

МЕТОД ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ ДЛЯ ПРОГНОЗИРОВАНИЯ ЗНАЧЕНИЙ ПОЛОСЫ КОГЕРЕНТНОСТИ ИЗМЕНЯЮЩЕГОСЯ ТРАНСИОНОСФЕРНОГО РАДИОКАНАЛА

А.А. Кислицын, Н.В. Рябова, Н.А. Конкин

Поволжский государственный технологический университет, Йошкар-Ола, Россия
KislitsinAA@volgatech.netaleshka_89_@mail.ru

METHOD OF INTELLIGENT DATA ANALYSIS FOR PREDICTING THE COHERENCE BANDWIDTH OF A VARIABLE TRANSIONOSPHERIC RADIO CHANNEL

A.A. Kislitsin, N.V. Ryabova, N.A. Konkin

Volga State University of Technology, Yoshkar-Ola, Russia
KislitsinAA@volgatech.netaleshka_89_@mail.ru

Аннотация. Решается задача прогнозирования значений полос когерентности транссионосферного канала связи с применением методов интеллектуального анализа данных. Для предварительной обработки суточного хода полосы когерентности транссионосферных радиоканалов применена декомпозиция временного ряда на трендовую, сезонную и вариационную составляющие. Для выделения трендовой составляющей используется фильтр Ходрика—Прескотта, к вариационной составляющей применен алгоритм машинного обучения Isolation Forest с целью детектирования и аппроксимации аномальных точек временного ряда. Для повышения точности прогнозирования использовалась модель машинного обучения с применением алгоритма XGBoost, основанного на методе сдвигов значений полосы когерентности с вычитанием. Формирование новых признаков за счет метода сдвига позволило повысить точность прогноза примерно на 11 % и снизить среднеквадратичное отклонение (СКО) прогнозного значения полосы когерентности на 8 МГц. Обработка трендовой и вариационной составляющей позволила повысить коэффициент детерминации (КД) на 4 % и снизить СКО прогнозного значения полосы когерентности на 2 МГц.

Ключевые слова: полоса когерентности, транссионосферный радиоканал, декомпозиция временного ряда, машинное обучение, XGBoost, коэффициент детерминации.

Abstract. Paper presents the findings of the research into the problem of predicting the values of the coherence bandwidth of a transionospheric communication channel with the use of intelligent data analysis methods. To perform preliminary processing of the diurnal variations of the coherence bandwidth of transionospheric radio channels, we applied the decomposition of the time series into trend, seasonal, and variational components. The Hodrick—Prescott filter was used to get the trend component, and the Isolation Forest machine learning algorithm was applied to the variational component to detect and approximate anomalous samples in the time series. To improve the accuracy of forecasting, we used a machine learning model based on the XGBoost algorithm, which is based on the method of shifting the values of the coherence bandwidth with subtraction. The formation of new features due to the shift method allowed us to increase the forecast accuracy by roughly 11 % and to reduce RMS of the predicted value of the coherence bandwidth by 8 MHz. The processing of the trend and variational components allowed us to rise the coefficient of determination by 4 % and to reduce the RMS of the predicted value by 2 MHz.

Keywords: coherence bandwidth, transionospheric radio channel, time series decomposition, machine learning, XGBoost, coefficient of determination.

ВВЕДЕНИЕ

Применение широкополосных сигналов в системах спутниковой связи вызвано получением возможности увеличения скорости передачи информации в транссионосферном канале связи, повышения уровня помехоустойчивости и энергетической скрытности и снижения уровня энергопотребления. Однако положительные стороны использования широкополосных сигналов отчасти нивелируются необходимостью ограничивать полосу частот каналов из-за влияния ионосферы, изменяя тем самым информационно-технические характеристики спутниковых систем связи. Влияние неоднородной во времени и пространстве ионосферы приводит к возникновению частотной дисперсии в транссионосферном канале передачи данных. В работах [Ivanov et al., 2017; Кислицын, 2019] показано, что степень искажения системных характеристик транссионосферного радиоканала можно оценить, сравнивая параметр, называемый полосой когерентности канала (предельная полоса частот), с полосой пропускания канала связи.

В свою очередь, полоса когерентности транссионосферного радиоканала определяется полным электронным содержанием (ПЭС) ионосферы, значение которого варьирует в зависимости от различных факторов: времени суток, сезона, геомагнитной обстановки, солнечной активности. Из-за различных вариаций ПЭС при обработке данных возникают выбросы (аномалии) в суточных вариациях полосы когерентности. Важной проблемой в задаче прогнозирования на основе применения интеллектуальных методов является отсутствие дополнительных предикторов суточных вариаций прогнозируемого параметра (полосы когерентности), которые бы влияли на точность итоговой модели машинного обучения. Для решения задачи прогнозирования полос когерентности предлагается использовать интеллектуальный анализ большого объема данных (Big Data) от навигационных приемников сети ГНСС. Целью работы является прогнозирование значений полосы когерентности на основе интеллектуального анализа данных.

