

На правах рукописи



Залаяев Тимур Ленарович

**МЕТОД И АЛГОРИТМЫ АНАЛИЗА ДАННЫХ КОСМИЧЕСКИХ ЛУЧЕЙ
В ЗАДАЧАХ ВЫДЕЛЕНИЯ СПОРАДИЧЕСКИХ ЭФФЕКТОВ**

**05.13.18 Математическое моделирование, численные методы
и комплексы программ**

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

кандидата технических наук

Санкт-Петербург, 2019

Работа выполнена в ФГБУН «Институт космических исследований и распространения радиоволн Дальневосточного отделения Российской академии наук» (ИКИР ДВО РАН)

Научный руководитель: **Мандрикова Оксана Викторовна**

доктор технических наук, доцент

Официальные оппоненты: **Чье Ен Ун**

доктор технических наук, профессор, заведующий кафедрой «Автоматика и системотехника» ФГБОУ ВО «Тихоокеанский государственный университет», г. Хабаровск

Маркелов Олег Александрович

кандидат технических наук, доцент, заместитель заведующего кафедрой «Радиотехнические системы» ФГАОУ ВО «Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)» по научной работе, г. Санкт-Петербург

Ведущая организация:

ФГБУН «Институт земного магнетизма, ионосферы и распространения радиоволн им. Н.В. Пушкова Российской академии наук» (ИЗМИРАН), г. Троицк

Защита диссертации состоится 27 марта 2019 года в 15 часов 30 минут на заседании диссертационного совета Д 212.238.01 Санкт-Петербургского государственного электротехнического университета «ЛЭТИ» им. В.И. Ульянова (Ленина) по адресу: 197376, Санкт-Петербург, ул. Проф. Попова, 5.

С диссертацией можно ознакомиться в библиотеке ФГАОУ ВО «Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В.И. Ульянова (Ленина)» и на сайте университета www.etu.ru в разделе «Подготовка кадров высшей квалификации» – «Объявление о защитах».

Автореферат разослан 9 января 2019 года.

Ученый секретарь
диссертационного совета Д 212.238.01
К. Т. Н., доцент

Пазников А.А.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. Работа посвящена разработке метода и алгоритмов анализа сложно структурированных природных временных рядов и направлена на исследование характерных свойств и выделение аномальных эффектов в периоды воздействия возмущающих факторов. Объектом исследования являются регистрируемые на Земле данные галактических космических лучей (КЛ). Поток галактических КЛ в межпланетном пространстве испытывает воздействие атмосферы Земли, земного магнитного поля и неоднородного по структуре магнитного поля Солнца и солнечного ветра. Вследствие влияния указанных факторов регистрируемые вариации КЛ имеют сложную нестационарную структуру, включающую рекуррентные и спорадические составляющие. К рекуррентным вариациям относят 22-летние, 11-летние, 27-дневные и суточные вариации (например, Топтыгин, Крымский). К спорадическим вариациям, которые являются предметом данного исследования, относятся Форбуш-понижения и сильные наземные возрастания интенсивности КЛ. Форбуш-понижения – изменения интенсивности галактических КЛ, которые происходят в результате их взаимодействия с неоднородностями магнитного поля в солнечном ветре.

Применяемые в настоящее время методы анализа вариаций КЛ имеют следующие возможности и ограничения:

1. Используемые процедуры сглаживания, например усредняющие методы (SWS Australian Antarctic Division, Alania, Lockwood и др.), позволяют выделить устойчивые характеристики процесса, но ведут к потере части важной информации. В частности, в результате искажения информации полученные характеристики спорадических вариаций КЛ могут оказаться как завышенными, так и заниженными. Некоторые Форбуш-понижения имеют малую амплитуду и сложно детектируются из-за наличия высокого уровня шума (Kota).

2. Применение классических спектральных методов (Крымский, Козлов, Нымик, Стародубцев) позволяет определять некоторые характеристики данных КЛ, но не позволяет изучать их временную эволюцию. По аналогии с процедурой сглаживания также теряется возможность выделения тонких спорадических эффектов.

3. Использование широко применяемых для анализа данных КЛ глобальных методов, например метода глобальной съемки (Крымский, Nagashima, Белов, Янке, Дворников, Дорман, Григорьев и др.), позволяет с приемлемой точностью выделять определенные динамические особенности вариаций КЛ, а именно величину и направление их анизотропии. Однако данные методы требуют трудоемких расчетов, и их автоматизация весьма затруднена.

Для описания временного хода данных КЛ в диссертационной работе предложена обобщенная модель, включающая рекуррентную составляющую и разномасштабные особенности. Для идентификации модели предложен метод, основанный на вейвлет-преобразовании и нейронных сетях. Вейвлет-преобразование (Daubechies, Mallat) является эффективным средством отображения локальных свойств сложно структурированных нестационарных данных. Преимущество нейросетевого представления аппроксимируемой функции заключается в большой гибкости базовых функций и их способности к адаптации (Haykin).

В работе используется схема кратномасштабных вейвлет-разложений и архитектура прямонаправленных многослойных нейронных сетей. Показано, что предлагаемый метод позволяет с высокой эффективностью выделять рекуррентные компоненты временных рядов данных КЛ и аномальные эффекты, возникающие в периоды гелиосферных возмущений. С использованием непрерывного вейвлет-преобразования разработан алгоритм детального анализа вариаций КЛ, позволяющий выделять мелкомасштабные аномальные эффекты и оценивать их энергетические характеристики.

Таким образом, задачи построения методов и алгоритмов анализа вариаций КЛ, а также разработка соответствующих прикладных программных средств являются актуальными и решаются в данной работе.

Целью работы являются:

- построение модели данных нейтронных мониторов, описывающей их рекуррентные и спорадические свойства;
- разработка метода и алгоритмов анализа данных нейтронных мониторов и выделения спорадических эффектов.

Для достижения цели исследования **сформулированы и решены следующие задачи:**

1. По данным нейтронных мониторов выполнен анализ структуры вариаций космических лучей и построена их обобщенная модель, позволяющая описать рекуррентные и спорадические свойства.
2. Разработан метод идентификации рекуррентных и спорадических компонент модели.
3. Построены алгоритмы анализа данных космических лучей и оценки характеристик спорадических эффектов солнечной природы.
4. Разработаны прикладные программные системы открытого доступа (<http://www.ikir.ru/ru/Departments/Paratunka/lisa/programs/art005.html>; <http://aurorasa.ikir.ru:8580>), реализующие разработанный метод и алгоритмы.

