

# АДАПТАЦИОННЫЙ ПОДХОД К ИМПУТАЦИИ ОДНОМЕРНЫХ ВРЕМЕННЫХ РЯДОВ СПУТНИКОВЫХ ДАННЫХ НА ОСНОВЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

**Ханнанов Наиль Камилевич**

аспирант, Уфимский университет науки и технологий  
nael20000@yandex.ru

## ADAPTIVE APPROACH TO THE IMPUTATION OF ONE-DIMENSIONAL TIME SERIES OF SATELLITE DATA BASED ON MACHINE LEARNING METHODS

**N. Khananov**

*Summary.* Modern satellite observation systems form continuous time series of measured parameters, which are used in the tasks of analyzing and predicting the dynamics of observed phenomena, however, the presence of gaps resulting from equipment failures, interference in communication channels, etc. increases the level of uncertainty in the interpretation of satellite data and reduces the effectiveness of sources of information of this kind. Simple linear interpolation, classical statistical models, and neural network models are most often used in such situations. However, classical interpolation schemes ignore the complex dynamics of signals and have significant methodological errors, because of which they reduce accuracy on long fragments of gaps. Statistical models are sensitive to the assumption of stationarity and accumulate error in the long absence of data. Neural network models demonstrate excessive complexity for short passes with high demands on computing resources. The paper considers an adaptive method of imputation of one-dimensional satellite time series, which includes preprocessing and comparing three imputation tools — linear interpolation, an autoregressive integrated moving average model, and a recurrent neural network. It is shown that with short gaps, simple methods provide a quality comparable to a neural network, but with an increase in the length of gaps, the neural network demonstrates a stable advantage. The values of the coefficient of determination remain positive and retain practical value, whereas the autoregressive model and linear interpolation are characterized by an increase in the average absolute error and a decrease in the coefficient of determination up to negative values. Based on the revealed dependencies, an algorithm for selecting a method based on the length of the pass is proposed, which forms the basis for adaptive imputation of satellite time series.

*Keywords:* data imputation, time series, satellite data, linear interpolation, statistical models, recurrent neural networks, adaptive method selection.

*Аннотация.* Современные спутниковые системы наблюдения формируют непрерывные временные ряды измеряемых параметров, которые используются в задачах анализа и прогноза динамики наблюдаемых явлений, однако наличие пропусков, возникающих вследствие сбоев аппаратуры, помех в каналах связи и пр. способствует повышению уровня неопределенности при интерпретации спутниковых данных и снижает эффективность источников информации такого рода. Чаще всего в таких ситуациях применяются простая линейная интерполяция, классические статистические модели и нейросетевые модели. Однако, классические интерполяционные схемы игнорируют сложную динамику сигналов и имеют существенные методические ошибки, в результате чего снижают точность на длинных фрагментах пропусков. Статистические модели чувствительны к предположению стационарности и накапливает ошибку при продолжительном отсутствии данных. Нейросетевые модели демонстрируют избыточную сложность для коротких пропусков при высоких требованиях к вычислительным ресурсам. В работе рассматривается адапционный метод импутации одномерных спутниковых временных рядов, включающий предобработку и сравнение трёх инструментов импутации — линейной интерполяции, авторегрессионной интегрированной модели скользящего среднего, а также рекуррентной нейронной сети. Показано, что при коротких пропусках простые методы обеспечивают сопоставимое с нейросетью качество, однако при увеличении длины разрывов устойчивое преимущество демонстрирует нейросеть. Значения коэффициента детерминации остаются положительными и сохраняют практическую ценность, тогда как авторегрессионная модель и линейная интерполяция характеризуются ростом средней абсолютной ошибки и снижением коэффициента детерминации вплоть до отрицательных значений. На основе выявленных зависимостей предложен алгоритм выбора метода по длине пропуска, формирующий основу для адаптивной импутации спутниковых временных рядов.

*Ключевые слова:* импутация данных, временные ряды, спутниковые данные, линейная интерполяция, статистические модели, рекуррентные нейронные сети, адаптивный выбор метода.

### Введение

Современные спутниковые системы формируют непрерывные временные ряды измерений, которые используются в широком спектре прикладных задач. В частности, космические аппараты ACE и DSCOVR

выступают ключевыми источниками потоковых данных для оперативных систем предупреждения, защищающих энергетические сети и спутниковую навигацию от внешних возмущений [1]. Поскольку алгоритмы прогнозирования в таких системах требуют строго регулярных входных данных, задача импутации пропущенных зна-

чений является обязательным этапом предварительной обработки данных для сохранения работоспособности моделей [2]. В задачах атмосферного мониторинга данные спутника DSCOVR применяются для количественной оценки аэрозольной нагрузки и анализа состава среды [3]. В этом контексте методы реконструкции сигналов необходимы для устранения шумов и пробелов, вызванных оптическими помехами, что позволяет получить целостную структуру данных для дальнейшего анализа. Аналогичным образом телеметрия группы спутников Swarm используется для калибровки ионосферных моделей и детектирования аномалий в регистрируемых сигналах [4,5]. Временные ряды спутниковых измерений часто содержат пропуски, обусловленные отказами приборов, потерями телеметрии, орбитальными затенениями и помехами в каналах связи из-за чего возникает потребность восстановления данных обеспечивающего согласованную агрегацию и доступ к разнородным источникам для корректных научных и прикладных выводов [6,7]. Выбор метода заполнения пропусков напрямую влияет на качество последующего прогнозирования и статистических выводов. В частности, для одномерных временных рядов разные методы могут давать существенно отличающиеся оценки точности прогноза [8].

