

НЕЙРОСЕТЕВАЯ МОДЕЛЬ ПРОГНОЗИРОВАНИЯ ФЕНОЛОГИИ СКОРОСПЕЛЫХ СОРТОВ СОИ ПО КЛИМАТИЧЕСКИМ ФАКТОРАМ

© 2020 г. О.Д. Таратухин*, Л.Ю. Новикова*, **, И.В. Сеферова**, Герасимова Т.В.**, С.В. Нуждин*, ***, М.Г. Самсонова*, 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

***Университет Южной Калифорнии, СА 90089, Лос-Анжелес, США

E-mail: kozlov_kn@spbstu.ru

Поступила в редакцию 30.10.2019 г.

После доработки 30.10.2019 г.

Принята к публикации 01.11.2019 г.

Фенология сои сильно зависит от температуры и продолжительности светового дня и четко указывает на изменения климата. Авторами разработана модель, которая включает в себя три искусственные нейронные сети для прогнозирования временных интервалов между посевом, всходами, цветением и созреванием в зависимости климатических факторов, а также ансамбли регрессионных моделей для прогнозирования урожайности, содержания белка и масла в семенах. Проанализированы данные по созреванию скороспелых образцов сои, фенотипированных на двух опытных станциях ВИР имени Н.И. Вавилова в Северо-Кавказском и Северо-Западном регионах России. Модель реализована на языке Python с использованием библиотек Keras и TensorFlow.

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

DOI: 10.31857/S0006302920010159

Благодаря комплексу ценных свойств растений и зерна, соя (*Glycine max* (L.) Merr.) является одной из главнейших белково-масличных культур в мире. Высокое содержание белка (40–45%) и масла (20–25%) в семенах сои обеспечивает постоянный рост значения этой культуры в мировой экономике. Соя – растение короткого дня теплообеспеченных регионов умеренных широт, однако ее биологический потенциал позволяет выращивать сорта в широком диапазоне климатических условий. В условиях глобальных и региональных климатических трендов научной основой для изучения механизмов адаптации растений к изменению условий обитания является математическое моделирование и прогнозирование [1–3]. Индикатором климатических изменений и степени пригодности сортов к складывающимся климатическим условиям является фенология растения [4].

Математические модели хозяйственно ценных показателей, таких как, например, длина периодов «посев–всходы», «всходы–цветение», «цветение–созревание», могут строиться различными методами. Часто используются регрессионные модели с различными наборами факторов [5].

Распространенным подходом является расчет суммы накопленных «единиц тепла» в сутки. Продолжительность периода между фазами развития определяется нахождением суммы вкладов всех дней до достижения заданной величины, необходимой для завершения фазы. Вклад за день часто принимается равным произведению функций длины дня и средней суточной температуры воздуха [6–9]. Динамические и имитационные модели [10] позволяют исследовать развитие растения.

Искусственные нейронные сети являются одним из широко применяемых методов машинного анализа, который используется в различных областях, в том числе для описания хозяйственно-ценных признаков растений [11–13]. Искусственные нейронные сети основываются на концепции слоев искусственных нейронов. Исходные данные подаются в нейронную сеть на входной слой, с которого поступают на произвольное число промежуточных слоев, а результат считывается с выходного слоя. Такие параметры, как размерность и количество скрытых слоев, должны быть заданы с учетом объема имеющихся экспериментальных данных. Модели на основе

Таблица 1. Группировка сортов по областям селекции

Группа	Сорта
ЕС	Sito, Major
Канада	Maple Ridge, KG-20
Север	ПЭП 26, ПЭП 27, ПЭП 28, ПЭП 2, ПЭП 17, ПЭП 18
Сибирь	СибНИИСХ 6, Степная 90, СибНИИК 15/83, Алтом, Степная 85
Швеция	Fiskeby 840-2-7, Fiskeby 840-7-3, Fiskeby V, Fiskeby 1040-4-2, Bravalla
Центр	Соер 13-91, Мадева, Восход*1191/79, Белор, Магева, УСХИ 6, Соер 4, Окская, Светлая

нейронных сетей используются и для исследования фенологических показателей [14].

ОБЪЕКТ ИССЛЕДОВАНИЯ И ЭКСПЕРИМЕНТАЛЬНЫЕ ДАННЫЕ

Сбор фенотипических данных, показателей продуктивности и качества семян проводили в ФИЦ «Всероссийский институт генетических ресурсов растений им. Н.И. Вавилова» (ВИР) в двух географических районах — в Санкт-Петербурге (г. Пушкин) и Краснодарском крае (на Кубанской опытной станции) в 2004–2006 гг. В изучение были включены скороспелые сорта, способные вызревать в условиях Краснодарского края и формировать выполненные семена в условиях Северо-Западного региона. В Пушкине проведено два посева — основной и ранний, 25 и 11 мая соответственно, на Кубани посев проводили с 7 по 13 мая. В 2004 г. в Пушкинском филиале сумма активных температур (выше +10°C) за вегетацию в зависимости от сроков посева составляла от 1718,9 до 1721,5°C; сумма осадков за вегетацию составила от 181,1 до 182,3 мм. В 2005 г. сумма активных температур составляла от 1768,3 до 1924,5°C, сумма осадков за вегетацию — от 82,8 до 90,6 мм. В 2006 г. сумма активных температур составляла от 1789,2 до 1930,8°C, сумма осадков за вегетацию — от 58,2 до 84,7 мм. На Кубанской опытной станции (КОС ВИР) в 2004 г. сумма активных температур за вегетацию была выше и составляла 2020°C при сумме осадков 339,7 мм; в 2005 г. сумма активных температур составляла 2247°C, сумма осадков — 107 мм. Продолжительность светового дня составляла от 16 ч (в мае) до 18 ч (в июне) в Пушкине и от 14 ч (в мае) до 15 ч (в июне) на КОС ВИР.

Полевые исследования проводили общепринятыми методами в соответствии с «Методическими указаниями по изучению коллекции зерновых бобовых культур» [15] и «Международным классификатором СЭВ рода *Glycine Willd.*» [16]. У

изучаемых образцов за годы исследований и по вариантам опыта содержание белка и масла в семенах заметно варьировало. В условиях Северо-Запада содержание белка колебалось от 27,4 до 49,7%, а масла — от 13,8 до 20,9%. В Краснодарском крае содержание белка составляло от 32,8 до 46,1%, а масла — от 16,4 до 24,4%.

ГРУППИРОВКА ДАННЫХ

Исходные данные содержат информацию о сортах, которые были созданы в нескольких географических областях. Образцы были объединены в группы по областям селекции (см. табл. 1) для сравнения результатов моделирования.

