

*На правах рукописи*



**КОБЕЦ Дмитрий Александрович**

**МОДЕЛИ И ПРОГРАММНАЯ ИНФРАСТРУКТУРА ДЛЯ  
ПРОЕКТИРОВАНИЯ ИНТЕРФЕЙСОВ КОНТРОЛЯ И УПРАВЛЕНИЯ  
РАСПРЕДЕЛЕННОЙ ОБРАБОТКОЙ И АНАЛИЗА СПУТНИКОВЫХ ДАННЫХ  
НА ОСНОВЕ VI-ТЕХНОЛОГИЙ**

Специальность 05.13.11 – Математическое и программное обеспечение  
вычислительных машин, комплексов и  
компьютерных сетей

**АВТОРЕФЕРАТ**  
диссертации на соискание ученой степени  
кандидата технических наук

**Москва - 2017**

Работа выполнена в Федеральном государственном бюджетном учреждении науки Институте космических исследований Российской академии наук (ИКИ РАН).

Научный руководитель: **Лупян Евгений Аркадьевич**  
доктор технических наук, заведующий отделом «Технологии спутникового мониторинга», заместитель директора Федерального государственного бюджетного учреждения науки Института космических исследований Российской академии наук (ИКИ РАН)

Официальные оппоненты: **Романов Александр Алексеевич**  
доктор технических наук, профессор, главный конструктор направления АО «Российские космические системы», доцент кафедры «Системы, устройства и методы геокосмической физики» Федерального государственного автономного образовательного учреждения высшего образования «Московского физико-технического института (государственного университета)»

**Маглинец Юрий Анатольевич**  
кандидат технических наук, профессор, руководитель научно-учебной лаборатории информационной поддержки космического мониторинга Федерального государственного автономного образовательного учреждения высшего образования «Сибирского федерального университета»

Ведущая организация: **Федеральное государственное бюджетное учреждение науки Вычислительный центр Дальневосточного отделения Российской академии наук (ВЦ ДВО РАН), г. Хабаровск**

Защита состоится **28 февраля 2018 года в 13.00 часов** на заседании диссертационного совета Д 212.211.01 в Рязанском государственном радиотехническом университете по адресу: **390005, г. Рязань, ул. Гагарина, д. 59/1.**

С диссертацией можно ознакомиться в библиотеке и на сайте Федерального государственного бюджетного образовательного учреждения высшего образования "Рязанского государственного радиотехнического университета": <http://www.rsreu.ru/>

Автореферат разослан «\_\_» \_\_\_\_\_ 201\_ г.

Ученый секретарь  
диссертационного совета Д 212.211.01  
доктор технических наук



**Прущков Александр  
Викторович**

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность работы.** Данные дистанционного зондирования Земли (ДЗЗ) в настоящее время широко используются для получения количественной информации о различных объектах и территориях. Современные технологии дистанционного зондирования, активно развивающиеся сегодня как в России так и за рубежом, в частности, позволяют получать и накапливать, как пространственные распределения, так и временные ряды характеристик различных наблюдаемых объектов и явлений. Примерами такой информации являются, например, ряды ежегодно обновляющихся карт растительного покрова, пахотных земель, земель, занятых различными видами посевов, базы данных наблюдений природных пожаров и т.д. При этом возникает достаточно широкий класс задач, в котором требуется проведение анализа подобной информации интегрированной по различным параметрам, например, по конкретным территориям и/или временным периодам. Для эффективного проведения такого анализа с учетом быстрорастущих объемов получаемых данных возникает необходимость создания специальных инструментов, которые могли бы позволять работать с рядами долговременных наблюдений, обеспечивая, в том числе, возможности проведения быстрой агрегации информации по различным признакам и параметрам.

Подобные задачи возникают и при проведении анализа информации о работе систем автоматической обработки спутниковых данных, которые в настоящее время являются основой практически всех процессов, связанных с использованием данных, поступающих от современных систем ДЗЗ. Необходимо отметить, что в связи с фактически взрывным ростом объемов получаемой и обрабатываемой информации ДЗЗ, современные системы автоматизированной обработки являются многопоточковыми и используют большое число вычислительных ресурсов, в том числе и распределенных. Поэтому анализ информации об эффективности управления такими системами, а также оперативного выявления и диагностирования сбоев, становится нетривиальным и требует создания специальных инструментов анализа данных.

Следует отметить, что в последнее десятилетие достаточно быстро развивались различные технологии, связанные с созданием инструментов для проведения анализа различной информации. Одними из наиболее быстро развивающихся в этом направлении технологий являются BI-технологии (Business intelligence). Они основаны на построении хранилищ структурированной информации на основе OLAP (online analytical processing). Эти технологии в современных условиях позволяют строить достаточно сложные динамические инструменты для онлайн анализа больших объемов информации.

**Соответствие паспорту специальности 05.13.11.** Диссертация соответствует паспорту специальности 05.13.11 – «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей» в части:

- пункта 4 «Системы управления базами данных и знаний»;
- пункта 7 «Человеко-машинные интерфейсы; модели, методы, алгоритмы и программные средства машинной графики, визуализации, обработки изображений, систем виртуальной реальности, мультимедийного общения»;
- пункта 9 «Модели, методы, алгоритмы и программная инфраструктура для организации глобально распределенной обработки данных».

**Целью диссертации** является разработка моделей данных и создание на их основе программной инфраструктуры построения динамических человеко-машинных интерфейсов, ориентированных на контроль распределенных систем обработки данных дистанционного зондирования Земли и анализа результатов обработки спутниковых данных. Разработка таких моделей и программной инфраструктуры за счет использования подходов работы с агрегированной информацией (BI-технологий) должна обеспечить повышение эффективности процессов обработки и передачи данных и знаний в распределенных

вычислительных комплексах, ориентированных на обработку и анализ спутниковых данных и результатов их обработки.

Для достижения поставленной цели, решаются следующие **задачи**:

*Задача 1.* Анализ особенностей, подходов и тенденций к организации распределенных комплексов потоковой обработки спутниковых данных и постановка основных проблем организации эффективного анализа информации о состоянии элементов таких комплексов и выполняющихся в них процессов обработки данных.

*Задача 2.* Разработка модели организации данных для построения эффективных динамических интерфейсов, необходимых для контроля и управления распределенными системами автоматизированной обработки спутниковой информации, с учетом использования подходов организации работы с агрегированной информацией (OLAP).

*Задача 3.* Создание программной инфраструктуры (в том числе для управления базами данных), на основе разработанной модели данных, для разработки блоков контроля и управления подсистемами потоковой обработки данных в современных системах дистанционного мониторинга.

