

МОДЕЛИРОВАНИЕ ФЕНОЛОГИИ СОИ С ПОМОЩЬЮ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

© 2019 г. О.Д. Таратухин*, Л.Ю. Новикова* **, И.В. Сеферова**, К.Н. Козлов*

*Санкт-Петербургский политехнический университет Петра Великого, 195251, Санкт-Петербург, Политехническая ул., 29

**Федеральный исследовательский центр «Всероссийский институт генетических ресурсов растений имени Н.И. Вавилова», 190000, Санкт-Петербург, ул. Большая Морская, 42–44

E-mail: kozlov_kn@spbstu.ru

Поступила в редакцию 11.02.2019 г.

После доработки 11.02.2019 г.

Принята к публикации 25.02.2019 г.

Индикатором климатических изменений и степени пригодности сои к климатическим условиям является ее фенология, основные регуляторы которой – температура и длина дня. Для предсказания времени до цветения у девяти скороспелых образцов сои с пониженной фотопериодической чувствительностью в условиях Северо-Западного региона России была адаптирована модель на основе искусственной нейронной сети. Введены масштабирующие множители для сетевых входов, оптимизированы верхняя и нижняя температурные границы и базовая длина дня, создана программа на языке Питон с использованием Керас и ТензорФлоу. Экспериментальные данные, полученные в 1999–2013 гг. в Пушкине (Ленинградская область) и на Кубани были использованы для обучения 121 параметра модели, по завершению которого среднеквадратичная ошибка снизилась до 0,026. Образцы характеризовались пониженной верхней температурной границей по сравнению с литературными данными (23°C вместо 30°C) и повышенной нижней температурной границей (12°C вместо 5°C). Увеличение базовой длины дня с 12 до 13 часов подтвердило адаптированность образцов к более длинному дню. Средняя ошибка прогноза улучшилась примерно на двое суток по сравнению с предыдущей моделью температурных минимумов. Были сгенерированы ежедневные погодные данные для различных будущих сценариев выброса парниковых газов и предсказаны времена до цветения девяти образцов сои в условиях изменения климата в период с 2019 г. до 2030 г. для посевов 1 и 10 мая. Предсказанные значения уменьшаются до 2030 г. для большинства образцов и сценариев, но могут оставаться постоянными или в некоторых случаях испытывать колебания. Различие средних между 39,21 суток в данных и 36,33 суток в результатах моделирования для 2030 г. является статистически значимым согласно одностороннему тесту Манна–Уитни–Уилкоксона (5423,5, $P = 0,0097 < 0,01$). Таким образом, результаты подтвердили предсказательную силу разработанной модели.

Ключевые слова: климатические факторы, соя, искусственные нейронные сети.

DOI: 10.1134/S0006302919030177

Одной из мер минимизации потерь от потепления климата и превращения их в прибыль является осеверение сортов [1]. В условиях глобальных и региональных климатических трендов научной основой для изучения механизмов адаптации растений к изменению условий обитания является математическое моделирование и прогнозирование [2,3]. Индикатором климатических изменений и степени пригодности сортов к складывающимся климатическим условиям является фенология растения [4].

Соя является одной из наиболее рентабельных с/х культур, потребность в продуктах переработки сои продолжает расти. Соя – растение южных

областей умеренных широт. Основными регуляторами ее фенологии являются температуры и длина дня. Продвижению культуры на север мешают низкие температуры и длинный день [5]. С 2000-х годов посевы под соей в России неуклонно расширяются (Госкомстат России). В ФИЦ «Всероссийский институт генетических ресурсов растений имени Н.И. Вавилова» (ВИР) много лет проводятся опыты по осеверению сои, выводятся сорта с пониженной фотопериодической чувствительностью и пониженными температурными потребностями.

Разработанные в условиях последней четверти XX века модели нуждаются в параметризации

Таблица 1. Сравнение точности новой модели с моделью температурных минимумов [11]

Образец	Средняя ошибка прогноза в модели температурных минимумов, сут	Средняя ошибка прогноза новой модели, сут	Коэффициент корреляции Пирсона	R^2 – коэффициент детерминации
ПЭП 28	4,0	2,4	0,87	0,75
Светлая	3,9	2,9	0,87	0,75
ПЭП 2	4,1	1,8	0,92	0,84
ПЭП 18	5,9	1,9	0,91	0,84
ПЭП 27	3,9	2,5	0,88	0,78
Fiskeby 1040-4-2	3,6	2,0	0,94	0,87
Окская	5,6	2,9	0,86	0,74
Алтом	8,0	3,1	0,83	0,70
KG-20	7,4	2,5	0,94	0,89
Среднее значение	5,2	3,3	0,82	0,69

применительно к новым условиям. Существует множество моделей фенологии сои. Математическое моделирование важных агрономических показателей, таких как, например, длина периода «всходы-цветение», может осуществляться различными способами. Строятся регрессионные модели с различными наборами предикторов [6]. Наиболее распространенный – расчет суммы накопленных «единиц тепла» за сутки. Число дней между фазами развития находится суммированием вклада каждого дня до достижения суммы, необходимой для завершения фазы. Скорость развития за день рассчитывается как произведение функций в основном от длины дня и средней суточной температуры воздуха [7–10]. Для исследования развития строятся динамические и имитационные модели [11].

