

ИНФОРМАЦИОННЫЕ
ТЕХНОЛОГИИ
И ТЕЛЕКОММУНИКАЦИИ

INFORMATION
TECHNOLOGIES
AND TELECOMMUNICATIONS

УДК 57.084.1:632.08

DOI: 10.17212/2782-2001-2026-2-69-82

Метод подготовки данных для фенотипирования проростков пшеницы различных сортов на примере сорта «Новосибирская 41»*

Г.В. СЕРОКЛИНОВ^{1,a}, А.В. ГУНЬКО^{2,b}

¹ РФ, 630501, Новосибирская область, р. п. Краснообск, ул. Центральная, 1, Сибирский федеральный научный центр агробиотехнологий Российской академии наук

² РФ, 630073, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет

^a seroklinov@mail.ru ^b gun@ait.cs.nstu.ru

В работе обсуждаются действия при подготовке экспериментальных данных, полученных при исследовании биопотенциалов проростков пшеницы сорта «Новосибирская 41» при воздействии на них повышенной и пониженной температуры, для последующего фенотипирования этих растений с использованием кластерного анализа. Отмечено, что такая подготовка необходима при длительных экспериментальных исследованиях (до 10 дней и более), в течение которых происходят метаболические изменения образцов проростков, влияющие на величину их биопотенциалов. Работа основана на экспериментальных данных, полученных в 2020 и 2022 годах, и их регрессионном анализе. Кратко описаны результаты мониторинга изменения биопотенциалов проростков в зависимости от их возраста и приведен алгоритм расчета корректирующих величин биопотенциалов для каждого возраста исследуемых объектов. Сформированы два блока данных для оценки фенотипа объектов: исходный, полученный в результате первичной обработки реализаций изменения биопотенциалов проростков при воздействии повышенной и пониженной температуры, и откорректированный относительно параметра \hat{U}_{\max} и максимального фильтрованного центрированного значения cf биопотенциалов проростков пшеницы при этих воздействиях. Проведено фенотипирование растений на основе этих блоков данных с применением оригинальной программы Eclaster, реализующей метод спектральной кластеризации из библиотеки sklearn.cluster среды программирования Python. Представленные результаты кластеризации в виде диаграмм рассеивания показывают более качественное разделение по кластерам для откорректированных данных.

Ключевые слова: математическая модель, экспериментальные данные, биопотенциал, фенотипирование, регрессионный анализ данных, кластерный анализ, температура, возраст растений, алгоритм коррекции

* Стаття получена 10 октября 2025 г.

Исследование выполнено при государственном финансировании в соответствии с планом работ СФНЦА РАН по теме FNUU-2024-0004.

ВВЕДЕНИЕ

Для полной оценки фенома сельскохозяйственных растений необходим значительный объем разнообразных данных, которые формируются с использованием различных технических средств, предусмотренных используемой технологией [1–3]. Полученное в результате огромное количество разнообразных данных требует дополнительной оптимизации, после которой проведение автоматического интеллектуального анализа позволяет сделать из огромного количества измерений биологически значимые выводы [1, 4]. Для этого применяются различные методы предварительной обработки экспериментальных данных в соответствии с их видом и формой [5]. В отдельных случаях известны методы преобразования в адекватные диаграммы для визуализации полученных эффектов [6–8]. Известен также ряд математических методов, позволяющих обрабатывать экспериментальные данные для их дальнейшего использования, которые описаны в источниках [9, 10].

Получение экспериментальных данных при фенотипировании злаковых культур на платформе измерения биопотенциалов растений на начальной стадии их развития (проростков), как и при использовании других платформ [11, 12], сопряжено с созданием специализированных технических средств для измерения и накопления экспериментальных данных потенциалов действия (ПД), а также средств формирования возмущающих воздействий, вызывающих процессы изменения биопотенциалов. Однако сформированные в результате одинаковых адиабатических процессов электрические сигналы для разных растений одного вида могут существенно отличаться в зависимости от возраста растений (особенно для их проростков в первые несколько дней их развития). Это утверждение было экспериментально доказано при проведении суточного мониторинга изменения биопотенциалов проростков пшеницы сорта «Новосибирская 41» при воздействии на них повышенной и пониженной температуры [13]. В результате приведенных в работе [13] исследований были выявлены регрессионные зависимости изменения биопотенциала проростков пшеницы сорта «Новосибирская 41» от их возраста и воздействия повышенной и пониженной температуры. Эти зависимости были нормированы по уровню биопотенциала первого эксперимента (начального возраста проростков), что позволяет рассчитать величину коррекции для каждого возраста проростков, участвующих в экспериментах. Таким образом, можно сформировать блок данных, необходимых для более качественного фенотипирования пшеницы сорта «Новосибирская 41».

