

doi 10.18699/vjgb-26-36

## Применимость инструмента для анализа временных рядов StatFaRmer при цифровом фенотипировании сои (*Glycine max*)

Д.С. Ульянов , А.А. Ульянова , А.А. Кочешкова , А.О. Блинков , А.В. Архипов , Я.С. Меглицкая ,  
Н.Ю. Свистунова , Г.И. Карлов , М.Г. Дивашук 

Всероссийский научно-исследовательский институт сельскохозяйственной биотехнологии, Москва, Россия

 uldas1508@gmail.com

**Аннотация.** Современные исследования в области агробиотехнологий все чаще опираются на методы автоматизированной фиксации и интерпретации морфофизиологических и спектральных характеристик растений – направление, известное как цифровое фенотипирование. Этот подход направлен на обнаружение устойчивых различий между генотипами, культивируемыми в неидентичных условиях среды. Ранее был предложен StatFaRmer – инструмент с открытым кодом, совершенствующийся в рамках настоящей работы для комплексного анализа временных фенотипических наборов данных, преимущественно ориентированный на изучение сельскохозяйственных культур, таких как соя (*Glycine max*). Разработанный инструмент реализует автоматизированные процедуры предобработки данных, включая синхронизацию временных меток между образцами и устранение шумовых артефактов и выбросов. Эти функции особенно актуальны для многомесячных экспериментов, связанных с оценкой параметров роста, колебаний площади фотосинтетического аппарата или других биометрических показателей. Поддержка стандартизированных форматов данных (XLSX, CSV) обеспечивает совместимость с распространенными системами фенотипирования, упрощая кроссплатформенную интеграцию. Таким образом, инструмент может поддерживать интеграцию с популярными HTPP-платформами (например, Traitmill, HyperAlxpert, Plant Accelerator), что позволяет использовать данные, полученные из различных источников, в едином аналитическом конвейере. Для экспериментов с соей StatFaRmer предоставляет настраиваемый дисперсионный анализ (ANOVA) с визуализацией диагностических параметров (нормальность распределения, гомогенность дисперсий) и оценкой значимости эффектов между пользовательскими группами. Пример применения включает сравнение параметров роста 20 сортов сои в условиях контролируемого стресса: инструмент автоматически агрегировал данные с неравномерной частотой измерений (от 1 часа до 3 суток), идентифицировал аномалии в динамике удлинения гипокотилия и рассчитал статистическую значимость различий между группами ( $p < 0.01$ ). Инструмент протестирован на масштабных наборах данных (более 2000 измерений на эксперимент). StatFaRmer реализован в виде веб-приложения на платформе Shiny с пошаговыми инструкциями для установки и запуска в ОС Windows и Linux. Все этапы обработки – от первичных данных до итоговых графиков – документируются, что обеспечивает прозрачность анализа и соответствие требованиям воспроизводимости исследований. Таким образом, StatFaRmer предлагает специализированное решение для статистической верификации гипотез в цифровом фенотипировании сои, сокращая время на подготовку данных и минимизируя риски ошибок при работе с нестационарными временными рядами.

**Ключевые слова:** высокопроизводительное фенотипирование растений; визуализация фенотипических данных; анализ временных рядов; цифровые платформы фенотипирования; генотип-фенотипический анализ; статистический анализ фенотипических данных; программное обеспечение с открытым исходным кодом; автоматизированный анализ данных

**Для цитирования:** Ульянов Д.С., Ульянова А.А., Кочешкова А.А., Блинков А.О., Архипов А.В., Меглицкая Я.С., Свистунова Н.Ю., Карлов Г.И., Дивашук М.Г. Применимость инструмента для анализа временных рядов StatFaRmer при цифровом фенотипировании сои (*Glycine max*). *Вавиловский журнал генетики и селекции*. 2026;30(2):321-329. doi 10.18699/vjgb-26-36

**Финансирование.** Исследование проведено при поддержке Министерства науки и высшего образования Российской Федерации (госзадание FGUM-2025-0010).

## Applicability of the StatFaRmer time series analysis tool in soybean (*Glycine max*) digital phenotyping

D.S. Ulyanov , A.A. Ulyanova , A.A. Kocheshkova , A.O. Blinkov , A.V. Arkhipov , Ya.S. Meglitskaya ,  
N.Yu. Svistunova , G.I. Karlov , M.G. Divashuk 

All-Russia Research Institute of Agricultural Biotechnology, Moscow, Russia

 uldas1508@gmail.com

**Abstract.** Contemporary agrobiotechnology research increasingly relies on automated methods for capturing and interpreting morphophysiological and spectral plant characteristics – a field known as digital phenotyping. This approach aims to identify stable differences between genotypes cultivated under non-identical environmental conditions. We previously introduced StatFaRmer, an open-source tool that we further develop here for comprehensive analysis of temporal phenotypic datasets, with a primary focus on crops such as soybean (*Glycine max*). The tool implements automated data preprocessing procedures, including synchronization of timestamps across samples and removal of noise artifacts and outliers. These features are particularly relevant for multi-month experiments involving assessments of growth parameters, fluctuations in photosynthetic apparatus area, or other biometric indicators. Support for standardized data formats (XLSX, CSV) ensures compatibility with common phenotyping systems, simplifying cross-platform integration. Thus, the tool can integrate with widely used HTPP platforms (e.g., Traitmill, HyperAlxpert, Plant Accelerator), enabling data from diverse sources to be analyzed within a single pipeline. For soybean experiments, StatFaRmer provides customizable analysis of variance (ANOVA) with visualization of diagnostic parameters (normality of distribution, homogeneity of variances) and evaluation of effect significance between user-defined groups. An example application compares growth parameters across 20 soybean cultivars under controlled stress: the tool automatically aggregated data with uneven measurement frequencies (from 1 hour to 3 days), identified anomalies in hypocotyl elongation dynamics, and computed statistical significance between groups ( $p < 0.01$ ). The tool has been tested on large-scale datasets (over 2,000 measurements per experiment). StatFaRmer is implemented as a Shiny-based web application, with step-by-step deployment guides for Windows and Linux. All processing stages – from raw data to final plots – are documented to ensure transparency and compliance with research reproducibility standards. Thus, StatFaRmer offers a specialized solution for statistical hypothesis testing in soybean digital phenotyping, reducing data preparation time and minimizing risks of error when handling non-stationary time series.