Искусственные нейронные сети широко применяются в машинном обучении в различных областях, в том числе для предсказания важных агрономических показателей [12–14]. Модели на основе нейронных сетей основываются на концепции слоев и перцептронов. Нейронная сеть состоит из входного слоя, произвольного числа промежуточных слоев и выходного слоя. Размерности и количество скрытых слоев определяются исследователем, который учитывает объем имеющихся данных и их качество. Модели на основе нейронных сетей используются и для исследования фенологии [15].

Целью данной работы было исследование применимости модели нейронных сетей к моделированию фенологии сои в условиях низких температур и длинного дня Северо-Западного региона России.

ОБЪЕКТ ИССЛЕДОВАНИЯ И ЭКСПЕРИМЕНТАЛЬНЫЕ ДАННЫЕ

Были исследованы девять скороспелых образцов сои с пониженной фотопериодической чувствительностью, перспективные для продвижения в северные регионы (таблица). В качестве исходных были использованы данные по длине периода «всходы–цветение» у этих сортов, полученные на опытных станциях ВИР в г. Пушкин (Ленинградская область) и на Кубани в 1999–2013 гг. Проводились опыты с несколькими посевами в течение года (от 2 до 15), что позволило расширить диапазон условий температуры и фотопериода.

На рис. 1 представлена гистограмма длин периода в наборе данных.

МОДЕЛЬ ВРЕМЕНИ ЦВЕТЕНИЯ НА ОСНОВЕ НЕЙРОННЫХ СЕТЕЙ

В работе [15] предложена модель на основе нейронной сети, которая состоит из 4 входных нейронов, 20 перцептронов в единственном

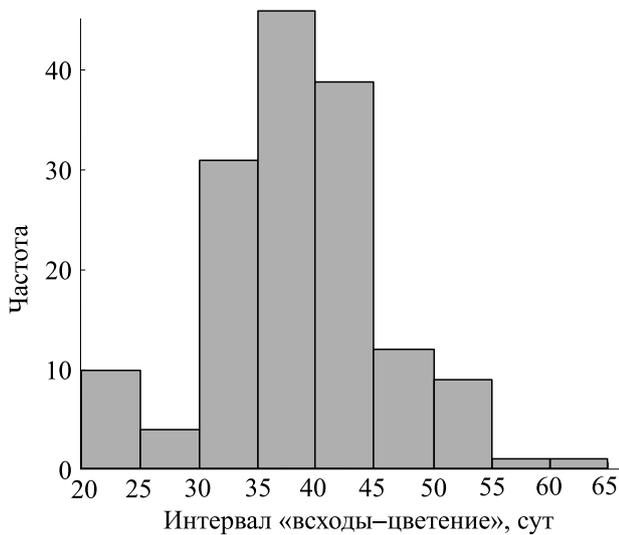


Рис. 1. Гистограмма длин периода «всходы–цветение» в наборе данных.

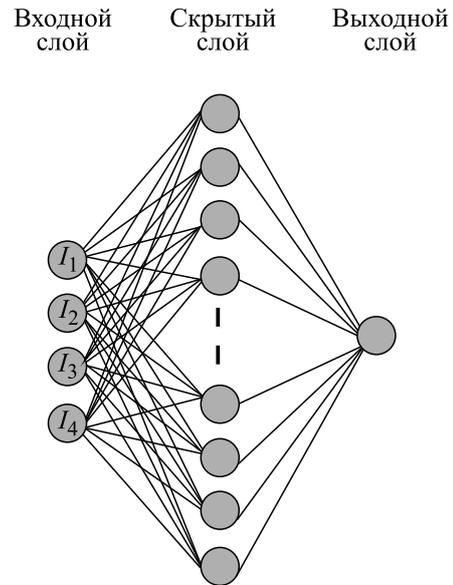


Рис. 2. Нейронная сеть модели времени цветения.

скрытом слое и единственного выходного нейрона (см. рис. 2).

