

УДК 004.942:004.45
Код ГРНТИ 50.41.25

doi: 10.54708/19926502_2026_30311370

Программная система регрессионного анализа и визуализации многомерных данных стендовых испытаний ДВС

В.В. Антонов¹, Л.Е. Родионова^{1*}, В.А. Суворова¹, Н.Ю. Дударева¹,
Е.В. Пальчевский², А.С. Дьячков²

¹Уфимский университет науки и технологий», г. Уфа, Республика Башкортостан, Россия

²МИРЭА - Российский технологический университет (РТУ МИРЭА), г. Москва, Россия

³ООО «Газпромнефть-Цифровые решения», г. Уфа, Республика Башкортостан, Россия

Аннотация. Представлен программный модуль на языке Python, реализующий комплексную обработку многомерных экспериментальных данных, характерных для испытаний технических систем (на примере поршневых двигателей внутреннего сгорания с керамическими покрытиями). Модуль включает алгоритмы: двухступенчатого иерархического усреднения, автоматического определения размерности физических величин, парного t-теста, расчета индекса воспроизводимости, а также полиномиальной регрессии второго порядка и гауссовской регрессии с RBF-ядром для построения поверхностей отклика и интервальных прогнозов. Реализованы средства визуализации (3D-поверхности, контурные графики, тренды, корреляционные диаграммы) и автоматической генерации выходных отчетов. Описана архитектура модуля, основные алгоритмы, сценарии верификации. Приведены результаты применения к реальным данным, подтверждающие корректность и эффективность работы. Программный модуль зарегистрирован в Роспатенте (свидетельство № 2026661664).

Ключевые слова: программный модуль, регрессионный анализ, гауссовская регрессия, визуализация данных, Python, pandas, scikit-learn, двигатель внутреннего сгорания, обработка экспериментальных данных.

*rodioniovakf@yandex.ru

Введение

Обработка многомерных экспериментальных данных, получаемых при испытаниях сложных технических систем (двигателей, турбин, химических реакторов), требует автоматизации, статистической обоснованности и воспроизводимости результатов. Существующие решения – коммерческие пакеты (MATLAB), табличные процессоры или разрозненные скрипты – имеют ограничения по открытости, стоимости, возможности интеграции и обеспечению полноты научной воспроизводимости. Обработка таких данных вручную или с помощью табличных процессоров трудоемка, чревата ошибками и не позволяет в полной мере использовать современные методы регрессионного анализа и машинного обучения [1, 2].

В области математического и программного обеспечения актуальна разработка открытых модулей, сочетающих алгоритмы статистической обработки и регрессионного анализа, методы машинного обучения с оценкой неопределенности, средства автоматической визуализации и генерации отчетов. Существующие программные средства (например, MATLAB с пакетами Curve Fitting Toolbox, Statistics and Machine Learning Toolbox) обладают широкими возможностями, но требуют лицензионных затрат и специальной подготовки. В существующей практике обработки данных моторных испытаний преобладают разрозненные скрипты, каждый из которых решает частную задачу – пересчет размерности, усреднение, расчет КПД или построение графика. Интегрального, открытого и задокументированного модуля, который бы охватывал полный цикл от извлечения данных из файлов до построения регрессионных поверхностей отклика с интервальной оценкой,

до сих пор не представлено. Настоящая работа восполняет этот пробел. Разработан программный модуль на языке Python, поддерживающий иерархическую структуру эксперимента (режим $\rightarrow$ прогон $\rightarrow$ замер). В его составе: взвешенное двухступенчатое усреднение, парный t -тест, физически обоснованная тепловая модель полноты сгорания, расчет эффективного КПД, полиномиальная регрессия второго порядка и гауссовская регрессия с RBF-ядром, а также автоматическая генерация комплекта графиков (поверхности отклика, контурные диаграммы, тренды, корреляционные поля). Особое внимание уделено автоматическому определению единиц измерения углеводородов ($ppm \leftrightarrow \%$) и проверке статистической значимости эффектов. Экспериментальные данные получены в ходе испытаний одноцилиндрового бензинового двигателя УМЗ-341 [3].

