

ПАРАМЕТРИЗАЦИЯ МОДЕЛИ ВРЕМЕНИ ЦВЕТЕНИЯ ОБРАЗЦОВ ДИКОГО НУТА С ПЕРЕНОСОМ ЗНАНИЙ ОТ НЕСКОЛЬКИХ ИСТОЧНИКОВ

© 2024 г. З.А. Саранин*, М.Г. Самсонова*, К.Н. Козлов*.,#

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

#E-mail: kozlov_kn@spbstu.ru

Поступила в редакцию 05.09.2023 г.

После доработки 05.09.2023 г.

Принята к публикации 20.09.2023 г.

Построение прогноза времени цветения позволяет исследователям создавать сорта, достигающие максимальной эффективности и ценности в условиях изменения климата. В данной работе предложен алгоритм параметризации модели времени цветения образцов дикого нута, который использует метод машинного обучения «перенос знаний» (transfer learning) для комбинирования нескольких наборов исходных данных и целевых данных. Построенная модель, используя генетические и климатические данные только за первые 20 суток после посева, предсказывает время цветения образцов с высокой точностью — средняя абсолютная ошибка составляет чуть больше 5 суток, коэффициент корреляции Пирсона — 0.93. Было обнаружено, что максимальная и минимальная температура оказывают наиболее сильное влияние на время цветения. При этом все погодные факторы на 7–10-е сутки после посева оказывают влияние на решение модели.

Ключевые слова: климатические факторы, нут, математическое моделирование, перенос знаний.

DOI: 10.31857/S0006302924050108, EDN: MJZCUO

Совершенствование селекции зернобобовых культур и прогнозирование их фенотипических признаков в частности является важным условием для обеспечения продовольственной безопасности растущего населения земного шара. Нут (*Cicer arietinum* L.) — вторая по распространенности зернобобовая культура, богатая пищевым белком [1]. Дикий нут, *Cicer reticulatum*, является растением длинного дня (LD) и проявляет чувствительность к яровизации [2, 3]. Однако окультуривание этого растения привело к появлению генотипов, не требующих яровизации и пригодных для весеннего посева [3, 4].

Факторами, устанавливающими предел времени для достижения нутом репродуктивной фазы, являются температурный режим, продолжительность светового дня, доступность воды [2, 5]. Условием стабильного урожая является соответствие продолжительности роста доступному вегетационному периоду [6]. Сорта нута, различающиеся по продолжительности роста, получают путем селекции. Различие в продолжительности роста сортов позволяет адаптировать их к разным широтам и режимам посева [3, 7–9].

Для прогнозирования фенотипических признаков может быть использовано математическое моделирование. Распространенные математические модели бобовых культур SSM [11, 12], DSSAT [13–17], APSIM [18] и др. [19, 20] описывают фотосинтез, поглощение воды и другие биофизические и биохимические процессы с помощью дифференциальных уравнений. Скорость перехода к следующей фазе в зависимости от погодных условий вычисляется с использованием таких характеристик, как индекс тепловых единиц (HUI) [20], единицы тепла (CHI), тепловые дни (DD) или биологические дни (BD) [21]. Адаптация существующих моделей к новым сортам или погодным условиям требует значительных усилий и времени [22–27]. Таким образом, разработка новых гибких моделей, которые легко могут адаптироваться к новому сорту, остается актуальной задачей.

В данной работе предложен алгоритм параметризации модели времени цветения образцов дикого нута с использованием переноса знаний (transfer learning) — метода машинного обучения, позволяющего комбинировать несколько наборов исходных и целевых данных.

ПЕРЕНОС ЗНАНИЙ ОТ НЕСКОЛЬКИХ ИСТОЧНИКОВ

Перенос знаний (transfer learning) — прием в машинном обучении, который использует информацию о хорошо изученном наборе объектов (исходной области) для описания целевой области: нового набора данных, сходного по каким-то параметрам с объектами исходной области. При построении агроклиматических моделей часто встает вопрос об использовании данных из нескольких экспериментов для увеличения количества образцов при обучении модели. В алгоритме обучения с переносом знаний знания из исходной области переносятся на целевую область. При таком обучении точность предсказаний в значительной степени зависит от сходства исходной и целевой областей. Использование нескольких источников при описании целевой области позволяет полнее использовать разнообразие данных.

