

Исследование точности восстановления сезонных значений ПЭС на основе машинного обучения и динамических предикторов

Рябова Н.В., Конкин Н.А., Вершинин М.В.

*Поволжский государственный технологический университет
г. Йошкар-Ола, пл. Ленина, д. 3, 424000, Российская Федерация
konkinna@volgatech.net*

Работа посвящена исследованию методов восстановления пропущенных данных полного электронного содержания (ПЭС), используемых для коррекции ошибок в спутниковых системах связи и навигации. Пропуски в данных возникают из-за технических сбоев, атмосферных явлений, человеческого фактора и ограничений измерений, что снижает точность работы систем. Традиционные подходы (интерполяция, физические модели) имеют недостатки: слабую адаптацию к нелинейным зависимостям и экстремальным условиям. В работе предложено применение ансамблевых методов машинного обучения (МО) в сочетании с динамическими предикторами для повышения точности восстановления сезонных значений ПЭС. Цель исследования — оценка точности и применимости разработанного алгоритма в зависимости от времени года и количества восстанавливаемых отсчётов. Результаты сравнительного анализа получены с использованием метрик: коэффициент детерминации (R^2), средняя абсолютная ошибка (MAE), средняя абсолютная процентная ошибка (MAPE) и среднеквадратическая ошибка (RMSE). R^2 при восстановлении 48 отсчётов (практически значимый диапазон) продемонстрировал разницу менее 0,6% между реальными и восстановленными данными для всех сезонов. MAE не превысила 4,3%. Доверительные интервалы (CI) подтвердили устойчивость метода в трёх из четырёх сезонов (зима, весна, осень), однако летние значения MS) выходили за пределы CI из-за повышенной геомагнитной активности. Исследование подтвердило, что гибридные модели МО, объединяющие физические закономерности и данные, эффективны для восстановления ПЭС в реальном времени. Метод демонстрирует устойчивость к сезонным изменениям и может быть интегрирован в системы мониторинга космической погоды и навигации. Перспективы работы связаны с добавлением признаков, учитывающих экстремальные геомагнитные условия, и расширением набора данных для повышения точности в летний период.

Ключевые слова: полное электронное содержание, машинное обучение, ансамблевые методы, динамические предикторы, сезонный анализ, восстановление данных.

Study of the accuracy of seasonal PES values recovery based on machine learning and dynamic predictors

Ryabova N.V., Konkin N.A., Vershinin M.V.

Volga State University of Technology

The study focuses on methods for restoring missing Total Electron Content (TEC) data used to correct errors in satellite communication and navigation systems. Data gaps arise due to technical failures, atmospheric phenomena, human factors, and measurement limitations, which reduce system accuracy. Traditional approaches (interpolation, physical models) have drawbacks: poor adaptation to nonlinear dependencies and extreme conditions. The paper proposes the use of ensemble machine learning (ML) methods combined with dynamic predictors to improve the accuracy of seasonal TEC value restoration. The research aims to evaluate the accuracy and applicability of the developed algorithm depending on the season and the number of restored samples. Comparative analysis results were obtained using metrics: determination coefficient (R^2), mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE). For 48 samples (practically significant range), R^2

showed less than 0.6% difference between real and restored data across all seasons. MAE did not exceed 4.3%. Confidence intervals (CI) confirmed the method's stability in three out of four seasons (winter, spring, autumn); however, summer MSE values exceeded CI limits due to increased geomagnetic activity. The study confirmed that hybrid ML models combining physical patterns and data are effective for real-time TEC restoration. The method demonstrates resilience to seasonal variations and can be integrated into space weather monitoring and navigation systems. Future work involves adding features accounting for extreme geomagnetic conditions and expanding the dataset to improve accuracy during summer.

Keywords: total electron content, machine learning, ensemble methods, dynamic predictors, seasonal analysis, data restoration.