Объект и предмет исследования

Объектом исследования являются сложно структурированные данные космических лучей, которые отражают свойства и характеристики гелиосферных процессов.

Предметом исследования являются модели временного хода регистрируемых на поверхности Земли вариаций КЛ, описывающие их характерные изменения и аномальные эффекты, возникающие в периоды гелиосферных возмущений. Также предметом исследования являются методы и алгоритмы анализа данных нейтронных мониторов, направленные на изучение их характерных свойств и локальных особенностей.

Методы исследования. В диссертационной работе используется аппарат теории случайных процессов, математической статистики, нейронных сетей, цифровой обработки сигналов, вейвлет-преобразования.

Научную новизну работы составляют:

1. Модель данных космических лучей, описывающая рекуррентные и спорадические изменения. В отличие от существующих, модель позволяет повысить точность отображения рекуррентных свойств процесса, обеспечивает возможность адаптации

к долговременным рекуррентным изменениям (22-летние, 11-летние вариации), отображает спорадические изменения и реализуема в автоматическом режиме.

2. Метод идентификации рекуррентных и спорадических компонент модели, основанный на комбинации методов вейвлет-преобразования и нейронных сетей. Метод основан на предложенной модели и позволяет эффективно выделять рекуррентные компоненты и идентифицировать спорадические особенности малых амплитуд.

3. Вычислительный алгоритм по обнаружению моментов возникновения спорадических особенностей, оценке их длительности и амплитуды. Алгоритм основан на совмещении кратномасштабного анализа и нейронных сетей; в отличие от аналогов, алгоритм позволяет более точно оценить параметры спорадических особенностей.

4. Впервые разработан автоматизированный алгоритм детального анализа структуры данных космических лучей, основанный на непрерывном вейвлет-преобразовании и позволяющий выделять тонкие локальные особенности и оценивать их параметры (моменты возникновения и интенсивность).

Положения, выносимые на защиту:

1. Обобщенная модель сложно структурированных данных космических лучей позволяет описать рекуррентные вариации и разномасштабные спорадические особенности, возникающие в периоды гелиосферных возмущений.

2. Разработанный метод идентификации обобщенной модели позволяет эффективно выделять рекуррентные компоненты и идентифицировать спорадические особенности, в т. ч. малых амплитуд.

3. Алгоритм выделения крупномасштабных спорадических эффектов позволяет с высокой точностью оценить моменты возникновения, длительность и амплитуду аномальных изменений в вариациях космических лучей.

4. Алгоритм детального анализа данных космических лучей позволяет выделить разномасштабные аномальные особенности и оценить их интенсивность и частотно-временные характеристики.

5. Разработанные прикладные программные средства могут быть использованы для получения важной информации в задачах прогноза космической погоды.

Практическая ценность работы заключается в следующем:

1. Разработаны прикладные программы анализа данных, позволяющие в автоматическом режиме идентифицировать рекуррентные компоненты и выделить аномальные эффекты. Программы введены в эксплуатацию в обсерваториях нейтронных мониторов ИКИР ДВО РАН и планируются к внедрению в другие обсерваторий.

2. Предложенная обобщенная модель временного хода данных космических лучей позволяет описать регулярные составляющие и разномасштабные аномальные особенности и применима для описания сложных данных при условии отсутствия полной априорной информации о виде функциональной зависимости.

3. Разработанный метод идентификации рекуррентной и спорадических составляющих вариаций КЛ учитывает внутреннюю структуру данных и обеспечивает выделение регулярных компонент, а также разномасштабных особенностей. Разработанный алгоритм детального анализа вариаций КЛ позволяет изучать локальные характеристики (моменты возникновения и энергетические характеристики) сложно структурированных нестационарных данных. Разработанные средства применимы для более широкого круга задач анализа сложных данных с нестационарной структурой, содержащих разномасштабные особенности.

Реализация и внедрение результатов исследований

Разработанные программные системы используются для анализа данных нейтронных мониторов в Институте космофизических исследований и распространения радиоволн ДВО РАН (ИКИР ДВО РАН).

Работа частично выполнена в рамках проектов по программам фундаментальных исследований Президиума РАН и отделений РАН № 12-И-П10-01 «Исследование модуляционных эффектов галактических и солнечных космических лучей в геосферах с помощью наземных, лидарных и космических наблюдений».

Исследования поддержаны грантом Президента РФ МД-2199.2011.9 «Средства и системы анализа ионосферных и геомагнитных данных» 2011–2012 гг.; грантом РФФИ – ДВО РАН № 11-07-98514-р_восток_a «Теоретические основы и алгоритмическое обеспечение систем анализа ионосферных и геомагнитных данных»

2011–2013 гг.; грантом федеральной программы «Участник молодежного научно-исследовательского конкурса», госконтракт № 862ГУ1/2013 от 09.12.2013 г. по теме «Разработка системы интеллектуального анализа вариаций космических лучей»; грантами РФФИ № 14-11-00194 и № 14-11-00194-П «Интеллектуальные средства и программные системы анализа динамических процессов в магнитосферно-ионосферной системе в периоды возмущений».

Апробация работы. Основные положения и результаты работы докладывались и обсуждались на следующих научно-технических конференциях:

1. 9-й, 10-й международной конференции «Интеллектуализация обработки информации», Черногория, г. Будва, 2012 г.; Греция о. Крит 2014 г.

2. 6-й международной конференции «Солнечно-земные связи и физика предвестников землетрясений», Петропавловск-Камчатский, 2013 г.

3. 3-й всероссийской конференции «Радиоэлектронные средства получения, обработки и визуализации информации», Смоленск, 2013 г.

4. 2-й международной научно-технической конференции «Вычислительный интеллект – 2013 (результаты, проблемы, перспективы)», г. Черкассы, Украина, 2013 г.

5. Международной молодежной конференции «Прикладная математика, управление и информатика», Белгород, 2012 г.

6. 2-й международной конференции International CAWSES-II Symposium, Japan, Nagoya, 2015 г.