По происхождению образцы могут быть разделены на созданные в Европе (Sito, Major), в Канаде (Maple Ridge, KG-20), в сибирских селекционных центрах (СибНИИСХ 6, Степная 90, СибНИИК 15/83, Алтом, Степная 85), в Швеции (Fiskeby 840-2-7, Fiskeby 840-7-3, Fiskeby V, Fiskeby 1040-4-2, Bravalla), в селекционных центрах средней полосы РФ (Соер 13-91, Мадева, Восход*1191/79, Белор, Магева, УСХИ 6, Соер 4, Окская, Светлая), а также экспериментальные образцы, созданные в ВИРе в Ленинградской области (ПЭП 26, ПЭП 27, ПЭП 28, ПЭП 2, ПЭП 17, ПЭП 18).

МОДЕЛЬ ДЛИНЫ ПЕРИОДОВ НА ОСНОВЕ НЕЙРОННЫХ СЕТЕЙ

В работах [14, 17] была предложена модель на основе нейронной сети, которая состоит из четырех входных нейронов, двадцати нейронов в единственном скрытом слое и единственного выходного нейрона. В данной работе сеть была модернизирована для включения информации о принадлежности образца одной из групп G_k , $k = 1, \dots, 6$. Для каждого образца $G_k = 1$, если образец принадлежит к группе k , в иных случаях $G_k = 0$. Схема сети показана на рис. 1. Показаны

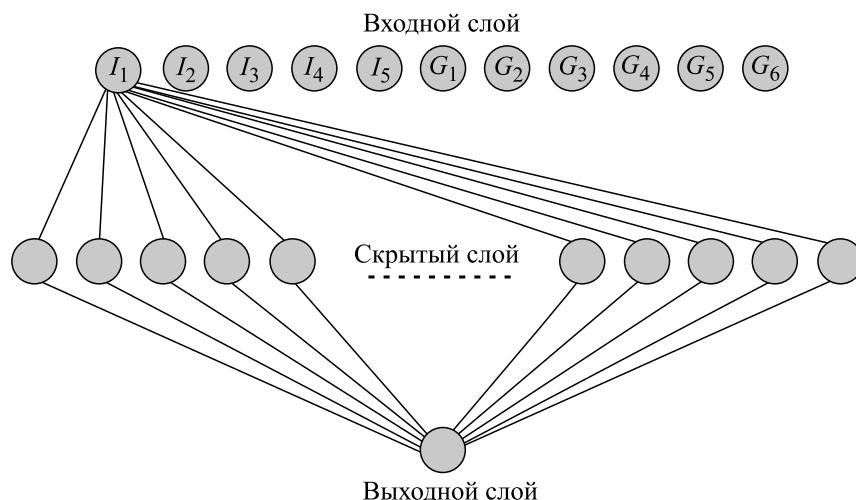


Рис. 1. Нейронная сеть модели длины периодов «посев–всходы», «всходы–цветение» и «цветение–созревание». На входы сети I_1 – I_5 подаются результаты применения специально подобранных функций климатических факторов, G_1 – G_6 получают информацию о принадлежности к группе. Показаны связи только первого нейрона входного слоя с нейронами скрытого слоя, чтобы не загромождать рисунок. Все нейроны входного слоя связаны с каждым нейроном скрытого слоя.

связи только первого нейрона входного слоя с нейронами скрытого слоя, чтобы не загромождать рисунок. Все нейроны входного слоя связаны с каждым нейроном скрытого слоя.

На входы сети I_1, I_2, I_3, I_4, I_5 подаются результаты применения специально подобранных функций, которые обобщают влияние ежедневных показателей максимальной и минимальной температуры, осадков, солнечной радиации и длины светового дня на состояние растения. Число входных функций по сравнению с предыдущими моделями было увеличено до пяти.

Входные данные I_i вычисляются по формуле (1) для каждого дня D суммированием по всем предыдущим дням j неотрицательных значений функций f_i от g_j – вектора ежедневных погодных данных и параметров:

$$I_i = \sum_{j=1}^D H(f_i(g_j))f_i(g_j), \quad (1)$$

$$g_j = \{T_{x,j}, T_{m,j}, P_j, R_j, L_j, T_{x,c}, T_{m,c}, P_c, R_c, L_c\}, \quad (2)$$

где H – функция Хевисайда (равна единице для неотрицательных аргументов, иначе 0); $T_{m,c}, T_{x,c}$ – нижняя и верхняя температурные границы роста; L_c – базовая длина светового дня; $T_{m,j}, T_{x,j}, L_j$ – минимальная температура, максимальная температура и длина светового дня за день j ; P_c, P_j – базовое и ежедневное количество осадков; R_c, R_j – базовая и ежедневная солнечная радиация. Конкретные значения параметров в векторе g_j и вид функций f_i должны быть выяснены в ходе

обучения сети для достижения наилучших результатов.

МОДЕЛЬ УРОЖАЙНОСТИ, СОДЕРЖАНИЯ БЕЛКА И МАСЛА В СЕМЕНАХ

Модели прогнозирования урожайности, содержания белка и масла в семенах строили с помощью программы autoML [18] на языке Python 3 по признакам: «Группа», «Длина периода «посев–всходы»», «Длина периода «всходы–цветение»», «Длина периода «цветение–созревание»», «Температура «посев–всходы»», «Осадки «посев–всходы»», «Температура «всходы–цветение»», «Осадки «всходы–цветение»», «Температура «всходы–созревание»», «Осадки «всходы–созревание»». AutoML использует автоматическое машинное обучение на основе байесовской оптимизации для построения ансамбля регрессионных моделей. Такой подход позволяет снизить риск переобучения [18].

ПРОГНОЗЫ ИЗМЕНЕНИЯ КЛИМАТА

Моделирование погоды на каждый день на Кубани с 2019 до 2030 года было проведено с помощью генератора погоды – программы МаркСим [19–23]. Были учтены социоэкономические сценарии развития, которые описываются четырьмя характерными профилями концентрации углекислого газа (Representative Concentration Pathways, RCPs), принятые IPCC (Intergovernmental Panel on Climate Change – Межправительственная группа экспертов по изменению климата) для