На входы сети I_1, I_2, I_3, I_4 подаются результаты применения специально подобранных в работе [15] функций (1)–(4), которые обобщают влияние ежедневных показателей максимальной и минимальной температуры и длины светового дня на состояние растения:

$$I_1 = \frac{\sum H(T_{m,i} - T_{m,c})(T_{m,i} - T_{m,c})}{\mu_m}, \quad (1)$$

$$I_2 = \frac{\sum H(T_{x,i} - T_{x,c})(T_{x,i} - T_{x,c})}{\mu_x}. \quad (2)$$

$$I_3 = \frac{\sum H(g_i - g_c)(g_i - g_c)}{\mu_g} \quad (3)$$

$$I_4 = \frac{D}{\mu_d}, \quad (4)$$

где H – функция Хевисайда, $T_{m,c}$ – нижняя граница роста, $T_{x,c}$ – верхняя, g_c – базовая длина светового дня, $T_{m,i}, T_{x,i}, g_i$ – минимальная, максимальная температура и длина светового дня за день i , а D – номер дня с даты всходов. По сравнению с исходной моделью введены масштабные константы μ_m, μ_x, μ_g и μ_d , чтобы не давать заранее больший вес одному из параметров, но при этом не вычислять эти константы для каждого эксперимента и каждого растения.

Конкретные значения параметров должны быть выяснены в ходе обучения сети для достижения наилучших результатов.

РЕЗУЛЬТАТЫ АДАПТАЦИИ МОДЕЛИ

Модель была запрограммирована на языке Питон с использованием Keras [16] и TensorFlow [17] в качестве вычислителя. Данные о ежедневных максимальных и минимальных температурах, осадках и длине дня были взяты в открытом доступе в сети Интернет. Обучение сети состоит в определении 121 параметра. На рис. 3 представлен график изменения значения функции потерь при тренировке нейронной сети. Минимальное значение среднеквадратичной ошибки составило 0,026.

В модели [15] для условий, близких к оптимальным для возделывания сои, был получен оптимальный набор параметров: $T_{m,c} = 5, T_{x,c} = 30$ и $g_c = 12$. В новых условиях был проведен новый подбор параметров путем перебора с шагом 1 около предыдущих значений. Полученные для условий Северо-Запада России значения $T_{m,c} = 12, T_{x,c} = 23$ и $g_c = 13$ показывают, что исследованные сорта характеризуются более низкой верхней границей температуры роста (23°C вместо 30°C). Температурные условия начала вегетации (12°C) в регионе определяются началом весенних полевых работ при температурах 10–15°C, связанным с избыточной влажностью почвы, что и опреде-

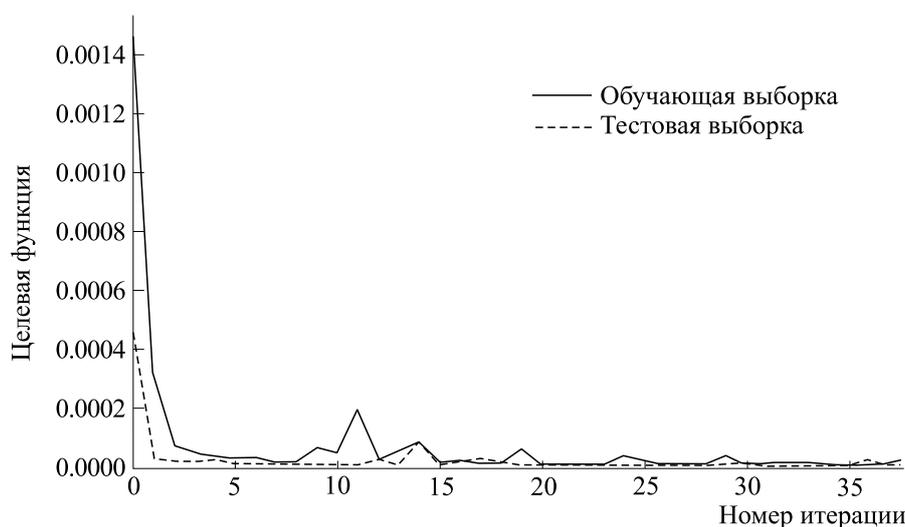


Рис. 3. График изменения значения функции потерь.

лило нижнюю границу температурной зоны роста исследованных образцов. Параметр фотопериодической чувствительности g_c подтвердил адаптированность образцов к более длинному дню (13 ч вместо 12 ч).

Для масштабных констант были выбраны значения $\mu_m = 80$, $\mu_x = 80$, $\mu_g = 80$ и $\mu_d = 36$.

На рис. 4 показано сравнение данных и решения модели. Видно, что модель достаточно точно описывает экспериментальные данные.

В таблице приведено значений сравнение средней ошибки прогноза, коэффициента корреляции и детерминации для новой модели и предыдущей модели температурных минимумов [11] для девяти скороспелых сортов сои. Видно, что точность новой модели выше, а средняя ошибка прогноза улучшилась примерно на двое суток.

