

УДК 550.384:004.9

***ИНФОРМАЦИОННО-АНАЛИТИЧЕСКОЕ ОБЕСПЕЧЕНИЕ
ИНТЕРПОЛЯЦИИ ВРЕМЕННЫХ РЯДОВ ГЕОМАГНИТНЫХ ДАННЫХ***

Абдуллин А.А.

студент,

ФГБОУ ВО «Уфимский университет науки и технологий»,

Уфа, Россия

Воробьева Г.Р.

д.т.н., доцент,

ФГБОУ ВО «Уфимский университет науки и технологий»,

Уфа, Россия

Аннотация

В статье рассматривается задача восстановления пропущенных значений в геомагнитных временных рядах наземных магнитометрических станций. Предложено информационно-аналитическое обеспечение, объединяющее получение данных SuperMAG, отбор высокоширотных станций, предварительное сглаживание сигнала, формирование контрольных пропусков и сравнение методов восстановления. Для оценки результата используются метрики R^2 , RMSE и MAE, а также визуальное сопоставление исходного и восстановленного фрагментов ряда. Показано, что выбор метода зависит от длины пропуска и локальной динамики сигнала; поэтому программная реализация должна поддерживать несколько подходов и единые условия вычислительного эксперимента.

Ключевые слова: геомагнитные данные, временной ряд, интерполяция, восстановление пропусков, SuperMAG, сглаживание, SARIMAX, машинное обучение.

***INFORMATION AND ANALYTICAL SUPPORT FOR INTERPOLATION
OF GEOMAGNETIC DATA TIME SERIES***

Abdullin A.A.

student,

Ufa University of Science and Technology,

Ufa, Russia

Vorobyeva G.R.

Doctor of Technical Sciences, Associate Professor,

Ufa University of Science and Technology,

Ufa, Russia

Abstract

The article considers the problem of recovering missing values in geomagnetic time series obtained from ground magnetometer stations. The proposed information and analytical support combines SuperMAG data acquisition, selection of high-latitude stations, preliminary signal smoothing, controlled gap generation and comparative evaluation of recovery methods. The quality of reconstruction is assessed by R^2 , RMSE and MAE metrics, complemented by visual comparison of the original and recovered fragments. The results show that the preferred method depends on the gap length and local signal dynamics; therefore, the software implementation should support several approaches under unified experimental conditions.

Keywords: geomagnetic data, time series, interpolation, missing values, SuperMAG, smoothing, SARIMAX, machine learning.

Введение

Геомагнитные временные ряды отражают изменение компонент магнитного поля Земли во времени и используются при анализе процессов в магнитосфере, ионосфере и околоземной плазменной среде. Наземные магнитометрические станции формируют последовательности наблюдений с регулярной дискретностью, благодаря чему становится возможным исследование возмущений, связанных с космической погодой, авроральной активностью и быстрыми вариациями магнитного поля [3; 4; 6]. Для последующей обработки такие данные должны сохранять непрерывную временную сетку, однако практические наборы наблюдений часто содержат пропуски, выбросы и участки с повышенным уровнем шума.

Появление отсутствующих значений связано с техническим обслуживанием станций, сбоями измерительного оборудования, нарушениями передачи данных и ошибками выгрузки файлов. Даже короткий разрыв в ряду осложняет построение графиков, расчет статистических характеристик и применение моделей, рассчитанных на равномерную последовательность наблюдений. В исследованиях временных рядов и данных с пропусками такая ситуация рассматривается как отдельная задача предварительной обработки, влияющая на надежность дальнейшего анализа [1; 2; 12].

Наиболее показательными для проверки методов восстановления являются высокоширотные станции. В северных районах геомагнитные вариации выражены сильнее, поскольку на форму сигнала влияют авроральные токовые системы и интенсивное взаимодействие солнечного ветра с магнитосферой. Восстанавливаемый фрагмент в таких условиях должен соответствовать отдельным числовым значениям и общей динамике исходного ряда. Поэтому оценка качества требует совмещения количественных метрик и визуального анализа формы сигнала [3; 4; 5].

Формирование информационно-аналитического обеспечения обработки геомагнитных данных

В качестве источника данных в работе используется SuperMAG. Данная система объединяет наблюдения наземных магнитометров и предоставляет геомагнитные данные в унифицированном представлении, что облегчает совместную обработку рядов разных станций [10]. Применение SuperMAG позволяет перейти от разрозненной работы с отдельными файлами к воспроизводимому вычислительному процессу, где станция, компонент магнитного поля, временной интервал и параметры обработки задаются в едином программном контуре.

