

**Койкова Виктория Павловна**, студентка, кафедра информатики, Уфимский университет науки и технологий

**Ханнанов Наиль Камилевич**, аспирант, кафедра информатики, Уфимский университет науки и технологий

**Чубукова Кристина Юрьевна**, студентка, кафедра информатики, Уфимский университет науки и технологий

**Садыков Арсэн Тимурович**, студент, кафедра информатики, Уфимский университет науки и технологий

**Нгуен Тхи Диеу Линь**, студент, кафедра информатики, Уфимский университет науки и технологий

## **ПОДХОД К ВОССТАНОВЛЕНИЮ ВРЕМЕННЫХ РЯДОВ СПУТНИКОВЫХ ДАННЫХ**

Статья посвящена восстановлению пропусков в одноминутных спутниковых временных рядах геофизических параметров (на примере компоненты межпланетного магнитного поля  $B_z$ ). Сравняются три подхода, отражающие разные классы методов: линейная интерполяция, классическая модель ARIMA и рекуррентная нейросеть LSTM. Экспериментальная схема включает 3-часовые окна, в которых синтетически создаются непрерывные пропуски длиной 1–59 мин, качество импутации оценивается по MAE и коэффициенту детерминации  $R^2$  внутри разрывов. Показано, что для коротких пропусков простые методы сопоставимы по ошибке с LSTM, однако при увеличении длины окна преимущество LSTM становится устойчивым: сохраняются положительные значения  $R^2$  и умеренные MAE, тогда как ARIMA демонстрирует ухудшение вплоть до отрицательных  $R^2$ , а линейная интерполяция даёт возрастающую систематическую погрешность формы. Аффинная стыковка на границах снижает краевые несогласованности, но не компенсирует модельные ограничения линейных подходов на длинных

интервалах. На основе полученных зависимостей предлагается гибридный подход выбора метода по длине пропуска.

*Ключевые слова:* импутация данных; восстановление временных рядов; спутниковые наблюдения; межпланетное магнитное поле; геомагнитная активность; линейная интерполяция; аффинная стыковка; адаптивный выбор метода; космическая погода.

## **Введение**

Современные спутниковые системы наблюдения обеспечивают непрерывный поток данных о состоянии геофизической среды, включая параметры магнитного поля Земли, параметры солнечного ветра, ионосферные характеристики и климатические переменные. Эти данные используются в научных исследованиях и прикладных задачах (системы поддержки принятия решений при наклонно-направленном бурении, оперативный мониторинг космической погоды, управление энергосистемами и др.). При этом временные ряды спутниковых измерений часто содержат пропуски, обусловленные отказами приборов, потерями телеметрии, орбитальными затенениями и помехами в каналах связи -- явление, требующее системной обработки и восстановления данных для корректных научных и прикладных выводов [1].

Влияние пропусков особенно существенно при анализе динамических геофизических процессов: потеря даже небольшого фрагмента ряда может исказить оценку амплитуды и частоты вариаций, скрыть или исказить корреляционные связи и сезонные составляющие, а также снизить точность прогнозов. Работа по оценке влияния методов импутации показывает, что выбор способа заполнения пропусков напрямую сказывается на качестве последующего прогнозирования и статистических выводов; в частности, при анализе унитарных временных рядов различия между методами приводят к заметно отличающимся оценкам неопределённости прогноза [2].

Одновременно за последние годы была получена убедительная подтверждающая эмпирика: классические статистические методы (линейная интерполяция, ARIMA (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 глубокие модели такие как LSTM (англ. Long Short-Term Memory) зачастую показывают лучшую точность восстановления и прогнозирования, особенно при относительно небольшой доле разрозненных пропусков. Обзоры и сравнительные исследования по глубокому обучению для спутниковых временных рядов подчёркивают растущую роль методов глубокого обучения в gap-filling и пространственно-временном восстановлении наблюдений [3].

Практическая значимость качественного восстановления спутниковых временных рядов проявляется в прикладных задачах, связанных с оценкой и прогнозированием геомагнитной активности: ряд работ, использующих LSTM-архитектуры для прогноза индексов геомагнитной активности (DST, SYM-H и др.), демонстрирует высокую предсказательную способность глубоких рекуррентных сетей и указывает на необходимость обеспечения надёжности входных временных рядов для снижения неопределённости прогноза [4, 5].