РЕЗУЛЬТАТЫ ИСПОЛЬЗОВАНИЯ МОДЕЛИ

Актуальной проблемой соеводства являются ранние и сверхранние посевы [18], позволяющие растениям более эффективно использовать осенне-зимние запасы влаги. Проблема ранних посевов сои – замедление развития при более длинном дне, что сказывается даже на сортах с пониженной фотопериодической чувствительностью, и вероятность заморозков. По модели была исследована возможность более раннего посева в условиях потепления климата. Средняя температура посевов в Ленинградской области составила $16,1^\circ\text{C}$. При потеплении климата дата устойчиво-

го перехода температур выше необходимого для активного роста сои предела 15°C смещалась на более ранние сроки на 1,0 сут/год. В 2000–2013 гг. эта дата была 27 мая, к 2030 г. можно прогнозировать ее наступление в десятых числах мая.

С помощью генератора погоды – программы МаркСим – были построены прогнозы погоды на каждый день для Ленинградской области с 2019 до 2030 г. MarkSim был разработан для моделирования погоды из известных источников ежемесячных климатических данных [19–23]. Кроме моделей для прогнозирования климата следует учесть социоэкономические сценарии развития, которые описываются четырьмя характерными профилями концентрации углекислого газа (Representative Concentration Pathways, RCPs), принятые Межправительственной группой экспертов по изменению климата (Intergovernmental Panel on Climate Change, IPCC) для пятого оценочного отчета (Assessment Report, AR5) в 2014. Профили соответствуют широкому кругу возможных изменений будущих антропогенных выбросов парниковых газов и называются rcp26, rcp45, rcp60 и rcp85 в соответствии с возможными значениями нарушения радиационного баланса Земли в 2100 г. относительно преиндустриальной эры (+2,6, +4,5, +6,0 и +8,5 Вт/м² соответственно) [24].

Созданные погодные данные и построенная модель были использованы для предсказания длины периода «всходы–цветение» для девяти скороспелых сортов сои в Ленинградской области в 2019–2030 гг.

