



Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)



Институт космофизических исследований и распространения радиоволн Дальневосточного отделения Российской академии наук

Автоматизированный метод анализа данных космических лучей и выделения спорадических эффектов

Профессор В.В. Геппенер, Б.С. Мандрикова

Цель исследования

Разработка *автоматизированных методов анализа данных сети нейтронных мониторов (космических лучей) и обнаружения периодов возникновения спорадических эффектов.*

Прикладная задача

Изучение данных космических лучей представляет интерес в задачах солнечно-земной физики и прикладных исследованиях, связанных с космической погодой.

Гелиогеомагнитная активность воздействует практически на все объекты техносферы: электрические сети, трубопроводы нефтегазового комплекса и водоводы, системы теле-, радио и спутниковой связи и космического позиционирования GPS и ГЛОНАСС. А так же угрожает жизни и здоровью людей.

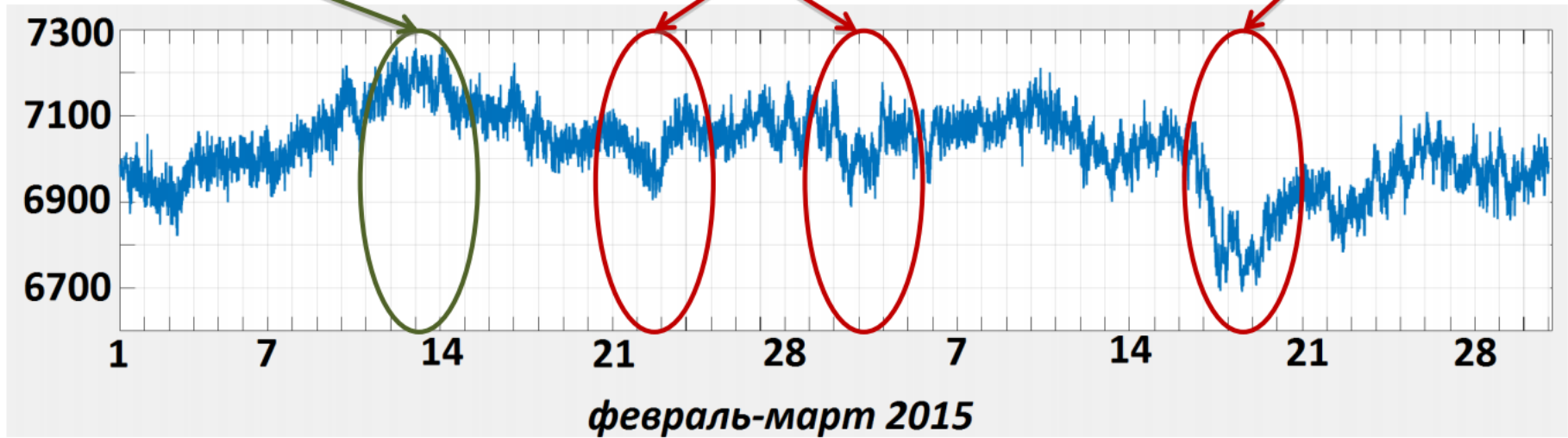


Структура данных космических лучей

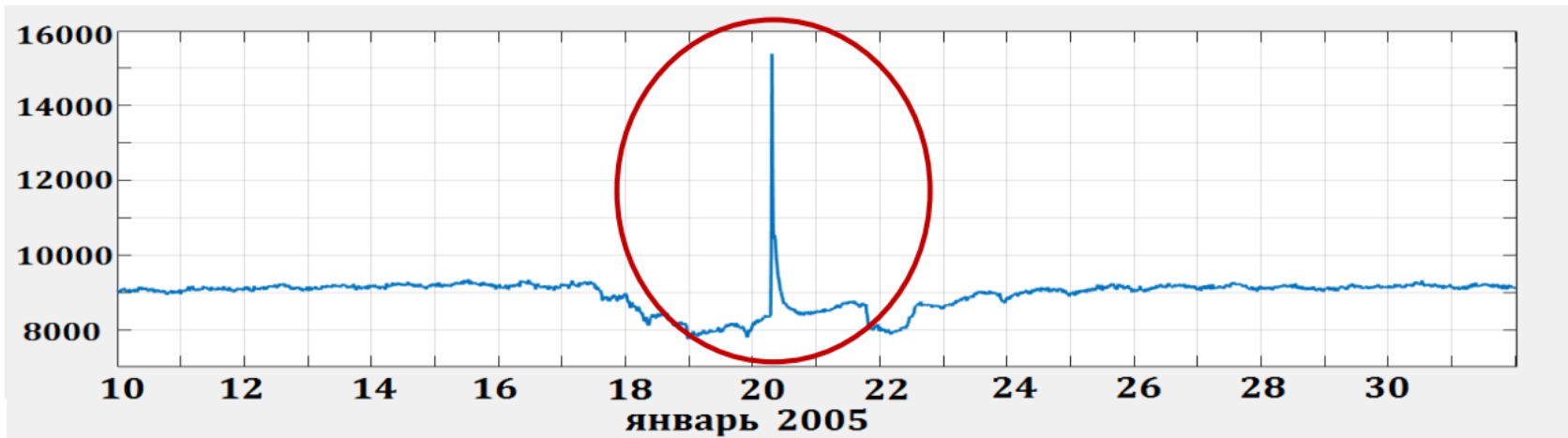
Рекуррентные
вариации

Мелкие спорадические
эффекты

Крупные спорадические
эффекты



Крупные GLE-события



1. Метод анализа данных космических лучей

1.1.Использование вейвлет-преобразования: позволяет детально изучить структуру данных, выделить информативные компоненты и подавить шум