Эмпирические исследования последних лет подтверждают, что классические методы импутации такие как линейная интерполяция и ARIMA (от англ. AutoRegressive Integrated Moving Average) и их вариации сохраняют свою практическую ценность благодаря интерпретируемости и низкой вычислительной сложности. Однако в задачах со сложной нелинейной динамикой и выраженной сезонностью современные нейросетевые модели такие как LSTM (от англ. Long Short-Term Memory) часто демонстрируют более высокую точность восстановления. Обзоры и сравнительные исследования по глубокому обучению для спутниковых временных рядов подчёркивают растущую роль методов глубокого обучения в импутации недостающих значений и пространственно-временном восстановлении наблюдений [9,10].

Практическая значимость качественного восстановления спутниковых временных рядов проявляется в прикладных задачах: ряд работ, использующих LSTM-архитектуры для прогноза индексов геомагнитной активности, демонстрирует высокую предсказательную способность рекуррентных сетей и указывает на необходимость обеспечения надёжности входных временных рядов для снижения неопределённости прогноза [9, 11]. В частности, стабильность восстановленных рядов критична для диагностики индуцированных токов по данным наземных магнитометров [12]. При этом динамическая визуализация разнородных геопространственных данных повышает ситуационную осведомленность и облегчает интерпретацию артефактов восстановления

[13]. Анализ подобных артефактов указывает на проблему слабой адаптированности известных методов восстановления временных рядов к разнородным пропускам спутниковых данных.

Целью исследования является разработка и апробация метода выбора оптимальной стратегии импутации временного ряда в зависимости от временной протяженности пропусков. Предполагается, что такая адаптивная стратегия позволит повысить качество восстановленных рядов и, как следствие, надёжность последующего анализа и прогнозирования в различных прикладных задачах.

**Исходные данные и их предварительная обработка**

В качестве экспериментальной базы исследования используются временные ряды параметров солнечного ветра и межпланетного магнитного поля (ММП), полученные из открытых источников Центра прогнозирования космической погоды NOAA. Исходный массив данных представляет собой последовательность изменений с минутным временным разрешением, содержащую значения ортогональных компонент вектора магнитной индукции ( $B_x, B_y, B_z$ ), выраженные в нанотеслах (нТл), модуль полного вектора  $B_T$ , нТл, а также пространственные координаты космического аппарата в системе GSM выраженные в градусах (табл. 1).

Таблица 1.

Фрагмент исходных данных

Дата/время, гггг-мм-дд чч:мм:сс	$B_x$ , нТл	$B_y$ , нТл	$B_z$ , нТл	LON <sub>GSM'</sub> град.	LAT <sub>GSM'</sub> град.	$B_T$ , нТл
2025-12-14 17:23:00	-1.15	-3.74	0.74	252.86	10.69	3.98
2025-12-14 17:24:00	-1.35	-3.68	0.90	249.92	12.93	4.02
2025-12-14 17:25:00	-1.40	-3.66	0.87	249.10	12.50	4.02
2025-12-14 17:26:00	-1.31	-3.69	0.94	250.41	13.50	4.03
2025-12-14 17:27:00	м1.33	-3.75	0.72	250.50	10.21	4.04
2025-12-14 17:28:00	-1.42	-3.68	0.97	248.90	13.81	4.06
2025-12-14 17:29:00	-1.43	-3.61	0.97	248.31	14.01	4.00

Для апробации рассматриваемых методов импутации в качестве целевой переменной выбрана вертикаль-

ная компонента ММП ( $B_z$ ), динамика которой является критически важной для задач геофизического мониторинга и прогнозирования. Перед процедурой импутации к данным применяется предварительная обработка для устранения высокочастотного шума и случайных выбросов. К временным рядам применяется фильтр низких частот Баттерворта 4-го порядка. Выбор данного типа фильтра обусловлен его плоской амплитудно-частотной характеристикой (АЧХ) в полосе пропускания, что обеспечивает минимальное искажение полезного сигнала. Квадрат АЧХ фильтра Баттерворта порядка  $n$  описывается выражением (1).

$$|H(w)|^2 = \frac{1}{1 + (f / f_c)^{2n}} \quad (1)$$

где  $n = 4$  — порядок фильтра, определяющий крутизну спада АЧХ в полосе подавления;  $f$  — частота сигнала, мГц;  $f_c \approx 0.55$  мГц — частота среза при выбранном периоде среза равном 30 минутам.

На рис. 1а представлена АЧХ фильтра, демонстрирующая резкий спад коэффициента передачи в области частот выше 0.55 мГц, что соответствует периодам колебаний короче 30 минут. На рис. 1б показан результат наложения фильтра на реальный временной ряд компоненты  $B_z$ : обработанный сигнал сохраняет основную динамику и структуру исходных данных, избавляясь при этом от шумовой составляющей.

### Методология адаптивной импутации данных на основе интеллектуальных подходов

В настоящее время для восстановления временных рядов используется широкий набор инструментов, начиная с простых статистических приёмов и заканчивая моделями глубокого обучения. При работе со спутниковыми данными чаще применяются гибридные нейросетевые схемы, способные учитывать пространственно-временную структуру сигнала [14,15]. Успешность применения таких подходов подтверждена и на смежных геофизических задачах [16]. В то же время, специфи-

ку геомагнитных данных эффективно учитывают методы, основанные на физике источника сигнала [17], а также современные подходы к реконструкции на базе концепции цифровых двойников [18].

