

Поиск ключевых управляющих параметров для оперативного прогноза полного электронного содержания ионосферы

А. В. Жуков¹, Д. Н. Сидоров^{1,2,3}, А. А. Мыльникова¹, Ю. В. Ясюкевич^{1,3}

¹ *Институт солнечно-земной физики СО РАН, Иркутск, 664033, Россия*

² *Институт систем энергетики им. Л. А. Мелентьева СО РАН
664033, Россия, Иркутск*

³ *Иркутский государственный университет, Иркутск, 664003, Россия
E-mail: zhukovalex13@gmail.com*

Оперативное прогнозирование динамики ионосферных параметров является актуальной и при этом достаточно сложной задачей. Один из основных вопросов состоит в выборе управляющих параметров для построения прогнозной модели. В работе предложен подход, основанный на технологии машинного обучения для решения указанной задачи. Данные вертикального абсолютного полного электронного содержания (ПЭС) с временным разрешением 30 мин используются в качестве экспериментальных. Данные получены с использованием фазовых и групповых измерений ПЭС на среднеширотной станции IRKJ (52° с. ш., 104° з. д.) за 2014 г. Результаты показали, что ключевыми управляющими параметрами для модели прогнозного значения ПЭС являются текущее значение ПЭС, оценка производной ПЭС, локальное время LT, текущие значения F10.7 и SYM/H, экспоненциально взвешенные скользящие средние ПЭС с периодами 12 и 24 ч и SYM/H с периодами 24 и 96 ч, а также ранее полученные данные со смещением, такие как значение ПЭС с задержкой 12 ч, F10.7 с задержкой 3 и 15 суток. В работе построены эмпирические модели оперативного прогноза на основе параметров с использованием рекурсивного отбора признаков с определением их значимости при помощи методов случайного леса и опорных векторов. При использовании указанных параметров линейная регрессионная модель позволяет получить оценку на интервале 4–7 ч со среднеквадратичным отклонением СКО ~4,5 TECU. Методы машинного обучения — случайный лес, метод опорных векторов и градиентный бустинг — позволяют сократить СКО до 3–3,5 TECU.

Ключевые слова: оперативный прогноз, абсолютное полное электронное содержание, машинное обучение, случайный лес, метод опорных векторов, градиентный бустинг

Одобрена к печати: 21.05.2018

DOI: 10.21046/2070-7401-2018-15-3-263-272

Введение

Ионосфера представляет собой высокодинамичную среду, значительно зависящую от потока солнечного излучения, метеорологических процессов, солнечного ветра (Forbes et al., 2000). Оперативный прогноз параметров ионосферы является достаточно актуальной задачей, особенно для систем радиосвязи и радиолокации (Куркин и др., 1997; Jakowski et al., 2002; Zolesi, Cander, 2014). В настоящее время активно разворачиваются комплексы мониторинга ионосферных параметров с использованием различных средств (Смирнов, Смирнова, 2010; Данилкин и др., 2014). В Европе создаётся беспрецедентная система мониторинга и предсказания ионосферных параметров, основная задача которой — поддержка пользователей КВ-радиосвязи (Belehaki et al., 2015; Stamper et al., 2004). Широкое развитие получило использование глобальных навигационных спутниковых систем для решения задач мониторинга ионосферы (Afraimovich et al., 2013) и корректировки радиотехнических систем (Ясюкевич и др., 2017), а также для построения глобальных (<ftp://cddis.gsfc.nasa.gov/gps/products/ionex/>) и региональных (<http://www.sws.bom.gov.au/Satellite/2/1>) ионосферных карт (Hernández-Pajares et al., 2009; Schaer et al., 1998).

На сегодняшний день существуют сложные методы прогнозирования ионосферных параметров, основанные, например, на нейронных сетях (El-naggar, 2013; Huang, Yuan, 2014; Tulunay et al., 2004). Так, авторы работы (Habarulema et al., 2011) использовали нейронные

сети для анализа пространственных и временных вариаций полного электронного содержания (ПЭС) ионосферы. Несмотря на это, задача оперативного прогноза в настоящее время полностью не решена. В ряде случаев активно и успешно используются более простые методы для предсказания динамики состояния ионосферы в регионе и в глобальном масштабе (García-Rigo et al., 2011).

Один из методов оперативного прогноза может использовать эмпирическое моделирование и корректировку прогноза по текущим данным. Например, модель SIRM (The Simplified Ionospheric Regional Model) основана на линейной регрессии фурье-коэффициентов измерений на индекс R12 (ряд чисел Вольфа, сглаженный с 12-месячным окном (Zolesi et al., 1993, 2004)). В работе (Hajra et al., 2016) используется многомерный регрессионный анализ, где в качестве параметров модели берутся индекс F10.7, сезон, локальное время и индекс аврорального электроджета EЕJ. При этом чаще интуитивно с использованием известных корреляционных связей выбираются ключевые параметры. В первую очередь такими параметрами являются индексы солнечной (R12, F10.7, UV) и/или геомагнитной активности (SYM/H, K_p , A_p , AE или др.). В таких условиях тот или иной важный параметр, способный улучшить качество прогноза, может быть просто пропущен.

В настоящей работе, продолжая наши предыдущие исследования (Zhukov et al., 2018), мы строим модель прогноза вертикального ПЭС I_v с использованием технологии машинного обучения. При построении модели мы определяем, какие из параметров являются ключевыми для построения, и проводим сравнение с простейшими «интуитивными аналогами».