*Задача 4.* Анализ особенностей, подходов и тенденций к организации работы с различными типами информационных продуктов, получаемых на основе долговременного постоянного спутникового мониторинга и постановка основные проблемы, связанные с проведением эффективного анализа результатов обработки спутниковых данных (в первую очередь долговременных рядов наблюдений).

*Задача 5.* Разработка модели организации данных, на основе проведенного анализа с использованием различных подходов и методов агрегации информации, для построения интерфейсов анализа временных рядов характеристик наблюдаемых процессов (объектов) и периодически обновляемых картографических материалов (временных серий пространственных данных).

*Задача 6.* Разработка необходимых для реализации предложенных моделей данных методов агрегирования временных рядов характеристик наблюдаемых процессов (объектов) и временных серий пространственных данных.

*Задача 7.* Разработка программной инфраструктуры (в том числе для управления базами данных), на основе разработанных моделей и методов агрегирования данных, для создания динамических интерфейсов распределенного анализа различных типов информационных продуктов, получаемых на основе данных долговременных спутниковых наблюдений в системах дистанционного мониторинга.

**Научная новизна** диссертации заключается в разработке новых моделей данных, обеспечивающих возможность создания с использованием VI-технологий эффективных динамических интерфейсов, для построения человеко-машинных интерфейсов контроля и управления распределенными системами автоматизированной обработки спутниковых данных и распределенного анализа результатов обработки данных долговременных спутниковых наблюдений.

**Степень разработанности темы.** Созданные на основе разработанных моделей организации данных и программной инфраструктуры динамические интерфейсы, не имеют аналогов, обеспечивающих схожие возможности анализа данных в существующих системах мониторинга.

**Основные положения, выносимые на защиту:**

1. На основе математического аппарата теории множеств, с учетом особенностей функционирования систем VI-аналитики, разработана модель организации данных, позволяющая вести сбор, хранение и динамический анализ информации, необходимой для контроля и управления распределенными системами автоматизированной обработки спутниковых данных. Данная модель позволила повысить эффективность управления процессами обработки данных в современных системах спутникового мониторинга.
2. На основе разработанной модели создана программная инфраструктура (в том числе для управления базами данных), обеспечивающая контроль и управление подсистемами

потоковой обработки данных в современных системах дистанционного мониторинга. Разработанная программная инфраструктура позволяет создавать и эффективно развивать интерфейсы анализа данных для управления крупными быстроразвивающимися распределенными системами автоматизированной обработки данных.

3. На основе математического аппарата теории множеств, с учетом особенностей функционирования систем VI-аналитики, разработана модель организации данных для создания интерфейсов (систем) анализа пространственной информации, полученной на основе спутниковых данных:
  - a. модель данных для работы с временными рядами характеристик наблюдаемых процессов (объектов), в частности, пространственной информацией;
  - b. модель данных для работы с периодически обновляемыми картографическими материалами (временными сериями пространственных данных).
4. На основе созданных моделей разработана программная инфраструктура (в том числе для управления базами данных), обеспечившая возможность создания новых динамических интерфейсов для распределенного анализа различных типов информационных продуктов, получаемых на основе данных долговременных спутниковых наблюдений в различных системах дистанционного мониторинга. Создаваемые на основе разработанной программной инфраструктуры динамические интерфейсы позволяют повысить эффективность процессов обработки и анализа результатов обработки спутниковых данных, особенно в проектах, выполняемых распределенными научными коллективами.

**Теоретическая и практическая значимость** диссертации состоит в том, что разработанные модели организации данных и программной инфраструктуры в информационных системах дистанционного мониторинга позволяют создавать, развивать и поддерживать:

- блоки анализа данных, необходимых для контроля и управления распределенных систем автоматизированной обработки спутниковых данных;
- блоки распределенного анализа различных типов информационных продуктов, получаемых на основе данных долговременных спутниковых наблюдений.

На основе представленных в работе разработок созданы и внедрены блоки анализа данных в различные системы дистанционного мониторинга, в том числе:

- «Информационная система дистанционного мониторинга» лесных пожаров Федерального агентства лесного хозяйства (<http://www.nffc.aviales.ru>).
- Система коллективного пользования «ВЕГА-Science», ориентированная на информационную поддержку научных исследований (<http://sci-vega.ru/>).
- Информационная система комплексного дистанционного мониторинга лесов Приморского края «ВЕГА-Приморье» (<http://primorsky.geosmis.ru/>).
- Центр коллективного пользования «ИКИ-Мониторинг» (<http://smiswww.iki.rssi.ru/default.aspx?page=357>).
- Центры приема и обработки спутниковых данных НИЦ «Планета» (<http://moscow.planeta.smislab.ru/>).

**Методы исследования.** Теоретико-методологической основой данной работы являются труды отечественных и зарубежных авторов, посвященные проблемам интеллектуального анализа данных и создания человеко-машинных интерфейсов. В работе на основе комплексного системного анализа процессов и данных, для которых должны быть созданы инструменты анализа, строятся модели данных, позволяющие организовать эффективную работу с информацией. На основе разработанных моделей данных и объектно-ориентированного подхода разрабатываются архитектура организации систем работы с данными и программная инфраструктура, необходимая для ее реализации.

**Реализация и внедрение.** Диссертационная работа выполнена в Институте космических исследований Российской академии наук (ИКИ РАН). Результаты работы использованы при выполнении проекта РАН № 01.20.0.2.00164 (тема "Мониторинг"); проектов Минобрнауки 14.607.21.0122, грантов РФФИ № 13-07-12017, 15-29-07953, 16-29-09615, 13-07-12116; проектов РНФ № 14-17-00389, 16-17-00042.

Результаты работы используются в различных организациях (ИКИ РАН, ФГБУ «НИЦ «Планета», ФБУ «Авиалесоохрана») и системах мониторинга («VEGA-Science», «ИСДМ-Рослесхоз», «VEGA-PRO», «VEGA-GEOGLAM», «Вега-Приморье», ЦКП «ИКИ-Мониторинг»). Акты внедрения представлены в диссертационной работе.

**Апробация работы.** Основные положения диссертации докладывались и обсуждались на всероссийских и международных научно-технических конференциях, посвященных дистанционному зондированию Земли. Всего сделано 17 докладов. Среди них всероссийские конференции: «Современные проблемы дистанционного зондирования Земли из космоса» (Москва, 2013-2017), «Фундаментальные и прикладные космические исследования» (Москва, 2014, 2016-2017); и международные конференции: «Космонавтика. Радиоэлектроника. Геоинформатика» (Москва, 2017), «Региональные проблемы дистанционного зондирования Земли» (Красноярск, 2017).

**Достоверность** результатов диссертации подтверждается успешным внедрением разработанных моделей и программной инфраструктуры в различные научные и прикладные системы дистанционного мониторинга, эффективность которых проверяется и доказывается.

