

Влияние гелио- и геофизических параметров на точность прогнозирования временной динамики полного электронного содержания ионосферы

А.М. Аппалонов¹, Ю.С. Масленникова²

¹ Казанский (Приволжский) Федеральный Университет
420008, г. Казань, ул. Кремлевская, 18
E-mail: artem309_97@mail.ru

² Казанский (Приволжский) Федеральный Университет
420008, г. Казань, ул. Кремлевская, 18
E-mail: jmaslennikova@kpfu.ru

Представлены результаты исследований влияния внешних солнечных параметров на краткосрочное прогнозирование временной динамики полного электронного содержания (ПЭС) ионосферы. Показано, что при учете таких факторов, как индексы солнечной и магнитной активности, а также информации о времени, простой авторегрессионный прогноз можно существенно улучшить, получив более высокую точность.

Ключевые слова: полное электронное содержание, ионосфера, прогнозирование, индекс F10.7, индекс Kp

The influence of helio- and geophysical parameters on the accuracy of forecasting the temporal dynamics of the total electron content of the ionosphere

A.M. Appalotov¹, Y.S. Maslennikova¹

¹ Kazan Federal university.

The results of studies of the influence of external solar parameters on the short-term prediction of the temporal dynamics of the total electron content (TEC) of the ionosphere are presented. It is shown that when taking into account factors such as solar and magnetic activity indices, as well as time information, a simple autoregressive forecast can be significantly improved by obtaining higher accuracy.

Keywords: total electronic content, ionosphere, forecasting, F10.7 index, Kp index

Введение

Анализ динамики ионосферы остается актуальной задачей радиофизики и радиотехники на протяжении последних десятилетий. Состояние ионосферы в значительной степени влияет на прохождение радиосигналов широкого диапазона частот и, как следствие, на функционирование многих радиотехнических устройств таких как, например, глобальные спутниковые системы радионавигации [1]. Полное электронное содержание (ПЭС) частиц в единичном атмосферном столбе получается путем сопоставления задержки наклонного пути распространения сигнала на двух частотах (1,545 и 1,226 ГГц).

Существующие модели прогнозирования полного содержания электронов в ионосфере Земли показывают удовлетворительные результаты в спокойной геомагнитной обстановке. Такие модели часто базируются на исходных временных рядах и являются авторегрессионными. Однако, при магнитных возмущениях наблюдается существенное снижение точности предсказания таких моделей [2]. Это связано со изменением активности Солнца и соответствующими параметрами, которые, данные модели не учитывают.

В связи с этим, целью данной работы является уточнение прогноза в условиях повышенной солнечной активности, используя матричное разложение и различные геомагнитные параметры при помощи методов машинного обучения.

База данных

Для обучения и тестирования моделей нами были выбраны глобальные карты ПЭС, полученные лабораторией JPL (NASA Jet Propulsion Laboratory), которая является одной из пяти ведущих аналитических лабораторий, обеспечивающих построение глобальных карт абсолютного вертикального значения ПЭС ионосферы путем интерполяции данных, получаемых на мировой сети приемников IGS. Прогностические модели, построенные по таким картам ПЭС, являются наиболее приемлемым с практической точки зрения способом, позволяющим обеспечить потребителей информацией о пространственно-временном распределении ПЭС [3].

Карты ПЭС представлены в географической системе координат с пространственным диапазоном по долготе -180° до 180° (с разрешением 5°) и диапазоном по широте -87.5° до 87.5° (с разрешением 2.5°). Карты доступны на официальном сайте JPL в формате IONEX (Ionosphere Map Exchange) по шкале мирового времени UT с временным разрешением 2 часа. В настоящей работе были использованы карты ПЭС за период с 1 января 1999 по 31 декабря 2018 год.

Предобработка данных

Перед тем как проводить разложения, данные необходимо предварительно отфильтровать того, чтобы уменьшить влияние суточной и сезонных периодичностей. Так как они будут вносить наибольший вклад в динамику первых компонент разложения ПЭС.

В предыдущих работах было показано, что результат анализа главных компонент существенно зависит от методики предварительной обработки данных, используемой системе координат и методике центрирования данных [4]. Следуя рекомендациям из этих же, при анализе главных компонент были предварительно исключены доминирующие периодичности (суточная и сезонная), а также учтены особенности сферической системы координат, которая используется при построении карт ПЭС. При задании условия ортогональности базисных векторов необходимо учитывать, что значения ПЭС представлены в сферической системе координат в точках, равноотстоящих по широте и долготе. Таким образом, различным узлам координатной сетки соответствуют различные площади на земной поверхности. Применительно к данным ПЭС учет координатной сетки производится путем домножения каждого отсчёта $X_{i,j}$ на величину $\sqrt{\cos\varphi_i}$, где φ_i - широта соответствующего узла сетки.