Представление данных космических лучей, основанное на кратномасштабных вейвлет-разложениях

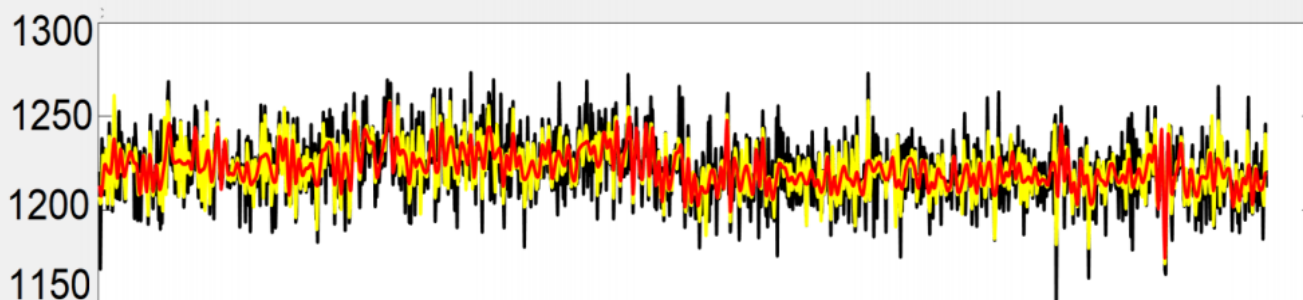
детализирующие
компоненты

сглаженная
компонента

$$f_j(t) = \underbrace{g_{j-1}(t) + g_{j-2}(t) + \dots + g_{j-m}(t)}_{\text{детализирующие компоненты}} + \underbrace{f_{j-m}(t)}_{\text{сглаженная компонента}} = \sum_{k=j-1}^{j-m} \sum_n d_{k,n} \Psi_{k,n}(t) + \sum_n c_{j-m,n} \varphi_{j-m,n}(t)$$

где $\bar{c}^j = \{c_{j,n}\}^{n \in \mathbb{Z}} \in V_j$ и $\bar{d}^j = \{d_{j,n}\}^{n \in \mathbb{Z}} \in W_j$: $c_{j,n} = \langle f, \varphi_{j,n} \rangle$, $d_{j,n} = \langle f, \Psi_{j,n} \rangle$

Применение операций вейвлет-разложения к данным космических лучей



3.08.2019

- исходные данные;
- 1 уровень разложения $m=1$;
- 2 уровень разложения $m=2$;

Выбор аппроксимирующего вейвлета

1. **Большое число нулевых моментов** s вейвлета Ψ (характеризует способность вейвлета выделять особенности высокого порядка): $\int_{-\infty}^{+\infty} t^k \Psi(t) dt = 0, k = \overline{0, s-1}$

2. **Малый носитель вейвлета** q : размер окрестности \bar{h} на масштабе j , содержащей краевой эффект, определяется по формуле:

$$h_j = 2^j * q$$

3. **Высокий порядок гладкости** вейвлета (позволяет выделять особенности высокого порядка).

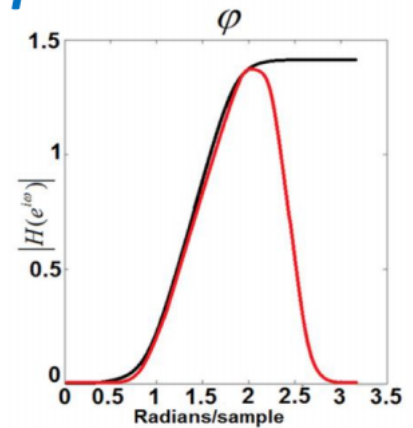
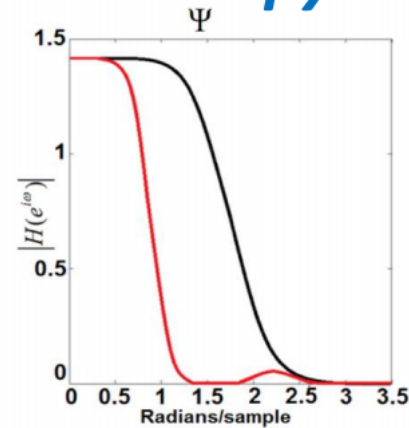
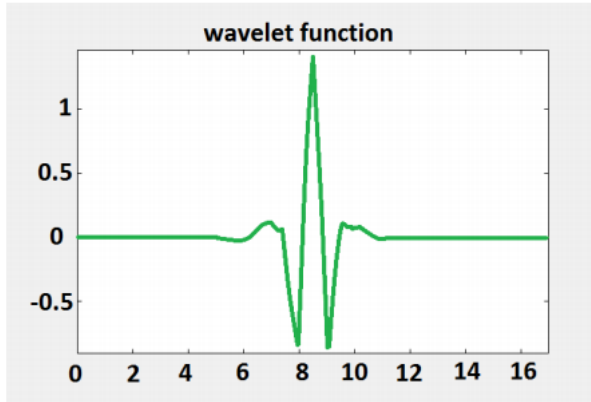
Для решения задачи определены семейства ортогональных вейвлетов Добеши и Койфлеты.