Целью статьи является разработка и апробация мета-модели, автоматически выбирающей оптимальную стратегию восстановления временного ряда в зависимости от его характеристик: доли и паттерна пропусков, скорости и интенсивности изменений сигнала, наличия трендов и сезонности, а также уровня шумов. Предполагается, что такая адаптивная стратегия позволит повысить качество восстановленных рядов и, как следствие, надёжность последующего анализа и прогнозирования в задачах геофизики и инженерных приложений.

### **Обзор существующих методов восстановления временных рядов**

Восстановление временных рядов – задача с широким спектром методов: от простых статистических приёмов до современных глубоких нейросетевых архитектур и гибридных схем, учитывающих пространственно-временную структуру данных (особенно важно для спутниковых временных рядов). Недавние обзоры и сравнительные исследования подчёркивают, что выбор метода должен учитывать природу пропусков, межпеременную корреляцию и пространственную структуру данных, для спутниковых продуктов всё чаще показаны преимущества методов, учитывающих пространственную информацию (CNN, частичные свёртки) в комбинации с рекуррентными блоками [6].

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.

На рисунке показана иерархическая схема, в основе которой лежат пять основных классов: статистические, интерполяционные, модели временных рядов, методы машинного и глубокого обучения, а также гибридные и ансамблевые подходы. Каждый класс содержит подуровни, отражающие конкретные техники – от простого заполнения средними и линейной интерполяции до ARIMA, рекуррентных сетей и диффузионных моделей. В

нижней части схемы указаны рекомендованные сценарии применения: статистика и интерполяция эффективны при малой доле пропусков, ARIMA – при выраженной сезонности, нейросетевые и генеративные модели – при нелинейной динамике и больших разрывах, а ансамбли – при смешанных сценариях. Схема наглядно демонстрирует, что выбор метода должен зависеть от характеристик данных, что обосновывает необходимость адаптивных мета-моделей.

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 аппроксимации 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 (1):

$$\hat{x}_t = x_{t_1} + \frac{t - t_1}{t_2 - t_1} \cdot (x_{t_2} - x_{t_1}), \quad (1)$$

где,  $t$  – момент времени с пропуском;

$t_1$  – индекс ближайшего наблюдения слева (до пропуска);

$t_2$  – индекс ближайшего наблюдения справа (после пропуска);

$x_{t_1}, x_{t_2}$  – значение ряда в этих точках;

$\hat{x}_t$  – восстановленное значение.

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 (2).

$$\Phi_p(B)(1-B)^d x_t = \Theta_q(B)\varepsilon_t, \quad (2)$$

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  $d$ , устраняющий 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 рекуррентных нейронных сетей, LSTM использует систему «ворот»,

регулирующих поток информации: входные, забывающие и выходные ворота.

Формально обновления состояния ячейки задаются выражениями (3) – (8).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \quad (3)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \quad (4)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \quad (5)$$

$$C_t = f_t \square C_{t-1} + i_t \square \tilde{C}_t, \quad (6)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \quad (7)$$

$$h_t = o_t \square \tanh(C_t). \quad (8)$$

Здесь:

$x_t$  – 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;

$\square$  – 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.

### **Методика предварительной обработки одноминутных геофизических рядов.**

Рассматривается вещественная последовательность наблюдений  $x_1, \dots, x_N$  со штампами времени  $t_1 < \dots < t_N$ . Медианный шаг дискретизации обозначим  $\Delta t_{med} = median(t_i - t_{i-1})$ . Цель обработки – подавление импульсных выбросов и высокочастотной помеховой составляющей при сохранении медленной динамики и локальной кривизны, без фазовых искажений и без «протечки» через разрывы данных.

Первым инструментом предварительной обработки данных является робастное подавление выбросов (фильтр Хэмпела).

Для каждого (i) на окне длины  $L=2k+1$  вычисляются локальная медиана согласно выражению 9.

$$m_i = median\{x_{i-k}, \dots, x_{i+k}\}. \quad (9)$$

и локальная медианная абсолютная девиация вычисляется согласно выражению 10.

$$D_i = median\{|x_j - m_i| : j \in [i-k, i+k]\}. \quad (10)$$

Точка  $x_i$  признаётся выбросом и заменяется на  $m_i$ , если выполняется условие (11).

$$|x_j - m_i| > \tau_i, \tau_i = \alpha \cdot 1.4826 \cdot D_i, \alpha \approx 3. \quad (11)$$

Коэффициент 1.4826 обеспечивает согласование с стандартным отклонением при нормальном шуме, при  $D_i \approx 0$  используется устойчивый положительный порог (например, медиана ненулевых значений  $D_i$  в окне).