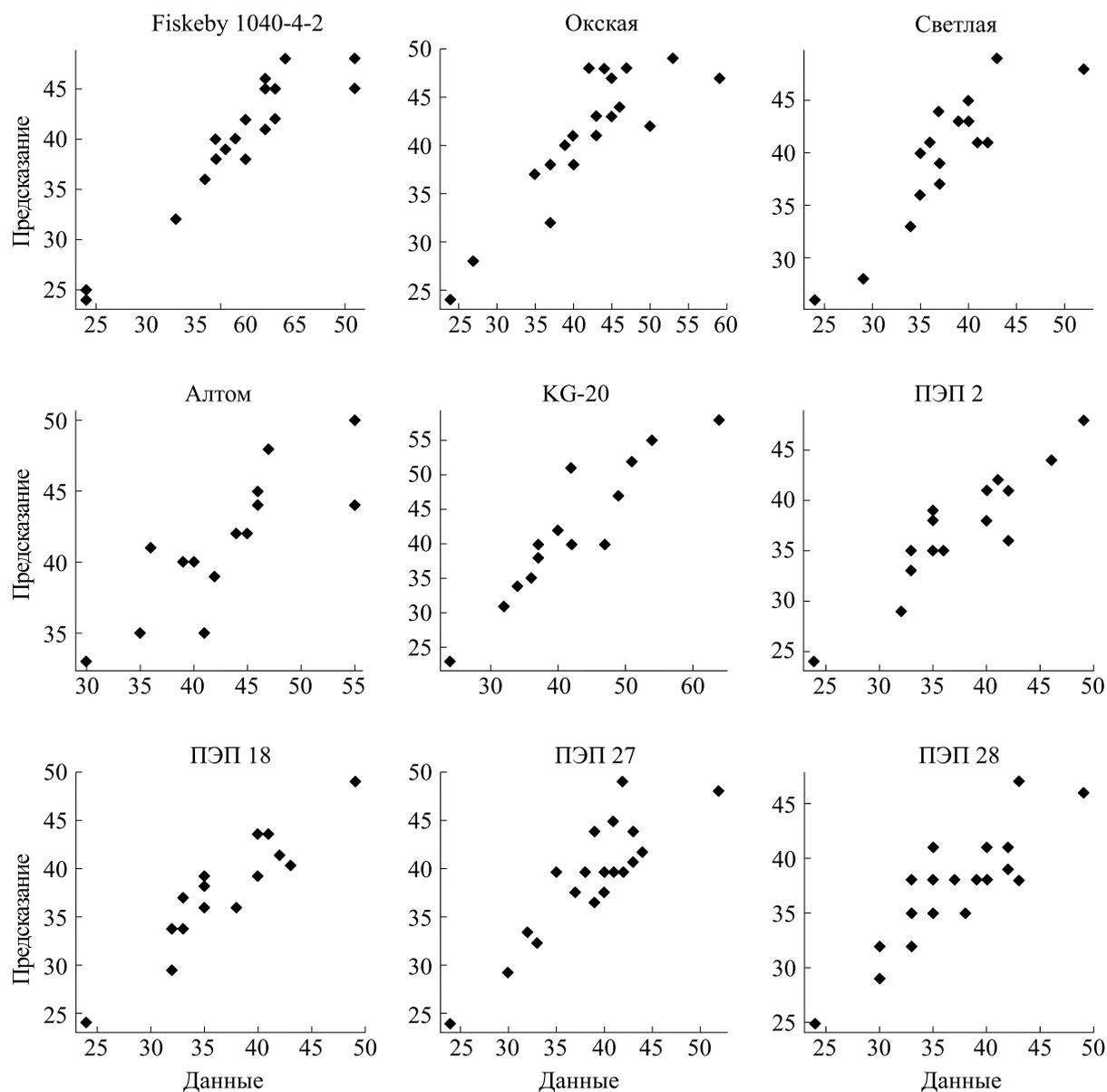


Рис. 4. Сравнение данных и решения модели.

На рис. 5 представлены результаты модельного исследования посева 1 мая. Результаты моделирования показывают, что время до цветения сильно сокращается по сравнению с годами проведения экспериментов. Для некоторых сортов (Fiskeby 1040-4-2, Окская, Светлая, Алтом, KG-20, ПЭП2) время до цветения будет заметно сокращаться в период до 2030 г. Для образцов ПЭП 27, ПЭП 28, ПЭП 18 возможен вариант, что время до цветения существенно не изменится. Именно эти образцы (и сорт Светлая) показали наименьшую разницу в продолжительности пе-

риода «всходы—цветение» в контрастных эколого-географических условиях г. Пушкина и Кубани, т. е. были наиболее стабильны.

Были модельно исследованы посева 10 мая в условиях потепления (см. рис. 6). Результаты моделирования показывают, что время до цветения сильно сокращается по сравнению с годами проведения экспериментов. Для некоторых сортов (ПЭП 27 и ПЭП 28) время до цветения будет заметно сокращаться в период до 2030 г. Для остальных сортов (Fiskeby 1040-4-2, Окская, Светлая, Алтом, KG-20, ПЭП 2 и ПЭП 18) время