ОБРАЗЦЫ РАСТЕНИЙ И ПОЛЕВЫЕ ИССЛЕДОВАНИЯ

Авторами работы [28] коллекция образцов дикого нута *Cicer reticulatum* L. и *Cicer echinospermum* была собрана в 21 географической локации вокруг 10 населенных пунктов в 5 регионах Турции. В районах сбора имеется значительный перепад высот, при этом наблюдения за климатом показывают, что на небольших высотах температуры выше, а количество осадков в среднем ниже по сравнению с более высокими точками. Отмечено, что популяции *C. echinospermum* обычно встречаются на более низких высотах, чем популяции *C. reticulatum*. Самые низкие среднемесячные температуры (от -4.8 до -2.2°C по сравнению с -2.2 до -0.6°C) зимой отмечаются на трех наиболее восточных и высокогорных участках. Весной на этих участках чаще бывают заморозки в вегетационный период. В ходе полевых экспериментов в Шанлыурфе и Анкаре (Турция) и Перте (Австралия) были исследованы 2174 образца с посевом во второй половине года на 270, 290, 294, 305, 325 или 339 сутки от его начала. Время цветения для этого набора данных колеблется от 117 до 221 суток. Эти данные были использованы в обучении с переносом знаний как источники 1 (турецкие образцы) и 2 (австралийские образцы) соответственно.

Были также проведены полевые эксперименты с 2088 образцами, высеванными на 32, 36, 78, 95, 161 сутки календарного года; время цветения для этого набора данных составило от 64 до 152 суток. Этот набор данных был использован в обучении с переносом знаний как целевая область.

КОМБИНИРОВАННАЯ МОДЕЛЬ СИМВОЛЬНОЙ РЕГРЕССИИ

Для прогнозирования времени цветения построим модель Q (1) как линейную комбинацию методом LASSO моделей символьной регрессии Q_k :

$$Q = b_0 Q_0 + b_1 Q_1 + \dots + b_k Q_k + \dots, \quad (1)$$

где b_k — коэффициенты, полученные методом LASSO, а каждая из моделей Q_k строится как суперпозиция элементарных функций «+», «−», «*», «/», связывающих переменные, описывающие генетические и климатические факторы.

Шесть однонуклеотидных полиморфизмов, предположительно связанных со временем цветения, были идентифицированы как лучшие однонуклеотидные полиморфизмы после запуска смешанной линейной модели (MLM) в TASSEL, которая ассоциировала время цветения (фенотип) с генотипами, используя место/год/сезон в качестве фактора для учета их влияния на фенотип.

В моделях используются переменные SNP1AA–SNP6RR, принимающие значения 0 или 1 для обозначения альтернативных (ALT) и референсных (REF) аллелей растений (ALT/ALT, REF/ALT и REF/REF) [27].

Данные о климатических условиях на каждые сутки, начиная с посева до 20 суток ($a = 0, \dots, 19$), взяты с общедоступного сайта «Радар погоды» (https://rp5.ru/Weather_in_the_world) и проекта POWER исследовательского центра NASA в Лэнгли (LaRC) [29]:

- $D(a)$ — длина дня;
- $T_n(a)$ — минимальная температура;
- $T_x(a)$ — максимальная температура;
- $P(a)$ — количество осадков;
- $R(a)$ — солнечное излучение.

АЛГОРИТМ ПАРАМЕТРИЗАЦИИ МОДЕЛИ

Будем строить комбинированную модель символьной регрессии Q , учитывающую климатические и генетические факторы и минимизирующую среднее относительное отклонение решения модели от данных (MAPE). Общая схема алгоритма представлена на рис. 1. Пусть исходные наборы данных $S_j = \{(x_i, y_i)_j, i = 1, \dots, N_j\}$, $j = 1, \dots, N$, а целевой набор данных $T = \{(x_i, y_i)_T, i = 1, \dots, N_T\}$ и L — число моделей символьной регрессии для каждого источника. Параметризация модели проводится для части целевого набора данных, от которого предварительно отделена часть для проверки — независимого тестирования финального результата.

1. Случайно разделим целевой набор на две равные половины: U, V .

