

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ НАУКИ  
ИНСТИТУТ КОСМИЧЕСКИХ ИССЛЕДОВАНИЙ  
РОССИЙСКОЙ АКАДЕМИИ НАУК  
(ИКИ РАН)

УДК 528.85/.87(15)

Номер государственной регистрации 122042500019-6

УТВЕРЖДАЮ  
И.о. директора ИКИ РАН  
Федерального государственного бюджетного  
учреждения науки Института космических  
исследований  
Российской академии наук  
Ученый корреспондент РАН



А.А. Петрукович

10 декабря 2023 г.

ОТЧЕТ

О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

Большие данные в космических исследованиях: астрофизика, солнечная система, геосфера

по теме:

КОСМОС-Д

FFWG-2022-0012

(промежуточный, этап 2023/2)

Научный руководитель

канд. техн. наук

А.А. Прошин

Москва

2023

## СПИСОК ИСПОЛНИТЕЛЕЙ

Руководитель темы, ведущий научный сотрудник, зав. лаб., к.т.н.	<hr/>	Прошин А.А. (раздел 1-3, введение, заключение)
Ответственный исполнитель, ведущий научный сотрудник д.ф.-м.н.	<hr/>	Лукьянова Р.Ю. (раздел 1)
Ответственный исполнитель, старший научный сотрудник к.ф.-м.н.	<hr/>	Мещеряков А.В. (раздел 2)
Ответственный исполнитель, ведущий научный сотрудник, зав. лаб., к.т.н.	<hr/>	Прошин А.А. (раздел 3)

Исполнители отдела 54

Зав. лабораторией

А.Т. Янаков  
(раздел 1)

\_\_\_\_\_

Подпись, дата

старший научный  
сотрудник,

А.М. Мерзлый  
(раздел 1)

\_\_\_\_\_

Подпись, дата

научный сотрудник, к.н.

В.А. Ожередов  
(раздел 1)

\_\_\_\_\_

Подпись, дата

ведущий инженер

О.В. Никифоров  
(раздел 1)

\_\_\_\_\_

Подпись, дата

Исполнители отдела 52

старший научный  
сотрудник, к.ф.-м.н.

\_\_\_\_\_

Подпись, дата

Р.А. Буренин  
(раздел 2)

инженер

\_\_\_\_\_

Подпись, дата

С.А. Воскресенская  
(раздел 2)

младший научный  
сотрудник

\_\_\_\_\_

Подпись, дата

И.А. Зазнобин  
(раздел 2)

инженер

\_\_\_\_\_

Подпись, дата

М.И. Бельведерский  
(раздел 2)

специалист

\_\_\_\_\_

Подпись, дата

А.Д. Самородова  
(раздел 2)

Исполнители отдела 56

Зав. отделом, д.т.н.

Е.А. Лупян  
(раздел 3)

\_\_\_\_\_  
Подпись, дата

ведущий научный  
сотрудник, к.ф.-м.н.

А.А. Мазуров  
(раздел 3)

\_\_\_\_\_  
Подпись, дата

Зав. лабораторией, к.ф.-м.н.

В.Н. Пырков  
(раздел 3)

\_\_\_\_\_  
Подпись, дата

старший научный  
сотрудник, зав. лаб.,к.т.н.

И.В. Балашов  
(раздел 3)

\_\_\_\_\_  
Подпись, дата

старший научный  
сотрудник, зав. лаб.,к.т.н.

А.В. Кашницкий  
(раздел 3)

\_\_\_\_\_  
Подпись, дата

старший научный  
сотрудник, зав. лаб.,к.т.н.

М.А. Бурцев  
(раздел 3)

\_\_\_\_\_  
Подпись, дата

старший научный  
сотрудник, к.ф.-м.н.

Ю.С. Крашенинникова  
(раздел 3)

\_\_\_\_\_  
Подпись, дата

старший научный  
сотрудник, к.т.н.

В.А. Толпин  
(раздел 3)

\_\_\_\_\_  
Подпись, дата

старший научный  
сотрудник, к.т.н.

И.А. Уваров  
(раздел 3)

\_\_\_\_\_  
Подпись, дата

научный сотрудник, к.т.н.

Д.А. Кобец  
(раздел 3)

\_\_\_\_\_  
Подпись, дата

младший научный  
сотрудник

А.А. Бриль  
(раздел 3)

\_\_\_\_\_  
Подпись, дата

младший научный  
сотрудник, к.т.н.

А.М. Константинова  
(раздел 3)

\_\_\_\_\_  
Подпись, дата

главный специалист

А.Ю. Дегай  
(раздел 3)

\_\_\_\_\_  
Подпись, дата

главный специалист, к.ф.-  
м.н.

А.А. Козочкина  
(раздел 3)

\_\_\_\_\_  
Подпись, дата

ведущий инженер

Ю.В. Артамонова  
(раздел 3)

\_\_\_\_\_  
Подпись, дата

ведущий инженер

А.М. Матвеев  
(раздел 3)

\_\_\_\_\_  
Подпись, дата

ведущий инженер

М.В. Радченко  
(раздел 3)

\_\_\_\_\_  
Подпись, дата

ведущий инженер	_____	О.А. Суднева (раздел 3)
	Подпись, дата	
ведущий конструктор	_____	В.В. Марченков (раздел 3)
	Подпись, дата	
инженер	_____	А.А. Антошкин (раздел 3)
	Подпись, дата	
инженер	_____	Е.Е. Волкова (раздел 3)
	Подпись, дата	
инженер	_____	М.В. Врублевский (раздел 3)
	Подпись, дата	
инженер	_____	П.В. Денисов (раздел 3)
	Подпись, дата	
инженер	_____	Д.С. Дианова (раздел 3)
	Подпись, дата	
инженер	_____	Е.С. Ёлкина (раздел 3)
	Подпись, дата	
инженер	_____	Д.В. Лозин (раздел 3)
	Подпись, дата	
инженер	_____	А.Ю. Полецкая (раздел 3)
	Подпись, дата	
инженер	_____	Б.П. Руткевич (раздел 3)
	Подпись, дата	
инженер	_____	К.С. Сенько (раздел 3)
	Подпись, дата	
инженер, к.г.н.	_____	К.А. Трошко (раздел 3)
	Подпись, дата	
инженер	_____	В.Н. Черных (раздел 3)
	Подпись, дата	
старший лаборант	_____	О.Е. Степанченко (раздел 3)
	Подпись, дата	
старший лаборант	_____	Д.А. Юдин (раздел 3)
	Подпись, дата	
нормоконтролер	_____	Е.А. Лупян
	Подпись, дата	

## РЕФЕРАТ

Отчет 78 стр., 48 рис., 33 источника.

НАСТОЯЩИЙ ОТЧЕТ ПОСВЯЩЕН ОПИСАНИЮ ОСНОВНЫХ РЕЗУЛЬТАТОВ РАБОТ, ПОЛУЧЕННЫХ В 2023 ГОДУ В РАМКАХ ТЕМЫ "КОСМОС-Д" (№ 122042500019-6), ВЫПОЛНЯЮЩЕЙСЯ В РАМКАХ ГОСУДАРСТВЕННОГО ЗАДАНИЯ ИНСТИТУТОМ КОСМИЧЕСКИХ ИССЛЕДОВАНИЙ РОССИЙСКОЙ АКАДЕМИИ НАУК.

Тема «Космос-Д» «Большие данные в космических исследованиях: астрофизика, солнечная система, геосфера» сформирована для развития новой парадигмы «больших данных» в науках о космосе и получения качественных новых результатов на этой основе. Тема впервые сформирована в 2021 г в рамках программы Министерства высшего образования и науки по созданию новых лабораторий, на основе задела, созданного в других «профильных» темах государственного задания ИКИ РАН.

В ОТЧЕТЕ ПРИВОДЯТСЯ ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТ, ЗАПЛАНИРОВАННЫХ НА 2023 ГОД ПО СЛЕДУЮЩИМ НАПРАВЛЕНИЯМ:

Развитие методов моделирования высокоширотной ионосферы и анализ массивов спутниковых и наземных данных измерений; построение численных и гибридных моделей распределения электронной концентрации в полярной ионосфере в зависимости от изменяющихся гелио-геофизических условий

Применение методов машинного обучения и глубокого обучения для обработки и анализа больших массивов астрономических данных в задачах наблюдательной астрофизики.

Разработка новых подходов и методов организации ведения и работы со сверхбольшими архивами данных спутниковых наблюдений Земли

### КЛЮЧЕВЫЕ СЛОВА

СВЕРХБОЛЬШИЕ АРХИВЫ СПУТНИКОВЫХ ДАННЫХ, ЭФФЕКТИВНАЯ ОРГАНИЗАЦИЯ ХРАНЕНИЯ СПУТНИКОВЫХ ДАННЫХ, СИСТЕМЫ ПАРАЛЛЕЛЬНОЙ ОБРАБОТКИ СПУТНИКОВЫХ ДАННЫХ, ОРГАНИЗАЦИЯ ОБРАБОТКИ СПУТНИКОВЫХ ДАННЫХ, КОНТРОЛЬ ФУНКЦИОНИРОВАНИЯ, МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ, ДОСТУП К АРХИВАМ СПУТНИКОВЫХ ДАННЫХ, АНАЛИЗ ДОЛГОВРЕМЕННЫХ СЕРИЙ НАБЛЮДЕНИЙ, МОНИТОРИНГ РАСТИТЕЛЬНОСТИ, МОНИТОРИНГ ВУЛКАНИЧЕСКОЙ АКТИВНОСТИ  
ИОНОСФЕРНО-МАГНИТОСФЕРНОЕ ВЗАИМОДЕЙСТВИЕ, ВЫСОКОШИРОТНАЯ ИОНОСФЕРА, ИОНЗАЦИЯ, АВРОРАЛЬНЫЕ ДУГИ, ОПТИЧЕСКИЕ МЕТОДЫ, ОБРАБОТКА МАССИВОВ СПУТНИКОВЫХ ДАННЫХ, ЧИСЛЕННОЕ МОДЕЛИРОВАНИЕ  
МАШИННОЕ ОБУЧЕНИЕ, ГЛУБОКОЕ ОБУЧЕНИЕ, ОБУЧЕНИЕ С ПОДКРЕПЛЕНИЕМ, ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ, КОСМИЧЕСКАЯ ОБСЕРВАТОРИЯ “СПЕКТР-РГ”, КОСМИЧЕСКАЯ ОБСЕРВАТОРИЯ “ПЛАНК”, АТАКАМСКИЙ КОСМОЛОГИЧЕСКИЙ ТЕЛЕСКОП, ОБСЕРВАТОРИЯ “SQUARE KILOMETRE ARRAY (SKA)”, РЕНТГЕНОВСКАЯ АСТРОНОМИЯ, НАБЛЮДАТЕЛЬНАЯ АСТРОФИЗИКА, ЦИФРОВЫЕ ОБЗОРЫ НЕБА, КОМБИНАТОРНАЯ ОПТИМИЗАЦИЯ, ОПТИМИЗАЦИЯ МАРШРУТОВ, СМЕШАНО-ЦЕЛОЧИСЛЕННОЕ ПРОГРАММИРОВАНИЕ, МЕТАЭВРИСТИКИ, НЕЙРОЭВРИСТИКИ

## СОДЕРЖАНИЕ

СПИСОК ИСПОЛНИТЕЛЕЙ .....	2
РЕФЕРАТ .....	7
СОДЕРЖАНИЕ .....	8
ВВЕДЕНИЕ .....	10
ГЛАВА 1 Развитие методов моделирования высокоширотной ионосферы и анализ массивов спутниковых и наземных данных измерений; построение численных и гибридных моделей распределения электронной концентрации в полярной ионосфере в зависимости от изменяющихся гелио-геофизических условий .....	12
1.1 Введение .....	12
1.2 Развитие численных моделей ионосферы, параметризация и валидация моделей с использованием массивов данных мировой сети наземных и спутниковых измерений.....	13
1.3 Разработка подходов к построению ассимиляционных блоков для ионосферного моделирования .....	14
1.4 Разработка прогноза состояния характеристик полярной ионосферы .....	16
1.5 Развитие методик комплексных орбитальных измерений, нацеленных на прямую диагностику характеристик плазмы и дистанционную глобальную и локальную оптическую диагностику распределений интенсивности авроральных эмиссий в видимой и ВУФ-областях спектра .....	18
1.6 Исследование статистических связей глобальных и локальных геомагнитных вариаций с процессами в ионосфере, магнитосфере и солнечном ветре с целью выявления новых закономерностей и прогноза экстремальных событий .....	22
1.7 Создание деревьев принятия решений для классификации изменений состояния организма в связи с воздействием на него факторов космической и земной погоды на основе алгоритмов дифференциальной эволюции.....	25
1.8 Разработка методов диагностики источников локальной ионизации, включая малые космические объекты, на основе анализа больших массивов данных радарных наблюдений .....	28
1.9 Заключение .....	31
ГЛАВА 2 Применение методов машинного обучения и глубокого обучения для обработки и анализа больших массивов астрономических данных в задачах наблюдательной астрофизики .....	33
2.1 Введение .....	33
2.2 Исследование, разработка и применение алгоритмов глубокого и классического машинного обучения для решения задач детектирования, кросс-отождествления, классификации, инкрементального моделирования и определения физических характеристик астрономических объектов по данным многоволновых обзоров всего неба .....	33
2.3 Исследование и разработка моделей обучения с подкреплением для задач комбинаторной оптимизации .....	35
2.4 Создание систем и технологий для обработки и анализа больших астрономических данных.....	38
2.7 Заключение .....	40
ГЛАВА 3 Разработка новых подходов и методов организации ведения и работы со сверхбольшими архивами данных спутниковых наблюдений Земли .....	42
3.1 Введение .....	42
3.2 Новые подходы организации автоматизированного ведения архивов и предоставления доступа к ним, включая организацию интерактивной и автоматизированной распределенной обработки предоставляемой ими информации .....	43



3.2.1 Анализ различных подходов к оптимизации хранения временных рядов данных ДЗЗ .....	44
3.2.2 Организация работы с данными архивов ЦКП «ИКИ-Мониторинг» на основе использования интерактивной среды разработки Jupyter Notebook .....	54
3.2.3 Развитие нейросетевых подходов для автоматического построения масок облачности и теней на спутниковых изображениях приборов серии КМСС .....	58
3.2.4 Разработка архитектуры и инструментов для модернизации распределенной многопоточной обработки спутниковых данных .....	61
3.2.5 Развитие системы ведения архивов и инструментов для контроля ее функционирования .....	63
3.3 Развитие с использованием созданных методов и подходов различных специализированных научных информационных систем, создаваемых и поддерживаемых в ИКИ РАН .....	65
3.3.1 Расширение функционала подсистемы анализа временных серий данных в информационных системах семейства «Созвездие-ВЕГА» .....	65
3.3.2 Развитие инструментария для системы доступа к временным рядам данных в системе мониторинга вулканической активности Камчатки и Курил (ИС VSV) .....	67
3.4 Заключение .....	70
ГЛАВА 4 Наиболее значимые результаты исследований ИКИ РАН .....	71
4.1 Вариации условий природной среды во время геомагнитной бури, приведшие к потере серии спутников Starlink .....	71
ЗАКЛЮЧЕНИЕ .....	72
ПУБЛИКАЦИИ .....	73

## ВВЕДЕНИЕ

Институт космических исследований Российской академии наук проводит научно-исследовательские и опытно-конструкторские работы в соответствии с государственным заданием ЧАСТЬ 2: ГОСУДАРСТВЕННЫЕ РАБОТЫ по следующим научным направлениям:

- Фундаментальные и прикладные научные исследования в области астрофизики и радиоастрономии;
- Фундаментальные и прикладные научные исследования в области Физики космической плазмы, энергичных частиц, Солнца и солнечно-земных связей;
- Фундаментальные и прикладные научные исследования планет и малых тел Солнечной системы;
- Фундаментальные и прикладные научные исследования планеты Земля;
- Фундаментальные и прикладные научные исследования в области механики, систем управления и информатики;
- Развитие исследовательской, конструкторской, опытно-экспериментальной базы научного космического приборостроения и методов экспериментальной физики.

Эти направления НИР и ОКР соответствуют следующим направлениям фундаментальных исследований, указанными в Программе фундаментальных исследований в Российской Федерации на долгосрочный период 2021-2030 гг.,, утвержденной распоряжением Правительства РФ от от 31 декабря 2020 г. № 3684-р

п/п	Направление фундаментальных исследований	Номер направления в «Программе»
1	Пламенные процессы в геофизике и астрофизике	1.3.4.3.
2	Фундаментальные проблемы распространения радиоволн, в том числе в ионосфере	1.3.6.4.
3	Происхождение, строение и эволюция Вселенной	1.3.7.1.
4	Физика галактик и межгалактической среды	1.3.7.2.
5	Развитие методов наземной и внеатмосферной астрономии	1.3.7.6.
6	Информационно-вычислительные системы и среды в науке и образовании	1.1.8
7	Алгоритмы и программные системы в космическом мониторинге Земли и экологии	1.1.7.6
8	Метрология и цифровизация в науках о Земле	1.5.12

Работы по теме сфокусированы на новой парадигме «больших данных». В космических исследованиях переход к новой парадигме обусловлен беспрецедентным ростом количества и разнообразия данных наблюдений. Существующие и будущие дальние космические миссии и системы наблюдения Земли из космоса становятся все более сложными, представляя высокое пространственное/временное/спектральное разрешение. Такое развитие экспериментальной базы требует адекватного усовершенствования подходов к анализу и моделированию соответствующих физических процессов. Перспективным представляется развитие интеллектуальной автоматизированной аналитики, усовершенствование методов моделирования с применением ассимиляции и глубокого машинного обучения, использование новых подходов организации хранения и управления данными (в том числе с использованием

облачных технологий и высокопроизводительных вычислительных систем для масштабируемого анализа). Также, результативными могут быть и новые применения традиционных методов анализа с учетом специфики больших данных.

В работах по теме «Космос-Д» предусматривается создание новых подходов к анализу данных космических экспериментов и мониторинга в астрофизике, солнечно-земной физике, физике солнечной системы, дистанционном зондировании Земли. Работы по этим направлениям объединяются разработкой уникальной единой системы хранения и анализа информации.

Работы по теме ведутся по следующим основным направлениям:

- Развитие методов моделирования высокоширотной ионосферы и анализ массивов спутниковых и наземных данных измерений; построение численных и гибридных моделей распределения электронной концентрации в полярной ионосфере в зависимости от изменяющихся гелио-геофизических условий;
- Применение методов машинного обучения и глубокого обучения для обработки и анализа больших массивов астрономических данных в задачах наблюдательной астрофизики;
- Разработка новых подходов и методов организации ведения и работы со сверхбольшими архивами данных спутниковых наблюдений Земли.

Представлению результатов, полученных в данных направлениях, посвящены соответствующие главы настоящего отчета.

# **ГЛАВА 1 Развитие методов моделирования высокоширотной ионосферы и анализ массивов спутниковых и наземных данных измерений; построение численных и гибридных моделей распределения электронной концентрации в полярной ионосфере в зависимости от изменяющихся гелио-геофизических условий**

## **1.1 Введение**

На 2023 г. было запланировано выполнение комплекса работ по решению следующих задач: (\*) Развитие численных моделей ионосферы, параметризация и валидация моделей с использованием массивов данных мировой сети наземных и спутниковых измерений; (\*) Разработка подходов к построению ассимиляционных блоков для ионосферного моделирования; (\*) Разработка прогноза состояния характеристик полярной ионосферы; (\*) Развитие методик комплексных орбитальных измерений, нацеленных на прямую диагностику характеристик плазмы и дистанционную глобальную и локальную оптическую диагностику распределений интенсивности авроральных эмиссий в видимой и ВУФ-областях спектра; (\*) Исследование статистических связей глобальных и локальных геомагнитных вариаций с процессами в ионосфере, магнитосфере и солнечном ветре с целью выявления новых закономерностей и прогноза экстремальных событий; (\*) Создание деревьев принятия решений для классификации изменений состояния организма в связи с воздействием на него факторов космической и земной погоды на основе алгоритмов дифференциальной эволюции; (\*) Разработка методов диагностики источников локальной ионизации, включая малые космические объекты, на основе анализа больших массивов данных радарных наблюдений.

Актуальность этих исследований обусловлена необходимостью изучения физических механизмов солнечно-земных связей и разработкой методов прогнозирования космической погоды, от состояния которой критически зависит современная высокотехнологичная инфраструктура. Исследования базируются на анализе больших массивов данных, полученных различными космическими миссиями, с привлечением данных измерений наземных приборов. В рамках парадигмы «больших данных» экспериментальные данные комплексуются с модельными расчетами.

В 2023 году исследования соответствовали плану и были направлены на усовершенствование ионосферных моделей и их использование для определения условий развития возмущений космической погоды, расширение и наполнение новыми данными атласа-справочника по классификации и определению авроральных структур по оптическим наблюдениям, изучение влияния продольных токов магнитосферного происхождения на локальные вариации электронной концентрации в высокоширотной ионосфере, анализ состояния ионосферы и термосферы, в частности, во время запусков спутников Старлинк, разработку алгоритмов для выявления факторов воздействия космической погоды на организм человека, а также оценку изменчивости региональной аэрозольной нагрузки по данным спутникового микроволнового зондирования. В данном разделе отчета описание результатов структурировано по тематическим пунктам плана, и каждой из задач соответствует свой подраздел.

Результаты, полученные в рамках выполненных работ, изложены в публикациях [1,2, 8-11, 15, 20-22]. Количество и рейтинг публикаций соответствуют показателям, запланированным на 2023 год.

## 1.2 Развитие численных моделей ионосферы, параметризация и валидация моделей с использованием массивов данных мировой сети наземных и спутниковых измерений

Изменчивость параметров высокоширотной ионосферы и ее зависимость от внешних условий – солнечного ветра и межпланетного магнитного поля (ММП), критически усложняют задачу построения моделей для этой области. Полярная ионосфера по силовым линиям геомагнитного поля связана с самыми внешними границами магнитосферы и непосредственно с ММП. Электрическое поле магнитосферной конвекции определяет особенности циркуляции плазменных трубок в высокоширотной ионосфере. Горизонтальный дрейф плазмы на высотах F-слоя непосредственно связан с магнитосферными электрическими полями, а ее плотность – с плазменным слоем в хвосте магнитосферы, который является источником высыпаний энергичных заряженных частиц. Это приводит к локальному увеличению ионизации в зоне аврорального овала, особенно на его ночной стороне.

В F-слое полярной ионосферы присутствуют не только области с увеличенной плотностью плазмы, но области ее истощения, когда в неосвещенной ионосфере медленная конвекция плазмы создает полярные дыры или полости с пониженной концентрацией плазмы. Ожидаемое расположение полярной дыры во время медленной конвекции находится в ночном секторе местного магнитного времени (MLT) на магнитной широте 70–80°. В этой области составляющая скорости антисолнечной конвекции будет примерно равна скорости коротации и противоположна ей, и поэтому плазма движется чрезвычайно медленно. В рамках решения задачи данного подраздела по численной модели трехмерного распределения электронной концентрации в F-области полярной ионосферы (описание модели PFIM можно найти в отчете за 2020 г.) были выявлены параметры ММП, при которых наиболее вероятно образование полярных дыр. Критерий наличия полярной дыры, где плотность электронов на несколько порядков ниже фоновых значений:  $NmF2 < 10^5 \text{ см}^{-3}$ .

На рисунке 1.2.1 представлены распределения максимума электронной концентрации  $NmF2$  в зимних условиях при различной полярности вертикальной компоненты ( $B_z$ ) ММП, а именно при  $B_z = 0, -5$  и  $+5$  нТл, азимутальная компонента ( $B_y$ ) отсутствует. Если ММП направлено к югу ( $B_z = -5$  нТл), истощения плазмы не наблюдается. Полярная дыра появляется при повороте ММП к северу. При этом при  $B_z = 0$  ее площадь даже несколько больше, чем при  $B_z = +5$  нТл.

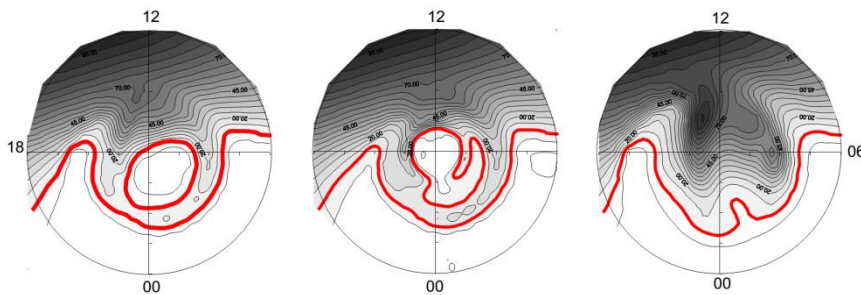


Рисунок 1.2.1 — Распределение  $NmF2 \times 10^5 \text{ см}^{-3}$  в координатах MLat-MLT зимой при ММП  $B_z = 0, -5$  и  $+5$  нТл,  $B_y = 0$  нТл. Другие параметры:  $UT = 5$ ,  $V_{sw} = 350 \text{ км/с}$ ,  $K_p = 1$ ,  $F10.7 = 100 \text{ sfu}$ .

Центром каждого графика является магнитный полюс, низкоширотная граница на 50° MLat.

Выделена изолиния  $10^5 \text{ см}^{-3}$

Влияние направления ММП  $B_y$  проявляется в изменении формы дыры. На рисунке 1.2.2 (а, б) показаны карты  $NmF2$  в зимних условиях при  $B_y = -5$  нТл и  $+5$  нТл, когда вертикальная компонента  $B_z$  направлена на север. При отрицательном  $B_y$  дыра локализована вокруг магнитного полюса и окружена гребнем ионизации. При положительном  $B_y$  истощение плазмы происходит не только в приполюсной области, но и на авроральных широтах. Здесь полярные дыры могут образовываться в вечерних условиях, попадая в замкнутые

конвекционные траектории. В этой области составляющая скорости антисолнечной конвекции примерно равна скорости коротации и противоположна ей. Поэтому плазма движется чрезвычайно медленно. При этом в спокойных геомагнитных условиях плазма истощается в ходе непрерывной рекомбинации.

На положение дыры оказывает некоторое влияние и суточная вариация проводимости, т.к. изменяется относительная площадь освещенной и неосвещенной частей ионосферы. На рисунке 1.2.2 (в, г) показаны карты  $NmF2$  при UT= 5 и 17 ч. При меньшей освещенности дыра, расположенная в секторе 00-06, расширяется до  $\sim 70^\circ$ . При большой освещенности дыра уменьшается и смещается от полюса в авроральные широты.

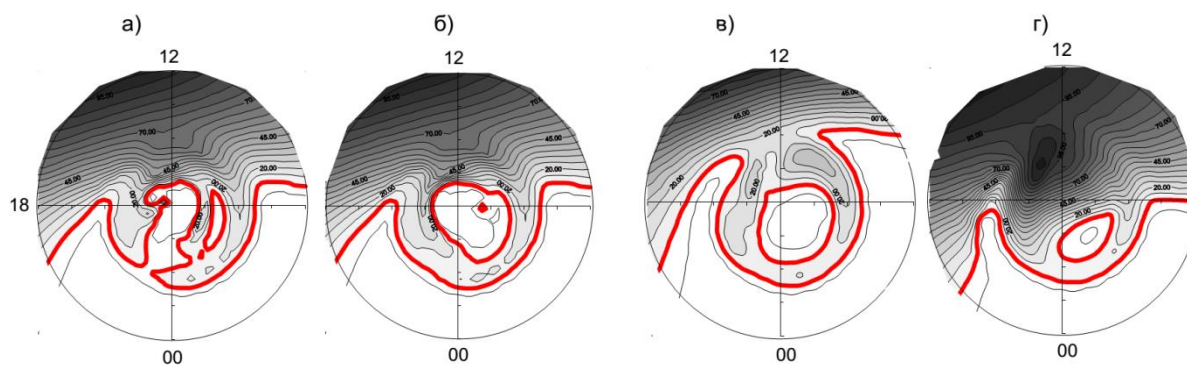


Рисунок 1.2.2 — Распределение  $NmF2 \times 10^5 \text{ см}^3$  при различных знаках азимутальной компоненты ММП (а)  $B_y = +5 \text{ нТл}$ ,  $B_z = +5 \text{ нТл}$ , (б)  $B_y = -5 \text{ нТл}$ ,  $B_z = +5 \text{ нТл}$ ; и в различные часы мирового времени (в) UT=5 и (г) UT=17 при ММП  $B_z = B_y = 0$

Таким образом, согласно модельным расчетам можно предсказать, что полярные дыры наиболее часто образуются в условиях неосвещенной ионосферы при северном направлении ММП. Эти структуры связаны с медленной антисолнечной конвекцией, когда плазма, истощается в отсутствие каких-либо источников ионизации. Диаметр дыры может достигать 2000 км. При положительной азимутальной компоненте ММП ( $B_y > 0$ ) в предполуденном секторе MLT область истощения плазмы расширяется на авроральный овал и сливается со среднеширотной ионосферой. Этому способствует смещение языка ионизации на утреннюю сторону, из-за чего расширяется область стагнации, где составляющая скорости антисолнечной конвекции примерно равна обратной направленной скорости, с которой ионосферная плазма коротатирует с Землей. При  $B_y < 0$  дыра локализована у геомагнитного полюса и окружена гребнем ионизации.

Таким образом, результаты модельных расчетов позволяют прогнозировать положение полярной полости и степень истощения плазмы в зависимости от параметров ММП.

### 1.3 Разработка подходов к построению ассимиляционных блоков для ионосферного моделирования

Ассимиляция данных рассматривается как процесс объединения данных измерений с моделью для оценки состояний системы, которые не измеряются напрямую. Посредством ассимиляции можно повысить эффективность ограниченных измерений за счет использования модели и одновременно повысить точность модельных оценок с помощью измерения. В общем случае, при использовании ассимиляции проводится корректировка расчетов численной модели, основанной на физических принципах, с учетом оперативно поступающих на вход модели массивов экспериментальных данных о выходных параметрах. Известно, что такой подход дает хорошие результаты при модельных прогнозах погоды. Метод решения задач усвоения, основанный на вариационных принципах (поиск минимума некоторого функционала, описывающего близость модельного решения и данных измерений в

некоторой метрике), сравнительно хорошо разработан и успешно применяется на практике. Также применяется динамико-стохастический метод с использованием фильтра Калмана, который основан на статистической теории оценивания и фильтрации процессов на фоне шума с известными статистическими свойствами.

Однако, испытания показали, что при использовании таких методов в блоках численной модели электронной концентрации ( $Ne$ ) в высокоширотной ионосфере возникают принципиальные трудности. Применению ассимиляционных методов препятствует то, что высокоширотная ионосфера демонстрирует достаточно быструю реакцию на изменения ее движущих сил и управляющих параметров (т.е. зависящих от времени потоков солнечного УФ излучения, уровня геомагнитной активности, параметров нейтральной атмосферы, обычно рассчитываемых по эмпирической модели, и траекторий конвекции плазмы за счет  $E \times B$  дрейфа), так что предыдущее состояние не дает эффективной информации о будущем состоянии системы. Кроме того, проблемы есть и в логистическом плане, т.к. в высокоширотной ионосфере, где пространственные градиенты  $Ne$  наибольшие, плотность измерительных станций наименьшая, а оперативность поставки данных низкая. В настоящее время эти проблемы только обострились. Определенные результаты были достигнуты при введении ассимиляционных блоков в климатические модели типа эмпирической модели IRI, которая описывает усредненное состояние среднеширотной ионосферы. При этом для решения прикладных задач использовались упрощенные процедуры ассимиляции экспериментальных данных о полном электронном содержании (ПЭС). В высоких широтах спутники системы ГНСС видны под высокими углами и редко проходят близко к зениту, что вместе с малоступностью зарубежной сети наземных приемников уменьшает количество возможных данных ПЭС. Что касается более детальных характеристик ионосферы, таких как вертикальный профиль  $Ne$  или максимум плотности плазмы  $NmF2$ , то на практике применение стандартных методов ассимиляции затруднительно в связи с малым числом точек инструментальных измерений. Скорее, можно говорить, что имеет смысл параметризация численной физико-математической модели по измерениям в отдельных географических локациях.