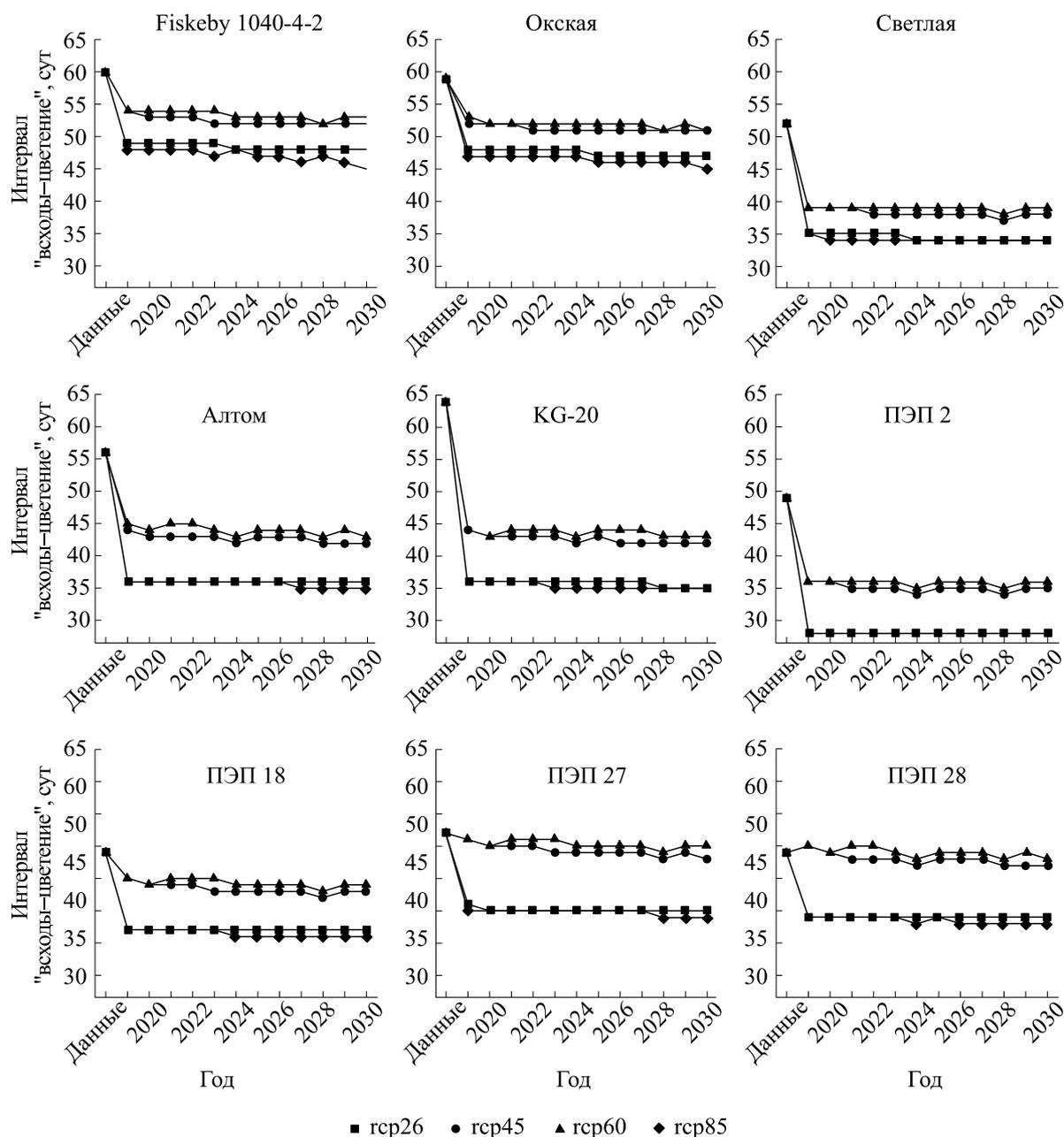


Рис. 5. Предсказание интервала «всходы—цветение» для даты посева 1 мая.

до цветения будет оставаться постоянным или незначительно колебаться.

Среднее значение длины периода «всходы—цветение» составило для экспериментальных данных 39,21 суток, что статистически значимо больше, чем среднее значение 36,33 суток для результатов моделирования исследуемых образцов в 2030 г. Значение одностороннего теста Манна—Уитни—Уилкоксона составило 5423,5, p -значение = 0,009669 < 0,01. Рас-

пределение модельных результатов представлено на гистограмме (рис. 7).

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Математическое моделирование хозяйственно ценных показателей растения, таких как длина периода «всходы—цветение» сои, является научной основой для выведения новых сортов культуры в условиях глобальных изменений климата. В

