

СОЗДАНИЕ ИНТЕГРАЛЬНОЙ ОБУЧАЮЩЕЙ БАЗЫ ДЛЯ КЛАССИФИКАЦИИ ВСПЫШЕК МЕТОДОМ МАШИННОГО ОБУЧЕНИЯ

Ожередов В.А.¹, Струминский А.Б.¹, Григорьева И.Ю.²

¹Институт космических исследований РАН, Москва, Россия

²Главная (Пулковская) астрономическая обсерватория РАН, Санкт-Петербург, Россия

CREATING OF THE INTEGRATIVE TRAINING DATABASE FOR THE MACHINE LEARNING-BASED SXR-FLARES CLASSIFICATION METHOD

Ozheredov V.A.¹, Struminsky A.B.¹, Grigoryeva I.Yu.²

¹Space Research Institute of RAS, Moscow, Russia

²Main Astronomical (Pulkovo) Observatory RAS, St.-Petersburg, Russia

Solar proton event (SPE) is a process of particle acceleration in the source and the propagation to the observer. This process is not understandable in details by now: its “pure” physical model allowing the SPE prediction is absent. Recently we developed a prediction algorithm as a hybrid of physical concept and machine learning method called Bayesian inference. When applied to the famous catalog of solar proton events (SWX), this method revealed “misses” on the several events within it and “hits” on the other events beyond it. So, we established a task to build up a hybrid training database constructed from SWX catalog and time series originated from SRG monitor in L₂ and Electron-Proton-Helium Instrument (EPHIN).

<https://doi.org/10.31725/0552-5829-2024-241-244>

Основанный на гибридизации физических идей из статьи [1] и Байесовского вывода алгоритм классификации вспышек как источников солнечных протонов показал удовлетворительные результаты на выборке [2], правильно определив все указанные там за период от 1996-го по 2017-й год события как SPE-источники (SPE = solar proton events) и дав примерно 1% ложных срабатываний на почти 30 000 «обычных» вспышек за тот же период. В рамках настоящей работы мы пытаемся создать новую базу, охватывающую период с 2017-го года по настоящее время для тестирования старых и создания новых версий методов классификации вспышек как SPE-источников.

На настоящий момент из открытых данных с источника [3] собрано 12407 вспышек за период с сентября 2017 года по октябрь 2024 года. Информация о каждой из них организована согласно схеме на рис. 1. Параметры вспышки делятся на признаки, то есть те, на основе которых будет производиться классификация новых вспышек, и отклики, то есть характеристики, по которым вычисляется мера уверенности эксперта в принадлежности вспышки к SPE-источникам. Признаки обозначены синим цветом, отклики – красным.

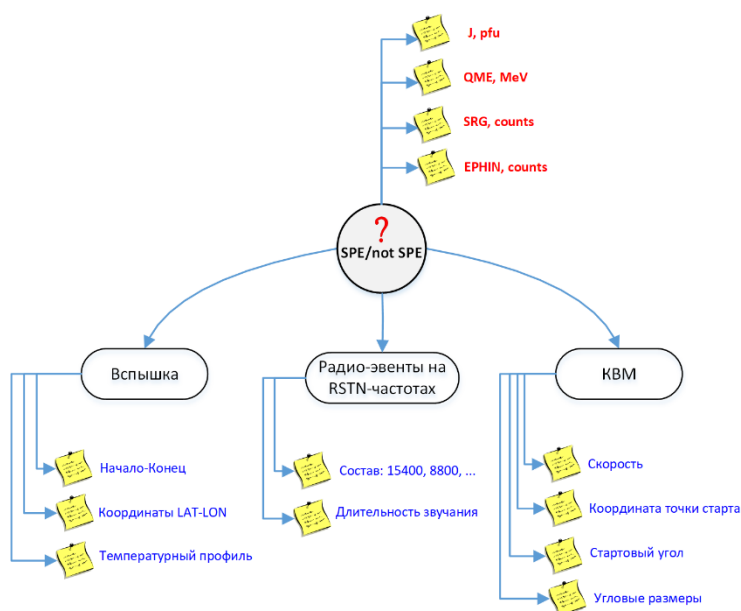


Рис. 1. Схема хранения информации о событиях в обучающей базе данных.

состав и продолжительность радио-событий внутри горячей фазы вспышки (т.е. интервала, на котором температура не опускается ниже 10 MK). Важное значение для распространения протонов в межпланетном пространстве к Земле имеет местоположение вспышки на солнечной сфере и обладающий достаточно удачной геометрией и высокой скоростью корональный выброс массы (КВМ).