7. 11-й, 12-й, 13-й международной конференции «Annual Meeting of Asia Oceania Geosciences Society» (AOGS), Japan, Sapporo, 2014; Singapore, 2015; China, Beijing, 2016.

8. 40th COSPAR Scientific Assembly. 2–10 August 2014, Moscow, Russia, Abstract S. 3-4-14.

9. Международной конференции SPITSE, Германия, Immenau, 2014 г.

10. 6-й, 9-й, 10-й, 11-й международной конференции «Международная конференция по мягким вычислениям и измерениям (SCM)», Санкт-Петербург, 2013, 2016, 2017, 2018 гг.

11. Международной конференции «Цифровая обработка сигналов и ее применение» (DSPA), Москва, 2013 г.

12. 11-й международной школе-конференции «Проблемы геокосмоса», Санкт-Петербург, 2017 г.

Публикации. Основные теоретические и практические результаты диссертации опубликованы в 30 печатных работах, среди которых 1 монография, 11 научных статей в рецензируемых изданиях, рекомендованных ВАК РФ, в т. ч. 6 статей опубликованы в зарубежных изданиях, индексируемых в базах WoS и Scopus, 1 свидетельство программ ЭВМ, 18 статей, опубликованных в других изданиях и материалах конференций.

Структура и объем диссертации. Диссертация состоит из введения, четырех глав, заключения, списка литературы, включающего 126 наименований. Содержание работы изложено на 159 страницах машинописного текста.

СОДЕРЖАНИЕ РАБОТЫ

Во **введении** дано обоснование актуальности рассматриваемой работы, сформулированы цели и задачи исследований, показана и обоснована актуальность работы, научная новизна и практическая ценность.

В первой главе даются необходимые сведения об исследуемом процессе. Рассматриваются основные используемые методы изучения регистрируемых на Земле данных КЛ, отмечены их возможности и недостатки. Для их устранения предложено использовать современные методы представления и анализа данных, такие как вейвлет-преобразование и нейронные сети. Предложена обобщенная модель вариаций КЛ, позволяющая описать рекуррентные и спорадические компоненты сигнала. Определены следующие основные задачи исследования:

1. Анализ природы данных нейтронных мониторов и создание обобщенной модели вариаций КЛ, описывающей его рекуррентные и спорадические составляющие.

2. На основе полученной модели разработка метода анализа данных нейтронных мониторов и выделения рекуррентных и спорадических составляющих.

3. Разработка алгоритмов по реализации метода.

4. Разработка программных систем, реализующих разработанные алгоритмы.

В рамках задач исследования вариация КЛ может быть представлена в следующем виде:

$$f_{HM}(t) = f_{рекк}(t) + f_{спор}(t) + e(t),$$

где $f_{рекк}(t) = f_{внз}(t) + f_{магн}(t)$, $f_{внз}(t)$ – внеземная составляющая, $f_{магн}(t)$ – магнитосферная составляющая; $f_{спор}(t)$ – спорадическая составляющая, $e(t) = f_{атм}(t) + f_{анн}(t) + f_{случ}(t)$, $f_{атм}(t)$ – атмосферная составляющая, $f_{анн}(t)$ – аппаратная составляющая, $f_{случ}(t)$ – случайная составляющая.

В соответствии с определенной структурой данных КЛ, содержащих рекуррентную составляющую и разномасштабные спорадические эффекты, они могут быть представлены в виде:

$$f_{HM}(t) = R(t) + \sum_k L_k(t) + \sum_g S_g(t) + e(t) = \sum_n r_n \phi_n + \sum_k \sum_n l_n^k \rho_n^k + \sum_g \sum_n s_n^g \tau_n^g + e(t), \quad (1)$$

где предполагается, что $f_{HM}(t) \in H$; рекуррентная составляющая $R(t)$ может быть представлена как $R(t) = \sum_n r_n \phi_n(t)$, $\phi_n = \{\phi_{j,n}\}_{j \in Z}$ – базис, $r_n = \langle f_{HM}(t), \phi_n \rangle$; спорадические составляющие $L_k(t)$ и $S_g(t)$ аддитивны и могут быть представлены: $\sum_k L_k(t) = \sum_k \sum_n l_n^k \rho_n^k$ – крупномасштабные спорадические составляющие, $\rho_n^k = \{\rho_{n,j}^k\}_{j \in Z}$ – базис, $l_n^k = \langle f_{HM}(t), \rho_n^k \rangle$, $\sum_g S_g(t) = \sum_g \sum_n s_n^g \tau_n^g$ – тонкие спорадические эффекты, $\tau_n^g = \{\tau_{n,j}^g\}_{j \in Z}$ – базис, $s_n^g = \langle f_{HM}(t), \tau_n^g \rangle$; $e(t)$ – шум.

Для идентификации модели (1) предложен метод, основанный на вейвлет-преобразовании и нейронных сетях:

• Вейвлет-преобразование обеспечивает численную устойчивость разложений:

$$\text{если } f(t) = f_{a,(-m)}(t) + \sum_{j=-m}^{-1} g_j(t), \quad (2)$$

$$\text{то } \|f(t) - (f_{a,(-m)}(t) + \sum_{j=-m}^{-1} g_j(t))\| = \|f(t) - \sum_n c_{-m,n} \phi_{-m,n}(t) - \sum_{j=-m}^{-1} \sum_n d_{j,n} \Psi_{j,n}(t)\|,$$

где $g_j(t) \in W_j$, $f_{a,-m}(t) \in V_{-m}$, $g_j(t) = \sum_n d_{j,n} \Psi_{j,n}(t)$, $f_{a,-m}(t) = \sum_n c_{-m,n} \phi_{-m,n}$, $\Psi_j = \{\Psi_{j,n}\}_{n \in \mathbb{Z}}$ – вейвлет-базис, $\phi_j = \{\phi_{j,n}\}_{n \in \mathbb{Z}}$ – базис, порожденный скейлинг-функцией.

- Полученное на основе КМА представление (2) позволяет изучить структуру данных и выделить шум.