Восстановление временных рядов представляет собой междисциплинарную задачу, в которой встречаются как традиционные статистические процедуры, ориентированные на простоту реализации и минимальные вычислительные затраты, так и сложные архитектуры глубокого обучения, предназначенные для работы с нелинейными динамическими процессами и большими массивами данных. Между этими полюсами располагаются интерполяционные алгоритмы и модели временных рядов, учитывающие сезонность и автокорреляционные зависимости, а также гибридные и ансамблевые подходы, объединяющие возможности разных методов. Такая классификация не только отражает эволюцию подходов от простых к более сложным, но и демонстрирует их взаимодополняемость, указывая на необходимость создания систем, способных гибко переключаться между алгоритмами в зависимости от характеристик исходных данных и структуры пропусков.

Среди множества методов восстановления временных рядов особый интерес представляют линейная интерполяция, ARIMA и LSTM, так как они отражают три различных подхода к решению задачи. Линейная интерполяция демонстрирует возможности простейших статистических процедур, ARIMA представляет классические модели с учётом автокорреляции и сезонности, а LSTM воплощает современные методы глубокого обучения, способные обрабатывать нелинейные зависимости и длинные последовательные пропуски. Рассмотрение именно этих методов позволяет проследить эволюцию подходов и сопоставить их эффективность в различных сценариях.

Линейная интерполяция относится к числу наиболее простых и широко используемых методов восстановления временных рядов. Её суть заключается в аппроксими-

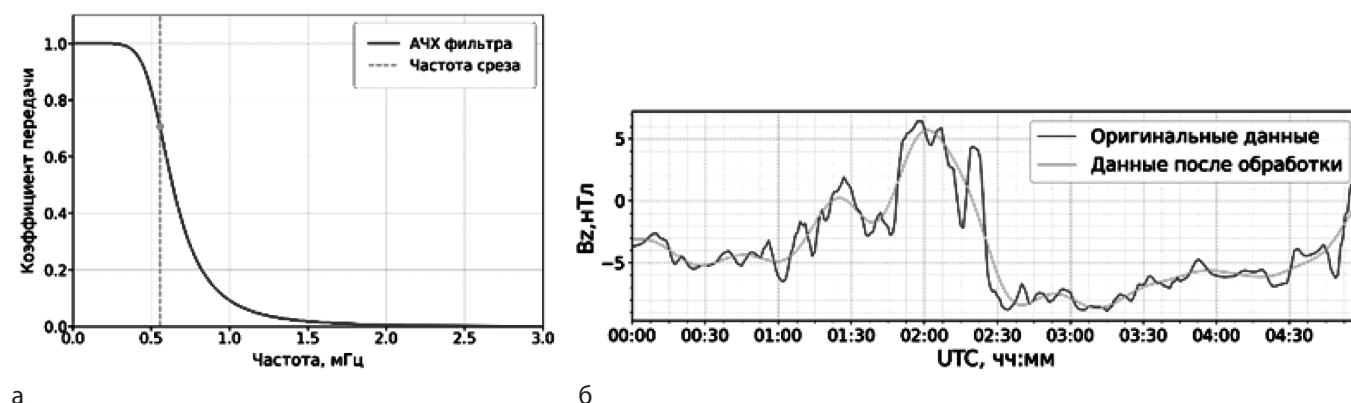


Рис. 1. АЧХ используемого фильтра Баттерворта 4-го порядка (а), фрагмент временного ряда компоненты  $B_z$  до и после фильтрации низких частот (б)

мации пропущенных значений отрезками прямых между двумя соседними известными точками. Для временного ряда  $\{x_t\}_{t=1}^T$  при наличии пропуска в момент времени  $t$ , значение восстанавливается согласно выражению 2:

$$\hat{x}_t = x_{t_1} + \frac{t - t_1}{t_2 - t_1} \cdot (x_{t_2} - x_{t_1}), \quad (2)$$

где  $t$  — момент времени с пропуском;  $t_1$  — индекс ближайшего наблюдения слева (до пропуска);  $t_2$  — индекс ближайшего наблюдения справа (после пропуска);  $x_{t_1}, x_{t_2}$  — значения ряда в этих точках;  $\hat{x}_t$  — восстановленное значение.

Данный метод обладает высокой вычислительной эффективностью и обеспечивает непрерывность ряда, однако он не учитывает ни долгосрочных зависимостей, ни сезонных компонентов, вследствие чего его точность резко снижается при длинных блоках пропусков и в условиях сложной динамики сигналов

Модели авторегрессии, интегрирования и скользящего среднего такие как ARIMA занимают важное место в задачах анализа и восстановления временных рядов. Их ключевая особенность состоит в способности учитывать автокорреляции и сезонные компоненты данных. Общий вид модели ARIMA  $(p, d, q)$  записывается согласно выражению 3.

$$\Phi_p(B)(1 - B)^d x_t = \Theta_q(B)\varepsilon_t, \quad (3)$$

где,  $x_t$  — значение временного ряда в момент времени  $t$ ;  $B$  — оператор сдвига, такой что  $Bx_t = x_{t-1}$ ;  $\Phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$  — оператор авторегрессии порядка  $p$ ;  $(1 - B)^d$  — оператор дифференцирования порядка  $d$ , устрояющий тренд и обеспечивающий стационарность ряда;  $\Theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$  — оператор скользящего среднего порядка  $q$ ;  $\varepsilon_t$  — белый шум (независимые случайные величины с нулевым математическим ожиданием и постоянной дисперсией).