• Семейство Добеши – это единственное семейство ортогональных вейвлетов, имеющих минимальный носитель при заданном числе нулевых моментов;

• Семейство Койфлеты - единственное семейство ортогональных вейвлетов, которые имеют носитель наименьшего размера при достаточном числе нулевых моментов в масштабирующей функции φ . Если $f \in C^r$ в окрестности $2^{-m}n$ с $r \leq s$ тогда:

$$2^{-m/2} \langle f, \varphi_{-m,n} \rangle \approx f(2^{-m}n) + O(2^{-m(r+1)})$$

Используемые вейвлет-функции



Алгоритм выбора «наилучшего» базиса для уровня разложения m может быть основан на следующих операциях:

Минимаксный риск : $r(\Theta) = \inf_D \sup_{y \in \Theta} E \left\{ \|F - Df\|^2 \right\}$, где $\tilde{F} = Df$. Оценка \tilde{F} в базисе β^λ :

$$\tilde{F}^\lambda = \sum_n P_T \left(\langle f, \beta_n^\lambda \rangle \right) \beta_n^\lambda, \text{ тогда наилучший базис } \beta^\alpha : E \left\{ \|F - \tilde{F}^\alpha\|^2 \right\} = \min_{\lambda \in \Lambda} E \left\{ \|F - \tilde{F}^\lambda\|^2 \right\}$$

Шаг 1. Выполняем разложение функции f до уровня m :

$$f_j(t) = g_{j-1}(t) + g_{j-2}(t) + \dots + g_{j-m}(t) + f_{j-m}(t) = \sum_{k=j-1}^{j-m} \sum_n d_{k,n} \Psi_{k,n}(t) + \sum_n c_{j-m,n} \varphi_{j-m,n}(t)$$

Шаг 2. Путем применения пороговых функций P_{T_j} получаем оценку:

$$\tilde{F}^m = \sum_{k=j-1, j-m} \sum_n P_{T_j}(d_{k,n}) \Psi_{k,n} + \sum_n P_{T_{j-m}}(c_{j-m,n}) \varphi_{j-m,n}, \text{ где } T_j = \text{Med}(\langle f, \beta^\lambda \rangle), \beta^\lambda = \{ \Psi_{k,n}, \varphi_{j-m,n} \}_{k=j-1, j-m}$$

Med – медиана

Шаг 3. Оцениваем величину $Q_m^\lambda = \sum_{n \in I^j} |c_{j-m,n}|^2 + \sum_{k=j-1, j-m} \sum_{n \in I^j} |d_{k,n}|^2$, где множество индексов I^j :

$n \in I^j$, если $|\langle f, \beta^\lambda \rangle| \geq T_j$ и определяем «наилучший» базис $\beta_m^\alpha : Q_m^\alpha = \max_{\lambda \in \Lambda} E \{ Q_m^\lambda \}$

Алгоритм предобработки данных космических лучей на основе КМА

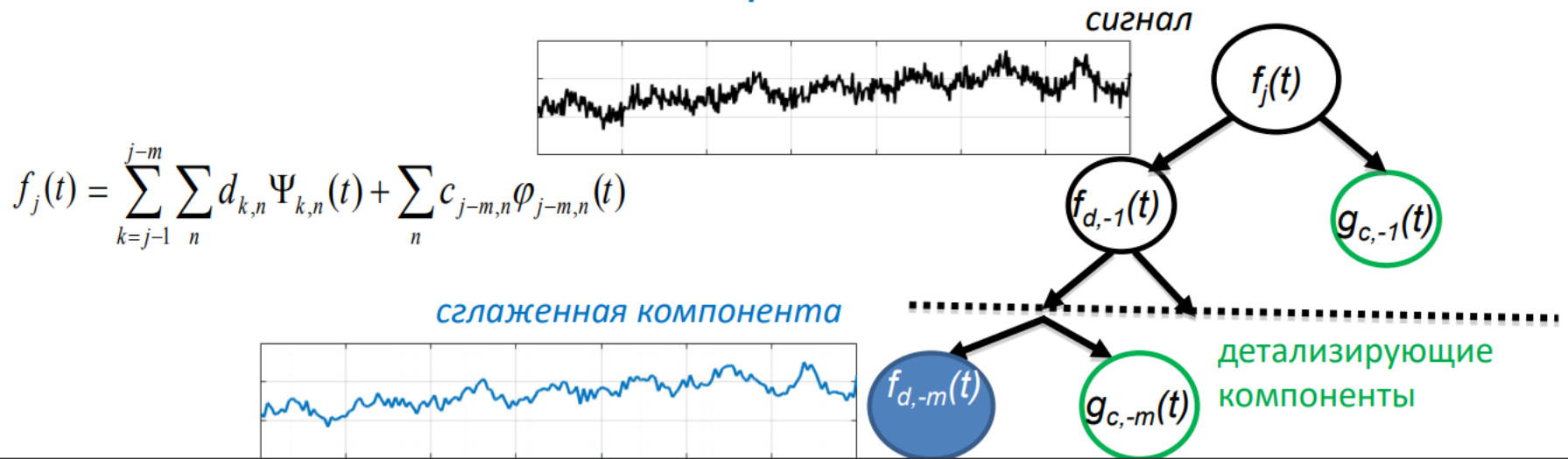
Шаг 1: на основе кратномасштабного вейвлет-разложения функции до уровня m получаем сглаженную компоненту:
$$f_{-m}(t) = \sum_n c_{-m,n} \varphi_{-m,n}(t)$$