- Из теоремы Жаффара следует:

$$|(W_\Psi f)(b, s)| \leq A s^{\alpha + \frac{1}{2}}, \quad (3)$$

где $A > 0$, α – показатель Липшица функции f в окрестности $t = b$, и непрерывное вейвлет-преобразование (НВП) позволяет локализовать мелкие разномасштабные особенности и определить их моменты возникновения, длительность и спектральную структуру.

- Основываясь на теореме Цыбенко, можно утверждать, что многослойные нейронные сети прямого распространения (МНСПР) позволяют для любой функции $p(f_{t_1}, \dots, f_{t_{\gamma_0}})$ построить аппроксимацию вида:

$$F(f_{t_1}, \dots, f_{t_{\gamma_0}}) = \sum_{i=1}^{\gamma_1} \alpha_i \Phi(\sum_{j=1}^{\gamma_0} w_{ij} f_{t_j} + \beta_i), \quad |F(f_{t_1}, \dots, f_{t_{\gamma_0}}) - p(f_{t_1}, \dots, f_{t_{\gamma_0}})| < \varepsilon, \quad (4)$$

где γ_0 , γ_1 – число нейронов входного и скрытого слоя НС соответственно; $\varepsilon > 0$, $\varepsilon \in \mathbb{R}$, $(f_{t_1}, \dots, f_{t_{\gamma_0}})$ – вектор входных значений сети, $\Phi(\lambda(n)) = \frac{1}{1 + e^{-a\lambda(n)}}$ – сигмоидальная функция активации.

Во **второй главе** описан предлагаемый метод идентификации модели данных КЛ. Приведен способ построения структуры нейронной сети и формирования обучающего множества. Описан разработанный алгоритм оценки параметров спорадических составляющих вариаций КЛ.

Предлагаемый метод идентификации модели данных КЛ включает следующие операции:

1. Совместное применение МНСПР и конструкции кратномасштабного анализа (КМА) для идентификации компоненты модели $f_\alpha(t) = R(t) + \sum_k L_k(t)$.

2. Процедуру анализа ошибок обученной нейронной сети для оценки спорадических компонент модели $L_k(t)$.

3. Применение непрерывного вейвлет-преобразования (НВП) для идентификации разномасштабных спорадических компонент модели $f_\beta = \sum_k L_k(t) + \sum_g S_g(t)$.

При построении процедуры идентификации компоненты модели $f_\alpha(t) = R(t) + \sum_k L_k(t)$ возникает задача найти оператор M , который задаст наилучшее соответствие между данными и построенным образом. В работе показано, что рекуррентные составляющие $R(t)$ могут быть определены на основе компоненты $f_{a,(-m)}$ КМА (см. соотн. (2)). В периоды аномальных изменений временного хода компонента $f_{a,(-m)}$ также может определять спорадические вариации $L_k(t)$, которые характеризуют возникновение Форбуш-эффектов. Детализирующие компоненты $g_j(t)$ (см. соотн. (2)) определяют мелкомасштабные спорадические возмущения (например, локальные понижения и повышения КЛ).

Поскольку мы имеем неполные априорные данные о компоненте $f_\alpha(t)$, имея представление функции в виде $f(t) = f_{a,(-m)}(t) + \sum_{j=-m}^{-1} g_j(t)$ (см. соотн. (2)), для оценки параметров $f_\alpha(t)$ предложено использовать МНСПР, которые позволяют определить наилучший оператор M , минимизирующий ошибку путем перекрытия признакового пространства:

$$F: f_{a,(-m)}(t) \rightarrow \overline{\overline{f_{a,(-m)}}(t)},$$

где $f_{a,(-m)}(t)$ – вход НС, $\overline{\overline{f_{a,(-m)}}}$ – выход НС.

В этом случае процедура определения наилучшего оператора M включает следующие операции:

1. На основе КМА получаем компоненту временного ряда в виде

$$f_{a,(-m)}(t) = \sum_n c_{-m,n} \phi_{-m,n}(t).$$

2. С учетом априорных знаний формируем массивы обучающих и тестовых данных для нейронной сети; используя обучающие данные $\{c_{-m,n}^{об}\}_{n=n_0}^N$, строим многослойную прямонаправленную сеть с фиксированным числом нейронов скрытого слоя $\gamma_1: F: f_{a,(-m)} \rightarrow \overline{\overline{f_{a,(-m)}}}$.

3. Используя тестовые данные $\{c_{-m,n}^{mest}\}_{n=n_0}^L$, оцениваем погрешность работы

$$E(m) = \frac{\|\overline{f_{a,(-m)}}(t) - f_{a,(-m)}(t)\|}{\|f_{a,(-m)}\|}.$$

4. Используя тестовые данные $\{c_{-m,n}^{mest}\}_{n=n_0}^L$, оцениваем потерю $\epsilon(m) =$

$$= \frac{\|f_{HM}(t) - f_{a,(-m)}(t)\|}{\|f_{HM}\|}.$$

5. На основе полученных оценок определяем наилучший уровень разложения m , обеспечивающий наименьшую погрешность аппроксимации данных.

Результаты оценки погрешности аппроксимации, полученные на основе выполнения операций 3 и 4, приведены в таблице 1. На основе полученных оценок в работе за наилучший уровень разложения принят уровень $m^* = 5$.

Таблица 1. Погрешность аппроксимации для различных уровней вейвлет-разложения m

m	γ_1	ϕ	$E(m)$	$\epsilon(m)$	Разм. кр. эфф.
0	8	Нет	1,31%	0,00%	0
1	8	Coif3	0,140%	0,80%	16
2	8	Coif3	0,080%	1,00%	32
3	8	Coif3	0,0133%	1,10%	64
4	8	Coif3	0,0023%	1,150%	126
5	8	Coif3	0,00035%	1,1700%	256
6	8	Coif3	0,00011%	1,1800%	512
7	8	Coif3	0,00003%	1,1805%	1024
8	8	Coif3	0,000015%	1,190%	2048

Таблица 2. Погрешность аппроксимации для различных базисных вейвлетов

m	γ_1	ϕ	$E(m)$	$\epsilon(m)$	Разм. кр. эфф.
5	8	Coif1	0,0125%	1,10%	64
5	8	Coif2	0,00147%	1,17%	160
5	8	Coif3	0,00035%	1,1700%	256
5	8	Coif4	0,00035%	1,1700%	352
5	8	Coif5	0,000295%	1,1700%	448
5	8	Db1	0,05273%	1,16%	32
5	8	Db2	0,0127%	1,1700%	96
5	8	Db3	0,00422%	1,1700%	160
5	8	Db4	0,00167%	1,160%	224
5	8	Db5	0,00132%	1,1700%	288

С использованием аналогичного подхода были получены оценки погрешности аппроксимации для различных вейвлетов (см. таблицу 2). Учитывая результаты оценок, за наилучшую базисную функцию принят вейвлет семейства Койфлеты порядка 3.