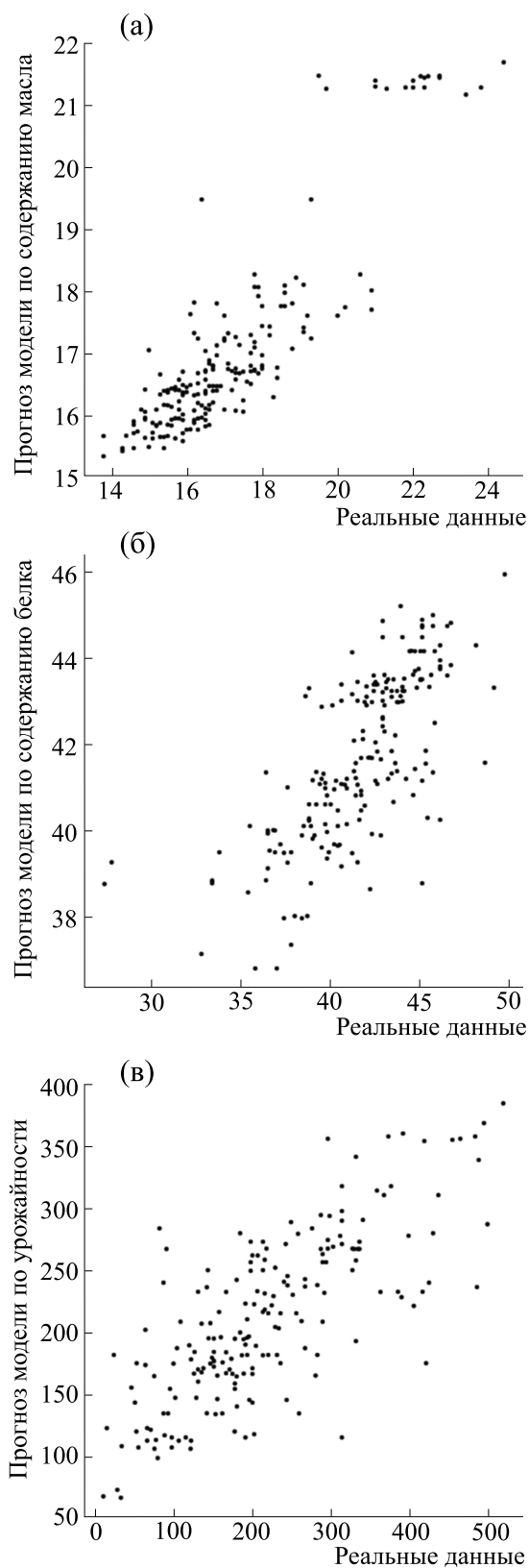


Рис. 2. Сравнение данных и решения модели для содержания масла (а), белка (б) и урожайности (в).

пятого оценочного отчета (Assessment Report, AR5) в 2014 г. 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].

РЕЗУЛЬТАТЫ АДАПТАЦИИ МОДЕЛЕЙ

С помощью программы autoML [18] построены модели урожайности, содержания белка и масла в семенах сои по признакам: «Группа», «Длина периода «посев–всходы»», «Длина периода «всходы–цветение»», «Длина периода «цветение–созревание»», «Температура «посев–всходы»», «Осадки «посев–всходы»», «Температура «всходы–цветение»», «Осадки «всходы–цветение»», «Температура «всходы–созревание»», «Осадки «всходы–созревание»». Набор данных включал 194 записи. В программе autoML были использованы параметры по умолчанию.

На рис. 2 показано сравнение данных о содержании масла и белка в семенах и урожайности и решений модели. Видно, что модель достаточно точно описывает экспериментальные данные. Коэффициент корреляции Пирсона на уровне значимости <0.01 составил для масла 0.88, для белка – 0.72, для урожайности – 0.75.

С использованием бутстрап-подхода в программе sklearn [25] были оценены относительные доли влияния признаков на модели (см. табл. 2). Урожайность и содержание белка в семенах в значительной степени зависели от генотипа, содержание масла – от тепловлагообеспеченности вегетации.

Модели для длин периодов «посев–всходы», «всходы–цветение» и «цветение–созревание» были запрограммированы на языке Python с использованием Keras [26] и TensorFlow [26] в качестве вычислителя. Данные о ежедневных максимальных и минимальных температурах, осадках и длине дня были взяты в открытом доступе в сети Интернет (<https://rp5.ru>). Аналитический вид функций f_i , а также значение констант $T_{m,c}$, $T_{x,c}$, L_c , P_c , R_c были получены путем минимизации отклонения решения модели от данных с использованием методов разностной эволюции [28, 29] и грамматической эволюции [17, 30].

Точность моделей оценивали с помощью коэффициента корреляции Пирсона между экспериментальными данными и решением модели. Модели с наилучшей точностью были отобраны

Таблица 2. Относительные доли влияния признаков в модели урожайности, содержания белка и масла в семенах сои

Признак	Модель содержания масла	Модель содержания белка	Модель урожайности
Группа	0.096	0.300	0.274
Длина периода «посев–всходы»	0.077	0.072	0.130
Длина периода «всходы–цветение»	0.044	0.072	0.096
Длина периода «цветение– созревание»	0.059	0.088	0.131
Температура периода «посев–всходы»	0.024	0.079	0.054
Осадки периода «посев–всходы»	0.088	0.091	0.037
Температура периода «всходы–цветение»	0.044	0.077	0.074
Осадки периода «всходы–цветение»	0.044	0.043	0.054
Температура периода «всходы–созревание»	0.200	0.089	0.090
Осадки периода «всходы–созревание»	0.324	0.089	0.059

Примечание. Значения округлены до трех знаков после запятой по правилам округления.

Таблица 3. Характеристика моделей периодов на основе нейронных сетей и

Имя	Период «посев–всходы»	Период «всходы–цветение»	Период «цветение–созревание»
f_1	$P_j - P_c$	$R_j - R_c - T_{x,j} - T_{x,c} + 1/T_{m,j}$	R_c
f_2	$(T_{x,j} - T_{x,c}) - (T_{x,j} - T_{x,c})/L_j$	$1/(T_{x,j} - T_{x,c})/(1/(R_j - R_c))$	$1/(T_{x,j} - T_{x,c})$
f_3	$T_{m,j}$	$1/(L_j - L_c) + (T_{m,j} - T_{m,c}) + 1/P_j$	$1/(P_j - P_c) - (P_j - P_c) + (L_j - L_c)$
f_4	$(T_{x,j} - T_{x,c})(T_{m,j} - T_{m,c})$	$1/(T_{m,j} - T_{m,c})$	$(T_{x,j} - T_{x,c}) + 1/(T_{x,j} - T_{m,c})$
f_5	$T_{x,j} - 1/(T_{x,j} - T_{x,c})$	1	$T_{m,j} - T_{x,c}$
$T_{m,c}$	6.683	10.174	10.975
$T_{x,c}$	26.136	29.868	23.805
L_c	–	15.493	15.544
P_c	7.288	–	0.376
R_c	–	6.479	0.090
r	0.79	0.78	0.067

Примечание. $T_{m,c}$, $T_{x,c}$ – нижняя верхняя граница роста; L_c – базовая длина светового дня; $T_{m,j}$, $T_{x,j}$, L_j – минимальная, максимальная температура и длина светового дня за день; P_c , P_j – базовое и ежедневное количество осадков; R_c , R_j – базовая и ежедневная солнечная радиация; r – коэффициент корреляции Пирсона между экспериментальными данными и решением модели..