Введение

Проблемой использования данных о полном электронном содержании (ПЭС) для задач коррекции ошибок в работе систем спутниковой связи (ССС) и навигации ССН, являются пропуски в данных. Пропуски могут возникать из-за технических сбоев оборудования (отказы приёмников, проблемы синхронизации, ошибки передачи), атмосферных и ионосферных явлений (геомагнитные бури, солнечные вспышки, полярные сияния), влияния окружающей среды (помехи от наземных объектов, многолучевое распространение сигналов), человеческого фактора (ошибки в настройке и обработке данных), а также ограничений измерений (низкая плотность станций, особенно над океанами, и снижение точности в условиях сильных возмущений). Для восстановления данных ПЭС применяют традиционные методы (интерполяция, физические модели, статистика), но они имеют ограничения: слабо учитывают нелинейные зависимости, требуют точных входных параметров и плохо адаптируются к экстремальным условиям. Машинное обучение (МО) становится более актуальным благодаря способности анализировать сложные паттерны в данных. Нейросети (LSTM, CNN) прогнозируют ПЭС, учитывая временные и пространственные связи, даже при пропусках. Ансамблевые методы (Random Forest) и гибридные модели объединяют физические законы с данными, повышая точность. МО эффективно работает с большими объёмами информации, автоматически адаптируется к изменениям (бури, сезонность) и интегрирует мультиисточниковые данные (спутники, сенсоры). Это позволяет восстанавливать ПЭС в реальном времени с высокой точностью, что критично для навигационных систем и мониторинга космической погоды, делая МО незаменимым инструментом в современных условиях [2, 6].

Цель — выполнить исследование точности и применимости восстановления сезонных значений ПЭС на основе ансамблевого метода машинного обучения и динамических предикторов в зависимости от времени года и количества восстанавливаемых отсчётов.

Для достижения поставленной цели необходимо решить следующие задачи:

- 1) Разработать алгоритма восстановления значений ПЭС.
- 2) Выполнить предварительную обработку данных ПЭС, полученных за 1 года на сети референчных станций глобальных навигационных спутниковых систем (ГНСС).
- 3) Выполнить анализ точности восстановления в зависимости от сезона года.
- 4) Выполнить сравнительный анализ результатов прогнозирования ПЭС на основе восстановленных и оригинальных данных в зависимости от сезона года для одной из ключевых станций.

2. Разработка алгоритма восстановления ПЭС на основе существующих данных и ранее восстановленных

Алгоритм восстановления реализуется в несколько этапов и включает в себя следующее:

1. Поиск «целого» набора данных, не содержащего пропусков ни в части с ключевыми станциями, ни в части с динамическими предикторами;

Данное условие необходимо, поскольку для первичного обучения модели необходимо сформировать окно обучения, не содержащего пропусков.

2. Формирование набора данных для анализа и потенциального восстановления;

Поиск «целого» набора данных априори предполагает наличие пропуска или пропусков после него, поэтому данные, следующие после, автоматически попадают в категорию данных для анализа и потенциального восстановления. В усреднённом представлении количество значений, попадающих в данный диапазон равно 95% от общего числа отсчётов.

3. Исследование отсчётов из набора для анализа и потенциального восстановления;
На данном этапе происходит поочерёдная, с единичным шагом (30 минут), проверка отсчётов на пропуски и в случае их нахождения в одной из строк – формирование её шаблона пропусков: списка с пропущенными столбцами.
4. Формирование тренировочной и восстанавливающей выборки;
Зная шаблон строки, можно правильно сформировать матрицу признаков и целевой вектор. Целевой вектор для обучения и восстановления формируется в соответствии с пропущенными столбцами или столбцом.
5. Тренировка модели машинного обучения;
6. Прогнозирование (восстановление) пропущенных отсчетов и их вставка в исходный датасет.

Упрощённая блок-схема алгоритма представлена на рис. 1. Таким образом разработанный алгоритм позволяет определять различные шаблоны пропусков в датасете и восстанавливать их на основе максимально возможной или заданной тренировочной части, при этом учитывая возможность использования значений ПЭС соседних станций, существующих за период прогноза (восстановления) [1, 3, 4].