В случае выраженной сезонности применяется расширение SARIMA, включающее сезонные операторы с периодом  $s$ . Достоинством ARIMA является её интерпретируемость и устойчивость при умеренной доле пропусков, однако ограничения проявляются при нелинейной динамике и значительных блоковых потерях данных.

LSTM представляет собой разновидность рекуррентных нейронных сетей, специально разработанную для обработки длинных временных зависимостей и нелинейной динамики. В отличие от традиционных рекур-

рентных нейронных сетей, LSTM использует систему «ворот», регулирующих поток информации: входные, забывающие и выходные ворота. Формально обновления состояния ячейки задаются набором выражений (4).

$$\left. \begin{aligned} f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\ i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \\ C_t &= f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \\ o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t &= o_t \odot \tanh(C_t) \end{aligned} \right\} \quad (4)$$

где  $x_t$  — входной вектор в момент времени  $t$ ;  $h_{t-1}$  — скрытое состояние на предыдущем шаге;  $C_{t-1}, C_t$  — состояние памяти на шаге  $t - 1$  и  $t$ ;  $f_t$  — забывающий вектор, контролирует, какая часть информации из памяти будет сохранена;  $i_t$  — входной вектор, контролирует, какая новая информация поступает в память;  $\tilde{C}_t$  — кандидатное значение ячейки памяти, которое формируется на основе текущего входа и предыдущего скрытого состояния и может быть добавлено к долгосрочному состоянию  $C_t$ ;  $o_t$  — выходной вектор, регулирует, какая часть информации памяти будет выдана наружу;  $\sigma(\cdot)$  — сигмоидальная функция активации;  $\tanh(\cdot)$  — гиперболический тангенс;  $\odot$  — поэлементное умножение;  $W_f, W_i, W_C, W_o$  — матрицы весов для каждого их ворот;  $b_f, b_i, b_C, b_o$  — векторы смещений.

Такая архитектура позволяет сети запоминать долгосрочные зависимости, игнорировать несущественные колебания и учитывать нелинейные эффекты, что делает LSTM особенно эффективной при работе с геофизическими временными рядами, где характерны сложные вариации и большие последовательные пропуски.

Однако для эффективного внедрения подобных вычислительно емких архитектур в реальный процесс мониторинга необходима единая процедурная схема, регламентирующая подготовку данных. Структурно этот подход реализован в виде методики представленной на рис. 2.

Схема метода иллюстрирует последовательный конвейер обработки данных, включающий этапы предобработки потока измерений, детектирования интервала пропусков и расчета признаков временного ряда. Ключевым элементом метода является блок адаптивного выбора модели, который направляет задачу импутации на вход оптимального алгоритма перед финальной постобработкой восстановленного ряда.

**Оценка погрешности восстановления в зависимости от длины импутируемых фрагментов**

Эксперимент направлен на сравнение точности восстановления пропусков в одноминутном ряду  $B_2(t)$

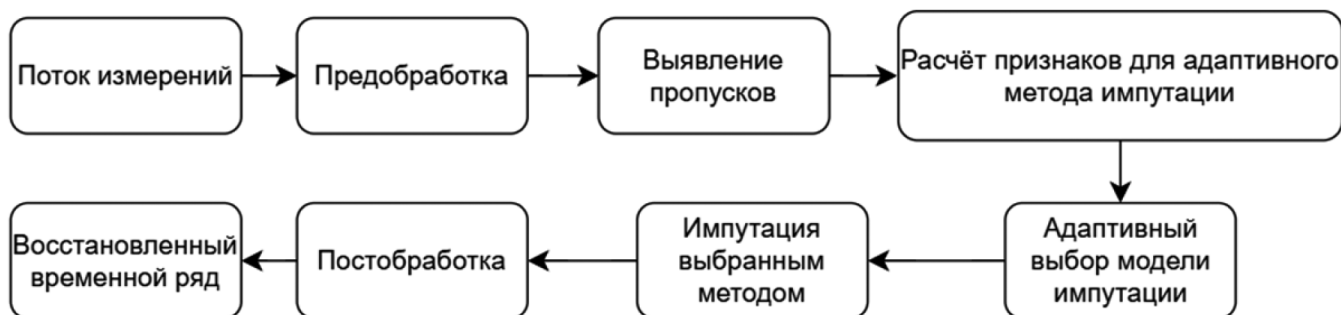


Рис. 2. Структурная схема адаптационного метода импутации

в фиксированных трёхчасовых окнах (180 точек) при различных длинах непрерывных пропусков. Для анализа отбираются  $N=100$  непересекающихся сегментов без пропусков; выборка стартовых индексов сегментов производится из множества допустимых окон длины 180 точек (равномерная частота «одна минута») с последующей случайной подвыборкой. Длины искусственно создаваемых пропусков варьируются в диапазоне  $L \in \{1, 3, 5, \dots, 59\}$ . Пропуски для расчёта метрик размещаются внутри сегмента с защитными отступами по 10 точек от краёв, а для визуализации формируется дополнительный сценарий с пропуском по центру окна. Вся процедура выполняется при фиксированном множестве случайных чисел, конфигурация и артефакты эксперимента сериализуются в отдельные файлы, что задаёт базовый ориентир для моделей машинного обучения и подчёркивает роль масштабируемых ARIMA-подходов на больших горизонтах [19].

