

СЕТЕВАЯ МАРКОВСКАЯ МОДЕЛЬ ПРОГНОЗИРОВАНИЯ ВЕСА ТЫСЯЧИ СЕМЯН У ГЕНОТИПОВ НУТА

© 2025 г. Д.Д. Мальцов*, 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 14.10.2024 г.

После доработки 14.10.2024 г.

Принята к публикации 30.10.2024 г.

Построение прогноза признаков, связанных с урожайностью, таких, например, как вес тысячи семян, позволяет исследователям создавать сорта, достигающие максимальной эффективности и ценности в условиях изменения климата. В данной работе предложена сетевая марковская модель прогнозирования важного фенотипического признака «вес тысячи семян» у генотипов нута, которая использует предварительно отобранные сніпы и погодные данные за 5 дней до и 20 дней после посева, такие как минимальная и максимальная температуры, количество осадков, влажность, инфракрасное излучение и длина светового дня. Построенная модель предсказывает признак «вес тысячи семян» с высокой точностью — коэффициент корреляции Пирсона составляет 0.83.

Ключевые слова: вес тысячи семян, климатические факторы, нут, математическое моделирование, сетевая марковская модель.

DOI: 10.31857/S0006302925010173, EDN: LUKWOO

Математическое моделирование зависимости фенотипических признаков от генетических и погодных факторов является важным инструментом совершенствования селекции зернобобовых культур, поскольку дает возможность строить прогнозы поведения растений в изменяющейся среде. Нут (*Cicer arietinum* L.) богат пищевым белком и занимает второе место по распространенности среди зернобобовых культур благодаря своей питательной ценности и способности фиксировать азот, тем самым повышая плодородие почв [1]. Совершенствование сортов сельскохозяйственных культур является важным условием для обеспечения продовольственной безопасности растущего населения земного шара. Дикорастущий предок культурного нута является растением длинного дня и проявляет чувствительность к яровизации [2, 3], в отличие от культивируемых генотипов, пригодных для весеннего посева [3, 4].

Факторами, влияющими на продуктивность генотипов нута, являются температурный режим, продолжительность светового дня, доступность воды [2, 5].

Разработаны несколько математических моделей для прогнозирования фенотипических признаков нута в зависимости от погодных условий — SSM [6, 7], DSSAT [8–12], APSIM [13] и др. [14, 15]. Для описания процессов фотосинтеза, погло-

щения воды и других биофизических и биохимических реакций используются дифференциальные уравнения, которые решаются с шагом по времени, равным 1 суткам либо 1 часу для быстрых процессов. Адаптация существующих моделей к новым сортам или погодным условиям требует значительных усилий и времени [16–21]. Таким образом, разработка новых гибких моделей, которые легко могут адаптироваться к новому сорту, остается актуальной задачей.

В данной работе предложена сетевая марковская модель прогнозирования важного фенотипического признака TSW (thousand seed weight — вес тысячи семян) у генотипов нута, которая использует предварительно отобранные сніпы (однонуклеотидные полиморфизмы) и погодные данные за 5 суток до и 20 суток после посева, такие как минимальная и максимальная температуры, количество осадков, влажность, инфракрасное излучение и длина светового дня.

СЕТЕВАЯ МАРКОВСКАЯ МОДЕЛЬ

Сетевую марковскую (англ. Markov Brain) модель можно считать классом нейронной сети [22]. От классической нейронной сети данная модель отличается тем, что, в отличие от слоев нейронной сети, в котором каждый узел выполняет одну и ту же функцию, марковская сетевая модель

состоит из отдельных компонент, называемых вентилями, каждый из которых может вести себя по-разному. Модель состоит из буфера данных, называемых ячейками модели, и вентилях, которые принимают значения из одних ячеек и возвращают в другие. Одной из особенностей данной модели является возможность моделировать процесс во времени. Так, модель последовательно получает на вход данные из разных временных промежутков. Вентили могут быть двух типов: детерминированными или вероятностными. Поведение вентиля определяется логической таблицей, которая устанавливает правила получения выходных значений из входных, для вероятностных вентилях определяются вероятности конкретных выходов.

