

Информационная система приема, обработки и доступа к спутниковым данным и ее применение для решения задач мониторинга окружающей среды

Ю. И. Шокин¹, Н. Н. Добрецов^{1,3}, Е. А. Мамаш¹, В. А. Кихтенко^{1,2},
П. В. Воронина^{1,2}, В. В. Смирнов¹, Д. Л. Чубаров^{1,*}

¹Институт вычислительных технологий СО РАН, Новосибирск, Россия

²Новосибирский государственный университет, Россия

³Институт геологии и минералогии СО РАН им. В.С. Соболева, Новосибирск, Россия

*Контактный e-mail: dchubarov@ict.nsc.ru

Рассмотрена информационная система для поддержки научных исследований с использованием данных спутникового мониторинга, в основе которой лежит технология представления файлового архива данных дистанционного зондирования Земли как единого многомерного информационного пространства — технология hVault. Возможности этой системы рассмотрены на примере решения нескольких прикладных задач. Продемонстрирована эффективность системы при работе с большими объемами геопространственных данных.

Ключевые слова: дистанционное зондирование Земли, MODIS, тепловые аномалии, геопространственные данные, метод RST (Robust Satellite Techniques).

Введение

В настоящее время привлечение больших объемов разнородной, в том числе спутниковой, информации в процессе изучения природной и окружающей среды становится практически нормой. Часто это сопряжено со значительными трудностями, связанными с разнообразием форматов хранения данных и вычислительной сложностью алгоритмов их анализа. Существует несколько подходов к работе с данными дистанционного зондирования Земли, обусловленных различием в организации приема, моделях хранения, способах обработки и сервисах доступа.

В рамках этого направления Институтом вычислительных технологий СО РАН на протяжении нескольких лет развивается информационная инфраструктура, обеспечивающая хранение, архивацию и пользовательский доступ к данным дистанционного зондирования и наземным наблюдениям [1]. В результате были созданы программный комплекс обработки данных спутникового мониторинга в оперативном режиме и архив данных дистанционного зондирования Земли, составляющие *распределенную инфраструктуру спутникового мониторинга* (РИСМ) [2, 3]. Эти работы направлены на обеспечение информационной поддержки научно-исследовательской деятельности и предоставление научным коллективам данных о состоянии и динамике окружающей среды и социально-экономических процессов. Однако эффективное использование

созданной инфраструктуры затруднено отсутствием удобного инструмента доступа к данным, пригодного для непосредственной работы с ними и их интеграции в различные информационные системы. Для решения этой задачи была предложена технология hVault виртуальной интеграции данных дистанционного зондирования, представленных в архиве, на основе реляционной СУБД. Эта технология позволяет реализовать сложные алгоритмы анализа временных рядов и пространственных распределений средствами СУБД, оснащенной специализированными расширениями [4].

Эффективность применения технологии демонстрируется в статье при выполнении исследований состояния природной среды, требующих обработки большого по пространственному и временному охвату объема спутниковых наблюдений. Технология позволяет реализовывать различные алгоритмы для анализа изменений распределения физических величин, характеризующих состояние подстилающей поверхности и атмосферы.

В статье обсуждаются результаты решения трех модельных задач. Первые две задачи связаны с анализом особенностей распределения температуры поверхности Земли: исследуются проявления локализованной в пространстве тепловой аномалии, обнаруженной на территории Семипалатинского полигона, и изучается вопрос о существовании временных тепловых аномалий на территории Горного Алтая в связи с Чуйским землетрясением, произошедшим 27.09.2003 г. Третьей модельной задачей является определение связи между пространственным распределением очагов природных пожаров и регистрируемых областей с повышенной яркостной температурой. Решение этих задач требует привлечения методов пространственного анализа, автоматической кластеризации и других алгоритмов, обладающих высокой вычислительной сложностью.

1. Технические решения для работы с архивом данных дистанционного зондирования Земли

Основные компоненты разрабатываемой информационной системы представлены на архитектурной диаграмме (рис. 1). Информационная система состоит из трех блоков. К *блоку приема и обработки* относятся элементы распределенной инфраструктуры спутникового мониторинга. Этот блок включает программные модули для передачи данных с приемных комплексов и сигнализации о получении новых данных спутникового мониторинга, а также подсистему их глубокой обработки, выполняющую геометрическую и радиометрическую коррекцию и производящую широкий ассортимент информационных продуктов уровней L2 и L3 по классификации Earth Observing System [5]. Все поступающие данные, а также информационные продукты, полученные на их основе, сохраняются в файловом архиве. Для обеспечения необходимых уровней производительности и надежности архива используется горизонтально масштабируемая система хранения, описанию которой посвящен разд. 1.1. По мере поступления данных инициируется их обработка в потоковом режиме на вычислительном кластере.