1. Исходные данные и предварительный выбор параметров модели прогноза

В качестве входных данных мы использовали данные абсолютного вертикального ПЭС I_v , полученные на основе двухчастотных совместных фазовых и групповых измерений (Yasyukevich et al., 2015a, b). Единица измерения ПЭС — TECU (total electron content unit, $1 \text{ TECU} = 10^{16} \text{ м}^{-2}$). Кроме ПЭС, проводились оценки его первых и вторых производных по времени (G_t и G_{tq}) и по пространству. Для анализа использовались данные в формате RINEX (Gurtner, Estey, 2007) станции IRKJ, входящей в международную службу ГНСС (IGS) (Dow et al., 2009), за полный 2014 г. Для характеристики состояния солнечной активности использовался индекс F10.7 (<ftp://spdf.gsfc.nasa.gov/pub/data/omni/>), а магнитосферной возмущённости — индексы SYM/H и AE (<ftp://spdf.gsfc.nasa.gov/pub/data/omni/>). Для построения модели брались данные за 1–273 дни года. Данные за 274–365 дни использовались для её тестирования.

Первым этапом является создание множества потенциально информативных признаков на основе исходных данных. При этом можно руководствоваться разными теоретическими и эмпирическими соображениями. В работе используются следующие признаки: I_v — абсолютное вертикальное ПЭС; G_t и G_{tq} — первая и вторая производные ПЭС по времени; AE — индекс аврорального электроджета; SYM/H — индекс геомагнитной возмущённости; F10.7 — поток солнечного радиоизлучения на длине волны 10,7 см (индекс F10.7); LT — местное время (ч).

В данной работе в качестве исходных признаков, помимо самих параметров I_v , G_t и G_{tq} , SYM/H, AE , F10.7, используются экспоненциально взвешенные скользящие средние с различными периодами и ранее полученные данные с разными запаздываниями. Запаздывание может быть выбрано с помощью анализа кросскорреляционной функции (АКФ) между анализируемым и целевым (I_v) параметрами. Например, максимумы кросскорреляционной функции рядов F10.7 и I_v имеют место при сдвиге, равном 3 сут, 15 сут и т. д. В данной работе для выбора запаздывающих значений анализируется АКФ сигнала. Период скользящих средних выбирался эмпирически. Для анализа выбраны следующие значения: I_v с временным лагом 30 мин, 12, 24 ч; AE с временным лагом 12, 24, 48 ч; SYM/H с временным лагом 25, 50, 125 ч; экспоненциально взвешенные скользящие средние I_v с периодами 2, 3, 4,

12, 24 ч; экспоненциально взвешенные скользящие средние SYM/H с периодами 24, 48, 72, 96 ч. Также используются априорные знания о периодическом характере рассматриваемых временных рядов. Для этого в качестве признака используется косинус локального времени $\cos(2\pi LT/24)$, обозначенный далее как $\cos(LT)$.

2. Методы машинного обучения

Машинное обучение — наиболее активно развивающийся в последние годы раздел теории искусственного интеллекта — направлен на создание алгоритмов, способных обучаться, выделять значимые признаки (параметры модели) и даже адаптировать свою структуру на основании обрабатываемой выборки данных (см., например, (Жуков, Сидоров, 2016; Kurbatsky et al., 2011; Zhukov et al., 2017)) в режиме реального или близкого к реальному времени.

Для решения задачи выбора ключевых параметров могут использоваться разные подходы: от полного перебора, который становится проблематичным при большом количестве признаков, до методов, основанных на итеративном добавлении или удалении признака из набора и дальнейшей проверке качества полученной модели (так называемый метод *leave-one-out*). Для эвристического сокращения перебора в последнем случае используются подходы ранжирования признаков, основанные на таких методах, как метод опорных векторов (МОВ) (Chen, Lin, 2006), композиционные методы на основе решающих деревьев типа случайного леса (СЛ) (Breiman, 2001), градиентный бустинг над решающими деревьями (ГБРД) (Saeys et al., 2008) и др.

В данной работе для поиска оптимального подмножества признаков используется метод рекурсивного удаления признаков на основе оценки значимости СЛ. Выбор именно этого метода обусловлен простотой его настройки и устойчивостью к влиянию неинформативных признаков, которая находит подтверждение во многих эмпирических исследованиях (Dáz-Uriarte, De Andres, 2006; Granitto et al., 2006).

3. Поиск ключевых параметров

Формирование данных и предварительная обработка. Первоначальный набор данных, описанный выше, был разделён на следующие части: 67,5 % — для обучения, 22,5 % — для тестирования и 10 % — для проверки. Учебный набор данных используется для подгонки модели, тестирование — для оценки производительности моделей и настройки параметров модели. Проверочный набор использовался для окончательной оценки точности и визуализации результатов.

Извлечение признаков (исходного набора) описано выше в разделе 1. Процедура позволила выделить 26 параметров в качестве признаков. Для удобства обозначим их следующим образом: приставка MAN означает экспоненциально взвешенное скользящее среднее с окном N ч; LN показывает, что используется величина с запаздыванием N ч. Например, G_t_MA4 означает экспоненциальное скользящее G_t с 4-часовым окном; I_v_L1 означает значение I_v с запаздыванием 1 ч.

Выбор признаков. Чтобы получить оптимальный набор признаков, мы используем метод рекурсивного исключения объектов со случайным лесом. Суть этого итеративного метода заключается в том, что мы получаем набор признаков, из них последовательно удаляем самые неинформативные, подгоняем модель и оцениваем её точность. Набор, показывающий максимальную точность, считается оптимальным, а признаки — ключевыми. На основе ключевых признаков формируется окончательная модель машинного обучения. Метод СЛ позволил выделить следующие признаки в качестве ключевых: I_v , G_t , F10.7, SYM/H, $\cos(LT)$, I_v_L12 , F10.7_L72, F10.7_L360, I_v_MA24 , I_v_MA12 , SYM/H_MA96, SYM/H_MA24.