Экспериментальные данные с высоким временным и вертикальным разрешением можно получить с помощью ионозондов и радаров некогерентного рассеяния. В арктической зоне имеются только несколько зарубежных радаров такого типа, к тому же работающих не постоянно, а по сеансам. Доступ к многим радарным данным ограничен. Наибольшее количество информации в открытую базу данных Madrigal поставляется после сеансов работы радара ESR системы EISCAT (геогр. коорд.  $78^\circ N$ ,  $16^\circ E$ ; геомагн. коорд.  $75^\circ$ ,  $113^\circ$ ). Локация ESR позволяет наблюдать ионосферу на границе полярной шапки. Были отобраны сеансы в период 2007-2014 г., когда антенна была направлена в зенит при спокойных геомагнитных условиях и затем проведено усреднение по сезонам. Для этих же дней был проведен модельный расчет. Сравнение наблюдаемых и модельных параметров  $NmF2$  и высоты максимума слоя F2 ( $HmF2$ ) приведено на рисунке 1.3.1. В целом, сезонные тенденции изменения параметров по радару и по модели совпадают, но модель регулярно занижает оценки. Для выявления экспериментального фонового уровня ионосферных параметров на широтах аврорального овала, были отобраны по тем же дням, что и ESR, и усреднены данные ионозонда обс. Соданкюля (геогр. коорд.  $78^\circ N$ ,  $16^\circ E$ ; геомагн. коорд.  $64^\circ$ ,  $119^\circ$ ). Величина критической частоты  $f_oF2$  (Гц) по ионозонду пересчитывалась в величину  $NmF2$  ( $m^{-3}$ ). Результат сравнения представлен на правой панели рисунка 1.2.1. В этом случае также модельные значения меньше измеренных, хотя общие сезонные тенденции совпадают. Результаты указывают на то, что модель имеет процентное отклонение до 30% от измерений  $Ne$ . Проведена калибровка модели по усредненным экспериментальным данным, в результате чего устранено занижение  $NmF2$ .

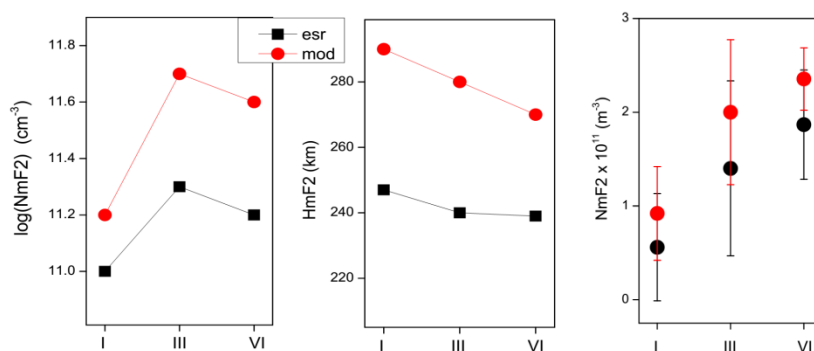


Рисунок 1.3.1 – Сезонные (зима, равноденствие и лето) значения  $NmF2$  и  $HmF2$  по данным измерений радара и модельным расчетам (левая и средняя панели), а также по данным ионозонда SGO и модели (правая панель)

#### 1.4 Разработка прогноза состояния характеристик полярной ионосферы

На текущем этапе работы по данной тематике были посвящены применению трансionoсферного зондирования на основе ГНСС-технологий для оперативной оценки параметров высокоширотной ионосферы. Соответствующие массивы измерений и экспериментальные результаты были получены во время экспедиции на архипелаге Земля Франца-Иосифа (ЗФИ) в марте 2021 г. Обработка данных и анализ полученных результатов были выполнены в 2023 г. Актуальность задачи обусловлена влиянием полярной ионосферы на функционирование высокоширотных средств связи, навигацию и радиолокацию. Возникающие магнитно-ионосферные возмущения приводят к изменению условий распространения радиоволн различного диапазона в арктической зоне. Ухудшение точностных характеристик радиотехнических средств различного назначения при неблагоприятных ионосферных условиях требует достоверного знания параметров среды распространения радиосигнала для разработки мер компенсации или уменьшения влияния возмущений полярной ионосферы. Изменения полного электронного содержания (ПЭС) вызывают пропорциональные изменения фазы и искажения амплитуды радиосигнала. Чем точнее будут определены значения ПЭС, тем точнее можно прогнозировать задержки радиосигнала, вызванные влиянием ионосферы.

В настоящее время распространенным способом, обеспечивающим потребителей информацией о пространственно-временном распределении ПЭС, является моделирование с помощью эмпирических или численных ионосферных моделей. Однако при возникновении возмущений при повышенной солнечной активности прогнозирование ПЭС становится менее точными. В этом случае полезную информацию можно получить по данным двухчастотных измерений, которые позволяют применять дисперсионный метод, основанный на оценке фазового набега и группового запаздывания навигационных сигналов в ионосфере на двух частотах.

Измерения на территории ЗФИ выполнялись методом трансionoсферного зондирования с использованием приемника GPStation-6 NovAtel. В результате наблюдений были получены данные мониторинга ионосферы за пять дней, которые содержали следующие основные параметры: время измерения в формате GPS Time; идентификатор спутника; измеренное значение ПЭС для каждого видимого в это время спутника; зенитное расстояние и азимут на каждый спутник; измеренные псевдодалности на каждый спутник. Пример ПЭС показан на рисунке 1.4.1.



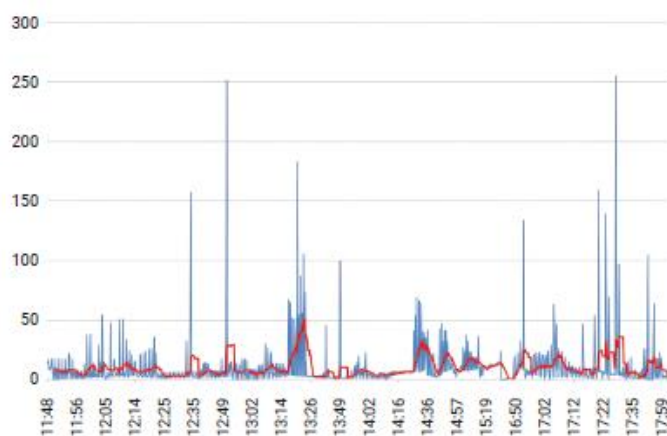


Рисунок 1.4.1 - Данные наблюдений ПЭС (синий цвет) и скользящее среднее (красный) за период 12 — 18 UT, 20.03.2021

Для апробации методики анализа полученные с высокой дискретностью данные были обработаны с помощью специальных алгоритмов, направленных на обеспечение полной автоматизации обработки непосредственно на станции. Из файлов формата \*.gps, в которых представлены результаты наблюдений, в результате первичной обработки получены измерения наклонного ПЭС, зенитные расстояния и азимуты на все спутники, для которых выполнялись измерения. По известным координатам фазового центра антенны приемника и фазовых центров антенн спутников на моменты измерений, определялись векторы, вдоль которых измерено ПЭС. Множества векторов фазовых центров «приемник — спутник» использовались для оценки распределения измерений в пространстве и сравнения полученных ПЭС с положением аврорального овала.

Визуализация ПЭС в трехмерной проекции показывает пути распространения радионавигационного сигнала от спутника до приемника. Пример ПЭС на картографической подложке за 6-часовой интервал представлен на рисунке 1.4.2. Векторы ограничены тонким слоем ионосферы на высоте 300 км. Точки на конце отрезка характеризуют значение ПЭС на данный момент времени в соответствии с цветовой шкалой.

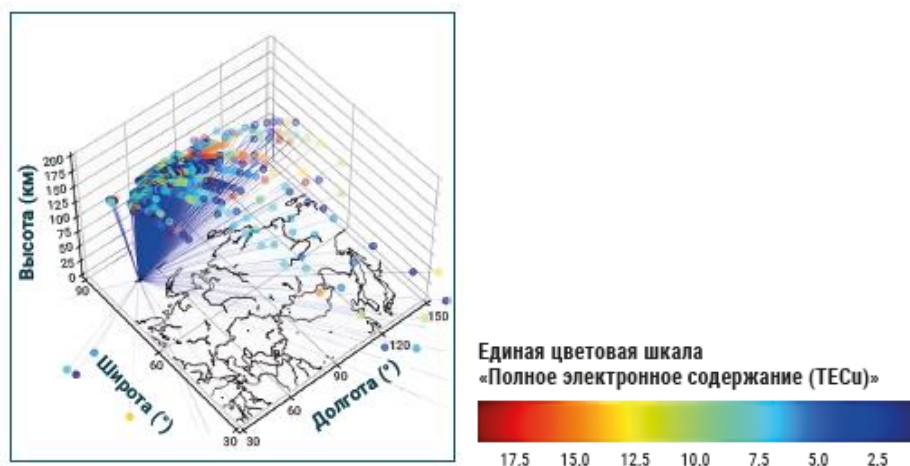


Рисунок 1.4.2 - Визуализация векторов «приемник — спутники» на промежутке 09-12 UT, 21.03.2021

Был создан алгоритм визуализации векторов в ПО Google Earth, где вместо картографической подложки используется изображение земной поверхности. Данный метод позволяет проводить сравнительный анализ пространственного распределения ПЭС с

данными снимков аврорального овала, соответствующим энергии высыпающихся частиц по наблюдениям спутника DMSP. Для планарного распределения ПЭС разработан алгоритм визуализации в виде тепловой карты. На рисунке 1.4.3 показан пример такой карты ПЭС; также представлено распределение высыпаний по DMSP в этом же географическом регионе. Можно видеть, что в областях с наибольшими ПЭС наблюдается и высокий уровень свечения ионосферы, а в областях с незначительным свечением — наименьшие значения ПЭС.

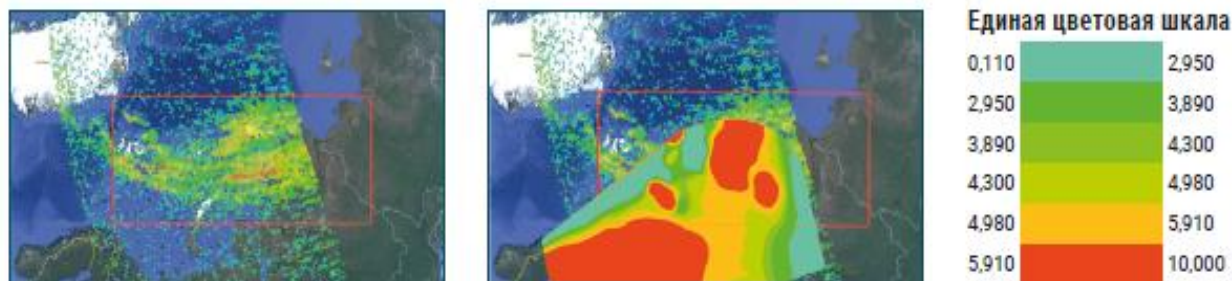


Рисунок 1.4.3 - ПЭС в виде тепловой карты и карта свечения DMSP 13:50 UT, 20.03.21.  
Красной рамкой выделена зона с наибольшими значениями ПЭС

Таким образом, по пункту плана “Разработка прогноза состояния характеристик полярной ионосферы” отработана технология наблюдений, оценки, прогноза и визуализации параметров высокоширотной ионосферы в районе ЗФИ в интересах решения научных и прикладных задач. Результаты будут использованы при организации постоянно действующей сети станций ионосферного мониторинга в реальном времени на основе мультиспектрных наблюдений ГНСС.

### **1.5 Развитие методик комплексных орбитальных измерений, нацеленных на прямую диагностику характеристик плазмы и дистанционную глобальную и локальную оптическую диагностику распределений интенсивности авроральных эмиссий в видимой и ВУФ-областях спектра**

В продолжение работы, начатой в 2022 г., разработана вторая часть аннотированного атласа примеров изображений структур эмиссий и характеристик плазмы в авроральном овале и субавроральной области во время событий, происходивших в разных секторах местного магнитного времени (MLT) в геомагнитно-возмущенных условиях. Был проанализирован большой объем информации нескольких космических миссий с разным разрешением и характеристиками аппаратуры и выявлены типичные сигнатуры суббуриевых событий. Атлас необходим для развития технологий комплексных экспериментов, нацеленных на создание системы картографирования и диагностики динамичных авроральных явлений в полярной ионосфере, например, с орбиты КА типа Молния и с полярной орбиты создаваемой российской орбитальной станции. Физические процессы, приводящие к авроральным структурам дециметрового масштаба и активным формам полярных сияний, до сих пор остаются недостаточно исследованными. Кроме того, авроральные формы отражают мгновенное поле градиентов электронной концентрации (Ne) в различных высотных слоях. Именно градиенты являются причиной искажений при распространении трансполярных радиосигналов.

Большинство примеров авроральных и аврора-подобных структур, каталогизированных в атласе, сопровождаются среднемасштабными изображениями частей аврорального овала и полярной шапки в разных секторах MLT, полученными в относительно близкое к рассматриваемым событиям время с орбит КА DMSP в авроральных эмиссиях в диапазоне вакуумного ультрафиолета (ВУФ) с помощью сканирующих по пространству

изображающих спектрографов SSUSI, а также прибора GUVI на борту КА TIMED. Примеры сопровождаются результатами спутниковых измерений характеристик потоков высыпавшихся частиц и градиентов магнитного поля, а также наблюдениями локальных градиентов Ne радаром некогерентного рассеяния. Рассмотрены следующие явления: (1) интенсификация сияний на полярной границе овала, известная как PBI (poleward boundary intensification); (2) авроральные лучевые структуры (flaming aurora - пылающие полярные сияния), возбуждаемые высыпаниями электронов разных энергий, и сопутствующее им усиление ионно-акустических линий NEIAL (naturally enhanced ion-acoustic lines); (3) аврора-подобные фрагментированные структуры FAE (Fragmented Aurora-like Emissions) на полярной стороне аврорального овала; (4) структуры STEVE (Strong Thermal Emission Velocity Enhancement) и часто сопровождающие их мелкомасштабных структур PF (Picket Fence) в субавроральной области на фазе восстановления суббури. К генерации этих структур причастны волны различных типов и турбулентность. Все авроральные и аврора-подобные структуры, так или иначе, отображают процессы динамики магнитосферы в областях ее проецирования в полярную ионосферу, а мелкомасштабные – отображают еще и процессы околоземной электродинамики и генерации неустойчивостей и волн в плазме.

*PBI.* На рисунке 1.5.1 показан пример сопряженных наблюдений во время события PBI четырех инструментов КА IMAGE, КА FAST и радара некогерентного рассеяния Sondrestrom. В высыпаниях электронов прослеживаются мелкомасштабные структуры. Данные радара ISR и сканирующего фотометра обнаруживают узкие вытянутые вдоль магнитного поля области с увеличенной концентрацией плазмы, которые согласуются с фотометрическими измерениями ионообразования и вытянутыми по полю электронными всплесками на орбите КА FAST.

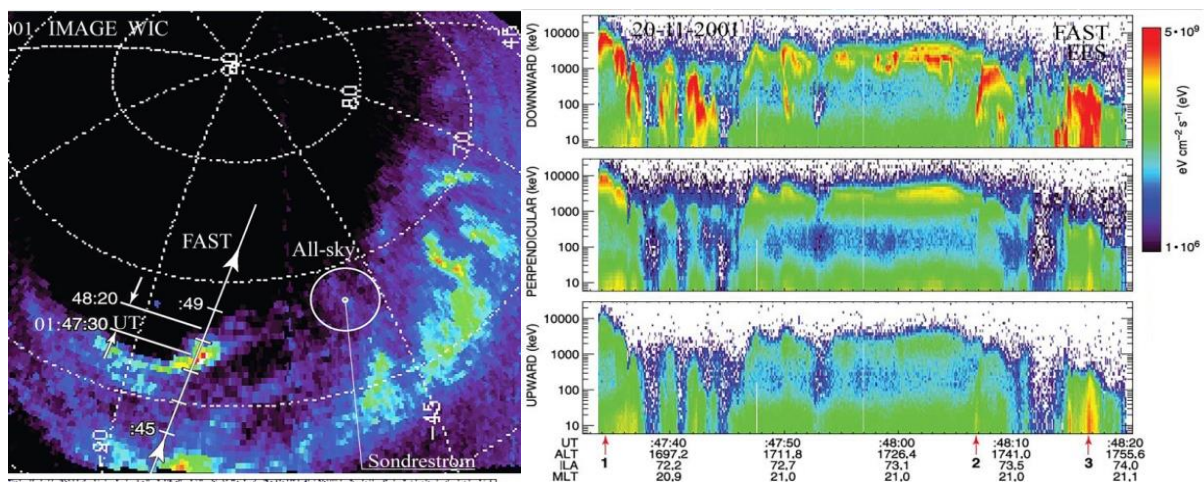


Рисунок 1.5.1 - Левая панель: оптическое изображение аврорального овала 20-11-2001 с орбиты КА IMAGE, на которое нанесен участок проекции орбиты КА FAST; белым кружком помечена проекция поля зрения (на высоте эмиссии) камеры всего неба исследовательского центра Sondrestrom вблизи радара некогерентного рассеяния ISR. Правая панель: спектрограммы электронов, полученные КА FAST

*Пылающая аврора и NEIAL.* В полярном сиянии, возбужденным потоками низкоэнергичных электронов, а также в оптических лучистых структурах обнаруживается усиление ионно-акустических линий NEIAL. Чтобы идентифицировать NEIALs в отраженных радиолокационных сигналах, используется их основная характеристика - отраженная энергия, которая на несколько порядков больше, чем фоновое некогерентное рассеяние. Один тип NEIAL встречается на относительно малых высотах в области F или ниже, при этом энергия распределяется между ионно-акустическими плечами (пример приведен на рисунок 1.5.2). Ко второму типу относятся NEIAL, которые происходят на относительно больших высотах (>600

км) и имеют большую асимметрию в усилении энергии эхо между двумя ионно-акустическими плечами.

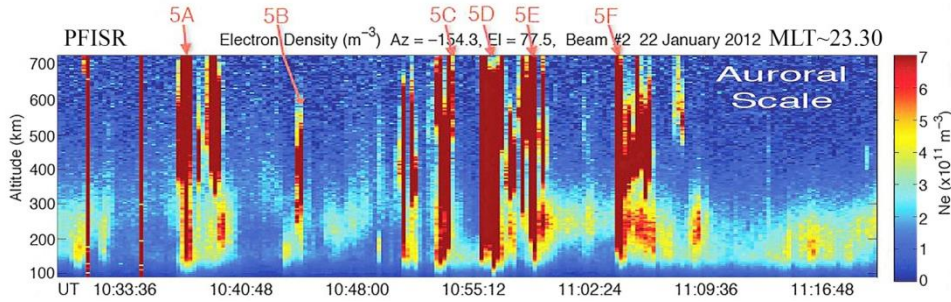


Рисунок 1.5.2 - Высотные структуры Ne в NEIALs, полученные радаром PFISR

*Аврора-подобные фрагментированные структуры FAE и STEVE.* К FAE относятся мелкие полосы в области полярной границы аврорального овала; характеристики красных SAR-дуг, и зеленых «протонных» дуг в субавроральной ионосфере, а также иногда сопровождающих их структуры STEVE (Strong Thermal Emission Velocity Enhancement - сильное увеличение скорости теплового излучения), структуры PF (Picket Fence), похожие на забор, встречающиеся в субавроральных областях и сопутствующие STEVE. Эти проявления авроральной активности фиксируются оптическими камерами всего неба как фрагменты зеленой эмиссии. FAE делятся на два типа: отдельные или неравномерно распределенные фрагменты свечения, и волнообразные структуры вблизи дуг полярного сияния. FAEs имеют горизонтальную, относительно малую (несколько км) протяженность излучения, вытянутую вдоль поля, и время жизни менее минуты. Отсутствие у них протяженности по магнитному полю указывает на иной механизм генерации, чем у полярных сияний (т.е. высыпание частиц). FAE сопровождаются повышенными электронными температурами в диапазоне высот 110-120 км и повышенными температурами ионов на высотах F-области. Было предложено два возможных типа волн, которые могут создавать FAE: электростатические ионные циклотронные волны и неустойчивость/волны Фарли–Бунемана в области сильных локальных токов. На рисунке 1.5.3 показана фрагментированная структура FAE. Эмиссия производится на больших высотах низкоэнергичными электронами, возбуждающими атомарный кислород. Механизм генерации фрагментов полярных сияний может быть связан с суперпозицией сходящихся электрических полей полярных сияний (перпендикулярно магнитному полю и дугам) и электрического поля атмосферного волнового поля.

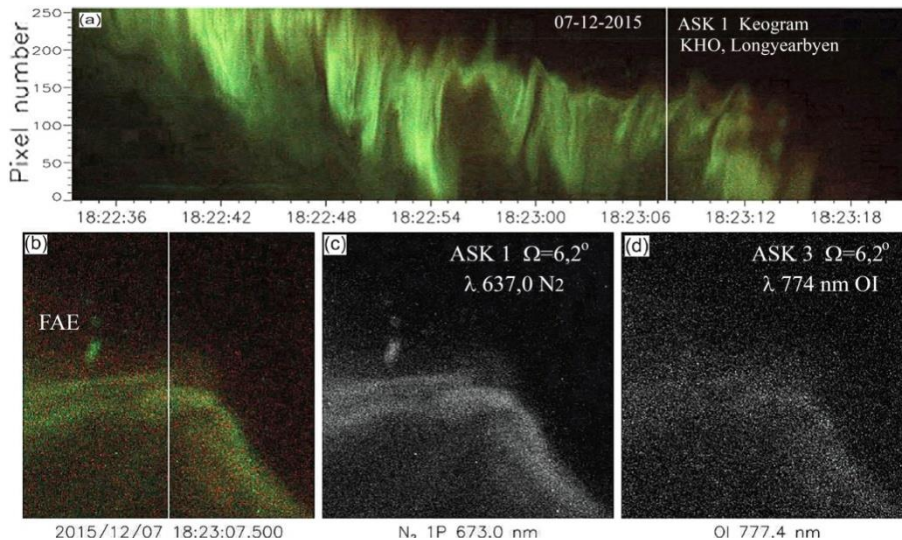


Рисунок 1.5.3 - Сверху вниз: кеограмма авроральной дуги, три изображения FAE, полученных узкоугольным имаджером в трех длинах волн

Структуры STEVE представляет собой дугу фиолетового цвета, которая образуется в верхних слоях атмосферы к экватору от овала полярных сияний. Модификацией STEVE считается пакет относительно маленьких зеленых лучей PF. На рисунке 1.5.4 показаны примеры этих структур. Наиболее часто STEVE появляется перед магнитной полночью во время фазы восстановления крупных суббурь и длится ~ 1-1,5 ч. В области самой дуги STEVE высыпания энергичных частиц отсутствуют. Считается, что усиленные экваториальные ветры могут переносить молекулы нейтрального газа, которые являются ключевыми для образования свечения STEVE на субавроральных широтах. Структура PF может быть локализована на небольшом участке пространства, образуя дугу со множеством лучей. Хотя структура зеленых лучей напоминает лучевые дуги полярных сияний, которые создаются энергичными электронами, высыпающимися из магнитосферы, предполагают, что за PF ответственны волны или их турбулентность в ионосфере. Дуга STEVE, дуга SAR и субавроральное излучение, возбужденное вторичными электронами от высыпающихся протонов, являются основными оптическими излучениями в ночной субавроральной ионосфере.



Рисунок 1.5.4 - Изображения структур STEVE и PF

Таким образом, в атласе на основе опубликованных примеров событий и явлений выбраны, систематизированы и проанализированы характеристики и условия генерации ряда авроральных и аврора-подобных структур, встречающихся в различных областях полярной ионосферы. В генерации всех структур, рассмотренных в части 2 атласа, принимают участие волны и/или плазменные неустойчивости разных типов. Эти структуры отражаются в неоднородностях плотности плазмы, возникающих в полярной ионосфере во время возмущенных геомагнитных условий и суббурь. Представленные примеры сопровождаются описанием обстановки в части областей аврорального овала, отображаемых в среднемасштабных изображениях ВУФ эмиссий, полученных с орбит КА, и рассчитанных из них характеристик высыпающихся электронов. В тематические примеры включены описание методик орбитальных измерений, включая *in situ* диагностику характеристик плазмы и дистанционную глобальную и локальную оптическую диагностику распределений интенсивности авроральных эмиссий в видимой и ВУФ-областях спектра.

## 1.6 Исследование статистических связей глобальных и локальных геомагнитных вариаций с процессами в ионосфере, магнитосфере и солнечном ветре с целью выявления новых закономерностей и прогноза экстремальных событий

На этапе 2023 г. было проведено исследование геомагнитных, термосферных и ионосферных возмущений при массовой потере спутников при очередном запуске Starlink в феврале 2022 г. По обычной процедуре проекта SpaceX ракета выводит спутники на начальную орбиту высотой около 210 км, далее они должны использовать электрические двигатели для подъема до рабочей высоты. Низкая начальная орбита предусмотрена, чтобы спутники, которые не прошли первоначальную проверку после запуска, быстро сбивались с орбиты из-за атмосферного сопротивления и не становились космическим мусором. На фазе роста 25-го солнечного цикла условия космической погоды стали меняться. В феврале 2022 г. при запуске с Мыса Канаверал очередной серии спутников системы Starlink произошла техногенная катастрофа. Этот запуск совпал с развитием двухфазной геомагнитной бури 3-4 февраля. После старта 3 февраля в 18 UT все спутники Starlink достигли начальной высоты 210 км в перигее и должны были после маневра достичь высоты ~340 км. Однако, 38 из 49 запущенных КА не смогли подняться до запланированной высоты, вследствие усиленного торможения сошли с орбиты и вернулись обратно в атмосферу 8 февраля. Пресс-релиз SpaceX информировал, что повышенное торможение и потеря большинства запущенных спутников произошли вследствие повышения на 50% плотности атмосферы во время геомагнитной бури по сравнению со спокойными условиями. Известно, что Сила торможения спутников пропорциональна плотности масс термосферы, а во время развития авроральных электроджетов и высыпаний частиц происходит джоулев нагрев с увеличением нейтральной плотности в расширяющейся термосфере. Эмпирическая модель термосферы NRLMSIS, исходя из уровня геомагнитной активности, предсказывала увеличение нейтральной плотности всего на 5-10%. Модельная оценка оказалась недостаточной для описания динамической реакции термосферы во время конкретной магнитной бури.

Был проведен сравнительный анализ изменения параметров термосферы и ионосферы для пяти последовательных запусков Starlink в условиях космической погоды, варьирующихся от спокойных до буревых возмущений. Запуски Starlink перечислены в таблице 1.6.1, причем событие с кодом S-36 выделено жирным шрифтом. Приведены суточные значения потока солнечного радио излучения F10.7, и геомагнитного индекса кольцевого тока Dst.

Таблица 1.6.1 – Список запусков Starlink

Starlink code	Date	Time UT hh:mm	F10.7	SSN2	Dstmin nT
S-31	2021.11.13	12:19	81.9	27	-4
S-36	<b>2022.02.03</b>	<b>18:13</b>	<b>122.9</b>	<b>80</b>	<b>-66</b>
S-37	2022.02.21	09:44	95.6	48	-4
S-57	2022.08.28	06:41	257.0	85	-12
S-58	2022.08.31	08:40	115.5	44	-11

На рисунке 1.6.1 на левой панели показан центрированный на нулевой день, соответствующий дате запуска, индекс Dst для 6-дневных периодов последовательных запусков 13 ноября 2021 г., 3 февраля 2022 г., 21 февраля 2022 г., 28 августа 2022 г., 31 августа 2022 г. Все запуски кроме 3 февраля 2022 г. проводились в геомагнитно спокойных условиях. Для события 31 августа 2022 г. Dst индекс достиг -70 нТл на четвертый день после запуска, когда спутники уже были на большой высоте. Окончившийся неудачей запуск 3 февраля 2022 г. проводился в условиях магнитной бури. Пики бури произошли 3 февраля в 11 UT (Dst = -66 нТл) и 4 февраля в 21 UT (Dst = -61 нТл). Второе усиление произошло, когда спутники все еще находились в более плотной атмосфере на небольшой высоте.

На рисунке 1.6.1 на правой панели показана плотность нейтрального термосферного газа ( $N$ ), измеренная на высоте  $\sim 450$  км вдоль орбиты одного из тандемных спутников Swarm (Sw-C), за 8-дневный период после запусков Starlink (день = 0). Видно, что с развитием бури 3–5 февраля  $N$  возрастает и на пиках Dst увеличивается до  $1,5 \text{ кг/м}^3$ , т.е. примерно на 50%. Данные Swarm подтверждают глобальное увеличение плотности нейтральной массы термосферы из-за усиленной активности авроральных электроджетов в этот период.

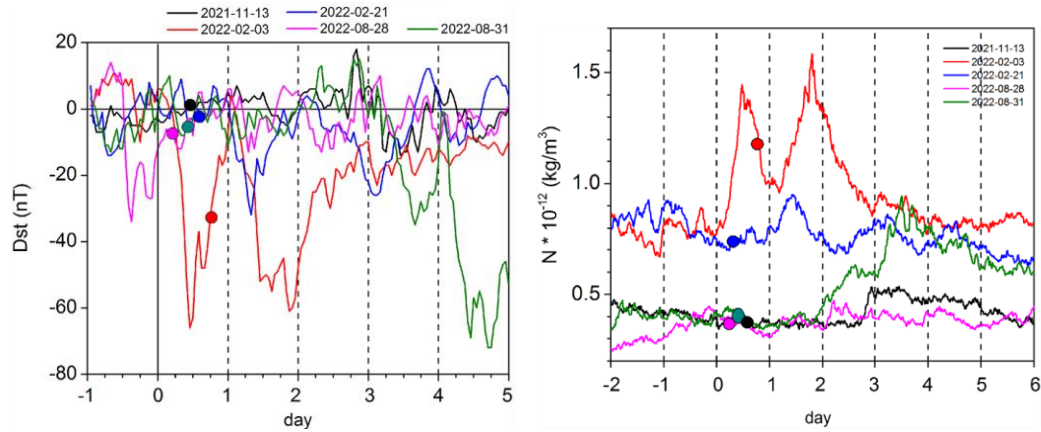
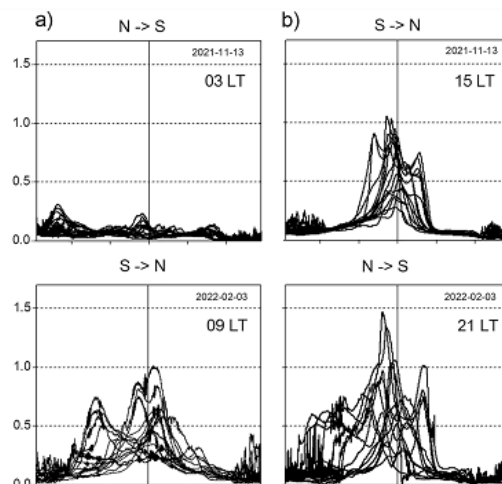


Рисунок 1.6.1 – (левая панель) Dst индекс для 6-дневного периода, центрированный на нулевой день, соответствующий дате запуска, для пяти рассматриваемых событий; (правая панель) плотность нейтральной массы термосферы за 8-дневный период. Точками отмечено время запуска

Зонд Ленгмюра на борту спутника Swarm проводит измерения электронной плотности ( $N_e$ ). На рисунке 1.6.2 показано широтное распределение  $N_e$  по измерениям Sw-C для тех же событий. В событии с кодом S-36 (3 февраля) увеличение  $N_e$  наблюдается на вечерней стороне больше, чем на утренней. Повышение распространяются на южное полушарие, захватывая большой диапазон широт. То есть, с развитием магнитной бури, во время которой был запущен Starlink S-36, происходило перераспределение ионосферной плазмы, порождающее глобальную асимметрию.



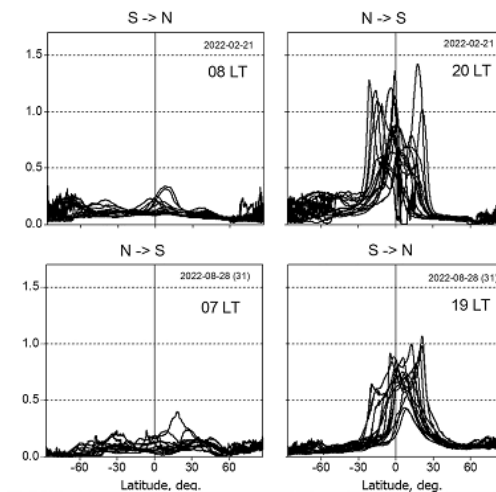


Рисунок 1.6.2 – Пары широтных профилей  $N_e$  для событий S-31, S-36, S-37 и S-57: (а) утренняя и (б) вечерняя сторона. На графиках указано местное время LT

Для того, чтобы оценить степень волновой активности, турбулентности и неустойчивости в ионосфере во время запусков Starlink, были рассчитаны средние значения ПЭС (интегральная плотность плазмы в столбе площадью  $1 \text{ м}^2$  на высоте от 65 км до 20200 км в единицах TECU  $\sim 10^{16}$  эл/м<sup>2</sup>) вдоль терминатора на его утренней и вечерней половинах на высоте 300 км. Использовались 15-минутные глобальные карты GIM-TEC, предоставленные Jet Propulsion Laboratory с пространственным разрешением  $1^\circ$  по широте и долготе. Результаты для двух событий, S-36 (1-8 февраля 2022 г.), S-57 и S-58 (27-31 августа 2022 г.), представлены на рисунке 1.6.3. В обоих случаях ПЭС на закате больше, чем на восходе. Это общая характеристика ионосферы, которая проявляется и в данных КА Swarm ( $N_e$  на высоте 450 км), и в общем количестве электронов от нижней части ионосферы до орбиты GPS на высоте 20 000 км.