При разложении на выходе получается представление исходной матрицы в виде произведения двух матриц, меньшей размерности U и V . Интерпретация векторов U_k и V_k (составляющих эти матрицы) зависит от исходного представления данных. Как было указано выше, строки в матрице X в нашем случае соответствовали пространственному распределению ПЭС, а столбцы – временной динамике. Поэтому матрица векторов U , отвечающая за пространственное распределение компонент разложения, имела размерность 5112 строк (по числу узлов карты) и 10 столбцов (по числу компонент разложения). Матрица векторов V – 10 строк (по числу компонент разложения) и 7305 столбцов (по числу временных отсчетов).

Таким образом, прогнозируя только временные вариации и затем вновь перемножив её с матрицей пространственных изменений, можно получить предсказанную карту ПЭС, затратив при этом меньше вычислительных ресурсов.

Авторегрессионная модель

Для отправной точки задачи в решении задачи прогнозирования используется модель авторегрессии.

Авторегрессионная модель временного ряда (AR) - это модель, которая используется для прогнозирования временных рядов на основе их собственных предыдущих значений [5]. В AR-модели текущее значение временного ряда предсказывается как линейная комбинация предыдущих значений этого же ряда. Количество предыдущих значений, учитываемых в модели (т.н. "лаг"), определяется порядком модели $AR(p)$, где p - количество лагов. таких моделях будущее значение есть нечто иное как линейная комбинация предыдущих значений временного ряда.

AR-модель может быть представлена следующим образом:

$$y_t = C + w_1 y_{t-1} + \dots + w_p y_{t-p} + \varepsilon_t, \quad (1)$$

где y_t – текущее значение временного ряда;

C - константа;

w_1, \dots, w_p – коэффициенты авторегрессии;

ε_t – остаток (шум) в момент времени t

AR-модель может быть расширена до модели ARIMA (авторегрессионная интегрированная скользящая средняя), добавляя в нее компоненты разности и скользящего среднего для учета трендов и сезонности в данных.

Для каждой из 10 компонент временного разложения была подобрана расширенная модель ARIMA. Прогноз производился на сутки вперед. Результаты приведены в таблице 1.

Таблица 1. Результаты сравнения двух подходов.

№ компоненты	Ошибка AP-модели	Ошибка MO-модели
1	0.11	0.09
2	0.16	0.11
3	0.24	0.19
4	0.24	0.17
5	0.31	0.25
6	0.34	0.25
7	0.42	0.33
8	0.53	0.45
9	0.67	0.53
10	0.81	0.65

Внешние данные

В работе [6] было показано, что временные компоненты разложения компоненты разложения исходных карт ПЭС коррелируют с данными солнечных параметров, такими, как, например индекс F10.7 солнечной активности и K_p магнитной активности. Это говорит о том, что, используя эти показатели как внешние можно уточнить прогноз, полученный при помощи авторегрессионной модели.

Для учета F10.7 были использованы данные, взятые с радиоастрофизической обсерватории Доминиона (DRAO) в Пентиктоне, Канада [7]. Глобальное суточное значение этого индекса измеряется в полдень по местному времени. Исторически этот индекс использовался в качестве входных данных для моделей ионосферы в качестве суррогата солнечного излучения на длинах волн, которые вызывают фотоионизацию в ионосфере Земли (в ультрафиолетовых диапазонах).

Данные о индексе K_p использовались с архивов Потсдамского центра наук о Земле (Немецкий центр исследования Земли). Он оценивается несколькими обсерваториями

расположены в США, Канаде, Великобритании, Германии и Австралии. Каждая станция калибруется по своей широте и сообщает определенный К-индекс в зависимости от геомагнитной активности, измеренной в месте расположения магнитометра. Сам К-индекс представляет собой трехчасовой квазилогарифмический локальный индекс геомагнитной активности в данном месте и времени по сравнению со спокойной дневной кривой. Магнитометр измеряет максимальное отклонение горизонтальной составляющей магнитного поля в месте его расположения и сообщает об этом. Затем глобальный К_р-индекс определяется с помощью алгоритма, который объединяет сообщаемые значения К каждой станции. Индекс К_р варьируется от 0 до 9, где значение 0 означает, что геомагнитная активность очень мала, а значение 9 означает сильную геомагнитную бурю [8].