**Публикации.** По результатам диссертационных исследований опубликовано 26 работ. Среди них: 9 статей (5 статей в изданиях по списку ВАК) и 17 тезисов докладов на международных и российских конференциях.

**Структура и объем работы.** Работа состоит из введения, четырех глав, заключения, списка литературы и пяти приложений, содержащих документы о внедрении результатов. Основной текст работы включает 128 страниц, 42 рисунка и 1 таблицу. Список литературы содержит 86 наименований на 13 с.

## ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

*В Главе 1 приведен обзор возможностей использования инструментов, предоставляемых системами BI-аналитики, для построения на их основе распределенных систем анализа данных.*

Одной из основных особенностей таких систем является то, что пользователь формулирует запрос, извлекает агрегированные данные из различных источников и просматривает полученные результаты в WEB-интерфейсе через динамически формируемые интерактивные отчеты. Интерактивные отчеты предоставляют пользователю возможность определять, какие именно данные требуются для решения стоящих перед ним задач анализа, и каким образом эти данные необходимо организовать (агрегировать). Таким образом, с помощью каждой интерактивной отчетной формы, пользователь способен решать широкий спектр аналитических задач (перестраивая её под свои нужды). Для реализации данных возможностей в системах BI-аналитики используются многомерные хранилища данных, опирающиеся на концепцию многомерных кубов (OLAP-кубов (On-Line Analytical Processing – оперативная аналитическая обработка)). Данные, формирующие OLAP куб, делятся на два вида:

- «Измерения» ( $D_i$ ) – дискретные характеристики, описывающие исследуемый объект или процесс, принимающие значения ( $d_{ij}$ ) из ограниченного набора:

$$\forall D_i = \left\{ d_{ij_i} \right\}, j_i = \{1 \dots m_i\}. \quad (1)$$

- «Факты» ( $F_{kd_1j_1d_2j_2\dots d_{nj_n}}$ ) – некоторые показатели (численные значения), которые соответствуют исследуемому объекту или процессу, полученные при фиксированных значениях (диапазонах значений) измерений. Т.е. каждой комбинации из значений  $(d_{1j_1}, d_{2j_2}, \dots, d_{nj_n})$  измерений  $(D_1, D_2, \dots, D_n)$ , соответствует определенный набор значений фактов:

$$\forall \left( d_{1j_1}, d_{2j_2}, \dots, d_{nj_n} \right), \exists F_{d_1j_1d_2j_2\dots d_{nj_n}} = \left\{ f_{kd_1j_1d_2j_2\dots d_{nj_n}} \right\}, k = \{1 \dots l\}. \quad (2)$$

$\begin{matrix} 1 \leq j_1 \leq m_1 \\ 1 \leq j_2 \leq m_2 \\ \vdots \\ 1 \leq j_n \leq m_n \end{matrix}$

OLAP куб (Рисунок 1) представляет собой систему координат, осями которой являются "Измерения". Значения "Измерений" являются координатами связанного с ними набора "Фактов". Достоинством OLAP кубов является то, что агрегированные значениями "Фактов", соответствующие некоторому диапазону "Измерений", в отличие от данных, располагающихся в плоских реляционных таблицах, могут вычисляться предварительно. Это позволяет достаточно быстро проводить агрегацию информации по разным "Измерениям". К недостаткам относится трудность их модификации, так как при необходимости добавления дополнительных измерений требуется физическая перестройка всего OLAP куба и достаточно большой объем вспомогательной технической информации, необходимой для реализации многомерности данных.

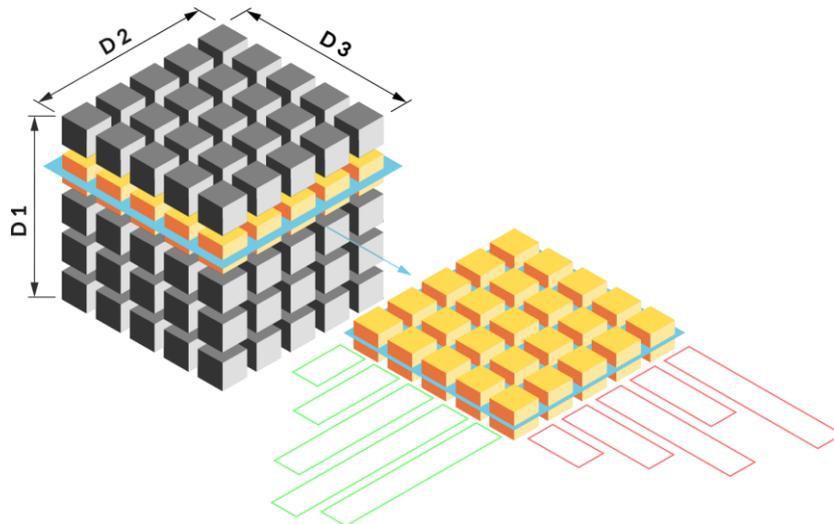


Рисунок 1 - Представление данных в виде 3-х мерного OLAP куба ( $i = 3$ ) и его среза, при фиксированном значении измерения D1

В главе также рассматриваются особенности, современные подходы и тенденции организации комплексов потоковой обработки данных в современных распределенных системах дистанционного мониторинга. При этом выделены основные уровни работы с данными, которые обычно присутствуют в таких комплексах. К ним относятся: *уровень получения данных*, осуществляющий получение данных из различных источников; *уровень хранения данных*, обеспечивающий ведение архивов спутниковых данных и результатов их обработки; *уровень подготовки данных*, обеспечивающий формирование наборов данных для конкретных процедур обработки; *уровень обработки данных*, обеспечивающий выполнение процедур обработки, в том числе и с использованием распределенных вычислительных ресурсов. На основе системного анализа процессов, возникающих при выполнении

поточковой автоматизированной обработки данных, выделены задачи, стоящие перед системой контроля и управления, для решения которых могут эффективно использоваться решения, создаваемые на основе VI-технологий.

В главе также проведен анализ особенностей различных типов информационных продуктов, получаемых на основе спутниковых данных, при решении задач долговременного мониторинга различных пространственных процессов и явлений. Проанализированы различные подходы организации хранения таких информационных продуктов и построения интерфейсов, обеспечивающих проведение анализа данных. Выделено два наиболее распространенных подхода к организации хранения данных: организация хранения наборов пространственных данных (карт, или их слоев), построенных для определенных территорий в фиксированные моменты времени, и организация БД, в которых хранится информация с описанием состояния (характеристик) конкретных объектов на различные моменты времени. Проводится анализ подходов, используемых для создания интерфейсов, обеспечивающих возможности распределенного анализа информации, предоставляемой подобными БД. Отмечены основные недостатки используемых подходов, связанные с необходимостью создания достаточно сложных индивидуальных запросов для получения конкретных наборов данных и, как следствие, значительными ограничениями в функциональности интерфейсов работы с данными. Определены задачи, которые наиболее часто возникают при анализе данных долговременных наблюдений при изучении различных пространственных процессов и явлений, для решения которых могут эффективно использоваться VI-технологии.