Представляет интерес сравнение значимости, получаемой различными методами, например СЛ и МОВ. Полученные оценки значимости показаны на *рис. 1* (см. с. 266). Отметим, что как в случае метода СЛ, так и МОВ наиболее значимыми оказались G_t , I_v , $\cos(LT)$.

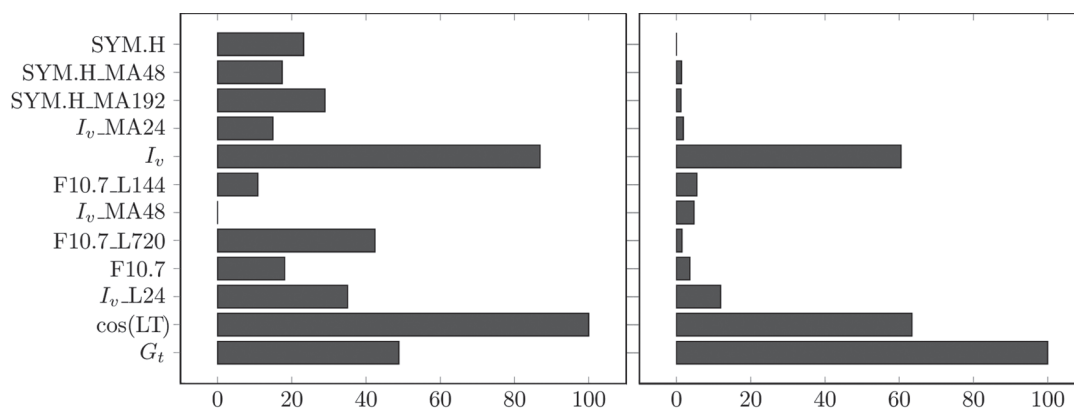


Рис. 1. Относительная значимость выбранных признаков согласно моделям СЛ (слева) и МОВ (справа)

4. Результаты прогнозирования

С использованием ретроспективных данных были построены три прогнозные модели ПЭС. На рис. 2 показаны исходные данные вертикального ПЭС и прогнозные значения для моделей. Качество оценивается на основе двух параметров: среднеквадратичного отклонения (СКО) и среднего абсолютного отклонения (САО). Результаты оценки качества работы моделей с использованием оставшихся 10 % данных приведены в таблице. Видно, что все модели на основе методов машинного обучения демонстрируют довольно хорошие результаты.

Характерные ошибки (СКО и САО) разных моделей для горизонта 4 ч. Значения даны в TECU

	СЛ	ГБДР	МОВ	N1	N2	ЛР
СКО	3,49	3,30	4,49	9,14	16,51	4,61
САО	2,49	2,35	3,50	6,74	12,23	3,66

СКО для моделей машинного обучения составило ~3–4,5 TECU. Наилучшие результаты показал метод ГБДР. Были протестированы также две «наивные» модели. Модель N1 соответствует отсутствию изменений в ионосфере в прогнозном периоде $I(t + \Delta h) = I(t)$. Для 4-часового прогноза СКО составило ~9,14 TECU.

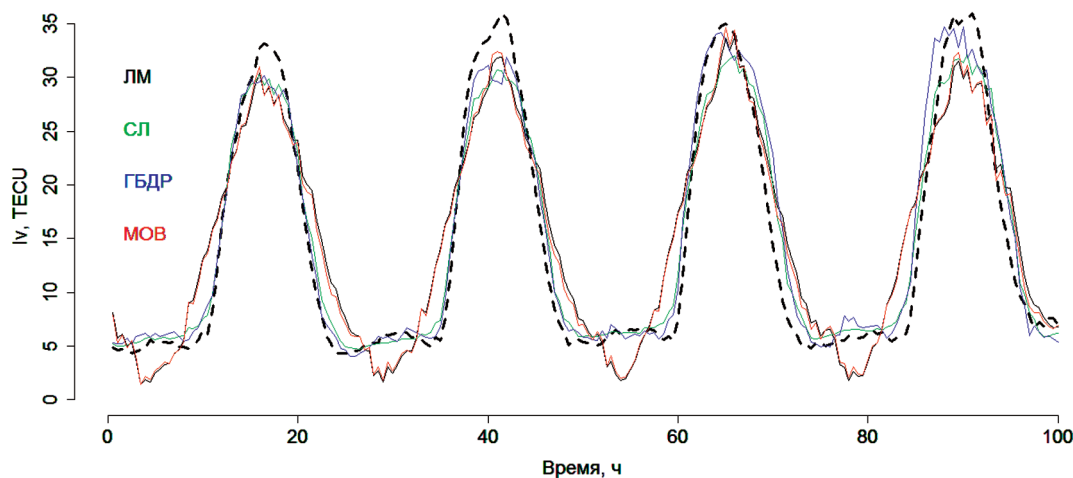


Рис. 2. Исходные данные вертикального ПЭС (пунктир) и прогнозные значения для моделей на основе машинного обучения: красная кривая — МОВ, синяя — ГБДР, зелёная — СЛ, чёрная — ЛМ