Наблюдается колебательная изменчивость ПЭС, при этом моменты экстремумов на кривой заката не совпадают с экстремумами на кривой восхода. Моменты запусков S-36, S-57 и S-58 отмечены вертикальными линиями. Запуск S-36 произошел во время резкого подъема ПЭС на закате и перед утренним повышением ПЭС. Амплитуда колебаний ПЭС во время S-36 существенно больше, чем в других случаях, что свидетельствует о различии волнового режима в штормовых (S-36) и спокойных (S-57 и S-58) условиях. При запуске S-36 резкий рост плотности плазмы на восходящей части волны при общем повышении плотности нейтральных частиц, вызвавшей рост силы сопротивления, способствовали отказу спутников S-36 при маневре их подъема на рабочую высоту. После неудачного запуска S-36 для последующих запусков SpaceX (включая S-57 и S-58) первоначальная высота старта была увеличена с 210 км до 300 км.

По результатам исследования можно сделать следующие выводы. Во-первых, сравнение данных пяти запусков Starlink показывает, что по мере развития магнитной бури 3–5 февраля 2022 г. на фоне падения Dst индекса нейтральная плотность термосферы увеличилась примерно на 50%. Плотность нейтральной компоненты и плотность электронов увеличивалась глобально и со смещением в южное полушарие. Во-вторых, наблюдались колебания ПЭС на солнечном терминаторе с асимметрией утренне-вечерних значений с преобладанием вечерних величин. Из пяти рассмотренных событий запусков Starlink амплитуда колебаний была наибольшей во время бури 3–5 февраля 2022 г. В-третьих, изменение среднего ПЭС на закате и восходе солнца имело волновой характер с периодами колебаний 24 часа, 12 часов и меньше. Ионосферно-термосферная динамика привела к увеличению сил сопротивления, в результате чего 38 из 49 запущенных спутников, повторно вошедших в атмосферу 8 февраля, были потеряны. решающее влияние на то, что запуск S-36 закончился потерей спутников оказали условия космической погоды.



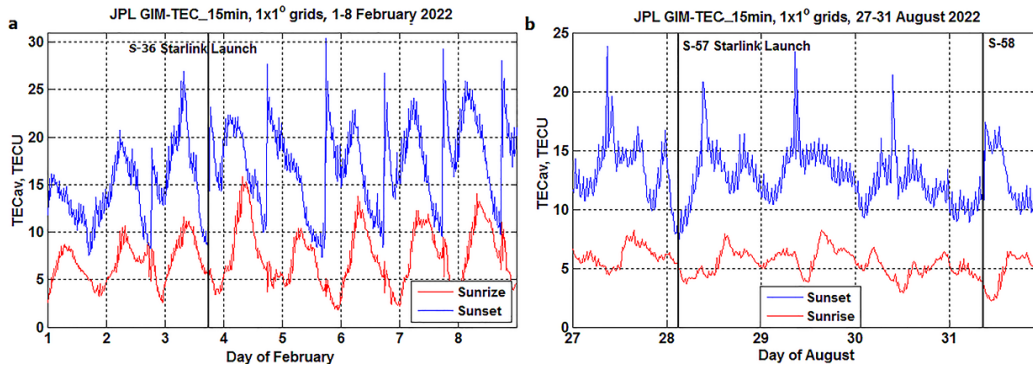


Рисунок 1.6.3 - Изменение ПЭС вдоль терминатора на восходе и заходе солнца на высоте 300 км; (а) запуск S-36, 2-8 февраля 2022 г.; (б) запуски S-57 и S-58, 27-31 августа 2022 г. Время запусков S-36, S-57 и S-58 обозначено вертикальными линиями

### 1.7 Создание деревьев принятия решений для классификации изменений состояния организма в связи с воздействием на него факторов космической и земной погоды на основе алгоритмов дифференциальной эволюции

Продолжалось изучение влияния космофизических факторов на состояние человеческого организма. В 2023 году мы сосредоточились на долгосрочной биотропности солнечных вспышек, а именно на вопросе о том, существует ли накопительный эффект влияния количества вспышек, появившихся за ограниченный отрезок времени, на состояние возбуждения нервной системы человека. Для этого были построены два временных ряда за 17 лет: количество произошедших за неделю вспышек, и число написанных за ту же неделю пользователями сообщений на форуме [dxdy.ru](http://dxdy.ru) (крупнейший в России научный форум), написанных в агрессивном ключе. Анализ настроений текста проводился нейросетью по особой программе и носил объективный характер.

И тот, и другой ряд содержат быстрые осцилляции, тогда как для выявления накопительного эффекта важен тренд. Традиционно используют полиномиальные модели трендов: если  $t$  – время, то модель тренда это  $X(t) = a_0 + a_1t + \dots + a_nt^n$ . Однако, невысокая степень  $n$  дает крайне низкодетализированные тренды, а большие  $n$  приводят к потере устойчивости алгоритма аппроксимации. Более того, полиномиальный тренд параметризован исключительно собственной степенью  $n$ , тогда как нам интересна отвечающая за его гладкость постоянная времени, связь которой с константой  $n$  неоднозначна. С целью параметризовать тренд величиной, имеющей размерность времени, мы ввели понятие постоянной релаксации  $T$ .  $T$  отвечает за негладкость тренда, обобщая понятие периода для аperiodических временных рядов. Определим негладкость тренда. Пусть  $\{X, t\}$  – совокупность отсчетов тренда  $X$  в моменты времени  $t$ . Рекурсивно введем вектор  $n$ -ных разностей временного ряда  $\{X, t\}$  как новый временной ряд  $\{\Delta_{(n)}, \tau_{(n)}\}$ :

$$\begin{cases} \Delta_{(n)}^k = \frac{\Delta_{(n-1)}^{k+1} - \Delta_{(n-1)}^k}{\tau_{(n-1)}^{k+1} - \tau_{(n-1)}^k}, & \Delta_{(0)}^k \equiv X^k \\ \tau_{(n)}^k = \frac{1}{2} [\tau_{(n-1)}^{k+1} + \tau_{(n-1)}^k] & \tau_{(0)}^k \equiv t^k \end{cases}, \quad (1.7.1)$$

Тогда негладкость тренда есть отношение квадрата нормы вектора  $\Delta_{(2)}$  к квадрату нормы вектора  $\Delta_{(1)}$ , а  $T$  обратно пропорциональна некоторой функции негладкости. То есть

$$\frac{1}{T} = f\left(\frac{\|\Delta_{(2)}\|^2}{\|\Delta_{(1)}\|^2}\right) \quad (1.7.2)$$

Конкретный вид функции  $f(\cdot)$  ищется исходя из условия совпадения  $T$  с полупериодом в случае синусоидального тренда: если  $X^k = A\cos(\omega t^k + \varphi)$ , то  $\Delta_{(1)}^k = \omega A\cos(\omega t^k + \pi/2 + \varphi)$  и

$$\|\Delta_{(n)}\|^2 = \frac{(A\omega)^2}{2} \left[ \|\Delta_{(n-1)}\|^2 + \sum_k \cos(2(\omega t_{(n-1)}^k + (n-1)\pi)) \right] \quad (1.7.3)$$

Для тренда, который похож на синусоиду (этот частный случай должен давать полупериод в качестве постоянной релаксации), вторым слагаемым в скобках в (1.7.3) можно пренебречь. Тогда для произвольного тренда

$$T = \pi \left( \frac{\|\Delta_{(2)}\|^2}{\|\Delta_{(1)}\|^2} \right)^{\frac{1}{2}} \quad (1.7.4)$$

Алгоритм построения тренда предполагает оптимизационную задачу, в которой ищется компромисс между его негладкостью и близостью к исходным данным в некоторой метрике. Выражение не-гладкости можно получить через квадратичную форму: действительно, вектор отсчетов временного ряда  $(n+1)$ -й разности линейно выражается через вектор отсчетов  $n$ -й разности:  $\Delta_{(n+1)} = D\Delta_{(n)}$ , где матрица  $D$

$$D_n = \begin{pmatrix} -\frac{1}{\tau_{(n)}^2 - \tau_{(n)}^1} & \frac{1}{\tau_{(n)}^2 - \tau_{(n)}^1} & 0 & \dots & 0 \\ 0 & -\frac{1}{\tau_{(n)}^3 - \tau_{(n)}^2} & \frac{1}{\tau_{(n)}^3 - \tau_{(n)}^2} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & -\frac{1}{\tau_{(n)}^{|\Delta_{(n)}|} - \tau_{(n)}^{|\Delta_{(n)}|-1}} & \frac{1}{\tau_{(n)}^{|\Delta_{(n)}|} - \tau_{(n)}^{|\Delta_{(n)}|-1}} \end{pmatrix} \quad (1.7.5)$$

Из этого следует, что  $\|\Delta_{(1)}\|^2 = X^T D_1^T X$  и  $\|\Delta_{(2)}\|^2 = X^T D_1^T D_2^T D_2 D_1 X$ . Целевая функция для оптимизационной задачи поиска тренда представляет собой выпуклую комбинацию близости и негладкости, и чем больший коэффициент получает негладкость – тем постоянная релаксации тренда выше. Можно подобрать коэффициент при негладкости так, чтобы добиться равенства параметра  $T$  практически любому заданному значению.

Метод позволяет искать не только тренд среднего значения, но и тренды огибающих. Огибающие временного ряда – это пара кривых, образующих коридор, целиком содержащий в себе временной ряд. Эти кривые можно искать итеративным образом. Пусть  $X_{(i)}$  – тренд, найденный на  $i$ -той итерации. Будем измерять близость искомого на  $(i+1)$ -й итерации тренда  $X_{(i+1)}$  к экспериментальным данным  $\xi$  при помощи функции:

$$\Psi = \sum_k \Re \left[ X_{(i)}^k - \xi^k \right] \left[ X_{(i+1)}^k - \xi^k \right]^2 \quad (1.7.6)$$

где  $\Re[\cdot]$  – немного сдвинутая вверх по вертикальной оси функция  $\text{gelu}$ , справа от нуля равная своему аргументу, а слева – нулю. В оптимизационной задаче находящиеся выше уже найденного тренда  $X_{(i)}$  (если ищется верхняя огибающая) или ниже тренда (нижняя огибающая) экспериментальные точки получают больший вес, чем остальные точки, что заставляет тренд сдвигаться, приближаясь к искомой границе коридора.

Был построен временной ряд концентрации солнечных вспышек (рисунок 1.7.1). Рассчитанный период по максимуму автокорреляционной последовательности оказался равным 3843 дня, или 10.53 лет.

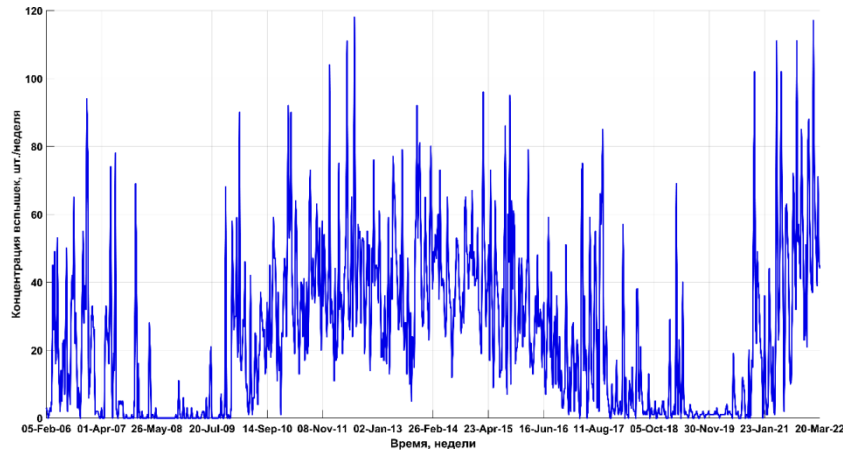


Рисунок 1.7.1 - Временной ряд недельной концентрации вспышек 2005-2022 гг.

Затем ряд пользовательской активности на форуме [dxdy.ru](http://dxdy.ru) (число сообщений в неделю) был разбит на обучающий и валидационный сегменты в соотношении 3 к 1 и построена зависимость погрешности  $P$  попадания валидационной выборки в коридор между огибающими от постоянной релаксации обеих огибающих. Если  $X$  - это исходные данные размера  $N$ , а  $R$ ,  $V$ , и  $U$  – соответственно тренд, верхняя и нижняя огибающая, то погрешность попадания определяется формулой:

$$P = \sqrt{16 \frac{1}{N} \sum_{k=1}^N \left( \frac{X^k - R^k}{V^k - U^k} \right)^2} + \sum_{k=1}^N [V^k > U^k] \quad (1.7.7)$$

где  $[ \cdot ]$  это скобка Айверсона. Эта зависимость изображена на рисунке 1.7.2.

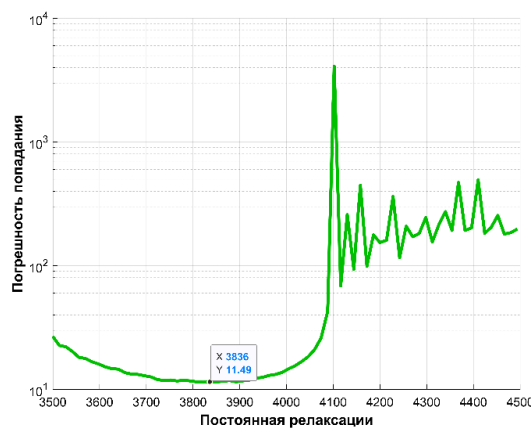


Рисунок 1.7.2 - Погрешность попадания валидационной выборки в коридор прогноза в зависимости от  $T$ . Флажком показан минимум, который приходится на 3836 дней

Построена модель неполиномиального тренда для выявления синхронизации параметров психофизического состояния человека с концентрацией вспышечной активности Солнца. Получено, что оптимальное время релаксации огибающих временного ряда недельного числа сообщений на 0.2% отличается от найденного по критерию максимума автокорреляционной последовательности периода временного ряда концентрации вспышек.

## 1.8 Разработка методов диагностики источников локальной ионизации, включая малые космические объекты, на основе анализа больших массивов данных радарных наблюдений

Локальная ионизация ионосферы приводит к появлению плазменных неоднородностей и увеличению градиентов электронной концентрации ( $N_e$ ). Так как в полярной ионосфере такие неоднородности особенно труднопредсказуемы, их формирование приводит к большим проблемам для радиосвязи, ухудшению точности ГНСС-позиционирования и другим проблемам. В 2023 году было проведено исследование влияния продольных токов (ПТ) магнитосферного происхождения на величину и локализацию неоднородностей  $N_e$  в высокоширотной ионосфере. Для этого были проанализированы результаты сопряженных измерений космических аппаратов (КА) Swarm и европейского радара некогерентного рассеяния EISCAT (ESR) на Шпицбергене для событий одновременных наблюдений в различных условиях: в ночной ионосфере во время активации суббури и в дневной спокойной ионосфере. Бортовые магнитометры КА Swarm обеспечивают измерения плотности ПТ над ионосферой. Радар, находящийся в это время под траекторией пролета, измеряет вертикальное распределение электронной концентрации  $N_e$ .

Носителями ПТ являются электроны. Направленные вверх ПТ (вытекающие из ионосферы) соответствуют нисходящему электронному тепловому потоку, а втекающие в ионосферу токи сопровождаются уходящим электронами. В области вытекающих ПТ происходит увеличение  $N_e$  и соответствующее повышение проводимости, а в области втекающего число электронов уменьшается. Для замыкания ионосферного тока, соединяющего области с разными проводимостями, в месте втекающего ПТ электрическое поле должно быть больше, чем в месте вытекающего. В ионосфере ПТ замыкаются за счет токов Педерсена, носителями которых являются в основном ионы. Таким образом, возникает трехмерная токовая система, а в области втекания и вытекания ПТ в ионосферной плазме должно происходить локальное изменение плотности плазмы. Теоретические оценки изменений  $N_e$  были сделаны на основе решения уравнений для движения ионосферной плазмы и замыкания токов. Однако до последнего времени так и не появилось экспериментальных подтверждений данного эффекта. Это связано с тем, что в эксперименте необходимо иметь данные одновременных измерений плотности ПТ и вертикального профиля  $N_e$ . Прямые измерения ПТ над ионосферой могут быть проведены только с помощью низкоорбитальных спутников, а распределение  $N_e$  по всей толщине ионосферы может быть получено только с помощью радара некогерентного рассеяния, находящегося в это время под траекторией пролета. Поскольку существует только два высокоширотных некогерентных радара – на архипелаге Шпицберген и на Аляске, а на полярной орбите в настоящее время находится только КА Swarm, и эти две измерительные системы, наземная и космическая, не проводят скоординированных экспериментов, обеспечить нужное сочетание наблюдений весьма проблематично. Тем не менее, используя открытую базу данных измерений MADRIGAL для европейского радара ESR и базу данных Европейского космического агентства по КА Swarm, удалось отобрать несколько репрезентативных событий.

Для выявления событий сопряженных спутниковых измерений ПТ и радарных измерений  $N_e$  были отобраны траектории пролетов КА, проходящие над локацией ESR на расстоянии не более, чем  $1^\circ$  по долготе. Затем для этих событий были проверена база данных ESR на предмет проведения в это время сеансов 42-метровой антенны в зенит. Ниже рассмотрены два репрезентативных события сопряженных измерений. На рисунке 1.8.1 для этих событий представлена полярная проекция траектории пролета тандема КА Swarm в геомагнитных координатах, наложенная на двумерную картину ПТ, которая получена по данным спутникового проекта AMPERE. Через вебпортал этого проекта доступны двумерные карты распределения плотности ПТ в высокоширотной области с 10-минутным разрешением, сформированные по данным спутников Iridium.

ПТ и Ne в (1) магнитоспокойный период 5 февраля 2017 г. и (2) во время активации суббури 9 января 2014 г. Для первого события на рисунке 1.8.1 на первой панели слева показана часть траектории пролета тандема Sw-A и Sw-C в прямоугольнике, ограниченном координатами 65°–85°с.ш. и 5°–27°в.д., и на второй панели слева - полярная проекция траектории, наложенная на карту ПТ по AMPERE. К югу от Шпицбергена расположен вытекающий ПТ с плотностью до 0.6  $\mu\text{A}/\text{m}^2$ . КА Sw-A и Sw-C пролетают вблизи луча ESR примерно в середине минуты 15:23 UT, а зону вытекающего ПТ попадают в конце этой минуты. На рисунке 1.8.1 на двух панелях справа показан пролет спутников над радаром в 20:26–20:36 UT, который совпал с пиком активации суббури, а траектория проходила через предполуночный сектор местного времени. Спутники наблюдают три слоя крупномасштабных ПТ: втекающий ток с плотностью до 2  $\mu\text{A}/\text{m}^2$  на 77-79° широты и вытекающий ток сравнимой плотности на 71-77°. Ближайшая к траектории позиция радара находится в зоне вытекающего из ионосферы ПТ.

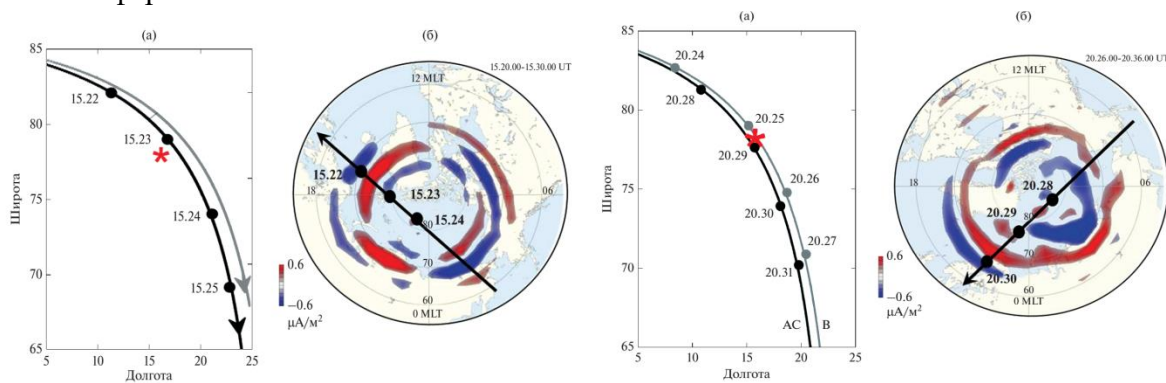


Рисунок 1.8.1 - Две левые панели относятся к пролетам в 15:22–15:25 UT 5 февраля 2017 г., а две правые – к пролетам в 20:26–20:36 UT 9 января 2014 г. Показаны: траектории пролета КА Sw-A (серая линия) и Sw-C (черная линия) в географических координатах; звездочкой обозначена локация радара ESR; ионосферная проекция ПТ по данным AMPERE (красный цвет – вытекающий из ионосферы ток, синий цвет – втекающий в ионосферу ток) и траектория пролета КА (черная линия и минутные реперные точки) в координатах MLT–геомагнитная долгота

Параметры токов, рассчитанные по измерениям бортовых магнитометров КА Sw-A и Sw-C, представлены на рисунке 1.8.2. Левая панель относится к магнитоспокойному периоду, а правая – суббуревому. В первом случае крупномасштабный втекающий (положительный) ПТ с плотностью до 0.8  $\mu\text{A}/\text{m}^2$  расположен на 66-70° широты, а вытекающий (отрицательный) ток – на 73-79°. Во время суббури наблюдаются три слоя крупномасштабных ПТ: втекающий ток с плотностью до 2  $\mu\text{A}/\text{m}^2$  на 77-79°, вытекающий ток на 71°-77°. Ближайшая к траектории позиция радара находится в зоне вытекающего из ионосферы тока.

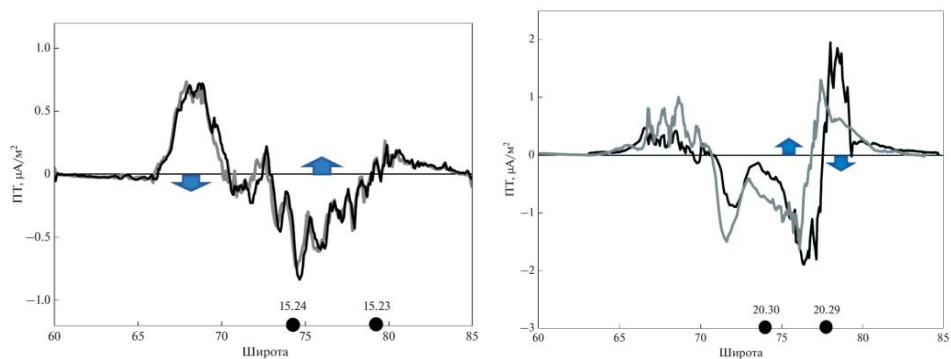


Рисунок 1.8.2 - Плотность ПТ по измерениям КА Swarm 5 февраля 2017 г. (слева) и 9 января 2014 г. (справа). Стрелкой, направленной вниз (вверх), обозначен ПТ, втекающий в ионосферу (вытекающий из ионосферы). На оси X точками обозначены минуты, когда спутник пролетал вблизи ESR

5 февраля 2017 г. радарные наблюдения проводились по программе CP в режиме *iru* с 13 до 18 UT. Рисунок 1.8.3, две панели слева, показывает 1-минутные вертикальные профили  $Ne$  в 15:21-15:28 UT. В 15:23 на высотах F-слоя (250–300 км) наблюдается резкое увеличение  $Ne$  до  $1.4 \cdot 10^{11} \text{ м}^{-3}$ . Для количественной оценки изменения  $Ne$  выделены три профиля: центральный в 15:23 UT и усредненные за три предыдущие и три последующие минуты. Разность между центральным и соседними, предыдущим и последующим, профилями в максимуме  $Ne$  на высоте 260 км составляет, соответственно,  $2.5 \cdot 10^{10}$  и  $7 \cdot 10^{10} \text{ м}^{-3}$ . На рисунке 1.8.3, справа, представлены последовательные минутные вертикальные профили  $Ne$  в 20:21-20:38 UT 9 января 2014 г. В 20:29 UT во время активизации суббури при пересечении траекторией спутников слоя вытекающего из ионосферы ПТ максимальное увеличение  $Ne$ , до  $8 \cdot 10^{11} \text{ м}^{-3}$ , наблюдалось на высотах 120 км, т.е. в E-слое. Увеличение плотности плазмы заметно и по всей высоте в F-слое. Разность между центральным и соседними профилями  $Ne$  в максимуме E-слоя достигает  $5.5 \cdot 10^{11} \text{ м}^{-3}$ . В F-слое эта разность - около  $2 \cdot 10^{11} \text{ м}^{-3}$ .

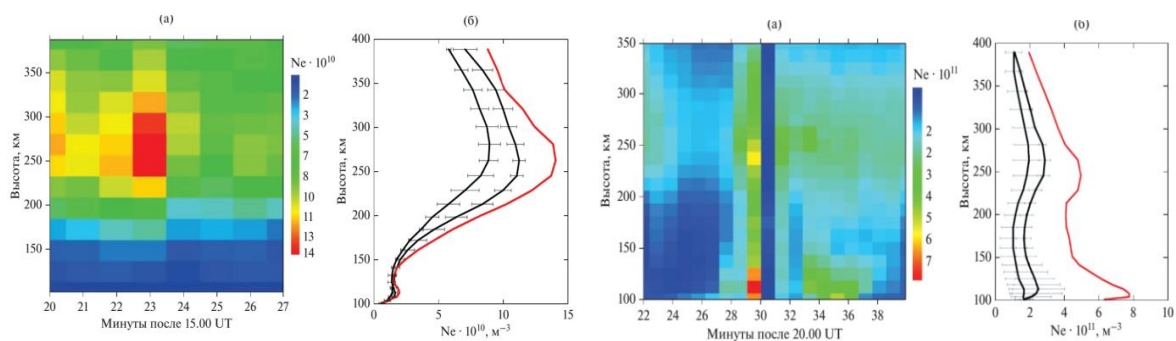


Рисунок 1.8.3 – (а) Вертикальное распределение и (б) профиль  $Ne$  (слева) 5 февраля 2017 г. и (справа) 9 января 2014 г. Красной линией показан профиль в 15:23 UT и 20:29 UT, соответственно, а черной - усредненные за три предыдущие и три последующие минуты

На основе аналитического решения системы уравнений непрерывности для концентрации плазмы и токов, и уравнения движения для ионов проведена количественная оценка соотношения между ПТ и  $Ne$ . В таблице 1.8.1 дана сводка значений параметров, измеренных во время пролетов спутников над областью радарного зондирования в рассмотренных выше событиях, и значений параметров, рассчитанных на основе уравнений.

Таблица 1.8.1 - Значения параметров во время событий 5 февраля 2017 г. и 9 января 2014 г.

Геомагнитные условия	спокойные		суббурия	
MLT пролета КА над радарным лучом	~16.30		~21.30	
Измеренная плотность восходящего ПТ ( $\mu\text{A}/\text{m}^2$ )	МАХ значение при 30-с усреднении (1-секундный экстремум)			
	0.8 (7.5)		1.8 (12.2)	
Тип оценки	Теория	Измерения	Теория	Измерения
E-слой $\Delta N$ ( $\text{m}^{-3}$ )	$2 \cdot 10^{11}$	$0.4 \cdot 10^{10}$	$4.5 \cdot 10^{11}$	$5.8 \cdot 10^{11}$
F-слой $\Delta N$ ( $\text{m}^{-3}$ )	$4 \cdot 10^{10}$	$4 \cdot 10^{10}$	$4.5 \cdot 10^{11}$	$2.3 \cdot 10^{11}$

Сравнение измеренных и рассчитанных значений изменения концентрации  $\Delta N$  в максимуме E- и F-слоев показывает хорошее согласие для события 9 января 2014 г. в ночной ионосфере в условиях активации суббурии. Для события 5 февраля 2017 г., когда измерения проходили в послеполуденном секторе, в слое F измеренные и рассчитанные значения  $\Delta N$  совпадают с большой точностью. В то же время, в E-слое значения  $\Delta N$  отличаются больше, чем на порядок. Наблюдаемые значения оказываются гораздо меньше теоретических. Несовпадение обусловлено тем, что в спокойный период ПТ, направленные из ионосферы, переносятся потоками самых низкоэнергичных электронов с энергией не больше первых единиц кэВ. Потоки высыпающихся частиц существенно зависят от уровня геомагнитной активности. Основной вклад в ионизацию в авроральной зоне вносят электроны с энергиями 1–10 кэВ. Электроны с энергиями 1 кэВ поглощаются в слое F, а электроны с энергией 10 кэВ в слое E. В возмущенной ночной ионосфере носители направленного вверх ПТ представлены всем энергетическим спектром, а в дневной спокойной ионосфере – только его низкоэнергичной частью. Результаты корреляционного анализа показали, что между проводимостью ионосферы и вытекающим из ионосферы ПТ существует более тесная связь по сравнению с ПТ, втекающим в ионосферу. Характер локальной модуляции  $Ne$  в области втекающего в ионосферу ПТ определяется не только величиной этого тока, но и энергией переносящих его электронов.

## 1.9 Заключение

Изложенные в главе 1 результаты показывают, что проведенные работы соответствуют перечню, установленному в плане работ ИКИ РАН на 2023 год по направлению «Развитие методов моделирования высокоширотной ионосферы и анализ больших массивов спутниковых и наземных данных измерений; построение численных и гибридных моделей распределения электронной концентрации в полярной ионосфере в зависимости от изменяющихся гелио-геофизических условий». Основные полученные результаты состоят в следующем.

В развитие численной модели распределения электронной концентрации в F-области высокоширотной ионосферы на основе численных экспериментов были проанализированы систематические изменения крупномасштабных ионосферных неоднородностей и выявлены условия, способствующие локальному истощению плазмы и образованию полярных дыр. Показано, что структура образуется в темной ионосфере в области стагнации под действием противоположно направленных скоростей конвекции и коротации.

Модель адекватно воспроизводит сезонные тенденции, но регулярно недооценивает абсолютные значения  $Ne$  выше максимума слоя F2. Требуется дальнейшая параметризация по данным радарных и ионозондовых измерений.

Отработана технология наблюдений, оценки и визуализации параметров высокоширотной ионосферы в районе экспериментальных радиофизических работ

трансионосферного мультимчастотного зондирования на основе ГНСС-технологий на Земле Франца Иосифа.

Подготовлена вторая часть аннотированного атласа изображений эмиссий в авроральных структурах, обусловленных определенными проявлениями магнитосферной динамики и волновыми процессами.

Проведено исследование геомагнитных, термосферных и ионосферных возмущений при неудачном запуске Starlink в феврале 2022 г., когда почти все запущенные спутники были потеряны из-за повторного входа в атмосферу. Причиной техногенной катастрофы было развитие магнитной бури, увеличение плотности термосферы и повышение волновой активности на ионосферных высотах на терминаторе.

На основе сопряженных спутниковых и радарных измерений выявлено влияние продольных токов магнитосферного происхождения на величину и локализацию неоднородностей электронной плотности в высокоширотной ионосфере. При возмущенных условиях в месте вытекающего из ионосферы продольного тока плотность плазмы растет и в слое E, и в слое F. В дневной спокойной ионосфере при вытекающем токе плотность плазмы увеличивается только в F-слое. Различия обусловлены тем, что в первом случае носители тока представлены полным энергетическим спектром авроральных электронов, а во втором – только его низкоэнергичной частью.

Построена модель неполиномиального тренда для выявления синхронизации параметров психофизического состояния человека (на примере активности в интернете) с концентрацией вспышечной активности Солнца. Получено, что огибающие временного ряда недельного числа сообщений подобны периоду временного ряда концентрации вспышек.

Материалы, изложенные в разделе, позволяют считать, что все работы по направлению выполнены. По результатам опубликовано запланированное число статей в реферируемых, в том числе высокорейтинговых, научных журналах и материалах конференций.



## **ГЛАВА 2 Применение методов машинного обучения и глубокого обучения для обработки и анализа больших массивов астрономических данных в задачах наблюдательной астрофизики**

### **2.1 Введение**

Направление 2 ("Космос-Д-Астрофизика") темы "Космос-Д" посвящено вопросам, связанным с разработкой и совершенствованием научных основ, методов, технологий и систем работы с большими данными для решения научных задач интеллектуального анализа больших данных в наблюдательной астрофизике.

По направлению «Космос-Д-Астрофизика», основные усилия в 2023 были сконцентрированы, во-первых, на создании моделей автоматического распознавания скоплений галактик в сырых данных изображения с микроволновых телескопов - Atacama Cosmology Telescope (ACT) и PLANCK; распознаванию радиисточников в линии HI 21см в больших данных моделирования SKAO. Во-вторых, исследовались модели глубокого обучения для уточнения космологических расстояний далеких квазаров. Далее, мы создали прототип системы SmartRoutes, предназначенной для решения задач комбинаторной оптимизации при помощи нейроэвристик на основе обучения с подкреплением. Последним направлением работы в 2023 году в рамках направления «Космос-Д-Астрофизика» стало создание программной библиотеки для оптического отождествления и классификации рентгеновских источников в больших данных многоволновых обзоров неба.

