

## Интеграция методов анализа временных рядов геофизических величин в программный комплекс с открытым кодом

М.Э. Арзангулян,<sup>\*</sup> Н.Е. Шапкина<sup>†</sup>

Московский государственный университет имени М.В. Ломоносова, физический факультет, кафедра математики  
Россия, 119991, Москва, Ленинские горы, д.1, стр. 2

(Поступила в редакцию 14.04.2025; подписана в печать 22.05.2025)

В статье представлен разработанный программный комплекс с открытым исходным кодом для анализа временных рядов геофизических величин. Комплекс объединяет различные методы анализа: статистические (ARIMA, SARIMA), спектральные (преобразование Фурье, вейвлет-анализ), нейросетевые (LSTM) и морфологические, а также включает инструменты обнаружения аномалий. Ключевыми особенностями разработки являются модульная архитектура, возможность генерации синтетических данных, гибкая параметризация алгоритмов и интерактивная визуализация результатов. Программный комплекс уже нашел применение в прикладных задачах нашей научной группы, в частности для анализа метеорологических данных с карбонового полигона. Открытость исходного кода обеспечивает возможность адаптации функционала под специфические задачи и дальнейшее развитие комплекса научным сообществом.

PACS: 93.85.-q, 93.30.-w. УДК: 550.3, 519.6

Ключевые слова: временные ряды, геофизические данные, программный комплекс, открытый исходный код, анализ данных, ARIMA, вейвлет-анализ, LSTM.

### ВВЕДЕНИЕ

Анализ временных рядов является фундаментальным инструментом современной науки, позволяющим исследовать динамические процессы в различных областях знаний. В условиях растущей сложности исследовательских задач и ограничения доступа к зарубежному программному обеспечению особую актуальность приобретает разработка собственных инструментов анализа временных рядов.

Существующие решения, такие как открытые библиотеки Python, хотя и предоставляют широкие возможности для анализа, требуют значительных навыков программирования и времени на интеграцию различных компонентов. Это создает существенные барьеры для специалистов, нацеленных на предметную область исследования, а не на технические аспекты реализации.

Для этого был разработан специализированный программный комплекс, пытающийся объединить в себе математическую строгость, простоту использования и гибкость модификации. Комплекс реализован как проект с открытым исходным кодом [1] и уже применяется в научной группе кафедры математики и кафедры математического моделирования и информатики физического факультета МГУ. Особую ценность представляет возможность его использования как опытными исследователями, так и начинающими специалистами из различных областей науки, включая физику, географию и экономику.

### 1. ОПИСАНИЕ МЕТОДОВ, ИСПОЛЬЗУЕМЫХ В КОМПЛЕКСЕ

Для понимания принципов работы разработанного комплекса и обоснования выбора реализованных в нем методов необходимо рассмотреть теоретические основы анализа и прогнозирования временных рядов. Это позволит не только описать математический аппарат, лежащий в основе инструментария, но и продемонстрировать логику построения комплекса, обеспечивающую последовательное применение различных методов анализа — от базовой декомпозиции до сложных прогнозистических моделей.

Временной ряд представляет собой упорядоченную последовательность наблюдений  $s_t$ , где  $t$  принадлежит некоторому множеству моментов времени  $T$ , как правило, дискретному. В отличие от классических выборочных данных, наблюдения временного ряда характеризуются существенной зависимостью от порядка их следования, что требует специальных методов анализа и обработки.

Классическая модель временного ряда включает три основные компоненты: тренд  $T_t$ , отражающий долгосрочную тенденцию изменения данных, циклическую составляющую  $C_t$ , характеризующую периодические колебания, и случайную компоненту  $I_t$ . Математически это может быть представлено в виде аддитивной модели  $y_t = T_t + C_t + I_t$  [2].

Важной характеристикой временных рядов является понятие стационарности, предполагающее неизменность вероятностных свойств ряда во времени. Стационарность требует инвариантности многомерных распределений относительно сдвигов во времени, тогда как слабая (ковариационная) стационарность предполагает лишь постоянство первых двух моментов распределения — математического ожидания и ковари-

<sup>\*</sup> arzanjulian.me19@physics.msu.ru

<sup>†</sup> neshapkina@mail.ru

ационной функции. Проверка стационарности часто осуществляется с помощью специальных статистических тестов, таких как расширенный тест Дики-Фуллера (ADF) [3], тест Квятковского-Филлипса-Шмидта-Шина (KPSS) и тест Филлипса-Перрона. В разработанном программном комплексе реализован тест Дики-Фуллера, который позволяет проверить нулевую гипотезу о наличии единичного корня против альтернативной гипотезы о стационарности ряда. Выбор данного теста обусловлен его широким применением в практике анализа временных рядов и наличием надежной реализации в статистических пакетах.