В периоды отсутствия спорадических эффектов в данных КЛ можно исходить из предположения, что вся информация о данных находится в самих данных, тогда значение функции в момент $t = n$ в соответствии с теоремой об универсальной аппроксимации может быть получено путем статистической экстраполяции:

$r_n = \rho(r_{n-1}, \dots, r_{n-\gamma_0})$, где ρ – некоторая нелинейная функция, r_n – значение аппроксимируемой функции $r(t)$ при $t = n$.

Тогда искомая нелинейная функция $\rho(r_{n-1}, \dots, r_{n-\gamma_0})$ может быть аппроксимирована описанным выше нейросетевым представлением $F(r_{n-1}, \dots, r_{n-\gamma_0})$ (см. соотн. (4)). С учетом введения обобщающего слоя НС получаем следующую нейросетевую модель статистической экстраполяции в момент $t = n$:

$$F: \bar{r}_n = \sum_{i=1}^{\gamma_2} \alpha_i \Phi_{\text{сигм}} \left(\sum_{j=1}^{\gamma_1} w_{ij} \Phi_{\text{сигм}} \left(\sum_{k=1}^{\gamma_0} w_{jk} r_{n-k} + \beta_k \right) + \beta_j \right) + \beta, \quad (5)$$

где $\Phi_{\text{сигм}}(\lambda(n)) = \frac{1}{1+e^{-\lambda(n)}}$ – сигмоидальная функция активации, w_{jk} – весовые коэффициенты 1-го скрытого слоя, w_{ij} – весовые коэффициенты 2-го скрытого слоя, α_i – весовые коэффициенты выходного нейрона, β_k – пороговые коэффициенты 1-го слоя скрытого, β_j – пороговые коэффициенты 2-го скрытого слоя, β – пороговый коэффициент выходного нейрона, γ_0 – размерность вектора входа, γ_1 и γ_2 – число нейронов 1-го и 2-го слоев.

Построение обучающего множества НС основано на априорных данных о процессе и выполнено с учетом уровня солнечной активности и гелиосферных параметров. Для определения размерности входного вектора НС в работе выполнены оценки погрешности модели НС (см. таблицу 3). Учитывая результаты полученных оценок, определена размерность входного вектора НС: $\gamma_0 = 5$.

На основании результатов полученных оценок (см. таблицы 1–3) и с учетом выбранной нейросетевой модели (см. соотн. (5)), в работе получена следующая модель данных КЛ:

$$\bar{c}_{-5,n} = \sum_{i=1}^{\gamma_2} \alpha_i \Phi_{\text{сигм}} \left(\sum_{j=1}^{\gamma_1} w_{ij} \Phi_{\text{сигм}} \left(\sum_{k=1}^6 w_{jk} c_{-m,n-k} + \beta_k \right) + \beta_j \right) + \beta, \quad (6)$$

где $c_{-5,n} = \langle f_{HM}, \phi_{-5,n} \rangle$.

На основе полученной модели (6) разработан следующий алгоритм 1 выделения и оценки параметров спорадических возмущений:

1. На основе модели получаем вектор ошибок НС: $e(n) = \bar{r}_n - r_n$, где r_n – реальное значение коэффициента в момент времени $t = n$, \bar{r}_n – экстраполированное значение.

Таблица 3. Погрешность модели НС для различной размерности входного вектора

ϕ	m	γ_0	$E(\gamma_0)$
Coif3	5	1	1.2072e-004
Coif3	5	2	1.0237e-005
Coif3	5	3	1.6214e-005
Coif3	5	4	4.0292e-006
Coif3	5	5	3.2607e-006
Coif3	5	6	2.2492e-006
Coif3	5	7	3.3631e-006
Coif3	5	8	3.6812e-006

2. Если $|e(t_0)| \geq T_s$, где T_s – некоторая пороговая функция, то t_0 – момент возникновения особенности.

3. Оцениваем временную длительность выделенной особенности: $L_{cn} = t_1 - t_0$, где $t_1 = \min_t(t > t_0 ; |e(t)| < T_s)$.

4. Оцениваем относительную амплитуду особенности $A_{cn} = \max_{t_0 < t < t_1} e(t)$.

В **третьей главе** описан способ идентификации компоненты модели $f_\beta = \sum_k L_k(t) + \sum_g S_g(t)$ и реализующий его *алгоритм 2* детального анализа данных КЛ и выделения спорадических особенностей. Показано, что предлагаемый *алгоритм 2* позволяет выполнять оценку следующих характеристик выделенных особенностей: моменты возникновения, длительность и интенсивность.

Предлагаемый *алгоритм 2* основан на применении НВП и включает следующие операции:

1. Выполнение НВП:

$$(W_\Psi f_{b,s}) := |s|^{-\frac{1}{2}} \int_{-\infty}^{+\infty} f(t) \Psi\left(\frac{t-b}{s}\right) dt, f \in L^2(R), s, b \in R, s \neq 0.$$

2. Применение пороговой функции P_{T_s} :

$$P_{T_s}(W_\Psi f_{b,s}) = \begin{cases} W_\Psi f_{b,s}, & \text{если } (W_\Psi f_{b,s} - W_\Psi f_{b,s}^{med,l}) \geq T_s^l \\ 0, & \text{если } |W_\Psi f_{b,s} - W_\Psi f_{b,s}^{med,l}| < T_s^l \\ -W_\Psi f_{b,s}, & \text{если } (W_\Psi f_{b,s} - W_\Psi f_{b,s}^{med,l}) < -T_s^l \end{cases}$$

где $W_\Psi f_{b,s}^{med,l}$ – медианное значение, рассчитанное в скользящем временном окне

длины l , $T_s^l = U * \sigma_s^l$ – порог, $\sigma_s^l = \sqrt{\left(\frac{1}{l-1} \sum_{k=1}^l (W_\Psi f_{b,s} - \overline{W_\Psi f_{b,s}})^2\right)}$ – стандартное отклонение, рассчитанное в скользящем временном окне длины l , $W_\Psi f_{b,s}$ – среднее значение, U – пороговый коэффициент.

