

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 АААА-А19-119072590030-9

УТВЕРЖДАЮ

Директор

Федерального государственного
бюджетного учреждения науки
Института космических исследований
Российской академии наук
Член-корреспондент РАН


А.А.Петрукович

« 25 » декабря 2019 г.

м.п.



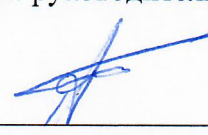
отчет о НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

Большие данные в космических исследованиях: астрофизика, солнечная система, геосфера
(промежуточный)

Тема Космос-Д

0024-2019-14

Научный руководитель


к.т.н. А.А. Прошин

« 25 » декабря 2019 г.

Москва

2019

СПИСОК ИСПОЛНИТЕЛЕЙ

СПИСОК ИСПОЛНИТЕЛЕЙ

Руководитель темы:

старший научный сотрудник
к.т.н.

11.12.19



Прошин А.А.

Ответственные исполнители разделов темы:

ведущий научный сотрудник
д.ф.-м.н.

11.12.19



Лукьянова Р.Ю.

научный сотрудник
к.ф.-м.н.

11.12.19



Мещеряков А.В.

старший научный сотрудник
к.т.н.

11.12.19



Прошин А.А.

СПИСОК ИСПОЛНИТЕЛЕЙ отдел 54

Должность, ученая степень, ученое звание	Дата	Подпись	ФИО	Номера разделов отчета в работе по которым принимал участие исполнитель
младший научный сотрудник	11.12.2019		Евдокимова М.А.	
младший научный сотрудник	11.12.2019		Лукин А.С.	1
ведущий научный сотрудник дн	11.12.2019		Лукьянова Р.Ю.	1
ведущий инженер	11.12.2019		Маевский Е.В.	1
старший научный сотрудник кн	11.12.2019		Мерзлый А.М.	1
ведущий инженер	11.12.2019		Никифоров О.В.	1
научный сотрудник кн	11.12.2019		Ожередов В.А.	1
младший научный сотрудник	11.12.2019		Шустов П.И.	1
ведущий инженер	11.12.2019		Янаков А.Т.	1

СПИСОК ИСПОЛНИТЕЛЕЙ отдел 52

Должность, ученая степень, ученое звание	Дата	Подпись	ФИО	Номера разделов отчета в работе по которым принимал участие исполнитель
старший лаборант	11.12.2019		Бельведерский М.И.	2
старший лаборант	11.12.2019		Борисов В.Д.	2
старший научный сотрудник кн	11.12.2019		Буренин Р.А.	2
старший лаборант	11.12.2019		Быков С.Д.	2
инженер	11.12.2019		Зазнобин И.А.	2
научный сотрудник кн	11.12.2019		Карасев Д.И.	2
ведущий инженер	11.12.2019		Лыскова Н.С.	2
инженер	11.12.2019		Ляпин А.Р.	2
младший научный сотрудник кн	11.12.2019		Медведев П.С.	2
научный сотрудник кн	11.12.2019		Мещеряков А.В.	2
старший лаборант	11.12.2019		Николаева С.М.	2
научный сотрудник кн	11.12.2019		Филиппова Е.В.	2
младший научный сотрудник кн	11.12.2019		Хорунжев Г.А.	2

СПИСОК ИСПОЛНИТЕЛЕЙ отдел 56

Должность, ученая степень, ученое звание	Дата	Подпись	ФИО	Номера разделов отчета, в работе по которым принимал участие исполнитель
старший научный сотрудник, к.т.н.	11.12.2019		Балашов И.В.	3
инженер	11.12.2019		Бриль А.А.	3
старший научный сотрудник, к.т.н.	11.12.2019		Бурцев М.А.	3
главный специалист	11.12.2019		Дегай А.Ю.	3
научный сотрудник, к.т.н.	11.12.2019		Кашницкий А.В.	3
младший научный сотрудник, к.т.н.	11.12.2019		Кобец Д.А.	3
главный специалист, к.ф.-м.н.	11.12.2019		Козочкина А.А.	3
инженер	11.12.2019		Константинова А.М.	3
старший научный сотрудник, к.ф.-м.н.	11.12.2019		Крашенинникова Ю.С.	3
заведующий отделом, д.т.н.	11.12.2019		Лупян Е.А.	3
ведущий научный сотрудник, к.ф.-м.н.	11.12.2019		Мазуров А.А.	3
ведущий конструктор.	11.12.2019		Марченков В.В.	3
младший научный сотрудник	11.12.2019		Матвеев А.М.	3
старший научный сотрудник, к.т.н.	11.12.2019		Прошин А.А.	3
заведующий лабораторией, к.ф.-м.н.	11.12.2019		Пыркков В.Н.	3
ведущий инженер	11.12.2019		Радченко М.В.	3
инженер	11.12.2019		Сенько К.С.	3
младший научный сотрудник	11.12.2019		Сычугов И.Г.	3
старший научный сотрудник, к.т.н.	11.12.2019		Толпин В.А.	3

старший научный сотрудник, к.т.н.	11.12.2019		Уваров И.А.	3
инженер	11.12.2019		Черных В.Н.	3
Старший лаборант	11.12.2019		Юдин Д.А.	3

РЕФЕРАТ

Отчет 48 стр., 23 рис., 23 источника.

НАСТОЯЩИЙ ОТЧЕТ ПОСВЯЩЕН ОПИСАНИЮ ОСНОВНЫХ РЕЗУЛЬТАТОВ РАБОТ, ПОЛУЧЕННЫХ В 2019 ГОДУ В РАМКАХ ТЕМЫ "КОСМОС-Д" (№ 0024-2019-14), 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 2019 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 2019 ГОД ПО СЛЕДУЮЩИМ НАПРАВЛЕНИЯМ:

Развитие комплекса взаимосвязанных физико-математических моделей для диагностики и прогнозирования состояния высокоширотной ионосферы. Разработки общей модели взаимодействия в системе солнечный ветер-магнитосфера-ионосфера-нейтральная атмосфера Земли.

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

Разработка новых методов и подходов организации хранения и параллельной обработки сверхбольших объемов данных для решения задач анализа долговременных наблюдений в интересах ведения мониторинга и изучения различных процессов и явлений.

КЛЮЧЕВЫЕ СЛОВА

СИСТЕМЫ ПАРАЛЛЕЛЬНОЙ ОБРАБОТКИ СПУТНИКОВЫХ ДАННЫХ, СВЕРХБОЛЬШИЕ ДОЛГОВРЕМЕННЫЕ АРХИВЫ СПУТНИКОВЫХ ДАННЫХ, БЛОЧНЫЙ ДОСТУП К ДАННЫМ В АРХИВАХ, АНАЛИЗ ДОЛГОВРЕМЕННЫХ СЕРИЙ НАБЛЮДЕНИЙ, ТЕХНОЛОГИЯ UNISAT, МИССИЯ ЭКЗОМАРС-2016, АППАРАТ TGO, MARS EXPRESS, ИНФОРМАЦИОННЫЙ СЕРВИС ARES, ОБРАБОТКА СПУТНИКОВЫХ ДАННЫХ, ДОСТУП К АРХИВАМ СПУТНИКОВЫХ ДАННЫХ, ИНФОРМАЦИОННАЯ СИСТЕМА "АВРОРА-АРКТИКА", АНАЛИЗ СОСТОЯНИЯ ИОНОСФЕРЫ ЗЕМЛИ, ПРОГНОЗЫ СОСТОЯНИЯ АВРОРАЛЬНОГО ОВАЛА.

ИОНОСФЕРНО-МАГНИТОСФЕРНОЕ ВЗАИМОДЕЙСТВИЕ И ЭЛЕКТРОДИНАМИКА, ВЫСОКОШИРОТНАЯ ИОНОСФЕРА, ПРОДОЛЬНЫЕ ТОКИ, ЭЛЕКТРОННАЯ КОНЦЕНТРАЦИЯ, ОБРАБОТКА МАССИВОВ ДАННЫХ СПУТНИКОВЫХ И РАДАРНЫХ ИЗМЕРЕНИЙ, ЧИСЛЕННОЕ МОДЕЛИРОВАНИЕ.

МАШИННОЕ ОБУЧЕНИЕ, ГЛУБОКОЕ ОБУЧЕНИЕ, ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ, КОСМИЧЕСКАЯ ОБСЕРВАТОРИЯ "СПЕКТР-РГ", РЕНТГЕНОВСКАЯ АСТРОФИЗИКА, ОБЗОРЫ НЕБА

СОДЕРЖАНИЕ

СПИСОК ИСПОЛНИТЕЛЕЙ	2
РЕФЕРАТ	7
СОДЕРЖАНИЕ	8
ВВЕДЕНИЕ	10
ГЛАВА 1 Развитие комплекса взаимосвязанных и дополняющих друг друга физико-математических моделей для диагностики и прогнозирования состояния высокоширотной ионосферы в северном и южном полушариях, которые электродинамически связаны между собой через область средних широт. Разработки общей модели взаимодействия в системе солнечный ветер-магнитосфера-ионосфера-нейтральная атмосфера Земли.	12
Введение	12
1.1. Эволюция продольных токов во время магнитной бури на основе анализа данных наблюдений группировки низкоорбитальных спутников SWARM	12
1.1.1 <i>Постановка задачи</i>	12
1.1.2 <i>Данные</i>	13
1.1.3 <i>Результаты</i>	13
1.1.3.1 <i>Эволюция плотности ПТ во время магнитной бури</i>	13
1.1.3.2 <i>Динамика экваториальной границы ПТ</i>	14
1.1.3.3 <i>Экстремумы мелкомасштабных ПТ</i>	15
1.2 Валидация модели распределения электронной концентрации в F-области полярной ионосферы по данным радара ESR	16
1.2.1 <i>Постановка задачи</i>	16
1.2.2 <i>Результаты</i>	17
Заключение	18
ГЛАВА 2 Применение методов машинного обучения и глубокого обучения для обработки и анализа больших массивов астрономических данных в задачах наблюдательной астрофизики	20
Введение	20
2.1 Разработка методов машинного обучения и глубокого обучения для обработки и анализа больших массивов астрономических данных	20
2.1.1 <i>Постановка задачи</i>	20
2.1.2 <i>Разработка алгоритмов машинного обучения для кросс-отождествления и классификации объектов в фотометрических каталогах обзоров неба в различных спектральных диапазонах, измерения фотометрических красных смещений внегалактических объектов</i>	22
2.2 Исследование и разработка горизонтально-масштабируемых алгоритмов обработки и анализа больших астрономических данных	23
Заключение	25
ГЛАВА 3 Разработка новых методов и подходов организации хранения и параллельной обработки сверхбольших объемов данных для решения задач анализа долговременных наблюдений в интересах ведения мониторинга и изучения различных процессов и явлений	26
Введение	26
3.1. Разработка подходов к созданию унифицированной технологии организации блочного доступа к данным в архивах для проведения их эффективной распределенной параллельной обработки	27
3.1.1 <i>Актуальность задачи</i>	27
3.1.2 <i>Унифицированная технология организации блочного доступа к данным в архивах для проведения их эффективной распределенной параллельной обработки</i>	28
3.1.3 <i>Вариант с использованием прямого доступа к файлам данных со стороны серверов обработки</i>	29

3.1.4 Вариант с использованием серверов предварительной подготовки необходимых наборов исходных спутниковых данных.....	30
3.1.5 Общие сведения о текущей программной реализации. Планы ее доработки и внедрения	32
3.1.6 Перспективы использования представленного подхода для формирования спутниковых изображений в картографических WEB интерфейсах	33
3.2. Разработка подходов к построению систем работы данными планетных миссий на основе методов и технологий работы с данными дистанционного мониторинга Земли на примере системы Ares Mars.....	34
3.2.1 Введение.....	34
3.2.2 Общая архитектура системы.....	35
3.2.3 Система ведения архивов данных.....	37
3.2.4 Интерфейсы работы с данными и примеры их использования.....	39
3.2.5 Работы по развитию системы, проведённые в 2109 году.....	41
3.2.6 Заключение	41
3.3. Развитие системы Аврора-Аркика	42
Заключение	44
ЗАКЛЮЧЕНИЕ	45
СПИСОК ЛИТЕРАТУРЫ	46

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 (номер направления в Программе 16, 14);
- Фундаментальные и прикладные научные исследования в области Физики космической плазмы, энергичных частиц, Солнца и солнечно-земных связей (номер направления в Программе 16, 14);
- Фундаментальные и прикладные научные исследования планет и малых тел Солнечной системы (номер направления в Программе 16, 71);
- Фундаментальные и прикладные научные исследования планеты Земля (номер направления в Программе 79,80);
- Фундаментальные и прикладные научные исследования в области механики, систем управления и информатики (номер направления в Программе 21);
- Развитие исследовательской, конструкторской, опытно-экспериментальной базы научного космического приборостроения и методов экспериментальной физики (номер направления в Программе 16).

Эти направления НИР и ОКР соответствуют следующим направлениям фундаментальных исследований, указанным в Программе фундаментальных исследований государственных академий наук на 2013-2020 годы, утвержденной распоряжением Правительства РФ от 03 декабря 2012 г., № 2237-р

п/п	Направление фундаментальных исследований	Номер направления в «Программе»
1	Современные проблемы астрономии, астрофизики и исследования космического пространства, в том числе происхождение, строение и эволюция Вселенной, природа темной материи и темной энергии, исследование Луны и планет, Солнца и солнечно-земных связей, развитие методов и аппаратуры внеатмосферной астрономии и исследований космоса, координатно-временное обеспечение фундаментальных исследований и практических задач	16
2	Современные проблемы физики плазмы, включая физику астрофизической плазмы, физику низкотемпературной плазмы и основы её применения в технологических процессах	14
3	Закономерности формирования минерального, химического и изотопного состава Земли. Космохимия планет и других тел Солнечной системы. Возникновение и эволюция биосферы Земли, биогеохимические циклы и геохимическая роль организмов.	129
4	Научные основы разработки методов, технологий и средств исследования поверхности и недр Земли, атмосферы, включая ионосферу и магнитосферу Земли, гидросферы и криосферы; численное моделирование и геоинформатика: инфраструктура пространственных данных и ГИС-технологии.	138
5	Эволюция окружающей среды и климата под воздействием природных и антропогенных факторов, научные основы рационального	137

	природопользования и устойчивого развития; территориальная организация хозяйства и общества	
6	Общая механика, навигационные системы, динамика космических тел, транспортных средств и управляемых аппаратов, механика живых систем.	21

Тема Космос-Д «Большие данные в космических исследованиях: астрофизика, солнечная система, геосфера» соответствует направлению 16.

Тема впервые сформирована в 2019 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.

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», начатой в 2019 г, 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 комплекса взаимосвязанных и дополняющих друг друга физико-математических моделей для диагностики и прогнозирования состояния высокоширотной ионосферы в северном и южном полушариях, которые электродинамически связаны между собой через область средних широт. Разработки общей модели взаимодействия в системе солнечный ветер-магнитосфера-ионосфера-нейтральная атмосфера Земли.
- Применение методов машинного обучения и глубокого обучения для обработки и анализа больших массивов астрономических данных в задачах наблюдательной астрофизики.
- Разработка новых методов и подходов организации хранения и параллельной обработки сверхбольших объемов данных для решения задач анализа долговременных наблюдений в интересах ведения мониторинга и изучения различных процессов и явлений.

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

ГЛАВА 1 Развитие комплекса взаимосвязанных и дополняющих друг друга физико-математических моделей для диагностики и прогнозирования состояния высокоширотной ионосферы в северном и южном полушариях, которые электродинамически связаны между собой через область средних широт. Разработки общей модели взаимодействия в системе солнечный ветер-магнитосфера-ионосфера-нейтральная атмосфера Земли.