### 1.1. Статистические методы анализа временных рядов

Одним из наиболее распространенных методов моделирования и прогнозирования временных рядов являются авторегрессионные модели, в частности модель ARMA (AutoRegressive Moving Average) [4]. Данная модель предполагает, что значение временного ряда в момент времени  $t$  может быть выражено через линейную комбинацию его  $p$  предыдущих значений:

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t, \quad (1)$$

где:

- $p$  — порядок авторегрессии, определяющий количество предыдущих значений ряда, которые используются для моделирования текущего значения;
- $q$  — порядок скользящего среднего, указывающий количество предыдущих ошибок прогноза (разностей между фактическими и предсказанными значениями), учитываемых при моделировании текущего значения временного ряда;
- $\phi_i$  и  $\theta_j$  — коэффициенты модели, характеризующие силу влияния соответствующих лаговых значений и ошибок на текущее значение ряда;
- $\varepsilon_t$  — белый шум.

Для применения модели ARMA критически важным является условие стационарности временного ряда. В реальных исследованиях это условие часто нарушается из-за наличия тренда или сезонности. Для преодоления данного ограничения используется модель ARIMA [2], включающая процедуру дифференцирования ряда для его приведения к стационарному виду:

$$(1 - B)^d X_t = \sum_{i=1}^p \phi_i (1 - B)^d X_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t, \quad (2)$$

где:

- $B$  — оператор сдвига назад (лаговый оператор), который при применении к временному ряду возвращает его предыдущее значение:  $BX_t = X_{t-1}$ ;
- $d$  — порядок дифференцирования ряда, который указывает, сколько раз нужно применить оператор  $(1 - B)$  к исходному ряду.

Так, например, для первого дифференцирования (разности первого порядка) получится следующее выражение:  $(1 - B)^1 X_t = X_t - X_{t-1}$ . Процесс дифференцирования ряда проводится до тех пор, пока полученный ряд не станет стационарным.

В обзоре статистических методов важно также рассмотреть понятия сезонности и цикличности временных рядов. Сезонность ряда является частным случаем цикличности и представляет собой регулярные, периодически повторяющиеся колебания с фиксированным и известным периодом (например, годовым, квартальным или месячным). Цикличность же характеризуется колебаниями с переменной длительностью цикла, которая может составлять несколько лет. Для учета сезонной компоненты применяется модификация предыдущей ARIMA модели — SARIMA (Seasonal ARIMA) [4].

Модель SARIMA ( $P, D, Q, s$ ) расширяет возможности ARIMA, позволяя учитывать как краткосрочные зависимости между последовательными наблюдениями, так и долгосрочные сезонные паттерны. Дополнительные параметры модели  $P, D, Q$  характеризуют ее сезонную составляющую модели, где  $P$  определяет порядок сезонной авторегрессии,  $D$  — порядок сезонного дифференцирования,  $Q$  — порядок сезонного скользящего среднего, а  $s$  — длину сезонного цикла. В отличие от несезонных параметров ( $p, d, q$ ), здесь учитываются зависимости между наблюдениями, отстоящими друг от друга на  $s$  периодов. Данная модель особенно эффективна при анализе метеорологических и экономических временных рядов с выраженной сезонной составляющей.

Выбор параметров  $p$  и  $q$  для базовой модели ARIMA осуществляется на основе анализа автокорреляционной (ACF) и частной автокорреляционной (PACF) функций, а также информационных критериев, таких как AIC (критерий Акаике) [5]. При этом порядок дифференцирования  $d$  определяется необходимым количеством преобразований для достижения стационарности ряда, что может быть проверено с помощью ранее упомянутого расширенного теста Дики-Фуллера.

На рис. 1 представлена реализация метода SARIMA в программном комплексе на примере синтетических данных. Комплекс успешно построил прогноз на тестовой выборке с погрешностью по метрике MAPE (средняя абсолютная ошибка в процентах) менее 5%.

**Дифференцирование временного ряда**



**Результаты модели SARIMA(1, 0, 1)(0, 1, 1, 12)**

MAE	RMSE	MAPE
4.6669	5.8560	4.81%

**Прогноз SARIMA и фактические значения**

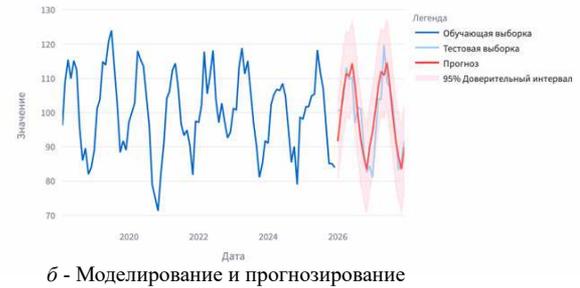


Рис. 1. Реализация метода SARIMA для моделирования и прогнозирования значений ряда

**1.2. Спектральные методы**

Спектральные методы анализа представляют собой мощный инструмент для исследования частотной структуры временных рядов, позволяющий выявлять периодические компоненты и локальные особенности сигналов. В рамках данной работы особое внимание уделяется двум ключевым спектральным методам: преобразованию Фурье и вейвлет-преобразованию.