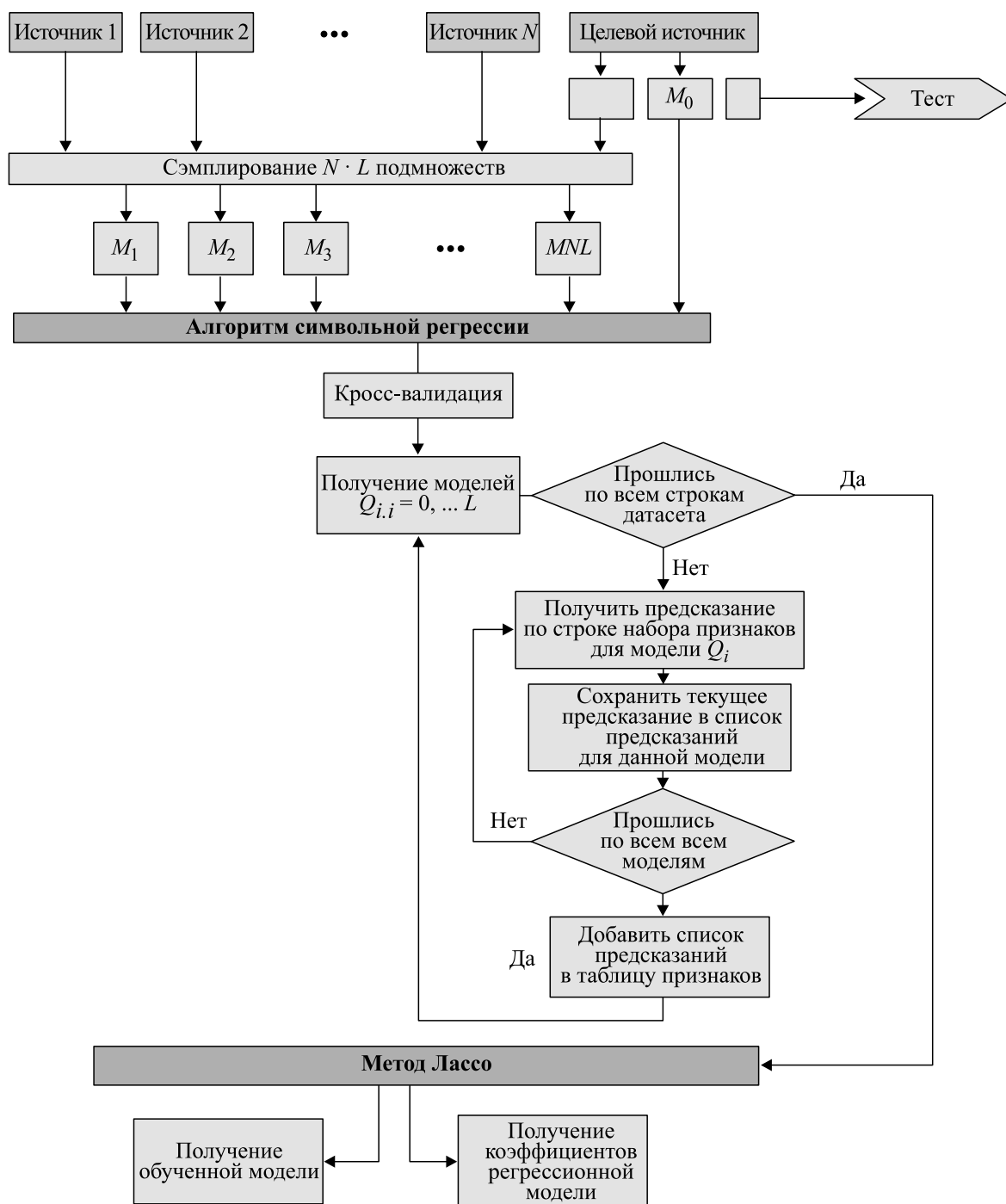


Рис. 1. Блок-схема алгоритма обучения с переносом знаний от нескольких источников.

2. Для каждого источника $j = 1, \dots, N$ создадим объединение половины набора целевых данных V и всех данных из других источников. Выберем из этого объединения L подмножества W_k для обучения моделей символьной регрессии случайной выборкой N_T образцов. Таким образом, $k = 0, \dots, NL$, $W_0 = U$.

3. По каждому набору данных W_k построим модель символьной регрессии Q_k , используя кросс-валидацию с разбиением 4 : 1 на обучающую и валидационную выборки.

4. С помощью метода LASSO [30] получим комбинированную модель Q как линейную комбинацию моделей Q_k .