Введение

Задачи работ по данному направлению на 2019 год были связаны с разработкой новых моделей ионосферы, атмосферно-ионосферных и ионосферных связей, привлечением для описания их динамики информации о солнечной активности и солнечном ветре. Создание таких моделей подразумевает использование больших объемов наблюдательных данных из разнородных источников (от Солнца до поверхности Земли). Ранее, в других работах ИКИ РАН, такого рода задачи не ставились. В дальнейшем разрабатываемые модели будут встроены в создаваемую (также в рамках данной темы, см. раздел 3.3) информационную систему «Аврора-Арктика». Предполагалось выполнить валидацию численной модели распределения электронной концентрации в F-области полярной ионосферы по данным сети ионозондов и радарных измерений. Провести анализ эволюции продольных токов во время магнитных бурь по данным высокоточных измерений вариаций магнитного поля и параметров плазмы в верхней ионосфере низкоорбитальными спутниками *SWARM*, оценки динамики широтного положения границ и плотностей продольных токов, выявление мелкомасштабных токовых структур с повышенной плотностью и их связи с локальными авроральными дугами и электроджетами.

1.1. Эволюция продольных токов во время магнитной бури на основе анализа данных наблюдений группировки низкоорбитальных спутников SWARM

1.1.1 Постановка задачи

Продольные электрические токи (ПТ), текущие вдоль силовых линий магнитного поля Земли, обеспечивают электродинамическую связь системы солнечный ветер-магнитосфера-ионосфера. Они расположены главным образом в авроральной зоне и являются основным источником ионосферных электрических полей, авроральных электроджетов и геомагнитных возмущений. Без знания (моделирования) величин продольных токов невозможно построить взаимосвязанные модели магнитосферы и ионосферы. Величина токов не может быть напрямую измерена с Земли с большинства космических аппаратов. Только в последние годы были запущены специализированные спутники *SWARM*.

Массив данных наблюдений группировки спутников *SWARM* впервые использовался для выявления параметров крупно- и мелкомасштабных ПТ в период магнитной бури. Такая морфология бури позволяет не только проследить эволюцию ПТ во время экстремального события космической погоды, но и попытаться разделить эффекты бури (кольцевого тока) и суббури (разрыв тока в хвосте магнитосферы).

В задачи исследования в 2019 г входило определение изменения плотности ПТ, динамики экваториальной границы ПТ в зависимости от интенсивности бури и сопровождающей ее суббулевой активности, а также выявление экстремумов мелкомасштабных ПТ.

1.1.2 Данные

Спутники *SWARM* имеют низкую круговую орбиту, лежащую в меридиональной плоскости и пересекающую полярные области Земли. Все КА *SWARM* оснащены идентичной аппаратурой. Основным модулем является комплекс для магнитных измерений: высокочувствительный векторный и скалярный магнетометры для определения величины и направления полного вектора и вариаций геомагнитного поля с точностью до 0.1 нТл и частотой 1 Гц. Кроме магнитометрической аппаратуры КА оснащены инструментом EFI, который включает в себя ленгмюровский зонд, и позволяет измерять плотность и температуру электронной компоненты ионосферной плазмы. По вариациям внешнего магнитного поля, измеряемым векторным магнитометром, рассчитываются 1-секундные значения плотности ПТ по закону Ампера.

1.1.3 Результаты

1.1.3.1. Эволюция плотности ПТ во время магнитной бури

Плотности ПТ изменяются от ~ 0.2 до $3 \mu\text{A}/\text{m}^2$ и имеют три ярко выраженных пика (рисунок 1.1.3.1.1.). Первый, сравнительно небольшой пик наблюдается 7 сентября, когда на магнитосферу воздействует фронт динамического давления солнечного ветра. Дневные ПТ (рисунок 1.1.3.1.1 a, b) реагируют без каких-либо задержек на этот шок, в то время как ночные ПТ (рисунок 1.1.3.1.1 e-h) начинают увеличиваться позже вместе с развитием суббуревой активности. Интенсивность конкретного пика плотности тока варьируется в разных секторах MLT, и наиболее четко пики проявляются в ранние утренние и предполуночные часы. На фазе восстановления 9 сентября плотность ПТ возвращается к значению $\sim 0.2 \mu\text{A}/\text{m}^2$.

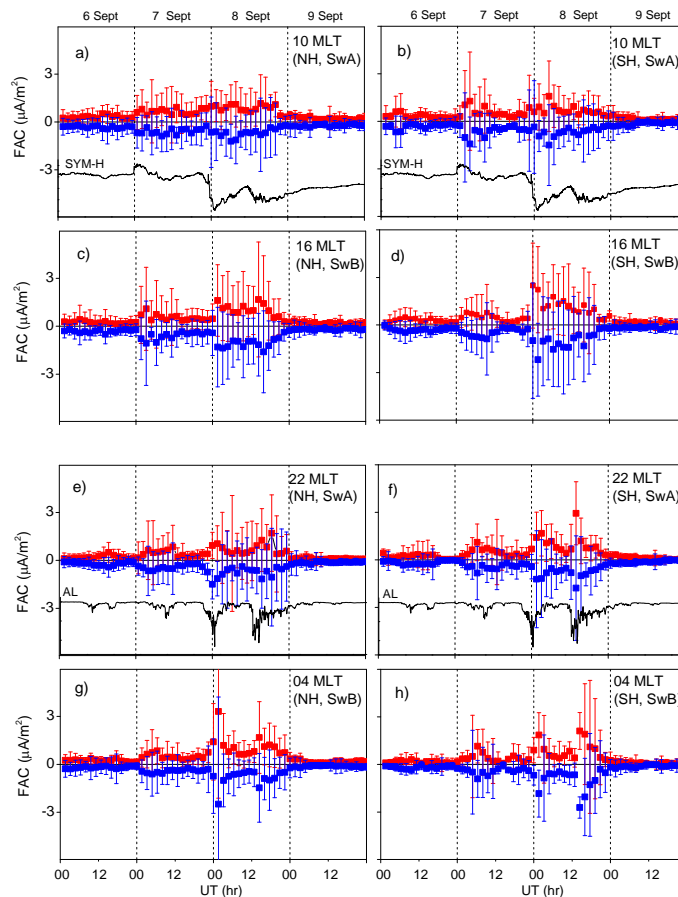


Рисунок 1.1.3.1.1 - Плотность ПТ в северном (слева) и в южном (справа) полушариях. Показаны также индексы *SYM-H* (черная линия на графиках а, б) и *AE* (е, ф)

1.1.3.2. Динамика экваториальной границы ПТ

Усиление воздействия солнечного ветра на магнитосферу и накопление открытого магнитного потока во время геомагнитной бури приводят к расширению полярной шапки и аврорального овала в целом. На рисунке 1.1.3.2.1 показана эволюция экваториальной границы ПТ. Граница области, занятой током, определялась как широта, выше которой по крайней мере в пяти точках измерений подряд модуль плотности ПТ $\geq 0,1 \mu\text{A}/\text{m}^2$. Из рисунка видно, что в предбуревой период ночная граница ПТ наблюдается на $\sim 65^\circ$ MLat, а дневная достигает $>70^\circ$ MLat. Во время главной фазы ПТ смещаются к экватору, так что ночная граница находится на $49\text{-}50^\circ$ MLat, а дневная – примерно на 54° MLat.

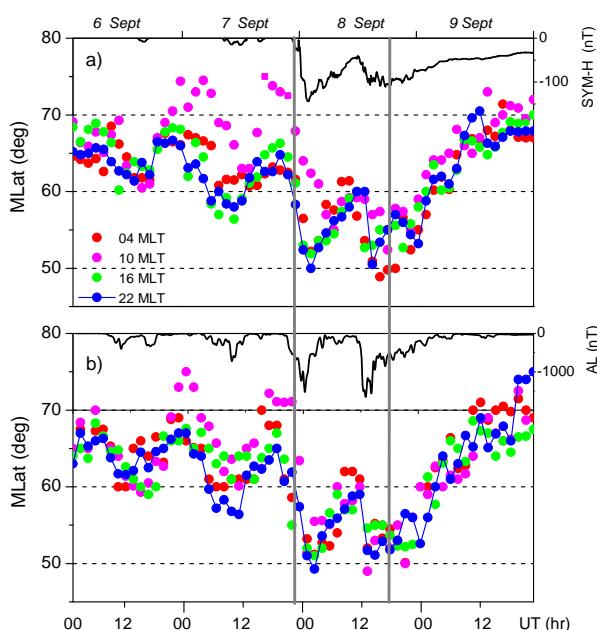


Рисунок 1.1.3.2.1 - Широтное положение экваториальных границ ПТ в северном (а) и в южном (б) полушариях.

Показаны также индексы *SYM-H* (черная линия на графике а) и *AE* (б)

Экваториальное смещение ПТ примерно коррелирует с интенсивностью бури (индекс *SYM-H*), но суббури модифицируют это соотношение. На рисунке 1.1.3.2.2 отдельно для главной фазы и фазы восстановления показаны корреляции между *SYM-H* и положением ночной границы. Коэффициенты корреляции для главной фазы и фазы восстановления близки ($r = 0,88$ и $0,87$), в то время как соответствующие уравнения регрессии значительно различаются. Во время главной фазы бури экваториальное смещение определяется уравнением $\text{MLat} = 63,1 + 0,1 \cdot \text{SYM-H}$. Когда начинается фаза восстановления, сдвиг описывается выражением $\text{MLat} = 79,5 + 0,3 \cdot \text{SYM-H}$. Более быстрое смещение ПТ к полюсу связано с быстрым снижением суббуревой активности 9 сентября.

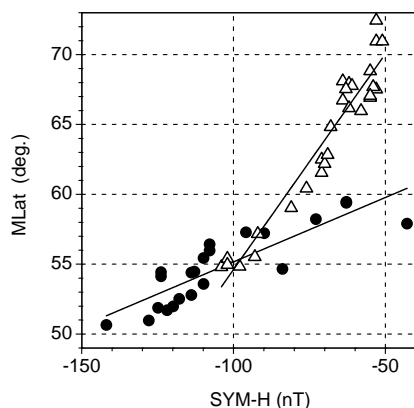


Рисунок 1.1.3.2.2 - Корреляции между индексом SYMH и широтным положением ночной (~22 MLT) экваториальной границы ПТ: черные точки (открытые треугольники) соответствуют главной фазе (фазе восстановления)

1.1.3.3. Экстремумы мелкомасштабных ПТ

Стандартные 1 Гц значения данных измерений КА *SWARM* обеспечивают пространственный масштаб около 7,5 км. Мелкомасштабные пики плотности ПТ могут быть более чем на порядок больше, чем плотности, усредненные вдоль траектории. Значения $>50 \mu\text{A}/\text{m}^2$ появляются в периоды общего повышения плотности в главную фазу бури во время активации суббурь. Самая большая плотность $80 \mu\text{A}/\text{m}^2$ - у втекающего ПТ в секторе ранних утренних часов MLT. В непосредственной близости к втекающему току обнаруживаются сравнимые с ним по величине значения вытекающего тока (рисунок 1.1.3.3.1).

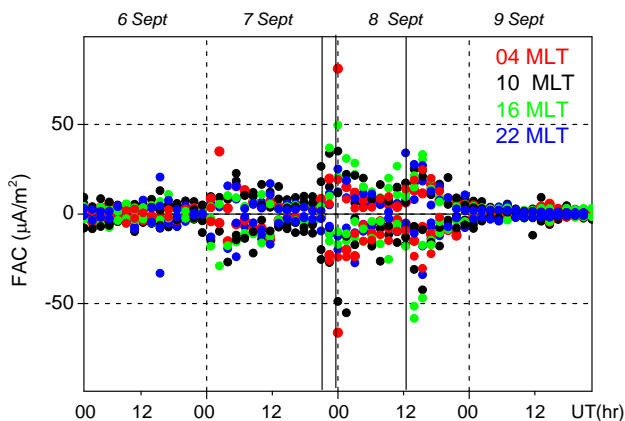


Рисунок 1.1.3.3.1 - Рис. 8. Экстремумы мелкомасштабных (7.5 км) ПТ

4-минутный отрезок измерений плотности тока и параметров плазмы (электронная концентрация и температура, а также электрический потенциал космического аппарата) для *SwB* показан на рисунке 1.1.3.3.2. В 00:10:18-00:10:19 UTC спутник наблюдает биполярную токовую структуру с экстремальной плотностью, состоящую из втекающего ($81 \mu\text{A}/\text{m}^2$) и вытекающего ($66 \mu\text{A}/\text{m}^2$) ПТ. Парные ПТ имеют сравнительно сопоставимые значения, т.е. они сбалансированы и вероятно замыкаются локально. Вокруг области вытекающего тока образуется узкая область повышенной электронной концентрации до $2 \cdot 10^5 \text{ cm}^{-3}$ и увеличенной до 10^4 K температуры, что примерно на 20% выше их значений в окружающей плазме. Наблюдаемая структура мелкомасштабных ПТ очень высокой плотности тока, которая сопровождается локальным увеличением N_e и T_e , вероятно, связана с мезомасштабной дискретной авроральной дугой/лучом. Анализ мелко-

масштабных плотностей ПТ для всех событий магнитных бурь в период работы *SWARM* показал, что биполярные структуры ПТ экстремальной плотности ($>50 \text{ мкА/м}^2$) обнаруживаются в каждом событии, но не более 1-2 раз.

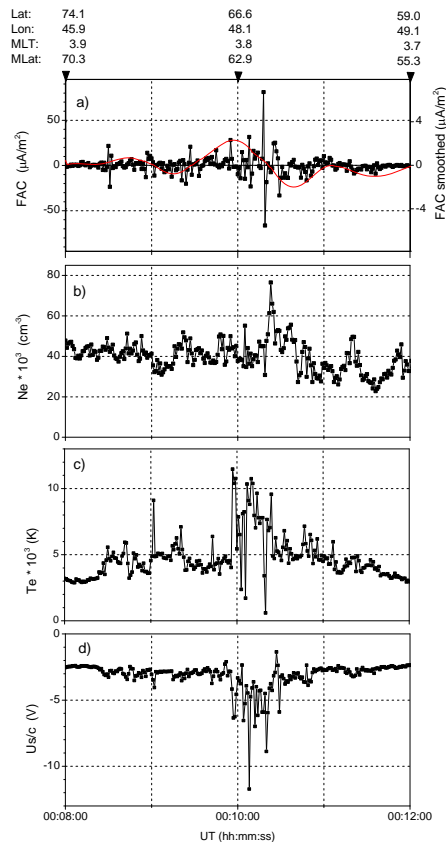


Рисунок 1.1.3.3.2 - 1-Гц значения (а) плотности ПТ, (b) Ne , (c) Te и Us/c

1.2 Валидация модели распределения электронной концентрации в F-области полярной ионосферы по данным радара ESR

1.2.1 Постановка задачи

Для сравнения результатов моделирования, полученных с помощью численной модели полярной ионосферы F-области, с экспериментальными данными был выбран период солнечного минимума и длительного геомагнитного спокойствия ($K_p < 1$), связанного с низкими значениями межпланетного магнитного поля ($B_T < 5 \text{ нТл}$). Можно считать, что эти условия соответствуют «базовому» состоянию ионосферы, когда роль магнитосферы в формировании полярной ионосферы стабильна и не нарушается спорадическими возмущениями. Задача состояла в определении сезонных закономерностей в распределении ионосферной плазмы для марта, июля 2007 года и января 2008 года. Эти периоды соответствуют обширным наблюдательным кампаниям, проведенным в Международный полярный год (МПГ). Используя возможности усреднения большого набора данных наблюдений радара некогерентного рассеяния EISCAT Svalbard (ESR), который расположен вблизи границы полярной шапки и почти непрерывно работал в течение МПГ, модельные результаты были проверены на соответствие радарным измерениям электронной концентрации в различные сезоны года. Также результаты моделирования сравнивались с данными ионозонда Геофизической обсерватории Соданкюля, которая расположена вблизи экваториальной границы аврального овала.