Дискретное преобразование Фурье (DFT) [6] является фундаментальным методом спектрального анализа, представляющим временной ряд в виде суммы гармонических составляющих. Это преобразование выражается формулой:

$$X(f) = \sum_{t=0}^{N-1} x(t)e^{-2\pi ift/N}, \quad (3)$$

где  $x_t$  обозначает исходный временной ряд,  $N$  — длину ряда,  $f$  — частоту, а  $X(f)$  — спектральные коэффициенты, которые отражают вклад гармонических колебаний различных частот в исходный сигнал. На основе этих коэффициентов вычисляется спектр мощности сигнала:

$$S(f) = |X(f)|^2. \quad (4)$$

Данный подход позволяет эффективно анализировать частотный состав сигнала, однако имеет ограничения при работе с нестационарными процессами, так как не отражает изменения частотных характеристик сигнала во времени.

На рис. 2 представлена реализация преобразования Фурье для обнаружения периодичностей в синтетических данных. Комплекс успешно определил периоды в 30 и 55 измерений без отклонений.

Для преодоления этих ограничений в исследовании применяется вейвлет-преобразование [7, 8], которое обеспечивает возможность анализа сигнала одно-

**Спектр мощности**

**Анализ периодичностей с помощью преобразования Фурье**

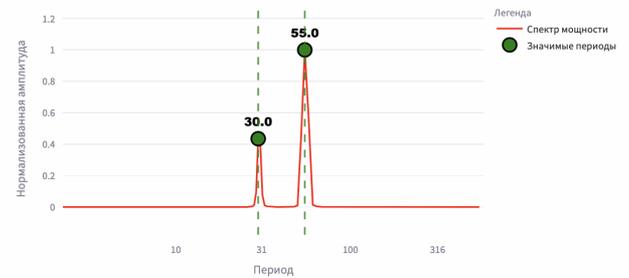


Рис. 2. Реализация Фурье преобразования в программном комплексе на синтетических данных

временно во временной и частотной областях. Непрерывное вейвлет-преобразование [9] определяется интегральным выражением:

$$W(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\psi^* \left( \frac{t-b}{a} \right) dt, \quad (5)$$

где  $a$  — параметр масштаба, обратно пропорциональный частоте (большие значения  $a$  соответствуют низким частотам и наоборот),  $b$  — параметр сдвига по времени,  $\psi(t)$  — материнский вейвлет (базисная функция с ограниченной энергией и нулевым средним значением, форма которой определяет свойства анализа), а  $\psi^*$  — его комплексно-сопряженная функция. Для дискретных временных рядов используется дискретная форма преобразования:

$$W(a, b) = \frac{1}{\sqrt{a}} \sum_{t=0}^{N-1} x(t)\psi^* \left( \frac{t-b}{a} \right). \quad (6)$$

Вейвлет-анализ особенно эффективен при исследова-

нии нестационарных процессов и выявлении локальных особенностей временного ряда, что расширяет его возможности для выявления циклическостей по сравнению с преобразованием Фурье.

На рис. 3 представлена реализация вейвлет-преобразования в программном комплексе на синтетических данных. Комплекс успешно определил периоды в 180 и 400 измерений с отклонением менее 1%.

В контексте разработанного аналитического комплекса спектральные методы находят широкое практическое применение. Они используются для выявления периодичностей в данных, что позволяет обнаруживать циклические паттерны различной природы. Анализ нестационарных характеристик временных рядов с помощью вейвлет-преобразования дает возможность исследовать динамику изменения частотных компонент во времени. Кроме того, спектральные методы эффективно применяются для фильтрации шумовых компонент, что повышает качество исходных данных для дальнейшего анализа. Наконец, декомпозиция сигнала на составляющие различных масштабов позволяет выделять и отдельно анализировать компоненты временного ряда, соответствующие различным временным горизонтам.

Таким образом, интеграция спектральных методов в разработанный аналитический комплекс расширяет его функциональные возможности, обеспечивая глубокий и многосторонний анализ временных рядов.

### 1.3. Нейросетевые методы

Модель Long Short-Term Memory (LSTM) [10] — это специализированный тип рекуррентных нейронных сетей, отличающийся своей способностью эффективно работать с длинными временными рядами. Обычные нейросети сталкиваются с так называемой «проблемой исчезающего градиента» — при анализе длинных последовательностей информация о далеко отстоящих во времени событиях теряется, поскольку влияние ранних данных на обучение сети становится крайне малым.

LSTM решает эту проблему с помощью специальной архитектуры с «памятью» и «воротами», контролирующими потоки информации. Подобно тому, как человек может помнить важные события годичной давности и игнорировать несущественные недельной, LSTM избирательно сохраняет значимую информацию на протяжении сотен временных шагов. Это критически важно для метеорологических прогнозов, где текущие значения могут зависеть от сезонных паттернов и долгосрочных климатических циклов.