Следующим инструментом предварительной обработки данных является низкочастотная фильтрация без фазового смещения (фильтр Баттерворта). Задаётся временной порог  $T_c$  (в минутах или секундах), определяющий границу сохраняемой динамики. Пусть частота Найквиста

$f_N = 1/(2\Delta t_{med})$ , а частота среза  $f_c = 1/T_c$ , нормированная частота среза  $k = f_c / f_N \in (0,1)$ . Применяется цифровой низкочастотный фильтр Баттерворта порядка ( $n$ ) с амплитудно-частотной характеристикой, вычисляемой согласно выражению 12.

$$|H(w)|^2 = \frac{1}{1 + (w/w_c)^{2n}} \quad (12)$$

где  $w_c = 2\pi f_c$ . Для исключения фазового сдвига выполняется двунаправленное фильтрование (прямой и обратный проход) последовательности после шага Хэмпела; тем самым результирующая фаза равна нулю, а модуль отклика по амплитуде – возведён в квадрат, что обеспечивает подавление компонент быстрее  $T_c$  без групповой задержки. Обозначим результат как  $y_i^{(B)}$ .

Ещё одним инструментом предварительной обработки данных является локальная полиномиальная аппроксимация (метод Савицкого-Голея). Для восстановления локальной кривизны поверх  $y^{(B)}$  используется окно  $L_{SG} = 2h + 1$  (нечётное) и полином степени  $p$ . Для каждого  $i$  находится полином  $q_i(\cdot)$ , минимизирующий выражение 13.

$$\sum_{j=-h}^h (y_{i+j}^{(B)} - q_i(j))^2, p \in \{2,3\}, L_{SG} > p + 2, \quad (13)$$

после чего  $y_i^{(SG)} = q_i(0)$ . Этот шаг малым окном «смягчает» остаточную шероховатость, практически не меняя тренда.

Множество валидных индексов  $M = \{i: x_i \text{ конечен}\}$  представляется объединением максимальных непрерывных интервалов  $[s_q, e_q)$ . Шаги 1–3 применяются независимо в каждом  $[s_q, e_q)$  при  $e_q - s_q \geq L_{min}$  (минимальная длина сегмента), а значения вне  $M$  сохраняются пропущенными. Это исключает проникновение информации через длительные разрывы и граничные артефакты.

Для формальной визуальной валидации формируются пять непересекающихся окон фиксированной длины ( $m=300$ ) последовательных

точек (примерно 5 часов при одноминутной дискретизации), равномерно распределённых по множеству допустимых стартов  $S = \{s: s, \dots, s+m-1 \in M\}$ . Для каждого окна строятся совмещённые графики «до обработки» ( $x_i$ ) и «после обработки» ( $y_i^{(SG)}$ ), что позволяет выявлять избыточное сглаживание, остаточные выбросы и краевые искажения.

Итоговый ряд определяется как выражение (14).

$$x_i^{(clean)} = y_i^{(SG)} \quad (13)$$

Параметры  $\alpha$ ,  $n$ ,  $T_c$ ,  $L$ ,  $L_{SG}$ ,  $p$ ,  $L_{min}$  задаются в физических единицах времени через  $\Delta t_{med}$ , что обеспечивает переносимость методики между наборами данных. В практических настройках использовались  $\alpha \approx 3$ ,  $n=4$ ,  $T_c$  порядка десятков минут, малое окно Савицкого-Голея с ( $p=3$ ). Для последующего моделирования формируется изолированный столбец очищенной компоненты ( $B_z$ ) в исходной временной разметке.