1. ПОСТАНОВКА ЗАДАЧИ

В результате проведенных исследований суточного мониторинга проростков пшеницы сорта «Новосибирская 41», приведенных в работе [13], были получены нормализованные усредненные зависимости скорости изменения сигнала биопотенциала:

– при воздействии повышенной температуры

$$\Delta_T = -0,0913 + 0,051t; \quad (1)$$

– при воздействии пониженной температуры

$$\Delta_x = -0,08195 + 0,00466t, \quad (2)$$

где Δ_T и Δ_x – нормализованные усредненные значения скорости изменения сигнала биопотенциала при воздействии повышенной и пониженной температуры; t – возраст проростка.

Уравнения (1) и (2) получены на основе регрессионных зависимостей, построенных по результатам двух экспериментов по суточному мониторингу изменения величины биопотенциала проростков пшеницы сорта «Новосибирская 41» при воздействии пониженной и повышенной температуры, проведенных в 2020 и 2022 годах. На основе этих нормированных выражений нужно рассчитать нормализованные значения корректирующих величин для экспериментально полученных данных биопотенциалов проростков пшеницы в более поздний период, чем данные предыдущих экспериментов, а затем определить физическую величину коррекции. Так, если при проведении первого эксперимента возраст растения был 10 суток, то коррекции подвергаются экспериментальные данные, полученные для растений возраста 12 и более суток.

2. РАЗРАБОТКА АЛГОРИТМА КОРРЕКЦИИ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Согласно данным, приведенным в работе [13], первая серия экспериментов была проведена с проростками возраста 10 дней, а следующие серии – с проростками возраста 12, 14, 16 и 18 дней. Для определения нормализованного значения отклонения биопотенциала для каждого из указанных возрастов вычислим определенный интеграл из уравнения (1) для экспериментов при воздействии повышенной температуры и из уравнения (2) для экспериментов при воздействии пониженной температуры. Пример вычисления нормализованного значения отклонения для экспериментальных данных, полученных для проростков пшеницы сорта «Новосибирская 41» возраста 12 суток, приведен в выражении (3):

$$\begin{aligned} \delta_{Т.Н} &= \int_{10}^{12} (-0,0913 + 0,00516t) dt = \\ &= -0,0913t_{12} + 0,00516 \frac{t_{12}^2}{2} + 0,0913t_{10} - 0,00516 \frac{t_{10}^2}{2} = -0,069. \end{aligned} \quad (3)$$

Вычисление нормализованного значения отклонения для экспериментальных данных, полученных для проростков пшеницы сорта «Новосибирская 41» возраста 12 суток при воздействии пониженной температуры, приведено в выражении (4):

$$\begin{aligned} \delta_{x.Н} &= \int_{10}^{12} (-0,08195 + 0,00466t) dt = \\ &= -0,08195t_{12} + 0,00466 \frac{t_{12}^2}{2} + 0,08195t_{10} - 0,00466 \frac{t_{10}^2}{2} = -0,07138. \end{aligned} \quad (4)$$

Вычисленные таким образом нормализованные значения отклонений для экспериментальных данных, полученных при воздействии повышенной температуры для растений других возрастов, приведены в табл. 1.

Таблица 1

Table 1

Значения отклонений для экспериментальных данных, полученных при воздействии повышенной температуры на проростки пшеницы сорта «Новосибирская 41»

Deviation values for experimental data obtained under the influence of an elevated temperature for Novosibirsk 41 wheat seedlings

Время (сутки)	10	12	14	16	18
$U_{\max \text{ т.ц.}}, (\text{В})$	0,02571	0,022164	0,015623	0,0140419	0,013049
$\delta_{\text{т.н}}$	–	–0,069	–0,11752	–0,14532	–0,15248
$\delta_{\text{т}}, (\text{В})$	–	0,005198	0,00885	0,01095	0,01149

Вычисленные по выражению (4) нормализованные значения отклонений для экспериментальных данных, полученных при воздействии пониженной температуры для проростков пшеницы сорта «Новосибирская 41» других возрастов, приведены в табл. 2.

Таблица 2

Table 2

Значения отклонений для экспериментальных данных, полученных при воздействии пониженной температуры на проростки пшеницы сорта «Новосибирская 41»

Deviation values for experimental data obtained under the influence of a low temperature for Novosibirsk 41 wheat seedlings

Время (сутки)	10	12	14	16	18
$U_{\max \text{ х.ц.}}, (\text{В})$	0,02714	0,0198	0,01495	0,01215	0,01222
$\delta_{\text{х.н}}^i$	–	–0,07138	–0,10462	–0,12872	–0,13418
$\delta_{\text{х}}^i, (\text{В})$	–	0,00582	0,00851	0,01048	0,01092

Полученные нормализованные отклонения являются общими для различных серий экспериментальных исследований биопотенциалов проростков пшеницы сорта «Новосибирская 41». Для расчета действительных метрологических величин отклонений для каждой отдельной серии экспериментов необходимо провести преобразование, учитывающее абсолютное среднестатистическое значение биопотенциала проростков пшеницы при воздействии повышенной или пониженной температуры для первого дня этой серии экспериментов по следующим формулам:

- для повышенной температуры

$$\delta_{\text{т}}^i = K_{\text{т}} \left| \delta_{\text{т.н}}^i \right| U_{\max \text{ т.ц.}}; \quad (5)$$

- для пониженной температуры

$$\delta_x^i = K_x \left| \delta_{x,n}^i \right| U_{\max x,ц}, \quad (6)$$

где $U_{\max т.ц}$ и $U_{\max х.ц}$ – абсолютные среднестатистические значения биопотенциала проростка при воздействии повышенной и пониженной температуры для первого дня серии экспериментов; $\delta_{т.н}^i$ и $\delta_{х.н}^i$ – нормализованные значения отклонения для экспериментальных данных, полученные при воздействии повышенной и пониженной температуры для второго и последующих дней серии экспериментов; i – число дней в серии экспериментов; $K_T = 2,93$ – коэффициент нормализации, характеризующий отношение свободного члена уравнения регрессии для экспериментальных данных, полученных при воздействии повышенной температуры, к $U_{\max т.ц}$ [13]; $K_x = 3,0$ – коэффициент нормализации, характеризующий отношение свободного члена уравнения регрессии для экспериментальных данных, полученных при воздействии пониженной температуры, к $U_{\max х.ц}$.

Полученные в результате вычисления по выражениям (5) и (6) абсолютные значения величин отклонений для коррекции экспериментальных данных 2022 года, приведенных в работе [13], представлены в табл. 1 для повышенной температуры и в табл. 2 для пониженной температуры.

Согласно исследованиям, приведенным в [13], корректирующие величины рассчитаны для параметра $U_{\max ц}$ – максимального фильтрованного центрированного значения биопотенциалов проростков пшеницы сорта «Новосибирская 41» при воздействии на них повышенной и пониженной температуры и проведении экспериментальных исследований в течение восьми суток.

Полученные в результате этих исследований данные предполагают коррекцию измерений биопотенциалов на величину приведенных в табл. 1 и 2 значений в зависимости от времени проведения этих измерений. Таким образом, расчет данных для фенотипирования проростков пшеницы сорта «Новосибирская 41» на платформе биопотенциалов по воздействию повышенной и пониженной температуры проводится для каждого эксперимента, привязанного к идентификационному номеру (Id), по следующим формулам:

- при воздействии повышенной температуры

$$U_{\max kт}^n = U_{\max т.цf}^n + \delta_t^i; \quad (7)$$

- при воздействии пониженной температуры

$$U_{\max kх}^n = U_{\max х.цf}^n + \delta_x^i, \quad (8)$$

где $U_{\max kт}^n$ – данные n -го эксперимента после коррекции; $U_{\max т.цf}^n$ – данные n -го эксперимента, полученные в результате эксперимента; δ_t^i и δ_x^i – корректирующие значения для экспериментальных данных выбранного возраста

проростков, полученных при воздействии повышенной и пониженной температуры.

Результаты коррекции данных, полученные по выражениям (7) и (8) для возраста проростков пшеницы сорта «Новосибирская 41» 12 дней по атрибуту cf платформы биопотенциалов, приведены в табл. 3 и 4.

3. ОЦЕНКА ВЛИЯНИЯ КОРРЕКЦИИ НА РЕЗУЛЬТАТЫ КЛАСТЕРИЗАЦИИ ДАННЫХ

Фенотипирование проростков пшеницы сорта «Новосибирская 41» проводится с помощью оригинальной программы Eclaster [14], разработанной по результатам исследований, приведенных в работах [15] и [16]. Подготовка данных для использования в программе производится путем замены в массиве, сформированном после предварительной обработки, всех полученных реализаций экспериментальных исследований параметра $U_{\max cf}$ максимального фильтрованного центрированного значения на значения $U_{\max k}$, приведенные в табл. 3 и 4 соответственно для повышенной и пониженной температуры.

Таблица 3

Table 3

Расчет абсолютных корректирующих значений максимального центрированного фильтрованного сигнала биопотенциала проростков пшеницы сорта «Новосибирская 41» в возрасте 12 дней при воздействии повышенной температуры

Calculation of absolute correction values of the maximum centered filtered biopotential signal of Novosibirsk 41 wheat seedlings at the age of 12 days when exposed to an elevated temperature

Id	$U_{\max Tcf}^{12}$	δ_T^{12}	$U_{\max kT}^{12}$
129 663	0,00934	0,005198	0,014538
129 671	0,00815	–	0,013344
129 677	0,01024	–	0,015441
129 683	0,01094	–	0,016136
129 690	0,00691	–	0,012107
129 697	0,01297	–	0,018163
129 703	0,05482	–	0,060016
129 709	0,00997	–	0,015168
129 716	0,0156	–	0,020794
129 724	0,08271	–	0,087909

Таблица 4

Table 4

Расчет абсолютных корректирующих значений максимального центрированного фильтрованного сигнала биопотенциала проростков пшеницы сорта «Новосибирская 41» в возрасте 12 дней при воздействии пониженной температуры