1.2.2 Результаты

Для проведения сравнительного анализа были выбраны месячные отрезки времени: март и июль 2007 г. и январь 2008 г., и далее из них были выбраны наиболее спокойные периоды, как показано на рисунке 1.2.2.1.

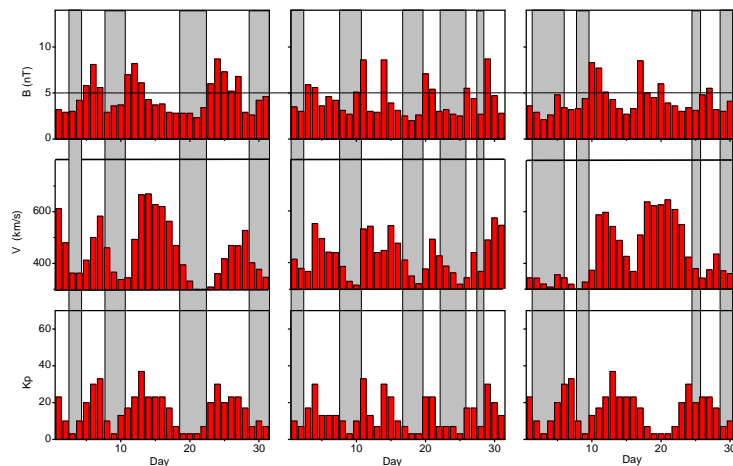


Рисунок 1.2.2.1 - Среднесуточные значения общей интенсивности ММП, скорости солнечного ветра и индекса Кр за (а) март, (б) июль 2007 г. и (с) январь 2008 г. Спокойные периоды выделены серым цветом.

Карты изолиний максимальных значений электронной концентрации ($NmF2$) в системе координат MLT-MLat представлены на рисунке 1.2.2.2 для трех рассмотренных сезонов. Из этого рисунка можно выделить несколько крупномасштабных неровностей F-области и проследить его сезонную эволюцию.

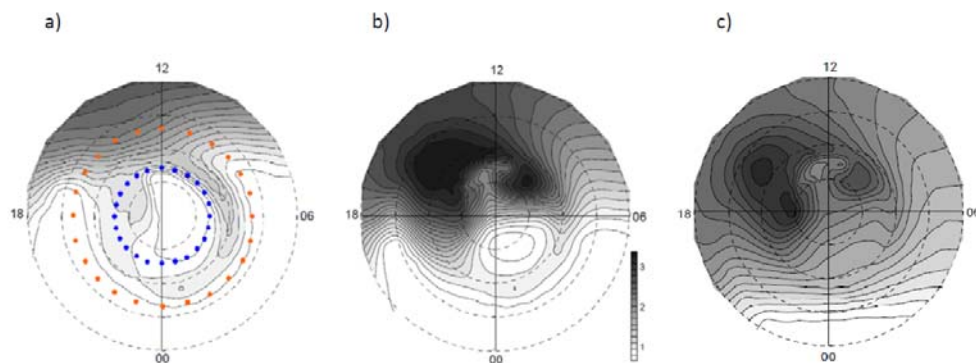


Рисунок 1.2.2.2 - Изолинии $NmF2$ (в 10^{11} м^{-3} , серая шкала является общей для всех трех графиков): (а) январь 2008 г., (б) март и (с) июль 2007 года. Расположение радара ESR и ионозонда в каждый час MLT отмечены синими и оранжевыми точками соответственно.

После выполнения серии модельных расчетов с часовым интервалом для каждого дня для периодов, представленных на рисунке 1.2.2.1, были собраны сезонные композиты для каждого часа MLT для широты, на которой расположен ESR радар. Такие же композиты были собраны для радарных данных. Двумерные распределения параметра $\log(Ne)$ (в координатах «высота-MLT») для трех различных сезонов показаны на рисунке 1.2.2.3. Верхний и нижний графики показывают наблюдаемые и смоделированные композиты $\log(Ne)$, соответственно.

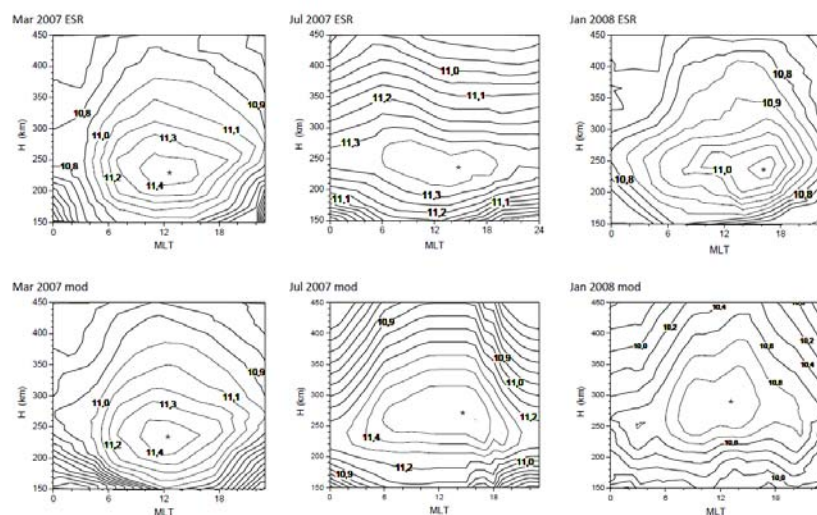


Рисунок 1.2.2.3 - Верхний и нижний ряды графиков показывают, соответственно, наблюдаемые и смоделированные композиты $\log(N_e)$ для марта, июля и января.

В спокойный период со стабильным уровнем солнечной активности распределения N_e демонстрируют только сезонную тенденцию. В полярной ионосфере N_e имеет глубокий минимум зимой, однако летом не достигает максимума, как можно было бы ожидать от фотохимических процессов; вместо этого наблюдаемый и смоделированный N_e имеют тенденцию быть самыми большими весной. Наблюдаемый (модельный) пик $\log(N_e)$ контур равен 11,4 (11,6), 11,4 (11,5), 11,0 (11,0) для равноденствия, летнего и зимнего месяца, соответственно, т.е. реально наблюдаемые значения регулярно несколько более низкие, чем модельные. Это может объясняться охлаждением и оседанием верхней атмосферы вследствие увеличения концентрации парниковых газов в реальной атмосфере.

Заключение

На основе анализа больших массивов данных измерений спутников *SWARM* получены следующие результаты. Плотность продольных токов в период бури достигает 3 мкА/м^2 , что более чем на два порядка превышает плотность в спокойных условиях. На дневной стороне интенсивность ПТ изменяется в соответствии с индексом кольцевого тока. На ночной стороне интенсивность повышается на порядок на пике суббури. Экваториальная граница ПТ следует изменению индекса кольцевого тока, суббури увеличивают смещение области, занятой током, на еще более низкие широты. В максимуме бури граница смешается до $49\text{-}50^\circ$ магнитной широты. В полуденном секторе преобладают биполярные нитевидные структуры (ПТ шириной $7,5 \text{ км}$ с противоположными полярностями, соседствующие друг с другом). Наиболее интенсивные мелкомасштабные ПТ, до 100 мкА/м^2 , наблюдаются только в послеполуночном секторе.

В качестве очередного шага в развитии численной модели F области полярной ионосферы была проведена валидация модели по экспериментальным данным. Модель рассчитывает 3D распределение электронной концентрации (N_e) и воспроизводит все основные крупномасштабные неоднородности F области, эволюция которых контролируется магнитосферной динамикой (электрическим полем магнитосферной конвекции и высыпаниями авроральных частиц) и солнечным зенитным углом, а также зависит от параметров нейтральной компоненты термосферы. Для сравнения и оценки точности результатов моделирования были отобраны данные измерений вертикальных профилей N_e с помощью высокоширотного Европейского радара некогерентного

рассеяния EISCAT, расположенного на Шпицбергене (ESR, 78.2°N). Сравнительный анализ показал, что до высот ~300 км вертикальные профили N_e , рассчитанные по модели и полученные по радарным данным, совпадают с точностью до 10%. Модель корректно воспроизводит форму суточной и сезонной вариаций максимума (N_mF2) и его высоты (h_mF2) – двух наиболее репрезентативных параметров F области. При этом наблюдается систематическое превышение (на 3-8%) модельных значений h_mF2 над измеренными. Причиной расхождений может быть охлаждение и оседание верхней атмосферы вследствие увеличения концентрации парниковых газов в реальной атмосфере.

Материалы, изложенные в данной главе, позволяют считать, что все работы по данному направлению выполнены в полном объеме. Публикации первого года приведены в списке использованных источников. На следующий год запланировано внедрение рассмотренной модели полярной ионосферы в систему «Аврора-Арктика».

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

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. В настоящей главе отчета представлены основные результаты работ по данному направлению, полученные по следующим разделам плана работ ИКИ РАН на 2019 год:

- Исследование и разработка методов машинного обучения и глубокого обучения для интеллектуального анализа больших массивов астрономических данных;
- Исследование и разработка горизонтально-масштабируемых алгоритмов обработки и анализа больших данных в астрофизике.

Результаты, полученные по данным направлениям, описаны в разделах 2.1 и 2.2 настоящей главы. Подробно полученные в данном направлении результаты изложены в следующих научных публикациях [4,21], а также представлены в рамках устного доклада на международной конференции "HEA-2019" [22].

2.1 Разработка методов машинного обучения и глубокого обучения для обработки и анализа больших массивов астрономических данных

2.1.1 Постановка задачи

13 июля 2019 года была запущена российская национальная космическая обсерватория «Спектр-РГ» («Спектр-Рентген-Гамма») с двумя рентгеновскими телескопами на борту (eROSITA и ART-XC). Главной задачей «Спектр-РГ» является 4-х летний рентгеновский обзор всего неба с беспрецедентной чувствительностью. 8 декабря 2019 года обсерватория начала сканирование неба в режиме обзора. Алгоритмы обработки данных, в частности, методы детектирования рентгеновских объектов с основного телескопа eROSITA, ключевым образом влияют на качество научных результатов, которые будут получены в ходе космической миссии. В нашей работе, проведенной в 2019 году, впервые исследуется применение нейросетевых алгоритмов для детектирования астрономических объектов на рентгеновских изображениях и делается сравнение со стандартными методами поиска объектов в составе программного комплекса eSASS рентгеновского телескопа eRosita. Основной задачей данной работы являлось:

- Исследование возможности применения сверточных нейронных сетей при детектировании источников на астрономических изображениях в рентгеновском диапазоне.

Процесс детектирования рентгеновских источников в eSASS состоит из следующих этапов (см. Рисунок 2.1.1.1, верхняя панель): (i) сегментирование областей с рентгеновскими объектами, (ii) оценка фоновой компоненты на изображении, (iii) повторное детектирование источников и измерение их характеристик (положение, яркость, размер) на основе модели максимального правдоподобия с учетом фоновой компоненты.

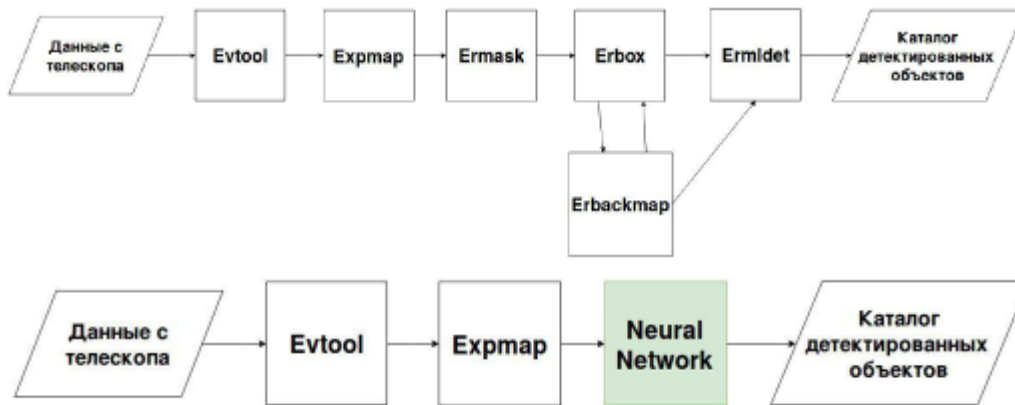


Рисунок 2.1.1.1 - Схема работы конвейера eSASS (верхняя панель) и схема предлагаемой модификации eSASS на базе сверточной нейронной сети (нижняя панель)

В рамках данной работы мы заменили один из этапов цепочки обработки данных eSASS, а именно блок сегментации источников (erbox), на обученную сверточную нейронную сеть с модифицированной архитектурой на базе нейросетевой архитектуры U-net. Результатом работы модели U-net является карта сегментации с вероятностью принадлежности астрономическому источнику в каждом пикселе. Архитектура U-net была нами адаптирована под астрономические данные телескопа eRosita, в частности, были внесены изменения в размерность входного слоя, увеличено количество сверточных слоев сети, применены взвешенная функция ошибок и современные техники глубокого обучения (Dropout и Batch Normalization) для регуляризации и улучшения сходимости при обучении нейросетевой модели. В качестве итогового результата работы конвейера детектирования рентгеновских источников нам необходимо получить каталог астрономических объектов. Нами был разработан алгоритм постобработки, который на основе карты сегментации создает астрономический каталог с информацией об источниках. Обучение и тестирование модели проводилось на астрономических симуляциях данных телескопа eRosita, подготовленных сотрудниками ИКИ РАН.

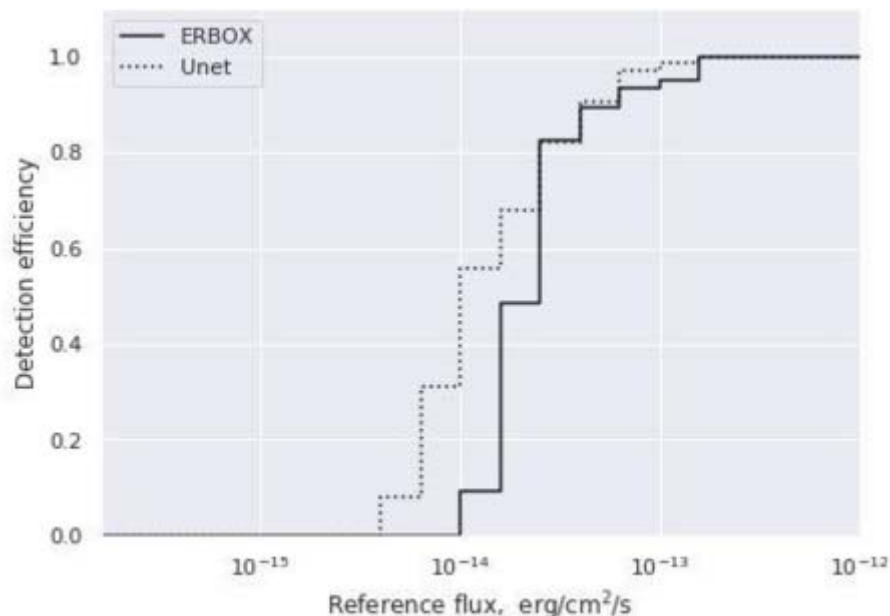


Рисунок 2.1.1.2 - Эффективность детектирования в зависимости от потока в диапазоне 0.5-2 кэВ от рентгеновского источника в симуляции данных eRosita

Эксперименты на данных компьютерных симуляций телескопа eRosita показали, что предложенная нейросетевая модель по качеству сегментации источников значительно превосходит стандартный алгоритм erbox eSASS, работающий по принципу скользящего

окна. Предложенный нами нейросетевой алгоритм находит в 5 раз больше объектов на потоке $1e-14$ эрг/с/см², что соответствует ожидаемой чувствительности в 4-х летнем обзоре eRosita в диапазоне 0.5-2 кэВ (см. Рисунок 2.1.1.2), не увеличивая при этом долю ложных детектирований.