Научная новизна представленной работы заключается в следующем:

1. Разработан новый алгоритм двухступенчатого иерархического усреднения экспериментальных данных с автоматическим определением размерности физических величин ($ppm \leftrightarrow \%$). В отличие от известных методов, алгоритм не требует ручной настройки единиц измерения и учитывает структуру многофакторного эксперимента (режимы $\rightarrow$ прогоны $\rightarrow$ замеры).

2. Впервые предложен и программно реализован индекс воспроизводимости как метрика стабильности экспериментальных данных. Индекс позволяет количественно сравнивать воспроизводимость результатов при различных модификациях объекта (например, при замене конструктивного элемента).

3. Выполнено систематическое сопоставление полиномиальной регрессии второго порядка и гауссовской регрессии с RBF-ядром в рамках единого программного модуля для обработки данных технических испытаний. Показано, что гауссовская регрессия, в отличие от полиномиальной, позволяет получать не только точечный прогноз, но и интервальную оценку неопределенности в каждой точке факторного пространства, что критически важно при малом объеме экспериментальных данных (6 режимов). Реализация на базе `scikit-learn` адаптирована для автоматической оптимизации гиперпараметров.

4. Предложена модульная архитектура программного модуля, состоящая из семи независимых компонентов: парсера данных, препроцессора (включая автоопределение размерности), статистического блока, расчетного блока с подключаемыми физическими моделями, регрессионного блока (полиномиальная и гауссовская регрессии), блока визуализации (3D-поверхности, контурные диаграммы, тренды, корреляционные поля) и генератора отчетов (Excel, CSV, PNG, текстовый протокол). В отличие от монолитных скриптов, компоненты допускают замену и адаптацию без изменения остального кода.

5. Разработан алгоритм автоматического распознавания размерности концентраций углеводородов (ppm или $\%$) на основе порогового значения среднего арифметического по массиву данных. Если среднее превышает 1,0, данные считаются заданными в ppm и переводятся в проценты делением на 10 000, в противном случае интерпретируются как проценты. Алгоритм исключает ручную настройку и повышает робастность при смене источников данных.

Архитектура программного модуля, основные алгоритмы и методы обработки

Модуль реализован на языке Python 3.9. В его основе лежит функционально-модульный принцип, позволяющий изолировать ключевые этапы обработки данных, взаимодействие между программными компонентами организовано через передачу данных в виде структур `pandas.DataFrame`.

Вычисления производятся по следующему алгоритму. Для каждого значения x_{ij} прогона i и замера j вычисляется среднее арифметическое по формуле:

$$\bar{x}_i = \frac{1}{m} \sum_{j=1}^m x_{ij}. \quad (1)$$

Дисперсия по прогону оценивается по формуле:

$$s_i^2 = \frac{\sum_{j=1}^m (\bar{x}_{ij} - \bar{x}_i)^2}{m-1}. \quad (2)$$

На втором шаге каждому прогону присваивается вес, обратный его дисперсии (с малой добавкой $\varepsilon = 10^{-9}$ для устойчивости):

$$w_i = \frac{1}{s_i^2 + \varepsilon}. \quad (3)$$

Такой вес уменьшает вклад прогонов с большим разбросом замеров. Итоговое среднее значение по всем прогонам находится как взвешенное среднее:

$$\bar{x} = \frac{\sum_{i=1}^n w_i \bar{x}_i}{\sum_{i=1}^n w_i}. \quad (4)$$

Взвешенное стандартное отклонение (корень из взвешенной дисперсии) вычисляется по формуле:

$$s = \sqrt{\frac{\sum_{i=1}^n w_i (\bar{x}_i - \bar{x})^2}{\frac{n-1}{n} \sum_{i=1}^n w_i}}. \quad (5)$$