Программная реализация разработана на основе пакета с открытым исходным кодом [23], в который были внесены модификации для определения типа вентиля в оптимизируемых параметрах, благодаря чему можно использовать разные типы вентилях в одной модели, а также было запрещено чтение из выходных ячеек, чтобы «подтолкнуть» вентили к более частому использованию памяти. Оптимизация сетевой марковской модели производилась с помощью генетического алгоритма, реализованного в пакете *pygad* [24]. Программный код разработки доступен по адресу <https://gitlab.com/dmitry-maltsov/mb-chickpea>.

После того как модель построена, происходит череда активаций модели. При каждой активации во входные ячейки записываются входные данные — в рамках данной работы генетические маркеры и погодные факторы. Ввиду того, что модель оперирует только битами, исходные данные категоризируются. Так, вместо подачи на вход снипа *Ca4_11241316_G_A* со значением 2, означающим, что замена произошла в обеих хромосомах, подаются три значения — *False*, *False*, *True*. Первое значение обозначает что утверждение «*Ca4_11241316_G_A* равно 0» ложное, второе, что «*Ca4_11241316_G_A* равно 1» тоже ложное, и, наконец, третье, что «*Ca4_11241316_G_A* равно 2» правдиво. Данные о погоде категоризируются другим образом. Каждый погодный фактор разбивается на 5 равных интервалов и каждое булево значение, подающееся на вход, отвечает на вопрос, выше ли реальное значение погодного фактора левой границы соответствующего интервала». Например, значения разбиения для *T2M_MAX* равны 19.35, 20.88, 22.4, 23.92 и 25.45, а значение *T2M_MAX* в двадцать пятые сутки, равное 22.04, представляется в виде *True*, *True*, *False*, *False*, *False* (больше 19.35, больше 20.88, не больше 22.4, не больше 23.92, не больше 25.45). После того, как записаны входные данные, происходит активация вентилях — они читают данные по своим входным индексам (из памяти или

из входных ячеек) и записывают значения по своим выходным индексам (в память или в выход модели), руководствуясь своими таблицами. При следующей активации данные о погоде заменятся на данные следующего дня, генетические данные не меняются. Всего происходит 25 активаций, по одной активации на каждый день. После прохождения всего цикла активаций значения, получившиеся в выходных ячейках, считаются окончательным выходом модели и определяют предсказание фенотипа нута.

ОБРАЗЦЫ РАСТЕНИЙ И ПОЛЕВЫЕ ИССЛЕДОВАНИЯ

407 образцов нута *Cicer arietinum*, собранных в Эфиопии, Ливане, Марокко, Турции, Индии, Узбекистане и Средиземноморском регионе, были фенотипированы на Кубанской опытной станции ВИР в 2016 г. В течение вегетационного периода были измерены 36 фенологических, морфологических, агрономических и биологических дескрипторов. Подробная информация об экспериментах по фенотипированию, генотипированию и последующем анализе, в результате которых было получено 6642 снипа, была представлена в работе [25].

В этой работе создана модель признака продуктивности нута, а именно — массы тысячи семян (TSW); этот признак варьируется от 110 до 440 г.

ОТБОР СНИПОВ

Так как 6642 снипа — это слишком много для построения модели, был проведен их отбор алгоритмом градиентного бустинга, а точнее — его реализацией XGBoost [26]. Градиентный бустинг представляет собой метод машинного обучения, применяемый для решения задач классификации и регрессии. Он основан на генерации случайного леса — ансамбля слабых моделей для предсказания деревьев решений. На каждом этапе вычисляются ошибки предсказаний уже существующего ансамбля на обучающей выборке. Следующая модель, добавляемая в ансамбль, будет предсказывать эти ошибки. Добавляя предсказания нового дерева к предсказаниям существующего ансамбля, мы уменьшаем среднюю ошибку модели. Новые деревья продолжают добавляться в ансамбль до тех пор, пока либо снижается ошибка, либо пока не достигнут критерий остановки.

