
ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОПРЕДЕЛЕНИЯ СПЕКТРАЛЬНЫХ ХАРАКТЕРИСТИК ИЗЛУЧЕНИЯ В ЭКСПЕРИМЕНТЕ «СОЛНЦЕ-ТЕРАГЕРЦ»

APPLICATION OF MACHINE LEARNING METHODS TO DETERMINE SPECTRAL CHARACTERISTICS OF RADIATION IN THE “SOLNTSE-TERAHERTZ” EXPERIMENT

Е.Д. Тульников

Физический институт им. П.Н. Лебедева РАН,
Москва, Россия, tulnikov.ed@yandex.ru

В.С. Махмутов 

Физический институт им. П.Н. Лебедева РАН,
Москва, Россия, mahmutovvs@lebedev.ru
Московский физико-технический институт,
Москва, Россия

М.В. Филиппов 

Физический институт им. П.Н. Лебедева РАН,
Москва, Россия, filippovmv@lebedev.ru

E.D. Tulnikov

P.N. Lebedev Physical Institute RAS,
Moscow, Russia, tulnikov.ed@yandex.ru

V.S. Makhmutov

P.N. Lebedev Physical Institute RAS,
Moscow, Russia, mahmutovvs@lebedev.ru
Moscow Institute of Physics and Technology,
Moscow, Russia

M.V. Philippov

P.N. Lebedev Physical Institute RAS,
Moscow, Russia, filippovmv@lebedev.ru

Аннотация. Исследуется возможность использования методов машинного обучения для анализа наблюдений с помощью научной аппаратуры «Солнце-Терагерц», изготовленной в ФИАН для установки на российском сегменте МКС. Научная аппаратура состоит из восьми детекторов, целевые частоты которых лежат в диапазоне 0.4–12.0 ТГц. Одной из основных целей эксперимента является изучение солнечных вспышек, спектр которых в этом диапазоне часто имеет U-образную форму. Основное внимание при определении параметров спектра уделено определению спектральных индексов спадающей и растущей частей спектра, а также положения точки перегиба. Для обучения алгоритмов использовались модельные данные интенсивности излучения, проходящего в оптические тракты научной аппаратуры. Данные были получены методами численного интегрирования. Проведенный анализ показал, что алгоритм Stacking демонстрирует наилучшую точность определения параметров спектра и может быть интегрирован в систему обработки данных будущего эксперимента на МКС, обеспечивая автоматическое предварительное восстановление параметров спектров солнечных вспышек.

Ключевые слова: Солнце, вспышка, субмиллиметровое излучение, машинное обучение.

Abstract. This paper explores the possibility of using machine learning methods for analyzing observations from the “Solntse-Terahertz” scientific equipment, developed at the Lebedev Physical Institute for installation on the Russian segment of the ISS. The scientific equipment consists of eight detectors, with target frequencies ranging from 0.4 to 12.0 THz. One of the primary goals of the experiment is to study solar flares whose spectra in this range often have a U-shaped form. The primary focus in determining the spectral parameters is on identifying spectral indices of the decaying and rising parts of the spectrum, as well as the position of the turnover point. The algorithms were trained using model data on the intensity of radiation passing through optical paths of the instrument. The data was obtained by numerical integration methods. The analysis has shown that the Stacking algorithm demonstrates the highest accuracy in determining the spectral parameters and can be integrated into the data processing system for future experiment on the ISS, enabling the automatic preliminary restoration of solar flare spectrum parameters.

Keywords: Sun, flare, submillimeter radiation, machine learning.

ВВЕДЕНИЕ

В лаборатории физики Солнца и космических лучей (Долгопрудненская научная станция (ДНС) ФИАН) была создана научная аппаратура (НА) «Солнце-Терагерц» [Kalinin et al., 2021], предназначенная для реализации космического эксперимента на российском сегменте МКС. Цель эксперимента — измерение солнечного излучения и вспышек в терагерцовом диапазоне (0.4–12 ТГц) с целью изучения

природы солнечной активности и механизмов ускорения заряженных частиц на Солнце [Wedemeyer et al., 2016; Kaufmann et al., 2004; Krucker et al., 2013] и других астрофизических объектах (например, сверхярких инфракрасных галактик и блазаров) [Tulnikov et al., 2025].

Аппаратура включает восемь детекторов, предназначенных для регистрации излучения на частотах 0.4, 0.8, 1.0, 3.0, 5.0, 7.0, 10.0 и 12.0 ТГц. Оптическая система каждого канала [Kvashnin et al., 2021] содержит

- зеркала с рассеивающим покрытием для ослабления излучения видимого и ИК-диапазонов;
- отрезающий и полосовой фильтры, выделяющие целевой частотный диапазон;
- оптический прерыватель, модулирующий сигнал 10 Гц для корректной работы детектора;
- оптоакустический преобразователь (ячейка Голея) с усилителем в качестве приемника излучения [Kalinin et al., 2021].

Подробные характеристики оптической системы и фильтров приведены в работе [Tulnikov et al., 2024]. Выбор ячейки Голея [Philippov et al., 2024a] обусловлен ее равномерным откликом в целевом спектральном диапазоне. Электронная часть [Philippov et al., 2024b] состоит из усилителей, драйверов прерывателей, блока питания и системы терморегуляции.