Шаг 2: для восстановления исходного разрешения функции выполняем операцию вейвлет-восстановления:
$$f_0^{(-m)}(t) = \sum_n c_{0,n}^{(-m)} \varphi_{0,n}(t)$$

Шаг 3: для исключения влияния краевого эффекта:

- У восстановленного набора данных удаляем первые и последние $2^m * q$ отсчетов.
- Для оперативного режима выполняем зеркальное отображение на величину краевого эффекта.

Представление данных космических лучей, основанное на кратномасштабных вейвлет-разложениях



2. Построение структуры нейронной сети на основе парадигмы сети LVQ

2.1. Использование нейронных сетей: позволяет выявить сложные априори неизвестные связи и зависимости, и легко реализуемы в автоматическом режиме.

1. Классы нейронной сети:

«Спокойный» класс – отсутствие спорадических эффектов, (А-индекс < 7, К-индекс < 3, Dst-индекс в пределах ± 4);

«Слабовозмущенный» класс – наличие спорадических эффектов малой амплитуды, (А-индекс < 18, К-индекс < 5, Dst-индекс в пределах ± 8);

«Возмущенный» класс – наличие спорадических эффектов большой амплитуды, (А-индекс > 18, К-индекс > 5, Dst-индекс за пределами ± 8).

2. Число кластеров сети: $m=20$ – определено эмпирическим путем.

- Первый слой сети – слой Кохонена: $d_k = d(X, W_k) = \|X - W_k\| = \sqrt{\sum_{i=1}^I (x_i - w_{ik})^2}$
- Второй слой сети линейный: $F_l = \sum_k w_{kl} y_k$



Результаты работы сети LVQ

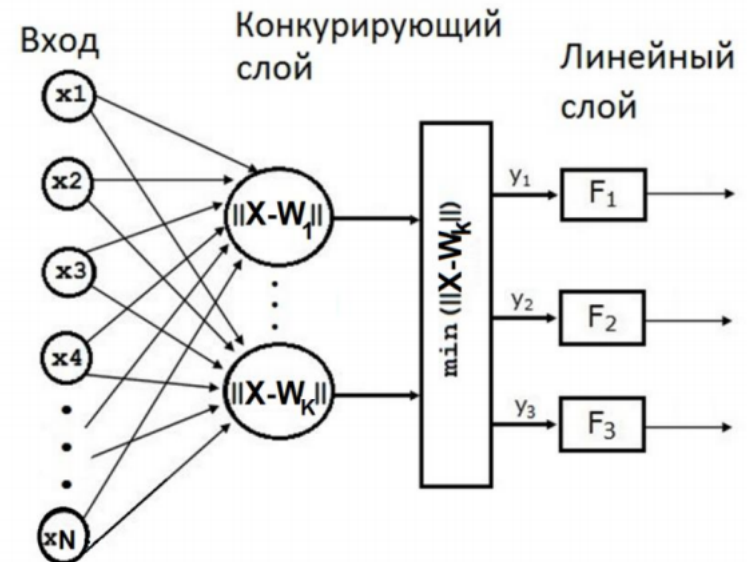
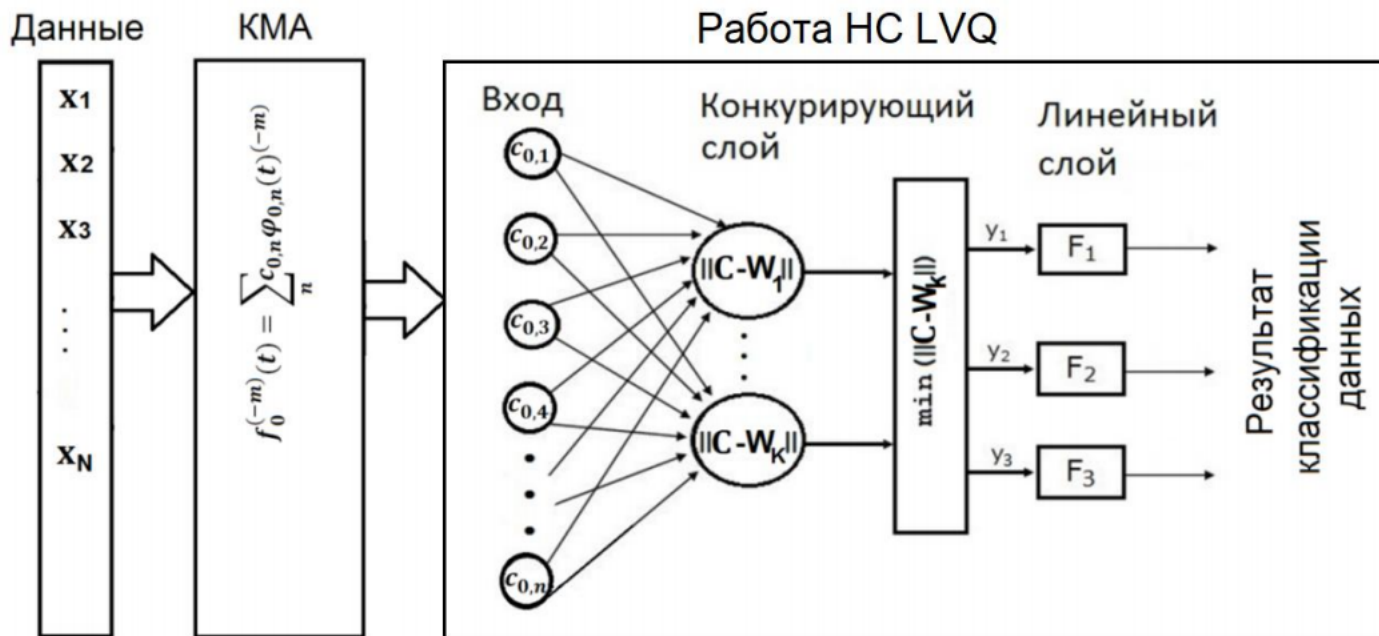