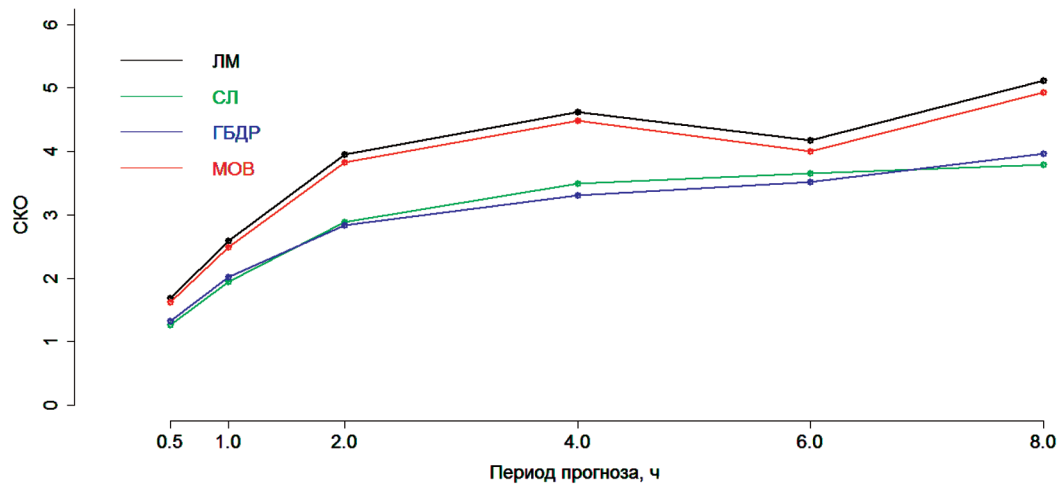


Рис. 3. Среднеквадратичная ошибка прогноза для различных моделей в зависимости от горизонта прогноза: красная кривая — МОВ, синяя — ГБДР, зелёная — СЛ, чёрная — ЛМ

Вторая наивная модель (N2) основана на первой производной ПЭС: $I(t + \Delta t) = I(t) + G_t \Delta t$, где t — время. Для данной простой модели СКО составило 16,5 TECU. Такая линейная модель может использоваться только для небольших периодов времени, а не для 4-часового интервала. Также была построена многопараметрическая модель линейной регрессии (ЛР) на значимых параметрах. Для этой модели СКО составило ~4,6 TECU, что является вполне приемлемым результатом для такого периода.

На рис. 3 показана зависимость СКО прогнозов, построенных на основе разных моделей, от горизонта прогноза (заблаговременности) в диапазоне 30 мин — 8 ч. Видно, что модели на основе композиционных методов ГБДР и СЛ дают наилучшие результаты. Эффективность МОВ можно повысить путём оптимизации процедуры выбора её параметров, в частности генетическим алгоритмом (см., например, работу (Li et al., 2017)).

На основании вышеизложенного можно сделать вывод, что динамика спрогнозированной ПЭС соответствует экспериментальным данным. При этом модели машинного обучения корректно определяют амплитуду колебаний (см. рис. 2).

Заключение

В работе рассмотрена задача поиска ключевых управляющих параметров для оперативного прогноза полного электронного содержания ионосферы. Результаты показывают, что современные методы машинного обучения позволяют получить СКО от 1,5 до 4 TECU для горизонтов прогноза от 30 мин до 6 ч. В качестве ключевых параметров были выделены I_p , G_p , F10.7, SYM/H, $\cos(LT)$, I_v _L12, F10.7_L72, F10.7_L360, I_v _MA24, I_v _MA12, SYM/H.MA96, SYM/H.MA24. Стоит отметить, что линейная регрессия, построенная на ключевых параметрах, показывает достаточно хорошие результаты с СКО ~4,5 TECU при 4-часовом горизонте прогноза. Методы машинного обучения позволяют уменьшить СКО до 3–3,5 TECU.

Аналогичный подход может быть использован для широкого круга задач, в том числе для оперативного прогноза геостационарного ПЭС (Kunitsyn et al., 2016), индексов возмущённости ионосферы (Nesterov et al., 2017) и сбоев функционирования ГНСС (Zakharov et al., 2016).

Авторы выражают благодарность Веснину А. М. и Сыроватскому С. В. за помощь в предварительной подготовке данных. В работе использованы данные Space Physics Data Facility (SPDF) и данные ГНСС сети IGS (Dow et al., 2009). Исследование выполнено при поддержке Российского научного фонда (проект № 17-77-20005). Экспериментальные данные ПЭС, используемые для построения моделей, получены в рамках базового финансирования программы ФНИ П.12.