Возможность детектирования терагерцового излучения вспышек подобной аппаратурой подтверждена в ходе антарктического эксперимента GRIPS (19–30 января 2016 г.) [Kaufmann et al., 2016; Duncan et al., 2016].

Как было показано в статье [Tulnikov et al., 2024], коэффициенты пропускания оптических трактов для каналов 3, 5, 7 и 12 ТГц в области низких частот соизмеримы с пропусканием в окрестности целевой частоты. Этот факт не усложняет анализа монотонно возрастающих во всем диапазоне (0.4–12 ТГц) спектров, так как интенсивность радиоизлучения в окрестности целевой частоты превосходит интенсивность излучения в области низких частот в 10–1000 раз в зависимости от спектрального индекса. Однако, при анализе частотных спектров, в которых интенсивность в низкочастотной области сравнима с интенсивностью в окрестности целевой частоты, это приводит к неоднозначному соотношению между интенсивностью, измеренной НА, и фактической интенсивностью излучения в окрестности целевой частоты, что затрудняет восстановление спектра.

Для определения параметров спектра в данном случае необходимо моделировать отклик НА на спектр с заданными параметрами и подбирать такой набор параметров, при которых модельный отклик НА совпадет с экспериментальным. Поскольку перебрать все возможные комбинации параметров спектра не представляется возможным в силу их непрерывности, необходимо разработать алгоритм, способный по зарегистрированной различными каналами НА интенсивности излучения определить параметры спектра излучения, попавшего на вход оптических трактов НА. Это естественным образом приводит к постановке задачи регрессии в области машинного обучения, где входными переменными служат сигналы с восьми каналов НА, а выходными переменными — параметры частотного спектра радиоизлучения.

В работах [Kaufmann et al., 2004; Tsap et al., 2016; De Castro et al., 2005; Cristiani et al., 2007, 2009] была обнаружена интересная особенность спектра радиоизлучения ряда солнечных вспышек: ожидаемое уменьшение потоков излучения на частотах выше 70–90 ГГц сменяется значительным ростом излучения на частотах выше ~200–400 ГГц, т. е. наблюдается U-образная форма спектра в области субмил-

лиметрового излучения солнечных вспышек. Целью данной работы является создание алгоритма, способного определять параметры спектра U-образной формы (спектральный индекс спадающей части спектра, положение точки перегиба и спектральный индекс растущей части спектра), основываясь на зарегистрированной прибором интенсивности излучения.

1. ПОЛУЧЕНИЕ И ОБРАБОТКА ДАННЫХ

Как отмечалось ранее, существующие наблюдения солнечных вспышек в субтерагерцовом диапазоне указывают на вероятное нахождение в терагерцовом диапазоне точки, после которой спадающий гиротронный спектр сменяется растущим [Kaufmann et al., 2004; Tsap et al., 2016; De Castro et al., 2005; Cristiani et al., 2007, 2009], т. е. точки перегиба. Поэтому решено рассматривать спектр U-образной формы, спадающий со спектральным индексом γ до частоты перегиба ν_0 в диапазоне 0.4–12.0 ТГц, а затем возрастающий со спектральным индексом α (рис. 1). Рассмотрены спектры в диапазоне 0.016–15 ТГц. Спектральная плотность энергии в спектре нормировалась по значению потока на частоте 0.405 ТГц, которое было получено в работе [Kaufmann et al., 2004].

Для получения модельных данных интенсивности излучения, прошедшего через оптическую систему прибора при описанном выше спектре излучения, использовалась методика, рассмотренная в работе [Tulnikov et al., 2024]. Она подразумевает численное интегрирование произведения спектра излучения, попадающего на вход оптического тракта каждого канала НА, и функции пропускания соответствующего оптического тракта. Расчеты были проведены для значений $\gamma = -1.0 \div -3.0$ с шагом 0.5 и значений положения точки перегиба в интервале 0.4–12.0 ТГц с шагом 0.2 ТГц до 1 ТГц и далее с шагом 1 ТГц и для значений $\alpha = 1.0 \div 4.0$ с шагом 0.5. Таким образом, итоговый размер массива рассматриваемых спектров составил 9765 записей. Разбиение на тренировочную (train) и тестовую (test) выборки производилось в отношении 4:1, т. е. размер тренировочной выборки составил 7812 записей, а тестовой — 1953 записи.