Стандартная ошибка среднего равна $SE = \frac{s}{\sqrt{n}}$, а 95%-й доверительный интервал строится как $\bar{x} \pm t_{0.975, n-1} \cdot SE$. Для сравнения двух состояний объекта (например, штатная конструкция и конструкция с покрытием) используется парный t -тест, адаптированный для взвешенных данных [4], в котором разность показателей для каждого режима рассматривается как парная выборка, вычисляется средняя разность и ее стандартное отклонение, затем рассчитывается t -статистика. Нулевая гипотеза об отсутствии эффекта отвергается при $p < 0.05$.

Для количественной оценки стабильности эксперимента в модуль введен индекс воспроизводимости, основанный на разложении общей изменчивости на внутрипрогонную и межпрогонную составляющие. Определяются:

$$\sigma_{intra} = \sqrt{\frac{\sum_{i=1}^n w_i s_i^2}{\sum_{i=1}^n w_i}}, \quad \sigma_{inter} = \sqrt{\frac{\sum_{i=1}^n w_i (\bar{x}_i - \bar{x})^2}{\frac{n-1}{n} \sum_{i=1}^n w_i}}.$$

Индекс воспроизводимости вычисляется как

$$IB = \frac{\sigma_{intra}}{\sigma_{intra} + \sigma_{inter}}, \quad (6)$$

где σ_{intra} – средняя внутрипрогонная дисперсия, σ_{inter} – дисперсия средних по прогонам. Индекс показывает, какая доля общей изменчивости приходится на неконтролируемые флуктуации внутри прогонов (данный индекс лежит в интервале $[0, 1]$). Меньшие значения IB свидетельствуют о лучшей воспроизводимости эксперимента. В известных аналогах (MATLAB, пакет `scipy`) отсутствует специализированная реализация такого индекса для иерархических данных.

Для визуализации и прогнозирования влияния двух входных факторов (частоты вращения коленчатого вала n и мощности нагрузки N) на выходные показатели (эффекты покрытия Δ) модуль поддерживает два альтернативных метода регрессии.

Полиномиальная регрессия второго порядка описывается моделью:

$$\Delta = \beta_0 + \beta_1 n + \beta_2 N + \beta_{11} n^2 + \beta_{22} N^2 + \beta_{12} nN + \varepsilon, \quad (7)$$

где Δ – эффект покрытия (разность показателей «штатный – покрытие»), n – частота вращения коленчатого вала (мин^{-1}), N – мощность (Вт), ε – случайная ошибка (необъясненная часть) с нулевым математическим ожиданием и постоянной дисперсией. Коэффициенты β оцениваются методом наименьших квадратов с использованием функции «`numpy.linalg.lstsq`».

Достоинство метода – простота интерпретации и гладкость получаемых поверхностей, что удобно для поиска экстремумов. Однако полиномиальная модель дает только точечный прогноз и не оценивает неопределенность.

Гауссовская регрессия (Gaussian process regression) реализована с помощью класса GaussianProcessRegressor из библиотеки scikit-learn [2] с ядром RBF (радиально-базисная функция). Ядро имеет вид:

$$k(x, x') = \sigma^2 e^{-\left(\frac{(n-n')^2}{2l_n^2} - \frac{(N-N')^2}{2l_N^2}\right)} + (\sigma_n^2 + \varepsilon_{jitter})\delta_{nn'}, \quad (8)$$

где σ^2 – дисперсия сигнала, l_n , l_N – масштабы длины, σ_n^2 – уровень шума, $\delta_{nn'}$ – символ Кронекера [5], ε_{jitter} – малая добавка для численной устойчивости (аналог ε из формулы (3)). Гиперпараметры ядра автоматически оптимизируются путем максимизации логарифмической функции правдоподобия на обучающих данных. В отличие от полиномиальной регрессии, гауссовская регрессия позволяет строить доверительные интервалы прогноза, и оценки его неопределенности $\sigma(y)$ [6, 7].