Проведен комплексный анализ возможностей построения схем организации работы с данными на основе VI-технологий для построения интерфейсов анализа информации, получаемой при организации контроля и управления системами потоковой обработки данных и долговременных рядов информации, полученной на основе спутниковых наблюдений. На основе проведенного анализа предложена базовая схема построения таких интерфейсов (см. Рисунок 2). Предложенная схема фактически стала основой, для разработки моделей данных, представленных в главах 2 и 3 настоящей работы.

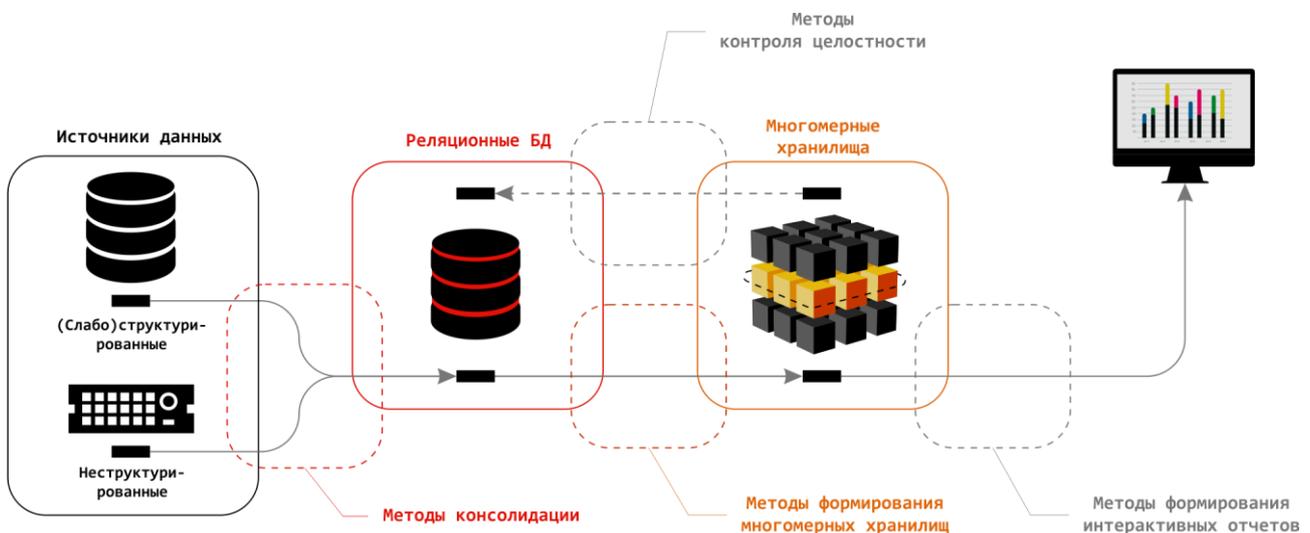


Рисунок 2 - Базовая функциональная схема систем VI-аналитики

*Глава 2 посвящена разработке моделей и программной инфраструктуры, необходимых для построения системы контроля и управления работоспособностью распределенных комплексов потоковой обработки спутниковых данных с использованием VI-технологий.*

Для этого предлагается ввести в такие системы дополнительный уровень контроля и управления (см. Рисунок 3), включающий следующие блоки: блок автоматизированного сбора данных, который обеспечивает документирование величин параметров, характеризующих работу уровней подготовки и обработки данных; блок хранения и