Поскольку по форме спектра интенсивность определяется с точностью до множителя, было принято решение в качестве признаков для обучения модели использовать все возможные отношения интенсивности излучения, проходящего через оптические тракты всех каналов НА, к интенсивности излучения, проходящего через оптические тракты каналов с меньшей целевой частотой. Таким образом, получено 28 признаков. Обнаружено большое количество коррелированных признаков, что при обучении модели может приводить к проблемам, связанным с неоднозначностью подбора коэффициентов у коррелированных признаков (см. рис. 2). Для уменьшения количества коррелированных признаков использован метод главных компонент (Principal Component Analysis, PCA) [Jolliffe, Cadima, 2016], поэтому средние значения всех признаков были приведены к 0,

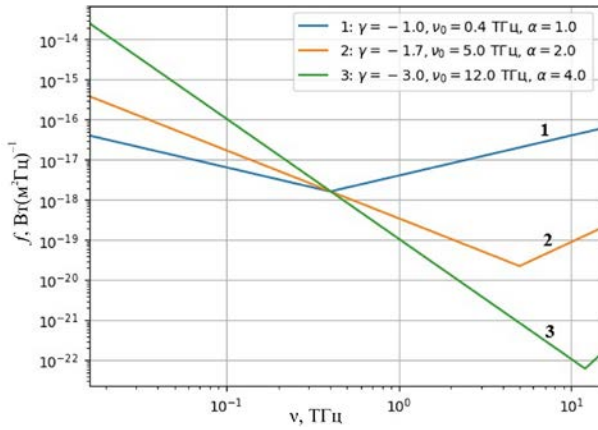


Рис. 1. Примеры рассматриваемых U-образных спектров радиоизлучения: излом в спектре на частоте 0.4, 5 и 12 ТГц

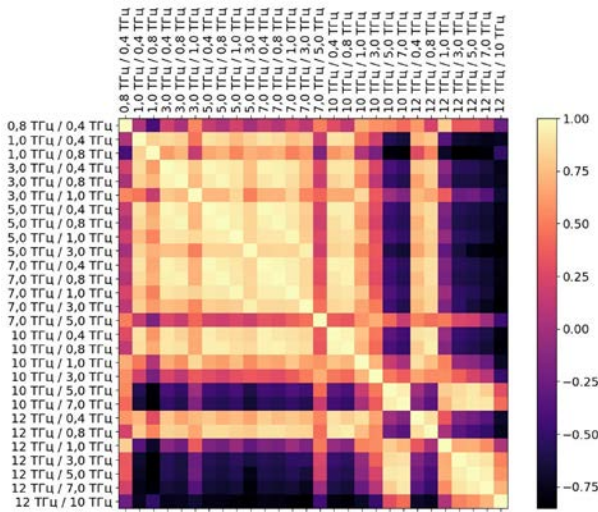


Рис. 2. Тепловая карта матрицы коэффициентов корреляции признаков друг с другом. Все признаки, у которых абсолютное значение коэффициента корреляции превосходит 0.5 (светлые и темные участки) считаются коррелированными. При обучении модели это может приводить к проблемам, связанным с неоднозначностью подбора коэффициентов у коррелированных признаков

а среднеквадратичные отклонения — к 1. Было взято 11 признаков, описывающих 99.9 % разброса данных. Средние значения всех признаков также были приведены к 0, а среднеквадратичные отклонения — к 1.

2. РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Для решения задачи регрессии применялось несколько алгоритмов: линейная регрессия (Linear Regression, LR) без регуляризации, метод k ближайших соседей (K-Nearest Neighbors, KNN) [Altman, 1992], метод опорных векторов (Support Vector Machine, SVM) [Cortes, Vapnik, 1995], решающее дерево (Decision Tree, DT) [Quinlan, 1986], случайный лес (Random Forest, RF) [Breiman, 2001], CatBoost [Prokhorenkova et al., 2018], XGBoost [Chen, Guestrin, 2016], LightGBM [Ke et al., 2017] и Stacking [Wolpert, 1992] с архитектурой, в которой случайный лес (RF) обучается на выводах методов KNN, RF, CatBoost, XGBoost и LightGBM. Гиперпараметры моделей, которые могут отличаться от стандартных

значений, представленных в библиотеке с открытым исходным кодом Scikit-learn [Pedregosa et al., 2011] для моделей LR, KNN, SVM, DT, RF и Stacking и в соответствующих библиотеках для моделей CatBoost, XGBoost и LightGBM, приведены в табл. 1. Принцип обучения метода Stacking заключается в том, что базовые модели (в нашем случае KNN, RF, CatBoost, XGBoost, LightGBM) обучаются на части тренировочной выборки. Их прогнозы на неиспользованной ранее части тренировочной выборки берутся в качестве мета-признаков для обучения мета-модели (в нашем случае RF), которая и выдает итоговый прогноз. При генерации прогнозов они сначала генерируются базовыми моделями, а затем на их основе делается итоговый прогноз мета-моделью. Полученные значения среднеквадратичного отклонения $RMSE$ и средней абсолютной относительной ошибки $MAPE$ на train- и test-выборках для γ , ν_0 , α показаны на рис. 3, 4 соответственно.

По полученным результатам на test-выборке видно, что с задачей лучшего всего справились алгоритмы KNN и Stacking. Однако равное нулю отклонение для метода KNN на train-выборке говорит о переобучении модели. Для дальнейшего улучшения точности были рассмотрены данные, на которых модели давали результаты с наибольшими отклонениями. Было обнаружено, что модели хуже всего различают данные спектров, у которых частота перегиба находится выше 8 ТГц. Этот факт согласуется с природой данных для обучения моделей, так как выше 8 ТГц имеется информация только с двух каналов, что затрудняет распознавание по отклику прибора положения точки перегиба вблизи правой границы изучаемого спектра.