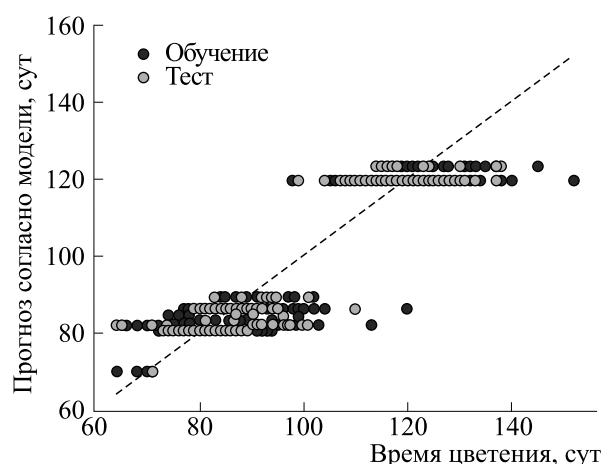


Рис. 2. Сравнение решения модели с данными для обучающей и тестовой выборок.

Из целевого набора данных тестовая выборка для проверки финальной модели составила 417 образцов (примерно 20% от общего числа 2088), таким образом объем множеств U и V равен 835 и 836 записей соответственно. Для кросс-валидации моделей символьной регрессии было использовано 8 разбиений, для метода LASSO — 10. Число моделей для каждого источника L было выбрано равным 15 в ходе тестирования, таким образом, общее число моделей на шаге 3 составляет $NL + 1 = 2 \times 15 + 1 = 31$ модель.

МОДЕЛЬ ДЛИНЫ ПЕРИОДА ПОСЕВ–ЦВЕТЕНИЕ

В данной работе построена модель (2) длины периода от посева до цветения в зависимости от ежедневных погодных факторов в первые 20 суток после посева ($a = 0, \dots, 19$) и заданных однонуклеотидных полиморфизмов:

$$Q = 0.025 Q_8 + 0.05 Q_9 + 0.50 Q_{14} + 0.26 Q_{23},$$

где

$$\begin{aligned} Q_8 &= T_n^2(5) + (\text{SNP4AA} - \text{SNP4AA}/(D(4)T_x(0)(\text{SNP2RR} + D(17)))) + \\ &+ (D(4) - \text{SNP4AA})/(D(7)T_x(6)) + (D(4) - \text{SNP4AA})/(D(2)T_x(6)) + \\ &+ (\text{SNP4AA} - D(4))(T_n(16) - R(3))/D(17))^2, \\ Q_9 &= D(7)T_x(5) - P(17)R(14) - 2P(4) + P(6) + R(2), \\ Q_{14} &= (\text{SNPAA}(4) + T_x(1))(R(9) + (D(17) - D(14)) - P(13)), \\ Q_{23} &= D(3)T_n(8) + 2D(4) + P(17)^2 + T_x(16), \end{aligned} \quad (2)$$

где $D(a)$ — длина дня, $T_n(a)$ — минимальная температура, $T_x(a)$ — максимальная температура, $P(a)$ — количество осадков, $R(a)$ — солнечное излучение.

Данная модель с оптимизированными параметрами описывает данные обучения и проверки с коэффициентом корреляции Пирсона, равным приблизительно 0.93 (см. рис. 2), коэффициент детерминации составил примерно 0.85. Для валидационной и тестовой выборок средняя абсолютная ошибка составляет 5.14 и 5.24 суток, максимальная абсолютная ошибка — 33.43 и 23.34, среднеквадратическая ошибка — 6.70 и 6.71 суток соответственно.

Надо отметить, что результирующие формулы являются нелинейными, используют дробно-рациональные функции, произведения предикторов и возведение в квадрат. При этом результиру-

ющая модель зависит от всех климатических показателей в разное время после посева.

Среди 31 модели символьной регрессии, построенной на шаге 3 алгоритма, в результате применения метода LASSO на шаге 4 было отобрано только 4 модели. Модель Q_0 , которая строится без использования данных источников, не вошла в результирующее выражение (2).

ИССЛЕДОВАНИЕ ВЛИЯНИЯ РАЗЛИЧНЫХ ФАКТОРОВ НА ТОЧНОСТЬ МОДЕЛИ

Использование математических моделей позволяет выявить факторы, наиболее влияющие на точность прогнозов. Для оценки важности каждого фактора из имеющихся в наборе данных будем применять разработанный алгоритм параметризации к набору данных, из которого