Сравниваются три метода восстановления: линейная временная интерполяция; двусторонний авторегрессионный метод ARIMA(5,0,0), двунаправленная рекуррентная нейросеть LSTM с последующей аффинной стыковкой на границах пропуска.

Анализ кривых на рис. 3 показывает устойчивое преимущество LSTM на средних и больших пропусках. Это преимущество связано с наличием длинной нелинейной памяти, позволяющей модели использовать широ-

кий контекст и типичные шаблоны динамики ряда, в том числе при смене режимов. В отличие от неё, ARIMA при заполнении длинных интервалов фактически выполняет многократный прогноз, накапливая ошибку и постепенно вырождаясь в усреднение, поэтому не воспроизводит реальные колебания сигнала внутри разрыва. Линейная интерполяция опирается только на две границы пропуска и игнорирует внутреннюю структуру процесса, поэтому остаётся конкурентоспособной лишь при очень коротких разрывах, где траектория реально близка к линейной. Таким образом, по мере роста длины пропуска преимущества методов, использующих широкий контекст и нелинейную динамику, последовательно увеличиваются, тогда как линейные и локальные схемы демонстрируют резкое ухудшение качества.

На рис. 4 представлены результаты импутации пропусков во временном ряду компоненты  $B_z$  при длительности разрыва 69 отсчётов. В обоих примерах линейная интерполяция формирует монотонную траекторию между границами и, как следствие, не воспроизводит внутреннюю динамику сигнала при наличии выраженной нелинейности и смены тренда. ARIMA частично восстанавливает форму за счёт учета автокорреляционной структуры, однако в условиях нестационарности демонстрирует сглаживание экстремумов, что проявляется в недооценке амплитуды и искажении фазы. Реконструкция LSTM в большей степени сохраняет характерные перегибы и экстремумы внутри разрыва, визуально со-

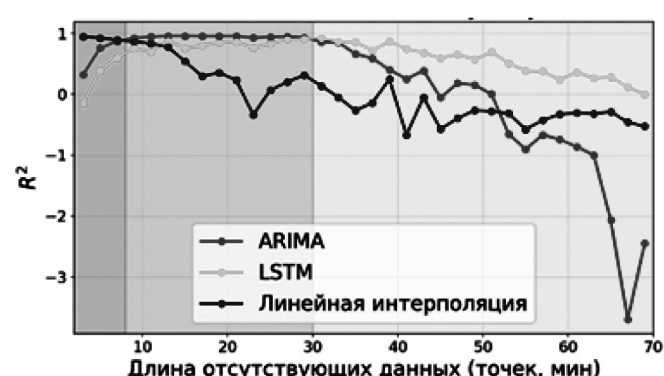
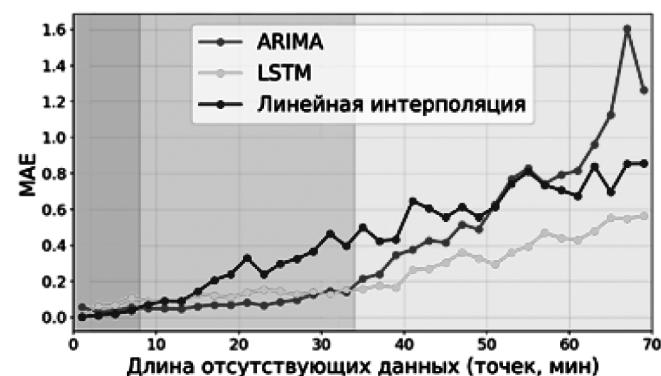
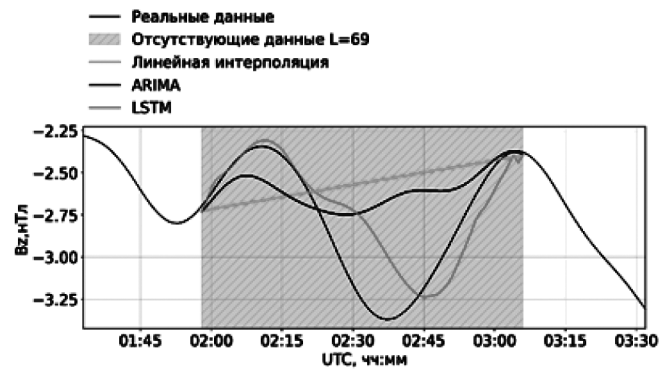
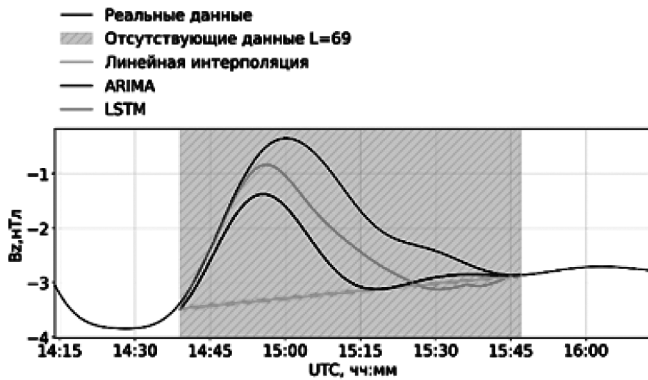


Рис. 3. Зависимость MAE (а) и  $R^2$  (б) восстановления от длины пропуска наблюдений



а б  
 Рис. 4. Результаты импутации отсутствующего фрагмента 17.01.2022 (а) и 22.03.2022 (б)

гласовываясь с реальной траекторией при сложной динамике, хотя в отдельных участках возможны локальные отклонения, связанные с неопределённостью внутри длинного интервала без наблюдений.

