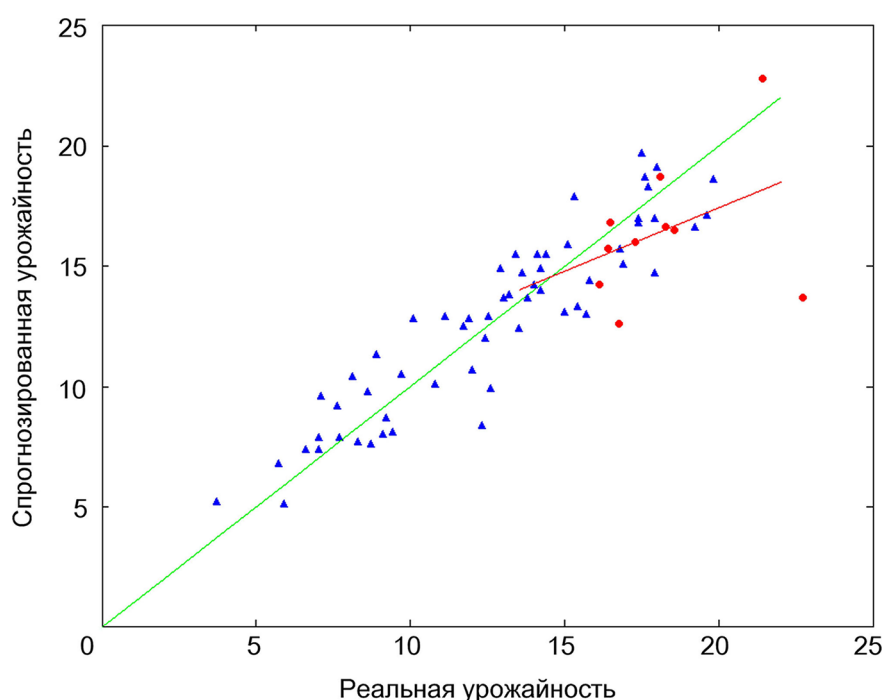


Рисунок 3 – Результаты прогнозирования урожайности ИНС в сравнении с фактическими данными

СПИСОК ЛИТЕРАТУРЫ

- Phalan, B., Green, R., & Balmford, A. (2014). Closing yield gaps: perils and possibilities for biodiversity conservation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369.
- Nandram, B., Berg, E., Barboza, W., 2014. A hierarchical Bayesian model for forecasting state-level corn yield. *Environ. Ecol. Stat.* 21, 507-530.
- Johnson, D.M., 2014. An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. *Remote Sens. Environ.* 141, 116-128.
- Агентство Республики Казахстан по статистике, статистический сборник. Портрет села. – Астана, 2011. – 92 с.
- Сайт Бюро национальной статистики Агентства по стратегическому планированию и реформам Республики Казахстан (https://stat.gov.kz/region/264023/statistical_information/industry/5914).
- Khoshnevisan, B.; Rafiee, S.; Omid, M.; Mousazadeh, H. Development of an intelligent system based on ANFIS for predicting wheat grain yield on the basis of energy inputs. *Inf. Process. Agric.* 2014, 1, 14-22.
- Peng, J.; Kim, M.; Kim, Y.; Jo, M.; Kim, B.; Sung, K.; Lv, S. Constructing Italian ryegrass yield prediction model based on climatic data by locations in South Korea. *Grassl. Sci.* 2017, 63, 184-195.
- Kim, S.; Kim, H. A new metric of absolute percentage error for intermittent demand forecasts. *Int. J. Forecast.* 2016, 32, 669-679.
- Schwalbert, R.A.; Amado, T.; Corassa, G.; Pott, L.P.; Prasad, P.V.V.; Ciampitti, I.A. Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. *Agric. For. Meteorol.* 2020, 284, 107886.

Бидай өнімділігін болжау үшін жасанды нейрондық желіні қолдану

¹*ӘУБАКИРОВА Гауля Фаритқызы, аға оқытушы, geli8778@mail.ru,

¹МОЛДАХМЕТОВ Саят Сайранұлы, PhD, доцент, sayatmoldakhmetov@gmail.com,

¹ЗЫКОВА Наталья Владимировна, аға оқытушы, zykovanata1978@gmail.com,

¹ГАГОЛИНА Ольга Сергеевна, аға оқытушы, olenka-russ@mail.ru,

¹ДАРИЙ Елена Михайловна, аға оқытушы, dariyelena@mail.ru,

¹«Манаш Қозыбаев атындағы Солтүстік Қазақстан университеті» КеАҚ, Қазақстан, Петропавл, Пушкин көшесі, 86,

*автор-корреспондент.