Для выделения данных, обработка которых заведомо сложна для алгоритмов, была решена задача бинарной классификации (класс 1 — положение точки перегиба не выше 8 ТГц, класс 2 — положение точки перегиба выше 8 ТГц). Для ее решения применялось несколько алгоритмов: логистическая регрессия (Logistic Regression, LogR) [Hosmer et al., 2013] без регуляризации, KNN, DT и RF. Гиперпараметры моделей, которые могут отличаться от стандартных значений, представленных в библиотеке с открытым исходным кодом Scikit-learn для моделей LogR, KNN, DT и RF, приведены в табл. 2. Значения метрик accuracy и F1 на train- и test-выборках приведены в табл. 3.

Полученные результаты показывают, что все выбранные алгоритмы, за исключением LogR, практически безошибочно определяют спектры, для которых $\nu_0 \leq 8$ ТГц.

После выделения данных с $\nu_0 \leq 8$ ТГц размер train-выборки составил 5729 записей, а test-выборки — 1432 записи. К этому набору данных был также применен метод главных компонент [Jolliffe, Cadima, 2016], поэтому средние значения всех признаков были приведены к 0, а среднеквадратичные отклонения — к 1. Было оставлено 11 признаков, описывающих 99.9 % разброса данных. После этого средние значения всех признаков также были приведены к 0, а среднеквадратичные отклонения — к 1.

Таблица 1

Гиперпараметры моделей для решения задачи регрессии для различных параметров спектра

Модель	γ	ν_0	α
LR	–	–	–
KNN	n_neighbors = 2 weights = 'distance' metric = 'minkowski'	n_neighbors = 2 weights = 'distance' metric = 'minkowski'	n_neighbors = 2 weights = 'distance' metric = 'minkowski'
SVM	kernel = 'rbf'	kernel = 'rbf'	kernel = 'rbf'
DT	max_depth = 18 criterion = 'squared_error'	max_depth = 14 criterion = 'squared_error'	max_depth = 19 criterion = 'squared_error'
RF	max_depth = 36 n_estimators = 300 criterion = 'squared_error'	max_depth = 21 n_estimators = 250 criterion = 'squared_error'	max_depth = 46 n_estimators = 250 criterion = 'squared_error'
CatBoost	loss_function = 'RMSE'	loss_function = 'RMSE'	loss_function = 'RMSE'
XGBoost	objective = 'reg:squarederror'	objective = 'reg:squarederror'	objective = 'reg:squarederror'
LightGBM	boosting_type = 'gbdt' n_estimators = 100 num_leaves = 31	boosting_type = 'gbdt' n_estimators = 100 num_leaves = 31	boosting_type = 'gbdt' n_estimators = 100 num_leaves = 31
Stacking	гиперпараметры KNN, RF, CatBoost, XGBoost, LightGBM указаны выше; гиперпараметры мета-модели RF: max_depth = 35 n_estimators = 100	гиперпараметры KNN, RF, CatBoost, XGBoost, LightGBM указаны выше; гиперпараметры мета-модели RF: max_depth = 20 n_estimators = 100 criterion = 'squared_error'	гиперпараметры KNN, RF, CatBoost, XGBoost, LightGBM указаны выше; гиперпараметры мета-модели RF: max_depth = 36 n_estimators = 100