В дальнейшем, мы планируем совершенствовать предложенные нейросетевые модели. Главной целью является замена всей цепочки детектирования астрономических объектов на рентгеновских изображениях на нейросетевую "end-to-end" архитектуру (см. Рисунок 2.1.1.1, нижняя панель), что позволит в дальнейшем значительно улучшить эффективность обработки данных национальной обсерватории «Спектр-РГ».

2.1.2 Разработка алгоритмов машинного обучения для кросс-отождествления и классификации объектов в фотометрических каталогах обзоров неба в различных спектральных диапазонах, измерения фотометрических красных смещений внегалактических объектов

В рамках темы Космос-Д в 2019 году была начата работа над конвейером оптических отождествлений SRGz (см. Рисунок 2.1.2.1), который призван решать задачи кросс-отождествления рентгеновских источников, обнаруженных телескопами eRosita и ART-XC в фотометрических каталогах современных обзоров неба (SDSS, PanSTARPS, Legacy Surveys, WISE и др.), а также классификации объектов и измерения красных смещений (photo-z) внегалактических источников по фотометрическим данным на базе современных алгоритмов машинного обучения (случайный лес, градиентный бустинг, глубокие нейросетевые модели).

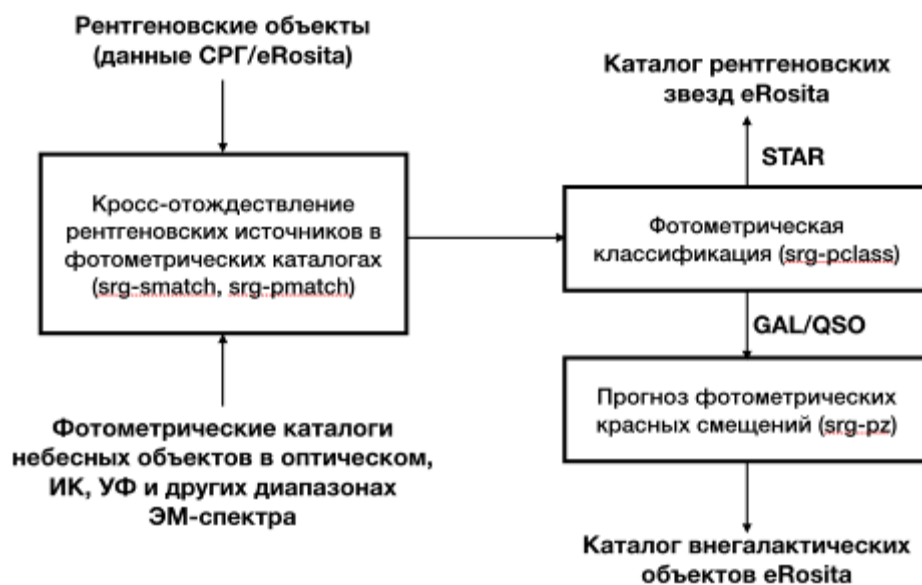


Рисунок 2.1.2.1 - Предлагаемая схема работы конвейера оптических отождествлений SRGz для интеллектуального анализа данных телескопа SRG/eRosita

К концу 2019 года в пакете srg-pz в рамках общей архитектуры конвейера оптических отождествлений SRGz были реализованы методы вероятностных прогнозов фотометрических красных смещений на основе ансамблей деревьев решений и разнообразных метрик оценки прогнозов photo-z различных типов: точечных прогнозов, прогнозов доверительных интервалов, прогнозов полных распределений. Архитектура SRGz и результаты измерений фотометрических красных смещений рентгеновских квазаров в тестовых полях (XMM-XXL-N, Stripe82X, Lockman Hole) были представлены А.В.Мещеряковым в рамках устного доклада на конференции Астрофизика высоких энергий сегодня и завтра - 2019.

В дальнейшем, мы планируем продолжить исследование и разработку алгоритмов машинного обучения для системы оптических отождествлений рентгеновских источников миссии Спектр-РГ. В частности, планируется разработка методов оптического отождествления рентгеновских источников eRosita на основе пространственной и фотометрической информации из обзоров неба, разработка методов классификации рентгеновских источников и дальнейшее развитие алгоритмов прогнозирования красных смещений рентгеновских объектов.

2.2 Исследование и разработка горизонтально-масштабируемых алгоритмов обработки и анализа больших астрономических данных

Еще одним важным направлением исследований в рамках темы «Космос-Д» в 2019 году стала разработка горизонтально-масштабируемого алгоритма для выполнения астрометрической калибровки массива астрономических изображений в реальном масштабе времени на базе фреймворка Apache Spark.

Объемы архивных данных, полученных в рамках цифровых обзоров неба в различных спектральных диапазонах, постоянно растут и в настоящее время составляют петабайты изображений, хранящихся в общедоступных астрономических архивах (таких как IRSA, SDSS Science Archive, MAST, HEASARC и др.). Для использования архивных изображений в современных астрофизических исследованиях часто требуется уточнить астрометрическую калибровку для интересующих изображений, хранящихся в архиве. При крупномасштабных исследованиях больших выборок объектов на небе, количество изображений, которые потребуется проанализировать, может исчисляться миллионами и занимать существенный объем в сотни терабайт. В качестве примера можно привести задачу исследования переменности рентгеновских объектов из обзора неба СРГ/eRosita в архивных данных спутника WISE. Для подобного анализа данных потребуется выполнить точную астрометрическую калибровку миллионов ИК-изображений, и в процессе обработки решать задачу хранения больших массивов изображений и их распределенной обработки на вычислительном кластере.

В 2019 году нами был разработан и протестирован горизонтально-масштабируемый астрономический конвейер астрометрической калибровки больших массивов изображений с современных космических и наземных телескопов. При его разработке была использована технология распределенного хранения, обработки и интеллектуального анализа больших массивов астрономических данных, ранее запатентованная нашей группой (патент на изобретение №2659493 "Способ формирования каталога небесных объектов из больших массивов астрономических изображений" от 02.07.2018).

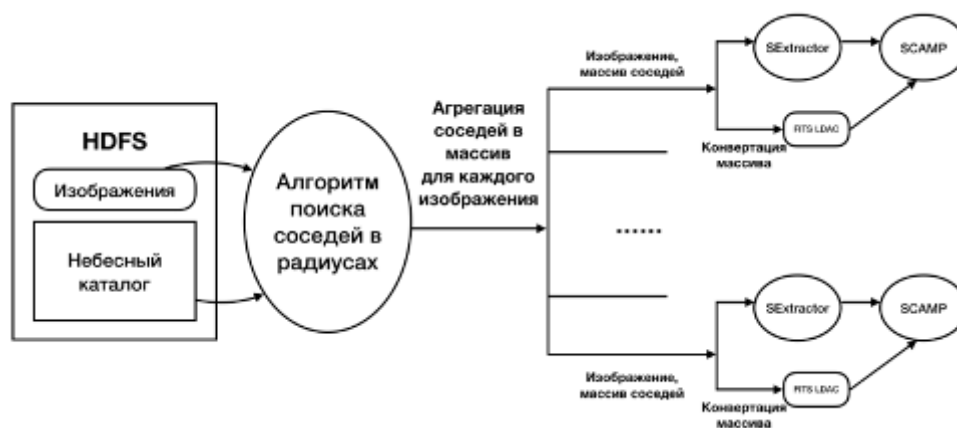


Рисунок 2.2.1 - Схема работы программного средства для астрометрической калибровки большого объема астрономических изображений

В основе архитектуры конвейера астрономической калибровки лежат современные принципы работы с большими массивами данных: модель вычислений MapReduce, каппа-архитектура, распределенная файловая система HDFS. Они обеспечивают горизонтальную масштабируемость, надежность хранения, эффективную утилизацию вычислительных ресурсов, унификацию пакетной и потоковой обработки в реальном времени. В качестве сквозных базовых технологий использованы технологии Apache Spark, HDFS, YARN. Программное средство для астрометрической калибровки изображений было реализовано нами в Apache Spark на языке Scala. В алгоритме поиска ближайших соседей в астрономических каталогах, являющемся частью программного решения, использовалась библиотека иерархической индексации HEALPIX. Логика астрономической калибровки изображений была реализована при помощи стандартных астрономических утилит из стека Astromatic. Схема разработанного программного средства представлена на Рисунке 2.2.1.

Астрономический конвейер астрометрической калибровки был протестирован в ходе серии экспериментов на Spark-кластере в облачном сервисе Microsoft Azure. В качестве калибровочного каталога звезд был выбран Gaia DR2, содержащий более 1,7 миллиарда звезд. Для хранения астрономических каталогов и изображений использовалось хранилище Azure Blob Storage, которое имитирует работу HDFS. Мы протестировали корректность и горизонтальную масштабируемость предложенного распределенного алгоритма астрометрической калибровки на одном из базовых сценариев, возникающих при много-волновом анализе данных миссии Спектр-РГ. В ходе эксперимента на кластере Microsoft Azure в составе 7 вычислительных узлов D13 v2 был продемонстрирован линейный рост скорости астрометрической калибровки астрометрических изображений (см. Рисунок 2.2.2) с увеличением количества процессов исполнителей в кластере.



Рисунок 2.2.2 - Скорость астрометрической калибровки изображений в зависимости от количества процессов исполнителей на кластере

В дальнейшем, мы планируем продолжить работы по исследованию алгоритмов распределенной обработки больших данных астрономических изображений в рамках темы «Космос-Д». Конечной целью здесь является построение универсальной системы

распределенной обработки и интеллектуального анализа больших массивов астрономических данных. Создание данной системы представляется особенно важным в контексте задач много-волнового анализа данных рентгеновского обзора всего неба национальной космической обсерватории Спектр-РГ.

Заключение

Представленные в данной главе материалы показывают, что работы по направлению «Космос-Д-Астрофизика» проводились в соответствии с утвержденным планом. Из наиболее значимых результатов, полученных по данному направлению, следует отметить следующие:

- Разработка горизонтально-масштабируемого астрономического конвейера для астрометрической калибровки больших массивов изображений с современных космических и наземных телескопов на базе фреймворка Apache Spark.
- Разработка алгоритмов измерения фотометрических красных смещений рентгеновских источников для космической обсерватории Спектр-РГ в рамках конвейера оптических отождествлений SRGz.
- Разработка нейросетевых алгоритмов для детектирования астрономических объектов на рентгеновских изображениях, их сравнение со стандартными методами поиска рентгеновских источников (на базе программного комплекса eSASS рентгеновского телескопа eRosita).

Материалы, изложенные в данной главе, позволяют считать, что все работы по направлению «Космос-Д-Астрофизика» выполнены в полном объеме.

ГЛАВА 3 Разработка новых методов и подходов организации хранения и параллельной обработки сверхбольших объемов данных для решения задач анализа долговременных наблюдений в интересах ведения мониторинга и изучения различных процессов и явлений

Введение

Настоящая глава посвящена описанию направления работ по разработке новых методов и подходов к организации хранения и параллельной обработки сверхбольших объемов данных для решения задач анализа долговременных наблюдений в интересах ведения мониторинга и изучения различных процессов и явлений. Полученные в рамках работ 2019 года результаты по данному направлению также изложены в следующих публикациях [1-2, 5, 6-20, 23].

Пожалуй, одной из основных проблем, которая возникает в настоящее время при организации работы с данными долговременных много-спутниковых наблюдений при решении различных задач, связанных с изучением и мониторингом различных процессов и явлений, является то, что в таких работах приходится иметь дело с достаточно разнородными архивами. Это приводит к существенному усложнению процедур совместного анализа и обработки различных типов данных. В первую очередь, эта разнородность вызвана объективными причинами, так как наиболее эффективные варианты организации хранения различных типов данных, как правило, существенно отличаются. Кроме того существуют и субъективные причины, связанные с различием подходов, используемых разными группами разработчиков для решения схожих задач по организации хранения данных.

Описанная выше разнородность спутниковых данных существенно усложняет разработку процедур обработки новых типов спутниковых данных, но в еще большей степени затрудняет совместный анализ и обработку различных типов данных. Так как пользователей, как правило, интересуют не только данные по отдельным спутниковым сценам, а также и агрегированная информация по заданной территории за указанные интервалы времени, возникает необходимость в получении различных композитных информационных продуктов. При этом отличия схем хранения различных типов данных не позволяют в полной мере использовать унифицированные процедуры обработки данных. В тех случаях, когда пользователей интересуют данные по значительным территориям, эффективное получение композитных информационных продуктов практически нереализуемо без использования распределенной параллельной обработки данных на кластере серверов обработки, но использование для этой цели исходных данных по сценам с нерегулярным пространственным покрытием существенно затрудняет разработку оптимальных по скорости выполнения процедур обработки.

В то же время при организации обработки и анализа данных (в особенности долговременных серий наблюдений) для большинства задач может быть успешно применен подход, позволяющий использовать регулярные пространственные разбиения различных типов данных на блоки информации (тайлы). Такой подход существенно упрощает задачи по их совместной обработке и анализу и при этом позволяет эффективно использовать механизмы распараллеливания процессов обработки и кеширования данных по тайлам данных при проведении их многократной обработки. Возможным вариантом реализации этого подхода является организация архивов на основе соответствующей блочной структуры. Однако такой подход имеет свои недостатки, которые в первую очередь связаны с потерей гибкости при необходимости организации работы с данными в различных проекциях, искажениями, которые неизбежно вносятся в исходные данные при подготовке их к блочному хранению и значительными вычислительными расходами при их ведении (наполнении архивов данных). Кроме этого в ряде случаев не представляется

возможным (разумным по временным и вычислительным затратам) проведение переобработки уже накопленных сверхбольших архивов спутниковых данных.

Поэтому в рамках данной тематики в 2019 году были начаты работы по разработке новых методов и подходов организации блочного доступа к сверхбольшим распределенным архивам спутниковых данных, основанные на предоставлении доступа к «виртуальным» блокам данных, которые формируются, когда в этом возникает необходимость. Такие подходы в перспективе должны:

- обеспечить все преимущества, связанные с блочной организацией архивов данных;
- исключить необходимость переобработки уже имеющихся сверхбольших архивов данных;
- за счет динамического формирования блоков обеспечить возможность оптимального выбора разбиения на блоки для решения конкретных задач обработки, и тем самым повысить эффективность и гибкость процедур обработки данных;
- сохранить возможность использования наиболее эффективных для конкретных типов данных схем организации их хранения.

Настоящая глава отчета посвящена описанию результатов работ, проведенных в 2019 году, и направленных на создание унифицированной технологии организации блочного доступа к данным в архивах для проведения их эффективной распределенной параллельной обработки (раздел 3.1).

В главе также представлены результаты работ по использованию подходов, созданных для работы с данными наблюдений Земли, для решения и других во многом схожих задач. На примере проведенных работ по созданию системы Ages-Mars (раздел 3.2) и «Аврора-Арктика» (раздел 3.3) показана возможность использования данных подходов для организации работы с данными планетных миссий и исследований космической погоды.