**Key words:** high-throughput plant phenotyping; phenotypic data visualization; time series analysis; digital phenotyping platforms; genotype-phenotype analysis; statistical analysis of phenotypic data; open-source software; automated data analysis

**For citation:** Ulyanov D.S., Ulyanova A.A., Kocheshkova A.A., Blinkov A.O., Arkhipov A.V., Meglitskaya Ya.S., Svistunova N.Yu., Karlov G.I., Divashuk M.G. Applicability of the StatFaRmer time series analysis tool in soybean (*Glycine max*) digital phenotyping. *Vavilovskii Zhurnal Genetiki i Seleksii* = *Vavilov J Genet Breed.* 2026;30(2):321-329. doi 10.18699/vjgb-26-36

## Введение

Несмотря на стремительное развитие геномики, фенотипирование – оценка внешних признаков и характеристик растений – остается ключевым узким местом в ускорении селекционного процесса (Fasoula et al., 2020; Patel et al., 2023). Традиционные методы, основанные на ручных измерениях, характеризуются высокой трудоемкостью, субъективностью, низкой пропускной способностью и необходимостью деструктивного отбора проб, что делает их непригодными для масштабных программ по созданию новых, более продуктивных и устойчивых сортов (Atefi et al., 2021; Li L. et al., 2023).

Выходом из этой ситуации является высокопроизводительное фенотипирование растений (high-throughput plant phenotyping, HTPP) – технология, представляющая собой прорыв в фенотипических исследованиях. Этот подход основан на интеграции современных сенсорных систем (гиперспектральная, тепловизионная или флуоресцентная съемка, 3D-реконструкция) и методов машинного обучения для автоматизированного, неразрушающего и точного анализа растений.

При традиционном ручном фенотипировании существуют такие недостатки, как большие временные и финансовые затраты, низкая масштабируемость из-за ограничения скорости и объема работ, субъективность и вариабельность, а также ограниченность измеряемых признаков (Thrash et al., 2022; Abebe et al., 2023; Anand et al., 2023; Yuan et al., 2023). В то же время цифровое фенотипирование имеет такие преимущества, как существенный прирост эффективности, непрерывность и недеструктивность измерений, улучшенная точность и объективность и возможность измерять десятки признаков одновременно,

существенно повышающая полученные объемы данных (Li D. et al., 2021; Anand et al., 2023; Buelvas et al., 2023; Gyan et al., 2023; Li L. et al., 2023; Lu et al., 2023; Wang et al., 2024).

Современные HTP-платформы генерируют огромные объемы гетерогенных данных, включая изображения (RGB, гиперспектральные, тепловые), 3D-модели и показания сенсоров, достигающие терабайтов в день. Ключевыми проблемами являются не только объем, но и многомодальность, пространственно-временной характер данных, а также необходимость хранения информации и управления ей. Отсутствие стандартизированных форматов и сложности обеспечения качества данных еще более осложняют анализ (Coppens et al., 2017; Morota et al., 2019; Danilevicz et al., 2021; Gill et al., 2022; Ninomiya, 2022).

Существующие статистические методы, используемые для данных фенотипирования, включают следующие модели:

- одномерные и многомерные методы анализа для проверки гипотез и изучения взаимосвязей между признаками растений. Включают корреляционный анализ, факторный анализ (ФА), главный компонентный анализ (PCA) для уменьшения размерности и выявления важных признаков (Rahaman et al., 2015);
- рандомизированные регрессионные модели и модели многофакторного взаимодействия (например, взаимодействие генотип–среда), применяемые для моделирования динамических фенотипических данных (Morota et al., 2019);
- кластерный анализ (иерархический, k-средних и т. д.) для оценки сходств и различий между растениями и стрессовыми реакциями (Rahaman et al., 2015);

- машинное обучение и глубокое обучение для обработки масштабных и сложных данных, включая классификацию, сегментацию и локализацию признаков на изображениях, а также прогнозирование фенотипических реакций. У таких подходов есть свои преимущества, например возможность обнаружить неочевидную заранее корреляцию; но часто такие методы требуют много обучающих данных, и их интерпретируемость ограничена (Coppens et al., 2017; Ubbens et al., 2025).

Для более комплексного применения вышеописанных методов существуют различные инструменты. Среди коммерческих решений одной из опций является Genstat, предоставляющий комплексные возможности статистического анализа, специально разработанные для приложения в селекции растений. Функции включают пространственный анализ, эксперименты со множественными вариантами условий, геномную селекцию и инструменты анализа стабильности, адаптированные для сельскохозяйственных данных. HortControl – это инструмент, по умолчанию поставляемый для платформы TraitFinder (HortControl – Plant Data Management Software, 2025), с которой работают авторы. Он предлагает интегрированное управление данными для цифрового фенотипирования с совместимостью с BgAPI и снабжен некоторыми инструментами анализа данных. Но для регрессионного анализа пользователю может понадобиться экспортировать файл .csv с полученными данными и проводить дальнейшие анализы сторонними инструментами.

Существуют также инструменты с открытым исходным кодом. К примеру, пакеты R, такие как ASReml-R и rcsvr, предоставляют специализированные функции для продольного моделирования роста и пространственного анализа данных фенотипирования (Sumner, 2025). Эти инструменты реализуют передовые модели смешанных эффектов и методы функционального анализа данных, но требуют опыта в программировании. Также существует PIPPA (plant image phenotyping and analysis/фенотипирование и анализ изображений растений), предоставляющий веб-доступ к высокопроизводительным вычислительным ресурсам для конвейеров анализа изображений (Coppens et al., 2017).