Архитектура LSTM основана на ячейке памяти с системой «ворот», каждое из которых выполняет определенную функцию по фильтрации и обновлению информации. Этот процесс описывается следующими уравнениями:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \quad (7)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \quad (8)$$

$$C_t = f_t \times C_{t-1} + i_t \times \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \quad (9)$$

$$h_t = o_t \times \tanh(C_t), \quad (10)$$

где:

- $f_t$  — вектор «забывания», определяющий долю информации, удаляемой из предыдущего состояния памяти  $C_{t-1}$ ;
- $i_t$  — вектор «входа», задающий значимость новой информации для обновления состояния ячейки;
- $C_t$  — текущее состояние памяти, объединяющее фильтрованную историческую и новую информацию;
- $h_t$  — выходное состояние, используемое для прогнозирования;
- $W$  и  $b$  — настраиваемые в процессе обучения параметры сети.

При работе с метеорологическими временными рядами LSTM демонстрирует ряд существенных преимуществ. Прежде всего, модель эффективно справляется с нестационарными данными, что позволяет напрямую анализировать ряды с переменной дисперсией и сложной сезонностью без предварительных преобразований. В отличие от традиционных статистических методов, LSTM автоматически адаптируется к изменяющимся статистическим свойствам данных, что особенно важно при работе с климатическими рядами, характеризующимися долгосрочными трендами и структурными сдвигами.

Другим важным преимуществом является устойчивость к шумам. LSTM может эффективно отфильтровывать случайные выбросы и некорректные измерения, сохраняя при этом значимые закономерности. Это свойство особенно ценно при работе с реальными метеорологическими наблюдениями, которые часто содержат ошибки измерений.

Однако применение LSTM сопряжено с определенными ограничениями. Для качественного обучения модели требуются значительные объемы данных — обычно не менее нескольких сотен наблюдений. Процесс обучения также требует заметно больших вычислительных ресурсов, нежели ранее рассмотренные методы, что также ограничивает его использование. Кроме того, результаты моделирования и прогнозирования с LSTM сложно интерпретировать для обоснования полученных прогнозов. Несмотря на эти ограничения, LSTM зарекомендовала себя как мощный инструмент анализа и прогнозирования временных рядов в задачах, где критически важны учет долговременных зависимостей и адаптация к изменяющимся условиям.

На рис. 4 представлена реализация метода LSTM в программном комплексе на примере синтетических

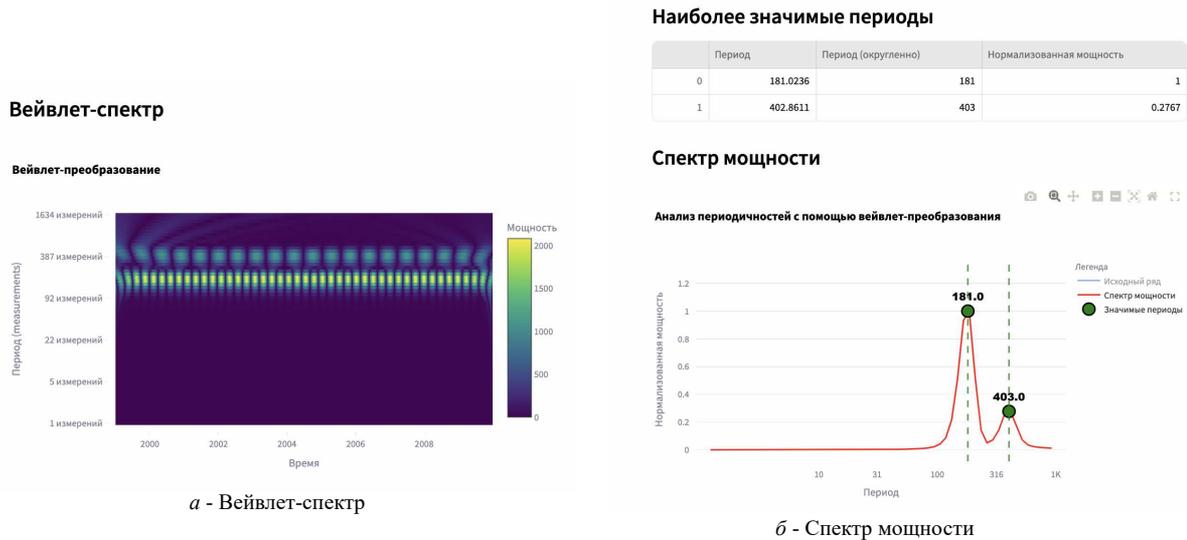


Рис. 3. Реализация вейвлет-преобразования в программном комплексе на синтетических данных

данных. Комплекс успешно построил прогноз на тестовой выборке. Например, коэффициент детерминации для нее равен 0.997.

**1.4. Морфологический анализ**

Морфологический анализ [11] представляет собой уникальный метод обработки сигналов, основанный на исследовании геометрических свойств данных. В отличие от классических прогнозно-ориентированных подходов, таких как ARIMA или нейросетевые архитектуры, данный метод фокусируется на декомпозиции и количественном описании структурных особенностей временного ряда, что делает его незаменимым инструментом для задач семантической сегментации и морфологической классификации сигналов.