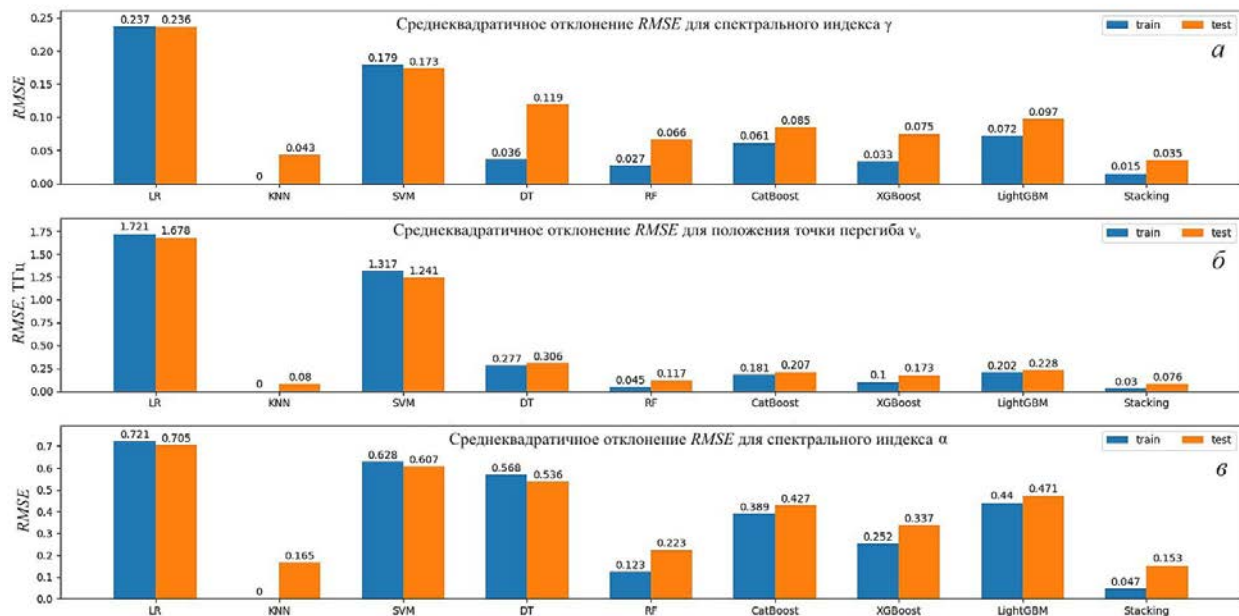


Рис. 3. Среднеквадратичное отклонение *RMSE* на train- и test-выборках для спектрального индекса γ (а), положения точки (частоты) перегиба ν_0 (б) и спектрального индекса α (в)

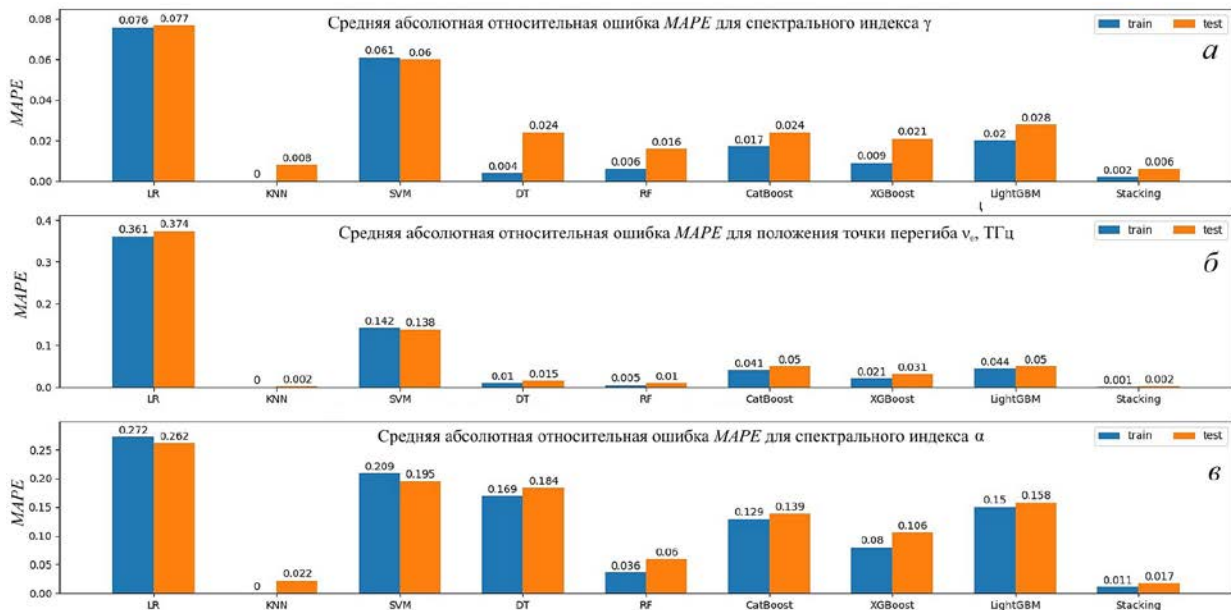


Рис. 4. Средняя абсолютная относительная ошибка MAPE на train- и test-выборках для спектрального индекса γ (а), положения точки (частоты) перегиба ν_0 (б) и спектрального индекса α (в)

Таблица 2

Гиперпараметры моделей для решения задачи классификации по положению точки перегиба

Модель	Гиперпараметры
LogR	—
KNN	n_neighbors = 2 weights = 'distance' metric = 'minkowski'
DT	max_depth = 17 criterion = 'gini'
RF	max_depth = 31 n_estimators = 100 criterion = 'gini'

Таблица 3

Значения метрик accuracy и F1 на train- и test-выборках

Модель	Метрика	train	test
LogR	accuracy	0.809	0.830
	F1	0.589	0.626
KNN	accuracy	0.999	0.998
	F1	0.997	0.994
DT	accuracy	1	0.994
	F1	1	0.989
RF	accuracy	1	0.997
	F1	1	0.995

Для решения задачи регрессии на новом наборе данных применялось также несколько алгоритмов: LR без регуляризации, KNN, SVM, DT, RF, CatBoost, XGBoost, LightGBM и Stacking. Гиперпараметры моделей, которые могут отличаться от стандартных значений, представленных в библиотеке с открытым исходным кодом Scikit-learn для моделей LR, KNN, SVM, DT, RF и Stacking и в соответствующих библиотеках для моделей CatBoost, XGBoost и LightGBM, приведены в табл. 1. Полученные значения $RMSE$ и $MAPE$ на train- и test-выборках для γ , ν_0 , α показаны на рис. 5 и 6 соответственно.