Для решения аналогичных задач в области продольных исследований (longitudinal studies – метод научного наблюдения, при котором одна и та же группа объектов изучается многократно в течение длительного времени) нами разработан StatFaRmer – открытый веб-инструмент на базе R с двухэтапной архитектурой: Master Wizard для обработки данных и Main Application для статистического анализа, не требующий навыков программирования. На сегодняшний день функционал StatFaRmer включает:

- Master Wizard: валидацию проектов, DBSCAN кластеризацию, фильтрацию выбросов (Z-критерий/IQR), техническую агрегацию данных;
- Main Application: автоматический выбор моделей (ANOVA/смешанные модели/сплайны), многофакторный анализ с post-hoc тестами, диагностику нормальности и гомогенности дисперсий, анализ величины эффектов и кривых роста;

- интерактивную визуализацию с Plotly, экспорт высокого разрешения (SVG, PNG, PDF);
- специализированные модули: анализ размеров эффектов, сводки роста;
- экспорт результатов в табличных форматах.

Инструмент оптимизирован для продольных исследований в контролируемых условиях, поддерживает двухфакторные взаимодействия в моделях ANOVA и обеспечивает прозрачность статистических выводов. StatFaRmer заполняет нишу между специализированным ПО и потребностями исследователей, предлагая гибкий анализ сложных фенотипических данных.

В сравнении со стандартным HortControl наш инструмент, будучи системой, развернутой на R/Shiny, предлагает, во-первых, большую гибкость в настройке анализа под конкретный эксперимент; во-вторых, – более продвинутую глубину и автоматизацию статистического конвейера, чем HortControl, более фокусирующийся на данных с датчиков и базовых расчетах.

Если сравнивать с уже существующими статистическими пакетами и платформами (Genstat, ASReml-R, rcsvr), то они представляют собой мощные, но общего назначения статистические среды. Они могут все, что нужно, но требуют от пользователя глубоких знаний статистики и/или навыков программирования для написания корректных моделей, особенно для сложных пространственных и временных данных. В то же время StatFaRmer автоматически генерирует сложные ANOVA-модели, включая все взаимодействия, что избавляет исследователя от рутинного и подверженного ошибкам кодирования. Также, в отличие от универсальных пакетов, StatFaRmer из коробки специализируется на фенотипических данных (повторные измерения, выбросы). Еще одно положительное отличие – интерактивность и визуализация: возможность интерактивного подмножества данных (по генотипам, обработкам, временным кластерам) и мгновенного получения результатов для проверки гипотез, что является ключевым преимуществом перед традиционными пакетами, где такой анализ требует постоянного переписывания кода.

В нашей работе StatFaRmer (Ulyanov et al., 2025) применен для анализа данных, полученных при цифровом фенотипировании сои (*Glycine max*) – ключевой сельскохозяйственной культуры. Инструмент предусматривает возможность интеграции с платформами, использующими различные сенсорные системы (мультиспектральные сканеры, 3D-камеры), благодаря поддержке стандартизированных форматов данных (XLSX, CSV) и гибкой настройке параметров. Это позволяет адаптировать его для работы с данными, собранными как в контролируемых условиях (например, климатических камерах), так и в полевых экспериментах, где нерегулярность измерений и разнородность источников требуют автоматизированной предобработки. Стандартизация входных данных и модульная архитектура StatFaRmer обеспечивают совместимость с такими системами, как Traitmill или PhytoOracle, что открывает перспективы для кроссплатформенных исследований.

Цель настоящей работы заключалась в представлении возможностей инструмента StatFaRmer, последовательной демонстрации его функционального потенциала для ознакомления читателя, а также в описании усовершенствований, реализованных в нем со времени предыдущей публикации.

## Материалы и методы

**Экспериментальный дизайн и растительный материал.** В исследовании использовано 50 сортов сои (*G. max*) различного происхождения (российская и зарубежная селекция), отличающихся по срокам созревания, типу роста и рекомендованным регионам возделывания. Семена перед посевом подвергали обработке фунгицидом, после чего высевали на глубину 2 см во влажный торф в горшки объемом 500 мл (по четыре семени на горшок); после всходов оставляли три растения на горшок.

Культивация растений осуществлялась в контролируемых условиях климатической камеры на протяжении всего вегетационного цикла. Эксперимент включал два варианта условий, различавшихся режимом освещенности при идентичных остальных параметрах микроклимата (температура, влажность, уровень питательных веществ). Для каждого варианта предусматривалось по две биологические повторности, что в совокупности обеспечило выборку из 200 растений (50 сортов × 2 варианта × 2 повторности). Основной уход включал регулярный полив водой комнатной температуры в первые 7–10 дней, затем – по мере необходимости, а после появления настоящих листьев – ежедневное внесение минерального удобрения. Для целей фенотипирования растения распределялись в случайно сформированные группы по 12 экземпляров, итоговое число групп составило по 9 для каждого варианта опыта. Подробное описание экспериментальных процедур и параметров культивации представлено в работе (Ulyanov et al., 2025).

**Цифровое фенотипирование.** Сканирование проводили на установке для фенотипирования TraitFinder (Pheposrex, Нидерланды) с двумя мультиспектральными сканерами в следующих диапазонах длин волн: красный канал (R): 624–634 нм, зеленый канал (G): 530–540 нм, синий канал (B): 465–485 нм и ближний инфракрасный канал (NIR): 720–750 нм.

Фенотипирование растений осуществляли по схеме, где распределение сортов и вариантов опыта по блокам фиксировалось во встроенном программном обеспечении HortControl и дублировалось в таблицах StatFaRmer для последующего анализа. Растения распределялись по экспериментальной площадке, представленной в виде layout – виртуальной схемы, отражающей физическое расположение блоков и растений. Все растения были разбиты на блоки, каждый из которых соответствует одному столу и имеет уникальный идентификатор; таких блоков получилось 9. Блоки дополнительно делились на единицы (units), соответствующие индивидуальным растениям для сбора данных. В HortControl блоки резервировались для эксперимента и связывались с биологической информацией (генотип, обработка) через присвоение метаданных.