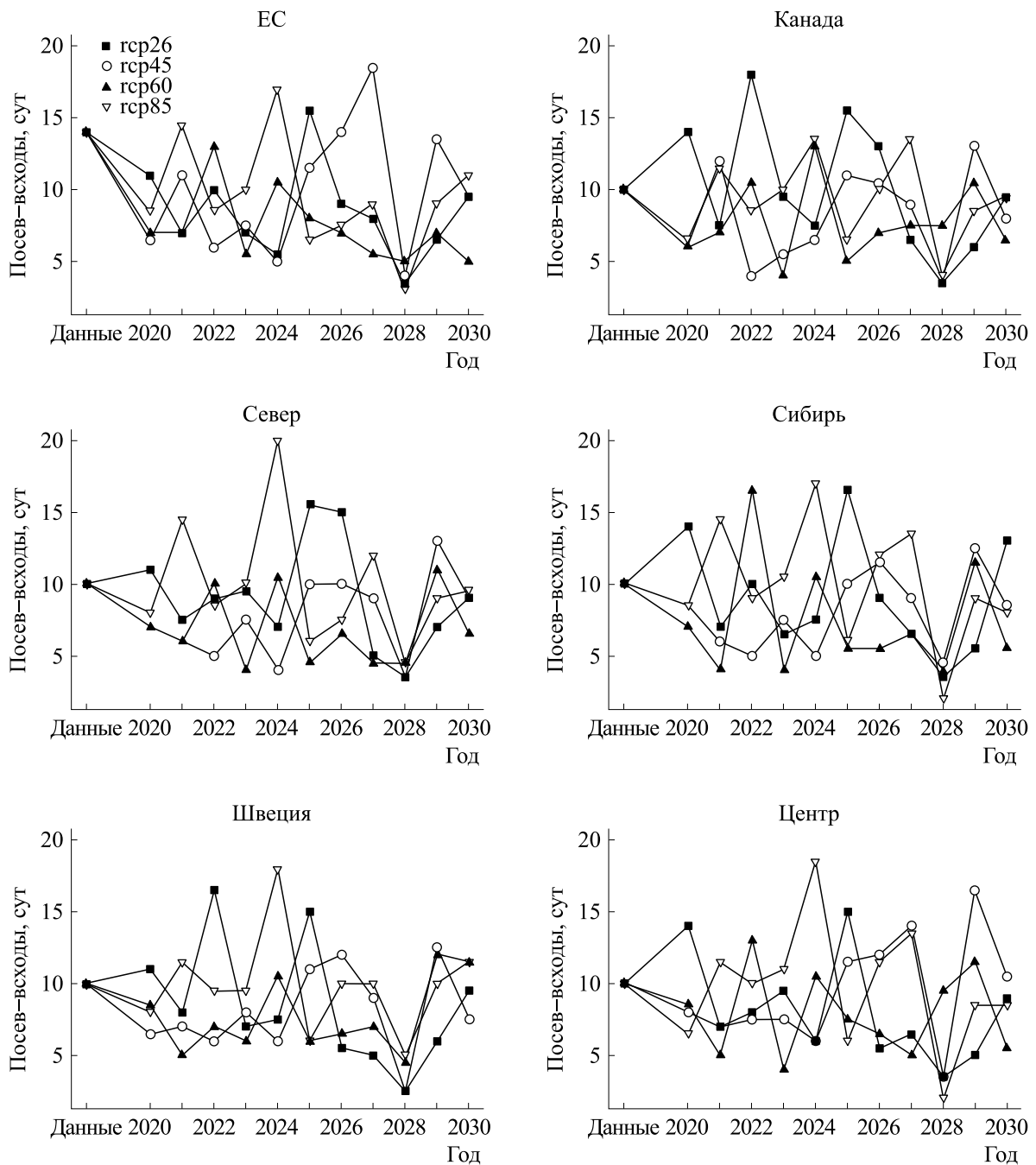


Рис. 3. Прогноз времени всходов сои по группам образцов разного происхождения. Различными символами — квадрат, круг, треугольник и обратный треугольник — помечены значения, соответствующие четырем характерным профилям концентрации углекислого газа: rcp26, rcp45, rcp60 и rcp85 соответственно.

для дальнейшего исследования (см. табл. 3). В каждой из моделей присутствуют линейные и нелинейные функции. Модель для длины периода «посев–всходы» оказалась независимой от солнечной радиации, а для периода «всходы–цветение» — независимой от осадков. Модели показывают удовлетворительную точность.

РЕЗУЛЬТАТЫ ИСПОЛЬЗОВАНИЯ МОДЕЛИ

Созданные погодные данные и построенные модели были использованы для прогнозирования длин периодов «посев–всходы», «всходы–цветение», «цветение–созревание», а затем для прогнозирования урожайности и содержания