В настоящей главе отчета представлены основные результаты работ по направлению «Космос-Д-Астрофизика», полученные по следующим разделам плана работ ИКИ РАН на 2023 год:

- Исследование, разработка и применение алгоритмов глубокого и классического машинного обучения для решения задач детектирования, кросс-отождествления, классификации, инкрементального моделирования и определения физических характеристик астрономических объектов по данным многоволновых обзоров всего неба.
- Исследование и разработка моделей обучения с подкреплением для задач комбинаторной оптимизации.
- Создание систем и технологий для обработки и анализа больших астрономических данных

Результаты полученные по данным направлениям описаны в разделах 2.2, 2.3, 2.4 настоящей главы. Подробно полученные в данном направлении результаты изложены в следующих научных публикациях [3,4,5,12], а также представлены в рамках 9-и устных докладов и тезисов на научных конференциях: «Ломоносовские чтения 2023» [23,24,25,26], «Тихоновские чтения 2023» [27-28], международной конференции «Астрофизика высоких энергий сегодня и завтра 2023 (HEA-2023)» [29-30] и XXV международной конференции "Data Analytics and Management in Data Intensive Domains" (DAMDID/RCDL 2023) [31].

### **2.2 Исследование, разработка и применение алгоритмов глубокого и классического машинного обучения для решения задач детектирования, кросс-отождествления, классификации, инкрементального моделирования и определения физических характеристик астрономических объектов по данным многоволновых обзоров всего неба**

Скопления галактик - это наиболее массивные гравитационно связанные системы во Вселенной, состоящие из темной материи, горячего барионного газа и звезд. Они играют важную роль в наблюдательной космологии и изучении эволюции галактик. В статье [4] мы

разработали модель глубокого обучения для сегментации сигнала Сюняева - Зельдовича в картах интенсивности ACT+Planck и представили в [30] новый каталог скоплений галактик в области покрытия ACT. Для повышения чистоты каталога скоплений мы ограничились публикацией только части полной выборки с наиболее вероятными скоплениями галактик, лежащими в направлениях кандидатов расширенного каталога скоплений Планка [25]. Каталог ComРАСТ содержит 2,934 скопления галактик, из которых до 1436 СЗ-объектов являются новыми по отношению к существующим выборкам скоплений ACT DR5 и PSZ2.

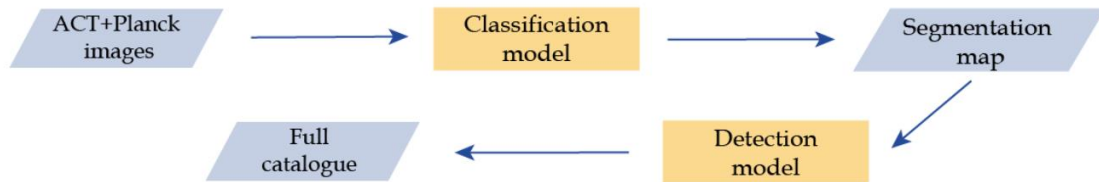


Рисунок 2.2.1 - На схеме из статьи [4] показаны основные шаги по созданию нейросетевого каталога объектов Сюняева-Зельдовича

	Layer	Output map size
1	Input	$32 \times 32 \times 3$
2	$3 \times 3$ convolution +ReLU	$32 \times 32 \times 16$
3	$3 \times 3$ convolution +ReLU	$32 \times 32 \times 16$
4	$2 \times 2$ MaxPooling	$16 \times 16 \times 16$
5	$3 \times 3$ convolution +ReLU	$16 \times 16 \times 32$
6	$3 \times 3$ convolution +ReLU	$16 \times 16 \times 32$
7	$2 \times 2$ MaxPooling	$8 \times 8 \times 32$
8	$3 \times 3$ convolution +ReLU	$8 \times 8 \times 64$
9	$3 \times 3$ convolution +ReLU	$8 \times 8 \times 64$
10	$2 \times 2$ MaxPooling	$4 \times 4 \times 64$
11	$3 \times 3$ convolution +ReLU	$4 \times 4 \times 128$
12	$3 \times 3$ convolution +ReLU	$4 \times 4 \times 128$
13	$2 \times 2$ MaxPooling	$2 \times 2 \times 128$
14	Flatten	512
15	Linear+ReLU	256
16	Linear+ReLU	128
17	Linear+ReLU	32
18	Linear+Sigmoid	1

Рисунок 2.2.2 - На схеме из статьи [4] показана архитектура нейросетевой модели для распознавания объектов Сюняева-Зельдовича

Данные о распределении вещества во Вселенной (карта крупномасштабной структуры Вселенной) позволяют делать выводы о процессах эволюции астрономических структур. Крупномасштабная структура Вселенной на больших космологических расстояниях (красных смещениях) состоит главным образом из квазаров (активные ядра галактик), которые являются наиболее светимыми объектами во Вселенной начиная с ранних эпох образования сверхмассивных черных дыр в ядрах галактик. Таким образом, для построения карты распределения вещества на больших космологических расстояниях необходимо уметь точно

оценивать расстояния (красные смещения,  $Z$ ) до этих астрономических объектов на основе доступных многоволновых фотометрических данных обзоров неба.

Большой интерес представляет применение моделей оценки расстояний до квазаров к современным большим каталогам рентгеновских источников, таких как обзор всего неба SRG/ePOZITA. В астрономии существует два основных способа определения космологических расстояний до квазаров (спектроскопический и фотометрический). Спектрографический способ дает точную оценку расстояния, однако снятые спектры есть только у  $\sim 30\%$  точечных рентгеновских источников ePOZITA. Фотометрические оценки красного смещения (photo-z) можно получить для всех наблюдаемых объектов, но для этого необходимо использовать модели машинного обучения или менее точные алгоритмы на основе спектральных шаблонов. В задаче оценки расстояния до рентгеновских источников используются фотометрические данные в ограниченном наборе широкополосных фильтрах: вероятностные прогнозы photo-z квазаров могут иметь мультимодальное распределение PDF(z). Стандартные нейросетевые модели регрессии на табличных данных не учитывают данные особенности задачи измерения photo-z квазаров. Целью работы [23] являлось построение модели измерения космологических расстояний до активных ядер галактик на основе нейронных сетей с учетом мультимодальных прогнозов PDF(z) и сравнение результатов с базовым алгоритмом SRGz.

Предложенная нами нейросетевая модель состояла из энкодера в виде многослойного перцептрона (MLP) и гауссовой смеси (GMM) на выходе модели [23]. Сравнение проводилось с моделью SRGz на базе квантильного случайного леса [29], обладающей на данный момент наибольшей точностью и являющейся базовой для photo-z рентгеновских источников ePOZITA. Важным требованием к поставленной задаче является способность модели оценивать калиброванные распределения получаемых оценок расстояний. Исходя из приведенного замечания, было решено обучить глубокий ансамбль моделей (MLP+GMM). Тренировочная выборка включала в себя оптические квазары SDSS DR14Q и далекие ( $Z > 5$ ) квазары VHzQ. Признаки объектов - из объединенных данных 4-х фотометрических обзоров: DESI Legacy Imaging Surveys, Pan-STARRS1, SDSS, WISE.

В результате [23], нами была создана нейросетевая модель прогнозов космологических расстояний (photo-z), имеющая на 17,5% меньше выбросов (на всей выборке квазаров), и на 50% меньше выбросов для далеких объектов ( $Z > 5$ ), чем модель на основе случайного леса SRGz. Таким образом показано, что в задаче измерения фотометрических красных смещений квазаров, модели на основе нейронных сетей, обученные оценивать смеси нормальных распределений, способны достигать лучшей точности, чем модели, использующие алгоритм квантильного случайного леса.

### **2.3 Исследование и разработка моделей обучения с подкреплением для задач комбинаторной оптимизации**

Задачи комбинаторной оптимизации возникают в астрофизике при кросс-отождествлении астрономических объектов в больших данных фотометрических каталогов, содержащих миллиарды объектов в различных диапазонах спектра. Создание моделей и технологий для решения крупномасштабных задач комбинаторной оптимизации важно как с точки зрения проведения фундаментальных исследований в области наблюдательной астрофизики, так и для приложений в других областях человеческой деятельности. В 2023 году мы сосредоточились на решении задач построения и оптимизации маршрутов - класса задач комбинаторной оптимизации с многочисленными практическими применениями: курьерская доставка, работа такси, логистика товаров между складами и точками продаж.

В 2023 году в рамках темы Космос-Д-Астрофизика нами были проведены исследования [5, 12, 24, 26, 27] по ряду перспективных направлений, которые объединены общей целью — созданием точных и быстрых нейроэвристик для задач оптимизации маршрутов с ограничениями реального мира с различной размерностью задач (от малых 50-100 до больших

>1000), а также, было проведено сравнение результатов нейроэвристик и других методов (e.g. точные алгоритмы SCIP, эвристические подходы в составе LKH, метаэвристики Google OR-Tools).

Проблема маршрутизации транспортных средств (англ. Vehicle Routing Problem) – это класс задач транспортной логистики, направленных на минимизацию затрат на транспортные ресурсы, стоимости маршрута и времени доставки груза группе клиентов. По мере увеличения числа городов и клиентов становится необходимым разрабатывать решения задачи доставки, которые позволят оптимально использовать выделенные ресурсы при сохранении качества услуг. В прикладных областях появляются ограничения: временные окна доступности, последовательность посещения, вместимости грузовых средств. Решение логистической задачи в условиях множества ограничений реального мира является актуальной проблемой для большинства логистических компаний. Классические подходы для решения подобных задач включают в себя: (i) поиск точных решений при помощи целочисленного программирования (главное достоинство - оптимальность; главный недостаток - экспоненциальная сложность) и (ii) эвристические подходы (являются субоптимальными по отношению к точным решениям, быстрый поиск решения). На практике, эвристика, как правило, - лучший выбор (так как она позволяет найти решение задачи пусть неоптимальное), но ее разработка требует большого количества проб и ошибок и во многом зависит от интуиции и опыта людей-экспертов.

Существуют инструменты, позволяющие получить субоптимальное решение для классической постановки задачи построения и оптимизации маршрутов ограниченного размера (например, Google OR-Tools, LKH). Однако, для больших задач (>1000 точек) в литературе не было предложено эффективных полиномиальных алгоритмов.

В 2023 году в работах [24, 27] мы провели сравнение различных эвристических алгоритмов, таких как имитация отжига, 2 Opt, 3 Opt, эвристика Лина-Кёрнигана, с программным обеспечением SCIP, выступающего в качестве точного метода. Эксперименты проводились с использованием случайно сгенерированных наборов данных (размерность 50-100), а также реальных наборов данных (200-900 точек) от компании ORTEC. Рассмотренные методы включают ограничения на вместимость транспортных средств и временные ограничения доставки товаров. По результатам проведенных экспериментов на наборах данных из 50, 100 и от 200 до 900 точек лучшее качество среди эвристик продемонстрировала эвристика Лина-Кёрнигана, качество которого хоть и хуже на ~ 20% чем у SCIP на примерах 50 и 100 точек, однако поиск решений осуществляется в ~ 35 раз быстрее, что мотивирует использовать эвристические подходы для решения задач VRP больших размерностей.

В работе [27] нами представлен прототип системы SmartRoutes для оптимизации маршрутов с учетом ограничений. Система содержит три основных блока:

- (1) Модуль набора данных — это точка входа для всех манипуляций с данными, а также хранит подготовленные наборы данных и позволяет загружать сторонние данные в указанном формате.
- (2) Модуль алгоритма, отвечающий за обучение и оценку все доступные алгоритмы на заранее подготовленных наборах данных. Есть возможность добавить пользователя алгоритм через специальные интерфейсы.
- (3) Модуль решения производит метрики, визуализирует их. и предлагает веб-интерфейс для связи с системой.

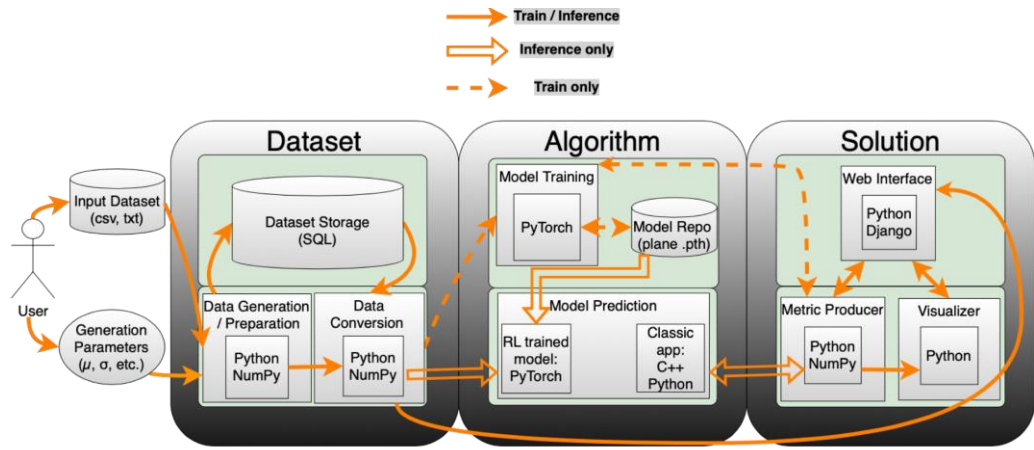


Рисунок 2.3.1 - Архитектура прототипа системы SmartRoutes [27]

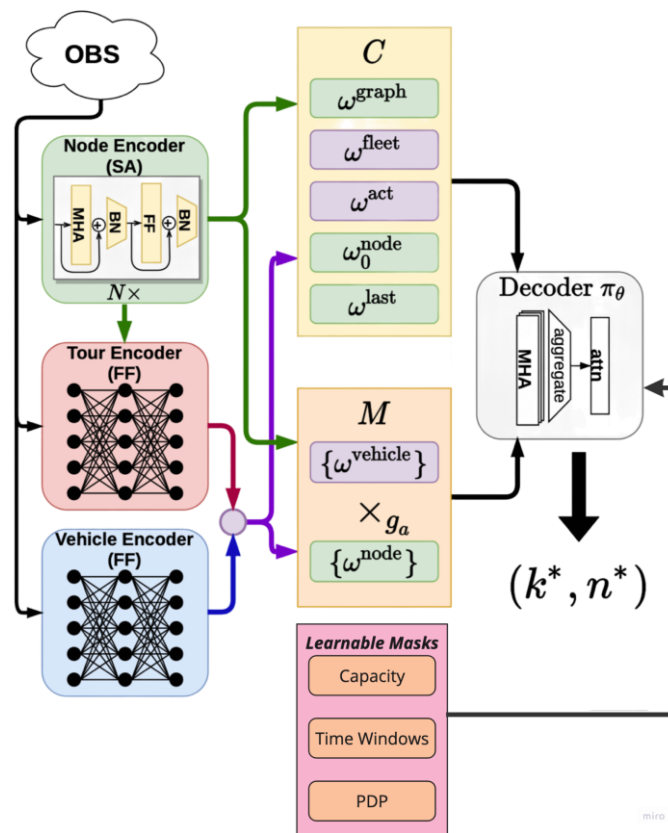


Рисунок 2.3.2 - Архитектура нейросетевой трансформерной модели JAMPR\*, предложенная нами в статье [5] для создания нейроэвристик для решения задач оптимизации маршрутов с ограничениями реального мира

В [31] мы приводим сравнительный анализ результатов работы нейросетевых и классических алгоритмов для задач оптимизации маршрутов с ограничениями реального мира. Данную систему в будущем планируется адаптировать для решения комбинаторных задач в астрофизических исследованиях.

В статье [5] мы создали нейроэвристики на базе алгоритма JAMPR\* для решения задач построения и оптимизации маршрутов с ограничениями реального мира (временные окна, вместимость транспортных средств, мультидепо) и сравнили ее результаты с популярным эвристическим подходом OrTools на задачах малой и средней размерности (50-1000). В работах [12, 26] мы впервые предложили и протестировали полностью нейросетевые

алгоритмы оптимизации маршрутов в условиях ограничений на задачах большой размерности ( $>1000$ ).

Создаваемые нами модели и технологии, в дальнейшем, планируется адаптировать для решения крупномасштабных задач комбинаторной оптимизации в астрофизике на базе системы SmartRoutes.

## **2.4 Создание систем и технологий для обработки и анализа больших астрономических данных**

В связи с развитием наблюдательной астрономии, в настоящее время, объемы данных полученных телескопами в ходе проведения обзоров всего неба в различных диапазонах ЭМ-спектра быстро растут. В 2019-2022 годах орбитальный телескоп SRG/eROZITA выполнил уникальный обзор всего неба в рентгеновских лучах и открыл новую эру в изучении Вселенной, обнаружив несколько млн. новых рентгеновских источников. Для понимания природы этих объектов необходимо определить их физический класс и другие характеристики, которые не могут быть получены из имеющихся рентгеновских наблюдений. Является актуальной задачи автоматического отождествления больших массивов рентгеновских источников в данных во всех доступных спектральных диапазонах, а также, создания и применения точных моделей физической классификации рентгеновских объектов. Машинное обучение предоставляет исследователям мощные инструменты для решения данного класса задач, но, в настоящее время, они не в полной степени учитывают особенности данных и задач в области рентгеновской астрофизики.

Целью проведенных исследований представленных в работе [28] было создание библиотеки моделей для интеллектуального анализа данных рентгеновских обзоров неба, и включающей в себя методы отождествления, классификации и измерения космологических расстояний (красных смещений) для внегалактических рентгеновских источников. Для задач отождествления рентгеновских источников астрономы используют различные пакеты: e.g. astropy, TOPCAT/STILTS, NWAY. Недостатком указанных подходов является то, что они не используют современные подходы на основе машинного обучения для анализа многомерного пространства признаков астрономических объектов. Библиотека AstroML использует модели scikit-learn для оптической астрономии, но он не адаптирована для работы с рентгеновскими данными. Таким образом, можно сделать вывод, что создание специализированной библиотеки SRGz для решения основных задач анализа рентгеновских данных, с применением современных подходов машинного обучения, является актуальной задачей.

Нами была создана библиотека на языке python, которая включает в себя модели для отождествления рентгеновских источников в оптических каталогах, модели для классификации рентгеновских источников из больших обзоров неба; ансамблевые древовидные модели для измерения фотометрических красных смещений рентгеновских объектов. Лучшие подходы SRGz для работы с рентгеновскими данными опираются на алгоритмы машинного обучения (градиентный бустинг, случайный лес, искусственные нейронные сети). Библиотека содержит инструменты для создания, обучения и тестирования новых моделей, а также, для обработки данных и генерации признаков. Качество моделей классификации рентгеновских источников было протестировано на нескольких выборках объектов eROZITA в области 1% обзора DESI с разным порогом по рентгеновскому потоку (соответствует чувствительности полугодового и 2-х летнего обзора неба eROZITA). Модели оптического отождествления рентгеновских источников были протестированы на данных рентгеновского обзора XMM-NEWTON и оптического обзора PAN-STARRS1. Разработанный пакет предоставляет новый инструмент для обработки и анализа данных в рентгеновской астрономии, позволяя отождествлять каталоги и классифицировать астрономические объекты на основе машинного обучения.

Сеть телескопов обсерватории SKAO (Square Kilometre Array Observatory) в ближайшем будущем дадут возможность проводить обзорные радионаблюдения неба с рекордной

чувствительность, что позволит значительно расширить наши знания о радиоисточниках. Продукты данных SKAO, предоставляемые астрономам, будут соответственно большими и сложными, что потребует применения передовых методов анализа для получения наиболее достоверных научных результатов на их основе. С этой целью SKAO проводит серию проектов Science Data Challenges, каждый из которых призван ознакомить научное сообщество с данными SKAO и стимулировать разработку новых методов анализа. В статье [3] представлены модели глубокого обучения, предложенные для анализа данных конкурса Science Data Challenge 2 (SDC2). Цель работы состояла в поиске объектов и определении их основных наблюдательных свойств для 233 245 источников нейтрального водорода (H I) в смоделированных данных, представляющих собой 2000-часовые наблюдения спектральных линий на красном смещении 0.25-0.5.

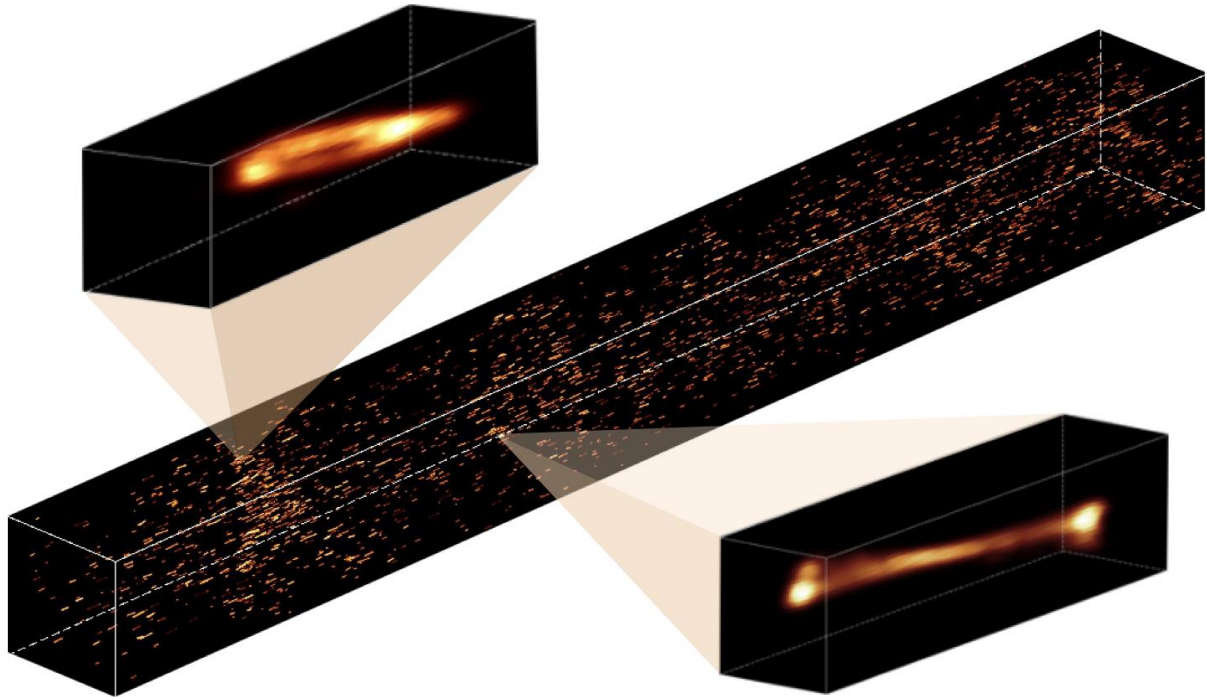


Рисунок 2.4.1 - 3D-изображение демонстрирует небольшой кусок больших радиоданных [3] на которых применялась предложенная нами модель. Куб данных по излучению в линии H I содержит 2683 H I-источника и содержит  $1286 \times 1286 \times 6668$  пикселей для представления поля зрения в 1 квадратный градус во всем рассматриваемом диапазоне частот 0,95-1,15 ГГц (красное смещение 0,235-0,495). Две короткие оси представляют пространственные размеры, а длинная ось - частотные

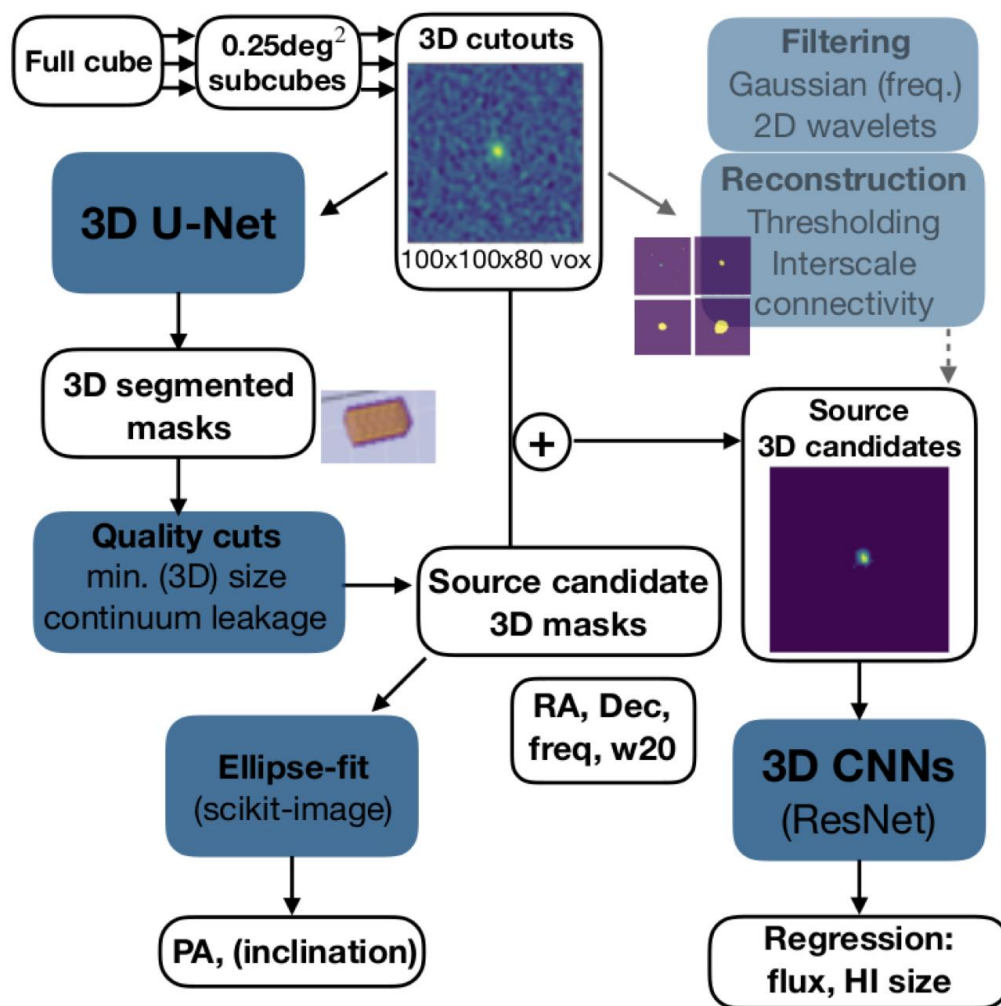


Рисунок 2.4.2 - Конвейер обработки больших радиоданных SKAO в линии 21см (HI), предложенный в статье [3]

В работе [3] был создан прототип конвейера обработки больших радиоданных SKAO. Предложенный конвейер позволил полностью обработать большие данные симуляций. Точность определения параметров радиоисточников составила ~60%. Работа показала, что комбинация алгоритмов машинного обучения дает возможность построить наиболее точные модели детектирования и измерения свойств радиоисточников по данным SKAO.

В дальнейшем, мы планируем развивать новые подходы на базе глубокого обучения в обработке и анализе больших массивов данных собираемых в радиодиапазоне

## 2.7 Заключение

Представленные в настоящей главе результаты показывают, что по направлению «Большие данные в космических исследованиях: астрофизика» проведенные работы соответствуют установленной в плане работ ИКИ РАН на 2023 г. тематике. Таким образом, работы выполнены полностью в соответствии с планом.

При этом получены следующие основные результаты:

- Создана модель глубокого обучения для сегментации сигнала Сюняева - Зельдовича в картах интенсивности ACT+Planck и с ее помощью подготовлен новый каталог скоплений галактик ComPACT содержит 2,934 скопления галактик, из которых до 1436 СЗ-объектов являются новыми (по отношению к существующим суб-мм выборкам скоплений).



- Создан конвейер для обработки и анализа больших радиоданных обсерватории SKAO, включающий в себя нейросетевые модели детектирования радиоисточников в линии 21 см и измерение их характеристик; работа конвейера была протестирована на данных моделирований Science Data Challenge 2 (SDC2).
- Создана библиотека моделей машинного обучения для оптического отождествления и классификации рентгеновских источников, а также, древовидные ансамблевые модели и модели глубокого обучения для измерения их космологических расстояний (красных смещений) по табличным данным многоволновых обзоров неба.
- Создан прототип системы SmartRoutes, предназначенной для решения задач комбинаторной оптимизации при построения маршрутов с учетом ограничений; сравнения нейроэвристик (на основе обучения с подкреплением и других подходов машинного обучения), классических эвристик и точных подходов решения логистических задач.
- Созданы нейроэвристики на базе алгоритма JAMPR\* для решения задач построения и оптимизации маршрутов с ограничениями реального мира (временные окна, вместимость транспортных средств, мультидепо) и проведено сравнение результатами нейроэвристик с популярным эвристическим подходом Google OrTools на задачах малой и средней размерности (50-1000); впервые предложены и протестированы полностью нейросетевые алгоритмы оптимизации маршрутов в условиях ограничений на задачах большой размерности (>1000). Создаваемые нами модели и технологии, в дальнейшем, планируется адаптировать для решения крупномасштабных задач комбинаторной оптимизации в астрофизике на базе системы SmartRoutes.

Таким образом, работы по направлению «Астрофизика» темы Космос-Д в 2023 году были выполнены в полном объеме.

## **ГЛАВА 3 Разработка новых подходов и методов организации ведения и работы со сверхбольшими архивами данных спутниковых наблюдений Земли**

### **3.1 Введение**

Настоящая глава посвящена описанию основных результатов, полученных в рамках проведения работ по направлению «Разработка новых подходов и методов организации сверхбольших архивов данных спутниковых наблюдений Земли и доступа к ним для обеспечения эффективной обработки и анализа данных ДЗЗ для решения научных и прикладных задач». В 2023 году по нему велись работы по следующим задачам плана работ по теме «Космос-Д»:

- Новые подходы к организации автоматизированного ведения архивов и предоставления доступа к ним, включая организацию интерактивной и автоматизированной распределенной обработки предоставляемой ими информации;
- Развитие с использованием созданных методов и подходов различных специализированных научных информационных систем, создаваемых и поддерживаемых в ИКИ РАН, в том числе: «Вега-Science» (система, обеспечивающая возможность распределенной работы со сверхбольшими архивами спутниковых данных ЦКП «ИКИ-Мониторинг»); «ВЕГА-Лес» (система, обеспечивающая ведение научного дистанционного мониторинга лесных ресурсов России), ИС STS (система, обеспечивающая ведение научного дистанционного мониторинга различных явлений в пограничных морях России), ИС VSV (система, обеспечивающая ведение научного дистанционного мониторинга вулканической активности Камчатки и Курил) и др.

В последние годы произошёл существенный качественный и количественный скачок в области применения данных дистанционного зондирования Земли (ДЗЗ) из космоса для решения различных научных и прикладных задач. Он был обусловлен тем, что увеличился объём доступных данных, выросло и дошло до измерительного качество съёмочных систем, устанавливаемых на космических аппаратах (КА), а также развитием технологий восстановления рядов данных, то есть создания непрерывных, очищенных от облачности, сбоев и шумов изображений различной временной скважности вплоть до ежедневной. Анализ таких рядов данных позволил успешно решать задачи анализа динамики сельскохозяйственной растительности, наблюдения наземных экосистем и многие другие. В то же время, несмотря на все плюсы использования такого подхода, у него есть один существенный недостаток: серьёзное, в некоторых случаях более чем трехкратное увеличение объёма хранимых данных, что приводит к дополнительным нагрузкам на системы хранения и обработки данных.

Таким образом, всё более актуальной становится задача оптимизации хранения подобных рядов данных, решение которой позволило бы, с одной стороны, сократить объёмы хранимой информации и, с другой, не потерять в её качестве и измерительных свойствах. При этом для новых схем организации хранения восстановленных рядов данных по-прежнему актуальны требования эффективного доступа к ним, что, в первую очередь, важно для работы пользовательских картографических веб-интерфейсов. В 2023 году был проведен комплексный анализ различных подходов к организации более эффективных схем хранения долговременных рядов спутниковых наблюдений.

Новым направлением работ в 2023 году стала организация работы с данными архивов ЦКП «ИКИ-Мониторинг» на основе использования интерактивной среды разработки Jupyter Notebook. В настоящее время эта среда наряду со следующим ее поколением Jupyter Lab широко используется исследователями применительно к самым разным задачам, связанным с

анализом и обработкой данных. Широкое применение она находит и для решения различных задач, связанных с анализом спутниковой информации ДЗЗ. Хотя рассматриваемая среда обычно не используется для разработки объемного программного кода, она хорошо подходит для решения различных частных задач, когда необходимо проверить какую-то гипотезу, алгоритмическое решение или быстро провести вычислительный эксперимент. Ключевым преимуществом такого решения является удобство визуализации получаемых результатов непосредственно рядом с программным кодом, что также позволяет упростить взаимодействие разных специалистов. В 2023 году был реализован базовый функционал по работе с данными архивов ЦКП «ИКИ-Мониторинг» непосредственно по месту их физического хранения.