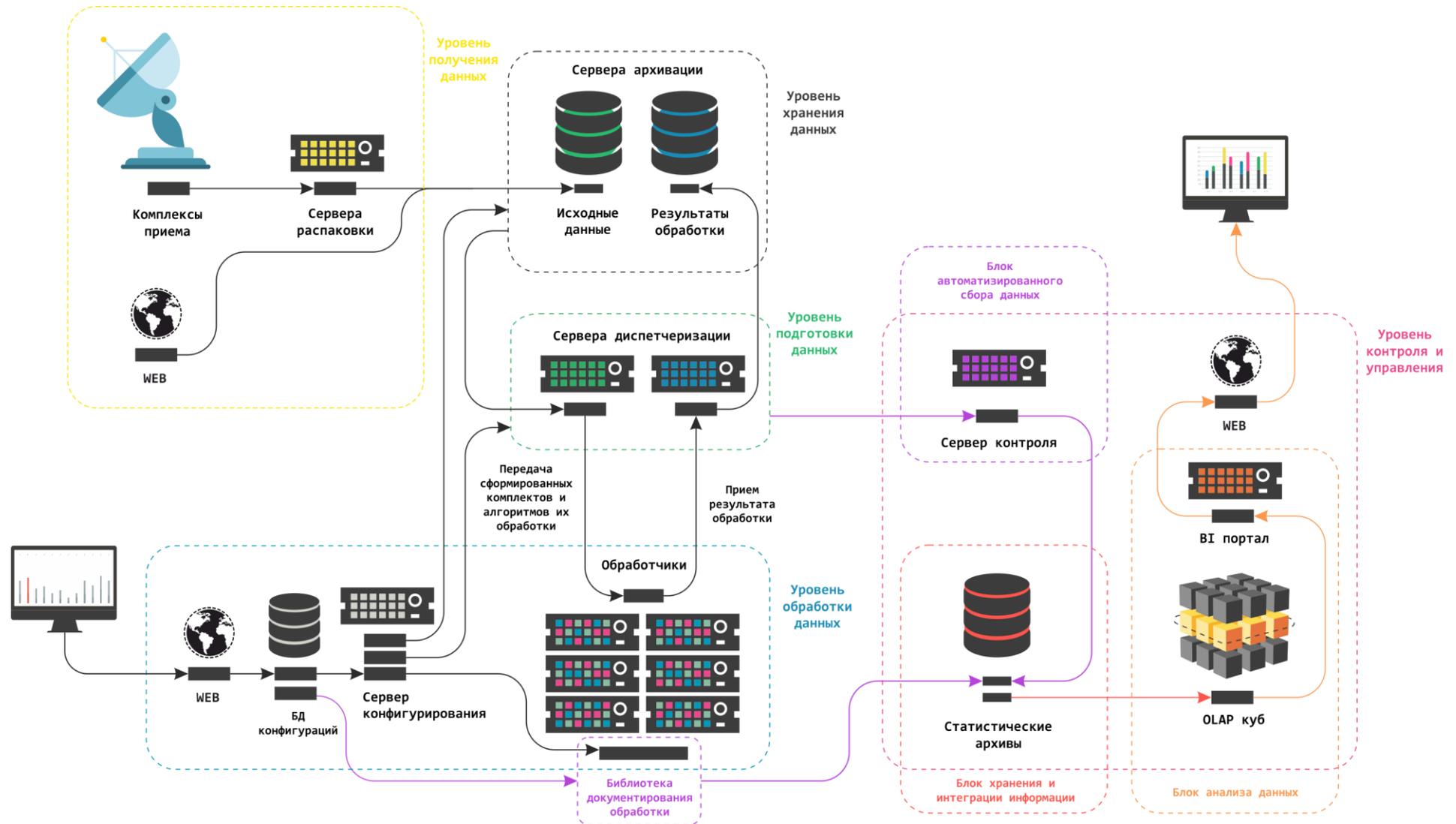


Рисунок 3 - Блок-схема, описывающая работу комплекса автоматической обработки спутниковых данных, включающего уровень контроля и управления

*интеграции информации*, обеспечивающий ведение реляционной БД для хранения собираемой информации о работе системы; *блок анализа данных*, который осуществляет преобразование информации, полученной из реляционных таблиц блока хранения и интеграции в многомерные OLAP-кубы, и строит на их основе интерактивные отчетные формы, предоставляя доступ к ним посредством WEB-интерфейса.

Рассматриваются особенности функционирования *уровней подготовки и обработки данных*, и разрабатываются их математические модели. Уровень подготовки данных осуществляет формирование множества исходных комплектов файлов  $kit_{ij}$ , для каждого типа обработки  $task_i$  (где  $task_i$  - множество параметров, характеризующих конкретный тип обработки). Каждому комплекту файлов соответствует флаг  $kf_{ij}$ , необходимый для информирования остальных блоков системы о стадиях обработки комплекта. В зависимости от значения флага комплект входит в одно из следующих множеств:

- $R_i$  – множество комплектов, готовых к обработке,
- $I_i$  – множество обрабатываемых комплектов,
- $F_i$  – множество удачно обработанных комплектов,
- $E_i$  – множество ошибочных комплектов.

Комплекты файлов обрабатываются посредством вычислительных узлов  $h()_k$ . Вычислительные узлы могут обрабатывать только те задания, которые входят в их конфигурацию  $\left\{ task_{i_k} \right\}_{1 \leq i_k \leq n_k}$  (являющуюся подмножеством всего множества видов обработки

$Task$ ). Процедуры управления автозапуском определяют сколько комплектов определенного типа заданий вычислительный узел может обработать одновременно. Также для каждого типа обработки определен комплект скриптов, отвечающих за организацию потока обработки комплекта ( $t_i$ ) и за саму обработку ( $\left\{ p_q() \right\}_i$ ). Таким образом,

математическая модель функционирования уровней подготовки и обработки данных выглядит следующим образом:

$$\left\{ \begin{array}{l} \forall task_i, \left\{ \begin{array}{l} \exists (R, I, F, E)_i; \\ \exists \left\{ \left\{ p_q() \right\}_i, t_i \right\}; \end{array} \right. \\ \forall h()_k, \exists \left\{ task_{i_k} \right\} \subseteq Task = \left\{ task_i \right\}. \end{array} \right. \quad (3)$$

На основе вышеприведенной математической модели разработаны метод консолидации статистической информации, характеризующей уровни подготовки и обработки данных, а также специализированная реляционная БД для их хранения. Так как реальные процессы  $process_r$ , происходящие на вычислительном узле, формируют потоки обработки  $thread_q$ , совокупность которых составляют сессии  $s_k$ , а каждому вычислительному узлу соответствует некоторое количество типов обработки (комплекты которых он может обрабатывать), то модель хранения документируемых параметров выглядит следующим образом,



Для этого предлагается ввести в таких системах дополнительный уровень анализа данных, включающий следующие блоки: *блок консолидации результатов обработки*, обеспечивающий структурирование различных видов пространственной информации, получаемой на основе спутниковых данных; *блок анализа данных*, полностью соответствующий одноименному блоку, описанному в предыдущей главе, но использующий в качестве источника данные, прошедшие консолидацию в предыдущем блоке.

Приведены особенности различных анализируемых типов информационных продуктов, полученных на основе спутниковых данных. Набор пространственных данных (карт или их слоев), построенных для определенных территорий в фиксированные моменты времени  $y_i$ , представлен отдельными локальными файлами  $\{f_i(n_1, n_2)\}$  (где  $f_i(n_1, n_2)$  – функция

$$\begin{matrix} 0 \leq n_1 \leq N_1 \\ 0 \leq n_2 \leq N_2 \end{matrix}$$

значения пикселя с координатами  $n_1$  по горизонтали и  $n_2$  по вертикали):

$$\forall y_i, \exists \{f_i(n_1, n_2)\} \Rightarrow \left\{ \begin{matrix} f_i(0,0) & f_i(1,0) & \dots & f_i(N_1,0) \\ f_i(0,1) & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ f_i(0,N_2) & \dots & \dots & f_i(N_1,N_2) \end{matrix} \right\}. \quad (6)$$

Временные ряды  $dt_k$  характеристик, наблюдаемых объектов и явлений  $obj_i$  располагаются в слабоструктурированных реляционных БД (где  $(cp_1, cp_2, \dots, cp_m)$  – не изменяющиеся во времени характеристики, а  $(vp_{1k}, vp_{2k}, \dots, vp_{hk})$  – изменяющиеся во времени характеристики):

$$\forall obj_i, \left\{ \begin{matrix} \exists (cp_1, cp_2, \dots, cp_m)_i; \\ \exists \{ (dt_k, vp_{1k}, vp_{2k}, \dots, vp_{hk}) \}_i. \end{matrix} \right. \quad (7)$$

Для каждого типа информационного продукта на основе их формализованных моделей разработан комплекс методов консолидации, направленных на извлечение данных из источников, обеспечение им необходимого уровня качества (отчистка от факторов, не позволяющих производить их корректный анализ) и информативности (обогащение дополнительной информацией), а также преобразование в единый формат.

