

РЕГРЕССИОННАЯ МОДЕЛЬ ВРЕМЕНИ ЦВЕТЕНИЯ
СТАРОМЕСТНЫХ СОРТОВ НУТА© 2019 г. К. Н. Козлов^{1, *}, М. Г. Самсонова¹, С. В. Нуждин^{1, 2}¹Санкт-Петербургский политехнический университет им. Петра Великого,
научно-исследовательская лаборатория Математическая биология и биоинформатика,
Институт прикладной математики и механики, Санкт-Петербург, 195251 Россия²Университет Южной Калифорнии, Программа молекулярной
и вычислительной биологии, Лос-Анжелес, 24105 США

*e-mail: kozlov_kn@spbstu.ru

Поступила в редакцию 01.12.2018 г.

После доработки 01.02.2019 г.

Принята к публикации 12.02.2019 г.

Разработаны регрессионные модели времени цветения староместных сортов нута из коллекции ВИР, собранных в Турции и Эфиопии. Предсказанное время цветения хорошо согласуется с опубликованными ранее данными. Различие моделей статистически значимо. Вклад температуры в модель для сортов из Эфиопии слабее, чем для сортов из Турции, 48 и 60% соответственно. Вклад осадков в модель составляет более 80%. Результаты согласуются с особенностями климата в местах сбора.

Ключевые слова: климатические факторы, нут, нелинейная регрессия, грамматическая эволюция, математическое моделирование.

DOI: 10.1134/S0016675819070099

Зернобобовые культуры, такие как нут, служат основным источником растительного белка. Нут, обладая высокими пищевыми и вкусовыми достоинствами, имеет и целый ряд биологических преимуществ перед другими культурами. Он отличается высокой засухоустойчивостью, жаро- и солеустойчивостью, устойчивостью к полеганию, осыпанию семян и вредителям, в частности такому злостному, как зерновка.

Время цветения является важным фенологическим признаком нута [1]. Точное его предсказание может помочь селекционерам в выборе родительских линий для создания сортов, оптимальных для конкретных регионов [2, 3].

Математическое моделирование стало одним из важнейших методов решения этой задачи. Стандартными методами моделирования фенологии являются расчет сумм температур выше температурного минимума и регрессионные модели [4–8]. Ранее нами был предложен метод построения нелинейных регрессионных моделей на основе грамматической эволюции, который был применен для моделирования фенологии сои [9].

Целью данного исследования было выявление различий реакции времени цветения сортов нута из коллекции ВИР, собранных в Турции и Эфиопии, на изменение агроклиматических факторов.

Образцы нута. Были рассмотрены агроклиматические данные, полученные в Сирии вблизи Алеппо для староместных сортов нута из коллекции ВИР, собранных в Турции и Эфиопии. Имеющиеся климатические данные представлены в табл. 1.

Математическая модель времени цветения. Модель времени цветения строится как линейная комбинация N управляющих функций, зависящих от агроклиматических признаков нелинейным образом. Таким образом модель имеет вид (1):

$$y_i = \beta_0 + \sum_{n=0}^{N-1} \beta_{n+1} F_n(X_i) + \varepsilon_i, \quad (1)$$

где y_i – моделируемый фенотип для каждого растения i , β_n – коэффициенты, X_i – вектор агроклиматических факторов и ε_i – стандартная ошибка. Отличительным признаком использованного подхода является автоматическое определение вида зависимости, что позволяет исследовать более широкий класс регрессионных функций.