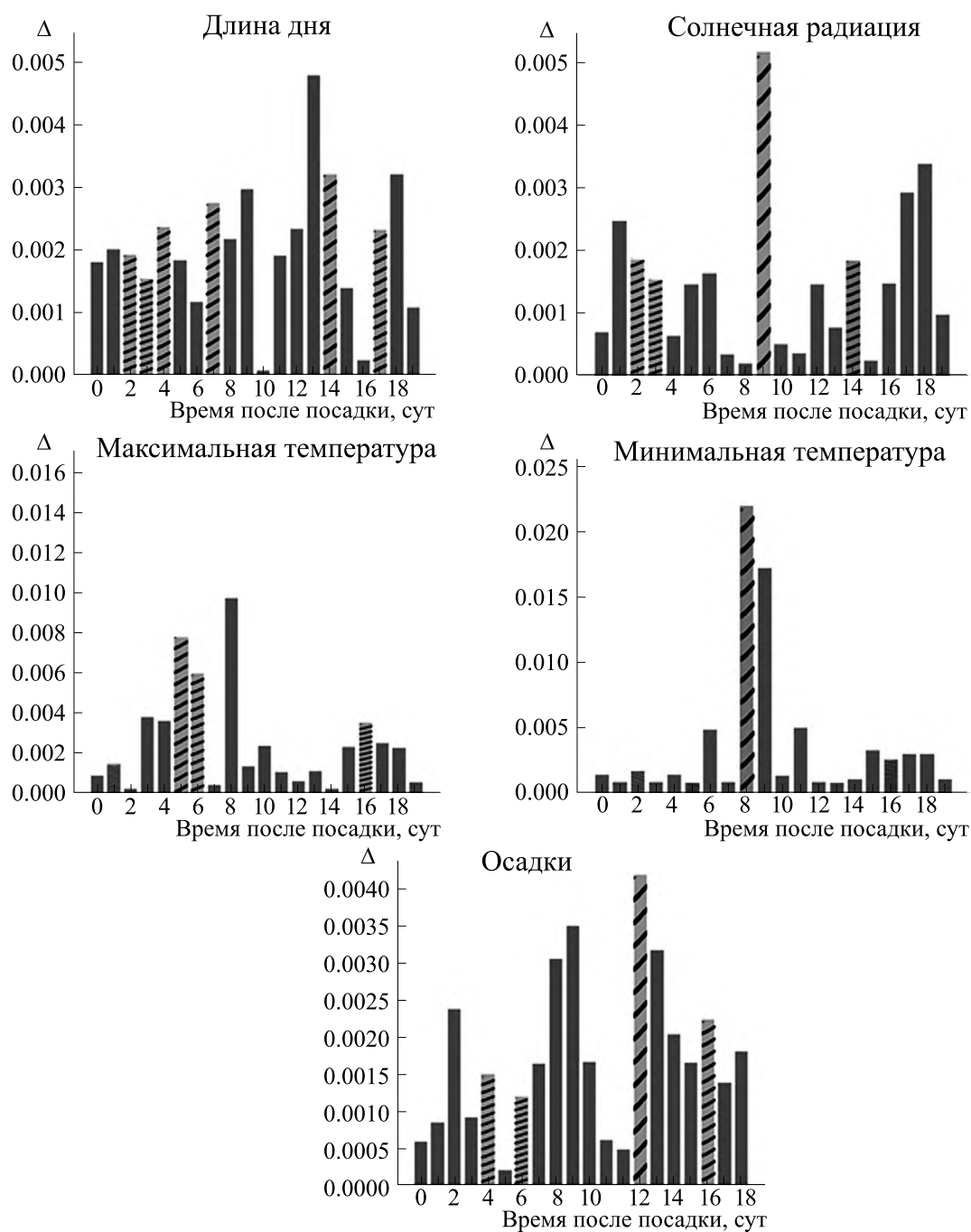


Рис. 3. Влияние климатических факторов на решение модели, вычисленное методом исключения.

интересующий нас фактор удален. Таким образом, модель будет обучаться в условиях полного отсутствия информации о рассматриваемом факторе, что является более сильным ограничением, чем широко используемый метод пермутации, в котором значения оцениваемого фактора переставляются между наблюдениями. Важность фактора будем оценивать по изменению ошибки решения модели.

Согласно полученным результатам, можно сделать вывод о том, что все погодные факторы на 7–10 сутки после посадки оказывают влияние на решение модели (см. рис. 3). Максимальное влияние оказывает минимальная температура на 8 и 9 сутки, и максимальная температура на 4 и 8 сутки. В том случае, когда данный фактор входит в итоговую модель Q (2), соответствующий ему столбик имеет штриховку (см. рис. 3).

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Использование математических моделей для прогнозирования хозяйственно ценных характеристик растений, таких как длины периодов между такими фенологическими фазами, как посев и цветение, закладывает научную основу селекционного улучшения сортов в условиях глобальных изменений климата.