Особенность консолидации временных рядов характеристик наблюдаемых объектов и явлений заключается в необходимости обогащения их информацией об изменяющихся во времени характеристиках, которые объект  $obj_i$  потенциально мог бы иметь в те моменты времени  $dt'_q$ , в которые его невозможно было наблюдать, чтобы избежать разрывов и получить непрерывный временной ряд наблюдений:

$$\left\{ \begin{matrix} \{dt_k\}_i \subseteq \{dt'_q\}_i; \\ \forall dt'_q, \left\{ \begin{matrix} dt'_{qi} < dt_{1i} \Rightarrow k = 0; \\ dt'_{qi} \geq dt_{1i} \Rightarrow \forall dt'_{qi}, \exists \{ (vp_{1i}, vp_{2i}, \dots, vp_{hi}) \}_i; \\ dt'_{qi} = dt_{k+1,i} \Rightarrow \left\{ \begin{matrix} \forall dt'_{qi}, \exists \{ (vp_{1,k+1}, vp_{2,k+1}, \dots, vp_{h,k+1}) \}_i; \\ k = k + 1; \end{matrix} \right. \\ dt'_{qi} \geq dt_{ki} \wedge dt'_{qi} < dt_{k+1,i} \Rightarrow \forall dt'_{qi}, \exists \{ (vp_{1k}, vp_{2k}, \dots, vp_{hk}) \}_i. \end{matrix} \right. \end{matrix} \right. \quad (8)$$

В случае с пространственной информацией, полученной на основе периодически обновляемых картографических материалов, особенность консолидации заключается в структурировании данного типа информационных продуктов. Только структурированные источники могут быть использованы для наполнения многомерных хранилищ, обеспечивающих работу систем ВІ-аналитики, а также для обогащения информацией о принадлежности пикселей картографического материала к различным пространственным



формировать, сохранять и обновлять (по мере актуализации структурированных данных в фоновом режиме) эти кубы, а также необходимые для функционирования интерфейсов анализа различной пространственной информации, получаемой на основе долговременных спутниковых наблюдений.

В Главе 4 описываются реальные примеры внедрения разработанных моделей и программной инфраструктуры при построении блоков анализа данных в различных системах дистанционного мониторинга и комплексах потоковой обработки спутниковых данных.

В главе описываются основные возможности и особенности:

- Систем контроля и управления комплексами потоковой обработки спутниковых данных в Центре коллективного пользования «ИКИ-Мониторинг» (<http://smiswww.iki.rssi.ru/default.aspx?page=357>) и в центрах приема и обработки НИЦ «Планета» (<http://moscow.planeta.smlab.ru/>), которые обеспечивают работу с результатами обработки спутниковых данных ведущих центров Федерального агентства по гидрометеорологии и мониторингу окружающей среды.
- Блока VI-аналитики информационной системы дистанционного мониторинга Федерального агентства лесного хозяйства (<http://www.nffc.aviales.ru>), предназначенного для комплексного информационного обеспечения мониторинга пожарной опасности, контроля распространения лесных пожаров, а также оценки их последствий на территории Российской Федерации.
- Блока VI-аналитики системы коллективного пользования Vega-Science (<http://sci-vega.ru/>), ориентированного на информационную поддержку научных исследований, позволяющих пользователям проводить динамический анализ ежегодно обновляющейся информации о растительном покрове на территории Российской Федерации, полученной на основе данных спутниковой системы Terra-MODIS.
- Блока VI-аналитики информационной системы "ВЕГА-Приморье" (<http://primorsky.geosmis.ru/>) ведения комплексного дистанционного мониторинга лесов Приморского края, предназначенного для комплексного информационного обеспечения мониторинга пожарной обстановки в крае и динамического анализа ежегодно обновляющейся информации о растительном покрове на его территории.

В главе подробно описываются возможности и особенности различных динамических web-интерфейсов, обеспечивающих анализ информации в системах контроля и управления распределенной потоковой обработкой спутниковых данных и анализа информационных продуктов, получаемых на основе данных долговременных спутниковых наблюдений. Примеры таких интерфейсов приведены на Рисунок 4, Рисунок 5 и Рисунок 6.



Рисунок 4 - Интерактивный инструмент анализа количества пожаров, произошедших на территории России, по годам

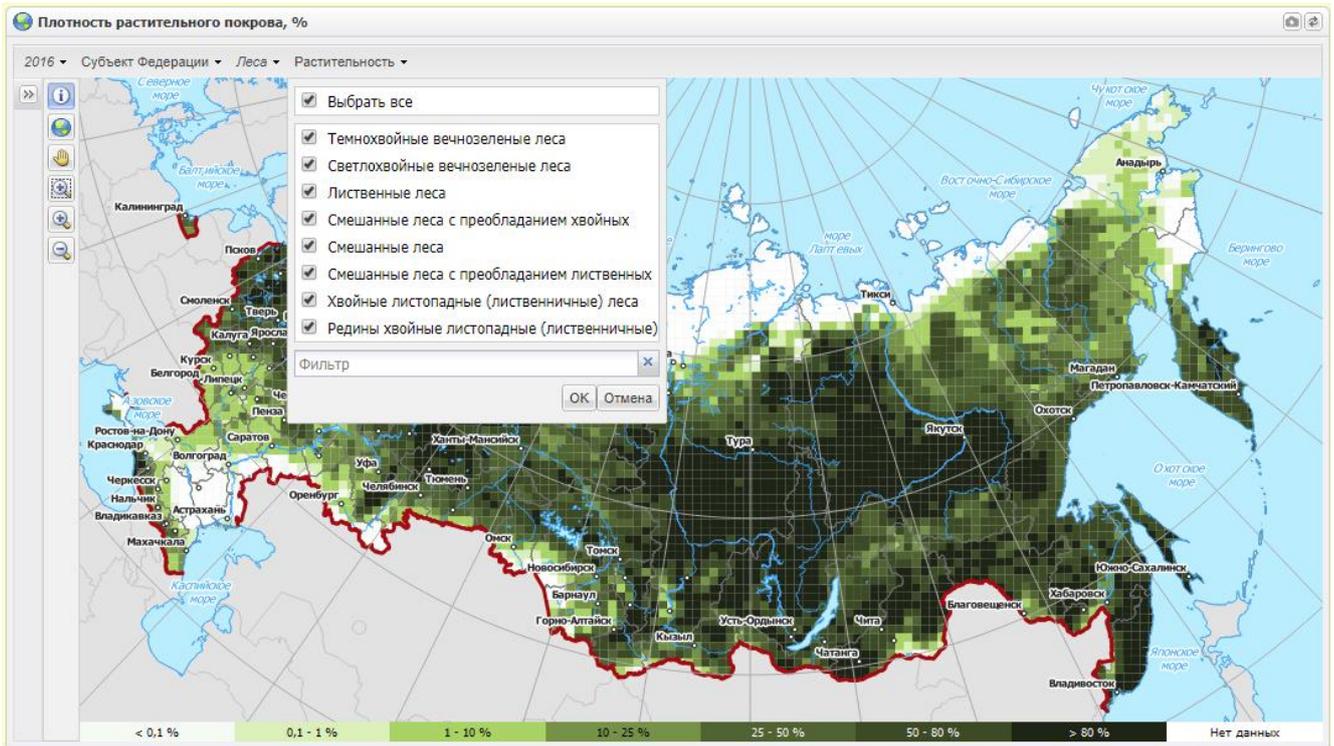


Рисунок 5 - Интерактивный инструмент анализа удельной площади растительного покрова Российской Федерации (по данным 2016 года)