Аналитический вид функции. В предыдущих работах были рассмотрены различные виды зависимости между фенотипом и климатическими факторами [10–14]. Нами предложен более общий подход, который позволяет исследовать широкий круг нелинейных зависимостей между фенотипом и факторами. Аналитический вид управляю-

Таблица 1. Климатические факторы

Символ	Описание	Турция	Эфиопия
		среднее \pm стандартное отклонение	
T^{mean}	Средняя годовая температура	130.12 \pm 30.59	173.30 \pm 22.57
ΔT^{mean}	Средний разброс температуры по месяцам	115.97 \pm 14.94	137.23 \pm 9.65
ΔT	Годовой разброс температуры	310.06 \pm 38.04	185.68 \pm 17.73
T^{sd}	Сезонность температуры (100 \times стандартное отклонение)	6899.28 \pm 1040.20	1172.47 \pm 381.10
p^{ann}	Годовые осадки	582.17 \pm 151.76	1002.39 \pm 255.42
p^{sd}	Сезонность осадков (100 \times стандартное отклонение)	53.60 \pm 19.88	95.84 \pm 21.37

Примечание. Значения всех признаков центрируются вычитанием среднего и делятся на модуль разности максимального и минимального значения. Пропущенные данные заменяются на средние значения соответствующего признака.

Таблица 2. Корреляции климатических факторов со временем цветения

Символ	Описание	Турция	Эфиопия
T^{mean}	Средняя годовая температура	0.104 ($p = 0.462$)	-0.236 ($p = 0.078$)
ΔT^{mean}	Средний разброс температуры по месяцам	-0.049 ($p = 0.731$)	0.026 ($p = 0.850$)
ΔT	Годовой разброс температуры	-0.061 ($p = 0.665$)	-0.077 ($p = 0.568$)
T^{sd}	Сезонность температуры (100 \times стандартное отклонение)	-0.043 ($p = 0.761$)	-0.192 ($p = 0.151$)
p^{ann}	Годовые осадки	-0.004 ($p = 0.975$)	0.172 ($p = 0.201$)
p^{sd}	Сезонность осадков (100 \times стандартное отклонение)	0.035 ($p = 0.708$)	-0.039 ($p = 0.771$)

Примечание. Все p -значения больше 0.05 (указаны в скобках).

шей функции определяется автоматически методом грамматической эволюции [15, 16] вместе с нахождением регрессионных коэффициентов методом LASSO [17] и набором предикторов путем стохастической минимизации отклонения решения модели от данных методом разностной эволюции [18]. При этом обучающая выборка составляла 75%, а ошибка рассчитывалась по неиспользованным при обучении данным.

В использованном подходе каждая модель характеризуется вектором из 19 чисел, 15 из которых описывают слова в грамматике, а 4, включая свободный член, являются коэффициентами построенной методом LASSO регрессионной модели. Таким образом, мы можем сравнить модели, сравнивая такие векторы.

Сравнение вкладов в модель различных факторов оценивается как отношение разницы сумм квадратов остатков без учета этого фактора и полной модели к сумме квадратов остатков без учета этого фактора.

Зависимость времени цветения от климатических факторов. В табл. 2 представлены значения коэффициента корреляции Пирсона между вре-

менем цветения и климатическими факторами. Статистически значимых зависимостей не обнаружено, все p -значения оказались значительно больше 0.05.

Таким образом, зависимость времени цветения от климатических факторов не может иметь простой вид.

Адаптация модели к данным. Была проведена серия численных экспериментов. Поскольку метод минимизации целевой функции является стохастическим, требуется проводить много запусков. Путем перебора было установлено, что лучшими значениями является $N = 3$ функций длиной $L = 15$. На последнем этапе было проведено по 20 запусков для данных староместных сортов, собранных в Турции и Эфиопии. На рис. 1 представлен типичный график сходимости процесса минимизации.

Староместные сорта, собранные в Эфиопии. Среди полученных результатов для староместных сортов, собранных в Эфиопии, была выбрана модель (2):