Оптимизацию градиентного бустинга проводили с использованием *k*-fold кросс-валидации с *n* повторами. Параметры кросс-валидации: *k* = 10, *n* = 3. Помимо этого, на каждой итерации процесса кросс-валидации применялся поиск по сетке оптимальных параметров случайного леса — числа деревьев *m* и максимальной глубины отдельного дерева *d*.

Таблица 1. Снимки, отобранные с помощью модели градиентного бустинга

TSW	Важность
Ca1_34754596_T_A	0.027955214
Ca2_6085175_G_A	0.061373600
Ca2_6419716_G_T	0.077856764
Ca4_3013266_G_T	0.019178191
Ca4_11240892_C_T	0.029107153
Ca4_12574358_T_A	0.038410958
Ca6_9039600_T_C	0.138557690
Ca6_18267475_C_T	0.023114156
Ca6_42759989_A_G	0.256156770
Ca7_5448007_T_C	0.028858041

В результате была получена модель с коэффициентом детерминации $R^2 = 0.65$. Основываясь на вычислении примеси Джини как важности предиктора, отобрали 10 снимков для моделирования TSW (табл. 1).

ПРОГНОЗИРОВАНИЕ ВЕСА ТЫСЯЧИ СЕМЯН У ГЕНОТИПОВ НУТА

Построение модели проводили с помощью процесса кросс-валидации, совмещенного с поиском оптимальных гиперпараметров по сетке, в данном случае подбирались оптимальные параметры типа скрещивания и стратегии отбора родителей в генетическом алгоритме.

На вход подавали все погодные факторы и 10 отобранных снимов. Параметры сетевой марковской модели:

- Число входных ячеек 60 (30 для однонуклеотидных полиморфизмов и 30 для погоды, по 5 для каждого фактора);
 - Число ячеек памяти 30;
 - Число выходных ячеек 16;
 - Выход модели переводится в целое число в десятичной системе и интерпретируется как прогнозируемый фенотип.
- Параметры генетического алгоритма:
- Размер популяции 50;
 - Число поколений 50;
 - Длина кодирующей модель последовательности 50000 чисел.

В результате была получена сетевая марковская модель (рис. 1), достаточно точно описывающая имеющиеся данные (см. рис. 2). Значение