Литература

1. Данилкин Н. П., Жбанков Г. А., Тасенко С. В. Восстановление трехмерного поля плотности электронов по результатам модельного эксперимента с участием бортового ионозонда и двух наземных ионозондов // Гелиогеофизические исследования. 2014. № 7. С. 43–55.
2. Жуков А. В., Сидоров Д. Н. Модификация алгоритма случайного леса для классификации нестационарных потоковых данных // Вестник ЮУрГУ. Сер. Математическое моделирование и программирование. 2016. Т. 9. № 4. С. 86–95. DOI: 10.14529/mmp160408.
3. Куркин В. И., Полех Н. М., Чистякова Л. В. Оперативный прогноз МПЧ при наклонном зондировании ионосферы // Исследования по геомагнетизму, аэронауке и физике Солнца. Изд-во СО РАН, 1997. Вып. 105. С. 168–174.
4. Смирнов В. М., Смирнова Е. В. Модуль ионосферного обеспечения на базе спутниковых систем GPS/ГЛОНАСС // Журн. радиоэлектроники. 2010. № 6.
5. Ясюкевич Ю. В., Оводенко В. Б., Мыльникова А. А., Живетьев И. В., Веснин А. М., Едемский И. К., Котова Д. С. Методы компенсации ионосферной составляющей ошибки радиотехнических систем с применением данных полного электронного содержания GPS/ГЛОНАСС // Вестник Поволжского гос. технологич. ун-та. Сер. Радиотехнические и инфокоммуникационные системы. 2017. Т. 2. № 34. С. 19–31. DOI: 10.15350/2306-2819.2017.1.19.
6. Afraimovich E. L., Astafyeva E. I., Demyanov V. V., Edemskiy I. K., Gavriluk N. S., Ishin A. B., Kosogorov E. A., Leonovich L. A., Lesyuta O. S., Palamartchouk K. S., Perevalova N. P., Polyakova A. S., Smolkov G. Y., Voeykov S. V., Yasyukevich Yu. V., Zhivetiev I. V. Review of GPS/GLONASS studies of the ionospheric response to natural and anthropogenic processes and phenomena // J. Space Weather and Space Climate. 2013. V. 3. A27. DOI: 10.1051/swsc/2013049.
7. Belehaki A., Tzagouri I., Kutiev I., Marinov P., Zolesi B., Pietrella M., Themelis K., Elias P., Tziotziou K. The European Ionosonde Service: nowcasting and forecasting ionospheric conditions over Europe for the ESA Space Situational Awareness services // J. Space Weather Space Climate. 2015. V. 5. A25. DOI: 10.1051/swsc/2015026.
8. Breiman L. Random forests // Machine learning. 2001. V. 45. No. 1. P. 5–32. DOI: 10.1023/A:101093340.
9. Chen Y. W., Lin C. J. Combining SVMs with various feature selection strategies // Feature Extraction. Studies in Fuzziness and Soft Computing / eds. Guyon I., Nikravesh M., Gunn S., Zadeh L. A. V. 207. Berlin: Springer, 2006. DOI: 10.1007/978-3-540-35488-8_13.
10. Dáz-Urriarte R., De Andres S. A. Gene selection and classification of microarray data using random forest // BMC bioinformatics. 2006. V. 7. No. 1. DOI: 3.10.1186/1471-2105-7-3.
11. Dow J. M., Neilan R. E., Rizos C. The International GNSS Service in a changing landscape of Global Navigation Satellite Systems // J. Geodesy. 2009. V. 83. No. 3–4. P. 191–198. DOI: 10.1007/s00190-0080300-3.
12. El-naggar A. M. Artificial neural network as a model for ionospheric TEC map to serve the single frequency receiver // Alexandria Engineering J. 2013. V. 52. No. 3. P. 425–432. DOI: 10.1016/j.aej.2013.05.007.
13. Forbes J. M., Palo S. E., Zhang X. Variability of the ionosphere // J. Atmospheric and Solar-Terrestrial Physics. 2000. V. 62. No. 8. P. 685–693.
14. García-Rigo A., Monte E., Hernández-Pajares M., Juan J. M., Sanz J., Aragón-Angel A., Salazar D. Global prediction of the vertical total electron content of the ionosphere based on GPS data // Radio Science. 2011. V. 46. No. 6. RS0D25. DOI: 10.1029/2010RS004643.
15. Granitto P. M., Furlanello C., Biasioli F., Gasperi F. Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products // Chemometrics and intelligent laboratory systems. 2006. V. 83. No. 2. P. 83–90. DOI: 10.1016/j.chemolab.2006.01.007.
16. Gurtner W., Estey L. RINEX: The receiver independent exchange format version 2.11. Bern: Astronomical Institute, University of Bern, 2007. hdl:10013/epic.43875.
17. Habarulema J. B., McKinnell L.-A., Opperman B. Regional GPS TEC modeling; Attempted spatial and temporal extrapolation of TEC using neural networks // J. Geophysical Research. 2011. V. 116. A04314. DOI: 10.1029/2010JA016269.
18. Hajra R., Chakraborty S. K., Tsurutani B. T., DasGupta A., Echer E., Brum C. G. M., Gonzalez W. D., Sobral J. H. A. An empirical model of ionospheric total electron content (TEC) near the crest of the equatorial ionization anomaly (EIA) // J. Space Weather Space Climate. 2016. V. 6. A29. DOI: 10.1051/swsc/2016023.
19. Hernández-Pajares M., Juan J. M., Sanz J., Orus R., García-Rigo A., Feltens J., Komjathy A., Schaer S. C., Krankowski A. The IGS VTEC maps: a reliable source of ionospheric information since 1998 // J. Geodesy. 2009. V. 83. No. 3–4. P. 263–275. DOI: 10.1007/s00190-008-0266-1.