Дополнительно были извлечены данные о месяце, дне и часе, на который делается прогноз.

Уточненная модель прогноза

Схематично предложенную модель можно представить так же, как на рисунке ниже:



Рис. 1. Блок-схема подхода для уточнения прогноза авторегрессионной модели

Таким образом расширяется признаковое пространство, на котором будет обучаться модель прогнозирования. На вход алгоритма машинного обучения (МО) подаются следующие данные:

- прогноз, сделанные авторегрессионной моделью;
- индекс F10.7 солнечной активности;
- индекс К_р магнитной активности;
- информация о месяце, дне и часе даты, для которой делается прогноз.

В качестве итогового алгоритма был выбран градиентный бустинг.

Градиентный бустинг – это, строго говоря, техника обучения машинного обучения. Она использует идею о том, что следующая модель будет учиться на ошибках предыдущей. Предсказателями выступают, как правило, слабые деревья решений. Из-за того, что предсказатели обучаются на ошибках, совершенных предыдущими, требуется меньше времени для того, чтобы добраться до реального ответа. Критерий остановки с осторожностью, иначе это может привести к переобучению. Градиентный бустинг — это пример бустинга, где каждая последующая модель старается уменьшить антиградиент предыдущей.

Результаты

Исходный набор векторов (87660) изначально был разбит на обучающую и тестовую выборки. 80 % (70128 векторов) было использовано для тренировки, итоговая же оценка алгоритма производилась на тестовых данных, которые составляли 20 % от всей первоначальной выборки (17532 вектора).

Сравнение двух различных подходов осуществлялось на основе метрики SMAPE (symmetric mean absolute error):

$$SMAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{(|y_i| + |\hat{y}_i|)/2}, \quad (2)$$

где y_i – истинное значение;
 \hat{y}_i – спрогнозированное значение;
 n – число элементов в выборке.

Результаты тестирования обоих подходов для каждой из 10 временных компонент полного электронного содержания приведены в таблице 1.

Выводы

Была реализована модель машинного обучения для уточнения краткосрочного прогноза в условиях повышенной солнечной активности. При учете дополнительных внешних параметров (как солнечных, так и земных), точность предсказания повышается для всех 10 главных компонент временных вариаций полного электронного содержания. Примечательно, что с ростом номера компоненты разница между точностями моделей по метрике SMAPE также увеличивается.

Литература

1. Пашинцев В.П., Ахмадеев Р.Р. Прогнозирование помехоустойчивости спутниковых радиосистем по данным GPS-мониторинга ионосферы // Электросвязь. 2015. № 11. С. 58–65.
2. Tsagouri I., Koutroumbas K., Elias P. A new short-term forecasting model for the total electron content storm time disturbances // Journal of Space Weather and Space Climate. 2018, no. 8, pp. 2–12. DOI: 10.1051/swsc/2018019.
3. Лаборатория JPL. URL: <https://www.jpl.nasa.gov> (дата обращения: 25.03.2024).
4. Аппалонов А. М., Масленникова Ю. С. Нейросетевое прогнозирование динамики экваториальной аномалии по данным полного электронного содержания ионосферы // Техника радиосвязи. 2021. № 3 (50). С. 29–42. DOI: 10.33286/2075-8693-2021-50-29-42
5. Тырсин А.Н., Построение моделей авторегрессии временных рядов при наличии помех / Математическое моделирование, 2005, № 5 (17), С. 10–16
6. Аппалонов А.М., Масленникова Ю.С. Анализ глобальной динамики полного электронного содержания за 23-й и 24-й циклы солнечной активности с использованием метода главных компонент // Радиотехника. 2023. Т. 87. № 12. С. 46–55. DOI: <https://doi.org/10.18127/j00338486-202312-06>
7. Электронный ресурс: лаборатория DRAO. URL: <https://nrc.canada.ca/en/research-development/nrc-facilities/dominion-radio-astrophysical-observatory-research-facility> (дата обращения: 25.03.2024)
8. Электронный ресурс: лаборатория в г. Постдам. URL: <https://kp.gfz-potsdam.de/en/> (дата обращения: 25.03.2024)