Рис 1. Алгоритм восстановления ПЭС на основе существующих данных и ранее восстановленных

2. Описание датасета (целевые периоды для восстановления)

Исследуемый в работе датасет содержит в себе данные полного электронного содержания за 2021 год. Сбор и сохранение полученных данных осуществляется на референчных станциях глобальной навигационной спутниковой системы компании ООО «Смартнет Геосистемс» (SmartNet), находящихся в Самарской области. Станции находятся в таких населённых пунктах, как: Самара, Кошки, Тольятти, Чапаевск, Кинель, Отрадный, Красноармейск, Нефтегорск и Елховка. Данный регион выбран из-за высокой плотности референчных станций. Частота сбора данных на станциях равна 30 секундам, однако при дальнейшей конвертации RINEX файлов в CSV формат представления часть значений усредняется, формируя частоту отсчётов, равную 30 минутам, поскольку основная изменчивость ПЭС варьируется в пределах от 15 до 30 минут.

Количество значений в исследуемом датасете без первичной обработки равно 17358 отсчётов. После восстановления всех временных меток значение равнялось 17520 отсчётам. Количество столбцов равно 9, что эквивалентно количеству исследуемых станций.

Для увеличения количества признаков, используемых для восстановления значений, использовались динамические предикторы, такие как: сдвиговые, дифференцированные сдвиговые и процентные дифференцированные сдвиговые. Данные предикторы формировались для каждой станции, поэтому общее количество столбцов, после их создания, равнялось 1315, поскольку каждый из них сдвигался до 48 раз, т. е. в каждом следующем отсчёте содержалась информация за весь прошлый период, равный по времени одному дню с учётом частоты отсчётов в 30 минут. Во избежание предупреждающих данных дифференцирование производилось между двумя прошлыми значениями и помещалось в следующий отсчёт, а не между текущим и прошлым отсчётом.

Во избежание ошибок при восстановлении было проведение фильтрации исходного датасета на аномальные значения и случайные всплески, учёт которых связан с геомагнитной и солнечной активностью [5, 7].

Обобщённое представление исследуемого датасета в общем виде представлено на рис. 2.

	Целевые станции			Предикторы		
1	1	2	3	*	*	*
2	4	5	6	1	2	3
n	7	8	9	4	5	6

Рис 2. Абстрактное представление исследуемого датасета и добавление к нему сдвиговых предикторов с единичным шагом

3. Апробация алгоритма: Сравнительный анализ результатов на основе реальных и восстановленных значений ПЭС (по метрикам)

Для апробации алгоритма было выполнено прогнозирование (восстановление) значений ПЭС с различным количеством восстанавливаемых отсчётов (6, 12, 24 и 48) во все сезоны на основе реальных и восстановленных значений для одной ключевой станции, находящейся в городе Самара. Полученные результаты можно наблюдать на рис. 3-4.

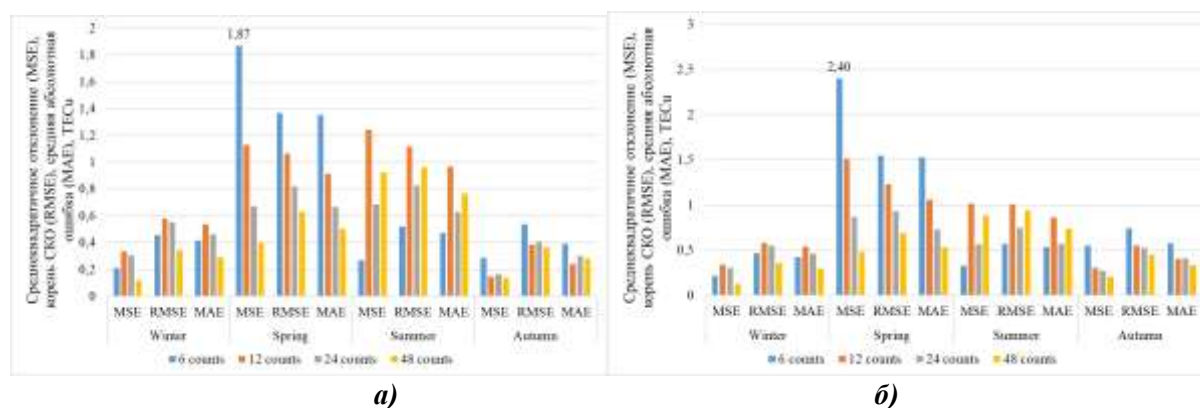


Рис 3. Абсолютные метрики в зависимости от сезона и количества восстанавливаемых отсчётов: а) на основе реальных данных; б) на основе восстановленных данных