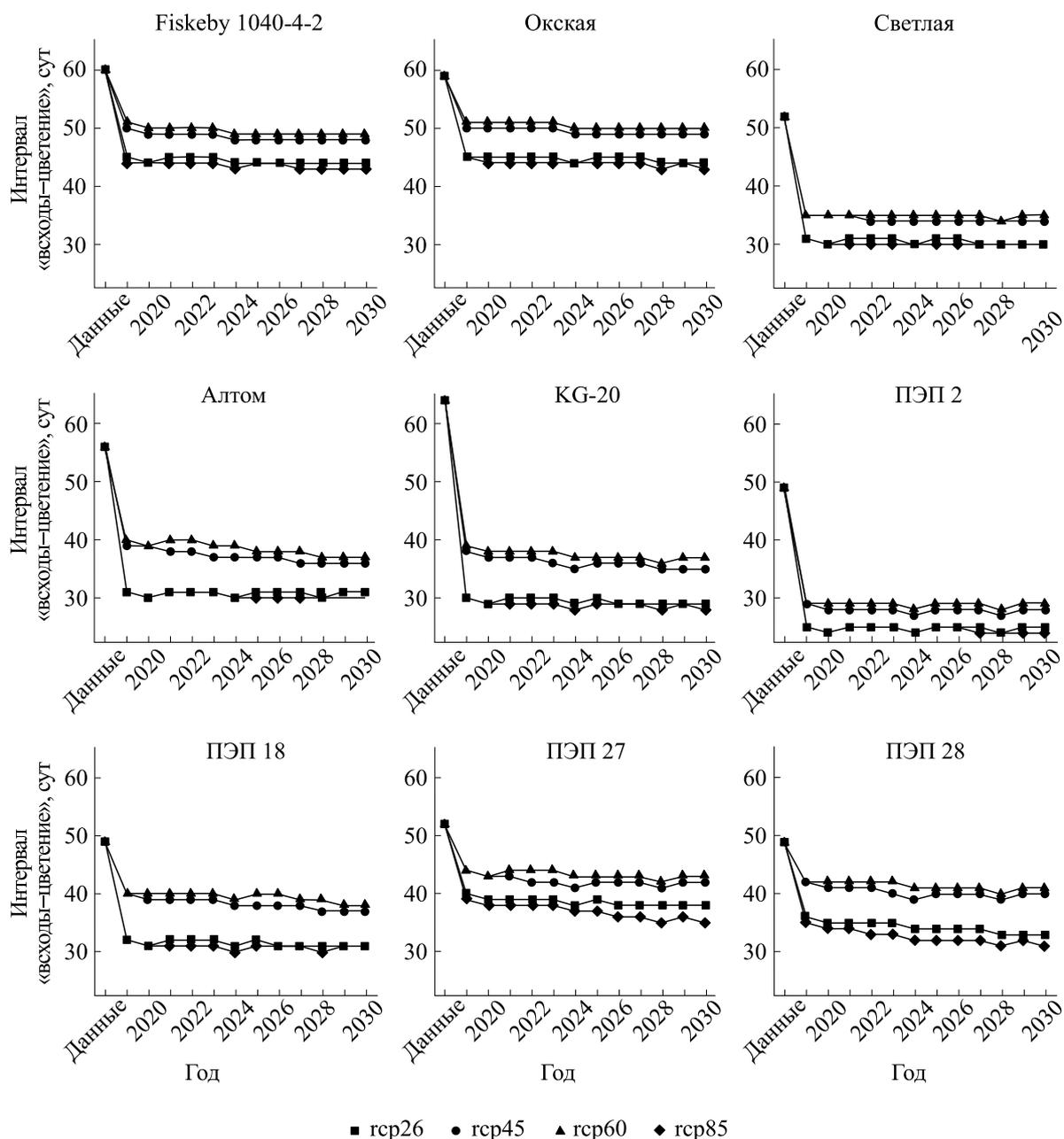


Рис. 6. Предсказание интервала «всходы—цветение» для даты посева 10 мая.

данной работе была построена модель предсказания времени до цветения девяти скороспелых сортов сои в условиях Северо-Запада России. Новая модель использует специально подобранные функции погодных данных и нейронную сеть. Расчеты по имеющимся экспериментальным данным показали, что новая модель имеет более высокую точность по сравнению с моделью температурных минимумов. Повышение точности может объясняться как более сложной структурой

модели, так и большим количеством свободных параметров.

Модельные прогнозы показали перспективность сверхранных посевов. Для исследованной группы сортов фактор температур был более важным, чем фактор длины дня, время цветения смещалось на более ранние сроки, обеспечивая условия более продолжительного и полного созревания.

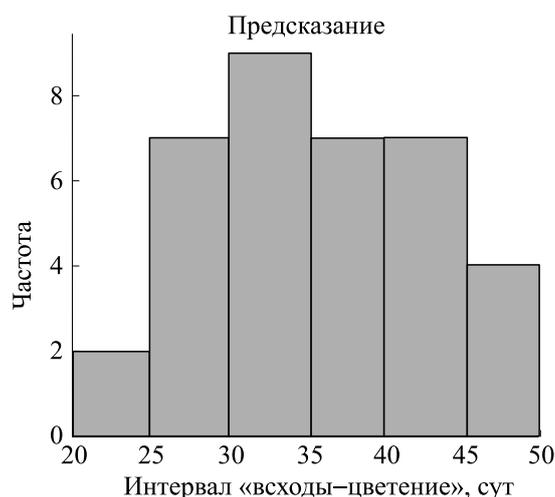


Рис. 7. Гистограмма длин периода «всходы-цветение» в результатах моделирования.

Таким образом, показана пригодность модели фенологии сои, созданной методом искусственных нейронных сетей, для моделирования в нетипичных для культуры условиях.

Работа выполнена в рамках Федеральной целевой программы (проект №14.575.21.0136 от 26.09.2017 г., уникальный идентификатор проекта RFMEFI57517X0136). Вычисления были проведены в Суперкомпьютерном центре «Политехнический» СПбПУ и кластере Университета Южной Калифорнии.

Исходные данные получены на базе уникальной научной установки «Коллекция генетических ресурсов растений» ВИР и любезно предоставлены И.В. Сеферовой.