В модуле реализована физически обоснованная тепловая модель для расчета полноты сгорания [8], которая устраняет артефакты эмпирической формулы ГОСТ Р 56163-2014 [9]:

$$\eta_{comb} = 1 - 0,31 \cdot CO + 1,34 \cdot CH, \quad (9)$$

где концентрации выражены в объемных процентах. Коэффициенты 0,31 и 1,34 получены как отношения теплот сгорания CO и CH к теплоте сгорания углерода (32,8 МДж/кг). Эффективный КПД [1] вычисляется по формуле:

$$\eta_e = \frac{N}{G_T \cdot Q_H}, \quad (10)$$

где N – мощность на валу (Вт), G_T – часовой расход топлива (кг/ч), $Q_H = 44$ МДж/кг – низшая теплота сгорания бензина.

При работе с исходными данными часто возникает проблема неоднозначного представления концентрации углеводородов CH : в одних протоколах она указана в объемных процентах (%), в других – в миллионных долях (ppm). Для устранения необходимости ручной настройки модуль реализует автоматический алгоритм определения размерности. Вычисляется среднее арифметическое всех значений CH по всему массиву данных. Если это среднее превышает порог 1,0, то данные трактуются как выраженные в ppm и каждый элемент делится на 10 000 для перевода в проценты. В противном случае значения считаются уже приведенными к процентам. Порог выбран эмпирически из физических соображений: содержание CH в отработавших газах не может превышать 1% (обычно десятые доли процента), тогда как в ppm оно достигает нескольких тысяч. Алгоритм робастен к выбросам, так как использует среднее по всей выборке.

Блок визуализации модуля автоматически генерирует следующие типы графиков для каждого анализируемого показателя:

- трехмерные поверхности отклика (как для полиномиальной, так и для гауссовской регрессии) с возможностью отображения стандартного отклонения для гауссовской модели;
- контурные диаграммы с линиями уровня и цветовой заливкой;
- столбчатые диаграммы эффектов покрытия с нанесенными 95% доверительными интервалами;
- графики внутрипрогонных трендов (отображение всех прогонов на одном поле для визуальной оценки воспроизводимости);
- диаграмму рассеяния с линейной регрессией для двух выбранных показателей (например, изменение расхода топлива vs изменение температуры ОГ) с выводом коэффициента корреляции Пирсона и уравнения регрессии.

Все графики сохраняются в формате PNG с разрешением 300 dpi, что пригодно для включения в научные статьи. Параметры отображения (цветовые схемы, размер шрифта, метки осей) могут быть настроены через конфигурационный файл.

Верификация и тестирование модуля

Для подтверждения корректности работы программного модуля проведен комплекс тестов, охватывающий модульное тестирование отдельных алгоритмов, интеграционное тестирование на реальных данных, проверку воспроизводимости на разных платформах и сравнение с эталонными реализациями.