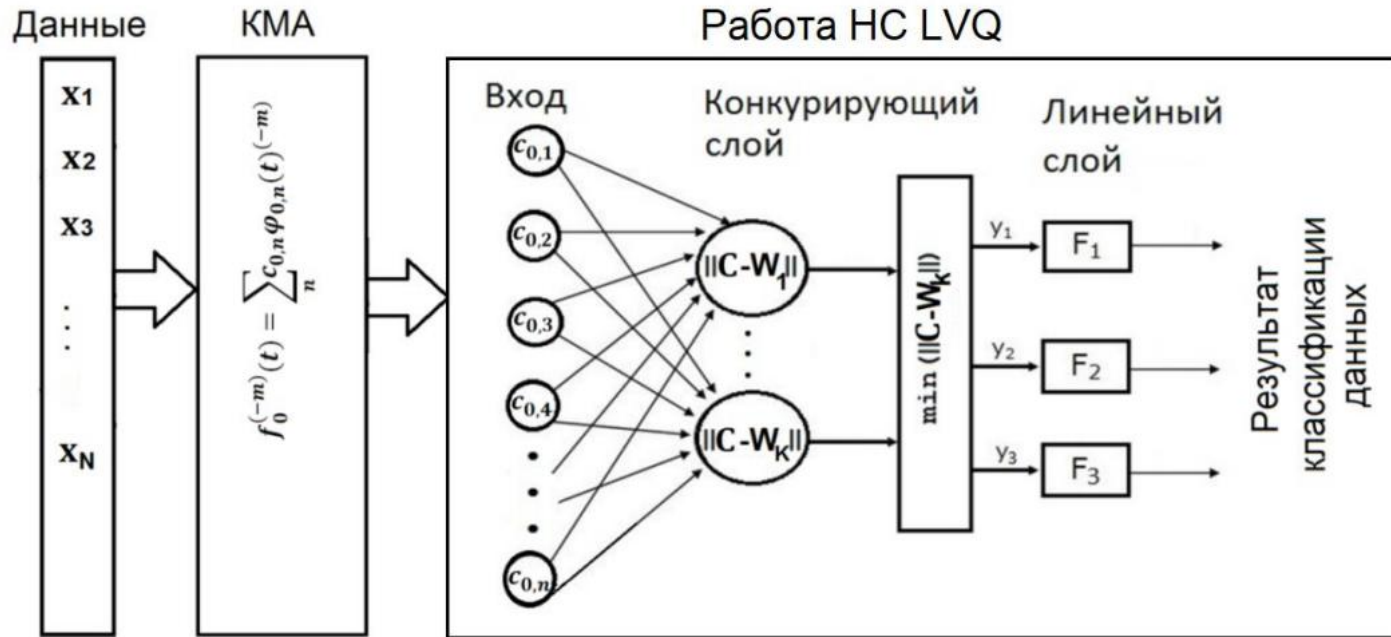
Схема решения задачи



**Результаты работы построенных нейронных сетей
(высокая солнечная активность)**

Входные данные НС	Производительность НС (двухсуточные интервалы)			Производительность НС (трехсуточные интервалы)		
	1 класс	2 класс	3 класс	1 класс	2 класс	3 класс
Исходные данные НМ	100%	60%	72%	100%	80%	93%
Предобработка db3_1	60%	53%	72%	100%	87%	93%
Предобработка db3_2	-	-	-	67%	47%	71%
Предобработка soif3_1	53%	53%	43%	100%	93%	93%
Предобработка soif3_2	-	-	-	60%	53%	57%