Calculation of absolute correction values of the maximum centered filtered biopotential signal of Novosibirsk 41 wheat seedlings at the age of 12 days when exposed to low temperature

Id	$U_{\max \times cf}^{12}$	δ_x^{12}	$U_{\max \times kx}^{12}$
129 662	0,01213	0,00582	0,017954
129 670	0,02264	–	0,028457
129 676	0,0069	–	0,012718
129 682	0,01863	–	0,024447
129 689	0,00623	–	0,01205
129 696	0,01976	–	0,025583
129 702	0,02238	–	0,028198
129 708	0,01146	–	0,017281
129 715	0,00851	–	0,014333
129 723	0,06936	–	0,075182

В результате проведенных операций формируется два блока (файла) данных – исходный и отредактированный с учетом изменения биопотенциалов в зависимости от возраста проростков пшеницы при воздействии на них повышенной и пониженной температуры. Результаты кластеризации этих данных в виде диаграмм рассеивания образцов приведены на рис. 1 для исходного значения при воздействии повышенной температуры и на рис. 2 – при этом же воздействии для данных после коррекции. На диаграммах область точек объектов, не вошедших в базовый кластер, обозначена 1, а вошедших – 2.

Результаты кластеризации исходных данных в виде диаграмм рассеивания образцов при воздействии пониженной температуры приведены на рис. 3, а результаты кластеризации данных после коррекции при этом же воздействии представлены на рис. 4.

Анализ результатов кластеризации представленных данных для образцов разного возраста по значениям атрибута *cf* для исходного и откорректированного по возрасту сигналов биопотенциалов проростков пшеницы сорта «Новосибирская 41» при воздействии повышенной температуры показывает, что в обоих случаях полученные кластеры содержат одинаковое количество объектов исследования. При этом объем базового кластера составляет 96 % от общего числа объектов. Однако плотность распределения объектов (компактность) для откорректированных значений данных выше, что при оценке фенотипа растения является определяющим.

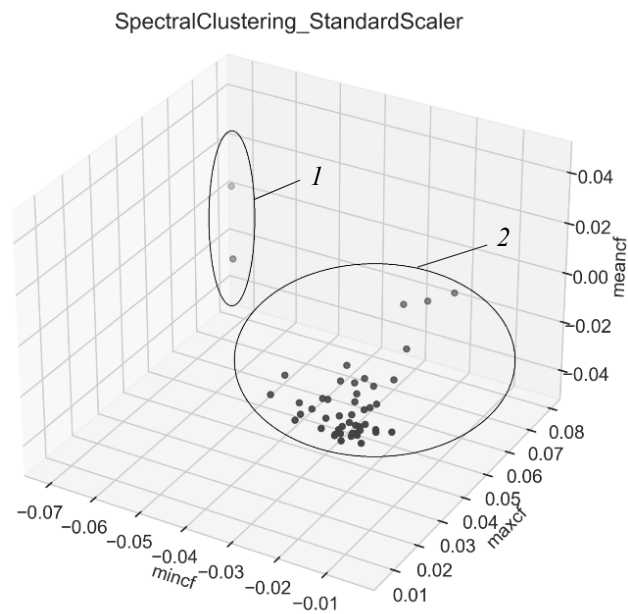


Рис. 1. Диаграмма рассеивания образцов для исходных данных при воздействии повышенной температуры

Fig. 1. Sample scattering diagram for the initial data when exposed to an elevated temperature

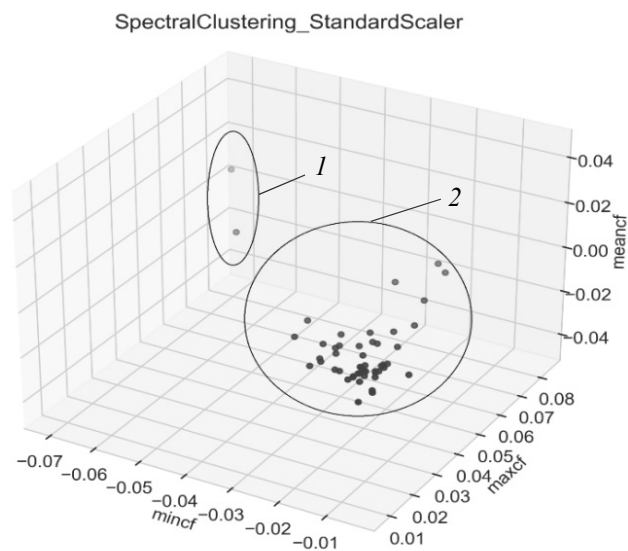


Рис. 2. Диаграмма рассеивания образцов для данных после коррекции при воздействии повышенной температуры

Fig. 2. Sample scattering diagram for data after correction when exposed to an elevated temperature