Перед сканированием растения транспортировали из климатических камер на специализированный фенотипический стол, соблюдая идентичное позиционирование в рамках заранее заданных цифровых координат. Процедура выполнялась в затемненном помещении с отключенной системой кондиционирования для исключения артефактов, вызванных воздушными потоками или изменением освещенности. Каждое растение размещали так, чтобы его центр массы совпадал с геометрическим центром сканируемой зоны, минимизируя перекрытие листьев и побегов с соседними образцами. Для стабилизации растений в горшках использовали опорные конструкции из зеленых пластиковых стержней и черной проволоки (спектрально нейтральной в диапазоне датчика PlantEye).

В ходе эксперимента выполнено пять последовательных сканирований, общая продолжительность наблюдения составляла 20 дней. Регистрировали общую высоту растения, а также спектральные индексы (NDVI, PSRI), рассчитанные по данным мультиспектрального сенсора следующим образом:

**Средняя высота растения (plant height average, мм):** усредненное значение высоты верхних 10 % точек растения, что минимизирует влияние внешних факторов (например, ветра).

**Спектральные индексы:**

NDVI (normalized difference vegetation index):

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

Диапазон значений: от –1 до 1. Высокие значения (>0.66) указывают на здоровые фотосинтезирующие ткани.

PSRI (plant senescence reflectance index):

$$PSRI = \frac{RED - BLUE}{NIR}$$

Используется для оценки старения листьев (сенесценции).

**Подготовка данных.** Пользователи могут подгружать собственные эксперименты, помещая проекты в папку StatFaRmer/data/project\_NAME/, предварительно подготовив данные следующим образом:

1. **Базовый архив эксперимента.** Проект должен содержать архивированный эксперимент в формате ZIP (\*\_data.zip), включающий CSV-таблицу, импортированную из HortControl с исходными данными.
2. **Таблица ручных корректировок.** Обязательно наличие файла \*\_handmade.csv со столбцами:
  - V.T.R – уникальный идентификатор биологического образца (сорт/обработка/повтор);
  - Treatment – параметры обработки (переопределяет соответствующий столбец в исходной таблице);
  - Cultivar – сорт растения.Эти CSV-файлы формируются ручным способом для внесения данных о растениях. Напротив уникального идентификатора каждого растения нужно указать его сорт и обработку, также можно внести любые другие данные о каждом растении.
3. **Таблица пространственной привязки.** Файл \*\_translation.csv должен включать:

- V.T.R – идентификатор образца;
- T:X:Y – координаты расположения образца, соответствующие столбцу unit в исходных данных.

Эти CSV-файлы формируются ручным способом для того, чтобы программа могла связать импортированные данные из пункта 1 с описанием экспериментальных образцов из пункта 2. Для получения такой таблицы необходимо соотнести идентификатор каждого растения V.T.R с соответствующими ему координатами в базе данных HortControl. Координаты unit кодируются следующим образом:

(номер блока соответствующей группы растений):

(номер ряда, начиная от баркода):

(номер растения в ряду, слева направо).

4. **Дополнительная группировка (опционально).** Файл groups.xlsx с обязательным столбцом Cultivar.

Прочие столбцы могут использоваться как факторы ANOVA и должны содержать только латинские буквы, цифры и подчеркивания.

Все дополнительные столбцы в указанных таблицах сохраняются и доступны для анализа в качестве независимых факторов.

Настройка анализа осуществляется через графический интерфейс Master Wizard, включая:

- выбор проекта из списка доступных с его валидацией;
- подбор DBSCAN eps (временной интервал между измерениями);
- выбор метода фильтрации/винзоризация/аннотации выбросов (IQR или Z-score) на основе указанной структуры ячеек опыта;
- настройку технической агрегации (медиана или среднее);
- экспорт и импорт конфигураций опыта для прозрачности обработки.

StatFaRmer эффективно обрабатывает высокочастотные временные ряды с неоднородными интервалами измерений, автоматически корректируя артефакты и агрегируя данные для снижения ошибок в продольных исследованиях. Инструмент включает модуль многофакторного дисперсионного анализа (ANOVA) с автоматической проверкой статистических предпосылок. Генерируемые интерактивные визуализации позволяют сосредоточиться на интерпретации результатов, минимизируя технические сложности анализа.

## Результаты

### Кластеризация, фильтрация выбросов и группировка данных

Кластеризация и фильтрация выбросов – важный этап первоначальной обработки полученных фенотипических данных. Под выбросами (аномалиями) понимаются точки данных, значительно отклоняющиеся от общего паттерна, что может быть вызвано техническими артефактами (например, временным затемнением сенсора) или биологическими факторами (локальные повреждения тканей). Она позволяет избавиться как от шумов, полученных из-за разного времени начала фенотипирования в различных

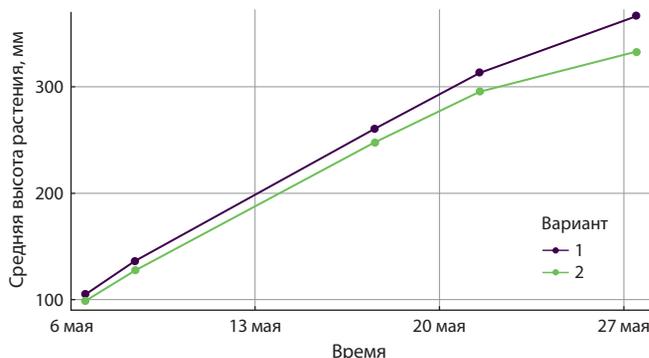


Рис. 1. Динамика изменения параметра «усредненная высота растения» между разными вариантами опыта.

Приведено два графика, показывающих изменение средней высоты растения во времени для вариантов опытов 1 и 2. На рисунке представлены все 50 сортов в двух повторностях для быстрого первичного анализа закономерностей в распределении высоты при вариантах опытов 1 и 2.