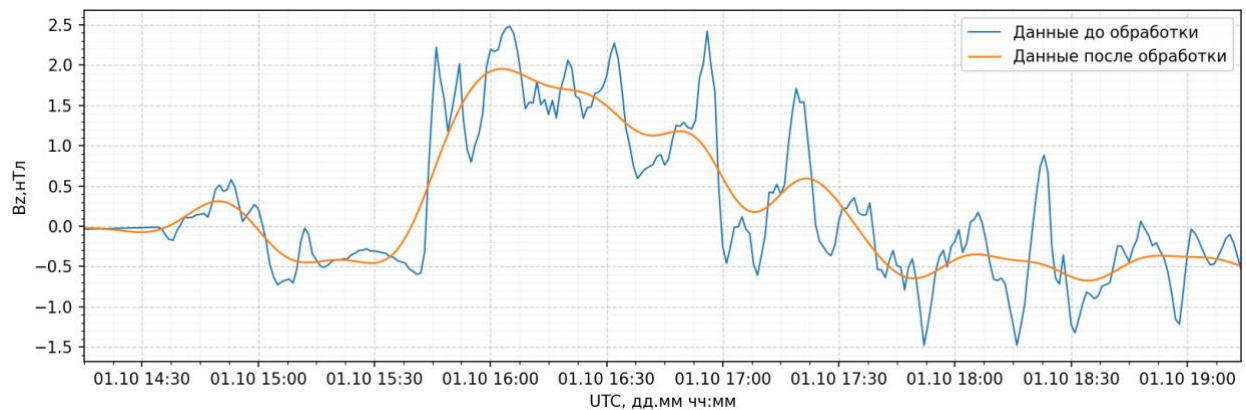


Рис. 1. Пример предобработки данных

### Вычислительный эксперимент.

Эксперимент направлен на сравнение точности восстановления пропусков в одноминутном геофизическом ряду  $B_z(t)$  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 зерне случайных чисел, конфигурация и артефакты эксперимента сериализуются в отдельные файлы.

Сравниваются три метода восстановления: линейная временная интерполяция; двусторонний авторегрессионный метод ARIMA(5,0,0) с симметричным контекстом (до 180 точек слева и справа), последующим плавным косинусным смешиванием прямого и обратного прогнозов и сглаженной стыковкой класса  $C^1$  в смысле кубического Эрмитова сплайна; двунаправленная рекуррентная нейросеть LSTM, принимающая на вход двухканальное представление  $[\tilde{x}(t) \cdot 1_{obs}(t), 1_{obs}(t)]$  с последующей аффинной стыковкой на границах пропуска. Косинусные веса задаются выражением 14.

$$w_{fwd}(u) = \frac{1}{2} + \frac{1}{2} \cos(\pi u), w_{bwd}(u) = \frac{1}{2} - \frac{1}{2} \cos(\pi u), u = \frac{j}{L-1}, \quad (14)$$

что обеспечивает плавный переход от левого прогноза к правому. Для обеспечения класса  $C^1$  применяется эрмитова коррекция базовой последовательности  $\{b_j\}_{j=0}^{L-1}$  с использованием стандартных базисных функций  $H_{00}, H_{10}, H_{01}, H_{11}$  и оценок краевых производных из ближайших наблюдений. Для LSTM итоговое заполнение в пропуске  $\{y_j^*\}$  подгоняется аффинно: при наличии обеих границ заданных выражением (15).

$$a = \frac{y_R - y_L}{y_{L-1}^* - y_0^*}, b = y_L - ay_0^*, \hat{y}_j = ay_j^* + b, \quad (15)$$

Где  $y_L$  и  $y_R$  – наблюдаемые значения на левой и правой границе пропуска.