**Синтез адаптивного алгоритма восстановления данных с учетом длины пропусков**

Как показал сравнительный анализ, ни один из рассмотренных методов восстановления не является универсальным. Статистические алгоритмы эффективны на коротких разрывах и обладают низкими вычислительными затратами, а нейросетевые архитектуры необходимы для для качественной реконструкции динамики сигнала на длительных интервалах отсутствия данных. Следовательно, применение единого алгоритма для всего потока данных является неоптимальным.

Для решения этой проблемы предлагается подход, основанный на адаптивном переключении алгоритмов. С формальной точки зрения, задача сводится к построению решающего правила, которое для каждого конкретного случая осуществляет выбор стратегии восстановления, минимизирующей ошибку.

Практическая мотивация подхода исходит из специфики анализа геофизических временных рядов, где корректность восстановления зависит от локальной топологии данных [17]. Предполагается, что учет характеристик разрыва позволяет снизить неопределенность и применить инструмент адекватный масштабу потери информации.

Цель метода — минимизировать риск восстановления на пропусках в соответствии с выражением 5.

$$\arg \min_{a \in A} L(X, \hat{X}^{(a)}; M) \tag{5}$$

где  $A$  — конечное множество методов восстановления данных;  $a \in A$  — конкретный метод из множества  $A$ ;  $X \in \mathbb{R}^{T \times d}$  — исходный многомерный временной ряд

(длина  $T$ , число признаков  $d$ );  $M \in \{0, 1\}^{T \times d}$  — маска наблюдаемости:  $M_{t,j} = 1$  если значение  $x_{t,j}$  наблюдалось, и 0 если оно пропущено;  $\hat{X}^{(a)} = f_a(X, M)$  — восстановленный ряд, полученный методом  $a$  при подаче ему исходных данных и маски;  $f_a(\cdot)$  — «исполнитель» выбранного метода;  $L(X, \hat{X}^{(a)}; M)$  — функция ошибки восстановления, которая сравнивает истинный ряд с восстановленным на позициях пропусков;  $\arg \min_{a \in A}$  — множество всех методов  $a$ , дающих минимальное значение ошибки.

В предлагаемом подходе выбор оптимального алгоритма  $a$  осуществляется на основе вектора характеристик пропуска, в частности, длительности непрерывного интервала отсутствия данных. Именно этот параметр является ключевым критерием для определения наиболее подходящей вычислительной стратегии.

На рис. 5 представлена блок-схема алгоритма, формализующего правило выбора метода восстановления в зависимости от длины пропуска и реализующего обработку каждого разрыва в цикле. Такой алгоритм позволяет объединить вычислительно простые методы для коротких пропусков и более выразительные модели для длинных разрывов, снижая риск ухудшения качества за счёт использования заведомо неподходящего метода восстановления.

**Обсуждение результатов**

В рамках развития предложенного подхода для повышения обобщающей способности выводов требуется переход от усреднённого анализа по совокупной выборке к стратифицированному рассмотрению различных типов динамики временных рядов. Например, режимов высокой и низкой вариабельности, присутствия или отсутствия выраженной сезонности, различной доли и структуры пропусков, а также влияния характеристик шума и возможных дрейфов. Такая стратификация позволит более точно описать области применимости каж-