СПИСОК ЛИТЕРАТУРЫ

1. А. Г. Папцов и др., *Адаптация сельского хозяйства России к глобальным изменениям климата* // Проект независимой международной организации Оксфам (2015). URL: <http://www.oxfam.ru/upload/iblock/f96/f9622b41f48549945438f2292f509d14.pdf> [Электронный ресурс, режим доступа: 06.01.2019].
2. J. E. Olesen, et al., *Eur. J. Agronomy* **34** (2), 96 (2011).
3. T. Carter and K. Mäkinen, *Approaches to climate change impact, adaptation and vulnerability assessment: towards a classification framework to serve decision-making: MEDIATION 2.1* (Finnish Environment Institute (SYKE), Helsinki, Finland: 2011).
4. A. D. Richardson, et al., *Global Change Biology* **18** (2), 566 (2012).
5. И. В. Сеферова, Т. В. Мисюрина и М. А. Никишкина, *Сельскохозяйственная биология*, № 5, 42 (2007).
6. К. Н. Козлов, Л. Ю. Новикова, И. В. Сеферова и М. Г. Самсонова, *Биофизика* **63** (1), 175 (2018).
7. D. J. Major, et al., *Crop Science* **15**, 174 (1975).
8. T. Hodges and V. French, *Agronomy J.* **77** (3), 500 (1985).
9. P. Pedersen, et al., *Agronomy J.* **96**, 556 (2004).
10. T. D. Setiyono, et al., *Field Crops Res.* **100** (2–3), 257 (2007).
11. Л. Ю. Новикова, И. В. Сеферова и К. Н. Козлов, *Биофизика* **63** (6), 1182 (2018).
12. M. Abdipour, et al., *J. Am. Oil Chemists' Soc.* **95** (3), 283 (2018).
13. M. Kaul, R. L. Hill, and C. Walthall, *Agricultural Systems* **85** (1), 1 (2005).
14. A. Bagherzadeh, et al., *Modeling Earth Systems and Environment* **2** (2), (2016).
15. D. A. Elizondo, R. W. McClendon, and G. Hoogenboom, *Trans. ASAE* **37** (3), 981 (1994).
16. F. Chollet, et al., *Keras* (GitHub, 2015). URL: <https://keras.io/> [Электронный ресурс, режим доступа: 06.01.2019 г.].
17. M. Abadi, et al., *TensorFlow: A system for large-scale machine learning* (2016), p. 21.
18. С. В. Зеленцов и Е. В. Мошенко, *Научный диалог*, № 7, 40 (2012).
19. M. Srinivasa Rao, et al., *PLoS One* **10** (2), e0116762 (2015).
20. P. G. Jones and P. K. Thornton, *Agricultural and Forest Meteorology* **86** (1–2), 127 (1997).
21. P. G. Jones and P. K. Thornton, *Agricultural and Forest Meteorology* **97** (3), 213 (1999).
22. P. G. Jones and P. K. Thornton, *Agronomy J.* **92**, 445 (2000).
23. P. G. Jones and A. L. Jones, *Centro Internacional de Agricultura Tropical. MarkSim: a computer tool that generates simulated weather data for crop modeling and risk assessment* (CIAT, 2002).
24. D. P. van Vuuren, et al., *Climatic Change* **109** (1–2), 5 (2011).

Simulation of Soybean Phenology with the Use of Artificial Neural Networks

O.D Taratuhin*, L.Yu. Novikova* **, I.V. Seferova**, and K.N. Kozlov*

*Peter the Great St. Petersburg Polytechnic University, ul. Polytekhnicheskaya 29, St. Petersburg, 195251 Russia

**Federal Research Center “N.I. Vavilov All-Russian Institute of Plant Genetic Resources”,
ul. Bolshaya Morskaya 42–44, St. Petersburg, 190000 Russia

Soybean’s phenology explicitly indicates environmental changes and strongly depends on temperature and day length. We adapted an artificial neural network model to predict time to flowering in nine early maturing soybean accessions in the North-West region of Russia. We added scaling constants for network inputs, optimized high and low temperature thresholds and base day length, and implemented a new model written in Python using Keras and TensorFlow libraries. Experimental data obtained in 1999–2013 in Pushkin (Leningrad region) and Kuban were used for training 121 model parameters and after training the mean-root-square error became smaller, 0.026. Investigated accessions had the reduced upper temperature threshold as opposed to literature data (23°C instead of 30°C) and increased low temperature threshold (12°C instead of 5°C). The extension of day length from 12h to 13h confirmed the adaptation to a longer day. The average prediction error was improved by approximately 2 days as compared to the previous model of temperature minima. We generated daily weather for different future greenhouse gas emission scenarios and predicted time to flowering for nine soybean accessions in changing climate for years 2019–2030 and two planting days – May 1 and May 10. Predicted time to flowering decreases to 2030 for most accessions and scenarios but may stay constant or fluctuate in several cases. The difference in mean between 39.21 days in experimental data and 36.33 days in modeling results for 2030 is statistically significant according to Mann–Witney–Wilcoxon criterion (5423.5, $P = 0.0097 < 0.01$). Consequently, the results confirmed the predictive power of the developed model.

Keywords: climatic factors, soybeans, artificial neural networks