Ключевая идея морфологического анализа заключается в построении гладкой аппроксимации исходного ряда, сохраняющей его топологические характеристики. Для этого применяются операции математической морфологии, работающие с сигналом как с множеством в метрическом пространстве.

На рис. 5 представлена реализация моделирования морфологическим методом с помощью кусочно-непрерывных функций одной выпуклости на примере синтетических данных.

**1.5. Обнаружение аномалий**

В анализе временных рядов существует проблема с данными, собираемыми с помощью измерительных приборов — они часто содержат аномалии, т.е. отклонения от нормального поведения, которые могут существенно влиять на результаты анализа. Аномалии

классифицируются на три основные категории: точечные (локальные отклонения отдельных наблюдений), контекстуальные (отклонения, обусловленные особенностями контекста) и коллективные (аномалии групп данных).

Выявление аномальных наблюдений является важной прикладной задачей в предобработке временных рядов. В метеорологических данных аномалии могут быть вызваны как реальными экстремальными явлениями (ураганы, резкие перепады температуры), так и техническими ошибками (сбой датчиков, ошибки записи). В разработанном комплексе реализованы различные методы обнаружения аномалий, которые мы рассмотрим далее. [12]

**Z-score** [13] метод служит базовым инструментом для идентификации выбросов в предположении нормальности распределения. Для временного ряда  $\{x_t\}$  аномальность наблюдения определяется через стандартизованное отклонение:

$$z_t = \frac{x_t - \mu}{\sigma}, \tag{11}$$

где  $\mu$  и  $\sigma$  — выборочные среднее и стандартное отклонение. Пороговое значение  $|z_t| > 3$  соответствует вероятности 0.27% в нормальном распределении. Однако чувствительность метода к экстремальным значениям при расчете параметров требует дополнительной валидации результатов.

Для данных с неизвестным распределением или существенной зашумленностью применяются методы, устойчивые к выбросам. К ним относится **метод Хампеля (MAD)** [13], который использует скользящее окно  $W_t$ . Оно заменяет среднее значение медианой  $\tilde{x}$ , а стандартное отклонение — медианным абсолютным отклонением (MAD):

$$MAD = 1.4826 \cdot \text{median}(|x_i - \tilde{x}|). \tag{12}$$

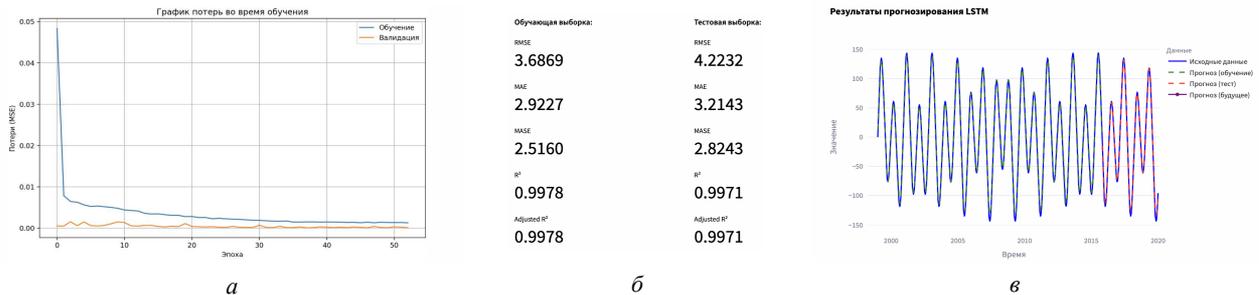


Рис. 4. Реализация метода LSTM в программном комплексе на синтетических данных

**Морфологический анализ**

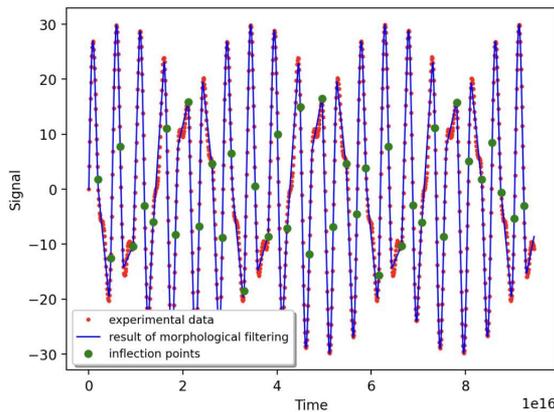


Рис. 5. Реализация моделирования морфологическим методом с помощью кусочно-непрерывных функций одной выпуклости на синтетических данных

Множитель 1.4826 обеспечивает согласованность с нормальным распределением. Наблюдение маркируется как аномалия при  $|x_t - \tilde{x}| > 3 \cdot \text{MAD}$ , что исключает влияние локальных выбросов на пороговое значение.