Модульное тестирование реализовано через запуск набора специально подготовленных примеров, для которых правильный ответ рассчитан заранее (вручную или с помощью эталонных программ). При тестировании двухступенчатого усреднения на искусственных данных (5 прогонов $\times$ 3 замера) проверялась корректность формул (1)–(5). Для прогона со значениями [10,10,10] среднее равно 10; для [0,10,20] среднее = 10. Усреднение по прогонам и стандартное отклонение сверялись с ручным расчетом в табличном процессоре – погрешность не превысила 1×10^{-12} . Парный t -тест проверялся на синтетических выборках с известными t -статистикой и p -уровнем; результаты модуля сравнивались с эталонной функцией `scipy.stats.ttest_rel` [10], максимальное различие не превысило 1×10^{-10} . Алгоритм автоматического пересчета $ppt \rightarrow \%$ тестировался на массивах: значения 100–500 ppt (среднее $> 1,0$) корректно делятся на 10 000; значения 0,01–0,5% (среднее $< 1,0$) остаются без изменений. Проверка робастности при наличии выброса 5000 ppt показала, что алгоритм корректно обрабатывает все значения. Для полиномиальной регрессии данные генерировались по известному полиному (5) с добавлением шума (5% амплитуды); оцененные коэффициенты β отличались от исходных в среднем на 0,2%. Для гауссовской регрессии использовались синтетические данные из примера в литературе [6]; модуль воспроизвел прогнозы с погрешностью менее 1×10^{-6} при идентичных гиперпараметрах. Интеграционное тестирование выполнено на полном наборе экспериментальных данных (6 режимов $\times$ 5 прогонов $\times$ 3 замера), описанном в разделе 4. Проверялась корректность парсинга Excel: модуль правильно идентифицировал все 6 режимов, для каждого – 5 прогонов, для каждого прогона – 3 замера. Ручная сверка с исходным файлом дала 100% совпадение. При трех последовательных запусках на одной машине все численные результаты (средние, стандартные отклонения, p -значения, коэффициенты регрессии) совпали до 8-го знака. Для двух режимов (3200/1500 и 2700/500) значения ΔCO_2 , ΔCO , Δg_e , $\Delta \eta_{comb}$ были рассчитаны вручную по формулам (1)–(5); расхождения с результатами модуля находятся в пределах погрешности округления.

Сводные количественные результаты тестирования по основным подсистемам модуля представлены в Табл. 1.

Таблица 1. Результаты верификации модуля.

Подсистема	Тестовый сценарий	Эталон / критерий	Достигнутый результат
Парсер Excel	6 режимов, 5 $\times$ 3 замера	100% совпадение с разметкой	100%
Усреднение (формулы 1–3)	Искусственные данные	Относительная погрешность $< 1e^{-10}$	$1,2e^{-12}$
Парный t -тест	Синтетические выборки	Совпадение с <code>scipy.stats.ttest_rel</code>	$\max diff 1e^{-10}$
Пересчет $ppt \rightarrow \%$	Смесь ppt и %	Автоопределение без ошибок	0 ошибок на 20 тестах
Полиномиальная регрессия	Данные по известному полиному	Погрешность коэфф. $< 1\%$ при шуме 5%	0,2%
Гауссовская регрессия	Пример из [6]	Совпадение прогноза	$diff < 1e^{-6}$
Интеграция (реальные данные)	Два режима	Ручной расчет	$\max diff 0,0005$
Парсер Excel	6 режимов, 5 $\times$ 3 замера	100% совпадение с разметкой	100%

Применение модуля к реальным экспериментальным данным и обсуждение

Модуль был апробирован на данных, полученных в ходе моторных испытаний одноцилиндрового бензинового двигателя УМЗ-341 со штатным поршнем и поршнем с МДО-покрытием (толщина 120 мкм). Эксперимент включал 6 режимов работы, для каждого режима – 5 прогонов по 3 замера. В Табл. 2 приведены вычисленные модулем средние эффекты покрытия (разность «штатный – покрытие») и их статистическая значимость по парному t -тесту (выделены $p < 0,05$).

Таблица 2. Статистические данные по результатам парного t -теста.

Режим (n/N)	ΔCO_2 , % / p	ΔCO , % / p	Δg_e , кг/ч / p	$\Delta\eta_{comb}$ / p	$\Delta\eta_e$ / p
3200/ 1500	+0,08/ 0,025	–0,12/ 0,106	–0,046/ 0,003	+0,020/ 0,048	–0,0022/ 0,003
3200/ 1000	+0,13/ 0,011	–0,09/ 0,020	–0,062/ 0,141	+0,026/ 0,013	–0,0024/ 0,153
3200/ 500	+0,09/ 0,041	–0,11/ 0,117	–0,117/ 0,036	+0,021/ 0,062	–0,0026/ 0,019
2700/ 1000	+0,14/ 0,083	–0,10/ 0,264	+0,078/ 0,219	+0,022/ 0,139	+0,0042/ 0,211
2700/ 500	+0,03/ 0,497	–0,04/ 0,911	–0,018/ 0,856	+0,005/ 0,786	–0,0006/ 0,755
2400/ 500	–0,06/ 0,949	+0,07/ 0,455	+0,001/ 0,990	–0,009/ 0,693	+0,0001/ 0,852