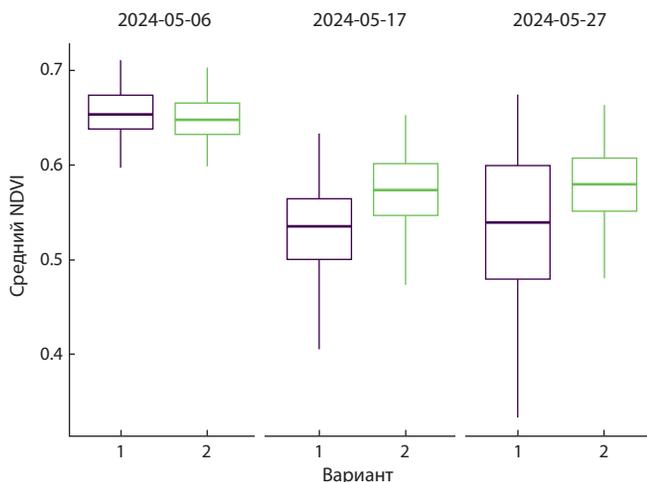
экспериментальных блоках, погрешностей сенсоров, а также влияния внешних факторов (например, ветра), так и от крупных случайных выбросов, таких как посторонние объекты в зоне видимости камеры. На данном этапе работы программы также производится кластеризация данных; можно произвести группировку и сравнить данные при разных параметрах – условиях, обработках и других. Фильтрацию выбросов в программе StatFaRmer можно включать и отключать через графический интерфейс, установив или убрав галочку “plot with outliers”, что позволяет пользователю самостоятельно контролировать данный аспект анализа. Этот этап обработки показан на рис. 1.

StatFaRmer предоставляет возможность проводить сравнительный анализ наборов растений, объединяя образцы по обработкам, сортам или любым дополнительным параметрам, указанным пользователем в сопутствующих таблицах данных. Графики демонстрируют, что усредненная высота растений, культивируемых при варианте опыта 1, в среднем превышает таковую при варианте опыта 2. Значительная разница наблюдается уже на начальном этапе, что при высоком количестве образцов (по 100 для каждого варианта опыта) и начале фенотипирования не с первого дня свидетельствует о том, что различия между вариантами опытов проявляются довольно быстро.

### ANOVA

StatFaRmer предоставляет автоматический выбор статистических моделей: классический ANOVA для сбалансированных планов ( $\leq 10$  уровней факторов), смешанные модели для несбалансированных планов (11–20 уровней факторов) и сплайновые модели для анализа временных рядов. Программа старается подобрать оптимальную схему анализа с учетом всех заданных факторов и их парных взаимодействий, при этом в продвинутых настройках у пользователя сохраняется возможность фиксировать тип модели.

При анализе рис. 1 возникает важный вопрос: является ли ускоренный рост при варианте опыта 1 положитель-



**Рис. 2.** Сравнение среднего NDVI всех растений в трех временных точках: в начале, середине и конце эксперимента для двух вариантов опыта.

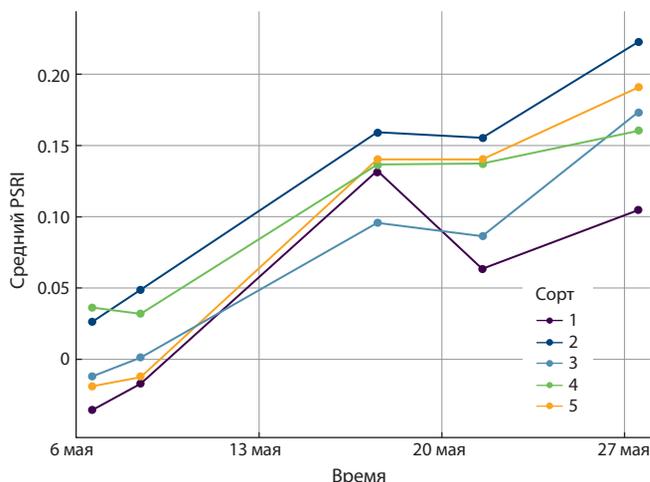
ным признаком, или свидетельствует о стрессе растений? Для оценки состояния растений можно использовать показатель NDVI, позволяющий приблизительно оценить уровень фотосинтетической активности и общего здоровья растений. Более высокие значения NDVI указывают на лучшее состояние растений и более интенсивный фотосинтез, в то время как низкие значения могут свидетельствовать о стрессе. Изменение среднего NDVI во времени, усредненное по всем сортам, показано на рис. 2.

На рис. 2 NDVI не демонстрирует существенных различий между вариантами опыта в первой временной точке, однако дальше NDVI растений, которые выращивали при варианте опыта 2, значительно выше, как видно из того, что медиана распределения соответствует более высокому NDVI. Далее это также можно будет увидеть из сравнительной таблицы групп в разделе «Экспорт результатов в табличных форматах». Это указывает на более благоприятные условия для роста в условиях варианта опыта 2.

### Выборка из данных для анализа временных кластеров

В условиях работы с большими наборами данных пользователям необходима возможность выборочного анализа для самостоятельного выявления закономерностей. В предыдущем примере мы сравнивали значения NDVI в разных вариантах опыта для всех сортов. Теперь предположим, что нам необходимо отобрать из этого набора сорта, более чувствительные и менее чувствительные к варианту опыта. Можно было бы выбрать такие значения полей, чтобы сравнивать один сорт с другим, но проверка таким образом каждого из 50 сортов могла бы оказаться трудоемкой.

Вместо этого можно использовать альтернативный подход. Функция StatFaRmer по созданию выборок данных позволяет, к примеру, сравнивать различные сорта в идентичных условиях, а не один сорт между вариантами опыта. В качестве примера рассмотрим пять сортов и оценим параметр PSRI, отражающий увядание растений. Чем выше



**Рис. 3.** Изменения показателя среднего PSRI в условиях варианта опыта 1 для пяти сортов.

этот параметр, тем больше стресса испытывает растение. Результаты этого сравнения для варианта опыта 1 представлены на рис. 3.