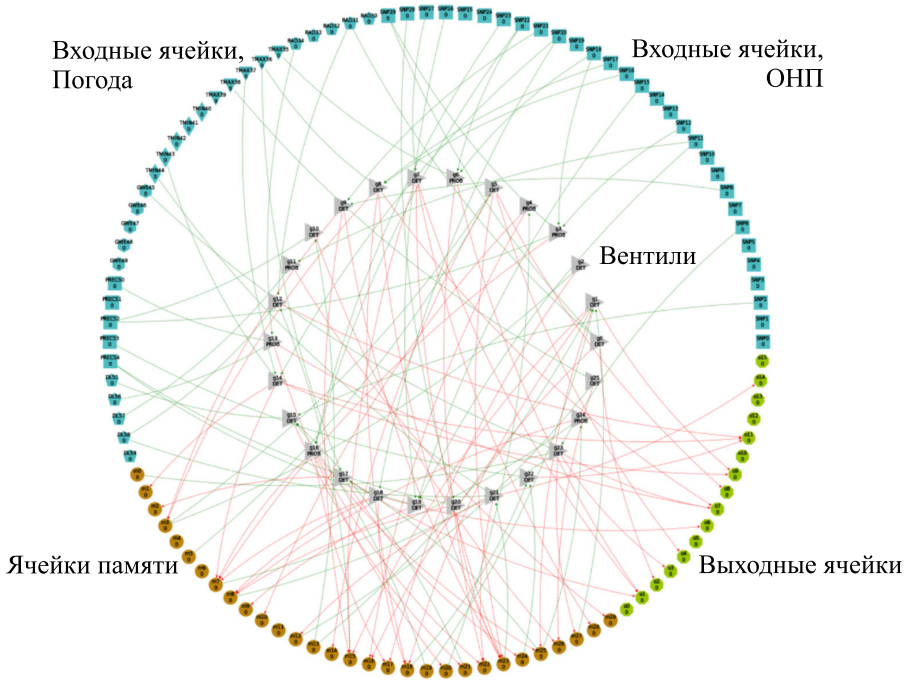


Рис. 1. Схема сетевой марковской модели. Входные ячейки для погодных данных о длине дня (DL), осадках (PREC), влажности (GWE), минимальной (TMIN) и максимальной (TMAX) температурах, инфракрасном излучении (RAD), а также для снимов (SNP) обозначены 5-угольником, квадратом, 6-угольником, ромбом, 5-угольником, квадратом соответственно. Венти́ли (g), ячейки памяти (m) и выходные ячейки (o) — треугольниками, кругами и 8-угольниками соответственно. Стрелки показывают направление передачи информации. Название ячейки погодных данных включает порог категоризации фактора. В названии вентиля DET и PROB обозначают детерминированный и вероятностный соответственно. Название вентиля, входных и выходных ячеек включает порядковый номер объекта данного типа. Ноль в строке 2 подписи ячейки соответствует начальному состоянию модели до активации.

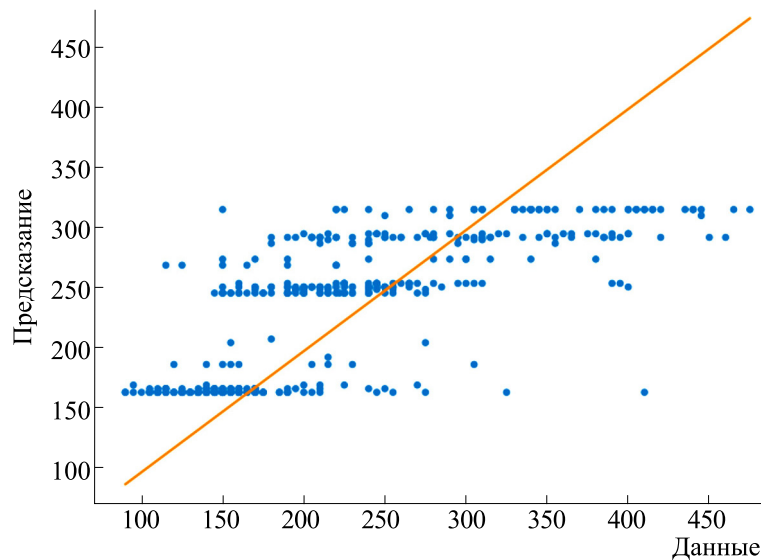


Рис. 2. Сравнение решения модели с данными для обучающей и тестовой выборки. Каждая точка соответствует одному экземпляру, по оси абсцисс отложено измеренное значение признака TSW, по оси ординат — решение модели. Для наглядности сплошной линией показана биссектриса угла.

функции ошибки на тестовой выборке составило 2872.93 с коэффициентом корреляции Пирсона между решением и тестовыми данными, равным 0.83, $P = 1.9 \cdot 10^{-11}$.

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 метод пермутации, в котором значения оцениваемого фактора переставляются между наблюдениями. Важность фактора оценивается по изменению ошибки решения модели.

Согласно полученным результатам, можно сделать вывод о том, что максимальное влияние

оказывает инфракрасное излучение и минимальная температура, для которых среднее изменение ошибки по 100 пермутациям составило 294.996 и 130.797 соответственно, при небольших значениях дисперсии — 1.63 и 0.95 соответственно (табл. 2).