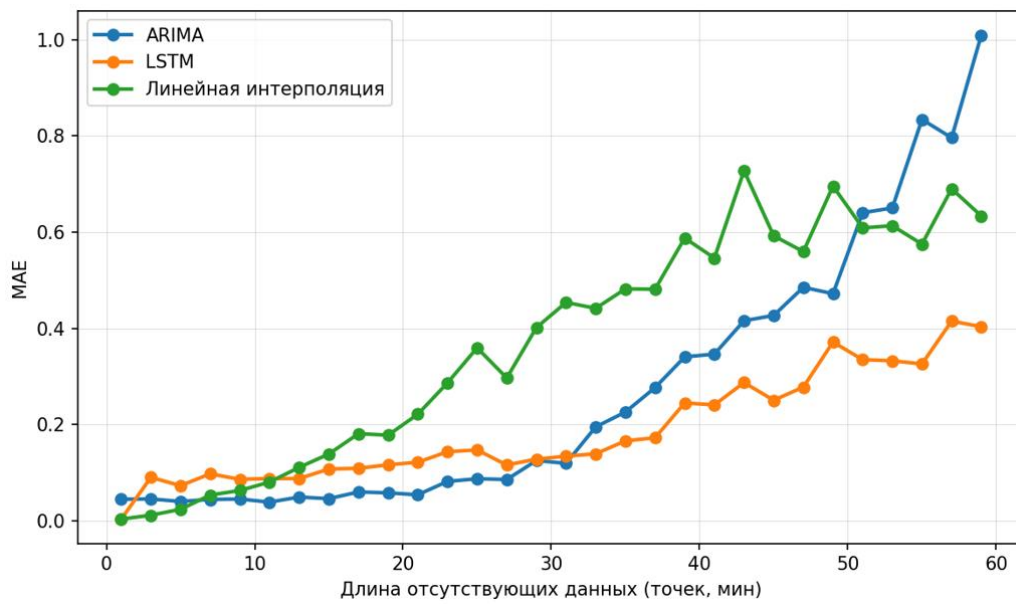


Рис. 2. Зависимость MAE восстановления от длины пропуска наблюдений

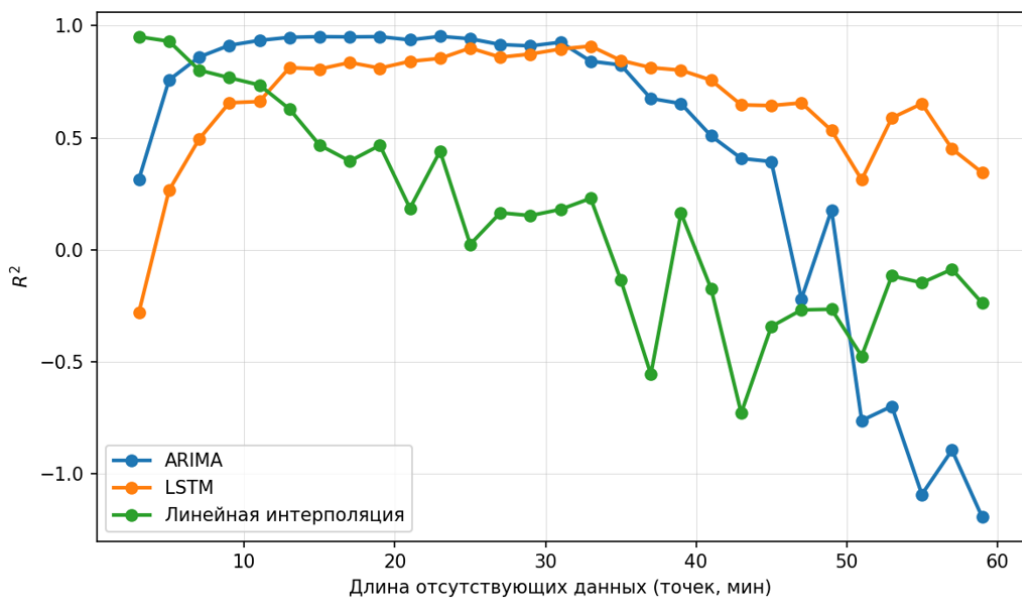


Рис.3. Зависимость  $R^2$  восстановления от длины пропуска наблюдений

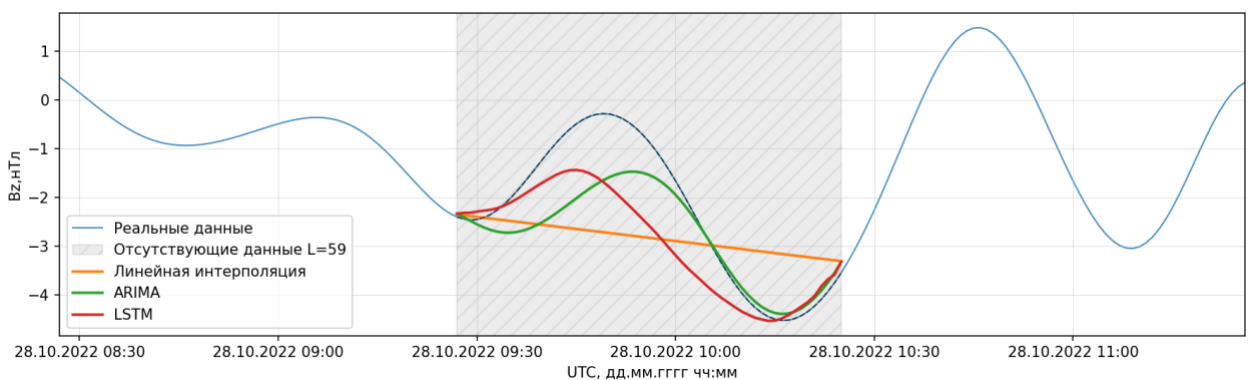


Рис.4. Пример восстановления отсутствующих данных

Анализ кривых на рис. 2 и рис. 3 показывает устойчивое превосходство нейросетевой модели LSTM при средних и больших длинах пропусков и резкое ухудшение качества ARIMA на длинных интервалах, отсутствующих данных, линейная интерполяция конкурентоспособна лишь на очень коротких пропусках.