По полученным результатам на test-выборке видно, что с задачей лучшего всего справились алгоритмы KNN и Stacking. Однако равное нулю отклонение для метода KNN на train-выборке говорит о переобучении модели. Отделение данных о спектрах с $\nu_0 > 8$ ТГц улучшило результаты методов LR, SVM, DT, CatBoost, XGBoost и LightGBM по определению ν_0 и α и результаты RF по определению α . В остальных случаях такое отделение к улучшению результатов не привело. Таким образом, алгоритм обработки данных по U-образным всплескам может состоять из следующих шагов.

1. Применение к экспериментальным данным преобразования, которое для всех модельных данных приводит средние значения всех признаков к 0, а среднеквадратичные отклонения — к 1.
2. Применение к результирующим данным преобразования, полученного методом главных компонент (PCA) для всех модельных данных.
3. Применение к полученным данным преобразования, которое приводит средние значения всех признаков, полученных после PCA для всех модельных данных, к 0, а среднеквадратичные отклонения — к 1.
4. Обработка полученных данных заранее обученными алгоритмами Stacking для получения γ ($RMSE=0.035$), ν_0 (0.076 ТГц) и α (0.153).

3. ВЫВОДЫ

В рамках данной работы разработана методика на основе алгоритмов машинного обучения для восстановления параметров U-образных спектров солнечных вспышек в терагерцовом диапазоне (0.4–12 ТГц) по данным, полученным при моделировании работы оптической системы научной аппаратуры «Солнце-Терагерц». Основное внимание уделено определению спектральных индексов γ спадающей части и α растущей части спектра, а также положения точки (частоты) перегиба ν_0 .