### Интерактивная визуализация с группировкой по признакам

Интерактивная визуализация играет ключевую роль для пользователей, желающих получить адаптивный инструмент для анализа экспериментов. Для анализа отобранных сортов можно построить графики зависимости PSRI от времени с учетом варианта опыта. При этом для изучения динамики множества групп рационально применять фасетирование – разделение графиков на несколько суб-графиков, создаваемых по одному принципу, для подгрупп этой же переменной/переменных. При использовании полей «фактор группировки», «выбранные сорта» и «формула фасетирования» (синтаксический шаблон, задействованный в визуализации данных для описания того, как набор данных разделяется на подгруппы (фасеты) при построении множества связанных графиков; краткое описание принципов построения формулы можно найти на странице github проекта), мы получим отдельный график для каждого сорта (рис. 4).

Такой подход позволяет детально изучить, как изменяется показатель увядания с течением времени для разных сортов. Например, можно заметить, что у некоторых сортов наблюдается резкое изменение динамики показателя – увеличение увядания до середины эксперимента с последующей стабилизацией, тогда как у других этот переломный момент либо слабо выражен, либо проявляется только на завершающих этапах.

### Экспорт результатов в табличных форматах

StatFaRmer позволяет представлять результаты анализов в табличном формате и экспортировать графики в высоком разрешении (300 DPI) в форматах SVG, PNG и PDF для публикаций. Ниже представлена таблица с описательной

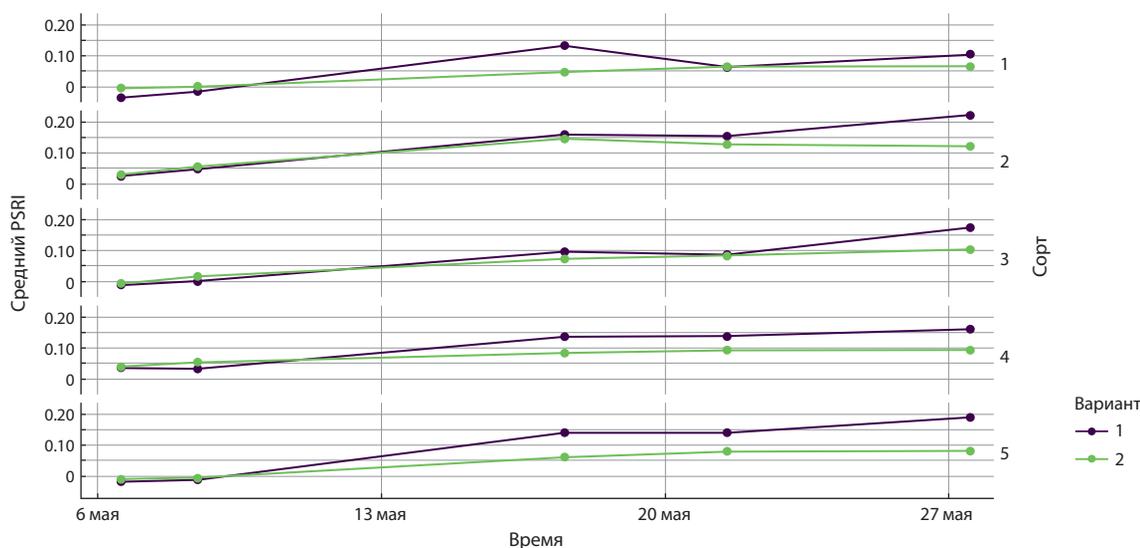


Рис. 4. Изменение среднего PSRI во времени для сортов 1–5.

Описательная статистика для рис. 2.

№	Временная точка	Вариант опыта	n	Среднее	Медиана	Стандартное отклонение	Стандартная ошибка среднего	min	max	25-й квартиль	75-й квартиль	Интер-квартильный размах	Коэффициент вариации	Коэффициент асимметрии	Куртозис
1	1	1	100	0.655	0.653	0.0259	0.00259	0.596	0.731	0.637	0.674	0.0369	3.96	-0.00113	0.0668
2	1	2	100	0.648	0.647	0.0268	0.00268	0.56	0.724	0.633	0.665	0.0315	4.14	-0.0214	1.3
3	4	1	100	0.532	0.535	0.0467	0.00467	0.395	0.633	0.5	0.564	0.0636	8.77	-0.411	0.263
4	4	2	100	0.57	0.573	0.0457	0.00457	0.421	0.653	0.547	0.601	0.0542	8.02	-0.823	1.16
5	6	1	100	0.528	0.539	0.0838	0.00838	0.333	0.674	0.479	0.6	0.12	15.9	-0.477	-0.537
6	6	2	100	0.579	0.58	0.0396	0.00396	0.48	0.664	0.551	0.607	0.0567	6.84	-0.0886	-0.274

Примечание. В колонке «Временная точка» отображены анализируемые группы, где 1, 4 и 6 – это порядковый номер временной точки, т. е. начало, середина и конец эксперимента.

статистикой для рис. 2, в которой указан набор параметров, характеризующий каждую из наблюдаемых групп в каждый момент времени.

### Производительность

Производительность StatFaRmer протестирована на ноутбуке со следующей конфигурацией: процессор AMD Ryzen 5 5500U с Radeon Graphics (2.10 ГГц) и 8 ГБ оперативной памяти.

- Master Wizard: обработка проекта project\_NO3 (58380 строк, 49 столбцов) занимает 8–16 секунд с медианной агрегацией, в зависимости файла конфигурации;
- Main Application: запуск и анализ происходят практически мгновенно благодаря кэшированию обработанных данных;
- визуализация оптимизирована для больших датасетов с использованием карт плотности для > 2000 измерений. Реализованы логирование (logger.R) и бенчмаркинг (benchmark.R) затрачиваемых памяти и времени при работе модулей.

Таким образом, применение автоматизированных методов анализа временных рядов позволяет выявить устойчивые динамические паттерны в данных сои. В отличие от средств, не поддерживающих анализ временных рядов, StatFaRmer справляется с этой задачей, будучи при этом гибким инструментом, не требующим навыков в программировании. При анализе длительных экспериментов StatFaRmer демонстрирует эффективность в обработке высокочастотных временных рядов с неоднородными интервалами измерений. Кроме того, инструмент автоматически выполняет коррекцию артефактов и агрегацию данных. Эти процедуры критически важны для минимизации ошибок в продольных исследованиях. StatFaRmer также включает модуль многофакторного дисперсионного анализа (ANOVA) с автоматической диагностикой статистических предпосылок (например, проверкой нормальности распределения и однородности дисперсий). Интерактивные визуализации высокого разрешения, генерируемые инструментом, выгодно отличают его от большинства других платформ, где графики обычно статичны. В результате

исследователи могут сосредоточиться на интерпретации полученных результатов, а не на технических деталях анализа.