3.1. Разработка подходов к созданию унифицированной технологии организации блочного доступа к данным в архивах для проведения их эффективной распределенной параллельной обработки

3.1.1 Актуальность задачи

Как правило, исходные спутниковые данные представляют собой информацию по пролетам, т.е. по виткам, которые совершает спутник, которые в случае прямого приема ограничены зоной видимости станции приема. Исключение составляют данные геостационарных спутников, которые постоянно находятся над одной и той же точкой поверхности Земли и поэтому всегда снимают одну и ту же территорию. Такая информация, особенно в случае данных высокого разрешения, имеет очень большой объем, поэтому для удобства использования данные делятся на сравнительно небольшие фрагменты. При этом способ этого деления зависит от реализации конкретной спутниковой системы и системы их архивации. В результате, архивы исходных спутниковых данных содержат сравнительно небольшие сцены различных спутниковых приборов наблюдения, имеющих сильно отличающееся пространственное разбиение.

В то же время, для решения многих задач мониторинга окружающей среды и антропогенных объектов пользователи заинтересованы в получении данных по значительным территориям, размер которых может радикально превышать размеры отдельных сцен. Возникает необходимость в построении композитных информационных продуктов по заданной территории за требуемые интервалы времени, например еженедельные, ежемесячные или годовые композиты. Важно отметить, что для получения таких информационных продуктов, как правило, требуется обработка очень большого количества сцен, особенно, когда интерес представляют только безоблачные данные или

когда композит за заданный интервал времени формируется на основе интерполяции, с использованием данных за предшествующий и последующий временной интервал.

Наиболее эффективным, а зачастую, и единственно возможным подходом для получения композитных информационных продуктов по значительным территориям (например, по территории России) является проведение распределенной параллельной обработки, когда каждый из вычислительных узлов оперирует с данными по одной из ячеек (гранул) пространственного разбиения региона интереса. Возникает задача эффективного предоставления узлу обработки данных по той или иной ячейке разбиения, при этом даже в рамках одной процедуры обработки могут использоваться одновременно данные разных приборов наблюдения. Важно отметить, что получение композитных информационных продуктов в общем случае является довольно-таки затратным в плане использования как вычислительных, так и дисковых ресурсов программно-аппаратного комплекса.

В результате, остро встает вопрос минимизации накладных расходов на чтение исходных спутниковых данных и передачу их по сети. При этом наиболее целесообразным на наш взгляд вариантом является использование прямого блочного доступа к файлам данных, позволяющим читать и передавать по сети только ту часть информации, которая реально необходима для проведения конкретной операции по их обработке по заданной ячейке разбиения. Важно отметить, что использование общепринятого для хранения спутниковых данных формата GeoTIFF позволяет эффективно реализовать такой механизм доступа к файлам и получать данные только по заданному региону и требуемому набору каналов (bands). Немаловажным достоинством этого формата в данном случае также является возможность работы со сжатыми данными. Кроме непосредственного формирования требуемых данных по ячейкам разбиения для эффективной реализации описанных выше процедур обработки необходимы также программные компоненты, обеспечивающие запрос требуемых метаданных, формирование заданий на получение данных. Необходима также выработка оптимальной схемы проведения такого рода процедур обработки, позволяющей эффективно использовать имеющиеся дисковые и вычислительные ресурсы.

В последние десятилетия наблюдается стремительное развитие спутниковых систем ДЗЗ, что, в частности, приводит к росту числа различных типов спутниковых данных, которые могут быть использованы для получения тех или иных информационных продуктов, в том числе, и композитных. Если до недавнего времени предварительная подготовка исходных спутниковых данных была неотъемлемой частью самой процедуры их обработки, то теперь более целесообразным является реализация их в рамках отдельных унифицированных процедур. Представленный выше подход позволяет не только повысить эффективность проведения распределенной параллельной обработки спутниковых данных, но и упростить задачу реализации новых алгоритмов обработки различных типов спутниковых данных.

3.1.2 Унифицированная технология организации блочного доступа к данным в архивах для проведения их эффективной распределенной параллельной обработки

На основе вышеописанного подхода в отделе «Технологии спутникового мониторинга» ИКИ РАН в 2019 году разработан базовый вариант технологии, обеспечивающей эффективное предоставление спутниковых данных для проведения их распределенной параллельной обработки. Основными требованиями при ее разработке были максимальная унификация и оптимизация всех процедур доступа к данным, что позволяет использовать ее для решения самых разных задач по обработке различных типов спутниковых данных, отдельные сцены или сеансы которых могут состоять из различного числа фрагментов. Важно отметить, что представленный подход может быть реализован только для архивов, поддерживающих прямой блочный доступ к файлам данных, в частности, по протоколу NFS, применяемому для этих целей нами.

Ниже представлены два основных варианта реализации подсистемы блочного доступа к файлам данным для проведения их распределенной параллельной обработки, реализованных в рамках рассматриваемой технологии. При этом целесообразность выбора одного из них в каждом конкретном случае зависит от специфики реализации систем хранения и обработки спутниковых данных. Также приводятся общие сведения о текущей реализации описываемой технологии и рассматриваются планы по ее развитию. Затем рассматриваются перспективы использования нового подхода для эффективного формирования спутниковых изображений, предоставляемых пользователям в картографических WEB-интерфейсах.

3.1.3 Вариант с использованием прямого доступа к файлам данных со стороны серверов обработки

Рассматриваемый вариант организации подсистемы блочного доступа к файлам данным применим в тех случаях, когда вычислительные узлы имеют прямой доступ к файлам данным в архивах спутниковых данных. В частности, нами прорабатывалась конфигурация, когда сервера обработки функционируют под операционной системой UNIX, а прямой доступ к файлам данным в архивах с их стороны обеспечивается по протоколу NFS.

Принципиальная схема построения подсистемы блочного доступа к файлам для проведения их распределенной параллельной обработки при использовании рассматриваемого варианта приведена на рисунке 3.1.3.1. Сервера диспетчеризации обработки отвечают за формирование заданий на проведение обработки по каждой из ячеек разбиения и предоставление их серверам обработки. Универсальная процедура формирования заданий на получение наборов исходных данных использует гибкую систему настроек, позволяющую задать географическую область, параметры ее разбиения на ячейки и наборы различных типов спутниковых данных, необходимых для проведения обработки, и географическую проекцию, в которой они должны быть предоставлены. Если требуемые файлы данных являются многоканальными, то также указываются только те из них, которые реально нужны для проведения обработки. Для каждой из ячеек разбиения производится поиск необходимых данных в указанных архивах и формируется задание, содержащее NFS-пути для доступа ко всем выбранным файлам данным, а также необходимые для их обработки метаданные. Важно отметить, что для реализации такого варианта доступа к файлам данным на каждом из серверов обработки единообразным образом должны быть примонтированы по NFS все разделы хранения данных. Распределение сформированных таким образом заданий по серверам обработки производится на базе подсистемы управления обработкой, рассмотрение которой выходит за рамки настоящего изложения.



Рисунок 3.1.3.1. Схема организации блочного доступа к данным с использованием прямого доступа к данным архивов со стороны серверов обработки.

На всех серверах обработки устанавливается универсальный программный модуль для подготовки исходных данных для обработки, который для повышения эффективности реализуется в многопоточном режиме. В результате его выполнения по каждому из спутниковых сеансов (пролетов) формируется GeoTIFF файл в заданной географической проекции. При этом его экстенд соответствует границам текущей ячейки разбиения или же несколько большей области, если это требуется для проведения обработки. Отметим, что в текущей реализации все операции по обработке спутниковых изображений реализованы на основе использования свободно распространяемого ПО GDAL.

Хотя рассмотренный вариант относительно прост в реализации, он обладает рядом недостатков. Во-первых, обеспечение прямого доступа к файлам данных со стороны серверов обработки связано с необходимостью монтирования на каждом из серверов всех разделов хранения данных, а в некоторых реальных конфигурациях может быть просто невозможным. В частности, такой вариант не подходит, когда мы имеем дело с распределенными архивами спутниковых данных, находящимися в различных информационных центрах. Во-вторых, такой вариант не позволяет использовать для предварительной подготовки данных вычислительные ресурсы самих серверов хранения данных, в том числе нет возможности обрабатывать данные непосредственно по месту их расположения в тех случаях, когда это возможно.

3.1.4 Вариант с использованием серверов предварительной подготовки необходимых наборов исходных спутниковых данных

Наиболее перспективным вариантом организации подсистемы предоставления данных для проведения их распределенной параллельной обработки, на наш взгляд, является подготовка требуемых наборов исходных данных на серверах предварительной обработки данных. В качестве таких серверов могут быть использованы сами сервера хранения данных, а также специально выделенные для этой задачи сервера, имеющие высокую скорость доступа к файлам данных в архивах. На каждом из этих серверов

реализуется универсальный сервис для формирования исходных наборов спутниковых данных, функционирующий по протоколу HTTP. Реализация его аналогична реализации описанного выше модуля для подготовки наборов данных по спутниковым сеансам и также основана на использовании ПО GDAL.

Принципиальная схема построения подсистемы блочного доступа к файлам для проведения их распределенной параллельной обработки при использовании рассматриваемого варианта приведена на рисунке 3.1.4.1. Для формирования заданий на проведение обработки используется универсальная процедура, выполняемая на серверах диспетчеризации. Но в этом случае, формируемые на серверах диспетчеризации обработки задания содержат не пути к файлам, а HTTP-запросы на получение готовых наборов данных по спутниковым сеансам. С целью равномерного распределения нагрузки по имеющимся серверам хранения и выделенным для предварительной обработки данных серверам в данной схеме используется сервер диспетчеризации запросов на получение наборов исходных данных, который перенаправляет (операция redirect) запросы на один из серверов подготовки данных. При этом реализуется алгоритм балансировки, позволяющий в приоритетном режиме направлять запросы непосредственно на те сервера, на которых непосредственно находятся исходные файлы данных, что позволяет повысить эффективность выполнения предварительной обработки данных и уменьшает сетевой трафик. Если требуемые для обработки исходные файлы находятся в локальной сети, то задания содержат запросы к локальным серверам диспетчеризации запросов. Если же эти файлы находятся в другом информационном центре, то они адресуются на расположенный в этом центре сервер балансировки с использованием специальной схемы проксирования запросов, в текущей реализации использующей возможности ПО nginx.



Рисунок 3.1.4.1. - Схема организации блочного доступа к данным с использованием серверов предварительной подготовки необходимых наборов спутниковых данных.

На серверах обработки однотипно устанавливается универсальный программный модуль для многопоточкового получения исходных наборов данных по сеансам, в текущей реализации построенный на базе использования свободно распространяемого ПО Aria2c.

Он предназначен для эффективного скачивания исходных наборов данных по спутниковым сеансам путем проведения запросов к локальным или внешним серверам диспетчеризации запросов на получение данных.

Благодаря использованию универсального сервиса подготовки исходных наборов спутниковых данных представленный вариант организации доступа к данным для проведения их обработки позволяет максимально упростить разработку и внедрение новых алгоритмов обработки различных типов спутниковых данных.

3.1.5 Общие сведения о текущей программной реализации. Планы ее доработки и внедрения

В 2019 году был реализован действующий макет подсистемы блочного доступа к данным для проведения их эффективной распределенной параллельной обработки. При его реализации использовалась инфраструктура и технологические решения ЦКП «ИКИ-Мониторинг», в первую очередь, связанные с обеспечением доступа к архивам спутниковых данных и проведением их распределенной параллельной обработки. Архивы ЦКП построены на базе использования разработанной в ИКИ РАН технологии UNISAT, предназначенной для построения сверхбольших распределенных архивов спутниковых данных. Взаимодействие с подсистемой обработки спутниковых данных реализовано с использованием технологии организации распределенной многопоточковой обработки спутниковых данных, также разработанной в ИКИ РАН. Ниже кратко описаны реализованные программные компоненты, а также ближайшие планы по их доработке и оптимизации.

Универсальная процедура формирования заданий на получение исходных наборов данных реализована на языке программирования Perl с использованием пакета библиотек SDB (Satellite Data Base), разработанного в ИКИ РАН. В текущей реализации параметры запусков процедуры определяются двумя типами текстовых конфигурационных файлов. Файл grids.ini описывает различные варианты регулярных разбиений и, в частности, содержит шаг сетки в градусах, диапазон координат и степень перекрытия соседних ячеек сетки разбиения. Для каждой обработки также формируется конфигурационный файл, содержащий информацию об используемом при его построении разбиении, полигон, по которому будет производиться обработка спутниковых данных, а также описания множества источников исходных данных, позволяющие задать большое число различных критериев поиска необходимых данных в архивах, а также информацию о том, какие именно каналы исходных файлов данных должны быть предоставлены для обработки. Хотя скорость выполнения процедуры по результатам тестирования является вполне приемлемой, в перспективе планируется модернизация отдельных компонент реализации технологии UNISAT, что позволит заметно ускорить и упростить задачу поиска данных по спутниковым сеансам. В результате выполнения утилиты формируется набор поддиректорий, соответствующих отдельным ячейкам разбиения, в каждой из которых создаются файлы, содержащие ссылки на получения необходимых наборов исходных данных в формате программы Aria2c, а также файлы с метаданными для каждого из источников исходных данных в формате JSON.

Универсальный программный модуль для многопоточкового скачивания исходных наборов данных также реализован на языке программирования Perl и при этом представляет собой относительно простую процедуру, позволяющую произвести запуск ПО Aria2c и убедиться в успешности его выполнения.

Сервис для оптимальной переадресации запросов на получение исходных наборов спутниковых данных в текущей реализации также реализован на языке программирования Perl. Он позволяет равномерно распределить нагрузку на сервера предварительной подготовки данных на основе модифицированного метода «Least connection», позволяющего в приоритетном режиме направлять запросы на те сервера, на которых непосредственно находятся исходные данные. Специальные параметры конфигурации позволяют задать степень приоритетности выполнения обработки на этих серверах, а

также выставить серверам веса, соответствующие уровню их производительности. Тестирование показало, что производительность и надежность реализованного сервиса находятся на вполне приемлемом уровне, однако в перспективе рассматриваются варианты более эффективной и потенциально более надежной его реализации с использованием языка программирования C++ или Python.

Унифицированный сервис формирования наборов данных по спутниковым сеансам в рамках описываемого действующего макета установлен на пяти различных серверах хранения спутниковых данных. В текущей реализации операции по обработке спутниковых изображений реализуются путем последовательных вызовов утилит, входящих в состав ПО GDAL. Отдельные вызовы используются для корректировки исходных данных и для получения мозаики исходных фрагментов данных по заданной территории в требуемой географической проекции. Так как этот вариант, на наш взгляд, является не самым оптимальным, в настоящее время ведется разработка процедуры на языке программирования C++, использующей GDAL API для проведения операций со спутниковыми изображениями. Предварительные результаты ее тестирования показывают существенное повышение скорости выполнения операций по предварительной подготовке исходных наборов спутниковых данных по сеансам, за счет минимизации издержек, связанных с использованием промежуточных временных файлов.

В настоящее время ведутся работы по внедрению представленной технологии, в частности, для задачи построения безоблачных атмосферно-скорректированных композитов по данным прибора КМСС, установленного на российских спутниках серии «Метеор». Также продолжаются работы по интеграции подсистемы блочного доступа к данным с автоматизированной системой распределенной многопоточковой обработки спутниковых данных.

3.1.6 Перспективы использования представленного подхода для формирования спутниковых изображений в картографических WEB интерфейсах

Современные картографические WEB-интерфейсы, как правило, используют динамическое формирование спутниковых изображений для предоставления их пользователям. И в полной мере это относится к интерфейсам доступа, реализуемым на основе возможностей, предоставляемых ЦКП «ИКИ-Мониторинг». Кроме того, что все спутниковые изображения формируются по заданной территории и географической проекции, большая часть информационных продуктов получается динамически на основе обработки имеющихся в архивах исходных спутниковых данных. Поэтому задачу формирования изображений для интерфейса можно с полным правом считать частным случаем проведения обработки спутниковых данных. При этом стандартным подходом, используемым для ускорения визуализации картографических данных, является формирование итогового изображения из небольших фрагментов, которые запрашиваются с сервера доступа в параллельном режиме. Поэтому мы считаем, что к данной задаче вполне может быть применен подход, во многом аналогичный описанному выше.