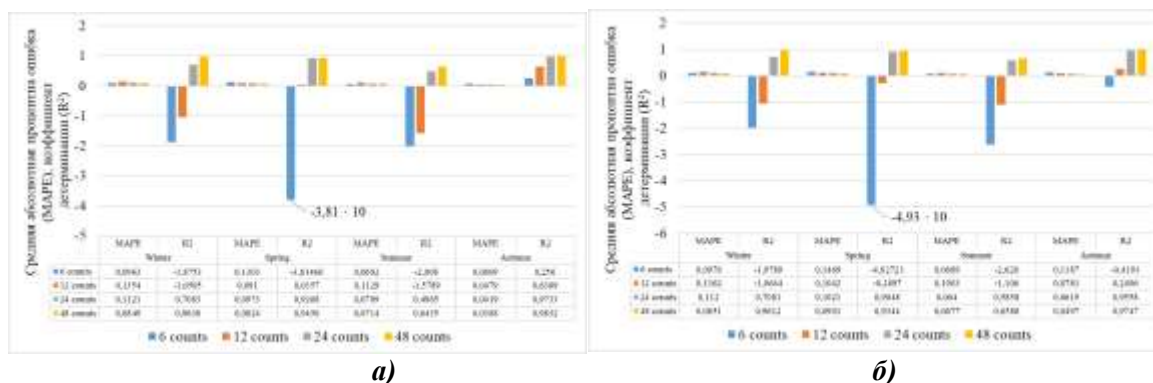


Рис 4. Относительные метрики в зависимости от сезона и количества восстанавливаемых отсчётов: а) на основе реальных данных; б) на основе восстановленных данных

Сравнительный анализ относительной метрики - коэффициента детерминации (R^2) - для практически значимого диапазона, равного 48 отсчётам во всех сезонах, показал, что разница при восстановлении с помощью реальных и восстановленных значений составляет менее 0,6% (для зимы - 0,068%, для весны - 0,298%, для лета - 0,587%, для осени - 0,217%).

Аналогичный анализ был проведён и для абсолютной метрики - средней абсолютной ошибки (MAE). Разница составила менее 4,3% (для зимы - 0,671%, для весны - 1,391%, для лета - 1,013%, для осени - 4,256%).

Дополнительно был вычислен доверительный интервал (CI) для практически значимого диапазона для всех сезонов. Анализ показал, что среднее квадратичное отклонение (MSE) укладывается в его пределы в трёх из четырёх сезонов, таких как: зима, весна и осень. Летом значение CKO (MSE) выходило за пределы доверительного интервала, однако средняя абсолютная процентная ошибка (MAPE) находится в пределах нормы относительно других рассматриваемых сезонов. Результаты анализа можно наблюдать в таблице 1.

Таблица 1 – Результат анализа с использованием доверительного интервала

Сезон	Зима		Весна		Лето		Осень	
	MSE	CI	MSE	CI	MSE	CI	MSE	CI
На основе существующих	0,118	0,516	0,397	0,772	0,924	0,460	0,135	0,810
На основе восстановленных	0,126		0,479		0,884		0,203	

Заключение

В данной работе было выполнено исследование точности восстановления сезонных значений ПЭС на основе ансамблевого метода машинного обучения и динамических предикторов. Для его реализации был разработан алгоритм восстановления значений полного электронного содержания, выполнена предварительная обработка исходных данных, обучена модель машинного обучения в соответствии с поставленными задачами и условиями, а также проведён сравнительный анализ абсолютных и относительных метрик в зависимости от количества восстанавливаемых отсчётов и выбранного сезона.