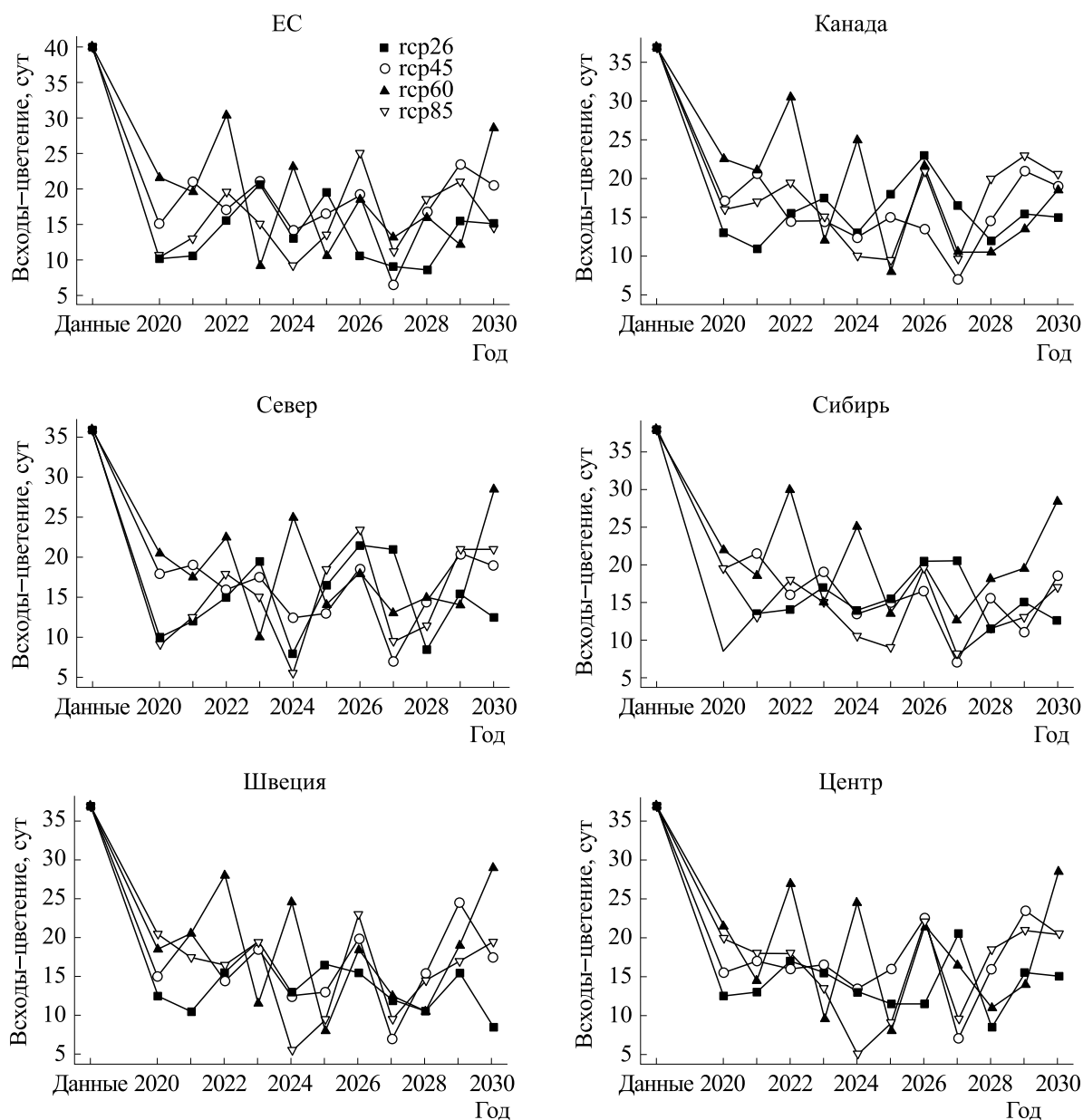


Рис. 4. Прогноз времени цветения сои, по группам образцов разного происхождения. Условные обозначения, как на рис. 3.

белка и масла в семенах для модельного посева 10 мая шести групп сортов сои в Кубани в 2019–2030 гг.

Результаты моделирования показывают, что время до всходов (см. рис. 3) сильно колеблется год от года в интервале примерно от –50% до +50% к значениям в данных. При этом максимальные и минимальные значения в большинстве случаев относятся к профилю rcp85. Прогнозные значения длины периода «всходы–цве-

тение» (см. рис. 4) имеют тенденцию к уменьшению, однако также подвержены колебаниям. В отличие от прогноза времени до всходов, верхняя граница колебаний существенно ниже значений в реальных данных, в то время как нижняя граница находится на уровне в три-четыре раза ниже. Модельные значения длины периода «цветение–созревание» (см. рис. 5) изменяются колебательно, однако не имеют выраженной тенденции к увеличению или уменьшению и лишь в

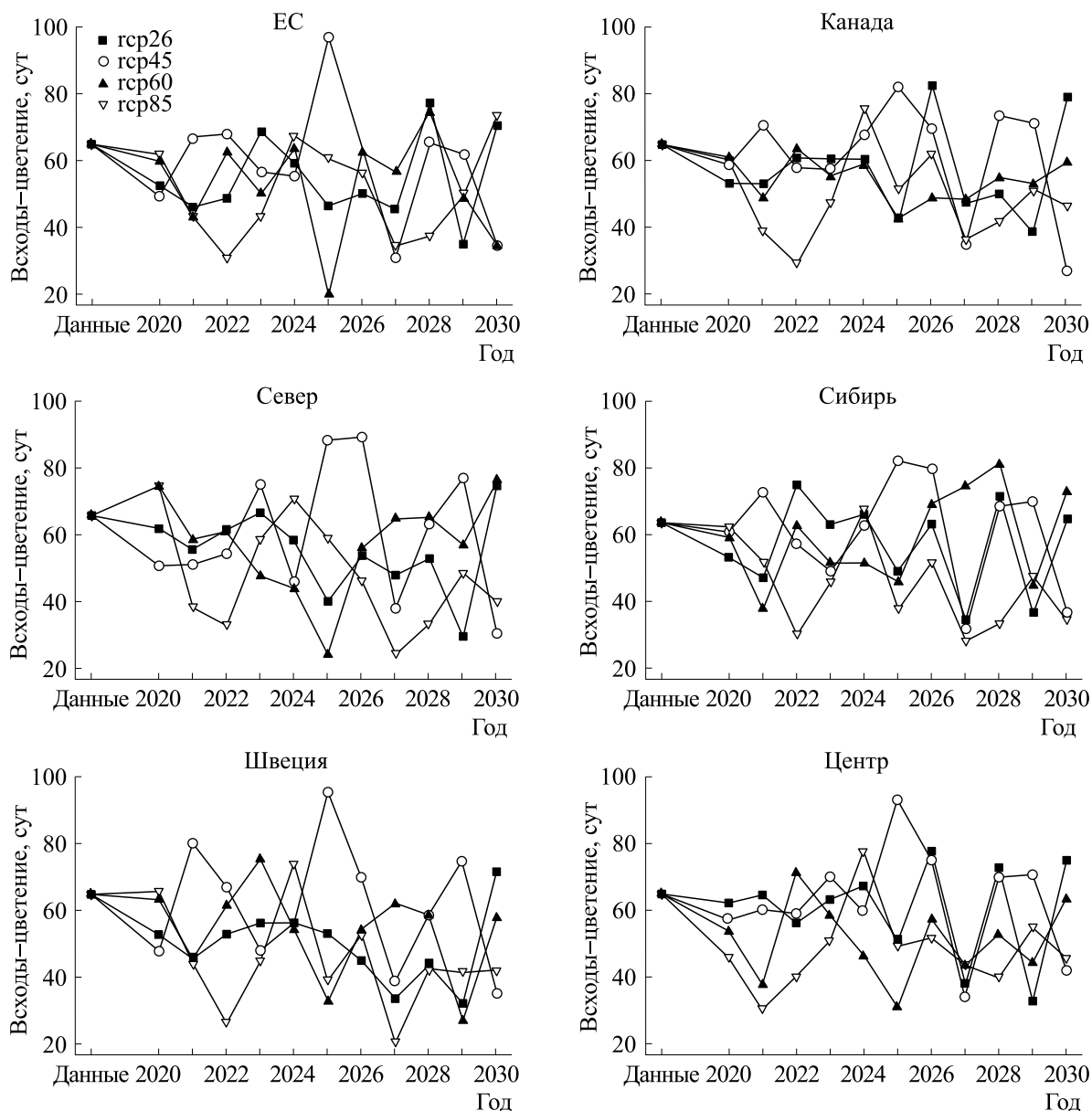


Рис. 5. Прогноз времени созревания сои, по группам образцов разного происхождения. Условные обозначения, как на рис. 3.