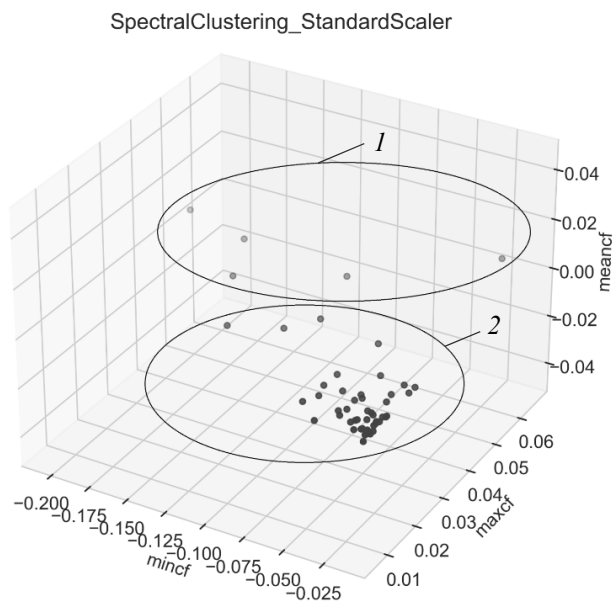


Рис. 3. Диаграмма рассеивания образцов исходных данных при воздействии пониженной температуры

Fig. 3. Sample scattering diagram of the initial data under the influence of lowered temperature

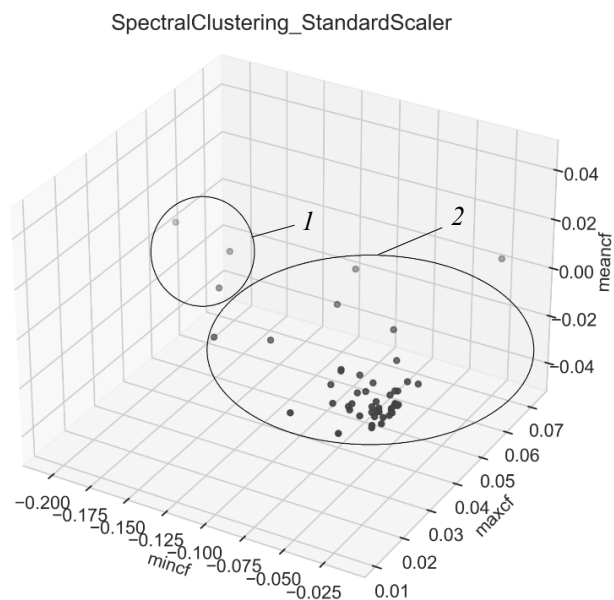


Рис. 4. Диаграмма рассеивания образцов данных после коррекции при воздействии пониженной температуры

Fig. 4. Diagram of the scattering of data samples after correction under the influence of low temperature

При анализе аналогичных результатов с учетом воздействия пониженной температуры распределение объектов в полученных кластерах существенно меняется. Для исходных данных количество объектов, не вошедших в базовый кластер, составляет 5 единиц, для аналогичного кластера при откорректированных по возрасту значениях данных – 3 единицы. Соответственно объем базового кластера возрастает на 2 единицы и составляет 94 % от общего числа объектов. При этом плотность распределения объектов основной части базового кластера для откорректированных значений данных также возрастает, что повышает качество оценки фенотипирования.

ЗАКЛЮЧЕНИЕ

В результате выполненного исследования построена математическая модель подготовки экспериментальных данных (выражения (7) и (8)), которая может быть использована при наличии статистической регрессии изменения величины биопотенциалов злаковых растений в зависимости от их возраста. Разработан алгоритм коррекции исходных величин биопотенциалов, позволяющий в последующих исследованиях проводить коррекцию данных без предварительного построения регрессии. Также показано, каким образом формируются файлы данных для выполнения операции кластеризации при оценке сорта растения. Получение необходимых корректирующих величин производится на основании среднего значения параметра U_{\max} для центрированного фильтрованного сигнала cf реализаций экспериментов первого дня. Выполненный сравнительный анализ показывает потребность проведения экспериментальных исследований посуточного мониторинга биопотенциалов проростков других сортов пшеницы аналогично сорту «Новосибирская 41» и согласно предложенной методике позволяет рассчитать корректирующие величины и подготовить данные для их совместного фенотипирования. В результате применения предложенного алгоритма количество объектов в базовом кластере увеличивается, а их плотность возрастает, и это еще раз подтверждает необходимость подготовки данных, полученных при измерении биопотенциалов проростков пшеницы различных сортов для улучшения процесса кластеризации, особенно при продолжительных экспериментах с проростками сельскохозяйственных культур. Учет всех особенностей метаболизма растений различных сортов позволит более качественно и полно провести оценку их фенотипа. Для численной оценки качества фенотипирования в дальнейшем предполагается применение метрик количественной оценки качества кластеризации из модуля `sklearn.metrics` либо подмодуля `sklearn.metrics.pairwise`, который реализует утилиты для оценки попарных расстояний или близости наборов данных.