Модуль написан на Python с использованием открытых библиотек [11, 12], что делает его доступным для широкого круга исследователей без дополнительных лицензионных затрат. Программный код может быть легко адаптирован для других типов двигателей (дизельных, газовых) путем изменения справочных параметров и структуры входных данных. Программный модуль зарегистрирован в Федеральной службе по интеллектуальной собственности (Роспатент) [13].

Ограничения и перспективы: на данный момент модуль ориентирован на обработку данных, полученных на одноцилиндровом двигателе с фиксированным набором режимов.

Регрессионные модели в текущей версии поддерживают только два входных фактора (n и N). Хотя это соответствует данному эксперименту, в общем случае число факторов может быть больше (например, температура охлаждающей жидкости, давление наддува). Расширение на многомерную регрессию потребует пересмотра визуализации (3D-графики невозможны для > 2 факторов) и увеличения объема данных. Для автоматического выбора между полиномиальной и гауссовской регрессией можно добавить критерии для автоматического сравнения качества моделей. В будущем планируется добавить поддержку многоцилиндровых двигателей, возможность импорта данных из распространенных форматов (CSV, TDMS, HDF5) без привязки к Excel, а также реализовать методы многокритериальной оптимизации (например, генетические алгоритмы) для поиска наилучших параметров покрытия или режимов работы двигателя.

Заключение

В работе представлен программный модуль, реализующий комплексную обработку экспериментальных данных на основе двухступенчатого статистического анализа, полиномиальной и гауссовской регрессии, автоматической визуализации. Разработана модульная архитектура, включающая парсер данных, препроцессор с автоопределением размерности физических величин ($p_{prt} \leftrightarrow \%$), статистический блок (взвешенное двухступенчатое усреднение, парный t -тест), регрессионный блок (полиномиальная

и гауссовская регрессии), блок визуализации и генератор отчетов. Компоненты независимы и могут быть адаптированы под другие предметные области. Предложен и реализован алгоритм взвешенного усреднения с весами, обратными внутрипрогонной дисперсии. Введен индекс воспроизводимости для количественной оценки стабильности экспериментальных данных. Разработан метод автоматического определения единиц измерения концентраций. Выполнена автоматическая оптимизация гиперпараметров.

Проведено модульное и интеграционное тестирование, сравнение с эталонными реализациями. Модуль является открытым (код может быть опубликован), документированным и воспроизводимым. Перспективы развития включают поддержку многомерной регрессии, импорт из различных форматов (CSV, TDMS), веб-интерфейс. Модуль может быть адаптирован для любых технических систем с повторными измерениями.

Литература:

1. Heywood J.B. Internal Combustion Engine Fundamentals. 2nd ed. McGraw-Hill Education, 2018. 1056 p. ISBN 1260116107, 9781260116106.
2. Pedregosa F., Varoquaux G., Gramfort A., et al. Scikit-learn: Machine learning in Python // Journal of Machine Learning Research. 2011. Vol. 12. P. 2825–2830. DOI: 10.48550/arXiv.1201.0490.
3. Дударева Н.Ю., Ситдигов В.М., Ишемгузин А.А. и др. Снижение токсичности отработавших газов в камере сгорания двигателя внутреннего сгорания // Труды НАМИ. 2022. № 4(291). С. 85–95. DOI: 10.51187/0135-3152-2022-4-83-95 [Dudareva N.Yu., Sitdikov V.M., Ishemguzhin A.A., et al. Reduction of exhaust gas toxicity in the combustion chamber of an internal combustion engine // Proceedings of NAMI. 2022. No. 4(291) P. 85–95. DOI: 10.51187/0135-3152-2022-4-83-95 (in Russian)].
4. McKinney W. Data Structures for Statistical Computing in Python. In: Proceedings of the 9th Python in Science Conference. 2010. P. 56–61. DOI: 10.25080/Majora-92bf1922-00a.
5. Korn G.A., Korn T.M. Mathematical Handbook for Scientists and Engineers: Definitions, Theorems, and Formulas for Reference and Review. 2nd ed. Courier Corporation, 2013. 1152 p. ISBN 0486320235, 9780486320236.
6. Rasmussen C.E., Williams C.K.I. Gaussian Processes for Machine Learning. MIT Press, 2006. 266 p.
7. Williams C.K.I., Rasmussen C.E. Gaussian processes for regression // Advances in Neural Information Processing Systems. 1996. Vol. 8. P. 514–520. ISBN: 0-262-20107-0.
8. Дударева Н.Ю., Коломейченко А.В., Деев В.Б. Влияние керамического покрытия, формируемого методом микродугового оксидирования, на напряженно-деформированное состояние поршней двигателей внутреннего сгорания // Цветные металлы. 2023. № 12. С. 57–62. [Dudareva N.Yu., Kolomeychenko A.V., Deev V.B. Effect of a ceramic coating formed by micro-arc oxidation on the stress-strain state of internal combustion engine pistons // Non-Ferrous Metals. 2023. No. 12. P. 57–62 (in Russian)]. DOI: 10.17580/tsm.2023.12.05.
9. ГОСТ Р 56163-2014. Выбросы загрязняющих веществ в атмосферу. Метод расчета выделений (выбросов) загрязняющих веществ при неорганизованном сжигании нефти и нефтепродуктов на объектах нефтедобычи. М.: Стандартинформ. 2015. [GOST R 56163-2014. Atmospheric emissions of pollutants. Method for calculating pollutant releases (emissions) from the uncontrolled burning of oil and petroleum products at oil production facilities. Moscow: Standartinform, 2015 (in Russian)].
10. Virtanen P., Gommers R., Oliphant T.E., et al. SciPy 1.0: fundamental algorithms for scientific computing in Python // Nature Methods. 2020. Vol. 17. P. 261–272. DOI: 10.1038/s41592-019-0686-2.
11. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. М.: Физматлит, 2006. 816 с. [Kobzar A.I. Applied Mathematical Statistics: For Engineers and Researchers. Moscow: Fizmatlit, 2006. 816 p. (in Russian)]. ISBN 5-9221-0707-0.
12. Hunter J.D. Matplotlib: A 2D Graphics Environment // Computing in Science & Engineering. 2007. Vol. 9. No. 3. P. 90–95. DOI: 10.1109/MCSE.2007.55.

13. Свидетельство о государственной регистрации программы для ЭВМ № 2026661664 «Программный модуль для комплексного статистического и регрессионного анализа экспериментальных данных ДВС» / Авторы: В.В. Антонов, Е.В. Пальчевский, Л.Е. Родионова, В.А. Суворова, Н.Ю. Дударева. Зарегистрировано 22 апреля 2026 г. [Certificate of State Registration of a Computer Program No. 2026661664 “Software Module for Comprehensive Statistical and Regression Analysis of Internal Combustion Engine Experimental Data” / Authors: V.V. Antonov, E.V. Palchevsky, L.E. Rodionova, V.A. Suvorova, N.Yu. Dudareva. Registered on April 22, 2026 (in Russian)].

Благодарности:

Исследование выполнено при финансовой поддержке Министерства образования и науки Российской Федерации в рамках основной части государственного задания высшим учебным заведениям № FRRR 2026-0006.

Об авторах:

АНТОНОВ Вячеслав Викторович, заведующий кафедры автоматизированных систем управления, УУНиТ, antonov.v@bashkortostan.ru

РОДИОНОВА Людмила Евгеньевна, кандидат технических наук, доцент, доцент кафедры автоматизированных систем управления, УУНиТ, rodionovakf@yandex.ru.