3. Оценку интенсивности аномалий: $Y(t) = \sum_s P_{T_s}(W_\Psi f_{b,s})$, которая в случае локального повышения КЛ будет положительной, а в случае локального понижения КЛ – отрицательной.

4. Определение момента возникновения аномалии t_k на масштабе $s_0 : t_k = b_0$, если $|W_\Psi f_{b_0,s_0} - W_\Psi f_{b_0,s_0}^{med,l}| \geq T_{s_0}^l$.

5. Оценку длительности аномалии на масштабе s_0 $L_{s_0} = t_j - t_k: \forall b_i \in \{t_1, t_2\}, t_j > t_k: |W_{\Psi} f_{b_i, s_0} - W_{\Psi} f_{b_i, s_0}^{med, l}| \geq T_{s_0}^l$.

Очевидно, погрешность *алгоритма 2* зависит от анализирующего вейвлета Ψ , длины скользящего временного окна l и порогового коэффициента U . Выбор анализирующего вейвлета Ψ в работе основывался на следующих критериях: число нулевых моментов, гладкость вейвлета и размер носителя. С учетом данных критериев и результатов статистического моделирования (приведены в главе 4), в работе определен вейвлет семейства Койфлеты порядка 1.

Определение порогового коэффициента U основано на оценке апостериорного риска (результаты оценок приведены в главе 4).

В **четвертой главе** описано разработанное программное обеспечение и приведены результаты оценки эффективности предлагаемого метода и алгоритмов (использовались модельные данные и данные станций Апатиты, Мыс Шмидта и Магадан за период 2002–2015 гг.).

Результаты экспериментов с модельными данными представлены в таблице 4. График зависимости вероятности обнаружения спорадической особенности от ее амплитуды показан на рисунке 1.

Таблица 4. Результаты оценки эффективности *алгоритма 2*

Амплитуда спорадической особенности, %	Процент выделенных спорадических особенностей
0,03	21%
0,04	31%
0,05	56%
0,06	93%

На рисунке 2 представлены результаты оценки работы *алгоритма 1*. В период Форбуш-понижения наблюдается существенное возрастание ошибок НС, что подтверждает эффективность разработанного *алгоритма 1*.

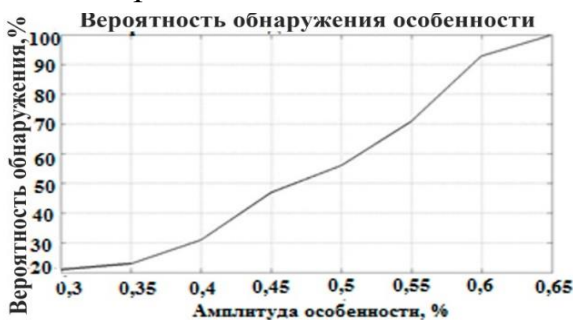


Рис. 1

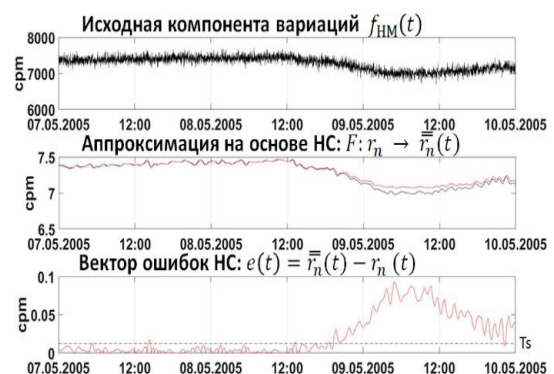


Рис. 2

На рисунке 3 представлены результаты работы алгоритма 2. Перед началом магнитной бури выделено возрастание интенсивности КЛ, в период главной фазы магнитной бури выделено Форбуш-понижение.

В таблице 5 показаны результаты работы предлагаемого метода и современного метода глобальной съемки (МГС) в задаче выделения малых Форбуш-понижений (с амплитудой 0,3% и менее). Предлагаемый метод показал сопоставимые результаты с методом МГС, однако метод МГС требует наличия данных сети станций и не автоматизирован.

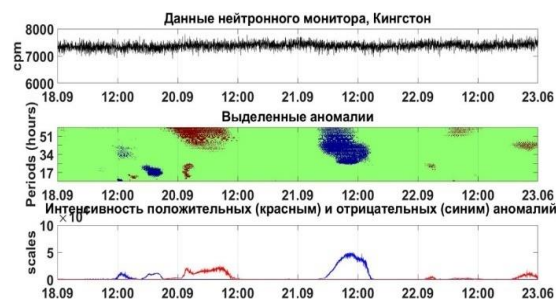


Рис. 3

Таблица 5. Сравнение с методом глобальной съемки

Даты	МГС	Предлагаемый метод	
		Алг. дет. анализа	КМА+НС
15.09.99	+	+	—
29.03.00	+	+	+
30.07.04	+	+	-
03.05.14	+	+	+

ОСНОВНЫЕ НАУЧНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

Основные научные и практические результаты работы можно сформулировать следующим образом:

1. Предложена обобщенная модель сложно структурированных данных КЛ. Модель позволяет с высокой точностью отображать рекуррентные и спорадические свойства процесса.

2. На основе комбинации аппаратов вейвлет-преобразования и нейронных сетей разработан метод идентификации предлагаемой обобщенной модели. Метод позволяет выделять рекуррентные составляющие и тонкие особенности сложно структурированных данных.

3. Разработан вычислительный алгоритм по обнаружению спорадических особенностей в данных КЛ. Алгоритм позволяет выделить особенности и с высокой точностью оценить их параметры.

4. Разработан вычислительный алгоритм детального анализа данных КЛ, позволяющий выполнять оценку их локальных характеристик (оцениваются моменты возникновения и энергетические характеристики спорадических особенностей).