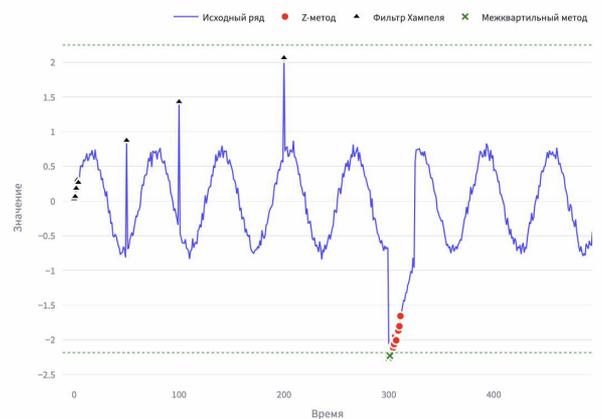
**Межквартильный размах (IQR)** [13] расширяет робастный подход, опираясь на квантили распределения. Для получения границ определяются 25-й ( $Q_1$ ) и 75-й ( $Q_3$ ) перцентили:

$$\text{Границы} = [Q_1 - 1.5 \cdot (Q_3 - Q_1), Q_3 + 1.5 \cdot (Q_3 - Q_1)]. \quad (13)$$

Особенность метода — эффективная работа с асимметричными распределениями, где классические параметрические критерии дают смещенные оценки.

В программном комплексе можно сравнить результаты работы разных методов обнаружения аномалий, в том числе на синтетически созданных аномалиях. На рис. 6 представлен пример анализа синтетического ряда на наличие аномалий. Комбинация описанных методов позволила обнаружить все добавленные во временной ряд аномальности.

**Обнаруженные аномалии**



**Статистика аномалий**

Z-метод	Фильтр Хампеля	Межквартильный метод
13	8	2

Рис. 6. Анализ синтетического ряда на наличие аномалий

**2. АРХИТЕКТУРА ПРОГРАММНОГО КОМПЛЕКСА**

Программный комплекс реализован с использованием современных технологий и архитектурных решений, обеспечивающих его эффективное функционирование и дальнейшее развитие.

**2.1. Технологическая основа**

Фундаментом комплекса служит Python 3.12, выбранный за его широкие возможности в области научных вычислений и анализа данных. Технологический стек включает специализированные библиотеки для анализа данных (numpy, pandas, statsmodels, scikit-learn), визуализации (altair, matplotlib, plotly, seaborn) и машинного обучения (tensorflow). Особое место занимают библиотеки для специализированных вычислений, такие как PyWavelets и cvxpy, обеспечивающие реализацию вейвлет преобразования и морфологического анализа.

## 2.2. Архитектура

Архитектура комплекса основана на принципах модульности и расширяемости. Важным архитектурным решением стало разделение кода на независимые модули, каждый из которых отвечает за конкретный метод анализа временных рядов. Такой подход обеспечивает не только четкую структуризацию кода, но и возможность легкого расширения функционала.

Модульная структура комплекса включает несколько основных компонентов: аналитическое ядро с реализацией методов анализа, утилитарный слой с вспомогательными функциями, документационный модуль с теоретическими описаниями и интерфейсный слой, реализованный на базе Streamlit.

## 2.3. Система обработки данных

В основе системы обработки данных лежит механизм работы с временными рядами в формате CSV, который является наиболее распространенным и универсальным способом хранения и обмена данными в области анализа временных рядов. Реализована валидация входных данных, обеспечивающая корректность временной индексации и числовых значений. Производительность системы оптимизирована за счет встроенного механизма кэширования результатов вычислений, предотвращающего избыточные расчеты.

## 2.4. Механизм расширения функционала

Особое внимание уделено обеспечению возможности расширения функционала комплекса. Реализованный механизм основан на шаблонном подходе к добавлению новых методов анализа через открытый репозиторий GitHub. Такое решение позволяет исследователям участвовать в развитии комплекса, дополняя его функционал под свои требования, сохраняя при этом целостность его архитектуры и согласованность интерфейса.

В отличие от существующих коммерческих решений, разработанный комплекс предоставляет полный контроль над инструментарием анализа и возможность его модификации под специфические исследовательские задачи, что особенно важно в контексте научных исследований.

# 3. РЕАЛИЗАЦИЯ ФУНКЦИОНАЛА КОМПЛЕКСА

## 3.1. Генерация синтетических данных

Помимо простой загрузки настоящего временного ряда, в комплексе реализован гибкий механизм генерации синтетических временных рядов, позволяющий

создавать ряды ежедневных измерений с различными характеристиками. Пользователь может формировать сложные временные ряды путем аддитивного наложения до шести гармонических компонент произвольной частоты и амплитуды, добавления белого шума контролируемой интенсивности и линейной трендовой составляющей с заданным наклоном.

## 3.2. Методы анализа и их параметризация

Каждый реализованный метод анализа предоставляет пользователю возможность настройки ключевых параметров. Для моделей семейства ARIMA доступна настройка параметров авторегрессии, скользящего среднего и порядка интегрирования  $(p, q, d, P, Q, D)$ . При работе с вейвлет-преобразованием пользователь может выбирать тип материнского вейвлета, а для нейросетевых моделей LSTM настраиваются параметры обучения и прогнозирования.