Принципиальная схема организация блочного доступа применительно к задаче формирования спутниковых изображений для картографических WEB-интерфейсов приведена на рисунке 3.1.6.1. Планируется реализовать интерфейс, в рамках которого спутниковые изображения будут формироваться из гранул фиксированного размера, запрашиваемых в параллельном режиме с группы серверов визуализации данных, в качестве которых могут использоваться также и сервера предварительной обработки данных под обработку. Для этого на них предполагается установить сервисы формирования информационных продуктов по ячейкам регулярных разбиений, которые на основе небольших доработок могут быть получены из реализованных на настоящий момент сервисов доступа к данным. При этом для поддержки визуализации данных в разных масштабах будет реализован целый ряд разбиений разного пространственного разрешения. Дополнительное ускорение визуализации изображений при смещении области видимости пользователем в интерфейсе может достигаться благодаря

кешированию полученных фрагментов изображения. Так как для работы интерфейсов ссылки на получение данных должны быть доступны извне, в предлагаемой схеме дополнительно задействованы прокси-сервера с включенной функцией кеширования, через которые производится запрос к серверам переадресации запросов, реализуемых по тому же принципу, что было уже описано выше.



Рисунок 3.1.6.1 – Схема организации блочного доступа к данным применительно к задаче формирования спутниковых изображений для картографических WEB-интерфейсов.

3.2. Разработка подходов к построению систем работы данными планетных миссий на основе методов и технологий работы с данными дистанционного мониторинга Земли на примере системы Ares Mars

3.2.1 Введение

Настоящий раздел посвящен описанию опыта использования технологий распределенной работы со сверхбольшими архивами данных спутниковых наблюдений Земли для создания информационных систем работы с данными, получаемыми в ходе реализации проектов по исследованию других планет, а именно, на примере разработанной в интересах миссии «ЭкзоМарс-2016» системы AresMars.

В 2018 году аппарат TGO миссии «ЭкзоМарс-2016» закончил аэроторможение и приступил к выполнению научных задач. Данные с приборов ACS, CASSIS, FRENД и NOMAD начали в штатном режиме поступать в архивы научных центров в Европе и России. Одной из основных целей данной миссии является обеспечение возможности разностороннего анализа атмосферы и поверхности Марса при помощи разнородных наблюдений одних и тех же явлений. Для обеспечения решения этой задачи, то есть обеспечения совместной работы с данными приборов TGO, в ИКИ РАН на основе технологий UNISAT и GEOSMIS, ранее применяемых для анализа данных дистанционного зондирования Земли, была создана специализированная веб-ГИС, получившая название ARES – Mars - Analysis, Research and Exploration Service.

Система ARES-Mars позволяет получить совместный доступ как данным, поступающим с TGO, так и к данным предыдущих миссий по исследованию Марса. Такой подход не только даёт возможность достичь поставленной цели по совместной работе с разнородными данными текущих миссий, но и позволяет ретроспективно решить эту задачу для исторических данных, по отдельности накопленных для разных миссий. На сегодняшний день помимо информации, поступающей с КА TGO, в систему интегрируются данные с оптических камер КА Mars Express, ведутся работы по интеграции оптических данных с КА MRO. Главной особенностью, отличающей миссии по наблюдению Марса от миссий по наблюдению Земли, является широкое использование спектрометров, предназначенных для определения состава атмосферы Марса. Характерными примерами таких приборов можно назвать спектрометр SPICAM, установленный на КА Mars Express, и приборы ACS и NOMAD, установленные на КА TGO. Основным режимом их работы, помимо наблюдений в надири, является режим солнечных затмений, при котором свет Солнца проходит через атмосферу Марса и регистрируется прибором. Данные, полученные в таком режиме, требуют отдельного подхода к представлению, так как непосредственного наблюдения поверхности не происходит. Кроме того, даже данные надирных наблюдений могут быть достаточно трудно визуализируемы без серьёзной обработки до уровня L2.

Опыт разработки и поддержки системы ARES Mars показал, что имеющиеся технические решения по работе с данными наблюдения Земли могут быть успешно применены к данным наблюдения других планет. Связано это, безусловно, с тем, что значительная часть данных, получаемых в ходе планетных миссий, фактически являются данными дистанционных наблюдений, которые по своей структуре и схемам организации обработки и анализа данных во многом аналогичны данным наблюдения Земли из космоса.

В следующих подразделах приводится описание архитектуры системы, её основных компонент и примеров работы с ней.

3.2.2 Общая архитектура системы

Основной задачей при создании системы стала разработка новых инструментов для исследователей, позволяющих проводить распределённую обработку и совместный анализ данных текущих и исторических миссий по изучению Марса. Главной особенностью создаваемых инструментов и решений является их глубокая интеграция с постоянно пополняемыми архивами данных. Это позволяет отказаться от традиционных подходов к работе с данными дистанционных исследований, требующих проведения исследователями большого количества процедур обработки данных на своих рабочих мощностях и перекачки больших объёмов данных для такой обработки. Новый подход, реализованный в созданной системе, даёт возможность обширной онлайн-обработки данных за счёт использования ресурсов системы, а также возможность работы с произвольными наборами данных строго по выбранным пространственным, временным и другим критериям. Кроме того, система ARES Mars обеспечивает возможность распределённой работы с данными миссий по изучению Марса посредством широко распространённых веб-браузеров, без использования дополнительного ПО для обработки данных.

Система включает в себя три основных компоненты:

- Блок ведения архивов;
- Блок обработки данных;
- Блок обеспечения работы с данными.

Общая схема архитектуры системы представлена на рисунке 3.2.2.1.

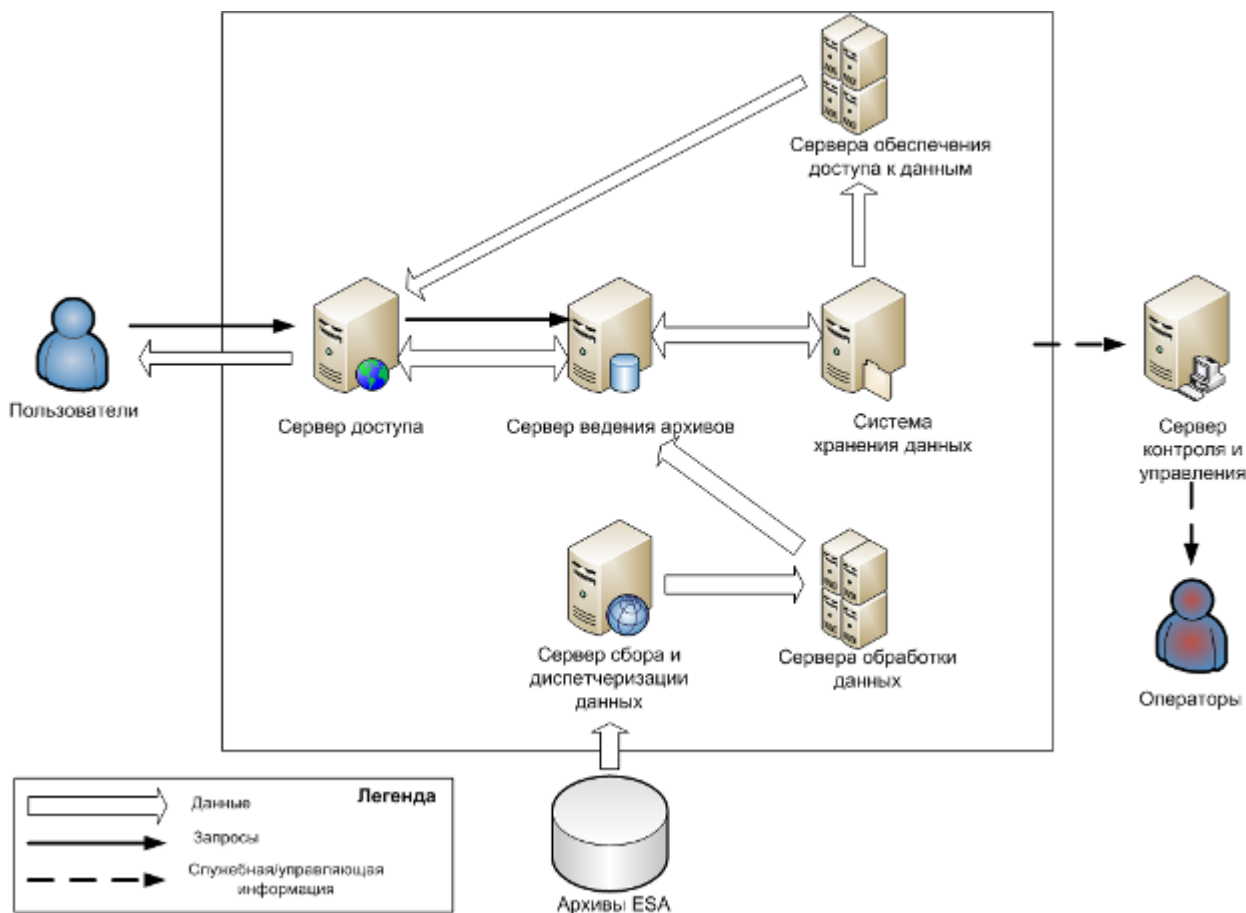


Рисунок 3.2.2.1. Общая схема архитектуры системы

Система ARES-Mars обеспечивает ведение постоянно пополняющихся оперативных и исторических архивов данных, прямой доступ к этим архивам, а также предоставляет средства для online-доступа, анализа и обработки имеющихся в архивах данных. Помимо этого, при создании системы ARES-Mars реализована программная платформа для создания и интеграции новых пользовательских инструментов для обработки и анализа данных.

Для обеспечения такой возможности при разработке системы ARES-Mars были решены следующие основные задачи:

- 1) обеспечение возможности единообразной работы с архивами данных различных миссий по исследованию Марса общей емкостью не менее 400 ТБ в непосредственном доступе (online);
 - 2) обеспечение возможности автоматического пополнения архивов данных со скоростью не менее 100 Гб в сутки;
 - 3) обеспечение возможности работы с информацией, поступающей от действующих, перспективных и исторических миссий по исследованию Марса, в том числе Exomars и Mars Express;
 - 4) обеспечение возможности удаленного доступа к информации (выбора, получения и проведения анализа);
 - 5) обеспечение возможности расширения инструментария по анализу и обработке данных.
- Кроме того, при создании системы ARES-Mars к ней были предъявлены следующие требования, которые были успешно выполнены:
- 1) создаваемые архивы данных должны быть совместимы (иметь инструменты взаимодействия) с действующими отечественными и зарубежными архивами данных миссий по исследованию Марса;
 - 2) Система ARES-Mars должна обеспечивать возможность одновременной работы с ним не менее 20 специалистов, ведущих научные разработки и вычисления;

- 3) Разработка элементов системы должна проводиться с использованием ОС семейства UNIX (подсистемы обработки данных, подсистемы архивации данных и подсистемы представления данных);

Система ARES-Mars была создана преимущественно с использованием ПО с открытым исходным кодом, а также специализированных программных пакетов разработки ИКИ РАН.

3.2.3 Система ведения архивов данных

Система ARES-Mars в первую очередь предназначена для работы с данными комплекта научной аппаратуры, установленной на аппарате TGO миссии Exomars-2016. В состав комплекта входят следующие приборы:

- ACS (Atmospheric Chemistry Suite) – набор ИК-спектрометров для анализа компонент атмосферы Марса;
- CaSSIS (Colour and Stereo Surface Imaging System) – оптическая широкозахватная камера высокого разрешения;
- FREND (Fine Resolution Epithermal Neutron Detector) – детектор нейтронов для обнаружения водорода;
- NOMAD (Nadir and Occultation for MArS Discovery) – спектрометр для анализа компонент атмосферы Марса.

Кроме того, система предполагает возможность интеграции данных приборов, установленных на посадочном модуле EDM (комплекты приборов AMELIA, DECA и DREAMS).

Для обеспечения возможности ретроспективного анализа данных исторических миссий по исследованию Марса система также имеет возможность обеспечивать работу с данными следующих приборов и миссий:

- Mars Express – HRSC, SPICAM, OMEGA;
- MRO – CRISM, HiRISE, MARCI, CTX, MCS;
- Mars Odyssey – THEMIS, HEND;

Этот список не является полным, так как архитектура системы является модульной и позволяет с минимальными затратами адаптировать её под новые типы данных.

Объёмы данных, с которыми должна иметь возможность работать системы, планировались, исходя из возможностей работы КА TGO. Запланированный срок работы аппарата TGO в рамках миссии Exomars-2016 составляет 687 дней. Исходя из этого, а также из номинальных характеристик приборов, минимальная конфигурация системы рассчитана на объёмы данных, представленные в таблице 3.2.3.1.

Таблица 3.2.3.1 - Запланированные суммарные объёмы данных научной аппаратуры миссии Exomars-2016.

Прибор	Суммарный объём данных за время миссии
ACS	1 ТБ
CaSSIS	7 ТБ
FREND	30 ГБ
NOMAD	140 ТБ
Итого:	180 ТБ

Данные научной аппаратуры миссии Exomars-2016 запланированы к распространению в формате PDS (Planetary Data System). Все данные, поступающие в архивы в таком формате, будут обрабатываться для конвертации их в форматы, более подходящие для визуализации и онлайн-обработки (например, в формат GeoTIFF для

растровых данных), после чего данные будут отправляться в систему архивации и будут доступны для работы с использованием картографического интерфейса.

Технологически система архивации построена на основе технологии UNISAT, разработанной в ИКИ РАН для работы с разнородными данными наблюдения Земли. Её общие принципы и подходы оказались достаточно универсальными, что позволило применить её к данным наблюдения других планет с минимальной адаптацией. Одним из основных направлений модернизации стала доработка, направленная на работу с данными спектрометров, предназначенных для определения параметров и состава атмосферы посредством затменных наблюдений. Для таких приборов крайне важным является не только визуализация полученных результатов обработки наблюдений, но и картирование самих сеансов наблюдения. Это необходимо как для понимания выполнения уже заданной научной программы, так и для планирования последующих наблюдений. Пример отображения сеанса затменного наблюдения прибора ACS MIR в картографическом интерфейсе системы ARES-Mars представлен на рисунке 3.2.3.1. Серая линия показывает подспутниковый трек КА TGO во время включения прибора и временные отсчеты сеанса наблюдения, цветная линия показывает трек подзатменной точки во время сеанса наблюдения и высоту Солнца над горизонтом во время сеанса наблюдения.



Рисунке 3.2.3.1 - Пример отображения сеанса наблюдения прибора ACS MIR

Ключевыми характеристиками технологии UNISAT можно назвать следующие:

- Использование унифицированной структуры БД для хранения спутниковых данных, которая позволяет обеспечить работу с широким набором различных спутниковых данных и результатов их обработки без существенного проигрыша в производительности;
- Использование единой справочной БД, содержащей всю необходимую информацию, в том числе информацию о центрах приема, спутниках, приборах, продуктах, включая «виртуальные» продукты и правила их построения;
- Предоставление доступа к спутниковым данным различных типов на основе унифицированного программного обеспечения, реализующего весь необходимый функционал, имеющийся в специализированных системах ведения архивов спутниковых данных;
- Поддержка единого для всех типов спутниковых данных механизма предоставления доступа к «виртуальным» продуктам, позволяющего гибко задавать правила получения новых информационных продуктов на основе имеющихся в архиве данных;

- Поддержка сервисов предоставления расширенных метаданных, необходимых для работы инструментов анализа данных и проведения обработки данных в режиме «реального времени»;
- Максимальная унификация программного обеспечения, предназначенного для архивации спутниковых данных и ведения распределенных архивов, при сохранении совместимости с существующими форматами поступления данных в архивы.