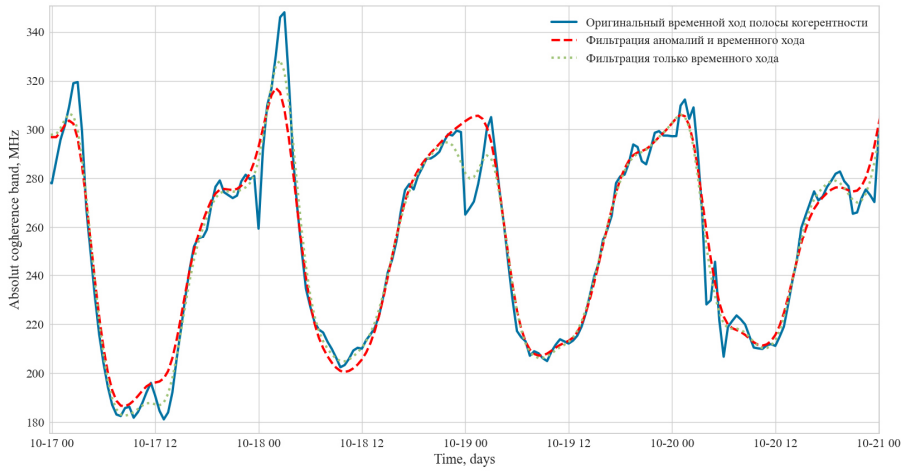


Рис. 1. Графики фильтрации временного ряда полосы когерентности

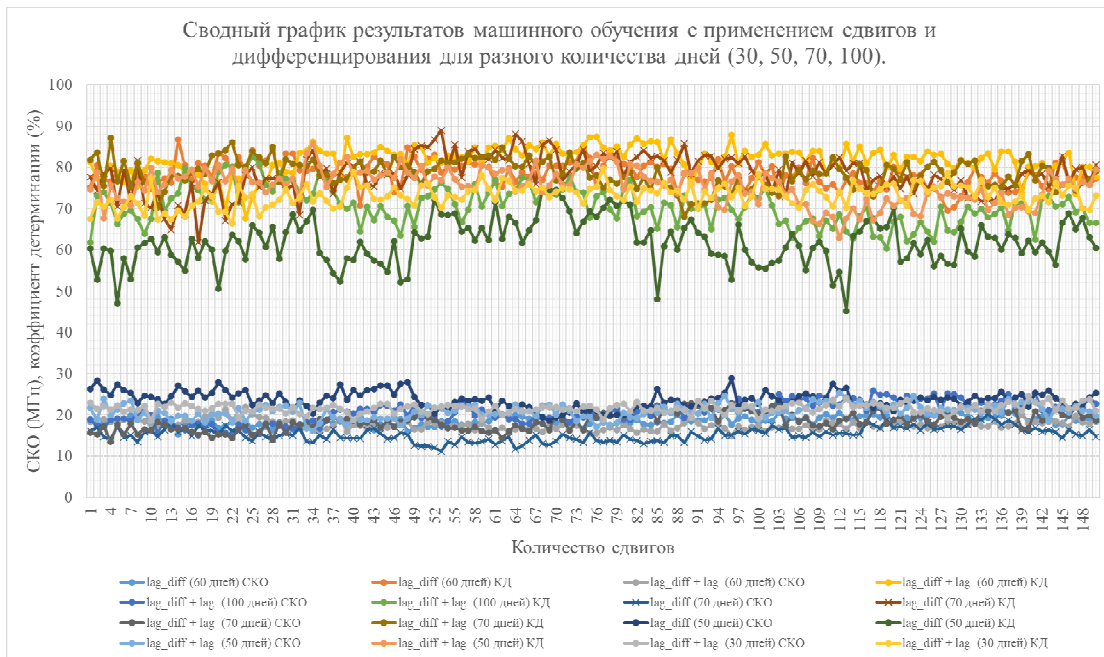


Рис. 2. Зависимость КД и SKO от количества сдвигов для случаев с комбинацией lag_diff + lag и только lag_diff, где lag_diff — сдвиг с вычитанием, lag — только сдвиг

ФИЛЬТРАЦИЯ ВАРИАЦИОННОЙ СОСТАВЛЯЮЩЕЙ ВРЕМЕННОГО РЯДА ПОЛОСЫ КОГЕРЕНТНОСТИ

Для работы радиотехнических систем, использующих трансионосферный радиоканал, важным параметром является полоса когерентности канала, т. е. полоса частот, для которой справедливо условие $B_{ch} < B_k$, где B_{ch} — полоса канала, B_k — полоса когерентности. В пределах этой полосы дисперсионными искажениями можно пренебречь.

Согласно [Ivanov et al., 2017; Ivanov, 2019], выражение для полосы когерентности можно представить в виде

$$B_k = \sqrt{(c\bar{f}^3) / (\pi k \bar{N}_t)}, \quad (1)$$

где $\bar{N}_t = \int_s N_e(z) dz$ — полное электронное содержание (истинное TEC); $k=80.5$ [m^3/c^2]; c — скорость света.