В данной работе была рассмотрена параметризация модели времени цветения образцов дикого нута, которая использует метод машинного обучения перенос знаний (transfer learning) для комбинирования нескольких наборов исходных и целевых данных. Расчеты по имеющимся экспериментальным данным показали высокую точность моделирования.

Полученные результаты о виде функциональных зависимостей и влиянии погодных факторов согласуются с ранее опубликованными данными [21, 27, 28], однако понимание роли температуры и продолжительности дня в адаптации к различным типам среды обитания все еще неполное и требует дальнейших исследований для проверки модельных решений [31].

ВЫВОДЫ

Использование данных полевых экспериментов, проведенных в различных средах, определяемых сочетанием локации, времени и других внешних факторов, с применением методов машинного обучения, таких как перенос знаний и символьная регрессия, позволило построить модель, с высокой точностью описывающую длину периода от посева до цветения у образцов дикого нута.

Показано, что максимальная и минимальная температура оказывают наиболее сильное влияние на время цветения. Полученные результаты о виде функциональных зависимостей и влиянии погодных факторов согласуются с ранее опубликованными данными.

БЛАГОДАРНОСТИ

Вычисления были проведены в Суперкомпьютерном центре «Политехнический» СПбПУ. Авторы выражают благодарность М.П. Банкину, С.Ю. Сурковой, С.В. Нуждину, Э.Дж. Бишопу-фон Веттбергу, А. Кахраману, А. Айдогану, Д. Куку и М.А. Вишняковой.

ФИНАНСИРОВАНИЕ РАБОТЫ

Работа выполнена при финансовой поддержке Российского научного фонда (грант № 22-46-02004).

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов, связанных с изложенными в статье данными.

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Настоящая статья не содержит описания собственных исследований с участием людей или животных в качестве объектов.

СПИСОК ЛИТЕРАТУРЫ

1. Varshney R. K., Song C., Saxena R. K., Azam S., Yu S., Sharpe A. G., Cannon S., Baek J., Rosen B. D., Tar'an B., Millan T., Zhang X., Ramsay L. D., Iwata A., Wang Y., Nelson W., Farmer A. D., Gaur P. M., Soderlund C., Penmetra R. V., Xu C., Bharti A. K., He W., Winter P., Zhao S., Hane J. K., Carrasquilla-Garcia N., Condie J. A., Upadhyaya H. D., Luo M. C., Thudi M., Gowda C. L., Singh N. P., Lichtenzweig J., Gali K. K., Rubio J., Nadarajan N., Dolezel J., Bansal K. C., Xu X., Edwards D., Zhang G., Kahl G., Gil J., Singh K. B., Datta S. K., Jackson S. A., Wang J., and Cook D. R. Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nature Biotechnol.*, **31** (3), 240–246 (2013). DOI: 10.1038/nbt.2491
2. Smithson J. B., Thompson J. A., and Summerfield R. J. Chickpea (*Cicer Arietinum* L.). In *Grain Legume Crops*. Ed. by R.J. Summerfield and R.E. Roberts (Collins, London, UK, 1985), pp. 312–390.
3. Abbo Sh., Berger J., and Turner N. C. Evolution of cultivated chickpea: four bottlenecks limit diversity and constrain adaptation. *Funct. Plant Biol.*, **30**, 1081–1087 (2003). DOI: 10.1071/FP03084
4. Kumar J. and Abbo Sh. Genetics of flowering time in chickpea and its bearing on productivity in the semi-arid environments. *Adv. Agron.*, **72**, 107–138 (2001).
5. Roberts E. H., Hadley P., and Summerfield R. J. Effects of temperature and photoperiod on flowering in chickpeas (*Cicer Arietinum* L.). *Ann. Botany*, **55** (6), 881–892 (1985).
6. Ellis R. H., Lawn R. J., Summerfield R. J., Qi A., Roberts E. H., Chay P. M., Brouwer J. B., Rose J. L., Yeates S. J., and Sandover S. Towards the reliable prediction of time to flowering in six annual crops. iv. cultivated and wild mung bean. *Exp. Agricult.*, **30** (3), 271–282 (1994).
7. Upadhyaya H. D., Bajaj D., Das S., Saxena M. S., Bandoni S., Kumar V., Tripathi S., Gowda C. L., Sharma S., Tyagi A. K., and Parida S. K. A genome-scale integrated approach aids in genetic dissection of complex flowering time trait in chickpea. *Plant Mol. Biol.*, **89** (4), 403–420 (2015). DOI: 10.1007/s11103-015-0377-z
8. Kumar V., Singh A., Mithra S. V., Krishnamurthy S. L., Parida S. K., Jain S., Tiwari K. K., Kumar P.,