Внедрение данной технологии, обеспечивающей эффективную работу со сверхбольшими распределёнными архивами спутниковых данных, позволило реализовать систему работы с планетными данными, позволяющую проводить комплексный анализ разнородной информации с использованием единого инструментария.

3.2.4 Интерфейсы работы с данными и примеры их использования

Основным инструментом работы с данными в системе ARES-Mars являются специализированные картографические веб-интерфейсы, созданные по разработанной в ИКИ РАН технологии GEOSMIS. Эти интерфейсы обеспечивают пользователю возможность работы с разнородными данными (данными разных приборов, пространственных разрешений, временной скважности и т.д.), получаемыми из архивов системы. Пользователь имеет возможность задавать произвольные временные и пространственные диапазоны интереса для дальнейшей работы. Данные, с которыми возможна работа, могут быть получены с разных аппаратов и приборов, иметь разное пространственное разрешение, быть как растровыми, так и векторными. Общий вид интерфейса представлен на рисунке 3.2.4.1

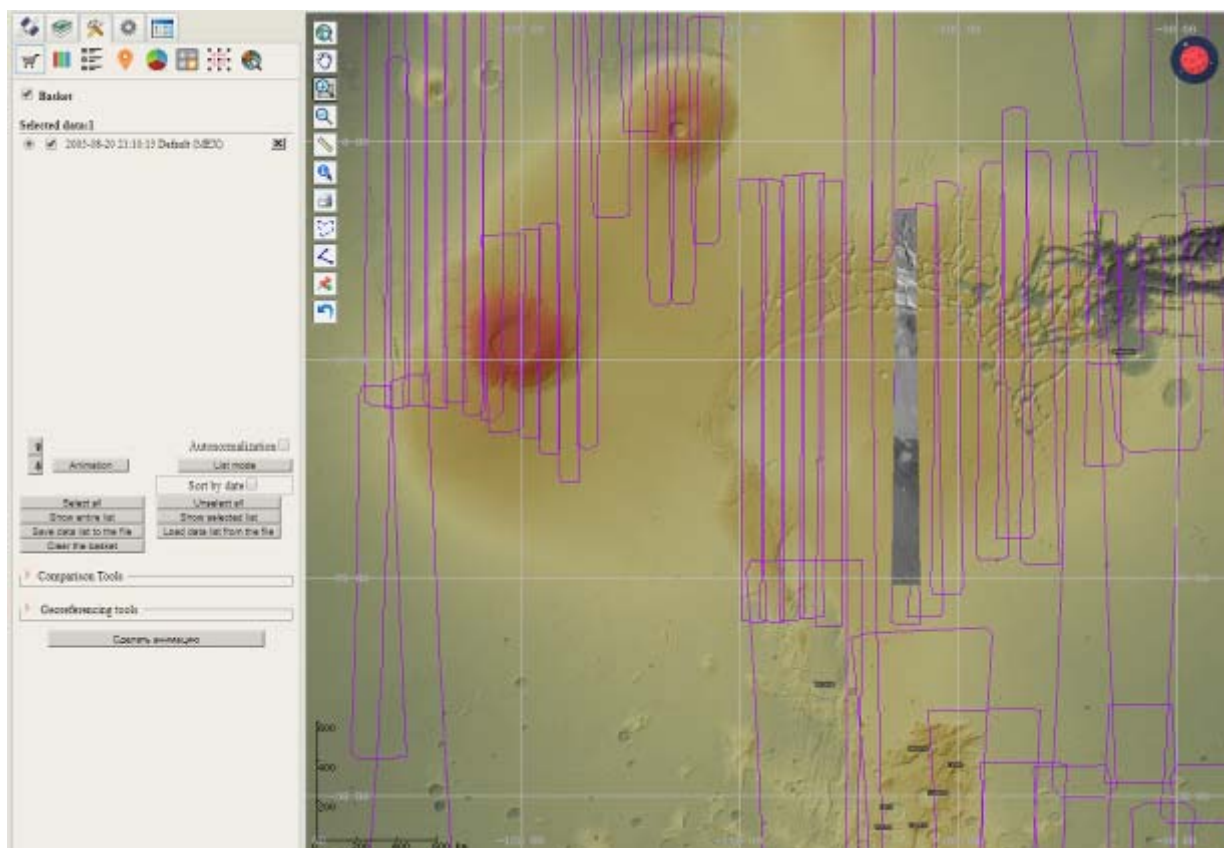


Рисунок 3.2.4.1 - Общий вид интерфейса с включённой картографической подложкой и выбранной сценой данных КА Mars Express, прибор HRSC

Помимо возможности выбора и просмотра данных, картографические интерфейсы обеспечивают возможность онлайн-обработки и анализа данных. В частности, пользователю доступны операции цветокоррекции и цветосинтеза над выбранными

данными, операции по их сравнению, а также операции по более сложной обработке данных, в том числе по проведению классификации (как с обучением, так и без) и математических операций над изображениями. Для созданных пользователем тематических продуктов реализованы механизмы экспорта для дальнейшей обработки и анализа в каких-либо других инструментах, а также механизмы сохранения и публикации в системе для дальнейшей работы с ними. Также пользователь имеет возможность загрузки в систему собственных тематических продуктов в формате GeoTIFF, полученных любым доступным ему способом, при помощи специализированного интерфейса.

Работа инструментов цветосинтеза и цветокоррекции на примере данных HRSC представлена на рис. 3.2.4.2. Работа инструментов классификации на примере данных HRSC представлена на рис. 3.2.4.3.

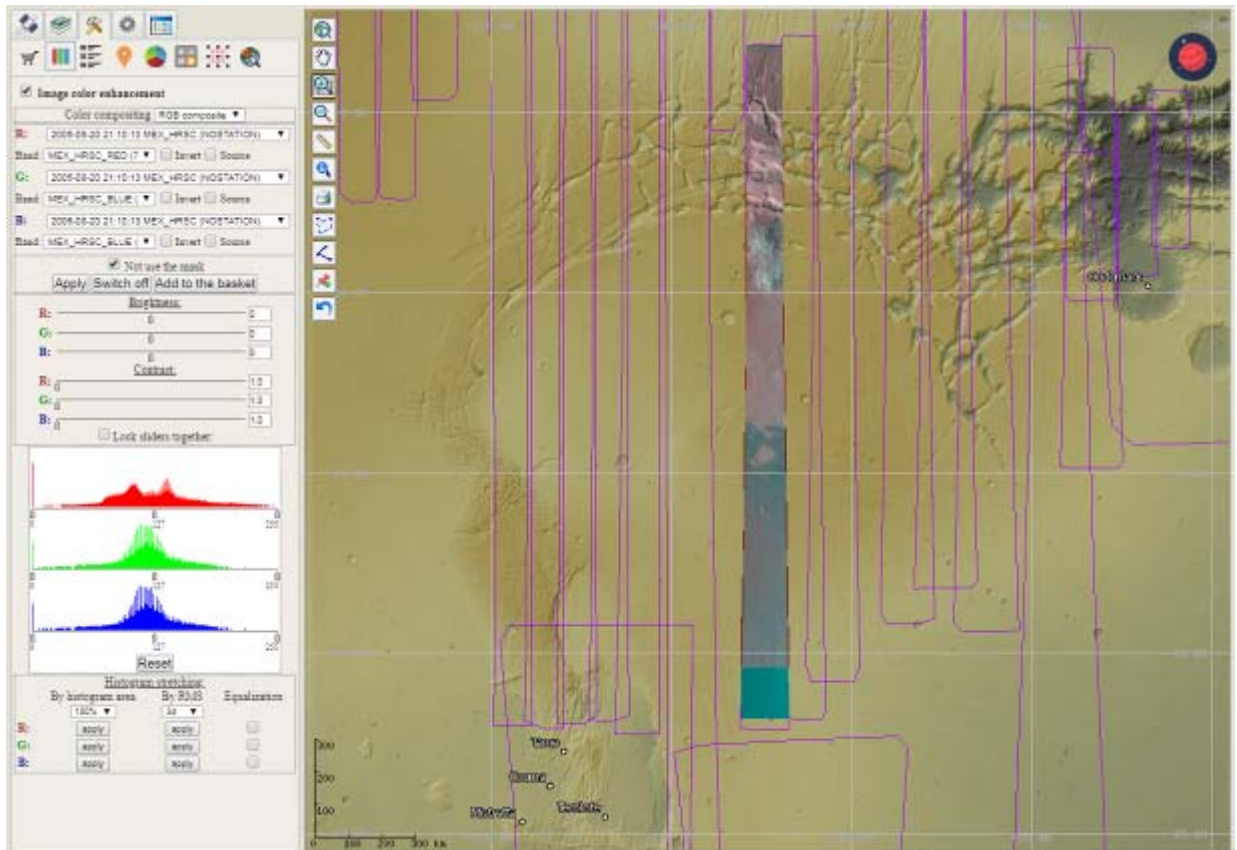


Рисунок 3.2.4.2 Пример использования инструментов цветосинтеза на данных HRSC.

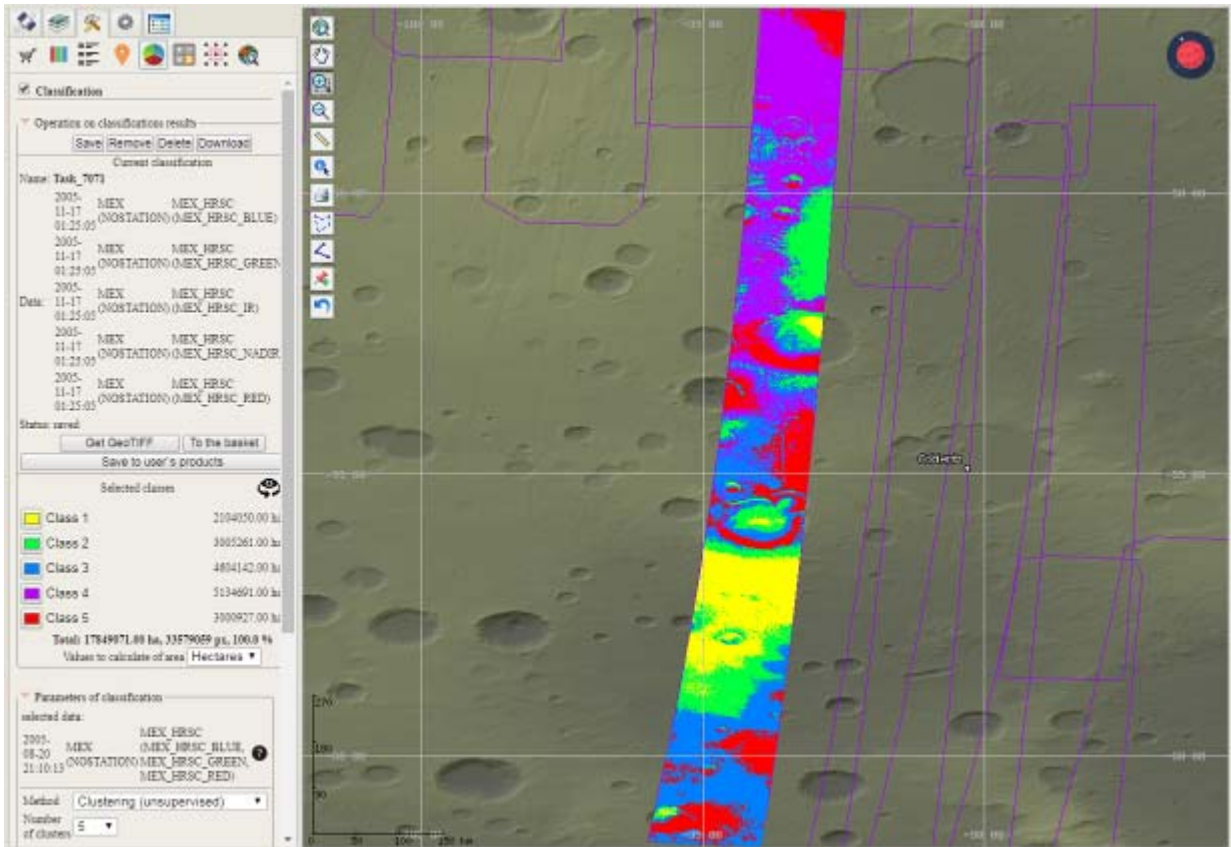


Рисунок 3.2.4.3 - пример использования инструментов классификации на данных HRSC.

Помимо визуального анализа данных, пользователю доступны механизмы анализа различных качественных и количественных показателей наблюдаемых территорий, таких как динамика различных показателей по заданным территориям или пространственные профили по заданным линиям разреза. Для этого используются интерфейсы работы с графиками, позволяющие отображать совместную динамику требуемых показателей за необходимый промежуток времени.

3.2.5 Работы по развитию системы, проведённые в 2109 году

В 2019 году основным направлением развития системы стала её адаптация под данные, характерные для планетных исследований, но нетипичных для исследования Земли. В первую очередь, таковыми стали данные спектрометров ACS и SPICAM, установленных на КА TGO и Mars Express соответственно. Для этих приборов была разработана схема картирования затменных и надирных наблюдений, адаптирована схема хранения таких данных для их совместного отображения с растровыми данными оптических приборов и реализована механика подготовки и занесения данных в БД системы. С использованием разработанных механик в системе были картированы наблюдения спектрометров ACS MIR (затменные наблюдения) и ACS TIRVIM (надирные наблюдения). Ведутся работы по интеграции данных SPICAM. Также в 2019 году были проведены работы по созданию механизмов автоматического получения данных из архивов PDS NASA, их обработки для конвертации в формат UNISAT и занесения обработанных данных в БД системы. Созданные механизмы были отработаны на данных камеры высокого разрешения Mars Express HRSC и гиперспектрометра MRO CRISM. Также ведётся работа по интеграцию в систему данных других приборов КА TGO, Mars Express и MRO.

3.2.6 Заключение

Разработанная и поддерживаемая ИКИ РАН система ARES-Mars стала уникальной веб-ГИС, обеспечивающей эффективную прозрачную работу с данными различных миссий по исследованию Марса. Система позволяет проводить поиск, визуализацию и онлайн-обработку и анализ разнородных данных средствами самой системы, без использования дополнительного ПО, установленного у пользователя. Система реализована на современных отечественных технологиях ведения распределённых, сверхбольших архивов и представления спутниковых данных, является легко масштабируемой как с точки зрения аппаратной составляющей, так и с точки зрения данных и инструментов обработки. Дальнейшее развитие системы и наполнение её архивов обеспечит исследователей Марса возможностью совместного анализа разнородных данных в едином легкодоступном окружении, что должно привести к повышению эффективности планетных исследований.

3.3. Развитие системы Аврора-Арктика

Информационная система "Аврора-Арктика" предназначена для анализа космической погоды - состояния ионосферы Земли, геомагнитной и геоэлектрической обстановки на поверхности и т.п. Она предусматривает автоматический сбор, визуализацию спутниковых и наземных данных, расчет моделей и прогнозов. В настоящее время в нее интегрированы данные спутников серии DMSP, DSCOVR, IRIDIUM, наблюдения с наземных камер. Ежечасно формируются прогнозы границ аврорального овала на основе модели Старкова. Система создана с использованием разработанной в ИКИ РАН технологии GEOSMIS, обеспечивающей работу со сверхбольшими распределенными многомерных спутниковых данных, результатами их обработки. Пользовательский картографический интерфейс системы доступен on-line через браузер и не требует установки специального программного обеспечения. Информационная система такого уровня сложности является уникальной для данной научной отрасли (солнечно-земная физика, космическая погода).