Блок управления базами данных построен на основе технологии hVault. Суть этой технологии состоит в представлении файлового архива в виде набора виртуальных таблиц базы данных, в которых каждой записи отвечает отдельное измерение. Перечень атрибутов измерения зависит от типа данных:

- для данных спутниковых сенсоров — значения в спектральных каналах для отдельных пикселей спутникового снимка и условия съемки;

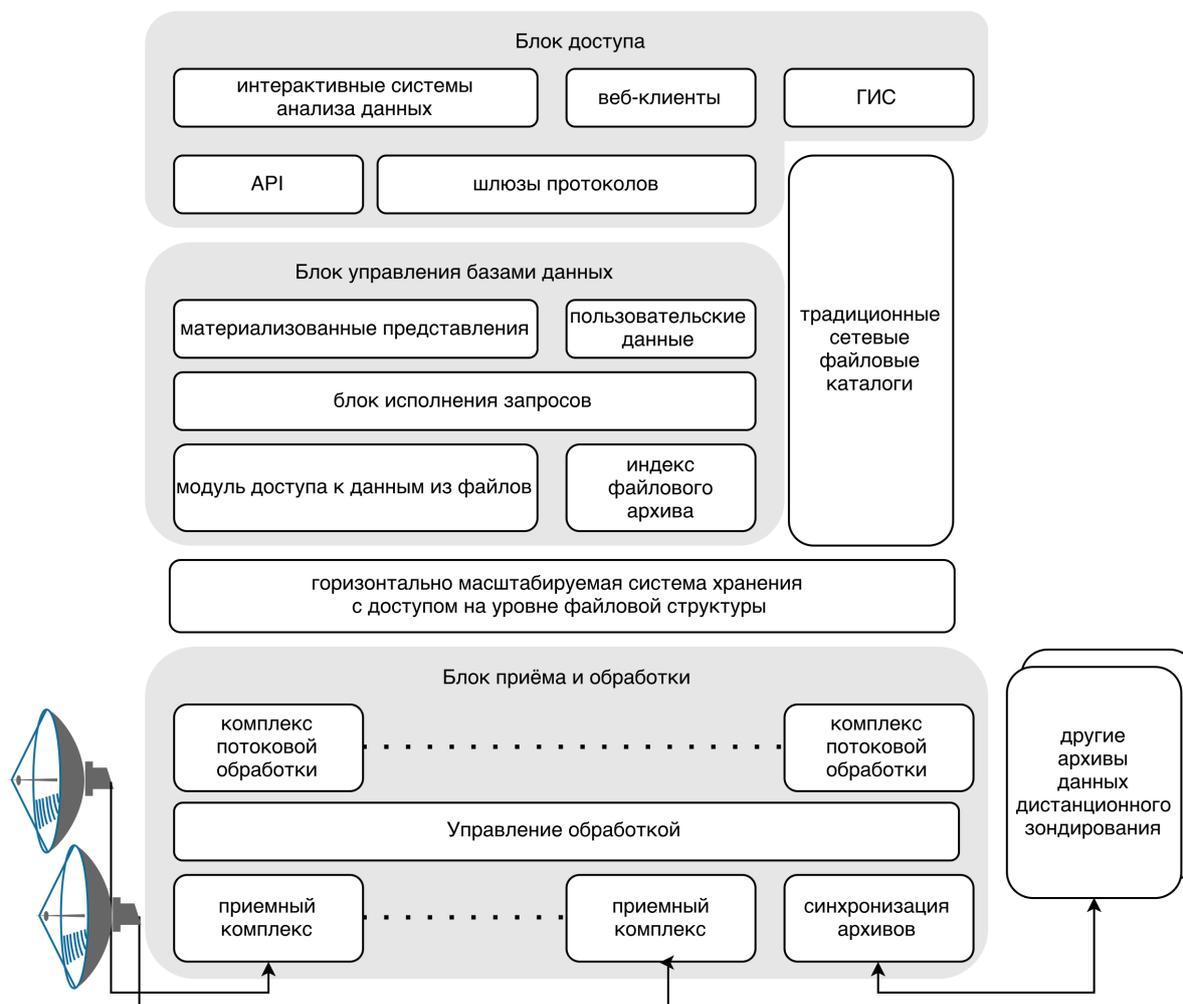


Рис. 1. Архитектурная диаграмма информационной системы приема, обработки и доступа к данным дистанционного зондирования

- для информационных продуктов высокого уровня обработки — набор физических величин, восстановленных по результатам спутниковых измерений на элементе поверхности и оценки точности восстановления;
- для результатов численного моделирования — значения в ячейке сетки и отдельные параметры модели.