редких случаях отличаются от значений в экспериментальных данных больше чем на треть.

Результаты прогнозирования урожайности, содержания белка и масла в семенах сои показаны соответственно на рис. 6, 7 и 8. Для всех сортов, кроме группы ЕС, в которой данные получены для Ленинградской области, содержание масла прогнозируется значительно ниже значений в реальных данных. Сорты делятся на три группы по прогнозу содержания белка. Для сортов Севера, Швеции и Центра прогнозируется содержание белка на уровне примерно 40%, для сортов Сиби-

ри – на уровне 41%, для остальных сортов (ЕС и Канада) – на уровне 42%. При этом для Сибири и Центра прогноз показывает лишь слабые колебания и незначительный прирост, в то время как для остальных сортов прогнозируется прирост содержания белка. Прогноз урожайности показывает ее рост для всех групп сортов, кроме группы ЕС, при наличии ярко выраженных колебаний.

Такие результаты можно объяснить тем, что для моделей содержания белка в семенах и урожайности принадлежность образца к группе име-

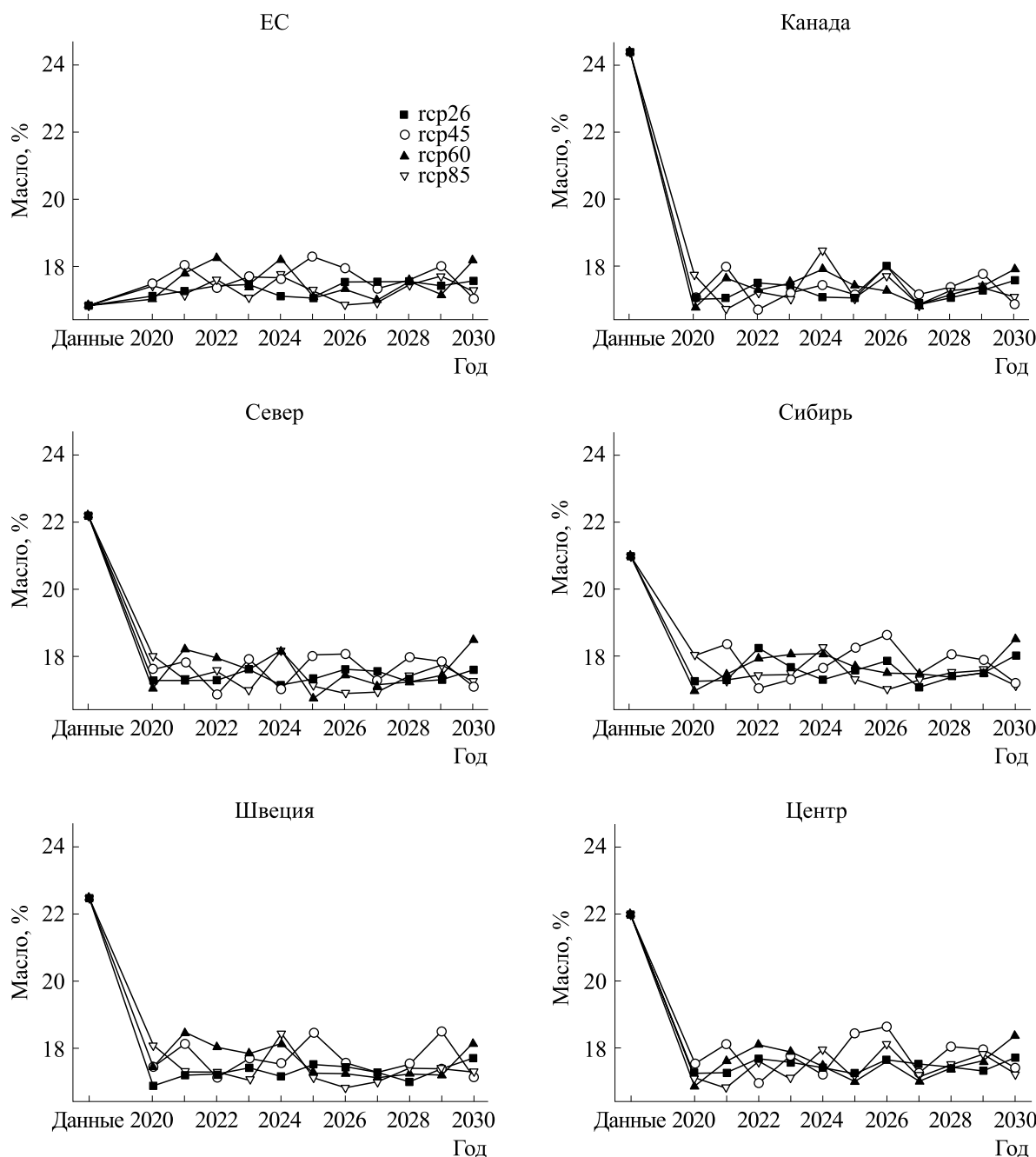


Рис. 6. Прогноз содержания масла в семенах сои, по группам образцов разного происхождения. Условные обозначения, как на рис. 3.

ет сравнительно высокую относительную долю влияния (см. табл. 2), поэтому прогнозы отличаются по группам (см. рис. 7 и 8). Колебания прогнозируемых значений урожайности, белка и масла порождаются колебаниями прогнозируемых значений длин периодов, которые можно объяснить стохастической природой моделей генерации погоды, а также наличием нелинейности

в моделях и сложным балансом погодных факторов. Одновременное снижение содержания масла в семенах сои и рост урожайности требует дополнительного объяснения. Мы предполагаем, что в условиях прогноза может формироваться больше семян. Однако скороспелые сорта попадают при созревании в условия избыточно высоких температур, при которых накопление масла понижено.