СПИСОК ЛИТЕРАТУРЫ

1. Чубукова А. Data Mining. – М.: ИНТУИТ, 2016. – 470 с. – ISBN 978-5-94774-819-2.
2. Феномика растений: фундаментальные основы, программно-аппаратные платформы и методы машинного обучения / В.В. Демидчик, А.Ю. Шашко, В.Ю. Бондаренко и др. // Физиология растений. – 2020. – Т. 67, № 3. – С. 227–245.

3. Crop phenomics and high-throughput phenotyping: past decades, current challenges and future perspectives / W. Yang, H. Feng, X. Zhang, J. Zhang, J.H. Doonan, W.D. Batchelor, L. Xiong, J. Yan // *Molecular Plant*. – 2020. – Vol. 13. – P. 187–214. – DOI: 10.1016/j.molp.2020.01.008.
4. Development of a novel aluminum tolerance phenotyping platform used for comparisons of cereal aluminum tolerance and investigations into rice aluminum tolerance mechanisms / A.N. Famoso, R.T. Clark, J.E. Shaff, E. Craft, S.R. McCouch, L.V. Kochian // *Plant Physiology*. – 2010. – Vol. 153. – P. 1678–1691.
5. *Chernoff H.* The use of faces to represent points in k-dimensional space graphically // *Journal of the American Statistical Association*. – 1973. – Vol. 68 (342). – P. 361–368.
6. *Thyssen A., Cubedo-Ruiz E., Winkler R.* Improved representation of biological information by using correlation as a distance function for cluster analysis of a heat map // *American Journal of Plant Sciences*. – 2017. – Vol. 8. – P. 502–516. – DOI: 10.4236/ajps.2017.83035.
7. OPTIMAS-DW: A comprehensive transcriptomics, metabolomics, ionomics, proteomics and phenomics data resource for maize / C. Colmsee, M. Mascher, T. Czauderna, A. Hartmann, U. Schlüter, N. Zellerhoff, et al. // *BMC Plant Biology*. – 2012. – Vol. 12. – P. 245. – DOI: 10.1186/1471-2229-12-245.
8. *Schauer N., Fernie A.R.* Plant metabolomics: towards biological function and mechanism // *Trends in Plant Science*. – 2006. – Vol. 11. – P. 508–516. – DOI: 10.1016/j.tplants.2006.08.007.
9. PENDISC: A simple method for constructing a mathematical model from time-series data of metabolite concentrations / K. Sriyudthsak, M. Iwata, M.Y. Hirai, F. Shiraishi // *Bulletin of Mathematical Biology*. – 2014. – Vol. 76. – P. 1333–1351. – DOI: 10.1007/s11538-014-9960-8.
10. *Бережной В.А., Иващук О.А., Семенов Д.С.* Обзор методов и алгоритмов автоматизированных систем фенотипирования растений // *Современные наукоемкие технологии*. – 2021. – № 4. – С. 111–116. – DOI: 10.17513/snt.38624.
11. New phenotyping methods for screening wheat and barley for beneficial responses to water deficit / R. Munns, R.A. James, X.R.R. Sirault, R.T. Furbank, H.G. Jones // *Journal of Experimental Botany*. – 2010. – Vol. 61. – P. 3499–3507.
12. *Furbank R.T., Tester M.* Phenomics – technologies to relieve the phenotyping bottleneck // *Trends in Plant Science*. – 2011. – Vol. 16 (12). – P. 635–644. – DOI: 10.1016/j.tplants.2011.09.005.
13. *Сероклинов Г.В., Гунько А.В.* Посуточный мониторинг биопотенциала проростков пшеницы как средство повышения качества оценки стрессоустойчивости растений // *Сибирский вестник сельскохозяйственной науки*. – 2024. – Т. 54, № 6. – С. 112–120.
14. Программа фенотипирования злаковых растений на платформе оценки биопотенциалов проростков «Eklaster»: свидетельство о гос. регистрации программы для ЭВМ № 2025688079; № 2025687303; заявл. 07.10.2025; опубл. 16.10.2025, Бюл. № 10 / Сероклинов Г.В., Гунько А.В.; Сибирский федеральный научный центр агробиотехнологий Российской академии наук. – 1 с.
15. *Seroklinov G.V., Gunko A.V.* Evaluation of the effectiveness of data preparation in phenotyping wheat seedlings of various varieties // *IEEE 3rd International Conference on Problems of Informatics, Electronics and Radio Engineering (PIERE-2024): proceedings, Novosibirsk, 15–17 Nov. 2024*. – IEEE, 2024. – P. 630–633.
16. *Seroklinov G.V., Gunko A.V.* Assessment of stress resistance of wheat varieties based on the cluster analysis of biopotential parameters // *IOP Conference Series: Earth and Environmental Science*. – 2021. – Vol. 848 (1). – P. 012183.
17. Application of data mining technologies for processing results of experimental studies / A.V. Goonko, G.V. Seroklinov, E.S. Devyatkin, F.S. Golubkov // *16 International Scientific and Technical Conference Actual Problems of Electronic Instrument Engineering (APEIE-2023): proceedings, Novosibirsk, 10–12 Nov. 2023*. – IEEE, 2023. – P. 860–863. – DOI: 10.1109/APEIE59731.2023.10347609.