На первом этапе предварительной обработки точного хода полосы когерентности решалась задача декомпозиции временного ряда. Временные вариации можно описать тремя составляющими: трендом, сезонностью и вариациями с применением методов аппроксимации и определения выбросов вариаций полос когерентности. Сглаживание долгосрочных циклических компонент временного хода полосы когерентности осуществлялось на основе фильтра Ходрика—Прескотта. Вариационный ряд вычислялся как разность между исходным временным рядом, трендом и сезонной компонентой, поэтому в его структуре отсутствуют постоянные составляющие. Одной из основных задач фильтрации вариационной компоненты является детектирование отсчетов, которые вносят максимальное отклонение от тренда временного ряда. Определение максимальных отклонений (выбросов) вариационной составляющей реализовано с помощью алгоритма Isolation Forest, который представляет собой модель машинного обучения для определения ано-

мальных точек за счет генерирования случайных деревьев (алгоритм случайного леса) на исследуемой выборке, пропускание каждой точки через сгенерированное дерево и оценку расстояния конкретной точки в структуре дерева. Точки, имеющие меньшее расстояние, относятся к разряду аномальных.

На рис. 1 показаны результаты фильтрации временного ряда полосы когерентности с применением обработки компонент этого ряда и без нее.

Анализируя полученные результаты, установлено, что фильтрация компонент временного ряда позволяет более точно аппроксимировать резкие скачки значений полосы когерентности.

ПРИМЕНЕНИЕ МЕТОДОВ ДИФФЕРЕНЦИРОВАНИЯ В ЗАДАЧЕ ПРОГНОЗИРОВАНИЯ ЗНАЧЕНИЙ ПОЛОСЫ КОГЕРЕНТНОСТИ

На следующем этапе было проведено исследование влияния количества дней тренировочной выборки данных на результаты алгоритма XGBoost, а также количества сдвигов с использованием дифференцирования и без него. Для обучения алгоритма были доступны данные за 100 дней с 15 августа 2021 г. В ходе анализа влияния количества циклов (дней) на точность прогнозирования были выделены выборки 30, 50, 70 и 100 дней. Для каждого набора данных проведены вычисления влияния количества сдвигов с учетом вычитания и без него. Общий анализ показал, что с количеством сдвигов, близким к 150, снижается КД алгоритма и среднеквадратическое отклонение, поэтому в исследовании ограничивается число сдвигов (150). На рис. 2 показаны полученные зависимости СКО и КД от количества сдвигов.

Наилучшие результаты для всех наборов данных находятся в диапазоне 50–80 сдвигов. Лучший результат с КД 90 % и СКО 11.1 МГц показала выборка 70 дней с использованием lag_diff (сдвиг с вычитанием). Наихудшие результаты получены для 50-дневной выборки.

ЗАКЛЮЧЕНИЕ

Решена задача интеллектуальной обработки суточных вариаций полосы когерентности трансionoсферных радиоканалов с целью прогнозирования. Применен подход декомпозиции временного ряда на трендовую, сезонную и вариационную (ошибки) составляю-

щие. К трендовой составляющей применен фильтр Ходрика—Прескотта, который позволил выполнить аппроксимацию краткосрочных вариаций, связанных с неоднородностями ионосферы. К вариационной составляющей применен алгоритм машинного обучения Isolation Forest для детектирования и аппроксимации аномальных точек временного ряда. Формирование новых признаков за счет метода сдвига позволило повысить точность прогноза примерно на 11 % и снизить СКО на 8 МГц, а обработка трендовой и вариационной составляющей позволили повысить коэффициент детерминации на 4 % и снизить СКО ошибки на 2 МГц.

Работа выполнена при поддержке гранта Российского научного фонда, проект № 22-19-00073.

СПИСОК ЛИТЕРАТУРЫ

Кислицын А.А. Комплексный подход к адаптивной компенсации дисперсионных искажений системных характеристик широкополосных трансionoсферных радиоканалов. *Вестник Поволжского государственного технологического университета*. Сер. Радиотехнические и инфокоммуникационные системы. 2019. № 3 (43). С. 6–21.

Ivanov D.V., Ivanov V.A., Ryabova N.V., et al. Dispersive distortions of system characteristics of broadband transionospheric radio channels. *J. Applied Engineering Sci.* 2017. Vol. 15, no 4. P. 500–555. DOI: [10.5937/jaes15-11784](https://doi.org/10.5937/jaes15-11784).

Ivanov D.V., Ivanov V.A., Ryabova N.V., et al. The plotting algorithm of coherence band maps of transionospheric radio channels. *Proc. SPIE 10466. 23rd International Symposium on Atmospheric and Ocean Optics: Atmospheric Physics*. 2017. Vol. 10466. P. 1–4. DOI: [10.1117/12.2285658](https://doi.org/10.1117/12.2285658).

Ivanov V.A., Ryabova N.V., Ryabova M.I., et al. Hardware software system for producing coherent bandwidth maps of transionospheric wideband radio channel. *Systems of Signal Synchronization, Generating and Processing in Telecommunications. (SYNCHROINFO)*. 2019. DOI: [10.1109/SYNCHROINFO.2019.8813934](https://doi.org/10.1109/SYNCHROINFO.2019.8813934).