Растительный покров по регионам

Факты

Год ↓

| Субъект Федерации       | Группа | Растительность                            | 2016        |             | 2015        |             | 2014        |             | 2013        |
|-------------------------|--------|---|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
|                         |        |   | Площадь, га | Удельная, % | Площадь, га | Удельная, % | Площадь, га | Удельная, % | Площадь, га |
| Республика Башкортостан |        |   |             |             | 6 018 581   | 42,2        | 5 964 956   | 41,8        | 5           |
| Республика Бурятия      |        |   |             |             | 19 095 503  | 57,4        | 20 540 599  | 61,8        | 20          |
| Республика Дагестан     |        |   |             |             | 338 803     | 6,6         | 362 645     | 7,1         |             |
| Республика Ингушетия    |        |   |             |             | 91 231      | 27,4        | 95 596      | 28,7        |             |
| Республика Калмыкия     |        |   |             |             | 323         | 0,0         | 175         | 0,0         |             |
| Республика Карелия      |        |   |             |             | 10 714 969  | 68,6        | 10 894 464  | 69,7        | 10          |
| Республика Коми         | Итого  |   |             |             | 30 457 984  | 73,1        | 30 451 457  | 73,1        | 30          |
|                         | Леса   | Итого                                     | 30 741 941  | 73,8        | 30 457 984  | 73,1        | 30 451 457  | 73,1        | 30          |
|                         |        | Темнохвойные вечнозеленые леса            | 13 452 063  | 32,3        | 13 561 465  | 32,5        | 14 171 899  | 34,0        | 13          |
|                         |        | Светлохвойные вечнозеленые леса           | 7 736 101   | 18,6        | 7 471 374   | 17,9        | 7 150 826   | 17,2        | 7           |
|                         |        | Лиственные леса                           | 2 212 098   | 5,3         | 2 033 719   | 4,9         | 1 950 005   | 4,7         | 1           |
|                         |        | Смешанные леса с преобладанием хвойных    | 3 600 623   | 8,6         | 3 823 273   | 9,2         | 3 636 923   | 8,7         | 3           |
|                         |        | Смешанные леса                            | 2 130 595   | 5,1         | 2 045 172   | 4,9         | 2 105 013   | 5,1         | 1           |
|                         |        | Смешанные леса с преобладанием лиственных | 1 563 745   | 3,8         | 1 486 416   | 3,6         | 1 406 495   | 3,4         | 1           |

Рисунок 6 - Интерактивный инструмент анализа динамики растительного покрова в республике Коми по типам растительности

## ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

1. Проведен анализ особенностей, подходов и тенденций к организации распределенных комплексов потоковой обработки спутниковых данных и сформулированы основные проблемы проведения эффективного анализа информации о состоянии элементов таких комплексов и выполняющихся в них процессов обработки данных.
2. С учетом проведенного анализа и использования подходов организации работы с агрегированной информацией разработана модель данных для построения эффективных динамических интерфейсов, необходимых для контроля и управления распределенных систем автоматизированной обработки спутниковой информации.
3. На основе разработанной модели данных создана программная инфраструктура для разработки блоков контроля и управления подсистемами потоковой обработки данных в современных системах дистанционного мониторинга.
4. Проведен анализ особенностей, подходов и тенденций к организации работы с различными типами информационных продуктов, получаемых на основе долговременного постоянного спутникового мониторинга, и сформулированы основные проблемы, связанные с проведением эффективного анализа результатов обработки спутниковых данных (в первую очередь долговременных рядов наблюдений).
5. На основе проведенного анализа с использованием различных подходов и методов агрегации информации разработаны модели организации данных для построения интерфейсов анализа временных рядов характеристик наблюдаемых процессов (объектов) и периодически обновляемых картографических материалов (временных серий пространственных данных).
6. Разработаны необходимые для реализации предложенных моделей данных методы агрегирования временных рядов характеристик наблюдаемых процессов (объектов) и временных серий пространственных данных.
7. На основе разработанных моделей и методов агрегирования данных создана программная инфраструктура (в том числе для управления базами данных) для организации динамических интерфейсов распределенного анализа различных типов информационных продуктов, получаемых на основе данных долговременных спутниковых наблюдений в системах дистанционного мониторинга.
8. При внедрении разработанных моделей и программной инфраструктуры в различные научные и прикладные системы дистанционного мониторинга показана их эффективность.

## ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

### Тезисы докладов на международных и всероссийских конференциях

1. **Кобец, Д.А.** Построение инструментов для контроля работоспособности и управления распределенными комплексами автоматической потоковой обработки спутниковых данных / *Д.А. Кобец* // 7-я Международная научно-техническая конференция К.Э. Циолковский : к 160 летию со дня рождения. Космонавтика. Радиоэлектроника. Геоинформатика : тезисы докладов (Рязань. 4-6 октября 2017 г.). – Рязань, 2017. – С. 353-357
2. **Кобец, Д.А.** VI-аналитика в анализе статистики работы процессов обработки спутниковых данных / *Д.А. Кобец, А.М. Матвеев, А.А. Мазуров, А.В. Кашицкий, М.А. Бурцев, А.А. Прошин* // XIII Конференция молодых ученых посвященная дню космонавтики «Фундаментальные и прикладные космические исследования» : тезисы докладов (Москва, 13-15 апреля 2016 г.). – Москва : ИКИ РАН, 2016. – С. 43.

3. **Кобец, Д.А.** VI-инструменты для анализа карт растительного покрова в информационных сервисах «Созвездие-ВЕГА» / *Д.А. Кобец, В.А. Толпин, И.В. Балашов* // Четырнадцатая Всероссийская открытая конференция «Современные проблемы дистанционного зондирования Земли из космоса» : тезисы докладов (Москва, 14-18 ноября 2016.). – Москва : ИКИ РАН, 2016. – С. 99.
4. **Кобец, Д.А.** Инструменты анализа спутниковых наблюдений природных пожаров на различных типах территорий / *Д.А. Кобец, И.В. Балашов* // Пятнадцатая Всероссийская открытая конференция «Современные проблемы дистанционного зондирования Земли из космоса» : тезисы докладов (Москва, 13 - 17 ноября 2017 г.). – Москва : ИКИ РАН, 2017. – С. 393.
5. **Кобец, Д.А.** Построение на основе VI-технологий инструментов анализа информации о состоянии растительного покрова в сервисах «Созвездие-ВЕГА» / *Д.А. Кобец, И.В. Балашов, И.Г. Сычугов, В.А. Толпин* // XIV Конференция молодых ученых посвященная дню космонавтики «Фундаментальные и прикладные космические исследования» : тезисы докладов (Москва, 12-14 апреля 2017 г.) – Москва : ИКИ РАН, 2017. – С. 142.
6. **Кобец, Д.А.** Управление работой крупных распределенных комплексов автоматической потоковой обработки спутниковых данных / *Д.А. Кобец, А.А. Мазуров, А.М. Матвеев, М.А. Бурцев, А.А. Прошин* // Четырнадцатая Всероссийская открытая конференция «Современные проблемы дистанционного зондирования Земли из космоса» : тезисы докладов (Москва, 14-18 ноября 2016 г.). – Москва : ИКИ РАН, 2016. – С. 98.