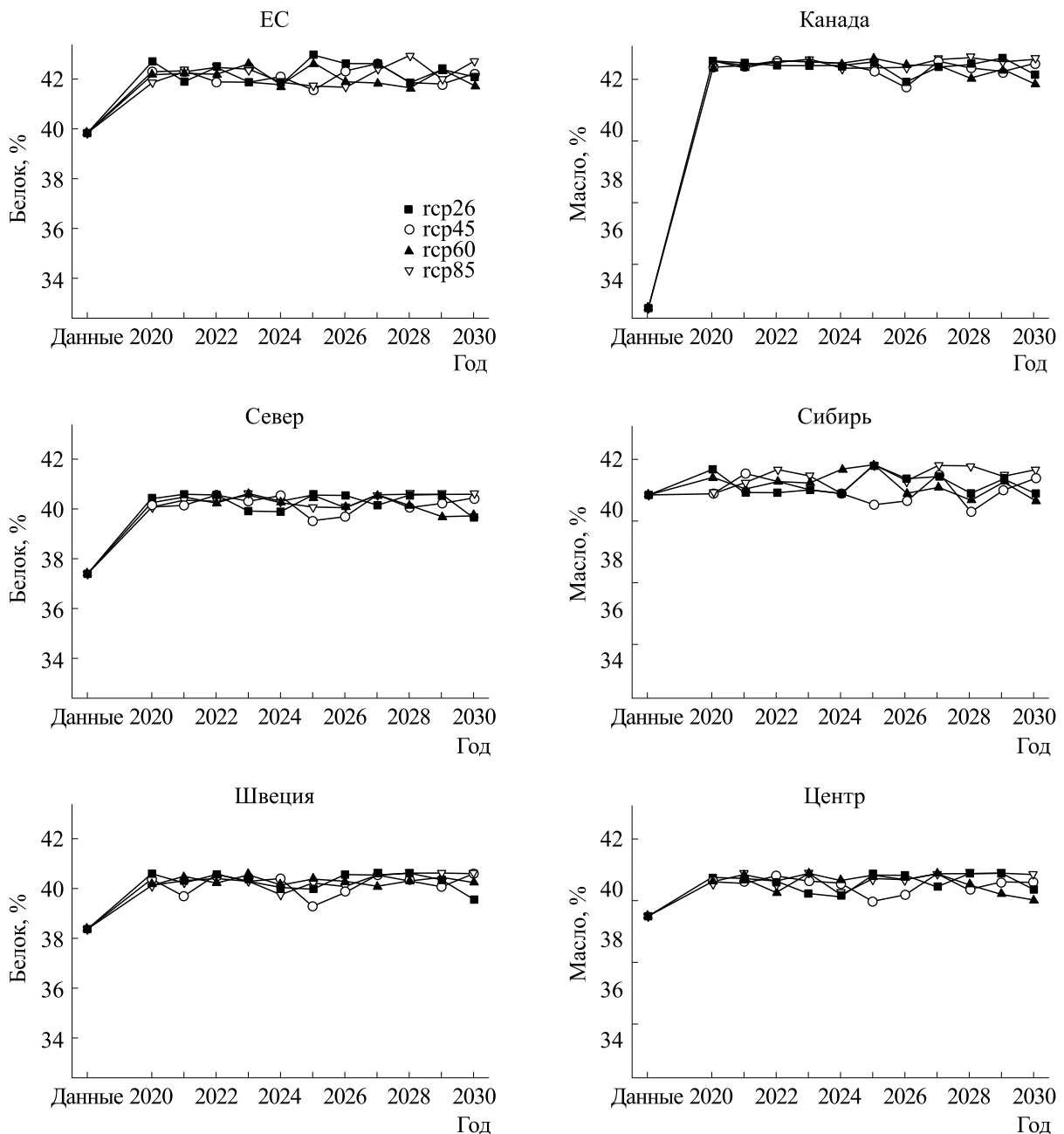


Рис. 7. Прогноз содержания белка в семенах сои, по группам образцов разного происхождения. Условные обозначения, как на рис. 3.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Использование математических моделей для прогнозирования хозяйственно-ценных характеристик растений, таких как длины периода «посев–всходы», «всходы–цветение», «цветение–созревание», а также урожайность, содержание белка и масла в семенах сои, должно являться научной основой селекционного улучшения сортов культуры в условиях глобальных изменений кли-

мата. В данной работе были построены модели для предсказания характеристик сортов сои, выведенных в шести регионах и фенотипированных в условиях Северо-Кавказского и Северо-Западного регионов. Разработанные модели используют новейшие подходы математического моделирования, такие как автоматическое машинное обучение, нейронные сети и грамматическую эволюцию, с помощью которой были подобраны