20. *Huang Z., Yuan H.* Ionospheric single-station TEC short-term forecast using RBF neural network // *Radio Science*. 2014. V. 49. P. 283–292. DOI: 10.1002/2013RS005247.
21. *Jakowski N., Wehrenpfennig A., Heise S., Kutiev I.* Space weather effects on transionospheric radio wave propagation on 6 April 2000 // *Acta Geodaetica et Geophysica Hungarica*. 2002. V. 37. No. 2–3. P. 213–220.
22. *Kunitsyn V.E., Padokhin A.M., Kurbatov G.A., Yasyukevich Y.V., Morozov Y.V.* Ionospheric TEC estimation with the signals of various geostationary navigational satellites // *GPS Solutions*. 2016. V. 20. No. 4. P. 877–884. DOI: 10.1007/s10291-015-0500-2.
23. *Kurbatsky V.G., Sidorov D.N., Spiryaev V.A., Tomin N.V.* The hybrid model based on Hilbert-Huang transform and neural networks for forecasting of short-term operation conditions of power system // *The Power of Technology for a Sustainable Society: Proc. IEEE PES Trondheim PowerTech 2011*. 2011. P. 1–7. DOI: 10.1109/PTC.2011.6019155.
24. *Li Y., Wen Z., Cao Y., Tan Y., Sidorov D., Panasetsky D.* A combined forecasting approach with model self-adjustment for renewable generations and energy loads in smart community // *Energy*. 2017. V. 129. P. 216–227.
25. *Nesterov I.A., Andreeva E.S., Padokhin A.M., Tumanova Yu.S., Nazarenko M.O.* Ionospheric perturbation indices based on the low- and high-orbiting satellite radio tomography data // *GPS Solutions*. 2017. V. 21. No. 4. P. 1679–1694. DOI: 10.1007/s10291-017-0646-1.
26. *Saeys Y., Abeel T., Van de Peer Y.* Robust feature selection using ensemble feature selection techniques // *Proc. European Conf. Machine Learning and Knowledge Discovery in Databases (ECML PKDD 2008)*. Part II / eds. W. Daelemans, B. Goethals, K. Morik. Berlin, Heidelberg: Springer-Verlag, 2008. P. 313–325. DOI: 10.1007/978-3-540-87481-2_21.
27. *Schaer S., Beutler G., Rothacher M.* Mapping and predicting the ionosphere // *Proc. IGS AC Workshop*. Darmstadt, Germany. Feb. 9–11, 1998. 1998. P. 307–320.
28. *Stamper R., Belehaki A., Buresová D., Cander L.R., Kutiev I., Pietrella M., Stanislawski I., Stankov S., Tsagouri I., Tulunay Y.K., Zolesi B.* Nowcasting, forecasting and warning for ionospheric propagation: tools and methods // *Annals of Geophysics*. 2004. V. 47. No. 2–3. P. 957–983.
29. *Tulunay E., Senalp E.T., Cander L.R., Tulunay Y.K., Bilge A.H., Mizrahi E., Kouris S.S., Jakowski N.* Development of algorithms and software for forecasting, nowcasting and variability of TEC // *Annals of geophysics*. 2004. V. 47. No. 2–3. P. 1201–1214.
30. *Yasyukevich Yu.V., Mylnikova A.A., Polyakova A.S.* (2015a) Estimating the total electron content absolute value from the GPS/GLONASS data // *Results in Physics*. 2015. V. 5. P. 32–33. DOI: 10.1016/j.rinp.2014.12.006.
31. *Yasyukevich Y.V., Mylnikova A.A., Kunitsyn V.E., Padokhin A.M.* (2015b) Influence of GPS/GLONASS differential code biases on the determination accuracy of the absolute total electron content in the ionosphere // *Geomagnetism and Aeronomy*. 2015. V. 55. No. 6. P. 763–769. DOI: 10.1134/S001679321506016X.
32. *Zakharov V.I., Yasyukevich Yu.V., Titova M.A.* Effect of magnetic storms and substorms on GPS slips at high latitudes // *Cosmic Research*. 2016. V. 54. No. 1. P. 20–30. DOI: 10.1134/S0010952516010147.
33. *Zhukov A.V., Sidorov D.N., Foley A.M.* Random forest based approach for concept drift handling // *Communications in Computer and Information Science*. 2017. V. 661. P. 69–77. DOI: 10.1007/978-3-319-52920-2_7.
34. *Zhukov A., Sidorov D., Mylnikova A., Yasyukevich Yu.* Machine learning methodology for ionosphere total electron content nowcasting // *Intern. J. Artificial Intelligence*. 2018. V. 16. No. 1. P. 144–157.
35. *Zolesi B., Cander L.* Ionospheric prediction and forecasting // *Springer Geophysics*. 2014. 240 p. DOI: 10.1007/978-3-642-38430-1_1.
36. *Zolesi B., Cander L.R., Franceschi G.D.* Simplified ionospheric regional model for telecommunication applications // *Radio Science*. 1993. V. 28. No. 4. P. 603–612. DOI: 10.1029/93RS00276.
37. *Zolesi B., Belehaki A., Tsagouri I., Cander L.R.* Real-time updating of the Simplified Ionospheric Regional Model for operational applications // *Radio Science*. 2004. V. 39. No. 2. RS2011. DOI: 10.1029/2003RS002936.

Selecting the key control parameters for the ionospheric total electron content nowcasting

A. V. Zhukov¹, D. N. Sidorov^{1,2,3}, A. A. Mylnikova¹, Yu. V. Yasyukevich^{1,3}

¹ Institute of Solar-Terrestrial Physics SB RAS, Irkutsk 664033, Russia

² Melentiev Energy Systems Institute SB RAS, Irkutsk 664033, Russia

³ Irkutsk State University, Irkutsk 664003, Russia

E-mail: zhukovalex13@gmail.com

Nowcasting the dynamics of ionospheric parameters is an actual and at the same time rather complicated task. One of the main issues is the selection of control parameters for constructing accurate predictive model (feature selection). The approach is based on the machine learning technology for this problem solution. The vertical absolute total electron content (TEC) with a time resolution of 30 minutes is used as experimental data. The data were obtained using phase and group measurements of TEC at the mid-latitude IRKJ station (52 N, 104 E) for 2014. The results showed that the key control parameters for TEC nowcasting model are the current value of the TEC, the estimated TEC derivative, the local time, the current values of $F10.7$ and the SYM/H, and the exponentially weighted moving averaged TEC values with periods of 12 and 24 hours and SYM/H with periods of 24 and 96 hours, as well as previously received data with some lag, such as the vertical TEC with 12 hours lag, $F10.7$ with 3 and 15 days lags. The proposed empirical nowcasting models are based on parameters selected by recursive selection of characteristics with determination of their significance using random forest and support vectors methods. Using these key parameters, the linear regression model allows obtaining an estimate on the interval of 4–7 hours with RMS ~ 4.5 TECU. The machine learning methods such as random forest, support vector method and gradient boosting allow to reduce RMS to 3–3.5 TECU.