Статьи в изданиях по списку ВАК

7. **Кобец, Д.А.** Использование VI-технологий для создания инструментов для анализа данных спутникового мониторинга / *Д.А. Кобец, И.В. Балашов, И.Д. Данилов, Е.А. Лупян, И.Г. Сычугов, В.А. Толпин* // Современные проблемы дистанционного зондирования Земли из космоса. – 2015. – Т. 12. – № 4. – С. 17-27.
8. **Кобец, Д.А.** Организация автоматизированной многопоточной обработки спутниковой информации в системах дистанционного мониторинга / *Д.А. Кобец, А.М. Матвеев, А.А. Мазуров, А.А. Прошин* // Современные проблемы дистанционного зондирования Земли из космоса. – 2015. – Т. 12. – № 1. – С. 145-155.
9. **Кобец, Д.А.** Организация контроля и анализа работоспособности систем автоматизированной обработки спутниковых данных с использованием VI-технологий / *Д.А. Кобец, И.В. Балашов, И.Г. Сычугов, В.А. Толпин* // Современные проблемы дистанционного зондирования Земли из космоса. Scopus. – 2017. – Т. 14. – № 3. – С. 92-103.
10. **Гордеев, Е.И.** ИС «VolSatView» : комплексный анализ данных об эксплозивных извержениях вулканов Камчатки / *Е.И. Гордеев, О.А. Гурина, Е.А. Лупян, А.А. Сорокин, Д.В. Мельников, А.Г. Маневич, И.М. Романова, Л.С. Крамарева, В.Ю. Ефремов, Д.А. Кобец, А.В. Кашицкий, А.Л. Верхотуров, М.А. Бурцев* // Вестник Дальневосточного отделения Российской академии наук. – 2016. – № 5. – С. 397-410.
11. **Лупян, Е.А.** Создание технологий построения информационных систем дистанционного мониторинга / *Е.А. Лупян, И.В. Балашов, М.А. Бурцев, В.Ю. Ефремов, А.В. Кашицкий, Д.А. Кобец, Ю.С. Крашенинникова, А.А. Мазуров, Р.Р. Назиров, А.А. Прошин, И.Г. Сычугов, В.А. Толпин, И.А. Уваров, Е.В. Флитман* // Современные проблемы дистанционного зондирования Земли из космоса. – 2015. – Т. 12. – № 5. – С. 53-75.

Статьи в сборниках

12. **Гордеев, Е.И.** Комплексный анализ данных об эксплозивных извержениях вулканов Камчатки в ИС VolSatView / *Е.И. Гордеев, О.А. Гурина, Е.А. Лупян, А.А. Сорокин, Д.В.*

*Мельников, А.Г. Маневич, И.М. Романова, Л.С. Крамарева, В.Ю. Ефремов, Д.А. Кобец, А.В. Кашицкий, С.П. Королев, М.А. Бурцев, С.Б. Самойленко* // Материалы региональной научной конференции «Вулканизм и связанные с ним процессы», посвящённой Дню вулканолога (29-30 марта 2016 г.). – Петропавловск-Камчатский : ИВиС ДВО РАН, 2016. – С. 53-64.

13. **Кобец, Д.А.** Контроль работоспособности и управление распределенными комплексами автоматической потоковой обработки спутниковых данных / *Д.А. Кобец, А.М. Матвеев, А.А. Прошин, А.А. Мазуров* // Материалы IV Международной научной конференции «Региональные проблемы дистанционного зондирования Земли» (Красноярск, 12-15 сентября 2017 г.). – Красноярск : СФУ, 2017. – С. 49-51.
14. **Кобец, Д.А.** Построение на основе VI-технологий инструментов анализа информации о состоянии лесов, получаемой на основе данных спутниковых наблюдений / *Д.А. Кобец, И.В. Балашов, И.Г. Сычугов, В.А. Толпин* // Аэрокосмические методы и геоинформационные технологии в лесоведении, лесном хозяйстве и экологии : Доклады VI Всероссийской конференции (Москва, 20-22 апреля 2016 г.). – Москва : ЦЭПЛ РАН, 2016. – С. 24-28.
15. *Лузян, Е.А.* Новые возможности технологий построения информационных систем дистанционного мониторинга / *Е.А. Лузян, И.В. Балашов, М.А. Бурцев, В.Ю. Ефремов, А.В. Кашицкий, Д.А. Кобец, Ю.С. Крашенинникова, А.А. Мазуров, Р.Р. Назиров, А.А. Прошин, И.Г. Сычугов, В.А. Толпин, И.А. Уваров, Е.В. Флитман* // Региональные проблемы дистанционного зондирования Земли : материалы III Международной научной конференции «Региональные проблемы дистанционного зондирования Земли» (Красноярск, 13-16 сентября 2016 г.) / Научный редактор Е.А. Ваганов; ответственный редактор М.В. Носков. – Красноярск : Сиб. федер. ун-т, 2016. – С. 24-28.

**Кобец Дмитрий Александрович**

**МОДЕЛИ И ПРОГРАММНАЯ ИНФРАСТРУКТУРА ДЛЯ  
ПРОЕКТИРОВАНИЯ ИНТЕРФЕЙСОВ КОНТРОЛЯ И УПРАВЛЕНИЯ  
РАСПРЕДЕЛЕННОЙ ОБРАБОТКОЙ И АНАЛИЗА СПУТНИКОВЫХ ДАННЫХ  
НА ОСНОВЕ VI-ТЕХНОЛОГИЙ**

**Автореферат**

диссертации на соискание ученой степени  
кандидата технических наук

055(02)2 ИКИ РАН  
Москва, 117997, Профсоюзная ул., 84/32  
Подписано к печати «\_\_» \_\_\_\_\_ 201\_ г  
Заказ \_\_\_\_ Формат 60×84/16 Тираж 120