Аңдатпа. Жасанды нейрондық желіге (ЖНЖ) негізделген бидай өнімділігін болжау нәтижелері келтірілген. Жасанды нейрондық желіні қолдану үшін ауа температурасы мен жауын-шашын сияқты сызықтық емес факторларды болжаушы ретінде қолдануға байланысты. Зерттеуде бидай өнімділігіне ең көп әсер ететін метеорологиялық, агрохимиялық және фитосанитарлық факторларды қолданды. Бұл факторлар қатенің кері таралуының нейрондық желісінің кіріс айнымалылары ретінде пайдаланылды. ЖНЖ оқытудан және болжамды деректерді алғаннан кейін бидай өнімділігін болжау үшін жақсы нәтижелерге қол жеткізілді ($p = 0,52$, $MAPE = 12,02\%$, $RMSE = 3,368$).

Кілт сөздер: астық шығымдылығын болжау, жасанды нейрондық желі, бидай өнімділігі, BPNN, ауылшаруашылық дақылдарының маусымдық өнімділігін болжау, тәуелсіз айнымалылар, нейрондық желі моделі, өнімділікті болжаушылар, кері таралудың нейрондық желісі.

Application of Artificial Neural Network for Wheat Yield Prediction

¹*AUBAKIROVA Gailya, Senior Lecturer, geli8778@mail.ru,

¹MOLDAKHMETOV Sayat, PhD, Associate Professor, sayatmoldakhmetov@gmail.com,

¹ZYKOVA Natalia, Senior Lecturer, zykovanata1978@gmail.com,

¹GAGOLINA Olga, Senior Lecturer, olenka-russ@mail.ru,

¹DARIY Elena, Senior Lecturer, dariyelena@mail.ru,

¹NPLC «Manash Kozybayev North Kazakhstan University», Kazakhstan, Petropavl, Pushkin Street, 86,

*corresponding author.

Abstract. The results of wheat yield forecasting based on an artificial neural network (ANN) are presented. The use of ANN is due to the use of such non-linear factors as air temperature and precipitation as predictors. The study also used meteorological, agrochemical and phytosanitary factors that have the greatest impact on wheat yield. These factors were used as input variables to the backpropagation neural network. After training the ANN and obtaining predictive data, good results were achieved for wheat yield prediction ($p = 0.52$, $MAPE = 12.02\%$, $RMSE = 3.368$). The developed model can be easily adapted to other cultures and locations, and can also be improved by using new data sources.

Keywords: yield forecasting, artificial neural network, wheat yield, BPNN, seasonal crop yield forecasting, independent variables, neural network model, yield predictors, Back Propagation Neural Network.

REFERENCES

1. Phalan, B., Green, R., & Balmford, A. (2014). Closing yield gaps: perils and possibilities for biodiversity conservation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369 (1639), 20120285.
2. Nandram, B., Berg, E., Barboza, W., 2014. A hierarchical Bayesian model for forecasting state-level corn yield. *Environ. Ecol. Stat.* 21, 507-530.
3. Johnson, D.M., 2014. An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. *Remote Sens. Environ.* 141, 116-128.
4. Agentstvo Respubliki Kazakhstan po statistike. Portret sela (2011). Astana, 92. Available at: <https://stat.gov.kz/api/getFile/?docId=WC16200032726>
5. Bureau of National Statistics of the Agency for Strategic Planning and Reforms of the Republic of Kazakhstan. Available at: <https://www.gov.kz/memleket/entities/stat?lang=ru>
6. Khoshnevisan, B.; Rafiee, S.; Omid, M.; Mousazadeh, H. Development of an intelligent system based on ANFIS for predicting wheat grain yield on the basis of energy inputs. *Inf. Process. Agric.* 2014, 1, 14-22.
7. Peng, J.; Kim, M.; Kim, Y.; Jo, M.; Kim, B.; Sung, K.; Lv, S. Constructing Italian ryegrass yield prediction model based on climatic data by locations in South Korea. *Grassl. Sci.* 2017, 63, 184-195.
8. Kim, S.; Kim, H. A new metric of absolute percentage error for intermittent demand forecasts. *Int. J. Forecast.* 2016, 32, 669-679.
9. Schwalbert, R.A.; Amado, T.; Corassa, G.; Pott, L.P.; Prasad, P.V.V.; Ciampitti, I.A. Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. *Agric. For. Meteorol.* 2020, 284, 107886.