Keywords: nowcasting, absolute total electron content, machine learning, random forest, support vector machine, gradient boosting

Accepted: 21.05.2018

DOI: 10.21046/2070-7401-2018-15-3-263-272

References

1. Danilkin N. P., Zhbakov G. A., Tassenko S. V., Vosstanovlenie trekhmernogo polya plotnosti elektronov po rezul'tatam model'nogo eksperimenta s uchastiem bortovogo ionozonda i dvukh nazemnykh ionozondov (Reconstruction of a three-dimensional electron density field based on the results of a model experiment involving an onboard ionosonde and two ground ionosondes), *Geliogeofizicheskie issledovaniya*, 2014, No. 7, pp. 43–55.
2. Zhukov A. V., Sidorov D. N., Modifikatsiya algoritma sluchainogo lesa dlya klassifikatsii nestatsionarnykh potokovykh dannykh (Modification of the random forest algorithm for the classification of non-stationary streaming data), *Vestnik YuUrGU. Ser. Matematicheskoe modelirovanie i programmirovaniye*, 2016, Vol. 9, No. 4, pp. 86–95, DOI: 10.14529/mmp160408.
3. Kurkin V. I., Polekh N. M., Chistyakova L. V., Operativnyi prognoz MPCh pri naklonnom zondirovanii ionosfery (Operational forecast of MUF for oblique ionospheric sounding), *Issledovaniya po geomagnetizmu, aeronomii i fizike Solntsa*, Izd. SO RAN, 1997, Issue 105, pp. 168–174.
4. Smirnov V. M., Smirnova E. V., Modul' ionosferного obespecheniya na baze sputnikovykh sistem GPS/GLONASS (Ionospheric support module based on GPS/GLONASS satellite systems), *Zhurnal radioelektroniki*, 2010, No. 6.
5. Yasyukevich Yu. V., Ovodenko V. B., Mylnikova A. A., Zhivet'ev I. V., Vesnin A. M., Edemskii I. K., Kotova D. S., Metody kompensatsii ionosfernoi sostavlyayushchei oshibki radiotekhnicheskikh sistem s primeneniem dannykh polnogo elektronного soderzhaniya GPS/GLONASS (GPS/GLONASS total electron content based methods for ionospheric error compensation for the radio communication systems), *Vestnik Povolzhskogo gosudarstvennogo tekhnologicheskogo universiteta. Ser. Radiotekhnicheskie i infokommunikatsionnye sistemy*, 2017, No. 2 (34), pp. 19–31, DOI: 10.15350/2306-2819.2017.2.19.

6. Afraimovich E. L., Astafyeva E. I., Demyanov V. V., Edemskiy I. K., Gavriluk N. S., Ishin A. B., Kosogorov E. A., Leonovich L. A., Lesyuta O. S., Palamartchouk K. S., Perevalova N. P., Polyakova A. S., Smolkov G. Y., Voeykov S. V., Yasyukevich Yu. V., Zhivetiev I. V., Review of GPS/GLONASS studies of the ionospheric response to natural and anthropogenic processes and phenomena, *J. Space Weather and Space Climate*, 2013, Vol. 3, A27, DOI: 10.1051/swsc/2013049.
7. Belehaki A., Tsagouri I., Kutiev I., Marinov P., Zolesi B., Pietrella M., Themelis K., Elias P., Tziotziou K., The European Ionosonde Service: nowcasting and forecasting ionospheric conditions over Europe for the ESA Space Situational Awareness services, *J. Space Weather Space Climate*, 2015, Vol. 5, A25, DOI: 10.1051/swsc/2015026.
8. Breiman L., Random forests, *Machine learning*, 2001, Vol. 45, No. 1, pp. 5–32, DOI: 10.1023/A:101093340.
9. Chen Y. W., Lin C. J., Combining SVMs with various feature selection strategies, In: *Feature Extraction. Studies in Fuzziness and Soft Computing*, Guyon I., Nikravesh M., Gunn S., Zadeh L. A. (eds.), Vol. 207, Berlin: Springer, 2006, DOI: 10.1007/978-3-540-35488-8_13.
10. Dáz-Uriarte R., De Andres S. A., Gene selection and classification of microarray data using random forest, *BMC bioinformatics*, 2006, Vol. 7, No. 1, DOI: 3.10.1186/1471-2105-7-3.
11. Dow J. M., Neilan R. E., Rizos C., The International GNSS Service in a changing landscape of Global Navigation Satellite Systems, *J. Geodes.*, 2009, Vol. 83, No. 3–4, pp. 191–198, DOI: 10.1007/s00190-0080300-3.
12. El-naggar A. M., Artificial neural network as a model for ionospheric TEC map to serve the single frequency receiver, *Alexandria Engineering J.*, 2013, Vol. 52, No. 3, pp. 425–432, DOI: 10.1016/j.aej.2013.05.007.
13. Forbes J. M., Palo S. E., Zhang X., Variability of the ionosphere, *J. Atmospheric and Solar-Terrestrial Physics*, 2000, Vol. 62, No. 8, pp. 685–693.
14. García-Rigo A., Monte E., Hernández-Pajares M., Juan J. M., Sanz J., Aragón-Angel A., Salazar D., Global prediction of the vertical total electron content of the ionosphere based on GPS data, *Radio Science*, 2011, Vol. 46, No. 6, RS0D25, DOI: 10.1029/2010RS004643.
15. Granitto P. M., Furlanello C., Biasioli F., Gasperi F., Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products, *Chemometrics and intelligent laboratory systems*, 2006, Vol. 83, No. 2, pp. 83–90, DOI: 10.1016/j.chemolab.2006.01.007.
16. Gurtner W., Estey L., *RINEX: The Receiver Independent Exchange Format Version 2.11*. Bern: Astronomical Institute, University of Bern, 2007, hdl:10013/epic.43875.
17. Habarulema J. B., McKinnell L.-A., Opperman B., Regional GPS TEC modeling; Attempted spatial and temporal extrapolation of TEC using neural networks, *J. Geophys. Res.*, 2011, Vol. 116, A04314, DOI: 10.1029/2010JA016269.
18. Hajra R., Chakraborty S. K., Tsurutani B. T., DasGupta A., Echer E., Brum C. G. M., Gonzalez W. D., Sobral J. H. A., An empirical model of ionospheric total electron content (TEC) near the crest of the equatorial ionization anomaly (EIA), *J. Space Weather Space Climate*, 2016, Vol. 6, A29, DOI: 10.1051/swsc/2016023.
19. Hernández-Pajares M., Juan J. M., Sanz J., Orus R., Garcia-Rigo A., Feltens J., Komjathy A., Scherer S. C., Krankowski A., The IGS VTEC maps: a reliable source of ionospheric information since 1998, *J. Geod.*, 2009, Vol. 83, No. 3–4, pp. 263–275. DOI: 10.1007/s00190-008-0266-1.
20. Huang Z., Yuan H., Ionospheric single-station TEC short-term forecast using RBF neural network, *Radio Sci.*, 2014, Vol. 49, pp. 283–292, DOI: 10.1002/2013RS005247.
21. Jakowski N., Wehrenpfennig A., Heise S., Kutiev I., Space weather effects on transionospheric radio wave propagation on 6 April 2000, *Acta Geod. Geophys. Hung.*, 2002, Vol. 37, No. 2–3, pp. 213–220.
22. Kunitsyn V. E., Padokhin A. M., Kurbatov G. A., Yasyukevich Y. V., Morozov Y. V., Ionospheric TEC estimation with the signals of various geostationary navigational satellites, *GPS Solutions*, 2016, Vol. 20, No. 4, pp. 877–884, DOI: 10.1007/s10291-015-0500-2.
23. Kurbatsky V. G., Sidorov D. N., Spiryaev V. A., Tomin N. V., The hybrid model based on Hilbert-Huang transform and neural networks for forecasting of short-term operation conditions of power system, *The Power of Technology for a Sustainable Society, Proc. IEEE PES Trondheim PowerTech 2011*, 2011, pp. 1–7, DOI: 10.1109/PTC.2011.6019155.
24. Li Y., Wen Z., Cao Y., Tan Y., Sidorov D., Panasetsky D., A combined forecasting approach with model self-adjustment for renewable generations and energy loads in smart community, *Energy*, 2017, Vol. 129, pp. 216–227.
25. Nesterov I. A., Andreeva E. S., Padokhin A. M., Tumanova Yu. S., Nazarenko M. O., Ionospheric perturbation indices based on the low- and high-orbiting satellite radio tomography data, *GPS Solutions*, 2017, Vol. 21, No. 4, pp. 1679–1694, DOI: 10.1007/s10291-017-0646-1.