Схема решения задачи

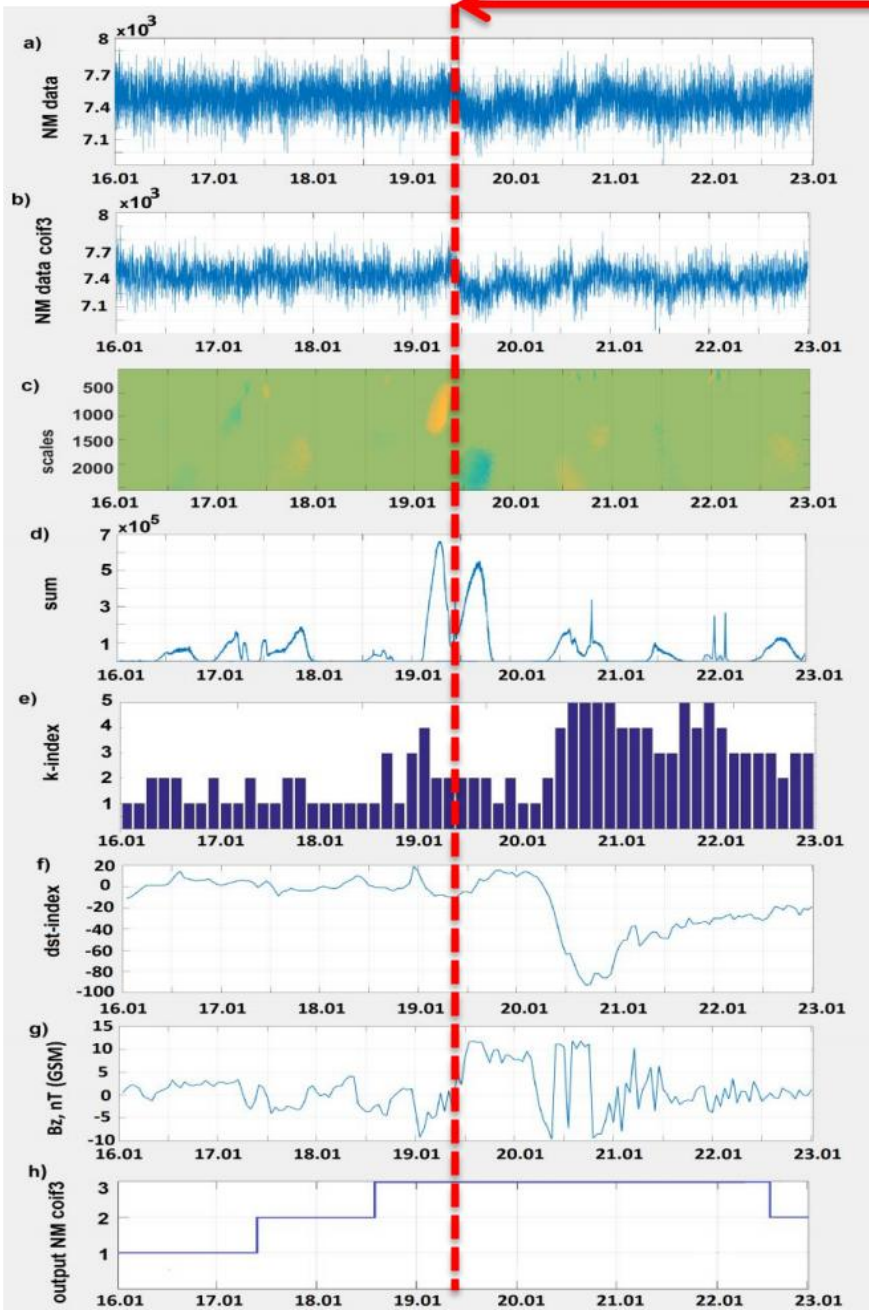


Результаты работы построенных нейронных сетей (низкая солнечная активность)

Входные данные НС	Работа НС (трехсуточные интервалы)		
	1 класс	2 класс	3 класс
Исходные данные НМ	73%	75%	67%
db3_1	-	-	-
coif3_1	80%	75%	83%

Результаты применения метода

Начало магнитной бури



данные НМ станции Кингстон, 2016г

данные НМ с применением вейвлета Койфлет 3, разложение до уровня $m=1$

результат применения порогового алгоритма

интенсивность выделенных аномальных изменения по результатам порогового алгоритма