### **Методологический подход к выбору метода восстановления**

Адаптивный выбор метода восстановления временных рядов формулируется как задача мета-моделирования, в которой описательным признакам ряда и маске пропусков требуется определить алгоритм или взвешенную комбинацию, минимизирующие ожидаемую погрешность восстановления. Пусть дан многомерный спутниковый временной ряд  $X \in \mathbb{R}^{T \times d}$  с маской наблюдаемости  $M \in \{0,1\}^{T \times d}$  (1 – наблюдаемое, 0 – пропуск). Требуется восстановить отсутствующие значения  $\hat{X}$  и одновременно адаптивно выбрать методы восстановления из фиксированного пула  $A = \{a_1, \dots, a_K\}$ . Мотивация адаптивности опирается на эмпирику бенчмарков: качество восстановления данных существенно зависит от доли пропусков и их паттерна, а также от свойств ряда (тренд, сезонность, нелинейность, шум)[1].

Цель – минимизировать риск восстановления на пропусках в соответствии с выражением (9).

$$\arg \min_{a \in A} L(X, \hat{X}^{(a)}; M) \quad (9)$$

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.

Мета-признаки  $\phi(X, M)$  описывают «характер» данных и пропусков. Их использование лежит в основе успешных систем мета-обучения и рекомендации моделей для рядов [3].

### **Заключение**

В работе рассмотрены три репрезентативных подхода к восстановлению пропусков в одноминутных геофизических временных рядах (Bz, нТл): линейная интерполяция, ARIMA и LSTM. Эксперименты на трёхчасовых окнах с искусственно созданными разрывами длиной 1–59 мин и последующей оценкой MAE и  $R^2$  показали, что:

1. Для коротких пропусков (до ~10 мин) линейная интерполяция и ARIMA обеспечивают близкие к LSTM ошибки; применение сложных моделей на этом горизонте необязательно.
2. В диапазоне 10–35 мин LSTM демонстрирует наименьший MAE и стабильно высокие значения  $R^2$  ( $\approx 0.8-0.9$ ), превосходя интерполяцию, которая систематически «выпрямляет» траектории внутри окна.
3. При длинных пропусках ( $\geq 40$  мин) преимущество LSTM становится выраженным:  $R^2$  остаётся положительным и сохраняет прогностическую ценность, тогда как у ARIMA наблюдается резкая деградация вплоть до отрицательных  $R^2$ , а у линейной интерполяции существенно растёт MAE.

4. Аффинная стыковка на границах разрыва уменьшает разрыв по уровню для всех методов, но не устраняет концептуальные ограничения линейных моделей на длинном горизонте.
5. С учётом полученных кривых обоснована гибридная политика выбора метода: линейная интерполяция – для очень малых «окон», LSTM – для средних и больших, ARIMA – как интерпретируемый ориентир на умеренных длинах при близкой к стационарной динамике.
6. Итоговое методологическое следствие для задач спутникового мониторинга: модели с богатым контекстом и нелинейным представлением (LSTM и их современные аналоги) должны рассматриваться как метод по умолчанию при доле длинных разрывов, тогда как простые методы – как быстрый базовый уровень и элемент ансамбля.

Ограничения: результаты агрегированы по выбранной выборке и настройкам; для большей обобщаемости требуется стратификация по уровням геомагнитной активности и расширение пула методов с единым протоколом честной валидации. Перспективы – внедрение мета-модели адаптивного выбора по признакам ряда и маске пропусков, а также ансамблевое сочетание локальных и глубинных моделей.

### **Источники литературы**

1. 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.
2. Miller L., Pelletier C., Webb G. I. Deep Learning for Satellite Image Time Series Analysis: A Review. arXiv:2404.03936, 2024.
3. Niako N. E., Kuttikat P., Patmanathan R., Neil M., Zhang X., Zahid H. Effects of missing data imputation methods on univariate blood pressure time series

analysis and forecasting with ARIMA and LSTM // BMC Medical Research Methodology. 2024. DOI: 10.1186/s12874-024-02448-3.

4. 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.

5. Shammi N. J., Ashraf T., Ifty 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.

6. Kazijevs M., Samad M. D. Deep Imputation of Missing Values in Time Series Health Data: A Review with Benchmarking. arXiv:2302.10902, 2023.

7. 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.

8. 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.

9. 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.

10. 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.

11. Zhou H., Yu J., Zuo L., Xu Y., Yi J., Qi Y., et al. TSI-Bench: Benchmarking Time Series Imputation. arXiv:2406.12747, 2024.