## Обсуждение

В отличие от ряда существующих решений, таких как HTPheno (Hartmann et al., 2011) и IAP (Yang et al., 2020), ориентированных преимущественно на обработку изображений, StatFaRmer обеспечивает расширенную интеграцию с внешними системами фенотипирования. Это позволяет исследователям комбинировать данные из разнородных источников, минимизируя ручной труд. Системы HTPheno и IAP имеют ограниченную поддержку автоматизированного анализа нестационарных временных рядов большого объема, поэтому для полноценной статистической обработки таких данных зачастую требуются дополнительные инструменты. StatFaRmer, напротив, успешно решает задачу автоматизации анализа сложных фенотипических временных рядов, генерируемых высокопроизводительными платформами фенотипирования.

Параллельно с развитием StatFaRmer появились и другие специализированные инструменты, решающие смежные задачи цифрового фенотипирования. Например, пакет AllInOne Pre-processing (Najafabadi et al., 2023) сосредоточен на предобработке полевых фенотипических данных и предоставляет функции нормализации изображений, пространственного анализа и базовой визуализации. Хотя AllInOne эффективно повышает качество исходных данных и способен обрабатывать большие наборы снимков, он не обеспечивает полноценного анализа временных рядов или статистической проверки гипотез. StatFaRmer дополняет такие решения, предлагая углубленный статистический анализ очищенных данных, включая сравнение групп по нескольким факторам и выявление динамических закономерностей во времени.

Современным подходом, ориентированным на масштабируемость, является также платформа PhytoOracle (Gonzalez et al., 2023) – модульный конвейер обработки фенотипических данных, оптимизированный для высокопроизводительных вычислительных кластеров. PhytoOracle способен параллельно обрабатывать мультимодальные данные (например, RGB-изображения, тепловые карты, 3D-точечные облака) и эффективно масштабируется на большие объемы информации. Однако высокая сложность и требование специализированной инфраструктуры делают PhytoOracle менее доступным для широкого круга биологов. StatFaRmer, напротив, ориентирован на доступность: он реализован в виде веб-приложения с интуитивно понятным интерфейсом, что упрощает его применение на стандартных настольных системах. При этом StatFaRmer поддерживает обработку крупных массивов численных данных и интеграцию разнородных источников информации, обеспечивая баланс между мощностью анализа и удобством использования.

Для быстрой оценки возможностей программы был создан демонстрационный датасет на основе эксперимента с азотным стрессом у злаковых культур. Он выложен в открытом доступе на GitHub (<https://github.com/Stathmin/>

StatFaRmer) и развернут для интерактивных тестов – на shinyapps.io (<https://stathmin.shinyapps.io/StatFaRmer/>), где можно ознакомиться с данными и протестировать рабочую копию StatFaRmer. Кроме того, мы провели масштабное тестирование на разнообразных датасетах, включая пшеницу, тритикале, сахарную свеклу, кукурузу, подсолнечник и сою. Программа показала себя как надежный и эффективный инструмент для анализа больших объемов фенотипических данных в различных экспериментальных условиях.

## Заключение

Современные методы цифрового фенотипирования растений генерируют колоссальные объемы данных, включая высокочастотные временные ряды, структурированные метаданные (сорт, обработки, повторности, вариант опыта) и внешние параметры, интегрируемые из сторонних источников. Это, а также многообразие измеряемых платформой признаков, требует специализированных решений для их интерпретации. Разработанный нами инструмент StatFaRmer решает эту задачу благодаря:

- интерактивной визуализации данных с поддержкой временных срезов и фильтрации артефактов;
- многофакторному сравнению групп (по сортам, обработкам, пользовательским тегам) с применением различных статистических анализов;
- гибкой настройке срезов данных для самостоятельного поиска закономерностей, доступного среднему пользователю.

Таким образом, StatFaRmer упрощает обработку сложных фенотипических данных, сокращая время на выявление закономерностей адаптации сои и других культур при различных вариантах опыта, стрессах и других воздействиях.

## Список литературы / References

- Abebe A.M., Kim Y., Kim J., Kim S.L., Baek J. Image-based high-throughput phenotyping in horticultural crops. *Plants*. 2023;12(10):2061. doi 10.3390/plants12102061
- Anand K.J., Nagre S.P., Shrivastava M.K., Amrate P.K., Patel T., Katarra V.K. Enhancing crop improvement through synergistic integration of advanced plant breeding and proximal remote sensing techniques: a review. *Int J Plant Soil Sci*. 2023;35(19):121-138. doi 10.9734/ijps/2023/v35i193533
- Atefi A., Ge Y., Pitla S., Schnable J. Robotic technologies for high-throughput plant phenotyping: contemporary reviews and future perspectives. *Front Plant Sci*. 2021;12:611940. doi 10.3389/fpls.2021.611940
- Buelvas R.M., Adamchuk V.I., Lan J., Hoyos-Villegas V., Whitmore A., Stromvik M.V. Development of a quick-install rapid phenotyping system. *Sensors (Basel)*. 2023;23(9):4253. doi 10.3390/s23094253
- Coppens F., Wuyts N., Inzé D., Dhondt S. Unlocking the potential of plant phenotyping data through integration and data-driven approaches. *Curr Opin Syst Biol*. 2017;4:58-63. doi 10.1016/j.coisb.2017.07.002
- Danilevich M.F., Bayer P.E., Nestor B.J., Bennamoun M., Edwards D. Resources for image-based high-throughput phenotyping in crops and data sharing challenges. *Plant Physiol*. 2021;187(2):699-715. doi 10.1093/plphys/kiab301
- Fasoula D., Ioannides I., Omirou M. Phenotyping and plant breeding: overcoming the barriers. *Front Plant Sci*. 2020;10:1713. doi 10.3389/fpls.2019.01713