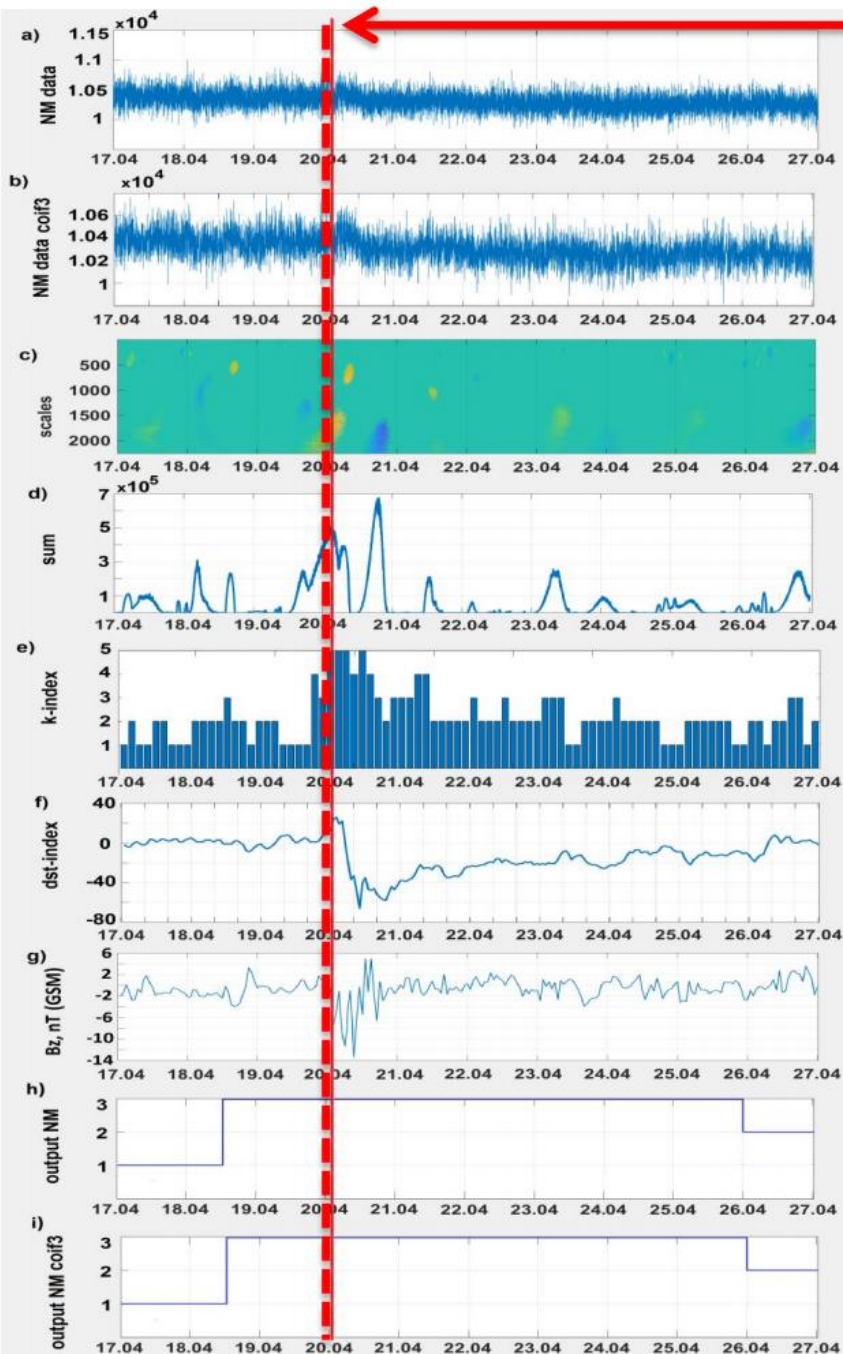
k-индекс

Dst-индекс

B_z компонента межпланетного магнитного поля

результаты работы нейронной сети LVQ с предобработкой сигнала на основе КМА, использовался вейвлет Койфлет3

Результаты применения метода



Начало магнитной бури

данные НМ станции Москва, 2018г

данные НМ с применением функции Койфлет 3, разложение до уровня $m=1$

результат применения порогового алгоритма

интенсивность выделенных аномальных изменения по результатам порогового алгоритма