Сероклинов Геннадий Васильевич, ведущий научный сотрудник лаборатории изучения физических процессов в агрофитоценозах Сибирского федерального научного центра агробиотехнологий Российской академии наук. Основное направление научных исследо-

ваний – автоматизация экспериментальных исследований. Имеет более 100 печатных работ. E-mail: seroklinov@mail.ru

Гунько Андрей Васильевич, доцент кафедры автоматизации Новосибирского государственного технического университета. Основное направление научных исследований – обработка больших массивов данных экспериментальных исследований. Имеет более 100 печатных работ. E-mail: gun@ait.cs.nstu.ru

Seroklinov Gennady V., a leading researcher at the Laboratory for the Study of Physical Processes in Agrophytocenoses, Siberian Federal Scientific Center for Agrobiotechnology of the Russian Academy of Sciences. The main area of scientific research is automation of experimental research. He has more than 100 published works. E-mail: seroklinov@mail.ru

Gunko Andrey V., associate professor at the Department of Automation in Novosibirsk State Technical University. The main area of his scientific research is the processing of large amounts of experimental research data. He has more than 100 published works. E-mail: gun@ait.cs.nstu.ru

DOI: 10.17212/2782-2001-2026-2-69-82

A method for preparing data for phenotyping wheat seedlings of different varieties using the example of variety "Novosibirskaya 41"

G.V. SEROKLINOV^{1,a}, A.V. GUNKO^{2,b}

¹ *Siberian Federal Scientific Center of Agrobiotechnics of the Russian Academy of Sciences, 1 Centralnaya Street, Krasnoobsk, Novosibirsk Region, 630501, Russian Federation*

² *Novosibirsk State Technical University, 20 K. Marx Prospekt, Novosibirsk, 630073, Russian Federation*

^a *seroklinov@mail.ru* ^b *gun@ait.cs.nstu.ru*

Abstract

The paper discusses the preparation of experimental data used to measure the biopotentials of Novosibirskaya 41 wheat seedlings exposed to elevated and lowered temperatures, in order to conduct phenotyping of these plants using cluster analysis. It is noted that such a preparation is necessary for long-term experimental studies that take several calendar days (up to 10 or more), during which metabolic changes in seedling samples occur, affecting their biopotential values.

The paper is based on experimental data obtained in 2020 and 2022 and their regression analysis, as reported in [13]. The results of changes in seedling biopotentials depending on their age are briefly described, and an algorithm for calculating corrective biopotential values for each magnification level of the objects is provided. Statistical regressions of changes in biopotential values depending on the need to preserve seedlings of these wheat varieties were obtained. This allowed the development of an algorithm for correcting the initial average biopotentials for these conditions without preliminary regression analysis of the data. Two data sets were generated for assessing the phenotype of the objects: the original data set, obtained through primary processing of changes in these seedling biopotentials under exposure to elevated and lowered temperatures, and the corrected data set, in the partial parameter ($\varphi_{\max.c.}$) of the maximum filtered centered value (cf) of the wheat seedling biopotentials under these conditions. Plant phenotyping was performed based on the data sets using the original Ecluster program, which implements this methodical spectral clustering from the sklearn.cluster library in the Python programming environment. The clustering results presented in the form of a scatterplot demonstrate improved cluster separation for the corrected data.

* Received 10 October 2025.

Keywords: mathematical model, experimental data, biopotential, phenotyping, data regression analysis, cluster analysis, temperature, plant age, correction algorithm