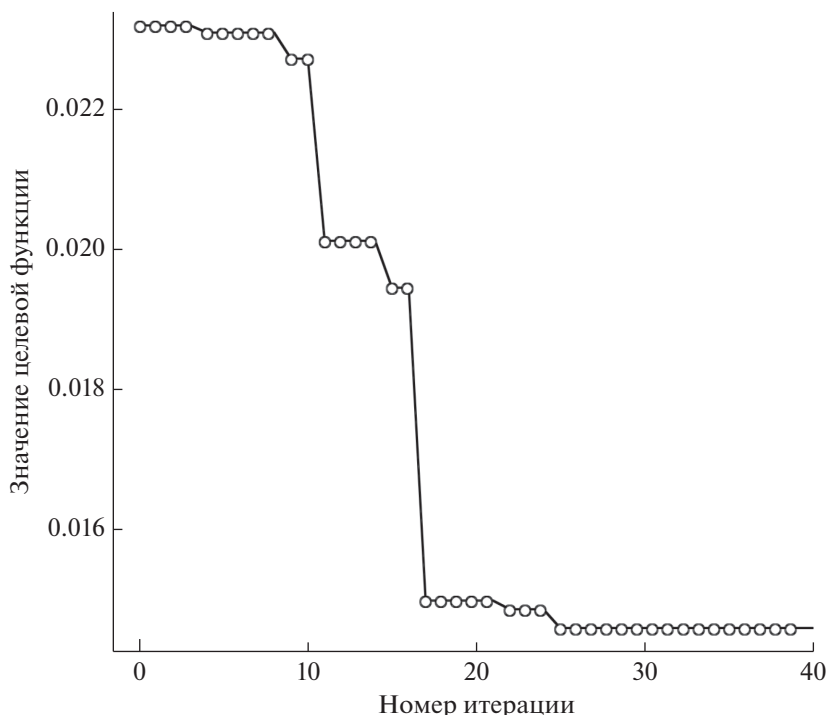


Рис. 1. Типичный график сходимости процесса минимизации методом ППРЭ.

$$F_E = 26.687 + 0.001\Delta T^{\text{mean}} - 0.013T^{\text{mean}} - 0.001T^{\text{sd}} + \frac{162.879(P^{\text{sd}} - 95.842)}{(T^{\text{sd}} - 1172.474)(T^{\text{mean}} - 219.798)}, \quad (2)$$

где T^{mean} — средняя годовая температура, ΔT^{mean} — средний разброс температуры по месяцам, T^{sd} — сезонность температуры ($100 \times$ стандартное отклонение), P^{sd} — сезонность осадков ($100 \times$ стандартное отклонение). Коэффициент детерминации равен $R^2 = 0.52$.

Староместные сорта, собранные в Турции. Среди полученных результатов для староместных сортов, собранных в Турции, была выбрана модель (3):

$$F_T = 29.602 - 0.004P^{\text{ann}} + \frac{20.667}{P^{\text{sd}} - 61.397} + \frac{10.449}{\Delta T^{\text{mean}} - 137.866}, \quad (3)$$

где P^{ann} — годовые осадки, ΔT^{mean} — средний разброс температуры по месяцам, P^{sd} — сезонность осадков ($100 \times$ стандартное отклонение). Коэффициент детерминации равен $R^2 = 0.45$.

Сравнение моделей. На рис. 2 представлено сравнение полученных предсказаний с данными для сортов из Эфиопии (рис. 2,а) и Турции (рис. 2,б). Коэффициент корреляции Пирсона между векторами, описывающими функции F_E и F_T , состав

вил 0.13 и оказался статистически незначим — $p\text{-value} = 0.38 > 0.05$, коэффициент корреляции Спирмана составил 0.15 и незначим, $p\text{-value} = 0.29 > 0.05$. Вклад температуры составил в модель для Турции 60%, для Эфиопии — 48%. Вклад осадков — 86 и 89%, для Турции и Эфиопии соответственно.

С помощью построенных нелинейных регрессионных моделей показано различие зависимости времени цветения от климатических факторов для староместных сортов, собранных в Эфиопии и Турции.

Особенностью модели для староместных сортов, собранных в Эфиопии, является наличие зависимости от сложной функции одновременно осадков и температуры. Вклад температуры в модель для сортов из Эфиопии слабее, чем для сортов из Турции, что может являться следствием того, что в Эфиопии среднегодовые температуры относительно постоянны и способность реагировать на их изменения была утрачена. Вклад осадков в модель составляет более 80%, хотя для сортов из Эфиопии на 3% сильнее, чем для Турции, что может объясняться чувствительностью сортов к влаге.

Работа выполнена в рамках проекта РНФ 16-16-00007.