В 2019 году была проведена работа по интеграцию в систему модели прогноза границ аврорального овала OVATION Prime. Модель OVATION Prime основана на данных об измерении потока электронов и протонов, а также о величине межпланетного магнитного поля и предсказывает пространственное распределение четырех компонент авроральной активности: вклад ионов, моноэнергетических электронов, электронов с широким диапазоном энергий и диффузного сияния (Рисунок 3.3.1). Разрешение модели – 15 минут MLT x 0.25° MLAT между 50°-89.5° для северной и южной полусфер. Модель была реализована на языке программирования Python 3 с дополнительными библиотеками NumPy, SciPy и MySQL Connector.

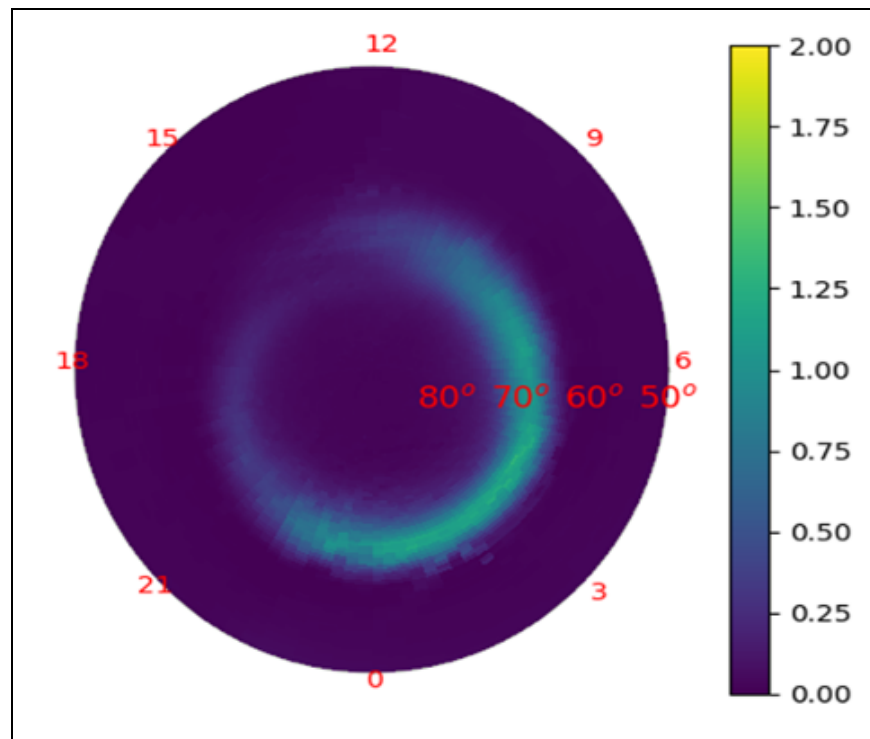


Рисунок 3.3.1 - Поле diffuse aurora над северным полюсом, полученное с помощью модели OVATION Prime

Также было разработано программное обеспечение для импорта, архивации и визуализации данных измерений когерентного радара декаметрового диапазона, расположенного на территории обсерватории Арти Института Геофизики УрО РАН (Рисунок 3.3.2). Основным результатом работы когерентного декаметрового радара является наблюдение за характеристиками мелкомасштабных (декаметровых) неоднородностей F-слоя ионосферы. В рамках существующих моделей таких неоднородностей их движение связывается со структурой электрических и магнитных полей в точке рассеяния и позволяет оценивать направление и абсолютную величину электрического поля в широком диапазоне азимутов и дальностей, а также общую радиопрозрачность ионосферы



Рисунок 3.3.2 - Пример визуализации данных когерентного радара декаметрового диапазона

Заключение

Представленные в настоящей главе результаты показывают, что по направлению «Разработка новых методов и подходов организации хранения и параллельной обработки сверхбольших объемов данных для решения задач анализа долговременных наблюдений в интересах ведения мониторинга и изучения различных процессов и явлений» проведенные работы соответствуют установленной в плане работ ИКИ РАН на 2019 г. тематике. Таким образом, работы выполнены полностью в соответствии с планом.

При этом получены следующие основные результаты:

- Разработана концепция и архитектура построения схемы динамического блочного доступа к разнородным сверхбольшим распределенным архивам данных спутниковых наблюдений Земли;
- на примере организации систем для распределенной работы с данными, получаемыми в ходе планетных миссий и исследований космической погоды, показано, что подходы, разрабатываемые для организации работы со спутниковыми данными наблюдения Земли, могут эффективно использоваться и в других направлениях космических исследований.

Таким образом, можно считать, что работы по данному направлению темы в 2019 году полностью выполнены.

ЗАКЛЮЧЕНИЕ

Таким образом, представленный отчет по теме «Большие данные в космических исследованиях: астрофизика, солнечная система, геосфера» (№ 0024-2019-0014) показывает, что в соответствии с планом ИКИ РАН на 2019 год они велись по следующим основным направлениям:

- Развитие комплекса взаимосвязанных и дополняющих друг друга физико-математических моделей для диагностики и прогнозирования состояния высокоширотной ионосферы в северном и южном полушариях, которые электродинамически связаны между собой через область средних широт. Разработки общей модели взаимодействия в системе солнечный ветер-магнитосфера-ионосферная нейтральная атмосфера Земли;
- Применение методов машинного обучения и глубокого обучения для обработки и анализа больших массивов астрономических данных в задачах наблюдательной астрофизики;
- Разработка новых методов и подходов организации хранения и параллельной обработки сверхбольших объемов данных для решения задач анализа долговременных наблюдений в интересах ведения мониторинга и изучения различных процессов и явлений.

В рамках данной тематики опубликовано в 2019 году 23 работы, из них 5 в рецензируемых изданиях, входящих в Scopus и WoS (плановый показатель 4).

Таким образом, можно считать, что поставленные в плане задачи на 2019 год в рамках данной темы полностью решены.

СПИСОК ЛИТЕРАТУРЫ

Опубликовано – 23 работы, из них:

- в зарубежных изданиях – 4
- книг – 0
- статьи в отечественных научных рецензируемых журналах – 1
- статьи в сборниках материалов конференций – 0
- материалы конференции - 9
- доклады, тезисы – 9
- публикаций по грантам РФФИ – 0
- в соавторстве с зарубежными авторами – 0
- статьи в WOS и Scopus – 5
- статьи в WOS и Scopus (Q1, Q2) – 1
- статьи в WOS и Scopus (Q3, Q4) - 1

Публикации в зарубежных изданиях

1. *Proshin A.A., Loupian E.A., Kashnitskii A.V., Balashov I.V.* UNISAT. The Technology for Development of United Systems of Maintaining Extra Large Distributed Archives of Heterogeneous Satellite Data // CEUR Workshop Proceedings, 2019. – в печати **Scopus**
2. *Proshin A.A., Loupian E.A., Kashnitskii A.V., Balashov I.V., Burcev M.A.* Current Capabilities of the "IKI-Monitoring" Center for Collective Use // CEUR Workshop Proceedings, 2019. – в печати **Scopus**
3. *Lukianova, R.* Swarm field-aligned currents during a severe magnetic storm of September 2017, *Annales Geophysicae*, <https://doi.org/10.5194/angeo-2019-40>, 2019. **WOS, Scopus (Q2)**
4. *Kolosov, I., Gerasimov, S., Meshcheryakov, A.* Architecture of Processing and Analysis Systems for Big Astronomical Data // Proceedings Paper, Astronomical Society of the Pacific Conference Series, Том: 521 Стр.: 428-430, Идентификационный номер: WOS:000495812900104 (**WOS**)

Публикации в отечественных научных рецензируемых изданиях

5. *Лупян Е.А., Прошин А.А., Бурцев М.А., Кашицкий А.В., Балашов И.В., Барталев С.А., Константинова А.М., Кобец Д.А., Мазуров А.А., Марченков В.В., Матвеев А.М., Радченко М.В., Сычугов И.Г., Толпин В.А., Уваров И.А.* Опыт эксплуатации и развития центра коллективного пользования системами архивации, обработки и анализа спутниковых данных (ЦКП «ИКИ-Мониторинг») // Современные проблемы дистанционного зондирования Земли из космоса. 2019. Т. 16. № 3. С. 151-170. DOI: 10.21046/2070-7401-2019-16-3-151-170. **Scopus (Q3)**

Материалы конференции

6. Умрихин А.Д., Никифоров О.В., Петрукович А.А., Уваров И.А. Интеграция моделей прогноза границ аврорального овала в информационную систему мониторинга ионосферы "Аврора-Арктика" // Материалы Семнадцатой Всероссийской Открытой конференции «Современные проблемы дистанционного зондирования Земли из космоса. 11-15 ноября 2019. ИКИ РАН, 2019. С. 499. DOI: 10.21046/17DZZconf-2019a.

7. Прошин А.А., Лупян Е.А., Кашицкий А.В., Балашов И.В. UNISAT. Технология построения унифицированных систем ведения сверхбольших распределенных архивов разнородных спутниковых данных // Обработка пространственных данных в задачах мониторинга природных и антропогенных процессов (SDM-2019) (Электронный ресурс): Сборник трудов всероссийской конференции (26-30 августа 2019 г., г. Бердск). Новосибирск: ИВТ СО РАН, 2019. С. 153-158.
8. Калашиников А.В., Успенский С.А., Крамарева Л.С., Бурцев М.А., Балашов И.В., Мазуров А.А., Прошин А.А., Пустынский И.С., Толпин В.А., Холодов Е.И. Новые возможности и перспективы развития объединённой системы работы с данными НИЦ "Планета // Материалы Семнадцатой Всероссийской Открытой конференции «Современные проблемы дистанционного зондирования Земли из космоса. 11-15 ноября 2019. ИКИ РАН, 2019. С. 91. DOI: 10.21046/17DZZconf-2019a.
9. Прошин А.А., Бурцев М.А., Балашов И.В., Лупян Е.А., Радченко М.В., Сычугов И.Г. Проблемы поддержки и развития ЦКП «ИКИ-Мониторинг» - возможные направления решения // Материалы Семнадцатой Всероссийской Открытой конференции «Современные проблемы дистанционного зондирования Земли из космоса. 11-15 ноября 2019. ИКИ РАН, 2019. С. 115. DOI: 10.21046/17DZZconf-2019a.
10. Марченков В.В., Толпин В.А., Уваров И.А. Анализ рядов данных в информационных системах семейства "Созвездие Вега" // Материалы Семнадцатой Всероссийской Открытой конференции «Современные проблемы дистанционного зондирования Земли из космоса. 11-15 ноября 2019. ИКИ РАН, 2019. С. 106. DOI: 10.21046/17DZZconf-2019a.
11. Прошин А.А., Бурцев М.А., Радченко М.В. Поддержка функционирования сверхбольших файловых хранилищ спутниковых данных ЦКП «ИКИ-Мониторинг» // Материалы Семнадцатой Всероссийской Открытой конференции «Современные проблемы дистанционного зондирования Земли из космоса. 11-15 ноября 2019. ИКИ РАН, 2019. С. 116. DOI: 10.21046/17DZZconf-2019a.
12. Прошин А.А., Лупян Е.А., Бурцев М.А., Кашицкий А.В., Кобец Д.А., Матвеев А.М. Унифицированная технология организации блочного доступа к данным в архивах для проведения их эффективной распределенной параллельной обработки // Материалы Семнадцатой Всероссийской Открытой конференции «Современные проблемы дистанционного зондирования Земли из космоса. 11-15 ноября 2019. ИКИ РАН, 2019. С. 118. DOI: 10.21046/17DZZconf-2019a.
13. Уваров И.А. Организация обработки и анализа метеорологических данных в ЦКП "ИКИ-Мониторинг" // Материалы Семнадцатой Всероссийской Открытой конференции «Современные проблемы дистанционного зондирования Земли из космоса. 11-15 ноября 2019. ИКИ РАН, 2019. С. 126. DOI: 10.21046/17DZZconf-2019a.
14. Бурцев М.А., Атрохов А.А., Вилкова Д.В., Шадрина А.В. Интеграция данных различных миссий по исследованию Марса в ИС ARES-MARS // Материалы Семнадцатой Всероссийской Открытой конференции «Современные проблемы дистанционного зондирования Земли из космоса. 11-15 ноября 2019. ИКИ РАН, 2019. С. 358. DOI: 10.21046/17DZZconf-2019a.

Тезисы, доклады

15. Лупян Е.А. Проблемы и подходы организации распределенного анализа сверхбольших объемов данных дистанционного зондирования Земли // выездное заседание "Методы искусственного интеллекта в космических исследованиях" научного семинара "Механика, управление и информатика". 19-21 февраля 2019 г. Таруса
16. Бурцев М. А., Кашицкий А. В., Константинова А.М. Современные возможности работы с данными автоматических и пилотируемых КА и комплексов // IX научно-техническая

- конференция молодых ученых и специалистов центра управления полетами. 9-12 апреля 2019. Москва: ЦНИИМАШ.
17. *Лупян Е. А., Балашов И.В., Барталев С.А., Сенько К.С.* Анализ динамики площадей лесных пожаров на территории России в 21 веке на основе данных спутниковых наблюдений горения // Доклады VII Всероссийской конференции (с международным участием) "Аэрокосмические методы и геоинформационные технологии в лесоведении, лесном хозяйстве и экологии", посвященной памяти выдающегося ученого-лесоведа, академика РАН А.С. Исаева. Москва, 22-24 апреля 2019. М.: ЦЭПЛ РАН, 2019.
 18. *Лупян Е.А. Барталев С.А., Балашов И.В., Егоров В.А., Жарко В.О., Стыценко Ф.В.* Возможности комплексного дистанционного мониторинга лесов (опыт ИКИ РАН) // V Международная научно-практическая конференция «Прикладные проблемы оптики, информатики, радиофизики и физики конденсированного состояния». 16-17 мая 2019, Минск (Беларусь).
 19. *Лупян Е.А.* Текущие возможности центра коллективного пользования ИКИ-Мониторинг // II Всероссийский форум с международным участием "Цифровизация-2019". 28-29 октября, Москва.
 20. *Лупян Е.А.* Современная платформа для решения задач исследований Земли из Космоса. Центр коллективного пользования системами архивации, обработки и анализа данных спутниковых наблюдений // Заседание Совета РАН по космосу, 11 декабря 2019 г.
 21. *Мещеряков А.В., Васильченко А.Д., Герасимов С.В.* Применение сверточных нейронных сетей в задаче детектирования астрономических объектов на рентгеновских изображениях // Сб. Научная конференция ЛОМОНОСОВСКИЕ ЧТЕНИЯ. Тезисы докладов. 15-25 апреля 2019 г, серия СЕКЦИЯ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И КИБЕРНЕТИКИ, Факультет вычислительной математики и кибернетики, Издательство ООО "МАКС Пресс" (Москва), тезисы, с. 32-33
 22. *Мещеряков А.В.* "СРГz - вероятностные измерения фото-z для рентгеновских источников на космологических расстояниях" Всероссийская конференция ", Научная конференция "Астрофизика высоких энергий сегодня и завтра (HEA-2019)".
 23. *Petrukovich, B. Kozelov, T. Podladchikova, I. Uvarov, O. Nikiforov* Monitoring and Forecasting of Space Weather in the Polar Region with use of the "Aurora Arctic" GIS // (14 International Conference on Substorms, Tromsø, Norway, 30.09.19 - 04.10.19.