- Rao A. R., Sharma S. K., Khurana J. P., Singh N. K., and Mohapatra T. Genome-wide association mapping of salinity tolerance in rice (*Oryza Sativa*). *DNA Res.*, **22** (2), 133–145 (2015). DOI: 10.1093/dnares/dsu046
9. Berger J., Milroy S., Turner N., Siddique K., Imtiaz M., and Malhotra R. Chickpea evolution has selected for contrasting phenological mechanisms among different habitats. *Euphytica*, **180**, 1–15 (2011).
10. Zhang X. and Cai X. Climate change impacts on global agricultural land availability. *Environ. Res. Lett.*, **6** (1), 014014 (2011).
11. Soltani A., Hammer G. L., Torabi B., Robertson M. J., and Zeinali E. Modeling chickpea growth and development: phenological development. *Field Crops Res.*, **99** (1), 1–13 (2006).
12. Soltani A., Robertson M. J., Mohammad-Nejad Y., and Rahemi-Karizaki A. Modeling chickpea growth and development: leaf production and senescence. *Field Crops Res.*, **99** (1), 14–23 (2006b).
13. Jones J. W., Antle J. M., Basso B., Boote K. J., Conant R. T., Foster I., Godfray H. C. J., Herrero M., Howitt R. E., Janssen S., Keating B. A., Munoz-Carpena R., Porter Ch. H., Rosenzweig C., and Wheeler T. R. Toward a new generation of agricultural system data, models, and knowledge products: state of agricultural systems science. *Agricult. Syst.*, **155**, 269–288 (2017). DOI: 10.1016/j.agry.2016.09.021
14. Jones J. W., Antle J. M., Basso B., Boote K. J., Conant R. T., Foster I., Godfray H. C. J., Herrero M., Howitt R. E., Janssen S., Keating B. A., Munoz-Carpena R., Porter C. H., Rosenzweig C., and Wheeler T. R. Brief history of agricultural systems modeling. *Agricultural Systems*, **155**, 240–254 (2016). DOI: 10.1016/j.agry.2016.05.014
15. Jones J. W., Hoogenboom G., Porter C. H., Boote K. J., Batchelor W. D., Hunt L. A., Wilkens P. W., Singh U., Gijsman A. J., and Ritchie J. T. The DSSAT cropping system model. *Eur. J. Agronomy*, **18** (3–4), 235–265 (2003).
16. Boote J., Jones K. W., and Pickering N. B. Potential uses and limitations of crop models. *Agronomy J.*, **88**, 704–716 (1996).
17. Boote K. J., Jones J. W., White J. W., Asseng S., and Lizaso J. I. Putting mechanisms into crop production models: putting mechanisms into crop production models. *Plant, Cell & Environment*, **36** (9), 1658–72 (2013).
18. Keating B., Carberry P. S., Hammer G., Probert M. E., Robertson M. J., Holzworth D., Huth N. I., Hargreaves J. N. G., Meinke H., Hochman Z., McLean G., Verburg K., Snow V., Dimes J. P., Silburn M., Wang E., Brown S., Bristow K. L., Asseng S., Chapman S., McCown R. L., Freebairn D. M., and Smith C. J. An overview of APSIM, a model designed for farming systems simulation. *Eur. J. Agronomy*, **18**, 267–288 (2003). DOI: 10.1016/S1161-0301(02)00108-9
19. Battisti R., Sentelhas P. C., and Boote K. J. Sensitivity and requirement of improvements of four soybean crop simulation models for climate change studies in Southern Brazil. *Int. J. Biometeorol.*, **62** (5), 823–832 (2018).
20. Williams J. R., Jones C. A., Kiniry J. R., and Spalton D. A. The EPIC crop growth model. *Transactions of the ASAE*, **32** (2), 497–511 (1989).
21. Vadez V., Soltani A., and Sinclair T. R. Crop simulation analysis of phenological adaptation of chickpea to different latitudes of India. *Field Crops Res.*, **146**, 1–9 (2013).
22. Lal M., Singh K. K., Srinivasan G., Rathore L. S., Naidu D., and Tripathi C. N. Growth and yield responses of soybean in Madhya Pradesh, India to climate variability and change. *Agricult. Forest Meteorol.*, **93** (1), 53–70 (1999).
23. Chung U., Kim Y. U., Seo B. S., and Seo M. C. Evaluation of variation and uncertainty in the potential yield of soybeans in South Korea using multi-model ensemble climate change scenarios. *Agrotechnology*, **06** (02), 1000158 (2017). DOI: 10.4172/2168-9881.1000158
24. Mohammed A., Tana T., Singh P., Molla A., and Seid A. Identifying best crop management practices for chickpea (*Cicer arietinum* L.) in Northeastern Ethiopia under climate change condition. *Agricult. Water Management*, **194**, 68–77 (2017). DOI: 10.1016/j.agwat.2017.08.022
25. Patil D. and Patel H. R. Calibration and validation of CROPGRO (DSSAT 4.6) model for chickpea under Middle Gujarat agroclimatic region. *Int. J. Agricult. Sci.*, **9** (27), 4342–4344 (2017).
26. Urgaya M. L. Modeling the impacts of climate change on chickpea production in Adaa Woreda (East Showa Zone) in the semi-arid Central Rift Valley of Ethiopia. *J. Pet. Environ. Biotechnol.*, **7**, 288 (2016).
27. Ageev A., Aydogan A., Bishop-von Wettberg E., Nuzhdin S. V., Samsonova M., and Kozlov K. Simulation model for time to flowering with climatic and genetic inputs for wild chickpea. *Agronomy*, **11**, 1389 (2021).
28. von Wettberg E. J. B., Chang P. L., Başdemir F., Carrasquilla-Garcia N., Korbu L. B., Moenga S. M., Bedada G., Greenlon A., Moriuchi K. S., Singh V., Cordeiro M. A., Noujdina N. V., Dinegde K. N., Shah Sani S. G. A., Getahun T., Vance L., Bergmann E., Lindsay D., Mamo B. E., Warschefsky E. J., Dacosta-Calheiros E., Marques E., Yilmaz M. A., Cakmak A., Rose J., Migneault A., Krieg C. P., Saylak S., Temel H., Friesen M. L., Siler E., Akhmetov Z., Ozcelik H., Kholova J., Can C., Gaur P., Yildirim M., Sharma H., Vadez V., Tesfaye K., Woldemedhin A. F., Tar'an B., Aydogan A., Bukun B., Penmetsa R. V., Berger J., Kahraman A., Nuzhdin S. V., and Cook D. R. Ecology and genomics of an important crop wild relative as a prelude to agricultural innovation. *Nature Commun.*, **9** (1), 649 (2018). DOI: 10.1038/s41467-018-02867-z