26. Saeys Y., Abeel T., Van de Peer Y., Robust feature selection using ensemble feature selection techniques, *Proc. European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD 2008)*, W. Daelemans, B. Goethals, K. Morik (eds.), Springer-Verlag, Berlin, Heidelberg, 2008, pp. 313–325, DOI: 10.1007/978-3-540-87481-2_21.
27. Schaer S., Beutler G., Rothacher M., Mapping and predicting the ionosphere, *Proc. IGS AC Workshop. Darmstadt, Germany. February 9–11, 1998*, 1998, pp. 307–320.
28. Stamper R., Belehaki A., Buresová D., Cander L. R., Kutiev I., Pietrella M., Stanislawski I., Stankov S., Tsagouri I., Tulunay Y. K., Zolesi B., Nowcasting, forecasting and warning for ionospheric propagation: tools and methods, *Annals of geophysics*, 2004, Vol. 47, No. 2–3, pp. 957–983.
29. Tulunay E., Senalp E. T., Cander L. R., Tulunay Y. K., Bilge A. H., Mizrahi E., Kouris S. S., Jakowski N., Development of algorithms and software for forecasting, nowcasting and variability of TEC, *Annals of geophysics*, 2004, Vol. 47 sup., No. 2–3, pp. 1201–1214.
30. Yasyukevich Yu. V., Mylnikova A. A., Polyakova A. S., Estimating the total electron content absolute value from the GPS/GLONASS data, *Results in Physics*, 2015, Vol. 5, pp. 32–33, DOI: 10.1016/j.rinp.2014.12.006.
31. Yasyukevich Y. V., Mylnikova A. A., Kunitsyn V. E., Padokhin A. M., Influence of GPS/GLONASS differential code biases on the determination accuracy of the absolute total electron content in the ionosphere, *Geomagnetism and Aeronomy*, 2015, Vol. 55, No. 6, pp. 763–769, DOI: 10.1134/S001679321506016X.
32. Zakharov V. I., Yasyukevich Yu. V., Titova M. A., Effect of magnetic storms and substorms on GPS slips at high latitudes, *Cosmic Research*, 2016, Vol. 54, No. 1, pp. 20–30, DOI: 10.1134/S0010952516010147.
33. Zhukov A. V., Sidorov D. N., Foley A. M., Random forest based approach for concept drift handling, *Communications in Computer and Information Science*, 2017, Vol. 661, pp. 69–77, DOI: 10.1007/978-3-319-52920-2_7.
34. Zhukov A., Sidorov D., Mylnikova A., Yasyukevich Yu., Machine learning methodology for ionosphere total electron content nowcasting, *Intern. J. Artificial Intelligence*, 2018, Vol. 16, No. 1, pp. 144–157.
35. Zolesi B., Cander L., Ionospheric prediction and forecasting, *Springer Geophysics*, 2014, 240 p., DOI: 10.1007/978-3-642-38430-1_1.
36. Zolesi B., Cander L. R., Franceschi G. D., Simplified ionospheric regional model for telecommunication applications, *Radio Sci.*, 1993, Vol. 28, No. 4, pp. 603–612, DOI: 10.1029/93RS00276.
37. Zolesi B., Belehaki A., Tsagouri I., Cander L. R., Real-time updating of the Simplified Ionospheric Regional Model for operational applications, *Radio Sci.*, 2004, Vol. 39, No. 2, RS2011, DOI: 10.1029/2003RS002936.