Целью выполненного исследования является разработка информационно-аналитического обеспечения интерполяции геомагнитных временных рядов, позволяющего получать исходные данные, формировать контрольные пропуски, восстанавливать отсутствующие значения несколькими методами и оценивать качество результата по общим критериям. Программная реализация получила название GeoGap SuperMAG. Она ориентирована на исследовательский сценарий, при котором пользователь выбирает станцию, компоненту поля, длительность пропуска и перечень методов восстановления.

Организация вычислительного эксперимента по восстановлению пропусков

Общая логика вычислительного эксперимента представлена на рисунке 1. Сначала временной ряд загружается из SuperMAG или локальной библиотеки, затем выполняются проверка временной сетки, выбор компоненты магнитного поля и сглаживание сигнала. После этого из непрерывного участка временно исключается контрольный фрагмент. Полученный ряд с пропуском передается в блок восстановления, а рассчитанные значения сравниваются с исходным фрагментом, сохраненным до удаления.

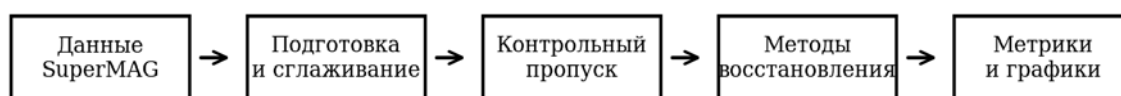


Рисунок 1 - Схема вычислительного эксперимента по восстановлению пропусков

Примечание – авторская разработка.

Формально исходный геомагнитный ряд можно представить в виде последовательности $X = \{x(t_i)\}_{i=1}^n$, где $x(t_i)$ обозначает значение выбранной компоненты магнитного поля в момент времени t_i . Контрольный пропуск задается множеством $G = \{x(t_{\text{к}}), x(t_{\text{к}+1}), \dots, x(t_{\text{к}+\text{к}-1})\}$, где k соответствует числу удаляемых отсчетов. После исключения фрагмента формируется ряд $X' = X/G$. Задача восстановления состоит в построении оценок $\hat{G} = \{\hat{x}(t_{\text{к}}), \hat{x}(t_{\text{к}+1}), \dots, \hat{x}(t_{\text{к}+\text{к}-1})\}$, наиболее близких к исходным значениям контрольного участка.

Применение методов интерполяции и моделирования временных рядов

В программном средстве сравниваются несколько групп методов. Классические интерполяционные подходы представлены линейной интерполяцией, кубическим сплайном, РСНР и Akima. Линейный метод служит базовым вариантом, кубический сплайн строит гладкую кусочно-полиномиальную аппроксимацию, РСНР ориентирован на сохранение формы ряда, а Akima учитывает локальные наклоны и уменьшает влияние резких переходов [7; 9; 15]. Статистический подход представлен моделью SARIMAX, описывающей зависимость текущего значения от предыдущих наблюдений и случайной составляющей [8]. Для сопоставления с методами машинного обучения используется многослойный персептрон MLP, реализованный как регрессионная модель по левому и правому контексту пропуска [13].

Состав методов восстановления, включенных в эксперимент, приведен в таблице 1. Такое разделение удобно для анализа, поскольку простые

интерполяционные методы хорошо интерпретируются и быстро выполняются, а модельные методы позволяют учитывать более сложные зависимости временного ряда. Универсальный метод для всех участков заранее не задается: итоговый выбор выполняется по результатам расчета метрик

Таблица 1 – Методы восстановления пропусков, включенные в сравнительный анализ

Группа методов	Методы	Назначение в эксперименте
Классическая интерполяция	Линейная интерполяция, кубический сплайн, PCHIP, Akima	Быстрое восстановление коротких пропусков и сравнение локальной формы сигнала
Статистическое моделирование	SARIMAX, SARIMAX с внешним рядом соседней станции	Учет временной структуры ряда и дополнительной информации от близкой станции
Машинное обучение	MLP-регрессор	Проверка нелинейного восстановления по левому и правому контексту пропуска

Перед формированием контрольного пропуска в программе применяется сглаживание. Основным вариантом является фильтр Савицкого–Голея, который позволяет уменьшить влияние кратковременных шумовых колебаний при сохранении общей формы сигнала [14]. Дополнительно предусмотрены гауссово сглаживание и скользящее среднее. Сглаженный ряд используется как рабочая основа эксперимента, поскольку восстановление одиночных резких пиков внутри отсутствующего фрагмента часто приводит к завышенной ошибке и затрудняет сравнение методов.