Вычисления проводились в СКЦ “Политехнический” ФГАОУ ВО СПбПУ.

Авторы заявляют, что у них нет конфликта интересов.

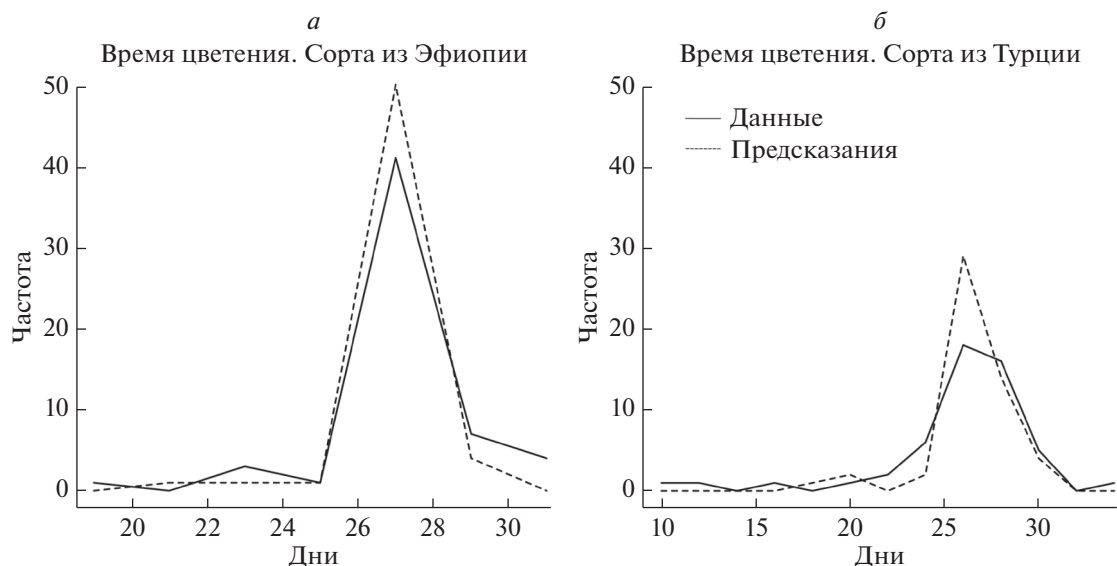


Рис. 2. Гистограммы времени цветения. Линии соединяют точки, ординаты которых соответствуют количеству измерений, попавших в интервал, центр которого совпадает с абсциссой этой точки. Предсказанное время цветения хорошо согласуется с данными.

Настоящая статья не содержит каких-либо исследований с использованием в качестве объекта животных.

Настоящая статья не содержит каких-либо исследований с участием в качестве объекта людей.