5. Разработаны программные системы по реализации разработанного метода и алгоритмов. Система Cosmic Rays Analysis Tool (<http://www.ikir.ru/ru/Departments/Paratunka/lisa/programs/art005.html>) позволяет в автоматическом режиме выделять крупномасштабные спорадические возмущения в данных КЛ. Компонента программной системы Aurora (<http://aurorasa.ikir.ru:8580>) позволяет выделять разномасштабные спорадические особенности и определять их моменты возникновения, длительность и интенсивность.

Основные результаты диссертации опубликованы в следующих работах:

Монография:

1. Мандрикова О.В., **Заляев Т.Л.** Методы анализа вариаций космических лучей в задачах исследования гелиосферных процессов и выделения спорадических эффектов: монография / О.В. Мандрикова, **Т.Л. Заляев.** – П.-Камч.: изд-во КамчатГТУ. – 2016. – 129 с.

Статьи, опубликованные в изданиях, включенных в перечень ВАК:

2. **Заляев Т.Л.** Алгоритм выделения аномалий в вариациях космических лучей в периоды гелиосферных возмущений / **Т.Л. Заляев** // Известия «ЛЭТИ». – 2015. – № 10. – С. 25–32.

3. Мандрикова О.В., Соловьев И.С., **Заляев Т.Л.** Методы анализа вариаций геомагнитного поля и данных космических лучей / О.В. Мандрикова, И.С. Соловьев, **Т.Л. Заляев** // Информационные технологии. – 2015. – № 10. – С. 849–855.

4. Мандрикова О.В., **Заляев Т.Л.** Моделирование вариаций космических лучей на основе совмещения кратномасштабных вейвлет-разложений и нейронных сетей переменной структуры / О.В. Мандрикова, **Т.Л. Заляев** // Цифровая обработка сигналов. – 2015. – № 1. – С. 11–16.

5. Мандрикова О.В. и др. Анализ динамики космических лучей и параметров ионосферы в периоды повышенной солнечной активности и магнитных бурь (по

данным наземных станций) / О.В. Мандрикова, Ю.А. Полозов, **Т.Л. Залаяев**, В.В. Геппенер // Известия СПбГЭТУ «ЛЭТИ». – 2018. – № 7. – С. 96–108.

Статьи, опубликованные в зарубежных изданиях, включенных в системы цитирования Scopus и Web of Science:

6. Mandrikova O.V., Solovev I.S., **Zalyaev T.L.** Methods of analysis of geomagnetic field variations and cosmic ray data / O.V. Mandrikova, I.S. Solovev, **T.L. Zalyaev** // Earth Planet Space. – 2014. – Vol. 66. – № 148. – DOI: 10.1186/s40623-014-0148-0.

7. Mandrikova O.V. et al. Methods of analysis of geophysical data during increased solar activity / O.V. Mandrikova, Yu. A. Polozov., I.S. Solovev, N.V. Fetisova (Glushkova), **T.L. Zalyaev**, M.S. Kupriyanov, A.V. Dmitriev // Pattern Recognition and Image Analysis (Advances in Mathematical Theory and Applications). – 2016. – Т. 26. – № 2. – P. 406–418.

8. Mandrikova O.V., Polozov Yu.A., **Zalyaev T.L.** Joint analysis of the ionospheric parameters and cosmic ray data during periods of magnetic storms 2015 / O.V. Mandrikova, Yu.A. Polozov, **T.L. Zalyaev** // Solar-Terrestrial Relations and Physics of Earthquakes Precursors. E3S Web of Conferences. – 2017. – Vol. 20. – DOI: 10.1051/e3sconf/20172001007.

9. Mandrikova O.V., **Zalyaev T.L.** Modeling and analysis of cosmic rays variations during periods of heliospheric disturbances / O.V. Mandrikova, **T.L. Zalyaev** // In: Proceedings of 2017 20th IEEE International Conference on Soft Computing and Measurements, SCM 2017. – 2017. – P. 410–413. – DOI: 10.1109/SCM.2017.7970601.

10. Mandrikova O.V. et al. Analysis of the Cosmic Rays dynamics on the basis of Neural Networks / O.V. Mandrikova, **T.L. Zalyaev**, B.S. Mandrikova, M.S. Kupriyanov // Proceedings of 2018 21th IEEE International Conference on Soft Computing and Measurements (SCM 2018). – 2018. – Pp. 683–686.

11. Mandrikova O.V. et al. Analysis of the dynamics of ionospheric parameters during periods of increased solar activity and magnetic storms / O.V. Mandrikova, Yu. A. Polozov, N.V. Fetisova (Glushkova), **T.L. Zalyaev** // Journal of Atmospheric and Solar-Terrestrial Physics. – 2018. – Vol. 181. – Pp. 116–126.

Статьи, опубликованные в других изданиях и материалах конференций:

12. Мандрикова О.В. и др. Анализ геомагнитных данных и космических лучей в периоды повышенной солнечной активности / О.В. Мандрикова, Т.Л. Заляев, И.С. Соловьев // Вестник КРАУНЦ. Физико-математические науки. – 2016. – № 3 (14). – С. 40–46.

13. Смирнов С.Э. и др. Анализ явлений космической погоды в 2014 году по наблюдениям в северо-восточном регионе России / С.Э. Смирнов, В.В. Бычков, О.В. Мандрикова, И.С. Соловьев, Ю.А. Полозов, **Т.Л. Заляев**, С.Ю. Хомутов, Д.Г. Баишев, И.Н. Поддельский, З.Ф. Думбрава // Вестник КРАУНЦ. Физико-математические науки. – 2016. – № 1 (12). – С. 66–84.

14. Мандрикова О.В., **Заляев Т.Л.** Моделирование вариаций космических лучей и выделение аномалий на основе совмещения вейвлет-преобразования с нейронными сетями / О.В. Мандрикова, **Т.Л. Заляев** // Машинное обучение и анализ данных. – 2014. – Т. 1. – № 9 – С. 1154–1167.

15. Мандрикова О.В. и др. Моделирование и анализ вариаций космических лучей в периоды повышенной солнечной и геомагнитной активности / О.В. Мандрикова, **Т.Л. Заляев**, Ю.А. Полозов, И.С. Соловьев // Машинное обучение и анализ данных. – 2016. – Т. 2. – № 1. – С. 89–103.