Оценка качества восстановления геомагнитных данных

Для количественной оценки результата используются три показателя. Среднеквадратическая ошибка RMSE вычисляется как корень из среднего квадрата отклонений восстановленных значений от исходных. Средняя абсолютная ошибка MAE отражает средний модуль расхождения. Коэффициент

детерминации R^2 характеризует соответствие восстановленного фрагмента исходной форме сигнала. В рамках выбора лучшего метода R^2 стремится к единице, а RMSE и MAE минимизируются. Совместное использование этих показателей снижает риск односторонней интерпретации качества восстановления.

Расчет метрик выполняется по формулам:

$$RMSE = \sqrt{(1/k) \cdot \sum_{i=1}^k (x_i - \hat{x}_i)^2},$$

$$MAE = 1/k \cdot \sum_{i=1}^k |x_i - \hat{x}_i|,$$

$$R^2 = 1 - \sum_{i=1}^k (x_i - \hat{x}_i)^2 / \sum_{i=1}^k (x_i - \bar{x})^2.$$

Программная реализация информационно-аналитического обеспечения

Программная реализация выполнена на языке Python. Для хранения и подготовки временных рядов используется pandas, численные операции выполняются средствами NumPy, интерполяционные методы реализуются с применением SciPy, а MLP-регрессор — средствами scikit-learn [11; 13; 15]. Статистическая модель SARIMAX применяется как инструмент анализа временной структуры ряда [8]. Графическое представление результатов построено средствами Python. В интерфейсе GeoGap SuperMAG пользователь может выбрать источник данных, станцию, компоненту магнитного поля, способ сглаживания, длины пропусков и набор методов восстановления.

Анализ результатов вычислительного эксперимента

Проверка работоспособности системы проводилась на вычислительном эксперименте с искусственными пропусками длительностью 5, 10, 15, 30, 45 и 60 минут. Для каждого варианта формировался ряд с удаленным контрольным участком, затем применялись все выбранные методы восстановления. В таблице 2 приведен пример итогового сравнения, полученный для рассматриваемого временного окна и выбранной станции. Значения не следует трактовать как универсальное ранжирование методов, поскольку геомагнитный сигнал

существенно меняется от станции к станции и от одного временного фрагмента к другому.

Таблица 2 – Пример результатов оценки качества восстановления пропусков

Длительность пропуска, мин	Лучший метод	R^2	RMSE	MAE
5	Кубический сплайн	0,281	6,975	6,109
10	Кубический сплайн	0,800	11,735	10,458
15	Кубический сплайн	-0,257	25,537	21,163
30	Akima	0,098	31,403	24,531
45	PCNIP	-0,998	42,545	35,215
60	Линейная интерполяция	-0,136	71,552	53,377

Анализ результатов показывает, что при коротких пропусках лучшими могут становиться гладкие интерполяционные методы, поскольку соседние опорные значения достаточно хорошо описывают локальный ход ряда. При увеличении длины пропуска ошибка возрастает, а коэффициент R^2 в отдельных случаях принимает отрицательные значения. Такая ситуация указывает на низкое соответствие восстановленного участка исходному контрольному фрагменту и связана с высокой изменчивостью геомагнитного сигнала. Для практического анализа отрицательный R^2 не является признаком сбоя программы; он отражает объективную сложность восстановления выбранного участка.

Практическое применение и развитие программного средства

Разработанное информационно-аналитическое обеспечение позволяет выполнять вычислительный эксперимент в воспроизводимой форме. В одном программном средстве объединены загрузка данных, отбор высокоширотных станций, сглаживание, формирование пропусков, восстановление, расчет метрик

и построение графиков. Благодаря этому пользователь получает таблицу ошибок и визуальное сопоставление исходного, удаленного и восстановленного фрагментов.

Практическая значимость работы состоит в возможности быстро оценивать пригодность различных методов восстановления для конкретной станции, компоненты магнитного поля и длительности пропуска. GeoGar SuperMAG может применяться в учебно-исследовательских задачах, связанных с обработкой геомагнитных данных, а также служить основой для дальнейшего расширения набора методов. Перспективным направлением развития является автоматический подбор параметров SARIMAX, анализ влияния соседних станций и включение более устойчивых моделей машинного обучения.