Перейдем к предварительным результатам классификации собранных вспышек. Важно отметить, что градация SPE/not SPE не является «черно-белой», и потому классификация вспышек не похожа, например, на задачу автоматического распознавания велосипедов на дороге, в которой транспортное средство не может быть «немножко велосипедом».

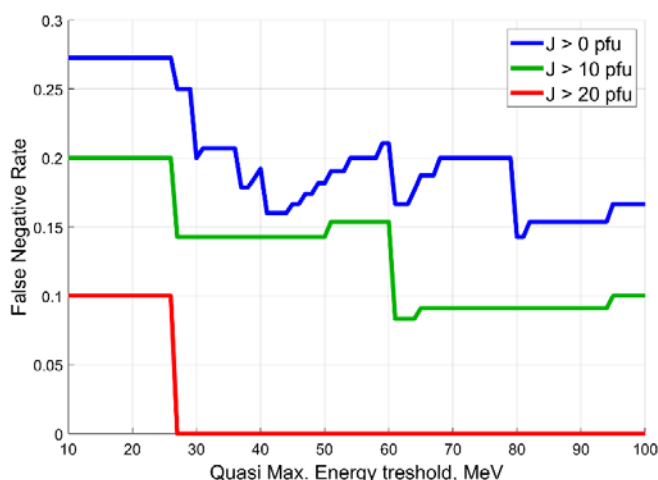


Рис. 2. Семейство зависимостей вероятности пропуска цели от QME протонного события.

Согласно идеям статьи [1], ускорение протонов происходит в достаточно хорошо прогретой плазме. Работа ускорителя сопровождается появлением радиоизлучения на гигагерцовых RSTN-частотах (2695 – 15400 МГц), тогда как мегагерцовые частоты и излучение II-IV типов являются индикатором сопутствующих эруптивных процессов. Поэтому для прогноза SPE имеет значение температурный профиль вспышки, а также со-

Входящее в любой каталог протонное событие имеет ряд «показателей протонности», характеризующих непрерывно распределенную от 0 до 1 меру уверенности эксперта в том, что данным событием не стоит пренебрегать в смысле его потенциальной опасности для наземных и космических устройств и пилотируемых полетов. В каталоге SWX [4] таких показателей два – интенсивность потока первого макси-

муму J и квазимаксимальная энергия (quasi-maximal energy) QME. Мы пытаемся дополнить этот ряд показателей профилями с монитора радиационной обстановки в точке Лагранжа L_2 [5] и с инструмента EPHIN [6].

Вышесказанное иллюстрирует рис. 2, на котором изображен пучок зависимостей ошибки первого рода (пропуск цели) по каталогу SWX от квазимаксимальной энергии (QME) протонного события для трех вариантов пороговой фильтрации событий по интенсивности максимального потока J .

Видно, что при отсутствии фильтра на поток и квази-максимальную энергию события вероятность пропуска цели составляет примерно 27%, следовательно, по точность распознавания при таких условиях алгоритм принятия решений слабо отличим от случайного выбора при подбрасывании монетки. Далее, при полном отсутствии фильтра на интенсивность потока ошибка с большими флуктуациями падает примерно до 15%. При введении стандартной отсечки на интенсивность потока $J = 10$ pfu (общепринято полагать, что протонные события с меньшей интенсивностью максимального потока не способны вызвать сколь-нибудь значимые последствия своего появления, следовательно – не особо интересны) ошибка снижается уже гораздо более монотонно до значений в 10% на 100 MeV, а при отсечке в 20 pfu ошибка уже на 28 MeV падает до нуля.

Самими благодатными с точки зрения *SPE case study* месяцами в текущем году выдались май и июнь. На рис. 3 показаны наложенные друг на друга временные ряды, полученные от инструментов Spectr-RG (серая кривая) и EPHIN (фиолетовая кривая). Жирная черная линия внизу обозначает глобальный тренд, связанный с модуляцией галактических космических лучей (ГКЛ), а тонкая черная линия соответствует отсечке, связанной с модуляцией ГКЛ в 25-цикле солнечной активности.