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Использование математических моделей для прогнозирования хозяйственно ценных характеристик растений, таких как связанный с продуктивностью вес тысячи семян, закладывает научную основу селекционного улучшения сортов в условиях глобальных изменений климата.

В данной работе с использованием генетического алгоритма и кросс-валидации была построена сетевая марковская модель в составе 26 вентелей, из них 7 вероятностные, 60 входных ячеек, 30 ячеек памяти и 16 выходных ячеек. Расчеты по имеющимся экспериментальным данным показали высокую точность моделирования.

Полученные результаты о влиянии погодных факторов, в целом, 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 [29].

ВЫВОДЫ

Применение экстремального градиентного бустинга для отбора заданного числа важных одонуклеотидных полиморфизмов позволило разработать сетевую марковскую модель для

Таблица 2. Важность погодных факторов

Фактор	Среднее изменение ошибки	σ^2
Мин. температура	130.797	0.95
Макс. температура	1.103	1.29
Кол-во осадков	11.302	2.23
Влажность почвы	1.002	1.04
Инфракрасное излучение	294.996	1.63
Длина св. дня	1.671	0.63

прогнозирования признака TSW нута с учетом погодных факторов для 5 до и 20 дней после посева.

Точность полученной модели характеризуется коэффициентом корреляции Пирсона между решением и данными, равным 0.83.

Показано, что солнечное излучение и минимальная температура оказывают наиболее сильное влияние на прогнозирование признака TSW. Полученные результаты в общем согласуются с ранее опубликованными данными.

ФИНАНСИРОВАНИЕ РАБОТЫ

Работа выполнена в рамках гранта РНФ 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., Penmetza 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, Lond., UK, 1985), pp. 312–390.
3. Shahal A., Berger J., and Turner N. Evolution of cultivated chickpea: four bottlenecks limit diversity and constrain adaptation. *Function. Plant Biol.*, **30** (10), 1081–1087 (2003). DOI: 10.1071/FP03084
4. Kumar J. and Abbo S. Genetics of flowering time in chickpea and its bearing on productivity in the semi-arid environments. *Adv. Agron.*, **72**, 107–138 (2001). DOI: 10.12691/wjar-4-1-1
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. 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). DOI: 10.1016/j.fcr.2006.02.004
7. Soltani A., Robertson M. J., Mohammad-Nejad Y., and Rahemi-Karizaki A. Modeling chickpea growth and development: Leaf production and senescence. *Field Crops Research*, **99** (1), 14–23 (2006).
8. 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. Systems*, **155**, 269–288 (2017). DOI: 10.1016/j.agry.2016.09.021
9. Jones J. W., Antle J. M., Basso B., Boote K. J., Conant R. T., Foster I., Charles H., Godfray J., Herrero M., Howitt R. E., Janssen S., Keating B. A., Munoz-Carpena R., Porter Ch. H., Rosenzweig C., and Wheeler T. R. Brief history of agricultural systems modeling. *Agricult. Systems*, **155**, 240–254 (2016). DOI: 10.1016/j.agry.2016.05.014
10. Jones J. W., Hoogenboom G., Porter C. H., Boote K. J., Batchelor W. D., Hunt L. A., Wilkens P. W., Singh U., Gijssman A. J., and Ritchie J. T. The DSSAT cropping system model. *Eur. J. Agronomy*, **18** (3–4), 235–265 (2003). DOI: 10.1016/S1161-0301(02)00107-7
11. Boote K. J., Jones J. W., and Pickering N. B. Potential uses and limitations of crop models. *Agron. J.*, **88**, 704–716 (1996). DOI: 10.2134/Agronj1996.00021962008800050005X
12. 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 Environ.*, **36** (9), 1658–72 (2013). DOI: 10.1111/pce.12119
13. 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., and Smith C. J. An overview of APSIM, a model designed for farming systems simulation. *Eur. J. Agronomy*, **18** (3–4), 267–288 (2003). DOI: 10.1016/S1161-0301(02)00108-9
14. 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). DOI: 10.1007/s00484-017-1483-1
15. Williams J. R., Jones C. A., Kiniry J. R., and Spindel D. A. The EPIC crop growth model. *Trans. ASAE*, **32** (2), 497–511 (1989).
16. 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). DOI: 10.1016/S0168-1923(98)00105-1
17. 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*, **6** (2), 1000158 (2017). DOI: 10.4172/2168-9881.1000158
18. 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 cli-