СПИСОК ЛИТЕРАТУРЫ

1. Berger J., Milroy S., Turner N. et al. Chickpea evolution has selected for contrasting phenological mechanisms among different habitats // *Euphytica*. 2011. V. 180. P. 1–15. doi 10.1007/s10681-011-0391-4
2. Abbo S., Berge, J., Turner N. Evolution of cultivated chickpea: Four bottlenecks limit diversity and constrain adaptation // *Functional Plant Biol.* 2003. V. 30. P. 1081–1087. doi 10.1071/FP03084
3. Kumar J., Abbo S. Genetics of flowering time in chickpea and its bearing on productivity in the semi-arid environments // *Adv. Agron.* 2001. V. 72. P. 107–138. doi 10.1038/35098564
4. Pedersen P., Boote K.J., Jones J.W., Lauer J.G. Modifying the CROPGRO-Soybean model to improve predictions for the upper midwest // *Agronomy J.* 2004. V. 96. P. 556–564. doi 10.2134/agronj2004.0556
5. Setiyono T.D., Weiss A., Specht J. et al. Understanding and modeling the effect of temperature and daylength on soybean phenology under high-yield conditions // *Field Crops Res.* 2007. V. 100. P. 257–271. doi 10.1016/j.fcr.2006.07.011
6. Major D.J., Johnson D.R., Tanner J.W., Anderson I.C. Effects of daylength and temperature on soybean development // *Crop Sci.* 1975. V. 15. P. 174–179. doi 10.2135/cropsci1975.0011183X001500020009x
7. Soltani A., Hammer G.L., Torabi B. et al. Modeling chickpea growth and development: Phenological development // *Field Crops Res.* 2006. V. 99. P. 1–13. doi 10.1016/j.fcr.2006.02.005
8. Vadez V., Soltani A., Sinclair T.R. Modelling possible benefits of root related traits to enhance terminal drought adaptation of chickpea // *Field Crops Res.* 2012. V. 137. P. 108–115. doi 10.1016/j.fcr.2012.07.022
9. Козлов К.Н., Новикова Л.Ю., Сеферова И.В., Самсонова М.Г. Математическая модель влияния климатических факторов на развитие сои // *Биофизика*. 2018. Т. 63. Вып. 1. С. 175–176.
10. Hammer G.L., Vaderlip R.L., Gibson G. et al. Genotype-by-environment interaction in grain sorghum. II. Effects of temperature and photoperiod on ontogeny // *Crop Sci.* 1989. V. 29. P. 376–384. doi 10.2135/cropsci1989.0011183X002900020029x
11. Horie T. Crop ontogeny and development // *Physiology and Determination of Crop Yield / Eds Boote K.J., Bennett J.M., Sinclair T.R., Paulsen G.M. Madison, 1994. P. 153–180.*
12. Piper E.L., Boote K.J., Jones J.W., Grimm S.S. Comparison of two phenology models for predicting flowering and maturity date of soybean // *Crop Sci.* 1996. V. 36. P. 1606–1614. doi 10.2135/cropsci1996.0011183X003600060033x
13. Yin X., Kropff M.J., Horie T. et al. A model for photo-thermal responses of flowering in rice. I. Model description and parameterization // *Field Crops Res.* 1997. V. 51. P. 189–200. doi 10.1016/S0378-4290(96)03456-9
14. Robertson M.J., Watkinson A.R., Kirkegaard J.A. et al. Environmental and genotypic control of time to flowering in canola and Indian mustard // *Australian J. Agricultural. Res.* 2002. V. 53. P. 793–809. doi 10.1071/AR01182

15. *O'Neill M., Ryan C.* Grammatical evolution // IEEE Transactions on Evolutionary Computation. 2001. V. 5. P. 349–358. doi 10.1109/4235.942529
16. *Noorian F., de Silva A.M., Leong P.H.W.* gramEvol: Grammatical Evolution in R // J. Statist. Software. 2016. V. 71. P. 1–26. doi 10.18637/jss.v071.i01
17. *Tibshirani R.* Regression shrinkage and selection via the LASSO // J. Royal Statistical Soc. Ser. B. 1996. V. 58. P. 267–288.
18. *Kozlov K., Samsonov A.M., Samsonova M.* A software for parameter optimization with Differential Evolution Entirely Parallel method // Peer J. Computer Sci. 2016. V. 2. P. E74 (2016). doi 10.7717/peerj-cs.74

Regression Model for Time to Flowering of Chickpea Landraces

K. N. Kozlov^{a,*}, M. G. Samsonova^a, and S. V. Nuzhdin^{a,b}

^a*Peter the Great St. Petersburg Polytechnic University, Mathematical Biology and Bioinformatics Laboratory, IAMM, St. Petersburg, 195251 Russia*

^b*University of Southern California, Molecular and Computation Biology Program, Los Angeles, 24105 USA*

**e-mail: kozlov_kn@spbstu.ru*

Regression models for time to flowering had been developed for VIR chickpea landraces collected in Turkey and Ethiopia. Predicted flowering time coincides closely with experimental data. The difference between models is statistically significant. The impact of the temperature to the model for Ethiopia landraces is weaker than that for Turkey landraces, 48 and 60%, respectively. The impact of precipitation on phenotype is more than 80%. The obtained results are in agreement with climate characteristics at collection sites.

Keywords: climatic factors, chickpea, non-linear regression, grammatical evolution, mathematical modeling.