## 3.3. Визуализация и экспорт результатов

Визуализация результатов анализа реализована с использованием библиотек Altair и Plotly, обеспечивающих создание интерактивных графиков. Комплекс поддерживает экспорт результатов в различных форматах:

- Числовые данные в формате CSV (например, для дальнейшего анализа).
- Графические материалы в форматах PNG и JPG.

## 3.4. Пользовательский интерфейс

Интерфейс комплекса построен на принципах простоты и интуитивной понятности [14]. Навигация между методами осуществляется через общее меню, доступное после загрузки или генерации временного ряда. Теоретическая информация о методах представлена в виде интерактивных раскрывающихся секций с использованием LaTeX-разметки для математических формул и описаний.

## 3.5. Производительность и масштабируемость

Производительность комплекса определяется аппаратными возможностями используемой системы. Для оптимизации работы реализовано кэширование результатов вычислений средствами фреймворка Streamlit, что позволяет избежать повторных расчетов при неизменных входных данных. Время обработки данных масштабируется в зависимости от размера временного ряда и сложности применяемого метода анализа.

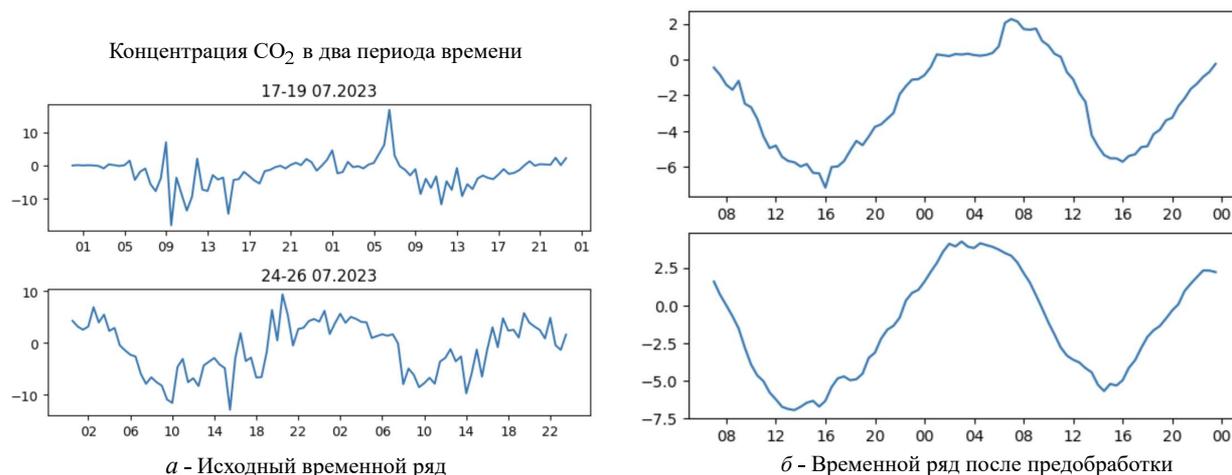


Рис. 7. Пример результатов использования программного комплекса в предобработке данных с карбонового полигона в Республике Коми

#### 4. ПРИМЕРЫ ИСПОЛЬЗОВАНИЯ КОМПЛЕКСА ДЛЯ РЕШЕНИЯ ФИЗИЧЕСКИХ ЗАДАЧ

Разработанный программный комплекс нашел практическое применение в двух ключевых направлениях: образовательном процессе и прикладных научных исследованиях. Его функциональные возможности были апробированы в рамках сотрудничества с научной группой географического факультета МГУ, занимающейся анализом данных карбонового полигона в Республике Коми [15].

В образовательной деятельности комплекс используется для формирования у студентов научной группы базовых навыков работы с временными рядами. Это позволяет на начале научного пути опробовать методы обработки данных, описанные ранее, и проверить методы на синтетических данных в рамках одного интерфейса с интегрированными теоретическими материалами.

При обработке метеорологических данных карбонового полигона комплекс использовался для первичного анализа многопараметрических временных рядов, включающих измерения концентрации  $\text{CO}_2$ , температуры воздуха, атмосферного давления, турбулентности, скорости и направления ветра, а также фотосинтетически активной радиации. На этапе предобработки проводилось сглаживание временных рядов методом скользящего среднего.

Перспективы внедрения комплекса связаны с его модульной архитектурой и кроссплатформенностью. Открытость исходного кода обеспечивает возможность адаптации функционала под специфические задачи, что особенно значимо для научных групп, работающих с уникальными типами данных, а также возможность поддержки развития комплекса представителями научного сообщества.

#### ЗАКЛЮЧЕНИЕ

Разработка специализированного программного комплекса для анализа временных рядов решает актуальную задачу создания доступного и функционального инструмента для исследователей. Комплекс сочетает математическую строгость, модульную архитектуру и удобство использования, что делает его универсальным решением как для опытных специалистов, так и для начинающих пользователей. Он успешно интегрирует методы классического статистического анализа, спектральные подходы, нейросетевые технологии и морфологический анализ, предоставляя широкий спектр возможностей для изучения временных рядов.