СУВОРОВА Вероника Александровна, кандидат технических наук, доцент, доцент кафедры автоматизированных систем управления, УУНиТ, milana_da@mail.ru.

ПАЛЬЧЕВСКИЙ Евгений Владимирович, доцент кафедры индустриального программирования РТУ МИРЭА, teelxp@inbox.ru.

ДУДАРЕВА Наталья Юрьевна, доктор технических наук, доцент, профессор, (Кафедра двигателей внутреннего сгорания), УУНиТ, natalia_jd@mail.ru

ДЬЯЧКОВ Алексей Сергеевич, кандидат технических наук, руководитель управления технологического развития, ООО «Газпромнефть-Цифровые решения», dyachkovAS@uust.ru.

Metadata:

Title: Software system for regression analysis and visualization of multidimensional data from bench tests of internal combustion engines.

Author 1: Vyacheslav Viktorovich Antonov, Dr. Sci. (Eng.), Head of the Department of Automated Control Systems, Ufa University of Science and Technology, 32 Zaki Validi st., 450076 Ufa, Republic of Bashkortostan, Russia, antonov.v@bashkortostan.ru ORCID 0000-0002-5402-9525, Researcher ID (WoS): AAH-5121-2019, Scopus Author ID: 57221337797.

Author 2: Lyudmila Evgenievna Rodionova, Associate Professor of the Department of Automated Control Systems, Ufa University of Science and Technology, 32 Zaki Validi st., 450076 Ufa, Republic of Bashkortostan, Russia, rodionovakf@yandex.ru, ORCID: 0000-0003-4041-0365; Researcher ID (WoS): AAU-3498-2020, Scopus Author ID: 57221335881.

Author 3: Veronika Alexandrovna Suvorova, Associate Professor of the Department of Automated Control Systems, Ufa University of Science and Technology, 32 Zaki Validi st., 450076 Ufa, Republic of Bashkortostan, Russia, milana_da@mail.ru, ORCID: 0009-0009-6005-0695, Scopus Author ID: 57263144400.

Author 4: Evgeniy Vladimirovich Palchevskiy, Cand. Sci. (Eng.), Associate Professor of the Department of Industrial Programming of the Russian Technical University MIREA, 78 Vernadsky pr., 119454 Moscow, Russia, teelxp@inbox.ru. ORCID: 0000-0001-6024-1074, Scopus Author ID: 57220744490.

Author 5: Natalya Yurievna Dudareva, Doctor of Technical Sciences, Associate Professor, Professor of the Department of Internal Combustion Engines, Ufa University of Science and Technology, 32 Zaki Validi st., 450076 Ufa, Republic of Bashkortostan, Russia, natalia_jd@mail.ru.

Author 6: Alexey Sergeevich Dyachkov, Candidate of Technical Sciences, Head of the Technological Development Department at Gazpromneft-Digital Solutions LLC, Bldg. 4, 5 Kyivskaya st., 190013 St. Petersburg, Russia, dyachkovAS@uust.ru.

Abstract: A Python-based software module is presented that implements complex processing of multidimensional experimental data typical for testing technical systems (using the example of internal combustion piston engines with ceramic coatings). The module includes algorithms for two-stage hierarchical averaging, automatic determination of the dimensionality of physical quantities, paired t-test, calculation of the reproducibility index, as well as second-order polynomial regression and Gaussian regression with an RBF kernel for constructing response surfaces and interval forecasts. Implemented visualization tools (3D surfaces, contour plots, trends, and correlation diagrams) and automatic generation of output reports. The module's architecture, basic algorithms, and verification scenarios are described. The results of applying the module to real data are presented, confirming its correctness and effectiveness. The software module is registered with Rospatent (certificate No. 2026661664).

Keywords: software module, regression analysis, Gaussian regression, data visualization, Python, pandas, scikit-learn, internal combustion engine, experimental data processing.