Также в 2023 году были продолжены работы по развитию нейросетевых подходов для автоматического построения масок облачности и теней на спутниковых изображениях приборов серии КМСС, установленных на российских спутниках серии «Метеор-М». Эти отечественные данные обладают большой ценностью благодаря редкому сочетанию достаточно высокого пространственного разрешения и широкой полосы захвата. Для их использования в целях мониторинга поверхности Земли, также как и практически для любых других данных ДЗЗ, очень актуальна задача фильтрации облачности и теней.

Важным направлением работ в 2023 году была модернизация функционирующей в составе ЦКП «ИКИ-Мониторинг» системы распределенной многопоточковой обработки спутниковых данных, в работе которой в настоящее время задействовано свыше 200 вычислительных узлов. Помимо сравнительно небольших улучшений в этом году были начаты работы по существенной модернизации этой системы, которая должна привести к повышению эффективности ее работы в ситуации со все возрастающими объемами обработки.

В 2023 году активно велись работы по модернизации системы ведения архивов, а также инструментов для контроля ее функционирования. Основной акцент был сделан на увеличении скорости усвоения данных в архивы, что позволило при необходимости пополнять архивы на объёмы вплоть до 15 Тб новых данных в сутки. В этом году были также разработаны новые инструменты для контроля поступления различных типов спутниковых данных в архивы ЦКП «ИКИ-Мониторинг».

В текущем году также активно велись работы по разработке новых подходов к созданию различных программных компонент, используемых для построения специализированных научных информационных систем, в том числе: ИС «ВЕГА-Лес» (комплексный научный мониторинг лесных ресурсов России), ИС STS (комплексный научный мониторинг различных явлений в пограничных морях России), ИС VSV (комплексный научный мониторинг вулканической активности Камчатки и Курил). Основные работы по теме «Космос-Д» велись в направлении модернизации системы анализа долговременных временных серий данных, получаемых по спутниковым наблюдениям. Благодаря этому в перечисленных информационных системах пользователям теперь доступен гораздо более широкий набор данных для анализа. А в перспективе планируется реализовать инструменты анализа временных серий для всех типов данных, содержащихся в архивах ЦКП «ИКИ-Мониторинг». Работы по расширению доступного пользователям функционала, а также по уменьшению задержек получения требуемых рядов данных были проведены и для информационной системы комплексного научного мониторинга вулканической активности Камчатки и Курил.

Полученные в рамках работ 2023 года результаты по данному направлению изложены в следующих публикациях [6, 7, 13, 14, 16-19].

### **3.2 Новые подходы организации автоматизированного ведения архивов и предоставления доступа к ним, включая организацию интерактивной и автоматизированной распределенной обработки предоставляемой ими информации**

### 3.2.1 Анализ различных подходов к оптимизации хранения временных рядов данных ДЗЗ

Многие задачи анализа состояния и динамики различных объектов и процессов на поверхности Земли практически невозможно решить путём анализа одного или нескольких изображений. Характерные периоды, в течение которых необходимо наблюдать за объектом исследования, могут составлять месяцы и годы, при этом зачастую необходимы наблюдения с заданной периодичностью, определяемой скоростью изменения во времени этого объекта. В частности, для анализа лесного покрова оптимальным является период наблюдения от недели до десяти дней, а для анализа динамики сельскохозяйственной растительности в общем случае достаточным является недельный промежуток, но в определённые моменты бурного роста или уборки урожая требуются как минимум четырёхдневные, а лучше даже ежедневные наблюдения. Другие задачи также могут требовать своей, отличной от указанной периодичности. По мере роста числа подобных задач использование уникальных временных рядов под каждую задачу становится неэффективным, так как количество используемых спутниковых систем и приборов является ограниченным, и такие расчёты просто дублируют друг друга на одних и тех же данных. Развитие технологий восстановления данных позволило перейти к использованию универсальных, ежедневных, очищенных от облачности, помех и шумов и восстановленных рядов данных.

Несмотря на все достоинства, подобные ряды имеют один существенный недостаток – серьёзный рост объёма по сравнению с исходными данными. В таблице 3.2.1.1 приведены характерные размеры исходных данных и восстановленных ежедневных композитов для приборов MSI, MODIS и KMCC, установленных на КА Sentinel-2, TERRA/AQUA и «Метеор-М» №2-2 соответственно. Все данные имеют формат Int16 и сжаты алгоритмом Deflate пакета GDAL с параметрами по умолчанию.

Таблица 3.2.1.1. Сравнение размеров исходных данных и восстановленных рядов. Данные взяты на территорию Московской области за 2023 год.

Данные	Объем исходных данных	Объем восстановленных композитов
Sentinel-2/MSI	~ 1 Тб	~1,3 Тб
MODIS	~ 33 Гб	~ 102 Гб
«Метеор-М»/KMCC	~ 123 Гб	~ 226 Гб

Представленная статистика показывает, что в ряде случаев объем восстановленных данных в ряде случаев больше объема исходных данных более чем в два раза. Учитывая тот факт, что исходные сцены спутниковых данных также представляют большой интерес и хранятся в архивах, это может привести к троекратному увеличению объема, занимаемого соответствующим типом данных. В настоящее время восстановленные ряды данных, как правило, формируются только по ограниченному набору каналов спутникового прибора, поэтому на общий объем архивов это влияет не так радикально. Так в настоящее время объем занимаемый рассматриваемыми рядами данных находится в пределах 500 терабайт, притом, что суммарный объем данных в архивах ЦКП «ИКИ-Мониторинг» приближается к 6 петабайтам. В то же время, в ближайшие годы планируется существенно увеличить как временное, так и пространственное покрытие ежедневными восстановленными рядами данных, получаемыми на основе данных высокого разрешения, что приведет к многократному увеличению объема, который они занимают в архивах. Таким образом, задача минимизации объемов ежедневных композитов становится крайне актуальной. При этом наряду с минимизацией объема данных важным требованием является максимально быстрая визуализация сжатых изображений. Кроме того, решение данной проблемы не должно приводить к серьёзной перестройке всей действующей инфраструктуры хранения, обработки и предоставления данных. Далее в статье рассмотрены варианты такого решения.

В 2023 году с целью выработки новых подходов к хранению долговременных рядов спутниковых наблюдений в архивах Центра коллективного пользования «ИКИ-Мониторинг» проводились работы по трем основным направлениям. Во-первых, был проведен анализ широко распространённых алгоритмов сжатия изображений без потери качества. Во-вторых, рассматривались варианты, связанные с изменением схемы хранения данных в архиве, а именно с использованием опорных и инкрементальных экземпляров данных. В рамках третьего направления работ рассматривались варианты с использованием алгоритмов сжатия растровых данных с потерями качества, но при этом с жестко лимитированной ошибкой данных в каждом пикселе.

На сегодняшний день реализовано достаточно большое количество алгоритмов сжатия растровых данных без потери их качества. Среди достаточно распространенных алгоритмов наибольшей степенью сжатия отличается алгоритм JPEG2000. Однако как сжатие данных по этому алгоритму, так и, что более критично, их восстановление требует существенных вычислительных ресурсов. Так как в рамках используемых нами сценариев использования рядов данных важна быстрая визуализация изображений в картографическом интерфейсе, этот алгоритм оказывается неприемлемым. Большинство остальных алгоритмов, таких как LZW, Deflate, LZMA, ZSTD и построено на базе алгоритмов Лемпеля-Зива-Уэлча и применения кодов Хаффмана. На текущем этапе развития их эффективность и быстродействие более-менее сопоставимы и зависят от природы сжимаемых данных. На рисунке 3.2.1.1 приведено сравнение эффективности различных алгоритмов без потери качества при разных сценариях использования. Практически все подобные алгоритмы имеют некоторый предел эффективности степени сжатия порядка 2,3 для данных Int16 и порядка 1,3 для данных Float32, выше которого можно перейти только в редких случаях (см., например <https://kokoalberti.com/articles/geotiff-compression-optimization-guide>). На основе сравнительного анализа оптимальным по совокупности характеристик степени сжатия и времени восстановления данных нами был выбран метод DEFLATE с указанием дополнительных параметров «TILED=YES PREDICTOR=2». Этот метод сжатия и ранее нами использовался как основной, но при этом не всегда были указаны оптимальные параметры, что ухудшало эффективность алгоритма примерно на 10%. В настоящее время ведутся работы по унификации этих параметров во множестве процедур обработки данных.

Compression ratio					
Algorithm			Test file		
	config.ini	predictor	Z-level	byte	int16
none	-	-	1.00	1.00	1.00
packbits	-	-	1.05	0.99	0.99
deflate	1	6	2.12	1.57	1.19
deflate_pred2	2	6	2.51	2.30	1.35
deflate_pred3	3	6	-	-	1.64
deflate_zlev9	1	9	2.12	1.57	1.19
deflate_zlev1	1	1	2.06	1.55	1.19
deflate_zlev1_pred2	2	1	2.34	2.24	1.35
deflate_zlev1_pred3	3	1	-	-	1.60
lzma	-	-	2.50	1.70	1.05
lzw	1	-	1.98	1.20	0.96
lzw_pred2	2	-	2.55	2.21	1.12
lzw_pred3	3	-	-	-	1.30
zstd	1	9	1.89	1.50	1.19
zstd_pred2	2	9	2.64	2.33	1.38
zstd_pred3	3	9	-	-	1.64
zstd_zlev15	1	15	1.89	1.50	1.19
zstd_zlev1	1	1	1.66	1.39	1.19
zstd_zlev1_pred2	2	1	2.85	2.18	1.38
zstd_zlev1_pred3	3	1	-	-	1.61

Рисунок 3.2.1.1 – Сравнение эффективности сжатия разных алгоритмов без потерь качества

Специфика временных рядов данных такова, что они, как правило, отражают процессы, изменяющиеся плавно, без серьёзных скачков. На рисунке 3.2.1.2 приведены графики годового хода яркости красного (RED) и ближнего инфракрасного (NIR) спектральных каналов прибора MODIS, а также рассчитанного на их основе нормализованного разностного вегетационного индекса (NDVI). Видно, что изменения каждого следующего значения относительно предыдущего сравнительно невелики.

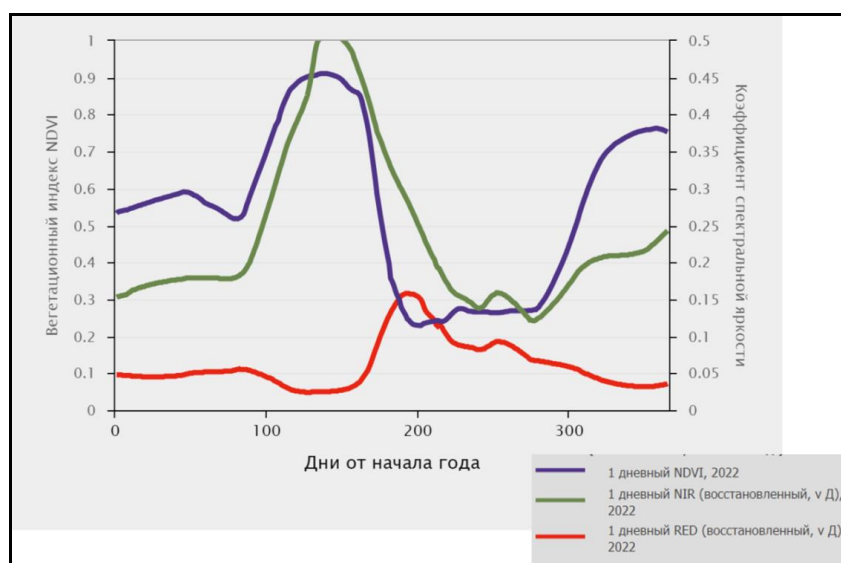


Рисунок 3.2.1.2 - Графики хода спектральных каналов и индекса NDVI

Исходя из этого, эффективной оказывается «разностная» схема хранения данных, в рамках которой в архив заносятся опорные изображения с заданным временным шагом, а в промежутках - разности между исходными и опорными изображениями (см. рисунок 3.2.1.3). Такой подход позволяет существенно снизить объём хранимых данных за счёт резкого снижения размерности и диапазона значений в промежуточных изображениях, тем самым увеличивая эффективность работы алгоритмов сжатия на них. Проверка работоспособности и эффективности предложенной схемы была проведена на восстановленных наборах данных прибора MODIS в каналах RED, NIR, а также индекса NDVI. Были взяты наборы за 2022 год по трём различным областям на территории России. Карта расположения тестовых наборов приведена на рисунке 3.2.1.3.

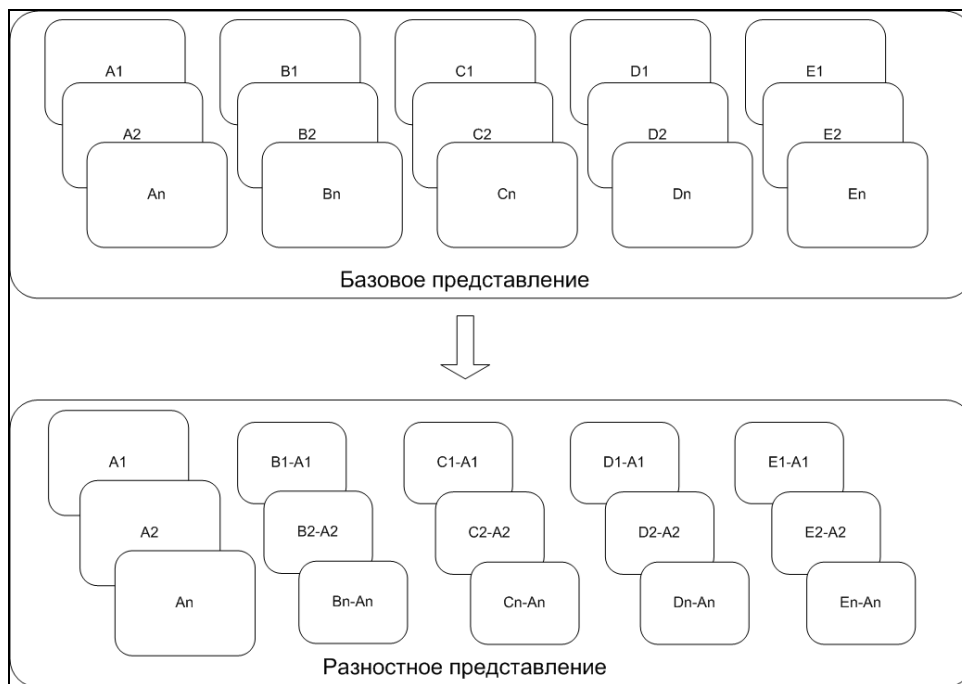


Рисунок 3.2.1.3 - Базовая и разностные схемы хранения данных. А, В, С, D, Е – экземпляры данных на временном промежутке

Проверка работоспособности и эффективности предложенной схемы была проведена на восстановленных наборах данных прибора MODIS в каналах RED, NIR, а также индекса NDVI. Были взяты наборы за 2022 год по трём различным областям на территории России, которые соответствуют существенно отличающимся по наблюдаемым характеристикам регионам. Карта расположения тестовых наборов приведена на рисунке 3.2.1.4.

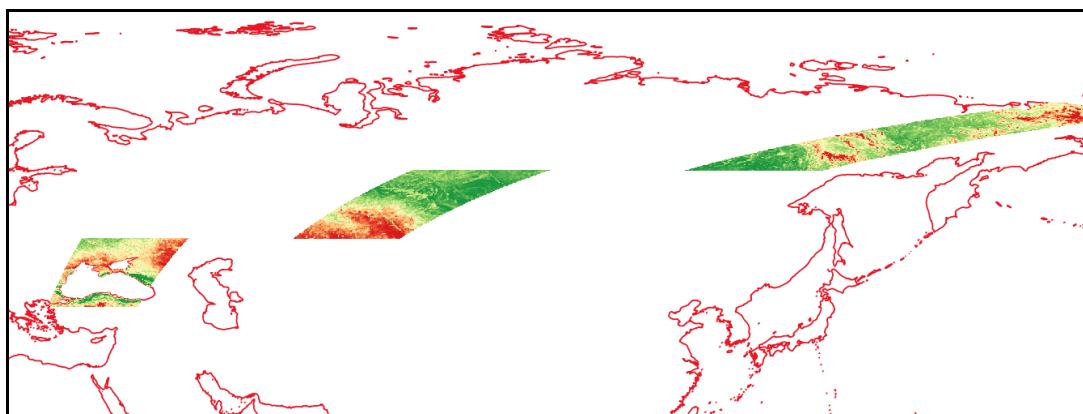


Рисунок 3.2.1.4 - Карта расположения тестовых наборов данных. Взяты гранулы h20v04 (слева), h22v03 (в центре) и h24v02 синусоидального разбиения данных MODIS

Для каждого набора были построены разностные временные ряды с шагами в 1, 2, 3, 4, 5, 10 и 15 дней. Полученные результаты по сжатию данных приведены в таблицах 3.2.1.2, 3.2.1.3. и 3.2.1.4. Как показали эксперименты, в результате применения предложенной схемы достигается экономия от 20% до 50% относительно исходного объема ряда. Очевидно, что степень сжатия варьируется в зависимости от выбранного шага между опорными изображениями, а также от территории наблюдения и особенностей расположенных на этих территориях объектов. Поэтому следующим шагом исследований в данном направлении является определение оптимального, возможно переменного, шага для разных типов данных и территорий. Кроме того, важным вопросом является оценка потерь в быстродействии при работе с данными, хранимыми по такой схеме.

Таблица 3.2.1.2. Результаты для индекса NDVI. Для каждого шага приведён объём полученного ряда в Гб, а также процент относительно исходного объёма

Тайл	Исходный ряд, объём в Гб	Шаг 2,	Шаг 3	Шаг 4	Шаг 5	Шаг 10	Шаг 15
h20v04	8,92	7,25 / 81%	6,96 / 78%	6,92 / 78%	6,95 / 78%	7,23 / 81%	7,45 / 84%
h22v03	10,75	8,11 / 75%	7,54 / 70%	7,39 / 69%	7,38 / 69%	7,64 / 71%	7,91 / 74%
h24v02	8,57	6,18 / 72%	5,61 / 65%	5,42 / 63%	5,37 / 63%	5,5 / 64%	5,68 / 66%

Таблица 3.2.1.3. Результаты для канала RED. Для каждого шага приведён объём полученного ряда в Гб, а также процент относительно исходного объёма

Тайл	Исходный ряд объём в Гб	Шаг 2	Шаг 3	Шаг 4	Шаг 5	Шаг 10	Шаг 15
h20v04	4,35	3,24 / 74%	3,03 / 70%	2,99 / 69%	3 / 69%	3,19 / 73%	3,35 / 77%
h22v03	6,53	4,37 / 67%	3,83 / 59%	3,65 / 56%	3,59 / 55%	3,71 / 57%	3,89 / 60%
h24v02	5,7	3,55 / 62%	2,97 / 52%	2,75 / 48%	2,65 / 46%	2,63 / 46%	2,75 / 48%

Таблица 3.2.1.4. Результаты для канала NIR. Для каждого шага приведён объём полученного ряда в Гб, а также процент относительно исходного объёма

Тайл	Исходный ряд объём в Гб	Шаг 2	Шаг 3	Шаг 4	Шаг 5	Шаг 10	Шаг 15
h20v04	4,79	3,78 / 79%	3,59 / 75%	3,57 / 75%	3,58 / 75%	3,76 / 78%	3,9 / 81%
h22v03	7,13	5,03 / 71%	4,54 / 64%	4,39 / 62%	4,35 / 61%	4,5 / 63%	4,68 / 66%
h24v02	6,37	4,17 / 65%	3,59 / 56%	3,38 / 53%	3,29 / 52%	3,3 / 52%	3,44 / 54%

Другим возможным вариантом оптимизации хранения временных рядом данных является использование алгоритмов сжатия с потерей качества данных, таких, как JPEG. Как правило, такие алгоритмы применяются только для задач визуализации данных, не требующих какого-либо анализа или вычислений, так как ошибка в данных не регламентируется явным образом, что неприемлемо для большинства задач численного анализа данных. Тем не менее, сегодня существуют алгоритмы, позволяющие задавать не целевой объём получаемого файла



при сжатии, а максимально допустимый размер ошибки, вносимой в данные при сжатии. Таким алгоритмом является LERC – Limited Error Raster Compression (алгоритм сжатия растров с фиксированной ошибкой). При сжатии изображений алгоритму явно указывается значение максимальной ошибки в любом из пикселей – MaxError. В процессе работы алгоритма сначала строится маска «данные – нет данных», затем данные разбиваются на небольшие блоки, а в каждом из блоков минимальное значащее значение пиксела берётся за основу, вычисляется разность остальных значащих пикселей с ним, делится на  $2 * \text{MaxError}$  и округляется, после чего блок сжимается без потерь. Таким образом, достигается потеря размерности данных за счет гарантированного перехода от данных любого формата к целочисленным данным, и существенное уменьшение словарей для последующего сжатия без потерь, что и приводит к очень высоким степеням сжатия. Результаты применения алгоритма LERC к тем же тестовым наборам данных индекса NDVI приведены в таблице 3.2.1.5.

Таблица 3.2.1.5. Результаты сжатия временных рядов индекса NDVI алгоритмом LERC с различными максимальными ошибками. Ошибка дана в абсолютных значениях, диапазон значений данных – от 0 до 1. Для каждого значения ошибки приведён объём полученного ряда в Гб, а также процент относительно исходного объёма

Тайл	Исходный ряд объём в Гб	Макс. ошибка 0,01	Макс. ошибка 0,02	Макс. ошибка 0,03	Макс. ошибка 0,04	Макс. ошибка 0,05
h20v04	8,92	4,6 / 52%	4 / 45%	3,6 / 40%	3,4 / 38%	3,2 / 36%
h22v03	10,75	4,8 / 45%	4,1 / 38%	3,7 / 34%	3,5 / 33%	3,3 / 31%
h24v02	8,57	4,2 / 49%	3,6 / 42%	3,3 / 39%	3,1 / 36%	2,9 / 34%

Получается, что даже сжатие с ошибкой 1% от диапазона возможных значений для тестовой выборки приводит к экономии порядка 50% от исходного объёма данных. Увеличение максимально допустимой ошибки, очевидно, приводит к дальнейшему росту степени сжатия, но уже не такому быстрому.

Так как наиболее критичными с точки зрения минимизации объема являются временные ряды, получаемые по данным высокого разрешения, для них были получена более детальная статистика. Для эксперимента использовался ежедневный композит индекса NDVI за лето 2022 года, полученные по данным спутников серии Sentinel 2 по небольшому фрагменту в европейской части России (всего 369 файлов с изображениями). Полученные результаты приведены на графике, приведенном на рисунке 3.2.1.5. На графике красной пунктирной линией как 100 % показан объем несжатых данных, зеленой – объем данных, сжатых при помощи алгоритма DEFLATE, а синей - процент от исходного объема, получающийся при сжатии алгоритмом LERC при разных значениях максимальной ошибки. При этом по оси абсцисс указана ошибка как процент от возможного диапазона значений в исходных файлах. Также на графике оранжевым цветом отображена кривая, показывающая объем, получающийся после применения алгоритма DEFLATE к данным, которые были сжаты алгоритмом LERC. Интерес к этой зависимости вызван соображениями обратной совместимости, так как алгоритм сжатия LERC поддерживается только в новых версиях используемых нами программных пакетов (прежде всего это пакет GDAL) и не во всех программных окружениях. Из графика видно, что при использовании алгоритма LERC с максимальной ошибкой в 1% от диапазона значений достигается экономия дискового пространства более чем в 3 раза, и даже при использовании компромиссного варианта – более чем в два раза. Даже при использовании минимальной ошибки в 0.01% алгоритм LERC позволяет сэкономить около 15% объема хранения.

Аргументация для выбора анализируемого диапазона ошибок была следующей. По оценкам как специалистов ИКИ РАН, так и информации из внешних источников, ошибка в пикселях при получении восстановленных ежедневных временных рядов составляет около 5%. Точно ее оценить проблематично, так как она очень существенным образом зависит от

количества безоблачных наблюдений в исходных данных для конкретной географической области. К примеру, если интересующий нас регион в летний сезон был целый месяц закрыт облаками, то ошибка может быть радикально больше.

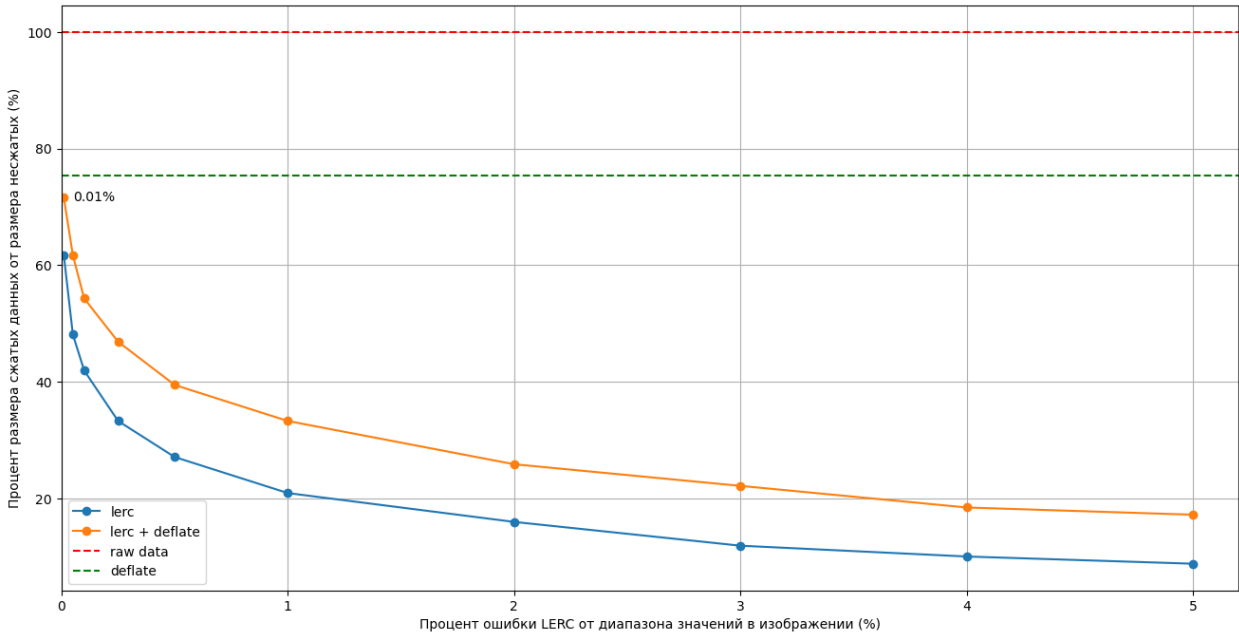


Рисунок 3.2.1.5 - Эффективность сжатия ежедневных композитов индекса NDVI, полученного по данным спутников серии Sentinel-2

Визуально изображения после применения сжатия алгоритма LERC с ошибкой в 1% практически неотличимы, а при величине ошибки 5% - различия уже достаточно хорошо заметны. Конечно, такой критерий для оценки качества данных не является подходящим, несмотря на то, что ошибка в каждом пикселе строго лимитирована. По этой причине были рассмотрены основные сценарии использования этих данных и влияние на них вносимых при сжатии ошибок.

Пожалуй, основным вариантом использования временных рядов спутниковых данных является отслеживание динамики средних значений по объектам наблюдения, в качестве которых могут выступать, к примеру, сельскохозяйственные поля, в качестве исследуемой величины – индекс NDVI. Так как ошибка в каждом пикселе лимитирована и разнонаправлена, то при получении средних значений по объектам наблюдения мы ожидаемо получаем гораздо меньшую ошибку. Для подтверждения этого факта на тестовом наборе данных MODIS, описанном выше, было проведено усреднение значений по всем сельскохозяйственным полям, попавшим на эти территории, на исходных и сжатых с разной степенью ошибки данных. Выборка включала в себя 768 472 поля разных размеров. При сопоставлении полученных результатов были получены следующие зависимости, представленные на рисунке 3.2.1.6. Видно, что при увеличении размера объекта средняя ошибка снижается и достигает значений, меньших половины от заданной. Для заданной ошибки 0,01 средняя ошибка по полям составляет около 0,005, что существенно превышает точность данных в восстановленных рядах.

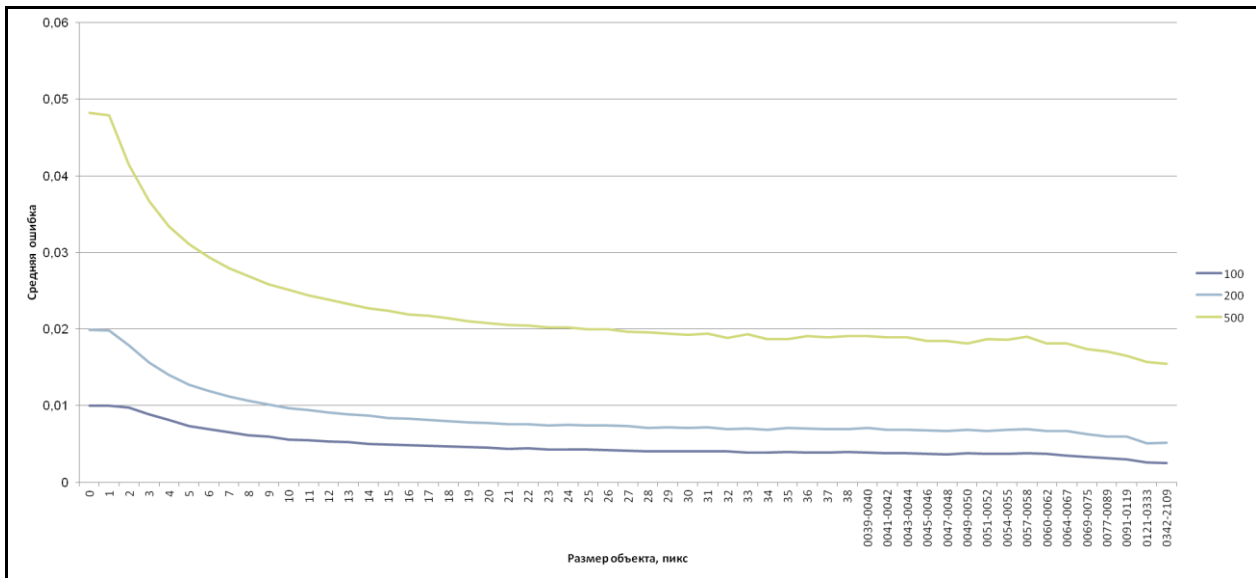


Рисунок 3.2.1.6 - Зависимость средней ошибки по полям от размера поля в пикселях для размеров ошибки 0,01, 0,02, 0,05 (на графике значения 0-1 перенормированы в 0-10000 для перехода к целым)

Другой, не менее важный класс задач связан с детектированием объектов разного типа на изображениях. И в этом случае нас интересует не только приемлемая ошибка в отдельных пикселях изображения, но и сохранение пространственной структуры изображений.

Для исследования влияния сжатия алгоритмом LERC на пространственную структуру изображений было проведено сравнение результатов выполнения различных классических алгоритмов классификации, кластеризации и сегментации, полученных по исходному фрагменту и по сжатым с разной ошибкой изображениям. Для сравнения полученных результатов был разработан специальный программный модуль, который позволяет получить процент пикселей, которые на сжатом изображении попали в такие же сегменты или классы, что и на исходном. При помощи этого модуля были получены результаты для большого количества различных фрагментов спутниковых изображений. Однако, хотя общая закономерность, заключающаяся в том, что процент совпадения падает с увеличением ошибки, наблюдалась во всех используемых нами алгоритмах анализа изображений, количественно оценки оказались неубедительными, так как очень сильно зависели как от используемых параметров алгоритмов, так и от конкретного изображения. Например, при использовании алгоритма сегментации методом k-means, разница в результатах, получаемых по исходному и сжатому изображениям, очень сильно зависит от заданного количества сегментов. В первую очередь это связано с неустойчивостью результатов этого класса алгоритмов применительно к реальным спутниковым изображениям, так на них может существенно повлиять даже самое незначительное изменение анализируемого изображения. По описанной причине от такого подхода к анализу пространственных искажений, вносимых алгоритмом LERC, пришлось отказаться.

В рамках проведения дальнейшего анализа были проанализированы изображения, получаемые как разница между исходными данными и сжатыми алгоритмом LERC с разной максимальной ошибкой. Существенно, что характеристики таких изображений сильно зависят не только от конкретных типов спутниковых данных, но и от конкретного региона наблюдения. Однако на качественном уровне все закономерности сохраняются. Для иллюстрации в настоящем отчете был взят фрагмент изображения 200x200 индекса NDVI, полученного по данным спутника Sentinel-2 (см. рисунок 3.2.1.7).