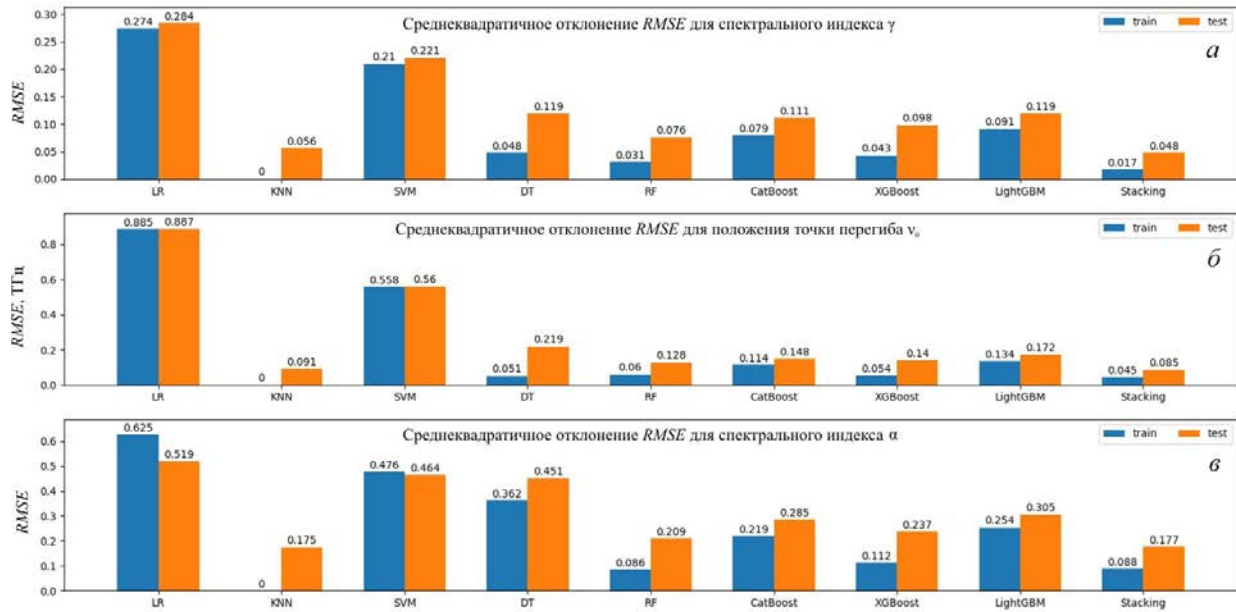


Рис. 5. То же, что на рис. 3. Данные с $\nu_0 \leq 8$ ТГц

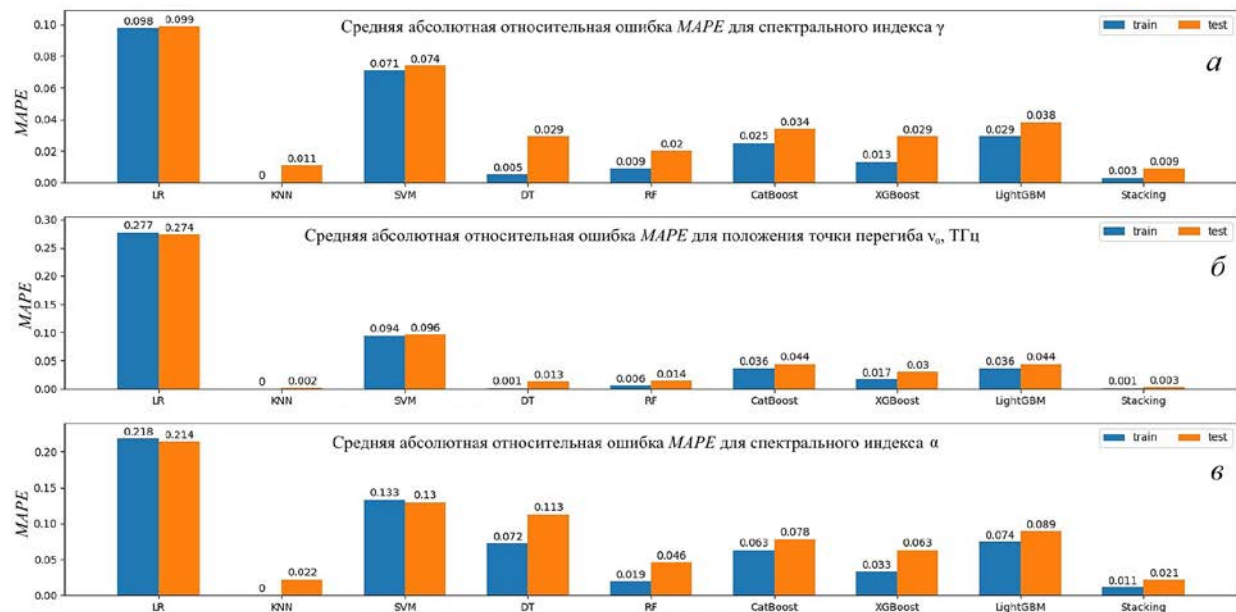


Рис. 6. То же, что на рис. 3. Данные с $\nu_0 \leq 8$ ТГц

Проведенный анализ показал, что алгоритм Stacking демонстрирует наилучшую точность при определении параметров спектра. Он показал наименьшие отклонения в определяемых параметрах при использовании спектров, для которых точка перегиба находится и выше, и ниже 8 ТГц. Это указывает на то, что алгоритм способен определить параметры даже слабо различающихся спектров, для которых точка перегиба находится выше 8 ТГц.

Разработанный алгоритм может быть интегрирован в систему обработки данных будущего эксперимента на МКС, обеспечивая автоматическое предварительное восстановление параметров спектров солнечных вспышек. Следует отметить, что для слабых вспышек значительным оказывается влияние шума измеряемого сигнала на форму регистри-

руемого спектра. В дальнейшем с целью улучшения интерпретации данных необходимо рассмотреть влияние шума, присутствующего в экспериментальных данных, на точность результатов работы алгоритма. Необходимо также рассмотреть определение параметров спектров других форм: монотонно растущего и монотонно убывающего с увеличением частоты излучения.

СПИСОК ЛИТЕРАТУРЫ

Altman N.S. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*. 1992, vol. 46, no. 3, pp. 175–185.
 Breiman L. Random forests. *Machine Learning*. 2001, vol. 45, pp. 5–32.