Основные результаты исследования

В результате проведенного исследования создано программное средство для информационно-аналитического обеспечения интерполяции временных рядов геомагнитных данных. Применение искусственных пропусков дает возможность объективно сравнивать восстановленные значения с исходным контрольным участком. Сочетание нескольких методов, метрик R^2 , RMSE и MAE, а также графического представления результата обеспечивает более надежный выбор подхода к восстановлению пропусков в высокоширотных геомагнитных наблюдениях.

Библиографический список:

1. Воробьев А. В. Индуктивный метод восстановления временных рядов геомагнитных данных / А. В. Воробьев, Г. Р. Воробьева // Труды СПИИРАН. – 2018. – № 2(57). – С. 104–133. – DOI: 10.15622/sp.57.5.
2. Воробьева Г. Р. Подход к восстановлению геомагнитных данных путем сопоставления суточных фрагментов временного ряда с равной геомагнитной активностью / Г. Р. Воробьева, А. В. Воробьев // Компьютерная оптика. – 2019. – Т. 43, № 6. – С. 1053–1063. – DOI: 10.18287/2412-6179-2019-43-6-1053-1063.

3. Гвишиани А. Д. Геомагнетизм: от ядра Земли до Солнца / А. Д. Гвишиани, Р. Ю. Лукьянова, А. А. Соловьев. – М.: Российская академия наук, 2019. – 186 с. – ISBN 978-5-907036-43-7.
4. Гвишиани А. Д. Геоинформатика и наблюдения магнитного поля Земли: российский сегмент / А. Д. Гвишиани, Р. Ю. Лукьянова // Физика Земли. – 2015. – № 2. – С. 3–20.
5. Мандрикова О. В. Автоматический способ оценки состояния геомагнитного поля / О. В. Мандрикова, Е. А. Жижикина // Компьютерная оптика. – 2015. – Т. 39, № 3. – С. 420–428.
6. Мировой Центр Данных по солнечно-земной физике: геомагнитные вариации [Электронный ресурс]. – Режим доступа — URL: https://www.wdcb.ru/stp/geomag/geomagnetic_variations.ru.html (дата обращения 22.06.2026).
7. Akima H. A new method of interpolation and smooth curve fitting based on local procedures / H. Akima // Journal of the ACM. – 1970. – Vol. 17, № 4. – P. 589–602. – DOI: 10.1145/321607.321609.
8. Box G. E. P. Time Series Analysis: Forecasting and Control / G. E. P. Box, G. M. Jenkins, G. C. Reinsel, G. M. Ljung. – 5th ed. – Hoboken: Wiley, 2015. – 712 p.
9. Fritsch F. N. Monotone Piecewise Cubic Interpolation / F. N. Fritsch, R. E. Carlson // SIAM Journal on Numerical Analysis. – 1980. – Vol. 17, № 2. – P. 238–246. – DOI: 10.1137/0717021.
10. Gjerloev J. W. The SuperMAG data processing technique / J. W. Gjerloev // Journal of Geophysical Research: Space Physics. – 2012. – Vol. 117. – Art. A09213. – DOI: 10.1029/2012JA017683.
11. Harris C. R. Array programming with NumPy / C. R. Harris [et al.] // Nature. – 2020. – Vol. 585. – P. 357–362. – DOI: 10.1038/s41586-020-2649-2.

12. Little R. J. A. Statistical Analysis with Missing Data / R. J. A. Little, D. B. Rubin. – 3rd ed. – Hoboken: Wiley, 2019. – 464 p.
13. Pedregosa F. Scikit-learn: Machine Learning in Python / F. Pedregosa [et al.] // Journal of Machine Learning Research. – 2011. – Vol. 12. – P. 2825–2830.
14. Savitzky A. Smoothing and Differentiation of Data by Simplified Least Squares Procedures / A. Savitzky, M. J. E. Golay // Analytical Chemistry. – 1964. – Vol. 36, № 8. – P. 1627–1639. – DOI: 10.1021/ac60214a047.
15. Virtanen P. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python / P. Virtanen [et al.] // Nature Methods. – 2020. – Vol. 17. – P. 261–272. – DOI: 10.1038/s41592-019-0686-2.