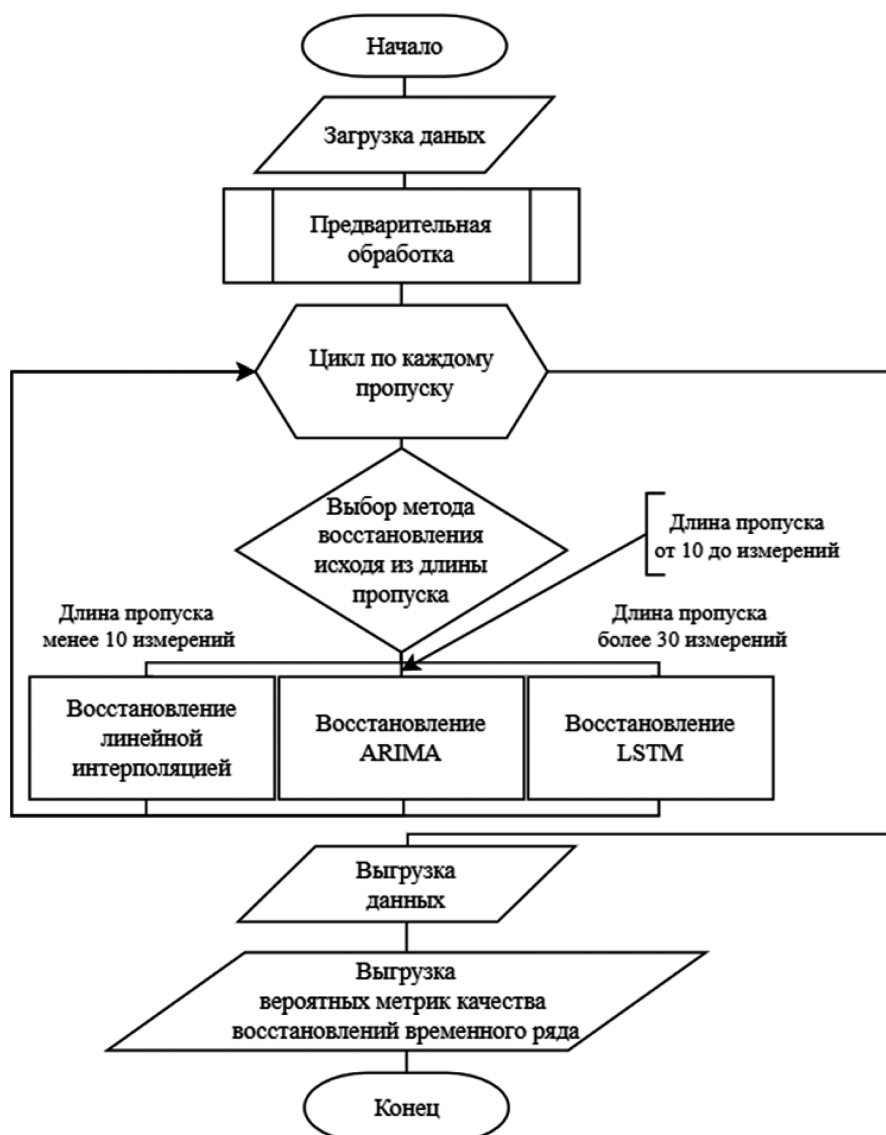


Рис. 5. Алгоритм восстановления пропусков временных рядов исходя из длины пропуска

ного метода и сформировать практические рекомендации, привязанные к конкретным классам спутниковых данных.

Отдельным направлением развития представляется расширение набора сравниваемых подходов с использованием единого протокола «честной» валидации, ориентированного на современные бенчмарки по импутации временных рядов [20]. В этот набор целесообразно включить как классические вероятностные модели и современные методы понижения размерности, так и специализированные архитектуры глубокого обучения для обработки неполных временных последовательностей. Также естественным шагом выглядит построение метамодели адаптивного выбора метода, которая, опираясь на признаки временного ряда и структуру маски пропусков, автоматически подбирает стратегию восстановления. Дополнительный потенциал связан с ансам-

блевыми схемами, сочетающими локальные (простые и интерпретируемые) и глубинные (контекстно-ориентированные) модели, что может обеспечить одновременно устойчивость к различным типам разрывов и повышение точности восстановления в реальных задачах обработки спутниковых временных рядов.

### Заключение

В работе предложен метод импутации одномерных временных рядов спутниковых данных на примере измерений  $B_z$  компоненты ММП. В ходе исследования осуществлено сопоставление эффективности трёх инструментов импутации — линейной интерполяции, стохастической модели ARIMA и рекуррентной нейронной сети LSTM. Численные эксперименты с искусственно сформированными разрывами длительностью от 1 до 59 минут и последующей оценкой качества по метри-

кам средней абсолютной ошибки (MAE) и коэффициента детерминации ( $R^2$ ) продемонстрировали выраженную зависимость эффективности указанных методов от длины интервала отсутствующих данных. Показано, что для коротких разрывов (до порядка 10 минут) линейная интерполяция и ARIMA обеспечивают сопоставимое с LSTM качество восстановления, что делает применение сложных моделей на таком горизонте избыточным и позволяет рассматривать простые методы в качестве рационального базового решения.

При увеличении продолжительности разрывов до диапазона порядка 10–35 минут выявлено устойчивое преимущество LSTM, проявляющееся в минимальных значениях MAE и стабильно высоких значениях  $R^2$  (около 0,8 — 0,9) на фоне систематического «выпрямления» траекторий линейной интерполяцией и, как следствие, потери части динамики сигнала внутри окна пропуска. Для длинных разрывов (порядка 40 минут и более) преимущество моделей с нелинейным представлением становится наиболее выраженным: LSTM обеспечивает положительные значения  $R^2$ , сохраняющие практическую прогностическую ценность, тогда как для ARIMA наблюдается резкое ухудшение вплоть до отрицательных значений  $R^2$ , а у линейной интерполяции

существенно возрастает MAE. Применение аффинной стыковки на границах разрывов снижает уровень скачков для всех рассмотренных методов, однако не устраняет принципиальные ограничения линейных моделей на больших интервалах восстановления.

Полученные зависимости качества восстановления от длины пропусков позволяют обосновать гибридную стратегию выбора метода импутации: линейная интерполяция оказывается целесообразной для очень малых окон, LSTM следует рассматривать как основной инструмент для средних и длинных разрывов, а ARIMA может использоваться как интерпретируемый ориентир на умеренных длинах при близкой к стационарной динамике временного ряда. Тем самым формируется методологический вывод для задач обработки спутниковых временных рядов: модели с богатым контекстным представлением и нелинейной аппроксимацией (LSTM и их современные модификации) должны рассматриваться в качестве метода по умолчанию при наличии значимой доли длинных пропусков, тогда как простые схемы импутации сохраняют значение быстрого базового уровня, эталона для сравнения и компонента ансамблевых решений.