16. Мандрикова О.В., **Заляев Т.Л.** Аппроксимация ионосферных параметров на основе совмещения вейвлет-преобразования и нейронных сетей / О.В. Мандрикова, **Т.Л. Заляев** // Вестник КамчатГТУ. – 2012. – № 21. – С. 18–24.

17. Мандрикова О.В., **Заляев Т.Л.** Выявление сильных наземных возрастных энергий космических лучей на основе совмещения кратномасштабного анализа и нейронных сетей / О.В. Мандрикова, **Т.Л. Заляев** // Вестник КамчатГТУ. – 2012. – № 22. – С. 26–31.

18. Мандрикова О.В., Полозов Ю.А., **Заляев Т.Л.** Анализ ионосферных параметров и данных космических лучей в периоды магнитных бурь 2015 года / О.В. Мандрикова, Ю.А. Полозов, Т.Л. Заляев // Вестник КамчатГТУ. – 2018. – № 43. – С. 22–29.

19. Mandrikova O.V., **Zalyaev T.L.** Modeling and analysis of cosmic ray variations during periods of heliospheric disturbances / O.V. Mandrikova, **T.L. Zalyaev** // PROBLEMS OF GEOCOSMOS Proceedings of the 11th International School and Conference. – 2016. – С. 270–277.

20. Mandrikova O.V., **Zalyaev T.L.** Modeling and analysis of the cosmic rays variations during periods of heliospheric disturbances on the basis of wavelet transform and neural networks / O.V. Mandrikova, **T.L. Zalyaev** // Proceedings of ITISE 2017. Granada, 18–20, September, 2017. – 2017. – Vol. 2. – P.1163–1174.

21. Мандрикова О.В., **Заляев Т.Л.** «Аппроксимация параметров космических лучей на основе совмещения вейвлет-преобразования и нейронных сетей» / О.В. Мандрикова, **Т.Л. Заляев** // Сборник докладов всероссийской молодежной конференции «Теория и практика системного анализа». – 2012. – С. 389–393.

22. Мандрикова О.В., **Заляев Т.Л.** Моделирование вариаций космических лучей на основе совмещения кратномасштабного анализа и сетей переменной структуры / О.В. Мандрикова, **Т.Л. Заляев** // Сборник тезисов докладов XVI Международной научно-технической конференции по мягким вычислениям и измерениям (SCM`2013). – 2013. – Т. 2. – С. 111–114.

23. Мандрикова О.В., **Заляев Т.Л.** Моделирование и анализ вариаций космических лучей в периоды гелиосферных возмущений / О.В. Мандрикова, **Т.Л. Заляев** // Сборник тезисов докладов XX Международной научно-технической конференции по мягким вычислениям и измерениям (SCM`2017). – 2017. – Т. 1. – С. 474–478.

24. Мандрикова О.В., **Заляев Т.Л.** Методы выделения и анализа тонких особенностей в вариациях космических лучей / О.В. Мандрикова, **Т.Л. Заляев** // Сборник тезисов докладов XIX Международной технической конференции по мягким вычислениям и измерениям (SCM`2016). – 2016. –Т. 1. – С. 485–488.

25. Мандрикова О.В. и др. Анализ динамики космических лучей на основе нейронной сети / О.В. Мандрикова, **Т.Л. Заляев**, Б.С. Мандрикова, М.С. Куприя-

нов // Сборник тезисов докладов XXI Международной научно-технической конференции по мягким вычислениям и измерениям (SCM`2018). – 2018. – Т. 1. – С. 716–719.

26. Мандрикова О.В. и др. Метод выявления аномалий в вариациях космических лучей на основе совмещения вейвлет-преобразования с нейронными сетями / О.В. Мандрикова, **Т.Л. Заляев**, А.В. Белов, В.Г. Янке // Сборник докладов VI международной конференции «Солнечно-земные связи и физика предвестников землетрясений». – 2013. – С. 303–307.

27. Мандрикова О.В., **Заляев Т.Л.** Аппроксимация параметров космических лучей на основе совмещения вейвлет-преобразования и нейронных сетей / О.В. Мандрикова, **Т.Л. Заляев** // Сборник трудов второй Международной научно-технической конференции «Вычислительный интеллект – 2013 (результаты, проблемы, перспективы)» (ComInt-2013). – 2013. – С. 215–217.

28. Мандрикова О.В., **Заляев Т.Л.** Вейвлет анализ временных рядов космических лучей / О.В. Мандрикова, **Т.Л. Заляев** // 17-я международная конференция «Цифровая обработка сигналов и ее применение DSPA 2015». – 2015. – Т. 2 – С. 579–583.

29. Мандрикова О.В. и др. Анализ вариаций космических лучей на основе нейронной сети / О.В. Мандрикова, **Т.Л. Заляев**, Б.С. Мандрикова // Информационные технологии и нанотехнологии: сборник трудов ИТНТ-2018. – 2018. – С. 2906–2916.

Свидетельство о государственной регистрации программы для ЭВМ:

30. **Заляев Т.Л.**, Мандрикова О.В. Свидетельство об отраслевой регистрации разработки № 12205 «Cosmic Rays Analysis Tool», дата регистрации 10.04.2016.

Заляев Тимур Ленарович

**МЕТОД И АЛГОРИТМЫ АНАЛИЗА ДАННЫХ КОСМИЧЕСКИХ ЛУЧЕЙ
В ЗАДАЧАХ ВЫДЕЛЕНИЯ СПОРАДИЧЕСКИХ ЭФФЕКТОВ**

*Автореферат
диссертации на соискание ученой степени
кандидата технических наук*

В авторской редакции
Набор текста Т.Л. Заляев
Верстка, оригинал-макет Е.Е. Бабух

Подписано в печать 27 декабря 2018 г.
Формат 60*84/16. Печать цифровая. Гарнитура Times New Roman
Авт. л. 1,41. Уч.-изд. л. 1,63. Усл. печ. л. 1,39
Тираж 100 экз. Заказ № 12

Издательство
Камчатского государственного технического университета

Отпечатано участком оперативной полиграфии издательства КамчатГТУ
683003, г. Петропавловск-Камчатский, ул. Ключевская, 35