REFERENCES

1. Chubukova A. *Data Mining*. Moscow, INTUIT Publ., 2016. 470 p. ISBN 978-5-94774-819-2. (In Russian).
2. Demidchik V.V., Shashko A.Y., Bandarenka U.Y., Przhevalskaya D.A., Charnysh M.A., Sokolik A.I., Yu M., Smolikova G.N., Pozhvanov G.A., Medvedev S.S., Barkosvkiy A.V. Fenomika rastenii: fundamental'nye osnovy, programmno-apparatnye platformy i metody mashinnogo obucheniya [Plant phenomics: fundamental bases, software and hardware platforms, and machine learning]. *Fiziologiya rastenii = Russian Journal of Plant Physiology*, 2020, vol. 67, no. 3, pp. 227–245. (In Russian).
3. Yang W., Feng H., Zhang X., Zhang J., Doonan J.H., Batchelor W.D., Xiong L., Yan J. Crop phenomics and high-throughput phenotyping: past decades, current challenges and future perspectives. *Molecular Plant*, 2020, vol. 13, pp. 187–214. DOI: 10.1016/j.molp.2020.01.008.
4. Famoso A.N., Clark R.T., Shaff J.E., Craft E., McCouch S.R., Kochian L.V. Development of a novel aluminum tolerance phenotyping platform used for comparisons of cereal aluminum tolerance and investigations into rice aluminum tolerance mechanisms. *Plant Physiology*, 2010, vol. 153, pp. 1678–1691.
5. Chernoff H. The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association*, 1973, vol. 68 (342), pp. 361–368.
6. Thyssen A., Cubedo-Ruiz E., Winkler R. Improved representation of biological information by using correlation as a distance function for cluster analysis of a heat map. *American Journal of Plant Sciences*, 2017, vol. 8, pp. 502–516. DOI: 10.4236/ajps.2017.83035.
7. Colmsee C., Mascher M., Czauderna T., Hartmann A., Schlüter U., Zellerhoff N., et al. OPTIMAS-DW: A comprehensive transcriptomics, metabolomics, ionomics, proteomics and phenomics data resource for maize. *BMC Plant Biology*, 2012, vol. 12, p. 245. DOI: 10.1186/1471-2229-12-245.
8. Schauer N., Fernie A.R. Plant metabolomics: towards biological function and mechanism. *Trends in Plant Science*, 2006, vol. 11, pp. 508–516. DOI: 10.1016/j.tplants.2006.08.007.
9. Sriyudthsak K., Iwata M., Hirai M.Y., Shiraishi F. PENDISC: A simple method for constructing a mathematical model from time-series data of metabolite concentrations. *Bulletin of Mathematical Biology*, 2014, vol. 76, pp. 1333–1351. DOI: 10.1007/s11538-014-9960-8.
10. Berezhnoy V.A., Ivaschuk O.A., Semenov D.S. Obzor metodov i algoritmov avtomatizirovannykh sistem fenotipirovaniya rastenii [Review of methods and algorithms of automated plant phenotyping systems]. *Sovremennye naukoemkie tekhnologii = Modern High Technologies*, 2021, no. 4, pp. 111–116. DOI: 10.17513/snt.38624.
11. Munns R., James R.A., Sirault X.R.R., Furbank R.T., Jones H.G. New phenotyping methods for screening wheat and barley for beneficial responses to water deficit. *Journal of Experimental Botany*, 2010, vol. 61, pp. 3499–3507.
12. Furbank R.T., Tester M. Phenomics – technologies to relieve the phenotyping bottleneck. *Trends in Plant Science*, 2011, vol. 16 (12), pp. 635–644. DOI: 10.1016/j.tplants.2011.09.005.
13. Seroklinov G.V., Gun'ko A.V. Posutochnyi monitoring biopotentsiala prorostkov pshenitsy kak sredstvo povysheniya kachestva otsenki stressoustoichivosti rastenii [Daily monitoring of the biopotential of wheat seedlings as a means of increasing the assessment of plant stress resistance]. *Sibirskii vestnik sel'skokhozyaistvennoi nauki = Siberian Herald of Agricultural Science*, 2024, vol. 54, no. 6, pp. 112–120.
14. Seroklinov G.V., Gun'ko A.V. *Programma fenotipirovaniya zlakovykh rastenii na platforme otsenki biopotentsialov prorostkov «Eklaster»* [A program for phenotyping cereal plants on the Eklaster seedling biopotential assessment platform]. Russian Federation. The Certificate on official registration of the computer program. No. 025688079, 2025.

15. Seroklinov G.V., Gunko A.V. Evaluation of the effectiveness of data preparation in phenotyping wheat seedlings of various varieties. *IEEE 3rd International Conference on Problems of Informatics, Electronics and Radio Engineering (PIERE-2024)*. Proceedings, Novosibirsk, 15–17 Nov. 2024, pp. 630–633.

16. Seroklinov G.V., Gunko A.V. Assessment of stress resistance of wheat varieties based on the cluster analysis of biopotential parameters. *IOP Conference Series: Earth and Environmental Science*, 2021, vol. 848 (1), p. 012183.

17. Goonko A.V., Seroklinov G.V., Devyatkin E.S., Golubkov F.S. Application of data mining technologies for processing results of experimental studies. *16 International Scientific and Technical Conference Actual Problems of Electronic Instrument Engineering (APEIE-2023)*. Proceedings, Novosibirsk, 10–12 Nov. 2023, pp. 860–863. DOI: 10.1109/APEIE59731.2023.10347609.

Для цитирования:

Сероклинов Г.В., Гунько А.В. Метод подготовки данных для фенотипирования проростков пшеницы различных сортов на примере сорта «Новосибирская 41» // Системы анализа и обработки данных. – 2026. – № 2 (102). – С. 69–82. – DOI: 10.17212/2782-2001-2026-2-69-82.

For citation:

Seroklinov G.V., Gunko A.V. Metod podgotovki dannykh dlya fenotipirovaniya prorstkov pshe-nitsy razlichnykh sortov na primere sorta «Novosibirskaya 41» [A method for preparing data for phenotyping wheat seedlings of different varieties using the example of variety "Novosibirskaya 41"]. *Sistemy analiza i obrabotki dannykh = Analysis and Data Processing Systems*, 2026, no. 2 (102), pp. 69–82. DOI: 10.17212/2782-2001-2026-2-69-82.