- Gill T., Gill S.K., Saini D., Chopra Y., de Koff J.P., Sandhu K. A comprehensive review of high throughput phenotyping and machine learning for plant stress phenotyping. *Phenomics*. 2022;2(3):156-183. doi 10.1007/s43657-022-00048-z
- Gonzalez E.M., Zarei A., Hendlar N., Simmons T., Zarei A., Demieville J., Strand R., ... Swetnam T.L., Merchant N., Michelmore R.W., Lyons E., Pauli D. PhytoOracle: scalable, modular phenomics data processing pipelines. *Front Plant Sci*. 2023;14:1112973. doi 10.3389/fpls.2023.1112973
- Gyan F., Cudjoe D., Sadeghi-Tehran P., Virlet N., Riche A., Castle M., Greche L., Mohareb F., Simms D., Mhada M., Hawkesford M. Machine learning methods for automatic segmentation of images of field- and glasshouse-based plants for high-throughput phenotyping. *Plants*. 2023;12(10):2035. doi 10.3390/plants12102035
- Hartmann A., Czaundera T., Hoffmann R., Stein N., Schreiber F. HTPPheno: an image analysis pipeline for high-throughput plant phenotyping. *BMC Bioinformatics*. 2011;12:148. doi 10.1186/1471-2105-12-148
- HortControl – Plant Data Management Software. *PHENOSPEX*. URL: <https://www.phenospex.com/products/plant-phenotyping/sciencehortcontrol-data-management-software/> (accessed 8.28.25)
- Li D., Quan C., Song Z., Li X., Yu G., Li C., Muhammad A. High-throughput plant phenotyping platform (HT3P) as a novel tool for estimating agronomic traits from the lab to the field. *Front Bioeng Biotechnol*. 2021;8:623705. doi 10.3389/fbioe.2020.623705
- Li L., Hassan M.A., Song J., Xie Y., Rasheed A., Yang S., Li H., Liu P., Xia X., He Z., Xiao Y. UAV-based RGB imagery and ground measurements for high-throughput phenotyping of senescence and QTL mapping in bread wheat. *Crop Sci*. 2023;63(6):3292-3309. doi 10.1002/csc2.21086
- Lu Y., Wang J., Fu L., Yu L., Liu Q. High-throughput and separating-free phenotyping method for on-panicle rice grains based on deep learning. *Front Plant Sci*. 2023;14:1219584. doi 10.3389/fpls.2023.1219584
- Morota G., Jarquín D., Campbell M.T., Iwata H. Statistical methods for the quantitative genetic analysis of high-throughput phenotyping data. In: Lorence A., Medina Jimenez K. (Eds) High-Throughput Plant Phenotyping. *Methods in Molecular Biology*. Vol. 2539. Humana, New York, 2019;269-296. doi 10.1007/978-1-0716-2537-8\_21
- Najafabadi M.Y., Heidari A., Rajcan I. AllInOne pre-processing: a comprehensive preprocessing framework in plant field phenotyping. *SoftwareX*. 2023;23:101464. doi 10.1016/j.softx.2023.101464
- Ninomiya S. High-throughput field crop phenotyping: current status and challenges. *Breed Sci*. 2022;72(1):3-18. doi 10.1270/jsbbs.21069
- Patel T., Babbar A., Behera K., Katara V.K., Anand K.J., Vyshnavi R.G., Pachori S., Bichewar N. Exploring the potential of proximal remote sensing in plant stress phenotyping: a comprehensive review. *Int J Environ Clim Change*. 2023;13(9):2602-2621. doi 10.9734/ijeccl/2023/v13i92511
- Rahaman M., Chen D., Gillani Z., Klukas C., Chen M. Advanced phenotyping and phenotype data analysis for the study of plant growth and development. *Front Plant Sci*. 2015;6:619. doi 10.3389/fpls.2015.00619
- Sumner J. Longitudinal Growth Modeling Options [WWW Document]. 2025. URL: <https://cran.r-project.org/web/packages/pcvr/vignettes/longitudinal.html> (accessed 28.08.25)
- Thrash T., Lee H., Baker R.L. A low-cost high-throughput phenotyping system for automatically quantifying foliar area and greenness. *Appl Plant Sci*. 2022;10(6):e11502. doi 10.1002/aps3.11502
- Ubbens J., Stavness I., Pound M.P., Guo W. Deep learning in plant phenotyping: the first ten years. *Plant Phenomics*. 2025;7(4):100062. doi 10.1016/j.plaphe.2025.100062
- Ulyanov D.S., Ulyanova A.A., Litvinov D., Kocheshkova A., Kroupina A.Yu., Syedina N.M., Voronezhskaya V.S., Vasilyev A.V., Karlov G.I., Divashuk M. StatFaRmer: cultivating insights with an advanced R shiny dashboard for digital phenotyping data analysis. *Front Plant Sci*. 2025;16:1475057. doi 10.3389/fpls.2025.1475057
- Wang Z., Hao J., Shi X., Wang Q., Zhang W., Li F., Mur L.A.J., Han Y., Hou S., Han J., Sun Z. Integrating dynamic high-throughput phenotyping and genetic analysis to monitor growth variation in foxtail millet. *Plant Methods*. 2024;20(1):168. doi 10.1186/s13007-024-01295-z
- Yang W., Feng H., Zhang X., Zhang J., Doonan J., Batchelor W., Xiong L., Yan J. Crop phenomics and high-throughput phenotyping: past decades, current challenges and future perspectives. *Mol Plant*. 2020;13(2):187-214. doi 10.1016/j.molp.2020.01.008
- Yuan H., Song M., Liu Y., Xie Q., Cao W., Zhu Y., Ni J. Field phenotyping monitoring systems for high-throughput: a survey of enabling technologies, equipment, and research challenges. *Agronomy*. 2023;13(11):2832. doi 10.3390/agronomy13112832

**Конфликт интересов.** Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 22.04.2025. После доработки 20.10.2025. Принята к публикации 01.11.2025.