k-индекс

Dst-индекс

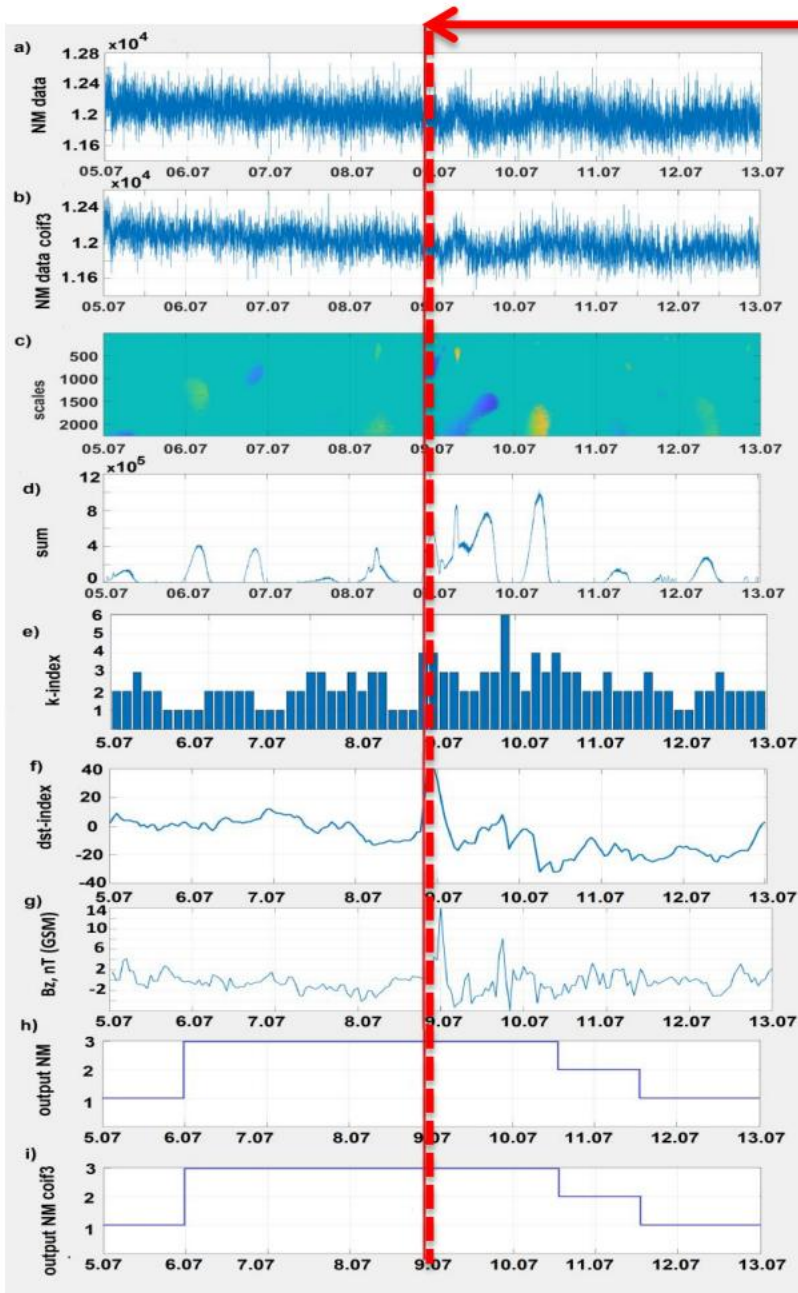
B_z компонента межпланетного магнитного поля

результаты работы нейронной сети LVQ

результаты работы нейронной сети LVQ с предобработкой сигнала на основе КМА, использовался вейвлет Койфлет3

Результаты применения метода

Начало магнитной бури



данные НМ станции Новосибирск, 2019г

данные НМ с применением функции Койфлет 3, разложение до уровня $m=1$

результат применения порогового алгоритма

интенсивность выделенных аномальных изменения по результатам порогового алгоритма

k-индекс

Dst-индекс

Bz компонента межпланетного магнитного поля

результаты работы нейронной сети LVQ

результаты работы нейронной сети LVQ с предобработкой сигнала на основе КМА, использовался вейвлет Койфлет3

Результаты применения метода в периоды сильных и умеренных магнитных бурь

Анализируемые магнитные бури (станция)	Аномалии накануне бури (класс / время до бури)	Основная фаза бури (класс)	Фаза восстановления (класс)
Июль 10-16 2013 Kingston	2 / 24 часа 3/ 12 часов	3	2
Март 15-20 2015 Kingston	2/ 48 часов 3/ 12 часов	3	2
Январь 16-22 2016 Kingston	2 / 24 часа	3	2
Август 21-28 2018 Moscow	2/ 18 часов	3	2
Март 12-19 2018 Moscow	2/ 48 часов 3/ 24 часа	3	2
Апрель 17-26 2018 Moscow	3/ 24 часа	3	1
Апрель 17-26 2018 Novosibirsk	3/ 24 часа	3	1
Октябрь 4-11 2018 Moscow	2/ 68 часов 3/ 9 часов	3	2
Июль 05-12 2019 Moscow	2/ 68 часов 3 / 12 часов	3	1
Июль 05-12 2019 Novosibirsk	3/ 68 часов	3	1
Июнь 4-11 2019 Moscow	2/9 часов	2	1

Выводы

1. Предлагаемый метод показал эффективность в задачах изучения динамики потока космических лучей и выделения спорадических эффектов.

Эмпирически доказана результативность совместного применения кратномасштабных вейвлет-разложений с нейронными сетями векторного квантования. Экспериментально подтверждена применимость метода для обнаружения спорадических эффектов малой амплитуды.

2. Предложен алгоритм определения «наилучшего» аппроксимирующего базиса в классе ортогональных вейвлетов, основанный на минимаксном подходе.

3. Результаты оценок показали:

- Погрешность метода в период высокой солнечной активности составила менее 7%, в период низкой солнечной активности – менее 25%;
- Показана возможность применения метода в оперативном режиме (на примере магнитных бурь 2015-2019 гг. по измерениям данных разных станций);
- Выявлено, что в периоды низкой солнечной активности амплитуды флуктуаций космических лучей имеют меньший размах, что усложняет задачу разделения признаков классов.

В будущем планируется расширение спектра анализируемых станций и увеличение статистического материала.