Сравнительный анализ показал, что при рассмотрении относительной метрики - коэффициента детерминации (R^2) - разница между её значениями, полученными при восстановлении на основе реальных и восстановленных данных, составила менее 0,6%, а при рассмотрении абсолютной метрики - средней абсолютной ошибки (MAE) - разница составила менее 4,3%.

Для оценки применимости данного подхода к прогнозированию (восстановлению) значений были рассчитаны доверительные интервалы (CI) для диапазонов из 48 отсчётов, находящихся в каждом сезоне. Результаты показали, что среднее квадратичные отклонения (MSE) укладываются в границы доверительных интервалов в трёх из четырёх сезонов: зимой, весной и осенью. Однако в летний период значения выходили за пределы доверительного интервала. Это

обусловлено повышенной изменчивостью показателей в летние месяцы, связанной с максимальной геомагнитной и солнечной активностью по сравнению с другими сезонами. Дополнительные признаки для учёта данных факторов не применялись. Тем не менее, полученные значения по метрике *MAPE* находятся в пределах 10%, что свидетельствует о приемлемом уровне точности прогнозирования (восстановления).

Работа выполнена при поддержке гранта Российского научного фонда № 23-19-00145.

Литература

1. Конкин, Н. А. Восстановление значений ПЭС с использованием методов машинного обучения / Н. А. Конкин, М. В. Вершинин // VII научный форум телекоммуникации: Теория и технологии ТТТ-2024 : Материалы XXI Международной научно-технической конференции, Самара, 06–08 ноября 2024 года. – Самара: Поволжский государственный университет телекоммуникаций и информатики, 2024. – С. 73-74. – EDN BFIEMH.
2. Исследование пошоритного влияния геомагнитных условий на вариации ПЭС над территорией Приволжского федерального округа / Н. В. Рябова, А. В. Зуев, А. А. Чернов, Е. В. Катков // СВЧ-техника и телекоммуникационные технологии. – 2020. – № 1-1. – С. 546-547. – EDN SRDWVT. 2.
3. Рябова, Н. В. Восстановление временного хода ПЭС с использованием методов машинного обучения / Н. В. Рябова, Н. А. Конкин, М. В. Вершинин // Распространение радиоволн : сборник докладов I Всероссийской молодежной научной школы-конференции, посвященной памяти Д. С. Лукина, Йошкар-Ола, 11–14 ноября 2024 года. – Йошкар-Ола: Поволжский государственный технологический университет, 2024. – С. 113-117. – EDN PTVDZQ.
4. Замогильный Д. Прогнозирование полного электронного содержания ионосферы на основе алгоритмов машинного обучения // Научные технологии в космических исследованиях Земли. 2022. №4.
5. Свидетельство о государственной регистрации программы для ЭВМ № 2021668835 Российская Федерация. Создание модели ионосферного полного электронного содержания на основе полносвязной искусственной нейронной сети : № 2021668314 : заявл. 15.11.2021 : опубл. 22.11.2021 / А. Е. Быков, Ю. В. Ясюкевич ; заявитель Федеральное государственное бюджетное учреждение науки Ордена трудового красного знамени Институт солнечно-земной физики Сибирского отделения Российской академии наук. – EDN GOGFHI.
6. Специализированная нейронная сеть для прогнозирования параметров широкополосных трансйонсферных радиоканалов в составе интеллектуального сенсора / Д. В. Иванов, А. А. Кислицын, Н. А. Конкин [и др.] // Вестник Поволжского государственного технологического университета. Серия: Радиотехнические и инфокоммуникационные системы. – 2023. – № 4(60). – С. 50-63. – DOI 10.25686/2306-2819.2023.4.50. – EDN NLCNFK.R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
7. Surakhi, Ola & Zaidan, Martha Arbayani & Fung, Pak Lun & Hossein Motlagh, Naser & Serhan, Sami & .Alkhanafseh, Mohammed & Ghoniem, Ranya & Hussein, Tareq. (2021). Time-Lag Selection for Time-Series Forecasting Using Neural Network and Heuristic Algorithm. Electronics. 10. 10.3390/electronics10202518.