- mate change condition. *Agricultural Water Management*, **194**, 68–77 (2017). DOI: 10.1016/j.agwat.2017.08.022
19. 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).
20. Mengesha U. 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). DOI: 10.4172/2157-7463.1000288
21. 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). DOI: 10.3390/agronomy11071389
22. Hintze A., Edlund J. A., Olson R. S., Knoester D. B., Schossau J., Albantakis L., Tehrani-Saleh A., Kvam P., Sheneman L., Goldsby H., Bohm C., and Adami Ch. Markov brains: A technical introduction. *arXiv*, 1709.05601 (2017). DOI: 10.48550/arXiv.1709.05601
23. Olson R. S., Knoester D. B., and Adami Ch. Evolution of swarming behavior is shaped by how predators attack. *Artificial Life*, **22** (3), 299–318 (2016). DOI: 10.1162/ARTL_a_00206
24. Gad A. F. PyGAD: An intuitive genetic algorithm python library. *Multimedia Tools and Applications*, **83**, 58029–58042 (2023). DOI: 10.1007/s11042-023-17167-y
25. Sokolkova A., Bulyntsev S. V., Chang P. L., Carrasquilla-Garcia N., Igolkina A. A., Noujdina N. V., von Wettberg E., Vishnyakova M. A., Cook D. R., Nuzhdin S. V., and Samsonova M. G. Genomic analysis of Vavilov's historic chickpea landraces reveals footprints of environmental and human selection. *Int. J. Mol. Sci.*, **21** (11), 3952 (2020). DOI: 10.3390/ijms21113952
26. Chen T. and Guestrin C. XGBoost: A scalable tree boosting system. In: *Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining* (ACM, San Francisco, USA, 2016), pp. 785–794. DOI: 10.1145/2939672.2939785
27. 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). DOI: 10.1016/j.fcr.2013.03.005
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 Ch. P., Saylak S., Temel H., Friesen M. L., Siler E., Akhmetov Zh., 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), (2018). DOI: 10.1038/s41467-018-02867-z
29. Vadez V., Berger J. D., Warkentin T., Asseng S., Ratnakumar P., Chandra Rao K. P., Gaur P. M., Munier-Jolain N., Larmure A., Voisin A.-S., Sharma H. C., Pande S., Sharma M., Krishnamurthy L., and Zaman A. M. Adaptation of grain legumes to climate change: A review. *Agron. Sustain. Dev.*, **32** (1), 31–44 (2012). DOI: 10.1007/s13593-011-0020-6

Markov Network Model for Predicting Thousand Seed Weight in Chickpea Genotypes

D.D. Maltsov*, M.G. Samsonova*, and K.N. Kozlov*

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

Predicting yield-related traits such as thousand seed weight (TSW) allows researchers to develop varieties that achieve maximum efficiency and value under changing climate conditions. In this paper, we propose a Markov network model for predicting the important phenotypic trait TSW in chickpea genotypes using pre-selected single nucleotide polymorphisms and weather data for 5 days before and 20 days after sowing, such as minimum and maximum temperatures, precipitation, humidity, infrared radiation, and daylength. The constructed model predicts the TSW trait with high accuracy – the Pearson correlation coefficient is 0.83.

Keywords: thousand seed weight, climate factors, chickpea, mathematical modeling, Markov network model