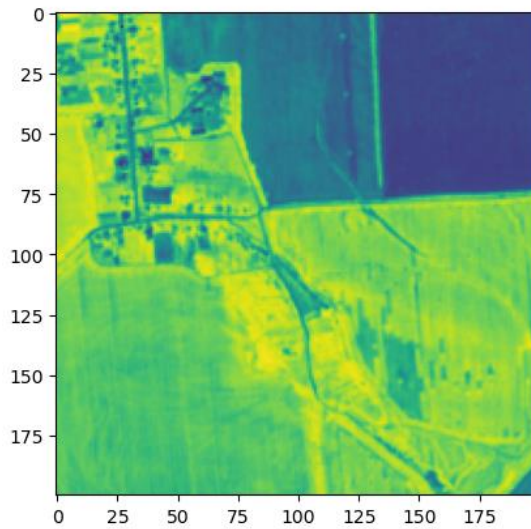


Рисунок 3.2.1.7 - Тестовый фрагмент изображения NDVI по данным спутника Sentinel-2A

На рисунке 3.2.1.8 приведено изображение, полученное вычитанием фрагмента, сжатого с ошибкой в 1%, из исходного фрагмента, а справа показана гистограмма значений для него. Визуально изображение похоже на шум, что подтверждается его гистограммой, на которой явно выделяется только нулевое значение. Отметим, что для иллюстрации использовался небольшой фрагмент, а при больших размерах изображений гистограмма получается еще более равномерной.

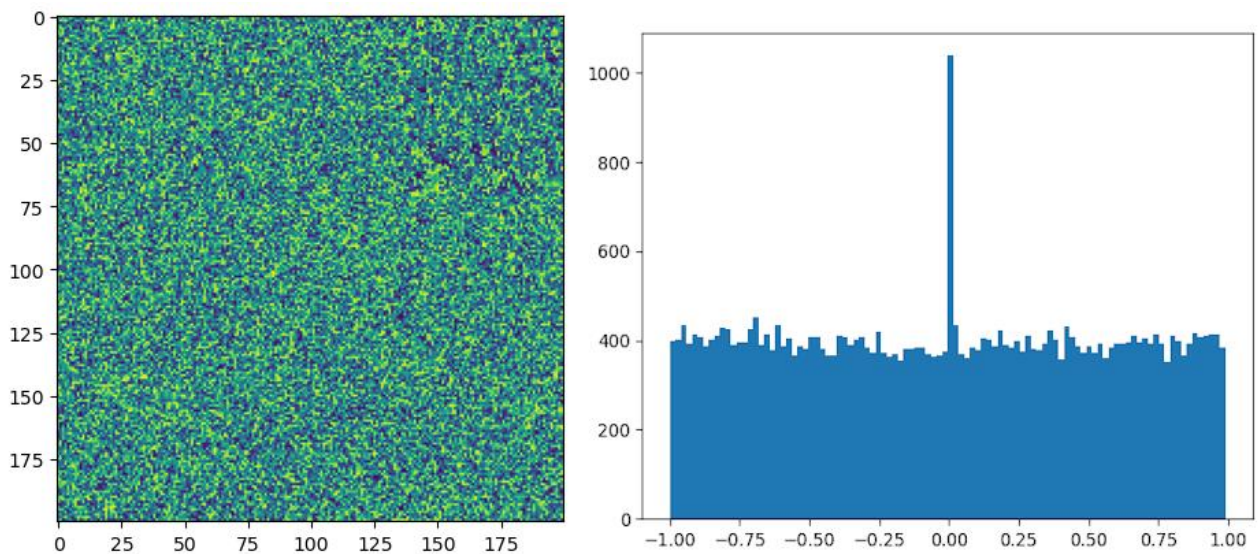


Рисунок 3.2.1.8 - Разница исходного фрагмента и фрагмента, сжатого алгоритмом LERC с ошибкой в 1% от диапазона, значений и гистограмма значений для неё

Однако при увеличении ошибки на изображениях с разницей начинают появляться неоднородности, так же, как и на соответствующих им гистограммах. На рисунке 3.2.1.9 приведено такое изображение, полученное вычитанием фрагмента, сжатого с ошибкой в 5%, из исходного фрагмента, а справа показана его гистограмма значений. Не вызывает сомнений тот факт, что при таком сжатии пространственная структура исходного изображения была существенно повреждена. Также отчетливо начинает проявляться эффект с занижением значений в сжатом изображении, который возрастает с увеличением заданной ошибки, что, судя по всему, является особенностью используемого алгоритма сжатия.

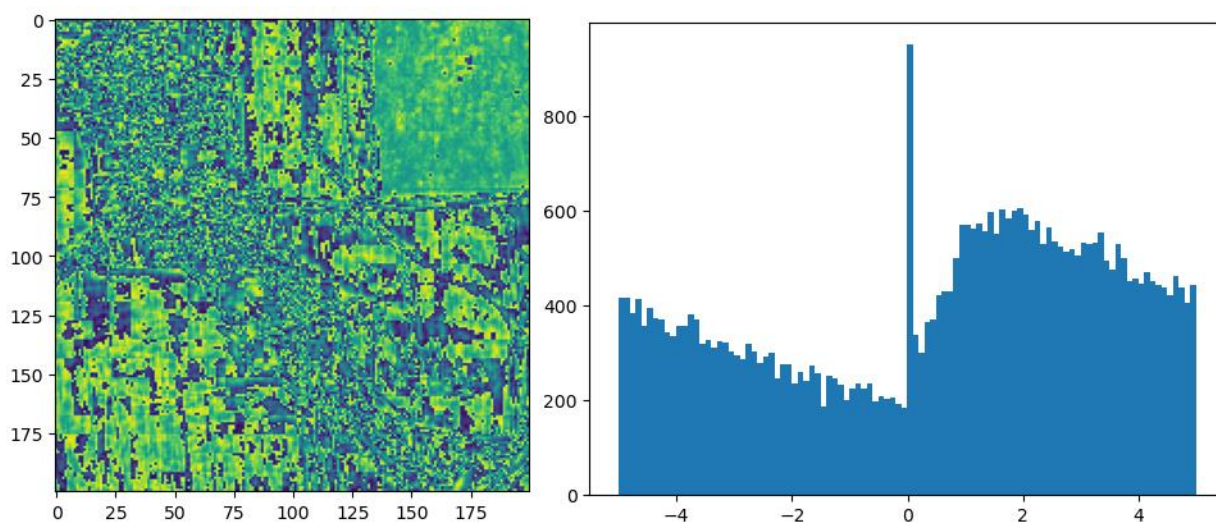


Рисунок 3.2.1.9 - Разница исходного фрагмента и фрагмента сжатого алгоритмом LERC с ошибкой в 5% от диапазона значений и соответствующая гистограмма значений

Для программного обнаружения неоднородностей в пространственной разнице между исходным и сжатым изображениями была разработана следующая методика. Изображение с разницей делится на блоки с заданным шагом и для каждого из блоков вычисляется среднее значение ошибки, после чего анализируется гистограмма полученных значений. При использовании различного шага, сопоставимого с размерами детектируемых объектов, были получены в целом схожие результаты. На рисунке 3.2.1.10 приведены графики ядерной оценки плотности вероятности, полученные по гистограммам, соответствующим разнице между исходным фрагментом и фрагментами, сжатыми при помощи алгоритма LERC с указанием ошибки в 1, 2 и 5 процентов. Видно, что график, соответствующий ошибке в 1% от диапазона, почти полностью симметричен и близок к нормальному распределению, что свидетельствует о том, что вносимая в блоки изображения ошибка распределена практически случайным образом. При увеличении ошибки алгоритма LERC начинает наблюдаться явная асимметрия графика, и для ошибки в 5% можно утверждать, что пространственная информация в сжатом изображении была существенно искажена. Таким образом, хотя для вычисления средних значений по объектам наблюдения ошибка алгоритма LERC в 5% является вполне приемлемой, такой вариант является неприемлемым для задач, связанных с детектированием объектов.

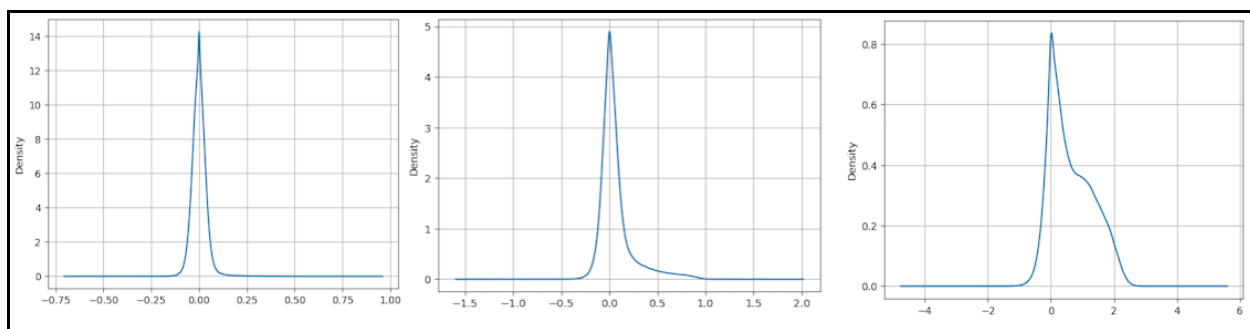


Рисунок 3.2.1.10 - Ядерная плотность вероятности значений в разнице исходного и сжатого изображений для ошибок LERC в 1, 2 и 5 процентов

В настоящее время ведутся работы по разработке методики сжатия восстановленных рядов данных с использованием как «разностной» схемы организации их хранения, так и хорошо показавшего себя алгоритма сжатия с регулируемой ошибкой LERC. Внедрение этой методики позволит уменьшить объем хранимых данных этого типа как минимум вдвое.

### ***3.2.2 Организация работы с данными архивов ЦКП «ИКИ-Мониторинг» на основе использования интерактивной среды разработки Jupyter Notebook***

Интерактивная среда разработки Jupyter Notebook в последние годы находит все более широкое применение для широкого спектра задач, связанных с анализом и обработкой данных, в том числе данных дистанционного зондирования Земли. Она, как и следующее ее поколение Jupyter Lab, реализована в виде удобного веб-интерфейса, позволяющего запускать отдельные участки программного кода и в этом же интерфейсе наглядно представлять результаты операций в виде графиков, изображений или таблиц. Этот инструмент особенно хорошо подходит для решения частных задач, связанных с анализом данных, выработкой алгоритмов, проверкой гипотез и т.п. При этом реализованные в среде документы могут быть наглядным и интерактивным представлением полученных результатов и для других исследователей и разработчиков. Благодаря наличию широкого спектра библиотек, предназначенных для анализа и обработки самых разных типов данных, язык программирования Python в настоящее время широко используется не только для прикладных, но и исследовательских задач, в том числе связанных с обработкой спутниковых изображений. Поэтому ниже рассматривается использование именно этого языка программирования в рамках среды Jupyter.

В 2023 году в отделе «Технологии спутникового мониторинга» ИКИ РАН в рамках использования среды Jupyter был реализован базовый функционал для получения доступа к данным сверхбольших распределенных архивов спутниковых данных ЦКП «ИКИ-Мониторинг», позволяющий производить анализ и обработку данных непосредственно по месту их физического хранения с использованием различных сторонних библиотек (gdal, numpy, matplotlib, scipy и др.). Далее более подробно рассмотрены вопросы организации доступа как к необходимой справочной информации, описывающей имеющиеся в архиве спутниковые данные, так и к самим файлам, содержащим спутниковые наблюдения.

Для эффективной работы с данными ДЗЗ необходим широкий спектр справочной информации, описывающей спутники, размещенные на них приборы наблюдения и характеристики конкретных информационных продуктов. В рамках разработанной в ИКИ РАН технологии построения сверхбольших распределенных архивов спутниковых данных UNISAT за хранение такой информации отвечает справочная база данных UNISAT\_CATALOG. Для доступа к этой информации на языке программирования python была разработана библиотека **unisat\_catalog.py**, позволяющая получать различную справочную информацию как в табличном виде, так и в виде структур данных. На рисунке 3.2.2.1 приведены примеры получения справочной информации о продукте, спектральных каналах прибора наблюдения, а также о характеристиках каналов конкретных информационных продуктов, имеющихся в архивах.

В рамках технологии UNISAT хранение метаданных, описывающих имеющиеся в архивах экземпляры спутниковых продуктов, реализовано в наборе баз данных unisat, имеющих одинаковую унифицированную структуру. В настоящее время за ведение метаданных архивов ЦКП «ИКИ-Мониторинг» отвечают пять основных баз данных: unisat\_hrsat (данные высокого разрешения), unisat\_mrsat (среднего разрешения), unisat\_comp (композиционные продукты), unisat\_radarsat (радарные наблюдения), unisat\_user (пользовательские продукты). Для доступа к метаданным таких архивов была разработана библиотека **unisat\_archive.py**.

```
[1]: import unisat_catalog as uc

[2]: uc.product(product="rgb", output="table")
```

product	product_type	product_group	name_en	desc_en	is_metadata
0	rgb	product	seance_syntheses	Surface Pseudo-color surface image	0

```
[3]: uc.device_bands("MSI", output="table", limit=30)
```

band_name	type	wl_min	wl_max	resolution	
0	1	Coastal aerosol	433.0	453.0	60.0
1	2	Blue	456.0	523.0	10.0
2	3	Green	542.0	577.0	10.0
3	4	Red1	650.0	680.0	10.0
4	5	Red2	698.0	713.0	20.0
5	6	Red3	732.0	747.0	20.0
6	7	Red4	793.0	773.0	20.0
7	8	NIR	785.0	900.0	10.0
8	8A	Narrow NIR1	855.0	875.0	20.0
9	9	Narrow NIR2	935.0	955.0	60.0
10	10	Cirrus	1365.0	1395.0	60.0
11	11	SWIR1	1565.0	1655.0	20.0
12	12	SWIR2	2100.0	2280.0	20.0

```
[4]: uc.device_product_channels(product="rgb", device="MODIS", output="table")
```

channel_number	datatype	nodata_value	def_rule	calibration_rule	band1_name	band2_name	band3_name
0	1	BYTE	255 {"min":0,"max":255}	{"type":"none"}	7	None	None
1	2	BYTE	255 {"min":0,"max":255}	{"type":"none"}	2	None	None
2	3	BYTE	255 {"min":0,"max":255}	{"type":"none"}	1	None	None

```
[5]: uc.device_product_channels(product="ash", device="MODIS", output="table")
```

channel_number	datatype	nodata_value	def_rule	calibration_rule	band1_name	band2_name
0	1	BYTE	255 {"min":-3,"max":3,"palette":"ash"}	{"type":"bdt","mult":0.0235294,"add":-3}	31	32

Рисунок 3.2.2.1 - Пример получения справочной информации о данных в архивах

На рисунке 3.2.2.2 приведены примеры запросов на получение информации из архива данных высокого разрешения unisat\_hrsat. Первый вызов products\_stat предоставляет статистику по имеющимся в архиве «физическим» продуктам для указанного прибора и диапазона дат. Второй вызов get\_metadata является основным запросом на получение информации об экземплярах, имеющихся в архивах продуктов, соответствующих заданному набору критериев поиска. Результат возвращается в формате JSON и содержит как общую информацию о запросе (секция INFO), так и информацию о найденных экземплярах продуктов (секция DATA).

```
[132]: from unisat_archive import Archive
ua = Archive("hrsat")

[133]: ua.products_stat(device='MSI', from_date='2023-01-01', to_date='2023-11-01', limit=5)

[133]:
```

	product	satellite	station	product_volume	min_date	max_date	count	size_gb
0	channel1	SENTINEL-2A	ESA	ONLINE	2015-07-04	2023-11-02	293329	5077
1	channel2	SENTINEL-2A	ESA	ONLINE	2015-07-04	2023-11-02	293354	177934
2	channel3	SENTINEL-2A	ESA	ONLINE	2015-07-04	2023-11-02	293351	179393
3	channel5	SENTINEL-2A	ESA	ONLINE	2015-07-04	2023-11-02	293323	47517
4	channel7	SENTINEL-2A	ESA	ONLINE	2015-07-04	2023-11-02	293321	48832

```
[134]: ua.get_metadata({
    'bbox': '45.3,45.3,45.5,45.5',
    'dt_from': '2023-07-01',
    'dt': '2023-08-01',
    'devices': 'MSI',
    'satellites': 'SENTINEL-2A',
    'limit': '1',
    'products': 'channel14,channel18',
    'get_online_base_files': '1',
    'volumes': 'ONLINE'
})

{'DATA': [{'common': {'daynight': 'unknown',
                      'device': 'MSI',
                      'dt': '2023-07-30 07:57:18',
                      'satellite': 'SENTINEL-2A',
                      'station': 'ESA'},
            'products': {'channel14': {'files': ['03737/99718.tif'],
                                          'id': '2307300757185235490110',
                                          'server': 'smisdev'},
                        'channel18': {'files': ['03737/99730.tif'],
                                          'id': '2307300757185235490820',
                                          'server': 'smisdev'}}}],
 'INFO': {'db_pkg_mode': 'hrsat',
          'project': 'smis',
          'query': {'count': 1,
                   'first': 1,
                   'last': 1,
                   'next': 2,
                   'predicted_count': 0,
                   'previous': None,
                   'total_count': 6},
          'servers': {'smisdev': {'accessibility': 1,
                                  'center': 'unknown',
                                  'is_local': 1,
                                  'url': 'dev.nffc.aviales.ru'}},
          'stations': {'ESA': {'tz': '0'}},
          'user': None}}
```

Рисунок 3.2.2.2 - Пример запросов к базе данных unisat\_hrsat.

На рисунке 3.2.2.3 показаны примеры операций над файлами данных, соответствующими выбранным на предыдущем слайде экземплярам спутниковых наблюдений. Существенно, что данные примеры являются чисто иллюстративными и предназначены только для демонстрации самых базовых возможностей среды Jupyter. Для получения полных путей к файлам используется получаемый из библиотеки unisat\_archive базовый путь к хранилищу файлов ua.store\_dir и относительные пути, полученные на предыдущем слайде.

В первом примере сначала при помощи библиотеки gdal из файлов вычитываются массивы значений для красного и инфракрасных каналов фрагмента изображения по данным прибора MSI спутника Sentinel-2A. Затем одной строкой на основе полученных массивов вычисляется вегетационный индекс NDVI. После чего при помощи библиотеки matplotlib в блокнот выдается полученное изображение. Во втором примере для полученного индекса NDVI также при помощи библиотеки matplotlib отображается гистограмма по частоте различных его значений.



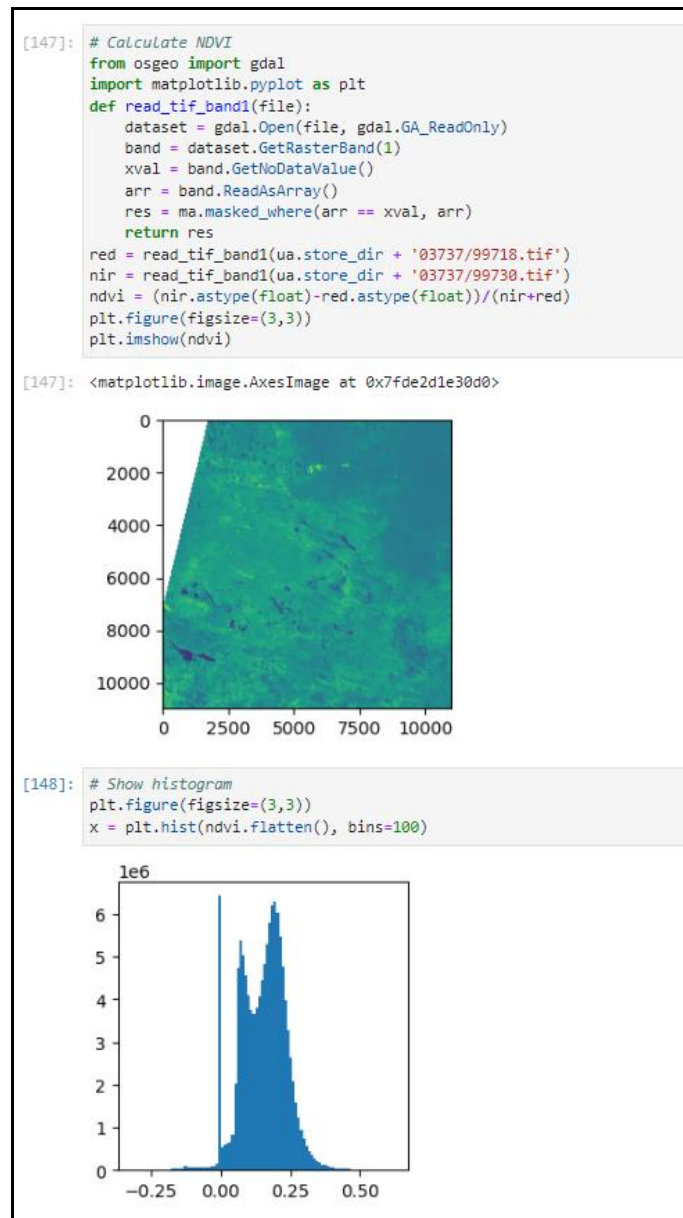


Рисунок 3.2.2.3 - Примеры тестовых операций с данными архива

В настоящее время в отделе «Технологии спутникового мониторинга» реализован только базовый функционал, обеспечивающий доступ к данным архивов ЦКП «ИКИ-Мониторинг» и позволяющий работать с ними по месту их физического хранения при помощи широкого спектра сторонних библиотек. При этом доступ к этому инструменту пока что предоставляется только узкому кругу специалистов и разработчиков.

В перспективе планируется расширить имеющийся функционал благодаря поддержке инструментов для анализа и обработки данных, разработанных в ИКИ РАН. Существенным шагом в развитии этого инструмента может быть также поддержка механизма «виртуальных» информационных продуктов, которые получаются с использованием заданных правил построения на основе базовых информационных продуктов и являются одним из ключевых преимуществ систем доступа к архивам ЦКП «ИКИ-Мониторинг». Также планируется внедрить специальные программные решения, которые позволили бы предоставить доступ к этому инструменту для более широкого круга пользователей, в том числе на основе кластерных решений.

### ***3.2.3 Развитие нейросетевых подходов для автоматического построения масок облачности и теней на спутниковых изображениях приборов серии КМСС***

В рамках проведенных в 2023 году работ получили развитие методы автоматического построения масок мешающих факторов на некалиброванных изображениях, полученных по данным приборов серии КМСС. В 2022 году было показано, что свёрточная нейросеть позволяет выявлять мешающие факторы на спутниковых изображениях различного уровня обработки за счёт использования текстурной информации. Дальнейшее развитие методов автоматического распознавания облачности и теней с учетом необходимости учёта различных условий наблюдения и освещения, а также при различных типах подстилающей поверхности требует использования пространственно-распределенной выборки, охватывающей большую неоднородную территорию. Кроме этого, для использования нейросети в потоковой обработке необходимо было провести анализ её эффективности при различных параметрах архитектуры и различных видах нормализации данных.

Для этих экспериментов, во-первых, был использован набор из около 5000 спутниковых изображений спутников Sentinel-2 (прибор MSI), полученных в 2021 году на территорию России, содержащий стандартные маски (scene classification map) для каждой сцены. Сцены были переведены в географическую проекцию (рабочую проекцию данных КМСС) и загружены до различных разрешений, включая угловой аналог 60 метров (разрешение КМСС) и 480 метров (восьмикратное разрешение КМСС) на экваторе. Во-вторых, были использованы наборы данных КМСС, содержащие штатные маски облачности и теней, полученные без использования методов машинного обучения. Сцены были разбиты на блоки размером 512 на 512 пикселей, 70% которых использовалась для обучения, остальная часть – для валидации. Второй набор данных КМСС был расширенным и получен, в том числе, с использованием лучших результатов предыдущих версий нейросети с экспертным отбором результатов.

В работе использовались два варианта архитектуры сети U-Net – вариант с 650 тыс. весов, четырьмя этапами понижения размера изображения и использованием оператора свёртки Conv2D (вариант А), а также вариант с 300 тыс. весов, двумя этапами понижения размера изображения и использованием оператора свёртки SeparableConv2D (вариант Б). Во всех экспериментах, кроме одного, в качестве функции потерь использовалась функция bce jaccard loss, в оставшемся эксперименте использовалась функция sparse categorical cross entropy. Кроме этого, в двух экспериментах варианта архитектуры А для повышения устойчивости к переобучению дополнительно использовался слой dropout.

Во время предварительной нормализации данных значения переводились в диапазон от 0 до 1 (код R) и опциональное умножение на случайный коэффициент (код M). Внутри сети опционально использовалась батч-нормализация. Таким образом, была проведена серия экспериментов, некоторые из которых приведены в Таблице 3.2.3.1 с описанием ключевых условий эксперимента. Пример динамики метрики общей точности на обучении приведен на рисунке 3.2.3.1.

Таблица 3.2.3.1 – Сводная таблица экспериментов для оценки эффективности различных схем обучения нейросетевых моделей для выявления облачности и теней на изображениях КМСС

Номер	Arch	Предварительная нормализация данных	Batch norm	Слой Dropout	Set train	Set val	Разрешение	Epochs	Fin OA train, %	Best OA val, %
1	A	R+M	+	+	Набор 1	Набор 1	480 м.	37	83.95	70.38
2	A	R+M	+	–	Набор 1	Набор 1	480 м.	44	88.21	87.26
3	A	R+M	–	+	Набор 1	Набор 1	480 м.	48	86.38	85.73
4	A	R+M	–	–	Набор 1	Набор 1	480 м.	61	88.29	87.14
5	A	R	–	–	Набор 2	Набор 2	60 м.	53	74.9	92.17
6	A	R	–	–	Набор 2	Набор 2	60 м.	667	95	95.07
7	A	R	–	–	Набор 3	Набор 2	60 м.	363	93.69	94.12
8	A	R+M	–	–	Набор 2+ Набор 3	Набор 2+ Набор 3	60 м.	249	96.14	86.28
9	A*	R+M	–	–	Набор 2+ Набор 3	Набор 2+ Набор 3	60 м.	256	95.88	86.5
10	Б	R+M	+	–	Набор 1	Набор 1	480 м.	500	91.32	77.66
11	Б	R+M	–	–	Набор 1	Набор 1	480 м.	23	76.47	76.82
12	A	R+M	+	–	Набор 1	Набор 1	480 м.	177	82.89	77.33
13	A	R+M	–	–	Набор 1	Набор 1	480 м.	1563	95.52	81.17
14	A	R+M	–	–	Набор 1	Набор 1	60 м.	339	95.73	78.1

**Примечание:** A\* – архитектура типа A, где в качестве функции потерь использовалась sparse categorical cross entropy; Arch – архитектура; R – решкалирование; M – умножение на случайное число; Batch norm – использование батч-нормализации; Set train – набор для обучения; Set val – набор для валидации; Набор 1 – набор подготовленных сцен Sentinel-2 (5000 сцен для обучения, 770 сцен для валидации); Набор 2 – набор вручную выбранных 60 сцен КМСС, которые нарезались блоками 512x512, 70% использовались для обучения, 30% – для валидации; Набор 3 – набор вручную выбранных 150 сцен КМСС, которые нарезались блоками 512x512, 70% использовались для обучения, 30% – для валидации; Epochs – число эпох к окончанию обучения, Fin OA train – финальная точность на обучении, Best OA val – лучшая точность на валидации.

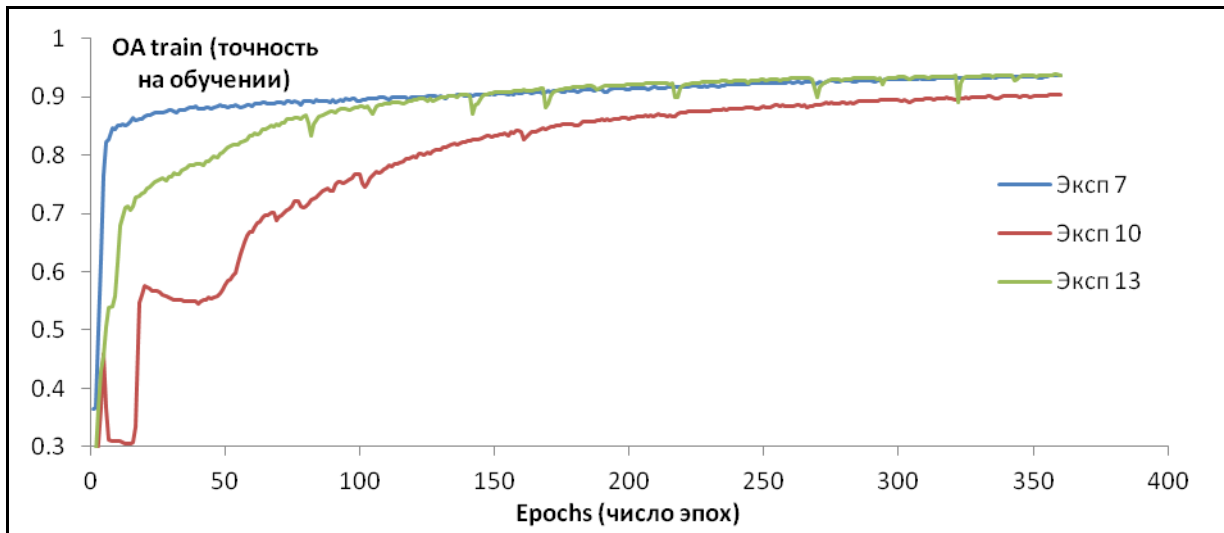


Рисунок 3.2.3.1 – Динамика точности на обучении для нескольких экспериментов из Таблицы 3.2.3.1

Динамика показателей общей точности на обучении и на валидации демонстрирует устойчивость сети и возможности её адаптации к особенностям входных данных. В частности, в процессе обработки порций загружаемых и выгружаемых порций данных может наблюдаться резкое кратковременное падение метрик точности, как, например, для эксперимента 13. Для учёта некалиброванности входных изображений КМСС, на которых требуется выявлять участки облачности и теней, в дизайн некоторых экспериментов была добавлена предварительная нормализация данных, а также батч-нормализация внутри сети.

Для оценки эффективности сети и условий схемы обучения оценивалась метрика лучшей точности на валидации, которая оказалась максимальной для эксперимента 6 и составила 95.07% для архитектуры типа А с рескалированием входных данных по Набору 2. Общая схема архитектуры нейросети из эксперимента 6 показана на рис. 3.2.3.2.

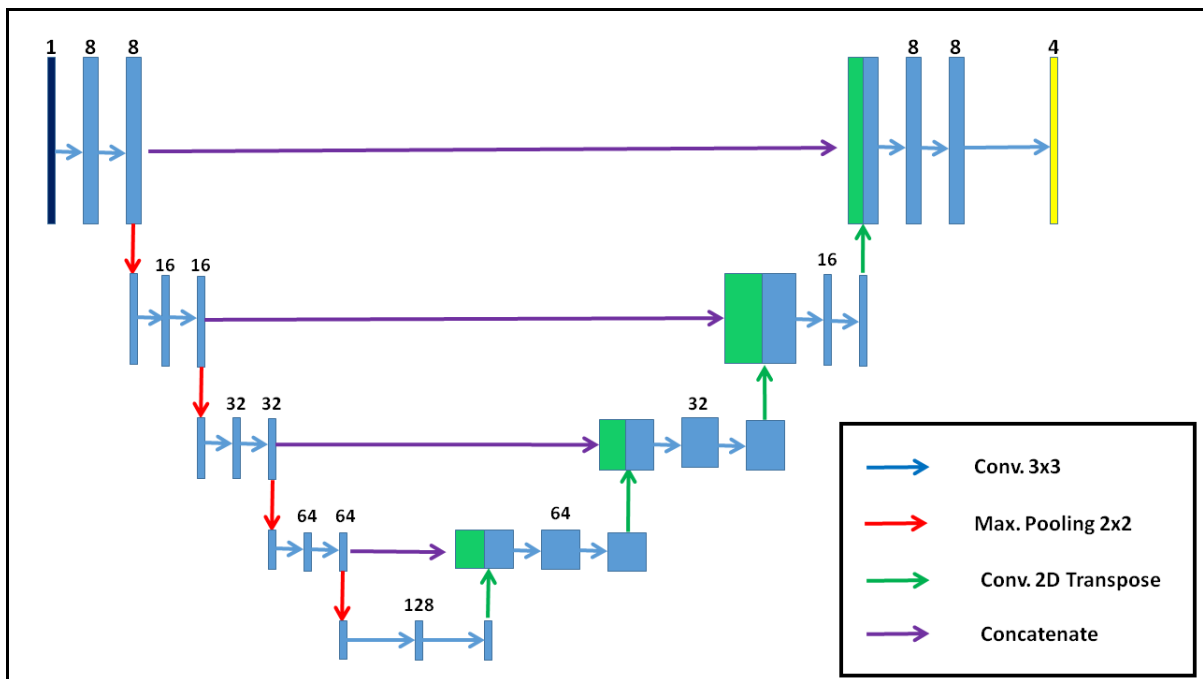


Рисунок 3.2.3.2 – Общая схема наиболее оптимальной архитектуры нейросети U-Net типа А для распознавания облачности и теней на изображениях КМСС

Следует отметить, что более простая версия этой же архитектуры Б демонстрирует заметно более низкие показатели максимальной точности на валидации, что указывает на то, что упрощение нейросети с целью её ускорения в данном случае не является оправданным.