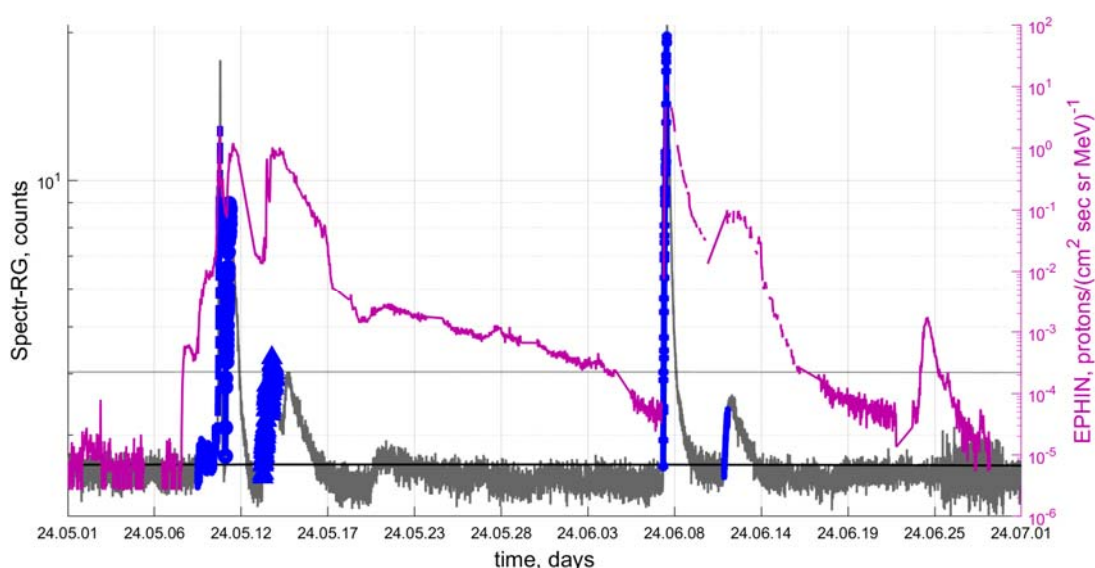


Рис. 3. Наложение временных рядов с инструментов Spectr-RG и EPHIN (канал 25–53 MeV) за период мая–июня 2024-го года.

Синим цветом заштрихованы участки кривой Spectr-RG, отвечающие пяти протонным всплескам, занесенным в каталог SWX. Видно, что величина всплесков соответствует вышеозначенной модуляции ГКЛ. На текущий момент наш алгоритм «ловит» не только все 5 из пяти протонных событий, занесенных в каталог SWX, включая залимбовое событие 11-Jun-2024, вспышка X3 каталога STIX [7], но и 20-May-2024, вспышка X12 каталога STIX [7], не включённая в каталог SWX (протонное событие было зарегистрировано даже на поверхности Марса!). Однако, визуальное регистрируемое на мониторе EPHIN протонное возрастание от 24-Jun-2024 не «ловится» нашим алгоритмом и не фиксируется протонным монитором GOES по причине его заведомо низкой квазимаксимальной энергией ($QME < 53\text{MeV}$).

- Принадлежность вспышки к протонному источнику является нечетким событием, характеризующимся индикаторной функцией со значениями от 0 (несколько не протонное событие) до 1 (определенно опасное протонное событие). Индикаторная функция зависит от интенсивности потока в максимуме, квази-максимальной энергии и силы реакции на событие таких инструментов, как Goes Proton Monitor, Spectr-RG и EPHIN.

- В связи с этим, наряду с достоверно протонными событиями их каталоги протонных событий содержат слабые и откровенно неинтересные экземпляры. В то же самое время, некоторые вспышки, появление которых отзывается в достаточно значимых всплесках на EPHIN и Spectr-RG, в каталогах часто отсутствуют.

- Текущая версия алгоритма классификации вспышек на основе Байесовского вывода работает достаточно неплохо.

Литература

1. *Struminsky A., Sadovsky A., Grigorieva I.* Criteria for Forecasting Proton Events by Real-Time Solar Observations // *Geomagnetism and Aeronomy*, 64(2), pp. 139-149. DOI:10.1134/S0016793223600984.
2. <https://umbra.nascom.nasa.gov/SEP/>
3. <https://aj4co.org/AJ4CODataArchive/NOAA%20SWPC%20Event%20Logs/>
4. https://swx.sinp.msu.ru/apps/sep_events_cat/
5. <https://monitor.srg.cosmos.ru/dates/>
6. <http://www2.physik.uni-kiel.de/SOHO/phpeph/EPHIN.htm>
7. <https://datacenter.stix.i4ds.net/view/flares/list>