Ключевыми преимуществами комплекса являются открытость исходного кода и гибкость архитектуры. Это позволяет пользователям адаптировать инструмент под свои задачи, расширять функционал и вносить вклад в его развитие. Реализованные методы — от ARIMA и SARIMA до LSTM и вейвлет-преобразований — обеспечивают обработку как стационарных, так и нестационарных данных. Особое внимание уделено визуализации результатов и экспорту данных, что упрощает интеграцию комплекса в исследовательские процессы.

Практическая ценность комплекса подтверждена его применением в образовательной деятельности и научных исследованиях. В образовательной среде он помогает студентам освоить базовые методы анализа временных рядов, а в прикладных задачах — эффективно обрабатывать метеорологические данные с экспериментальных площадок, например, карбоновых полигонов.

Перспективы развития комплекса связаны с добавлением новых методов анализа, улучшением интерфейса и повышением производительности. Открытая структура позволяет интегрировать инновационные подходы, обеспечивая актуальность инструмента на фоне быстро

меняющихся научных требований. Таким образом, разработанный комплекс не только отвечает современным

требованиям к анализу временных рядов, но и создает основу для дальнейшего прогресса в этой области.

- [1] Арзангулян М.Э. Универсальный программный комплекс для анализа временных рядов: репозиторий с открытым исходным кодом. GitHub, 2025. Дата обращения: 07.04.2025.
- [2] Бокс Дж., Дженкинс Г., Рейнсел Г., Льюнг Г. Анализ временных рядов: прогнозирование и управление. М.: Альпина Паблишер, 5-е изд., 2019. пер. с англ.
- [3] Dickey D.A., Fuller W.A. // Journal of the American Statistical Association, **74**(366). 427 (1979).
- [4] Носко В.П. Введение в регрессионный анализ временных рядов. Учебное пособие. М., 2002.
- [5] Akaike H. Information theory and an extension of the maximum likelihood principle. Proceedings of the 2nd International Symposium on Information Theory, pages 267–281, 1974.
- [6] Fourier Jean-Baptiste Joseph Theorie analytique de la chaleur. Firmin Didot, Paris, 1822.
- [7] Morlet J., Arens G., Fourgeau E., Giard D. // Geophysics, **47**(2). 203 (1982).
- [8] Morlet J., Arens G., Fourgeau E., Giard D. // Geophysics, **47**(2). 222 (1982).
- [9] Grossmann A., Morlet J. // SIAM Journal on Mathematical Analysis, **15**(4). 723 (1984).
- [10] Albeladi K., Zafar B., Mueen A. // International Journal of Advanced Computer Science and Applications, **14**(1). (2023).
- [11] Пытывев Ю.П., Чуличков А.И. Методы морфологического анализа изображений. М.: ФИЗМАТЛИТ, 2009.
- [12] Chandola V., Banerjee A., Kumar V. // ACM Computing Surveys, **41**(3). 1 (2009).
- [13] Shimizu Y. // Frontiers in Psychology, **12**. 819854 (2022).
- [14] Арзангулян М.Э. Универсальный программный комплекс для анализа временных рядов: веб-интерфейс. Streamlit Cloud, 2025. Дата обращения: 07.04.2025.
- [15] Министерство науки и высшего образования Российской Федерации Карбоновые полигоны. Электронный ресурс, 2025. Дата обращения: 07.04.2025.

## Integration of Geophysical Time Series Analysis Methods into an Open-Source Software Package

M.E. Arzangulyan<sup>a</sup>, N.E. Shapkina<sup>b</sup>

Department of Mathematics, Faculty of Physics, Lomonosov Moscow State University. Moscow 119991, Russia  
E-mail: <sup>a</sup>arzangulyan.me19@physics.msu.ru, <sup>b</sup>neshapkina@mail.ru

The article presents an open-source software complex for analyzing time series of geophysical quantities. The complex integrates various analytical methods: statistical (ARIMA, SARIMA), spectral (Fourier transform, wavelet analysis), neural network (LSTM), and morphological approaches, as well as tools for anomaly detection. The key features include modular architecture, synthetic data generation capabilities, flexible algorithm parameterization, and interactive visualization of results. The software complex is already applied in researches of our scientific group, particularly for analyzing meteorological data from carbon monitoring sites. The open-source nature ensures the possibility of adapting functionality for specific tasks and further development of the complex by the scientific community.

PACS: 93.85.-q, 93.30.-w.

*Keywords:* time series, geophysical data, software complex, open source, data analysis, ARIMA, wavelet analysis, LSTM.

*Received 14 April 2025.*

### Сведения об авторах

1. Арзангулян Марк Эмилевич — магистр; e-mail: arzangulyan.me19@physics.msu.ru.
2. Шапкина Наталья Евгеньевна — канд. физ.-мат. наук, доцент; e-mail: neshapkina@mail.ru.