#### ЛИТЕРАТУРА

- Smith A. W., Forsyth C., Rae I. J., et al. On the considerations of using near real time data for space weather hazard forecasting // *Space Weather*. 2022. Vol. 20, No. 7. e2022SW003098. DOI: 10.1029/2022SW003098.
- Kondrashov D., Shprits Y., Ghil M. Gap Filling of Solar Wind Data by Singular Spectrum Analysis // *Geophysical Research Letters*. 2010. Vol. 37. L15101. DOI: 10.1029/2010GL044138.
- Ahn C., Torres O., Jethva H., Tiruchirapalli R., Huang L.-K. Evaluation of aerosol properties observed by the DSCOVR/EPIC instrument from the Earth–Sun Lagrange 1 orbit // *Journal of Geophysical Research: Atmospheres*. 2021. Vol. 126, No. 12. e2020JD033651. DOI: 10.1029/2020JD033651.
- Cherniak I., Zakharenkova I. High-latitude ionospheric irregularities: differences between ground— and space-based GPS measurements during the 2015 St. Patrick's Day storm // *Earth, Planets and Space*. 2016. Vol. 68. Art. 136. DOI: 10.1186/s40623-016-0506-1.
- Fæhn Follstad A., Clausen L. B. N., van den IJssel J., et al. Two-Dimensional Reconstruction of Ionospheric Plasma Density Variations in the Polar Cap Ionosphere // *Space Weather*. 2020. Vol. 18, No. 2. e2019SW002406. DOI: 10.1029/2019SW002406.
- Alwateer M., Atlam E.-S., Abd El-Raouf M. M., Ghoneim O. A., Gad I. Missing Data Imputation: A Comprehensive Review // *Journal of Computer and Communications*. 2024. Vol. 12. P. 53–75. DOI: 10.4236/jcc.2024.1211004.
- Воробьев А. В., Воробьева Г. Р., Юсупова Н. И. Концепция единого пространства геомагнитных данных // *Труды СПИИРАН*. 2019. Т. 18, № 2. С. 390–415. DOI: 10.15622/sp.18.2.390–415.
- Park M. J., Park E. J., Sohn S. Y. Comparative Study of Time Series Analysis Algorithms (ARIMA, SARIMA, LSTM, SVM) for Short-Term Load Forecasting // *Sensors*. 2024. Vol. 24, No. 22. Art. 7205. DOI: 10.3390/s24227205.
- Miller L., Pelletier C., Webb G. I. Deep Learning for Satellite Image Time Series Analysis: A Review. arXiv:2404.03936, 2024.
- Kataoka R., Nakano S. Reconstructing Solar Wind Profiles Associated With Extreme Magnetic Storms: A Machine Learning Approach // *Geophysical Research Letters*. — 2021. — Vol. 48, No. 23. — e2021GL096275. — DOI: 10.1029/2021GL096275.
- Shammi N. J., Ashraf T., Iftly R. A., Rahman M. Z. Forecasting Geomagnetic Storms for Earth Using LSTM with Solar Wind and IMF Features // *Proc. 2024 International Conference on Innovations in Science, Engineering and Technology (ICISSET-2024)*. IEEE, 2024. DOI: 10.1109/ICISSET62123.2024.10939246.
- Vorobev A., Soloviev A., Pilipenko V. [et al.] An Approach to Diagnostics of Geomagnetically Induced Currents Based on Ground Magnetometers Data // *Applied Sciences (Switzerland)*. 2022. Vol. 12, No. 3. DOI 10.3390/app12031522.
- Воробьев А. В., Воробьева Г. Р. Подход к динамической визуализации разнородных геопроостранственных векторных изображений // *Компьютерная оптика*. 2024. Т. 48, № 1. С. 123–138. DOI: 10.18287/2412-6179-СО-1279.
- Appel M. Efficient data-driven gap filling of satellite image time series using deep neural networks with partial convolutions // *Artificial Intelligence for the Earth Systems*. 2024. Vol. 3, No. 2. DOI: 10.1175/AIES-D-22-0055.1.

15. Qian L., Wu L., Zhang Z., Cui Y. A gap filling method for daily evapotranspiration of global flux data sets based on deep learning // *Journal of Hydrology*. 2024. Vol. 641, No. 5. Art. 131787. DOI: 10.1016/j.jhydrol.2024.131787.
16. Bessenbacher V., Schumacher D. L., Hirschi M., Seneviratne S. I., Gudmundsson L. Gap-Filled Multivariate Observations of Global Land–Climate Interactions // *Journal of Geophysical Research: Atmospheres*. 2023. Vol. 128, No. 24. e2023JD039099. DOI: 10.1029/2023JD039099.
17. Воробьев А. В., Воробьева Г. Р. Индуктивный метод восстановления временных рядов геомагнитных данных // *Труды СПИИРАН*. 2018. № 2(57). С. 104–133. DOI: 10.15622/sp.57.5.
18. Воробьев А.В., Пилипенко В.А. Подход к восстановлению геомагнитных данных на базе концепции цифровых двойников // *Солнечно-земная физика*. 2021. Т. 7, № 2. С. 53–62. DOI: 10.12737/szf-72202105.
19. Wang X., Kang Y., Hyndman R. J., Li F. Distributed ARIMA models for ultra-long time series // *International Journal of Forecasting*. 2023. Vol. 39, No. 3. P. 1163–1184. DOI: 10.1016/j.ijforecast.2022.05.001.
20. Du W., Wang J., Qian L., Yang Y., Ibrahim Z., Liu F., Wang Z., Liu H., Zhao Z., Zhou Y., Wang W., Ding K., Liang Y., Prakash B. A., Wen Q. TSI-Bench: Benchmarking Time Series Imputation [Электронный ресурс]. — arXiv:2406.12747, 2024. — DOI: 10.48550/arXiv.2406.12747.

---

© Ханнанов Наиль Камилевич (nael20000@yandex.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»