### 3.2.4 Разработка архитектуры и инструментов для модернизации распределенной многопоточковой обработки спутниковых данных

Стремительное развитие систем спутникового мониторинга Земли в последние десятилетия привело к резкому увеличению объемов спутниковых данных и числа их типов. Это закономерно еще более повысило требования к эффективности организации систем обработки спутниковых данных. Необходимо максимально полно использовать имеющиеся вычислительные ресурсы и контролировать сотни различных процессов по обработке спутниковых данных. Для решения вышеперечисленных задач в рамках центра коллективного пользования ЦКП «ИКИ-Мониторинг» была разработана технология организации многопоточковой обработки спутниковых данных на кластере windows серверов. В настоящее время она успешно используется для проведения широкого спектра различных типов обработки спутниковых данных на кластере, включающем в себя более 200 серверов, включая виртуальные. Впервые эта технология была внедрена в 2016 году, при этом в нее регулярно вносились изменения, которые повышали ее эффективность. В 2023 году велись работы по совершенствованию схемы управления выполнением большого количества процедур обработки с разным приоритетом и продолжительностью выполнения. Как уже было отмечено выше, необходимо максимально полно использовать имеющиеся вычислительные ресурсы, однако это вступает в противоречие с требованиями на время выполнения «срочных» заданий на обработку. Одним из решений является указание максимального количества потоков выполнения для длительных, но не очень оперативных процедур обработки. Также в 2023 году были проведены работы по оптимизации запросов к базе данных, отвечающей за управление обработкой и контроль ее выполнения, что позволило минимизировать задержки при их проведении. В то же время на текущий момент реализация системы многопоточковой обработки перестала в полной мере отвечать предъявляемым к ней требованиям. Поэтому в 2023 году были начаты работы по более масштабной ее модернизации.

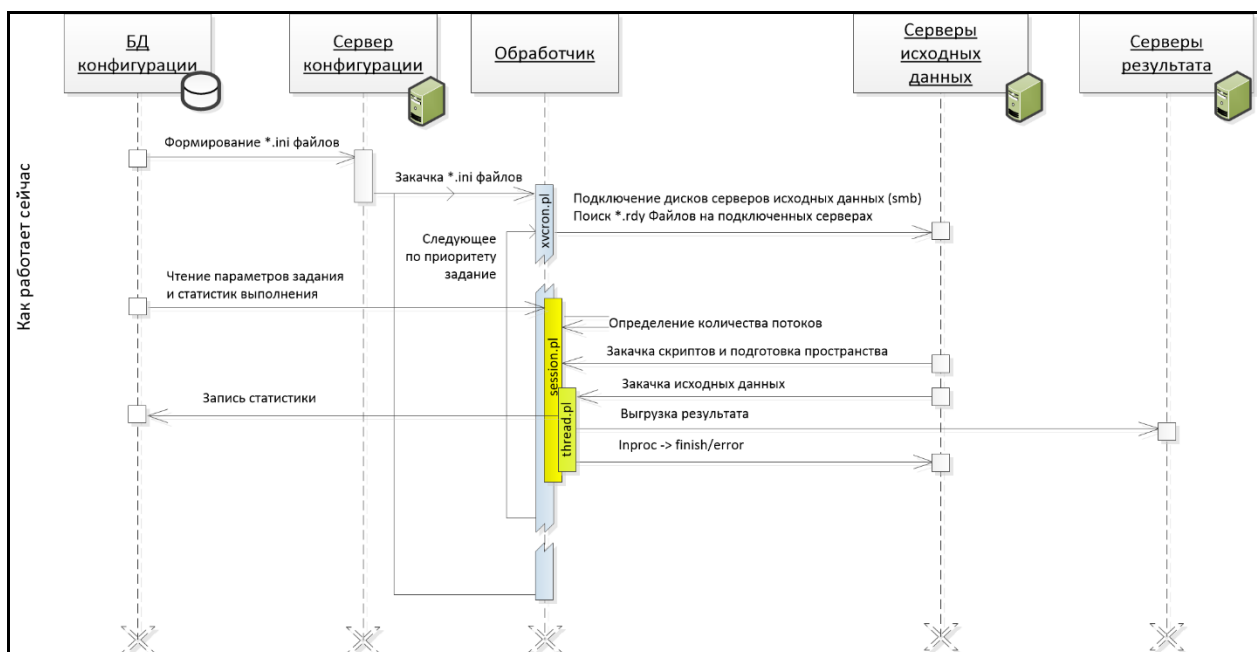


Рисунок 3.2.4.1 - Текущая схема организация обработки спутниковых данных

Текущая схема организации обработки спутниковых данных приведена на рисунке 3.2.4.1. Для хранения информации о типах обработки, в том числе и соответствующих им требований к вычислительным узлам реализована база данных конфигураций. В первом слое скриптов (`хvstop.pl`) вычислительный узел обращается к базе данных, скачивает себе комплект `.ini` файлов с метаданной, и в соответствии с прописанными внутри них путями размещения исходных данных, начинает опрашивать различные сервера на предмет наличия комплектов готовых к обработке. Если такой комплект находится, то запускается второй слой скриптов, соответствующий сессии выполнения (`session.pl`), функционал которого изначально сводился к определению количества комплектов, которые возможно обработать на вычислительном узле параллельно, подготовки скриптов для обработки этих комплектов и запуске скриптов третьего слоя потоков выполнения (`thread.pl`). В рамках последнего слоя уже непосредственно осуществляется обработка отобранных комплектов исходных данных. Но со временем во втором слое накопилось достаточно много функциональных особенностей, которые с точки зрения логики обработки, необходимо было включить в первый слой, но которые в силу различных обстоятельств оказались во втором. В некоторых случаях это приводит к ситуациям, когда факт наличия комплекта данных и вычислительного узла для его обработки, не гарантирует выполнение этой обработки. Это может происходить, когда на сервере для результатов закончилось свободное место, на вычислительном узле накопилось критическое количество ошибочно выполненных заданий этого вида или же вычислительный узел перестал отвечать характеристикам (по оперативной памяти, свободном месте на жестком диске и т.п.). Т.к. проверка всех этих характеристик осуществлялась из второго слоя, то на вычислительных узлах начали в массовом количестве появляться так называемые «пустые сессии», инициированные первым слоем по факту наличия комплекта на обработку, и завершённые вторым слоем из-за невозможности эту обработку осуществить.

Кроме того, в нынешней парадигме предполагается, что подавляющее большинство выполняемых в системе заданий на обработку данных, будет выполняться относительно быстро (минуты), а создаваемый трафик вычислительных узлов не будет упираться в пороговые значения пропускной способности коммутаторов и серверов, которые предоставляют исходные данные и принимают результат. В ситуациях, когда длительные задания (длящиеся часы, а в некоторых случаях десятки часов) начинают составлять значительную часть от всего массива обрабатываемых данных, а совместный трафик вычислительных узлов настолько велик, что может приводить к сбоям обмена данными с серверами, начинают возникать проблемы, парализующие работу всего пула обработчиков. Они связаны с тем, что, подхватив данные и проведя многочасовую обработку, вычислительный узел начинает пытаться выгрузить результат. Из-за проблем с пропускной способностью (т.к. большое количество вычислительных узлов пытается сделать то же самое, на тот же самый сервер), по таймауту задание завершается с ошибкой. После ошибки копирования, система выставляет комплект на повторную обработку. В случае с длительными обработками мы получаем ситуацию, когда большое число вычислительных узлов, мешая друг другу, пытается загрузить на один и тот же сервер результат своей обработки, уходят в итоге в ошибку копирования, после чего заново начинают обработку тех же самых комплектов.

Для устранения описанных выше проблемных ситуаций разработана модифицированная схема организации обработки, представленная на рисунке 3.2.4.2. В рамках нее копирование результатов обработки переведено в отдельный процесс. Функционал первых двух слоев будет выполнять один новый слой, который будет отвечать как за поиск комплектов на обработку, так и проверять, способен ли вычислительный узел их обработать.

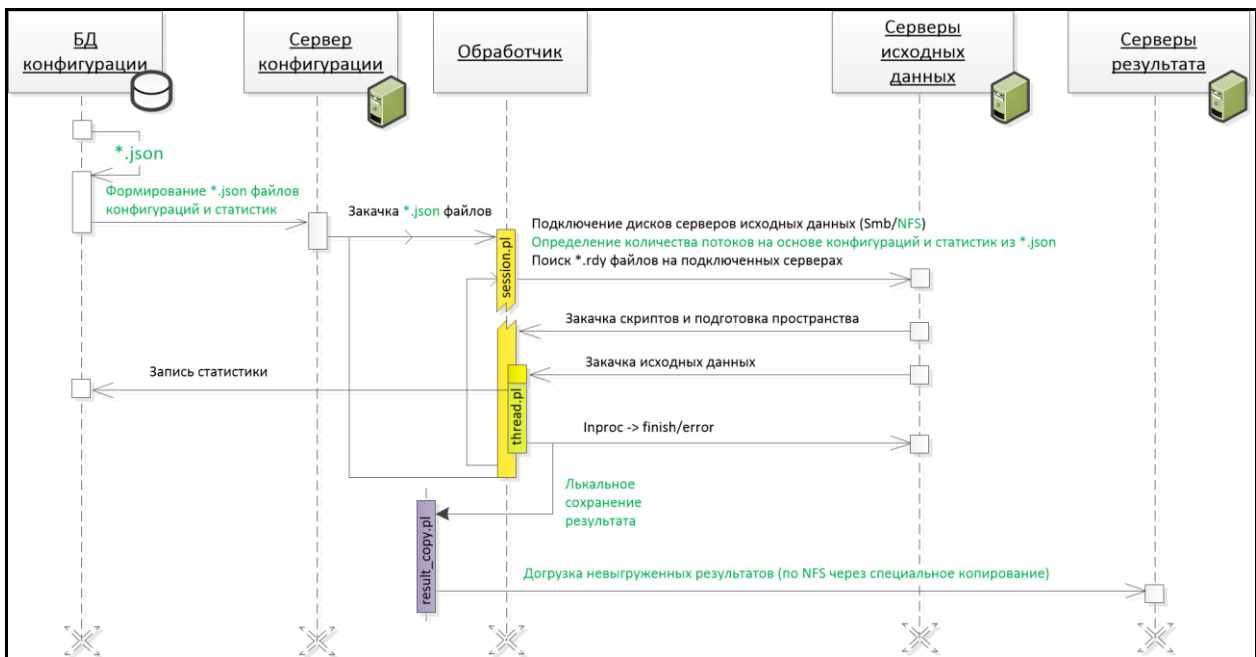


Рисунок 3.2.4.2 - Новая схема организация обработки спутниковых данных

При этом метainформация о различных видах обработки будет передаваться в виде json-файлов, что позволяет задавать иерархические структуры настроек в отличие от используемых в настоящее время стандартных ini-файлов. При этом процесс копирования результата будет активен только в моменты простоя сетевого интерфейса (чтобы не создавать на него излишнюю нагрузку при копировании исходных данных и выгрузке результата на сервер, см. рисунок 3.2.4.3).

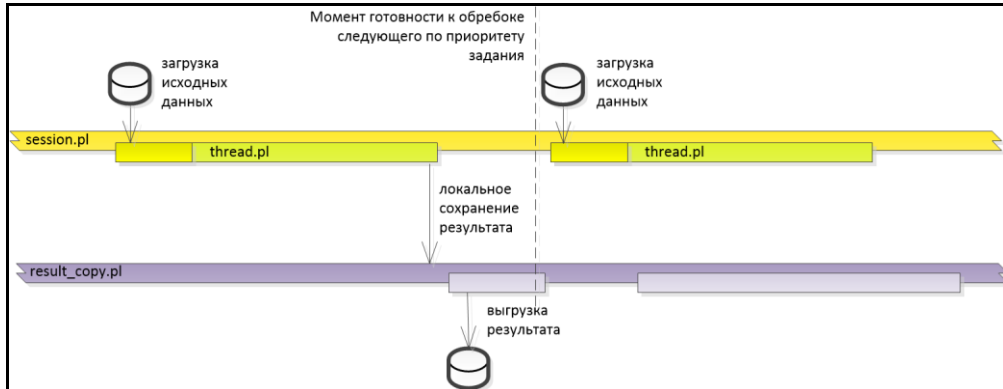


Рисунок 3.2.4.3 - Параллельная работа процессов обработки данных и выгрузки результата

В настоящее время ведутся работы по переводу конфигурации заданий по обработке на формат JSON и соответствующему изменению различных программных компонент. Также ведутся работы по разработке специального процесса для копирования результатов обработок.

### 3.2.5 Развитие системы ведения архивов и инструментов для контроля ее функционирования

В 2023 году существенное развитие претерпели процедуры, связанные с погружением спутниковых данных в архивы, реализованные на основе разработанной в ИКИ РАН технологии UNISAT. Основной причиной изменений была необходимость ускорения процессов усвоения данных в архивы, особенно в тех случаях, когда во входных директориях уже накопилось большое количество данных. Такие ситуации неоднократно случались в течение этого года и были связаны с одной стороны со сбоями в работе оборудования, а с

другой – с необходимостью быстрого восстановления больших объемов данных. В рамках имеющейся реализации распределенного файлового хранилища, построенной на базе протокола NFS, и используемых систем хранения данных (NAS) при накоплении во входных директориях десятков тысяч файлов существенно замедляются как операции получения списка файлов, так и сами операции по переносу их в архив. И хотя для ряда типов данных уже были реализованы процедуры для разнесения входных данных по разным поддиректориям, этого оказалось недостаточно. В рамках проведенных работ эта практика была распространена на большее число различных типов данных. Кроме этого утилиты архивации теперь не запрашивают полный список поступивших в директорию данных, а читают его в потоковом режиме. Это же относится и к процедуре запуска утилит архивации для различных типов входных данных (сейчас их около 100), которая ранее получала полный список файлов во входной директории для того, чтобы установить, какие из запусков необходимо выполнить.

Также в 2023 году базы данных, содержащие метаданные хранимых в архивах экземпляров спутниковых данных, были переведены с механизма хранения MyISAM на INNODB. Связано это было с тем, что хотя скорость добавления данных в монопольном режиме для первого из механизмов выше, в условиях большого количества пользовательских запросов к этим базам данных архивация замедляется из-за частого блокирования таблиц на запись. Второй из механизмов поддерживает блокировку на уровне записей в таблицах, что позволило существенно повысить скорость архивации данных. Также для основных по объему баз данных был развернут еще один вспомогательный сервер (slave), что позволило снять с него большую часть нагрузки по запросам данных.

Наряду с рядом других изменений вышеперечисленные меры позволили существенно ускорить архивацию данных и при необходимости заносить в архивы в сутки более 15Тб новых данных, содержащихся в сотнях тысяч файлов. Также был усовершенствован комплект утилит для быстрой диагностики как скорости передачи данных по сети на указанный том архива, так и накладных расходов на элементарные файловые операции, который активно использовался при анализе проблем с локальной сетью и системами хранения данных.

В рамках системы контроля функционирования программно-аппаратного комплекса ЦКП «ИКИ-Мониторинг» наряду с большим количеством различных инструментов реализовано формирование ежедневных сводных отчетов о работе различных подсистем. Такие отчеты рассылаются ответственным специалистам по электронной почте и позволяют руководителям быстро оценивать ситуацию с тем или иным направлением работ. Среди ежедневных отчетов наиболее востребованным является сводный отчет о поступлении в архивы различных типов данных из разных источников. При этом в отчете содержится не только информация за последние сутки, но и динамика поступления и архивации данных за последние несколько месяцев. Это позволяет автоматически детектировать ситуации с замедлением поступления или архивации тех или иных типов данных. В 2023 году в дополнении к этим отчетам был реализован специализированный служебный веб-интерфейс, позволяющий специалистам оперативно заносить информацию, описывающую причины неполадок, а руководителю – контролировать своевременное ее предоставление. Также в интерфейсе реализованы различные режимы для получения статистики по сбоям в поступлении данных, что позволяет анализировать как наиболее частые неполадки, так и их причины.

В 2023 году также была реализована детальная статистика по работе процессов занесения в архивы различных типов данных, которая рассылается специалистам как ежедневный отчет. Для каждого из заносимых в архивы типов данных приводится следующая информация: количество поступивших на вход данных, количество новых занесенных в архив данных, количество замененных в архиве экземпляров данных, время, затраченное на архивацию данных, текущая очередь во входной директории, а также статистика ошибок за сутки. При этом в отчете содержится как сводная информация за последние сутки, так и история для каждого из процессов архивации данных за последние 10 суток. Используя эту



информацию, можно легко оценить не только загруженность системы архивации, но и конкретные процессы, на которые тратится основное время. На основе ее, в частности, могут быть изменены приоритеты занесения в архивы тех или иных данных.

### **3.3 Развитие с использованием созданных методов и подходов различных специализированных научных информационных систем, создаваемых и поддерживаемых в ИКИ РАН**

#### ***3.3.1 Расширение функционала подсистемы анализа временных серий данных в информационных системах семейства «Созвездие-ВЕГА»***

В настоящее время для исследований Земли из космоса и мониторинга различных процессов и явлений доступны глобальные и протяженные во времени архивы спутниковых данных, такие как, например, архивы ЦКП «ИКИ-Мониторинг». Существующие инструменты позволяют «на лету» получать различные информационные продукты на основе данных архивов, а также рассчитывать историю необходимых для мониторинга характеристик для отдельных объектов наблюдений в оффлайн-режиме. Кроме них, однако, востребованными являются сервисы, позволяющие «на лету» получать данные по большому количеству разновременных наблюдений в произвольных точках Земной поверхности при решении целого ряда отдельных задач спутникового мониторинга. Опыт эксплуатации таких сервисов спутникового мониторинга показал востребованность инструмента анализа данных, представляющих протяженные временные серии в виде графиков.

До недавнего времени в рамках информационных систем семейства «Созвездие-Вега» был реализован ограниченный функционал по анализу временных серий данных на основе их представления в виде графиков. Пользователь мог получить график по точке или объекту наблюдения по ограниченному набору информационных продуктов, большую часть из которых составляли различные вегетационные индексы либо метеоданные.

В 2023 году функционал подсистемы анализа временных рядов был практически полностью переработан. Была поставлена задача обеспечения доступа пользователей к инструментам анализа временных рядов, полученных по всем имеющимся в архивах ЦКП «ИКИ-Мониторинг». Так как разные типы данных существенно отличаются по схемам хранения и построения, для реализации новых возможностей потребовалась существенная модернизация и унификация уже существующих программных модулей и разработка новых библиотек и программных компонент. Принципиальная схема взаимодействия сервисов задействованных при построении графиков временных рядов приведена на рисунке 3.3.1.1.



Рисунок 3.3.1.1 - Схема взаимодействия сервисов, задействованных при построении графиков временных рядов

На текущий момент новый функционал уже внедрен для работы с данными геостационарных спутников Himawari-8/9, «Электро-Л» №№1-3, GOES-E,W, METEOSAT 9-11 и высокоэллиптического КА «Арктика-М» №1. Пример использования нового интерфейса приведен на рисунке 3.3.1.2. Работы по внедрению аналогичного функционала для многих десятков различных композитных продуктов находятся на завершающей стадии, а именно на стадии тестирования. Именно композитные продукты, позволяющие получать регулярные временные ряды по точкам или полигонам, находящимся в зоне их покрытия (например, «вся Россия»), представляют наибольший интерес для пользователей. Реализация доступа к ним потребовала разработки нового, более сложного веб-интерфейса для интерактивного выбора требуемого композитного продукта по набору различных критериев (тип композита, его скважность, тип продукта и вариант его построения и совместного отображения соответствующих им временных рядов данных). Следующим этапом будет реализация в информационных системах семейства «Созвездие-Вега» инструментов для анализа рядов данных, получаемым по сценам среднего и высокого разрешения.

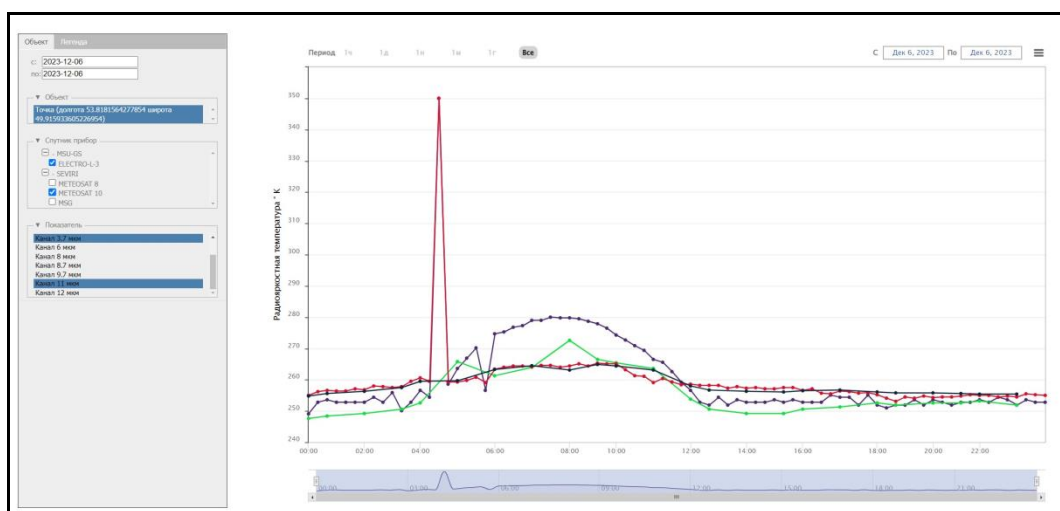


Рисунок 3.3.1.2 - Пример интерфейса для анализа временных рядов по данным геостационарных спутников

Для реализации вышеописанного функционала были существенно доработаны программные модули, отвечающие за предоставление в интерфейсы информации об имеющихся в архивах UNISAT данных и различных их характеристиках, таких как: наборы каналов, правила построения динамически формируемых («виртуальных продуктов»), пространственные характеристики и др.

Также в 2023 году были проведены работы по оптимизации блоков системы управления комплексными задачами по обработке данных TASK\_JOBS, которая в данном случае используется для диспетчеризации процессов подготовки временных рядов. Был существенно улучшен служебный интерфейс, который теперь позволяет проще найти информацию о выполнении конкретных процессов чтения рядов данных. С целью повышения скорости передачи по сети данных временных рядов теперь эта информация передается в сжатом виде. Дополнительно реализован функционал, позволяющий останавливать выполнение процессов обработки путем перевода соответствующего задания в системе TASK\_JOBS в статус «отменен». Такая необходимость иногда возникает при получении очень больших временных рядов данных, которое по тем или иным причинам оказывается ненужным.

Существенные изменения с целью повышения скорости работы были внесены непосредственно в «Блок получения временных рядов». Подробнее этот вопрос будет рассмотрен в следующем пункте отчета.

### ***3.3.2 Развитие инструментария для системы доступа к временным рядам данных в системе мониторинга вулканической активности Камчатки и Курил (ИС VSV)***

В рамках информационной системы VolSatView ведется анализ временных рядов данных спутниковых наблюдений. Разработана технология накопления данных дистанционного измерения радиояркостной температуры в зонах вулканической активности, а также пользовательский интерфейс интерактивного анализа рядов наблюдений. Крайне актуальными для такого рода исследований являются данные, полученные с геостационарных спутников, ввиду высокой частоты наблюдений, которая, например, в случае прибора АНІ составляет 10 минут.

В 2023 году проводились работы по развитию блока получения временных рядов с целью повышения скорости и надежности его работы. И хотя этот блок используется для получения всех временных серий данных по самым разным типам спутниковой информации, наиболее критичным его быстродействие оказывается именно для анализа временных рядов, получаемых на основе данных прибора АНІ, установленного на спутниках Himawari-8/9. Это связано как с большой частотой поступления таких данных, так и с необходимостью анализа многолетних рядов измерений. Для чтения из спутниковых изображений значений в точке и в ее окрестностях используется свободно распространяемый программный пакет GDAL (Geospatial Data Abstraction Library). В 2023 году вместо использования входящей в его состав утилиты gdallocationinfo чтение данных было переведено на непосредственное использование API, которое предоставляется этим пакетом, т.е. на прямое получение значений из массива значений, соответствующего каналу изображения. Это, в частности, потребовало изменений в процедуре перевода данных из проекции карты в проекцию сцены.

В результате скорость получения информации для построения временного ряда извлекаемой из исходных файлов стала сравнима со скоростью извлечения информации из баз метеоданных. Например, извлечение временного ряда за год по данным спутника Himawari-8 (с шагом 10 мин) теперь происходит примерно за 3 минуты, тогда как ранее на это уходило около 10 минут. Теперь это сравнимо со временем ожидания данных по температуре в точке, извлекаемых из базы данных со скоростью ~15-20 сек/год. При этом в первом случае мы получаем ряд из 50000 точек, а во втором - из 1400 точек. Однако при большом числе пользовательских запросов на получение долговременных рядов данных загрузка серверов

подготовки оказалась неприемлемо высокой, в первую очередь, из-за чрезмерного использования оперативной памяти. Поэтому дополнительно была реализована специальная программа-демон, которая анализирует текущее количество выполняемых процессов подготовки данных и ограничивает количество одновременно выполняемых параллельных процессов.

Также в 2023 году проводились работы по оптимизации инструментов анализа временных серий данных с целью повышения удобства их использования. Отметим, что в рамках рассматриваемой информационной системы задачи, связанные с временными рядами, можно разделить на два больших класса:

1. временные ряды в произвольной географической точке (доступной для наблюдения тем или иным прибором);
2. временные ряды, связанные с точками и областями постоянного наблюдения (в данном случае - окрестностями вулканов).

В первом случае скорость получения информации имеет решающее значение, особенно для геостационарных спутников, поставляющих информацию через каждые 10-15мин. Проведенные работы позволили сократить время получения информации за два года (около 100 тысяч значений) до 6 минут. Годом ранее аналогичный запрос занимал 24 мин. Пример предоставления результатов по такому запросу приведен на рисунке 3.3.2.1

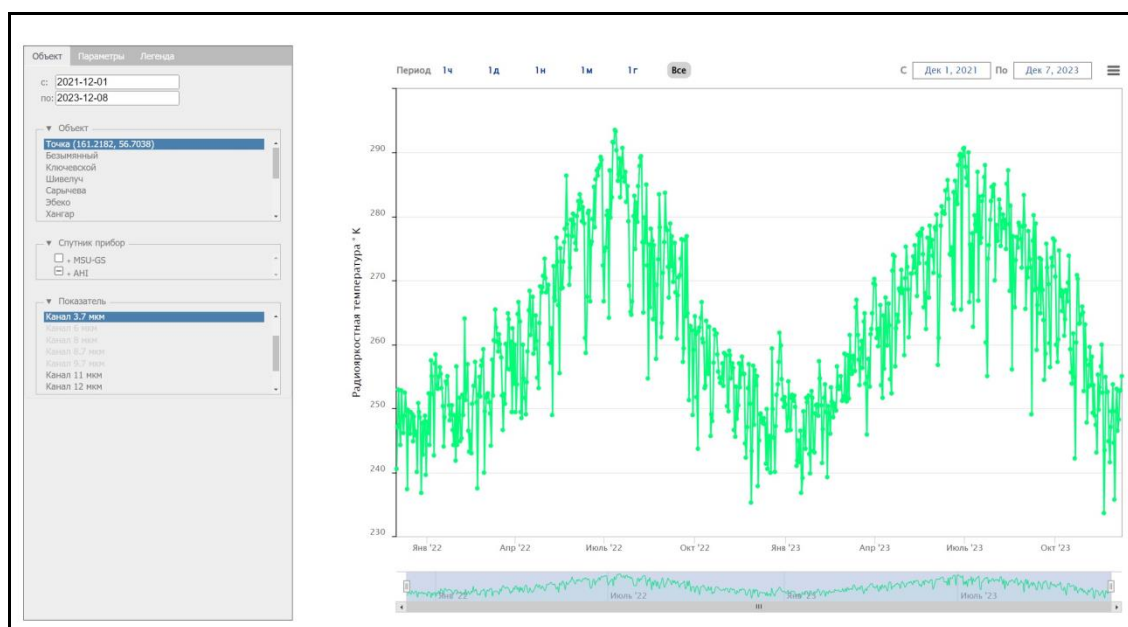


Рисунок 3.3.2.1 - Пример получения временного ряда по данным Himawari-8/9 за 2 года

Для второго варианта временных рядов повышение скорости загрузки позволило повысить информативность сервисов за счет накопления не только средних по областям, но и сбора статистических характеристик. Система обработки данных на участках наблюдения формирует выборку пикселей в окрестностях вулканов. Статистический анализ выборки позволяет фильтровать данные и анализировать производные величины, такие как максимум, минимум, среднее по окрестности или стандартное отклонение. В пользовательском интерфейсе реализованы возможности работы со всеми доступными для анализа величинами. Пример выполнения такого запроса приведен на рисунке 3.3.2.2. На нем представлен вывод информации о временном ряде, на котором можно проследить не только средние значения, но и минимальные и максимальные, а также стандартное отклонение значений радиояростной температуры в полигоне, определенном для постоянного мониторинга за вулканом Шивелуч.

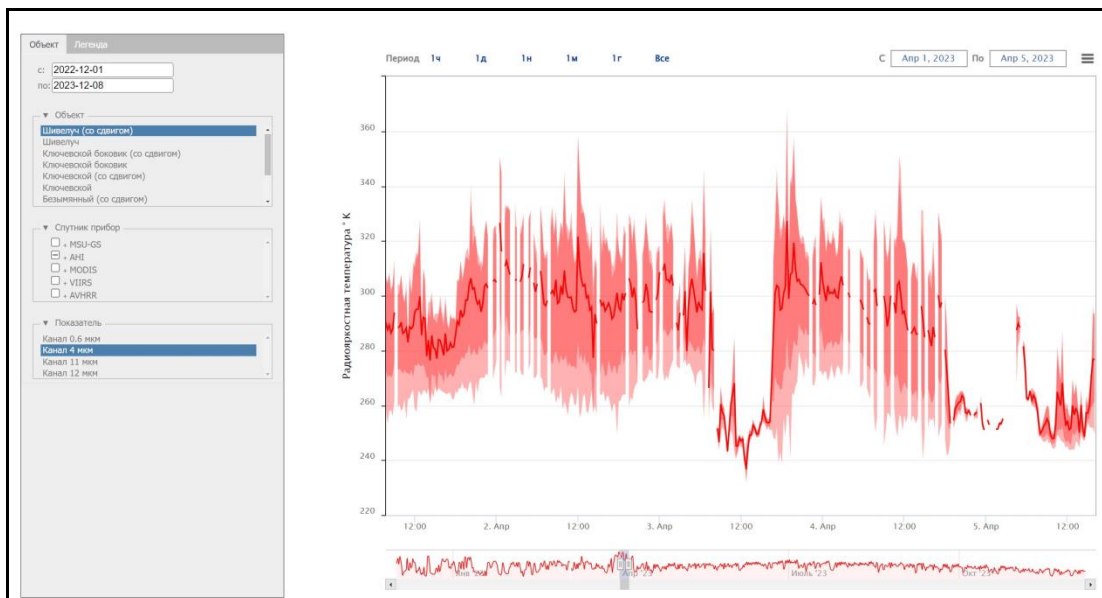


Рисунок 3.3.2.2 - Пример анализа временных рядов для вулкана Шивелуч

В качестве примера новых возможностей интерфейса можно привести реализованный алгоритм для подбора областей постоянного наблюдения за активностью вулканов. Ранее точки и полигоны для постоянного наблюдения выбирались субъективно, опираясь на визуальный анализ снимков. Теперь реализован алгоритм, который учитывает орбиту спутника, географическое положение точки наблюдения и ее геометрию (в данном случае высоту вулкана), что позволило исключить субъективный фактор. Для северных вулканов расстояние между центрами полигонов наблюдения для низкоорбитальных спутников и геостационарных спутников составляет километры. На рисунке 3.3.2.3 представлены контура наблюдения за вулканом Шивелуч. Красным цветом обозначен контур, выбираемый по субъективным представлениям, а зеленым – контур, полученный с помощью нового алгоритма. Для проведения этих работ потребовалось загрузить значительно больше информации за несколько лет наблюдений, чем грузилось ранее. Зеленый контур служит основой для выбора полигона наблюдения по снимкам низкоорбитальных спутников, а красный - для выбора полигона наблюдений по снимкам геостационарных спутников.