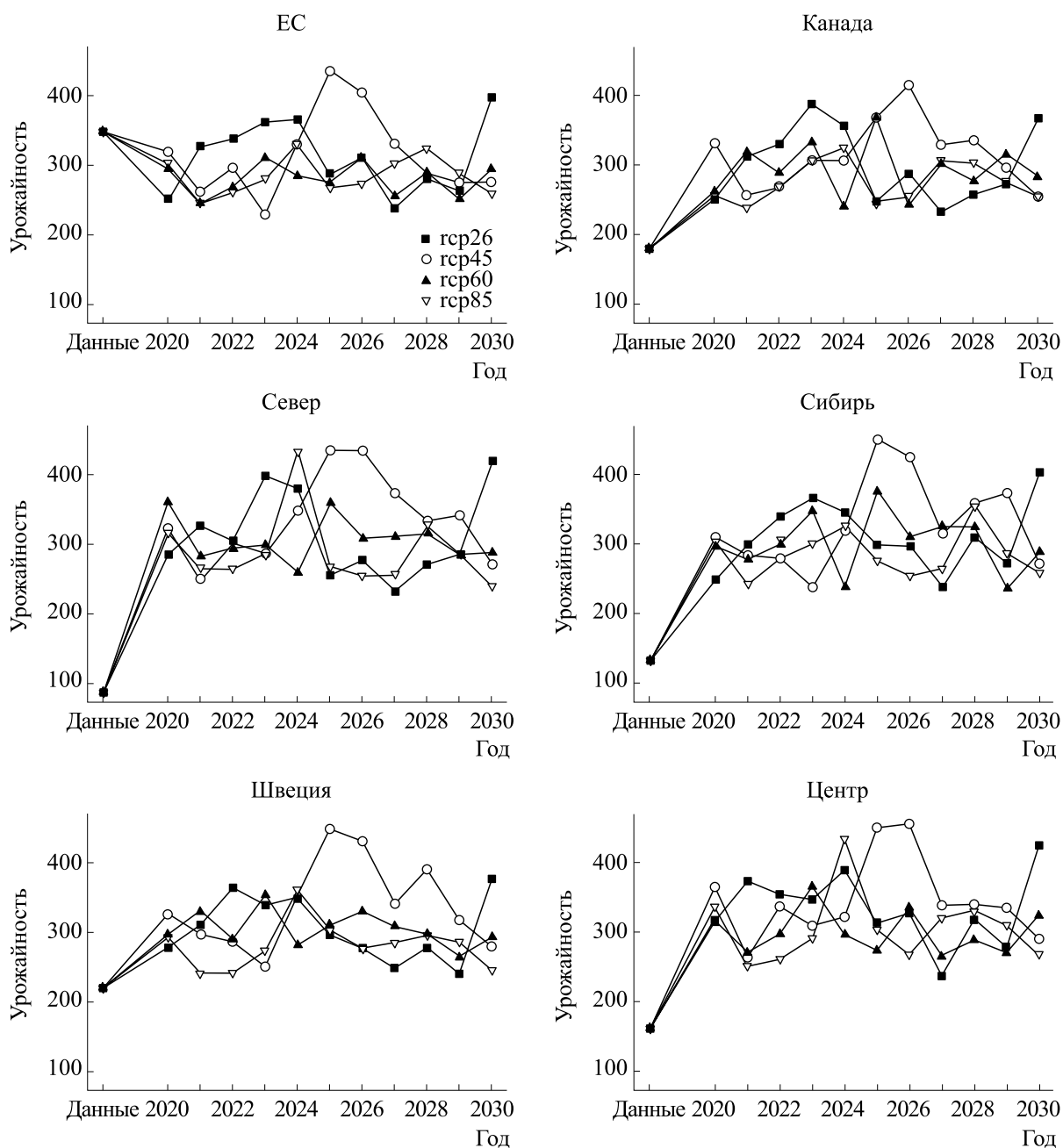


Рис. 8. Прогноз урожайности сои, по группам образцов разного происхождения. Условные обозначения, как на рис. 3.

функции погодных данных. Расчеты по имеющимся экспериментальным данным показали высокую точность моделирования. Прогноз роста содержания белка в семенах сои, не столь благоприятный прогноз для содержания масла, зависимость содержания белка от сорта совпадают с результатами, полученными по простым регрессионным уравнениям [31]. Однако новейшие подходы показали более высокие прогностические способности моделей.

Модельные прогнозы показали различную реакцию групп сортов на изменение климата. Однако для большинства сортов прогнозируется рост урожайности, что указывает на благоприятные перспективы использования изменений климата.

Таким образом, показана пригодность модели фенологии сои, созданной методом искусственных нейронных сетей, для прогнозирования в изменяющемся климате.

ФИНАНСИРОВАНИЕ РАБОТЫ

Работа выполнена в рамках Федеральной целевой программы (проект №14.575.21.0136 от 26.09.2017, уникальный идентификатор проекта RFMEFI57517X0136).

БЛАГОДАРНОСТИ

Вычисления были поведены в Суперкомпьютерном центре «Политехнический» СПбПУ и кластере Университета Южной Калифорнии.

Исходные данные получены на базе уникальной научной установки Коллекция генетических ресурсов растений ВИР.

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Настоящая работа не содержит описания каких-либо исследований с использованием людей и животных в качестве объектов.

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: MEDICATION 2.1* (Finnish Environment Institute, Helsinki, Finland, 2011).
4. A. D. Richardson, et al., *Global Change Biology* **18** (2) 566 (2012).
5. К. Н. Козлов, Л. Ю. Новикова, И. В. Сеферова и М. Г. Самсонова, *Биофизика* **63** (1), 175 (2018).
6. D. J. Major, et al., *Crop Science* **15**, 174 (1975).
7. T. Hodges and V. French, *Agronomy J.* **77** (3), 500 (1985).
8. P. Pedersen, et al., *Agronomy J.* **96**, 556 (2004).
9. T. D. Setiyono, et al., *Field Crops Res.* **100** (2–3), 257 (2007).
10. Л. Ю. Новикова, И. В. Сеферова и К. Н. Козлов, *Биофизика* **63** (6) 1182 (2018).
11. M. Abdipour, et al., *J. Am. Oil Chemists' Soc.* **95** (3), 283 (2018).
12. M. Kaul, R. L. Hill, and C. Walthall, *Agricult. Systems* **85** (1), 1 (2005).
13. A. Bagherzadeh, et al., *Modeling Earth Systems and Environment* **2** (2), (2016).
14. D. A. Elizondo, R. W. McClendon, and G. Hoogenboom, *Trans. ASAE* **37** (3), 981 (1994).
15. Н. И. Корсаков, О. П. Адамова и В. И. Будаова, *Методические указания по изучению коллекции зерновых бобовых культур* (ВНИИ растениеводства им. Н. И. Вавилова, Ленинград, 1975).
16. Л. Щелчко, *Международный классификатор СЭВ рода Glycine Willd* (Науч.-техн. совет стран-членов СЭВ по коллекциям диких и культурных видов растений, ВНИИ растениеводства им. Н. И. Вавилова, Ленинград, 1990).
17. О. Д. Таратухин, Л. Ю. Новикова, И. В. Сеферова и К. Н. Козлов, *Биофизика* **64** (3) 563 (2019).
18. M. Feurer, et al. in *Advances in Neural Information Processing Systems*, Ed. by C. Cortes et al. (Curran Associates, Inc., 2015), pp. 2962–2970.
19. M. Srinivasa Rao, et al., *PLoS One* **10** (2), e0116762 (2015).
20. P. G. Jones and P. K. Thornton, *Agricult. Forest Meteorol.* **86** (1–2), 127 (1997).
21. P. G. Jones and P. K. Thornton, *Agricult. Forest Meteorol.* **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, *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).
25. F. Pedregosa, et al., *J. Machine Learning Res.* **12**, 2825 (2011).
26. F. Chollet, et al., *Keras* (GitHub, 2015). URL: <https://github.com/keras-team/keras>.
27. M. Abadi, et al., in *Proc. 12th USENIX Conf. on Operating Systems Design and Implementation* (Savannah, GA, USA, 2016), pp. 265–283.
28. K. Kozlov, A. M. Samsonov, and M. Samsonova, *Peer J. Comp. Sci.* **2**, e74 (2016).
29. R. Storn and K. Price, *J. Global Optimization* **11**, 341 (1997).
30. M. O'Neill and C. Ryan, *IEEE Trans. Evolutionary Computation* **5** (4), 349 (2001).
31. Л. Ю. Новикова, И. В. Сеферова, А. Ю. Некрасов и др., *Вавиловский журн. генетики и селекции* **22** (6) 708 (2018).

An Artificial Neural Network Model for Prediction of Phenology of Early Maturing Soybean Varieties in Relation to Climate Factors

O.D. Taratuhin*, **L.Yu. Novikova***, ******, **I.V. Seferova****, **T.V. Gerasimova****, **S.V. Nuzhdin***, *******,
M.G. Samsonova*, 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*

****University of Southern California, Los Angeles, CA, 90089 USA*

Soybean phenology is strongly influenced by temperature and day length and phenological records clearly reflect changes in climatic conditions. We developed a model that includes four artificial neural networks for forecasting time intervals between sowing, emergence, flowering and maturity in relation to climatic factors and also formed assemblies of regression models to predict the yield, seed protein and oil content in soybean. We analyzed data on maturation of early ripening soybean accessions phenotyped at two experimental stations of N.I. Vavilov Institute of Plant Genetic Resources in North-Caucasian and North-Western districts of Russia. The model was implemented in Python using Keras and TensorFlow serving libraries.

Keywords: climatic factors, soybeans, artificial neural networks