- Chen T., Guestrin C. Xgboost: A scalable tree boosting system. *Proc. of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, 2016, pp. 785–794. DOI: [10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785).
- Cortes C., Vapnik V. Support-vector networks. *Machine Learning*. 1995, vol. 20, pp. 273–297.
- Cristiani G., De Castro C.G.G., Luoni M.L., et al. Observed flux density enhancement at submillimeter wavelengths during an X-class flare. *Adv. Space Res.* 2007, vol. 39, iss. 9, pp. 1445–1450. DOI: [10.1016/j.asr.2007.02.039](https://doi.org/10.1016/j.asr.2007.02.039).
- Cristiani G., De Castro C.G.G., Mandrini C.H., et al. Asymmetric precipitation in a coronal loop as explanation of a singular observed spectrum. *Adv. Space Res.* 2009, vol. 44, iss. 11, pp. 1314–1320. DOI: [10.1016/j.asr.2009.06.012](https://doi.org/10.1016/j.asr.2009.06.012).
- De Castro C.G.G., Kaufmann P., Raulin J.P. Recent results on solar activity at submillimeter wavelengths. *Adv. Space Res.* 2005, vol. 35, iss. 10, pp. 1769–1773. DOI: [10.1016/j.asr.2005.03.083](https://doi.org/10.1016/j.asr.2005.03.083).
- Duncan N., Saint-Hilaire P., Shih A.Y., et al. First flight of the Gamma-Ray Imager/Polarimeter for Solar flares (GRIPS) instrument. *Space Telescopes and Instrumentation 2016: Ultraviolet to Gamma Ray*. Edinburgh, 2016, vol. 9905, p. 876. DOI: [10.1117/12.2233859](https://doi.org/10.1117/12.2233859).
- Hosmer Jr D.W., Lemeshow S., Sturdivant R.X. *Applied logistic regression*. John Wiley & Sons, Inc., 2013. 510 p.
- Jolliffe I.T., Cadima J. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 2016, vol. 374:20150202, iss. 2065. DOI: [10.1098/rsta.2015.0202](https://doi.org/10.1098/rsta.2015.0202).
- Kalinin E.V., Philippov M.V., Makhmutov V.S., et al. A study of the characteristics of a terahertz radiation detector for the Solntse-Terahertz scientific apparatus. *Cosmic Res.* 2021, vol. 59, iss. 1, pp. 1–5. DOI: [10.1134/S0010952521010032](https://doi.org/10.1134/S0010952521010032).
- Kaufmann P., Raulin J.P., De Castro C.G.G., et al. A new solar burst spectral component emitting only in the terahertz range. *ApJ. Lett.* 2004, vol. 603, iss. 2, pp. L121–L124. DOI: [10.1086/383186](https://doi.org/10.1086/383186).
- Kaufmann P., Abrantes A., Bortolucci E.C., et al. THz solar observations on board of a trans-Antarctic stratospheric balloon flight. *41st International Conference on Infrared, Millimeter, and Terahertz waves (IRMMW-THz)*. Copenhagen, 2016, p. 1. DOI: [10.1109/IRMMW-THz.2016.7758395](https://doi.org/10.1109/IRMMW-THz.2016.7758395).
- Ke G., Meng Q., Finley T., et al. Lightgbm: A highly efficient gradient boosting decision tree. *31st Conference on Neural Information Processing Systems (NIPS–2017)*, Long Beach, 2017, vol. 30.
- Krucker S., De Castro C.G.G., Hudson H.S., et al. Solar flares at submillimeter wavelengths. *Astron. Astrophys. Rev.* 2013, vol. 21, iss. 1, pp. 1–45. DOI: [10.1007/s00159-013-0058-3](https://doi.org/10.1007/s00159-013-0058-3).
- Kvashnin A.A., Logachev V.I., Philippov M.V., et al. Optical system design of the detector for solar terahertz emission measurements. *Space Engineering and Technology*. 2021, vol. 35, iss. 4, pp. 22–30. DOI: [10.33950/spacetech-2308-7625-2021-4-22-30](https://doi.org/10.33950/spacetech-2308-7625-2021-4-22-30).
- Pedregosa F., Varoquaux G., Gramfort A., et al. Scikit-learn: Machine learning in Python. *J. Machine Learning Res.* 2011, vol. 12, pp. 2825–2830. DOI: [10.5555/1953048.2078195](https://doi.org/10.5555/1953048.2078195).
- Philippov M.V., Makhmutov V.S., Razumeyko M.V. Scientific equipment for the Sun-Terahertz space experiment: study of the temperature effect in the Golay cell. *Measurement Techniques*. 2024a, vol. 67, iss. 3, pp. 195–202. DOI: [10.1007/s11018-024-02335-9](https://doi.org/10.1007/s11018-024-02335-9).
- Philippov M.V., Makhmutov V.S., Maksumov O.S., et al. Electronics Unit for “Sun-Terahertz” Scientific Equipment. *Instrum. Exp Tech.* 2024b, vol. 67, iss. 3, pp. 545–553. DOI: [10.1134/S0020441224700829](https://doi.org/10.1134/S0020441224700829).
- Prokhorenkova L., Gusev G., Vorobev A., et al. CatBoost: unbiased boosting with categorical features. *32nd Conference on Neural Information Processing Systems (NeurIPS 2018)*. Montréal, 2018, vol. 31.
- Quinlan J.R. Induction of decision trees. *Machine Learning*. 1986, vol. 1, pp. 81–106.
- Tsap Y.T., Smirnova V.V., Morgachev, A.S., et al. On the origin of 140 GHz emission from the 4 July 2012 solar flare. *Adv. Space Res.* 2016, vol. 57, iss. 7, pp. 1449–1455. DOI: [10.1016/j.asr.2015.12.037](https://doi.org/10.1016/j.asr.2015.12.037).
- Tulnikov E.D., Logachev V.I., Makhmutov V.S., et al. Characteristics of the optical system of the Solntse-Terahertz scientific equipment. *Cosmic Res.* 2024, vol. 62, iss. 6, pp. 551–557. DOI: [10.1134/S0010952524600434](https://doi.org/10.1134/S0010952524600434).
- Tulnikov E.D., Makhmutov V.S., Philippov M.V. Review of studying submillimeter radiation from the Sun and astrophysical sources. *Bull. Russian Academy of Sciences: Physics*. 2025, vol. 89, iss. 6, pp. 854–857. DOI: [10.1134/S1062873825711298](https://doi.org/10.1134/S1062873825711298).
- Wedemeyer S., Bastian T., Brajša R., et al. Solar science with the Atacama Large Millimeter/Submillimeter Array — a new view of our Sun. *Space Sci. Rev.* 2016, vol. 200, pp. 1–73. DOI: [10.1007/s11214-015-0229-9](https://doi.org/10.1007/s11214-015-0229-9).
- Wolpert D.H. Stacked generalization. *Neural Networks*. 1992, vol. 5, iss. 2, pp. 241–259.
- Как цитировать эту статью:*
Тулников Е.Д., Махмутов В.С., Филиппов М.В. Применение методов машинного обучения для определения спектральных характеристик излучения в эксперименте «Солнце–Терагерц». *Солнечно-земная физика*. 2026, т. 12, № 1, с. 14–20. <https://doi.org/10.12737/szf-121202602>.