29. Stackhouse P. W., Perez R., Sengupta M., Knapp K., Mikovitz J. C., Schlemmer J., Scarino B., Zhang T., and Cox S. J. An assessment of new satellite data products for the development of a long-term global solar resource at 10–100 km. In *Proc. of the Solar 2016 Conf.* (San Francisco, CA, USA, 2016), pp. 1–6.
30. van der Laan M. J., Polley E. C., and Hubbard A. E. Super learner. *Stat. Applic. in Genetics and Mol. Biology*, **6** (1), Article 25 (2007).
31. Vadez V., Berger J., Warkentin T., Asseng S., Ratnakumar P., Rao K., Gaur P., Munier-Jolain N., Larmure A., Voisin A.-S., Sharma H. C., Pande S., Sharma M., Krishnamurthy L., and Zaman-Allah M. Adaptation of grain legumes to climate change: a review. *Agronomy for Sustainable Development*, **32** (1), 31–44 (2012).
DOI: 10.1007/s13593-011-0020-6

Parameterization of a Model for Wild Chickpea Flowering Time by Transferring the Knowledge Learned from Multiple Sources

Z.A. Saranin*, M.G. Samsonova*, and K.N. Kozlov*

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

Building forecasting the flowering time helps researchers to create varieties with maximum efficiency and value under a changing climate. This paper proposes an algorithm for parameterization of the wild chickpea flowering time model by using machine learning through knowledge transfer to combine multiple input-target sets. The resulting model showed high accuracy based on genetic and climatic data on only the first 20 days after sowing — the average absolute error is slightly greater than 5 days, the Pearson correlation coefficient is 0.93. It was found that maximum and minimum temperatures have the strongest effect on the timing of flowering. At the same time, all weather factors by the 7–10 day from the date of sowing affect a solution of the model.

Keywords: climatic factors, chickpea, mathematical modeling, transfer learning