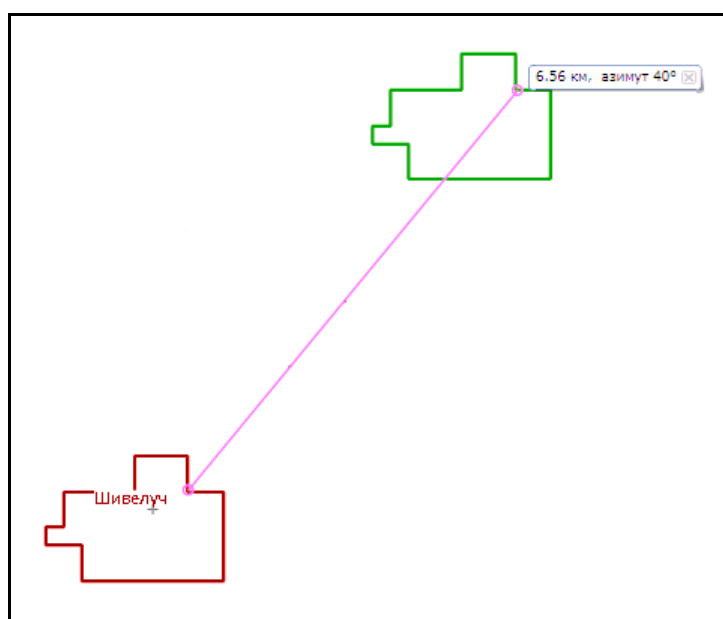


Рисунок 3.3.2.3 - Выбор полигона наблюдения с учетом орбиты спутника, географического положения и высоты вулкана

### 3.4 Заключение

Представленные в настоящей главе результаты показывают, что по направлению «Разработка новых подходов и методов организации сверхбольших архивов данных спутниковых наблюдений Земли и доступа к ним для обеспечения эффективной обработки и анализа данных ДЗЗ для решения научных и прикладных задач» проведенные работы соответствуют установленной в плане работ ИКИ РАН на 2023 год тематике.

Таким образом, работы выполнены полностью в соответствии с планом.

При этом получены следующие основные результаты:

- Проведен анализ новых подходов для оптимизации хранения долговременных рядов восстановленных рядов данных в архивах ЦКП «ИКИ-Мониторинг»;
- Разработан базовый вариант инструментария для работы с данными ЦКИ «ИКИ-Мониторинг» на базе интерактивной среды Jupyter;
- Получены новые результаты по получению масок облачности и теней на основе использования алгоритмов машинного обучения Deep Learning;
- Проведены работы по совершенствованию системы распределенной многопоточной обработки спутниковых данных и выработан план масштабной ее модернизации;
- Повышена скорость усвоения данных в архивы ЦКП «ИКИ-Мониторинг», а также созданы новые инструменты для контроля за архивацией разных типов спутниковой информации;
- Реализован новый функционал для анализа долговременных рядов данных в информационных системах «Созвездие-ВЕГА»;
- Проведены работы по оптимизации процедур подготовки данных временных рядов, что особенно актуально для ИС мониторинга вулканической активности Камчатки и Курил.

Таким образом, можно считать, что работы по данному направлению темы в 2023 году полностью выполнены.

## ГЛАВА 4 Наиболее значимые результаты исследований ИКИ РАН

### 4.1 Вариации условий природной среды во время геомагнитной бури, приведшие к потере серии спутников Starlink

В феврале 2022 г. при запуске с Мыса Канаверал очередной серии спутников системы Starlink произошла техногенная катастрофа. Этот запуск совпал с развитием двухфазной геомагнитной бури. Пики бури пришлись на 3 февраля 11 UT и 4 февраля 11 UT. После старта 3 февраля 38 из 49 запущенных КА не достигли запланированной высоты и вследствие усиленного торможения сошли с орбиты и вернулись обратно в атмосферу 8 февраля. Был проведен сравнительный анализ изменения параметров термосферы и ионосферы для пяти последовательных запусков Starlink в условиях космической погоды, варьирующихся от спокойных до буревых возмущений. Для оценки степени волновой активности, турбулентности и неустойчивости в ионосфере во время запусков Starlink, были рассчитаны средние значения ПЭС — интегральной плотности плазмы в столбе площадью  $1 \text{ м}^2$  в ионосфере и плазмосфере на высоте от 65 км до 20200 км — вдоль терминатора на его утренней и вечерней половинах на высоте 300 км.

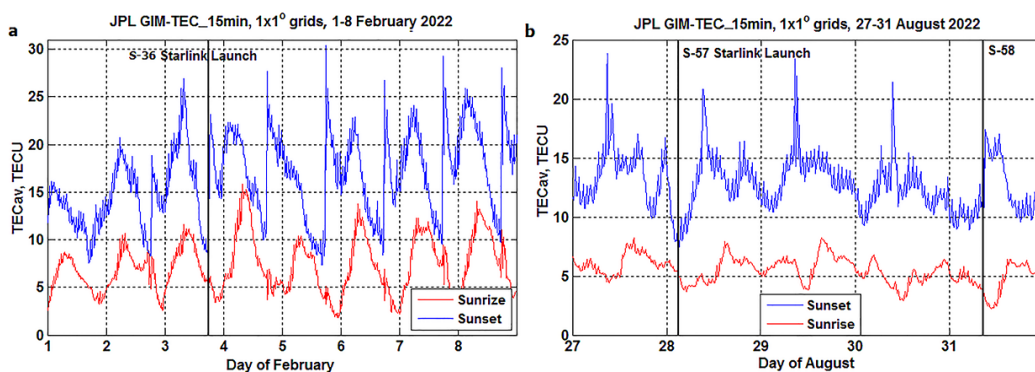


Рисунок 1.5.4 - Изменение ПЭС вдоль терминатора на восходе и заходе солнца на высоте 300 км; (а) запуск S-36, 2-8 февраля 2022 г.; (б) запуски S-57 и S-58, 27-31 августа 2022 г. Время запусков S-36, S-57 и S-58 обозначено вертикальными линиями

По результатам исследования можно сделать следующие выводы: 1) сравнение данных пяти запусков Starlink с геомагнитными индексами Dst показывает, что по мере развития магнитной бури 3–5 февраля 2022 г. нейтральная плотность термосферы увеличивается примерно на 50%, что приводит к увеличению сил сопротивления, в результате чего были потеряны 38 из 49 запущенных спутников; 2) во всех случаях наблюдались колебания ПЭС на солнечном терминаторе с асимметрией утренне-вечерних значений с преобладанием вечерних величин. Амплитуда колебаний была наибольшей во время бури 3–5 февраля 2022 г.; 3) изменение среднего ПЭС на закате и восходе солнца имело волновой характер с периодами колебаний 24 часа, 12 часов и меньше. То есть именно условия космической погоды оказали решающее влияние на то, что запуск S-36 закончился потерей спутников.

Gulyaeva T., Lukianova R., Haralambous H. (2023). Ionosphere heterogeneities at dawn–dusk terminator related to the Starlink satellites launch disaster on 3–8 February 2022 // *Journal of Geophysical Research: Space Physics*, 128, e2023JA031577. <https://doi.org/10.1029/2023JA031577>. (РИНЦ, Scopus, WoS, Q2)

Направления ПФНИ: 1.3.4.3. Пламенные процессы в геофизике и астрофизике; 1.3.7.4. Солнце и околоземное космическое пространство, солнечно-земные связи

## ЗАКЛЮЧЕНИЕ

Таким образом, представленный отчет по теме «Большие данные в космических исследованиях: астрофизика, солнечная система, геосфера» (№ 122042500019-6) показывает, что в соответствии с планом ИКИ РАН на 2023 год они велись по следующим основным направлениям:

- Развитие методов моделирования высокоширотной ионосферы и анализ массивов спутниковых и наземных данных измерений; построение численных и гибридных моделей распределения электронной концентрации в полярной ионосфере в зависимости от изменяющихся гелио-геофизических условий;
- Применение методов машинного обучения и глубокого обучения для обработки и анализа больших массивов астрономических данных в задачах наблюдательной астрофизики;
- Разработка новых подходов и методов организации ведения и работы со сверхбольшими архивами данных спутниковых наблюдений Земли.

В рамках данной тематики опубликовано в 2023 году 31 работа, из них 12 в рецензируемых изданиях, 10 - входящих в Scopus и WoS (плановый показатель 3).

Таким образом, можно считать, что поставленные в плане задачи на 2023 год в рамках данной темы полностью решены.



## ПУБЛИКАЦИИ

Опубликовано – 33 работа, из них:

- в зарубежных изданиях – 5
- книг – 0
- статьи в отечественных научных рецензируемых журналах – 7
- статьи в сборниках материалов конференций – 3
- материалы конференции - 5
- доклады, тезисы – 11
- публикаций по грантам РФФИ – 2
- в соавторстве с зарубежными авторами – 3
- статьи в WOS и Scopus – 10
- статьи в WOS и Scopus (Q1, Q2) – 4
- статьи в WOS и Scopus (Q3, Q4) - 6

### Публикации в зарубежных изданиях

1. *Gulyaeva T., Lukianova R., Haralambous H.* Ionosphere heterogeneities at dawn–dusk terminator related to the Starlink satellites launch disaster on 3–8 February 2022 // *Journal of Geophysical Research: Space Physics*, 128. 2023. e2023JA031577. <https://doi.org/10.1029/2023JA031577>. **(РИНЦ, Scopus, WoS, Q2)**
2. *Kozlovsky A., Myllymaa M., Lukianova R., Raita, T., Lester M.* Influence of atmospheric circulation on orientation of auroral arcs // *Journal of Geophysical Research: Space Physics*. 128. 2023. e2023JA031294. <https://doi.org/10.1029/2023JA031294>. **(РИНЦ, Scopus, WoS, Q2)**
3. *Hartley P., Bonaldi A., Braun R., Cornu D., Semelin B., Lu X., Aicardi S., Salomé P., Marchal A., Freundlich J., Combes F., Tasse C., Heneka C., Delli Veneri M., Soroka A., Gubanov F., Meshcheryakov A., Fraga B., Bom C.R., Brüggem M., Shaw A. K., Patra N., Chakraborty A., Mondal R., Choudhuri S., Mazumder A., Jagannath M., Hardcastle M. J., Forbrich J., Smith L., Stolyarov V., Ashdown M., Coles J., Håkansson H., Sjöberg A., Toribio M. C., Önnheim M., Olberg M., Gustavsson E., Lindqvist M., Jirstrand M., Conway J., Hess K. M., Jurek R. J., Kitaeff S., Serra P., Shen A. X., van der Hulst J. M., Westmeier T., Alberdi A., Cannon J., Darriba L., Garrido J., Gószta J., Herranz D., Jones M. G., Kamphuis P., Kleiner D., Márquez I., Moldón J., Pandey-Pommier M., Parra M., Sabater J., Sánchez S., Sorgho A., Verdes-Montenegro L., Fourestey G., Galan A., Gheller C., Korber D., Peel A., Sargent M., Tolley E., Liu B., Chen R., Peng B., Yu L., Xi H., Yu K., Guo Q., Pei W., Liu Y., Wang Y., Chen X., Zhang X., Ni S., Zhang J., Gao L., Zhao M., Zhang L, Zhang H., Wang X., Ding J., Zuo S., Mao Y., Vafaei Sadr A., Kunz M., Bassett B., Nistane V., Oozeer N., Jaiswal S., Lao B., Aditya J. N. H. S., Zhang Y., Wang A., Yang X.* SKA Science Data Challenge 2: analysis and results. // *Monthly Notices of the Royal Astronomical Society*, 2023, Volume 523, Issue 2, pp.1967–1993. DOI:10.1093/mnras/stad1375 **(РИНЦ, WoS, Scopus Q1)**
4. *Voskresenskaia, S. ; Meshcheryakov, A. ; Lyskova, N.* ComPACT: ACT+Planck galaxy cluster catalogue. // *Monthly Notices of the Royal Astronomical Society*, 2023 **(в печати)**. DOI: 10.48550/arXiv.2309.17077 **(РИНЦ, WoS, Scopus Q1)**
5. *Soroka A. G., Meshcheryakov A. V., Gerasimov S. V.* Deep Reinforcement Learning for the Capacitated Pickup and Delivery Problem with Time Windows. // *Pattern Recognition and Image Analysis*, 2023, Vol. 33, No. 2, pp. 169–178. DOI: 10.1134/S1054661823020165 **(РИНЦ, WOS, Scopus, Q3)**

## Публикации в отечественных научных рецензируемых изданиях

6. Прошин А.А., Бурцев М.А., Кобец Д.А., Сычугов И.Г., Радченко М.В. Система контроля функционирования распределённого программно-аппаратного комплекса ЦКП «ИКИ Мониторинг» // Современные проблемы дистанционного зондирования Земли из космоса. 2023. Т. 20. № 1. С. 95-110. DOI: 10.21046/2070-7401-2023-20-1-95-110. (**Scopus, Q3, РИНЦ**)
7. Балашов И.В., Марченков В.В., Прошин А.А., Руткевич Б.П., Уваров И.А. Разработка сервисов для эффективного предоставления доступа к временным сериям данных сверхбольших распределённых архивов // Современные проблемы дистанционного зондирования Земли из космоса. 2023. Т. 20. № 2. С. 113-122. DOI: 10.21046/2070-7401-2023-20-2-113-122. (**Scopus, Q3, РИНЦ**)
8. Лукьянова Р.Ю. Влияние продольных токов на электронную концентрацию в ионосфере: сопряженные наблюдения спутников SWARM и радара ESR // Космические исследования, Т. 66. № 6, с. 466–475, 2023, doi:10.31857/S0023420623600083. 10.1134/S0010952523700454. (РИНЦ) Переводная версия: The Influence of Field-Aligned Currents on Electron Density in the Ionosphere: Combined Observations of SWARM Satellites and ESR Radar // Cosmic Research, Vol. 61, No. 6, pp. 491–500, 2023, doi:10.1134/S0010952523700454. (**WOS, Scopus, Q4**)
9. Лукьянова Р.Ю. Стратосферный аэрозоль над регионом Казахстана по данным спутника Suomi NPP // Современные проблемы дистанционного зондирования Земли из космоса, Т. 20. № 2, с. 229-237, 2023, doi:10.21046/2070-7401-2023-20-229-237. (**РИНЦ, Scopus, Q3**)
10. Куприянов А.О., Петрукович А.А., Мёрзлый А.М., Никифоров О.В., Сахаров Я.А., Чернышев А.А., Янаков А.Т., Замогильный Д. Экспериментальные результаты применения трансionoсферного зондирования на основе ГНСС для оперативной оценки параметров высокоширотной ионосферы во время экспедиции на архипелаге земля Франца-Иосифа // Известия вузов «Геодезия и аэрофотосъемка». 2023. doi:10.30533/GiA-2023-001. (**РИНЦ**)
11. Кузьмин А.К., Мерзлый А.М., Никифоров О.В., Петрукович А.А., Потанин Ю.Н., Садовский А.М., Соколов А.Д., Янаков А.Т. Аннотированный атлас примеров изображений эмиссий в авроральных структурах, зарегистрированных имаджерами и изображающими спектрографами с разных орбит и поверхности земли. Часть 2. Авроральные и аврора-подобные структуры, возбужденные природными источниками, включая волны нескольких типов // Гелиогеофизические Исследования. 2023. - Принято в печать (**РИНЦ**)
12. Сорока А.Г., Мещеряков А.В. Оптимизация маршрутов большой размерности с использованием глубоких нейронных сетей // Доклады Российской академии наук. Математика, информатика, процессы управления, 2023 (**в печати**). (**РИНЦ, WOS, Scopus**)

## Статьи в сборниках материалов конференций

13. Прошин А.А., Кобец Д.А. Организация распределенной многопоточковой обработки спутниковых данных в интересах систем мониторинга окружающей среды // Сборник трудов всероссийской конференции с международным участием "Обработка пространственных данных в задачах мониторинга природных и антропогенных процессов" (SDM-2023) [Электронный ресурс]. 22-25 августа 2023 г. Бердск. Новосибирск: ФИЦ ИВТ, 2023. С. 132-136. DOI: 10.25743/sdm.2023.55.18.023. (**РИНЦ**)
14. Лупян Е.А., Бурцев М.А., Прошин А.А. Оптимизация хранения временных рядов данных // Материалы VII международной научно-практической конференции "Информационные технологии и высокопроизводительные вычисления". 11-13 сентября 2023 г. Хабаровск: ХФИЦ ДВО РАН, 2023.
15. Lukianova R., Siylkanova A., Daurbaeva G. Variations in ionospheric electron density and tropospheric aerosol during the November 14, 2014 Kajisai earthquake: conjugated observations of the SWARM and SUOMI NPP satellites // Сборник трудов XI Казахстанско-китайского

международного симпозиума «Прогноз землетрясений, оценка сейсмической опасности и сейсмического риска в центральной Азии», Алматы, 2023 г., с. 317-323.

## Материалы конференции

16. *Кобец Д.А., Прошин А.А.* Текущее состояние и перспективы развития подсистемы организации распределенной многопоточной обработки спутниковых данных ЦКП «ИКИ-Мониторинг» // Материалы 21-й международной конференции "Современные проблемы дистанционного зондирования Земли из космоса". Электронный сборник материалов конференции. 13 - 17 ноября 2023. ИКИ РАН. Москва, 2023. С. 30. DOI: 10.21046/21DZZconf-2023a. **(РИНЦ)**
17. *Прошин А.А., Бурцев М.А., Лупян Е.А.* Возможности оптимизации хранения временных рядов данных ДЗЗ // Материалы 21-й международной конференции "Современные проблемы дистанционного зондирования Земли из космоса". Электронный сборник материалов конференции. 13 - 17 ноября 2023. ИКИ РАН. Москва, 2023. С. 102. DOI: 10.21046/21DZZconf-2023a. **(РИНЦ)**
18. *Марченков В.В., Руткевич Б.П., Балашов И.В., Уваров И.А., Прошин А.А.* Возможности подсистемы анализа временных серий данных сверхбольших распределенных архивов в информационных системах семейства «Созвездие-ВЕГА» // Материалы 21-й международной конференции "Современные проблемы дистанционного зондирования Земли из космоса". Электронный сборник материалов конференции. 13 - 17 ноября 2023. ИКИ РАН. Москва, 2023. С. 98. DOI: 10.21046/21DZZconf-2023a. **(РИНЦ)**
19. *Прошин А.А., Бурцев М.А.* Организация работы с данными архивов ЦКП «ИКИ-Мониторинг» на основе использования интерактивной среды разработки Jupyter Notebook // Материалы 21-й международной конференции "Современные проблемы дистанционного зондирования Земли из космоса". Электронный сборник материалов конференции. 13 - 17 ноября 2023. ИКИ РАН. Москва, 2023. С. 101. DOI: 10.21046/21DZZconf-2023a. **(РИНЦ)**
20. *Лукьянова Р.Ю.* Глобальные аномалии нейтральной плотности термосферы во время магнитных бурь по наблюдениям КА SWARM // Материалы 21-й Международной конференции «Современные проблемы дистанционного зондирования Земли из космоса». Москва: ИКИ РАН, 2023, с. 429, doi:10.21046/21DZZconf-2023a. **(РИНЦ)**

## Тезисы, доклады

21. *Лукьянова Р.Ю.* Влияние продольных токов на ионосферные неоднородности: сопряженные наблюдения низкоорбитальных полярных спутников Swarm и радара EISCAT // 18-я конференция "Физика плазмы в солнечной системе", 6-10 февраля 2023 г., Москва, ИКИ РАН. тезисы докладов. устный доклад.
22. *Lukianova R.* Contribution of mesosphere-lower thermosphere perturbations to the formation of the extreme Arctic polar vortex // 28th IUGG General Assembly, 11-20 July 2023, Berlin, Germany, <https://doi.org/10.57757/IUGG23-3944>. тезисы докладов. устный доклад.
23. *Мещеряков А.В., Герасимов С.В., Васильев С.М.* Применение нейросетевых моделей для построения карты далеких ядер галактик по данным цифровых обзоров неба. // Ломоносовские чтения. Тезисы докладов научной конференции. Москва, 2023. С. 103-104. Устный доклад.
24. *Мещеряков А.В., Герасимов С.В., Сорока А.Г., Михельсон Г.В.* Применение классических эвристических и точных методов оптимизации маршрутов в статической постановке задачи маршрутизации транспортных средств с реальными ограничениями // Ломоносовские чтения. Тезисы докладов научной конференции. Москва, 2023. С. 110-111. Устный доклад.
25. *Мещеряков А.В., Немешаева А.А., Герасимов С.В.* Нейросетевые методы сегментации скоплений галактик и радиоисточников в данных микроволнового космического обзора

- ПЛАНК // Ломоносовские чтения. Тезисы докладов научной конференции. Москва, 2023. С. 112-113. Устный доклад.
26. *Мещеряков А.В., Сорока А.Г., Герасимов С.В.* Подходы к решению задач оптимизации маршрутов с большой размерностью (>1000) // Ломоносовские чтения. Тезисы докладов научной конференции. Москва, 2023. С. 113-114. Устный доклад.
  27. *Мещеряков А.В., Сорока А.Г., Герасимов С.В., Михельсон Г.В.* SmartRoutes: нейросетевые решения задач оптимизации маршрутов с ограничениями реального мира // Тихоновские чтения. Тезисы докладов научной конференции. Москва, 2023. С. 122. Устный доклад.
  28. *Мещеряков А.В., Тэаро К.А., Герасимов С.В.* SRGz: библиотека моделей машинного обучения для задач рентгеновской астрофизики // Тихоновские чтения. Тезисы докладов научной конференции. Москва, 2023. С. 125-126. Устный доклад.
  29. *Мещеряков А.В.* Применение машинного обучения для определения физических характеристик рентгеновских источников: результаты для СРГ/еРОЗИТА // Астрофизика высоких энергий сегодня и завтра 2023 (HEA-2023). 18-21 декабря 2023. Москва. Электронные тезисы докладов. Устный доклад.
  30. *Воскресенская С., Мещеряков А.В., Лыскова Н.С.* ComPACT: создание комбинированного каталога скоплений галактик АКТ + Планк с использованием глубокого обучения // Астрофизика высоких энергий сегодня и завтра 2023 (HEA-2023). 18-21 декабря 2023. Москва. Электронные тезисы докладов. Устный доклад.
  31. *Andrew Soroka, Alex Meshcheryakov, Sergey Gerasimov, German Mikhelson.* Smart Routes: a system for development and comparison of algorithms for solving vehicle routing problems with realistic constraints // XXV International Conference "Data Analytics and Management in Data Intensive Domains" (DAMDID/RCDL 2023), HSE University, Moscow October 24-27, 2023.

#### Публикации, подготовленные в соавторстве с зарубежными учеными

1. *Gulyaeva T., Lukianova R., Haralambous H.* Ionosphere heterogeneities at dawn–dusk terminator related to the Starlink satellites launch disaster on 3–8 February 2022 // *Journal of Geophysical Research: Space Physics*, 128, e2023JA031577. 2023. <https://doi.org/10.1029/2023JA031577>. (РИНЦ, Scopus, WoS, Q2)
2. *Kozlovsky A., Myllymaa M., Lukianova R., Raita, T., Lester M.* Influence of atmospheric circulation on orientation of auroral arcs // *Journal of Geophysical Research: Space Physics*, 128, e2023JA031294. 2023. <https://doi.org/10.1029/2023JA031294>. (РИНЦ, Scopus, WoS, Q2)
3. *Hartley P., Bonaldi A., Braun R., Cornu D., Semelin B., Lu X., Aicardi S., Salomé P., Marchal A., Freundlich J., Combes F., Tasse C., Heneka C., Delli Veneri M., Soroka A., Gubanov F., Meshcheryakov A., Fraga B., Bom C.R., Brüggem M., Shaw A. K., Patra N., Chakraborty A., Mondal R., Choudhuri S., Mazumder A., Jagannath M., Hardcastle M. J., Forbrich J., Smith L., Stolyarov V., Ashdown M., Coles J., Håkansson H., Sjöberg A., Toribio M. C., Önnheim M., Olberg M., Gustavsson E., Lindqvist M., Jirstrand M., Conway J., Hess K. M., Jurek R. J., Kitaeff S., Serra P., Shen A. X., van der Hulst J. M., Westmeier T., Alberdi A., Cannon J., Darriba L., Garrido J., Gószta J., Herranz D., Jones M. G., Kamphuis P., Kleiner D., Márquez I., Moldón J., Pandey-Pommier M., Parra M., Sabater J., Sánchez S., Sorgho A., Verdes-Montenegro L., Fourestey G., Galan A., Gheller C., Korber D., Peel A., Sargent M., Tolley E., Liu B., Chen R., Peng B., Yu L., Xi H., Yu K., Guo Q., Pei W., Liu Y., Wang Y., Chen X., Zhang X., Ni S., Zhang J., Gao L., Zhao M., Zhang L, Zhang H., Wang X., Ding J., Zuo S., Mao Y., Vafaei Sadr A., Kunz M., Bassett B., Nistane V., Oozeer N., Jaiswal S., Lao B., Aditya J. N. H. S., Zhang Y., Wang A., Yang X.* SKA Science Data Challenge 2: analysis and results. // *Monthly Notices of the Royal Astronomical Society*, 2023, Volume 523, Issue 2, pp.1967-1993. DOI:10.1093/mnras/stad1375 (РИНЦ, WoS, Scopus Q1)

## Публикации по грантам РФФ

1. Meshcheryakov, A. V. ; Borisov, V. D. ; Khorunzhev, G. A. ; Medvedev, P. A. ; Gilfanov, M. R. ; Belvedersky, M. I. ; Sazonov, S. Yu. ; Burenin, R. A. ; Krivonos, R. A. ; Bikmaev, I. F. ; Khamitov, I. M. ; Gerasimov, S. V. ; Mashechkin, I. V. ; Sunyaev, R. A. SRGz: Machine Learning Methods and Properties of the Catalog of SRG/eROSITA Point X-ray Source Optical Counterparts in the DESI Legacy Imaging Surveys Footprint // *Astronomy Letters*, Volume 49, Issue 7, p.359-409 (2023). DOI: 10.1134/S1063773723070022 **(РИНЦ, WoS, Scopus)**
2. Meshcheryakov, A.V.; Khorunzhev, G. A.; Voskresenskaia, S.; Medvedev, P.S., Gil'fanov, M. R.; Sunyaev, R.A. SRGz: classification of eROSITA point X-ray sources in the 1%DESI area and calibration of photometric redshifts // *Astronomy Letters*, Volume 49, Issue 11, p.?? (2023). **(в печати)**. DOI: **(РИНЦ, WoS, Scopus)**

## Список публикаций по теме «Космос-Д» в рецензируемых изданиях

1. Gulyaeva T., Lukianova R., Haralambous H. Ionosphere heterogeneities at dawn–dusk terminator related to the Starlink satellites launch disaster on 3–8 February 2022 // *Journal of Geophysical Research: Space Physics*, 128. 2023. e2023JA031577. <https://doi.org/10.1029/2023JA031577>. **(РИНЦ, Scopus, WoS, Q2)**
2. Kozlovsky A., Myllymaa M., Lukianova R., Raita, T., Lester M. Influence of atmospheric circulation on orientation of auroral arcs // *Journal of Geophysical Research: Space Physics*. 128. 2023. e2023JA031294. <https://doi.org/10.1029/2023JA031294>. **(РИНЦ, Scopus, WoS, Q2)**
3. Hartley P., Bonaldi A., Braun R., Cornu D., Semelin B., Lu X., Aicardi S., Salomé P., Marchal A., Freundlich J., Combes F., Tasse C., Heneka C., Delli Veneri M., Soroka A., Gubanov F., Meshcheryakov A., Fraga B., Bom C.R., Brüggem M., Shaw A. K., Patra N., Chakraborty A., Mondal R., Choudhuri S., Mazumder A., Jagannath M., Hardcastle M. J., Forbrich J., Smith L., Stolyarov V., Ashdown M., Coles J., Håkansson H., Sjöberg A., Toribio M. C., Önnheim M., Olberg M., Gustavsson E., Lindqvist M., Jirstrand M., Conway J., Hess K. M., Jurek R. J., Kitaeff S., Serra P., Shen A. X., van der Hulst J. M., Westmeier T., Alberdi A., Cannon J., Darriba L., Garrido J., Gósz J., Herranz D., Jones M. G, Kamphuis P., Kleiner D., Márquez I., Moldón J., Pandey-Pommier M., Parra M., Sabater J., Sánchez S., Sorgho A., Verdes-Montenegro L., Fourestey G., Galan A., Gheller C., Korber D., Peel A., Sargent M., Tolley E., Liu B., Chen R., Peng B., Yu L., Xi H., Yu K., Guo Q., Pei W., Liu Y., Wang Y., Chen X., Zhang X., Ni S., Zhang J., Gao L., Zhao M., Zhang L, Zhang H., Wang X., Ding J., Zuo S., Mao Y., Vafaei Sadr A., Kunz M., Bassett B., Nistane V., Oozeer N., Jaiswal S., Lao B., Aditya J. N. H. S., Zhang Y., Wang A., Yang X.. SKA Science Data Challenge 2: analysis and results. // *Monthly Notices of the Royal Astronomical Society*, 2023, Volume 523, Issue 2, pp.1967-1993. DOI:10.1093/mnras/stad1375 **(РИНЦ, WoS, Scopus Q1)**
4. Voskresenskaia, S. ; Meshcheryakov, A. ; Lyskova, N. ComPACT: ACT+Planck galaxy cluster catalogue. // *Monthly Notices of the Royal Astronomical Society*, 2023 **(в печати)**. DOI: 10.48550/arXiv.2309.17077 **(РИНЦ, WoS, Scopus Q1)**
5. Soroka A. G., Meshcheryakov A. V., Gerasimov S. V.. Deep Reinforcement Learning for the Capacitated Pickup and Delivery Problem with Time Windows. // *Pattern Recognition and Image Analysis*, 2023, Vol. 33, No. 2, pp. 169–178. DOI: 10.1134/S1054661823020165 **(РИНЦ, WOS, Scopus)**
6. Прошин А.А., Бурцев М.А., Кобец Д.А., Сычугов И.Г., Радченко М.В. Система контроля функционирования распределённого программно-аппаратного комплекса ЦКП «ИКИ Мониторинг» // *Современные проблемы дистанционного зондирования Земли из космоса*. 2023. Т. 20. № 1. С. 95-110. DOI: 10.21046/2070-7401-2023-20-1-95-110. **(Scopus, Q3, РИНЦ)**
7. Балашов И.В., Марченков В.В., Прошин А.А., Руткевич Б.П., Уваров И.А. Разработка сервисов для эффективного предоставления доступа к временным сериям данных

- сверхбольших распределённых архивов // Современные проблемы дистанционного зондирования Земли из космоса. 2023. Т. 20. № 2. С. 113-122. DOI: 10.21046/2070-7401-2023-20-2-113-122. (**Scopus, Q3, РИНЦ**)
8. Лукьянова Р.Ю. Влияние продольных токов на электронную концентрацию в ионосфере: сопряженные наблюдения спутников SWARM и радара ESR // Космические исследования, Т. 66. № 6, с. 466–475, 2023, doi:10.31857/S0023420623600083. 10.1134/S0010952523700454. (РИНЦ) Переводная версия: The Influence of Field-Aligned Currents on Electron Density in the Ionosphere: Combined Observations of SWARM Satellites and ESR Radar // Cosmic Research, Vol. 61, No. 6, pp. 491–500, 2023, doi:10.1134/S0010952523700454. (**WOS, Scopus, Q4**)
  9. Лукьянова Р.Ю. Стратосферный аэрозоль над регионом Казахстана по данным спутника Suomi NPP // Современные проблемы дистанционного зондирования Земли из космоса, Т. 20. № 2, с. 229-237, 2023, doi:10.21046/2070-7401-2023-20-229-237. (**РИНЦ, Scopus, Q3**)
  10. Куприянов А.О., Петрукович А.А., Мёрзлый А.М., Никифоров О.В., Сахаров Я.А., Чернышев А.А., Янаков А.Т., Замогильный Д. Экспериментальные результаты применения трансionoсферного зондирования на основе ГНСС для оперативной оценки параметров высокоширотной ионосферы во время экспедиции на архипелаге земля Франца-Иосифа // Известия вузов «Геодезия и аэрофотосъемка». 2023. doi:10.30533/GiA-2023-001. (**РИНЦ**)
  11. Кузьмин А.К., Мерзлый А.М., Никифоров О.В., Петрукович А.А., Потанин Ю.Н., Садовский А.М., Соколов А.Д., Янаков А.Т. Аннотированный атлас примеров изображений эмиссий в авроральных структурах, зарегистрированных имаджерами и изображающими спектрографами с разных орбит и поверхности земли. Часть 2. Авроральные и аврора-подобные структуры, возбужденные природными источниками, включая волны нескольких типов // Гелиогеофизические Исследования. 2023. - Принято в печать (**РИНЦ**)
  12. Сорока А.Г., Мещеряков А.В. Оптимизация маршрутов большой размерности с использованием глубоких нейронных сетей // Доклады Российской академии наук. Математика, информатика, процессы управления, 2023 (в печати). (**РИНЦ, WOS, Scopus**)