Таким образом, работа с отдельными файлами снимков заменяется на работу с базой данных, содержащей сразу все данные архива с возможностью их извлечения и преобразования посредством запросов на языке SQL. Эта технология реализована с помощью расширения СУБД PostgreSQL.

Использование стандартного языка запросов SQL предоставляет исследователям большое количество различных программных инструментов как для извлечения данных, так и для их анализа средствами СУБД. К ним относятся:

- интерактивные системы анализа данных, такие как Jupyter (интерактивная оболочка на основе интерпретатора IPython);
- языки программирования, поддерживающие работу с базами данных;
- настольные ГИС и др.

За интеграцию с системами, использующими для доступа к данным специализированные протоколы, отвечают шлюзы протоколов. На основе технологии hVault реализован веб-сервер протокола WMS (Web Map Service), предоставляющий доступ к данным архива. Веб-сервер hVault-WMS транслирует WMS-запросы в произвольные SQL-запросы к системе hVault, что позволяет реализовать необходимые вычисления на стороне сервера перед отображением карты, при этом вычисления производятся по запросу клиента. Предложенная система обладает структурной гибкостью и может развиваться вслед за эволюцией предъявляемых к ней требований. Это свойство особенно важно для информационных систем поддержки научных исследований, поскольку такие системы существуют в условиях постоянного потока новых запросов и пожеланий, исходящих от пользователей, применяющих технологию для решения конкретных практических задач. Таким образом, система находится в режиме постоянного совершенствования применяемых технических решений.

1.1. Хранение данных

Как показано на рис. 1, все информационные продукты, производимые РИСМ, аккумулируются в системе хранения, обладающей горизонтальной масштабируемостью и поддерживающей единое пространство имен. Система хранения построена на основе сетевого хранилища Isilon, конфигурация которого оптимизирована для решения задачи долговременного хранения сотен терабайт данных с постоянной доступностью.

Мировой опыт эксплуатации систем хранения данных указывает на то, что построение файловых архивов большого объема за счет простого масштабирования связано с рядом трудностей, для преодоления которых требуются оригинальные решения. Анализ сложностей, возникающих при построении больших архивов, позволяет сформулировать следующие требования:

- обеспечение устойчивости системы хранения при отказе отдельных компонентов, вероятность которого возрастает по мере увеличения их числа;
- сохранение времени восстановления при сбоях при росте числа компонентов и емкости системы в заданных пределах;
- необходимость повышения пропускной способности внешних интерфейсов системы при увеличении ее объема.

Горизонтальное масштабируемое сетевое хранилище Isilon представляет собой кластер — распределенную систему, состоящую из множества равноправных узлов, каждый из которых управляет своим пулом жестких дисков. Узлы объединены коммутационным полем с высокой пропускной способностью на основе технологии InfiniBand. Таким образом, гарантируются целостность данных и сохранение их доступности при выходе из строя отдельных узлов [6].

Программно-аппаратный комплекс на базе технологии OneFS фирмы Isilon в ИВТ СО РАН был приобретен в 2012 г. Впоследствии объем дисковой подсистемы расширен с 324 ТиБ до 0.5 ПБ.

1.2. Поточковая обработка данных

Для выполнения полного цикла производства информационных продуктов на основе данных оперативного спутникового мониторинга, поступающих в режиме онлайн, развернут вычислительный кластер. Технологические цепочки получения стандартизиро-

ванных и собственных тематических продуктов реализуются на кластере оригинальным комплексом потоковой обработки, построенным на основе системы Taverna [7].

Комплекс основан на принципах конвейерной обработки данных, в соответствии с которыми вычислительный процесс представляется в виде графа потоков данных, связывающего отдельные модули, отвечающие за получение каждого информационного продукта (рис. 2). В расчетных модулях сосредоточены экспертные знания о предметной области. В их описаниях указываются требования к входным и выходным данным. Затем модули, разработанные для решения конкретных задач, могут быть организованы в различные цепочки обработки. Контроль за тем, чтобы все необходимые данные были доступны перед запуском каждого модуля, а также определение порядка запуска модулей осуществляются автоматически системой Taverna Workbench.

Схема процесса потоковой обработки представляет собой направленный граф без циклов. Каждая вершина графа отвечает входному порту, вычислительному модулю

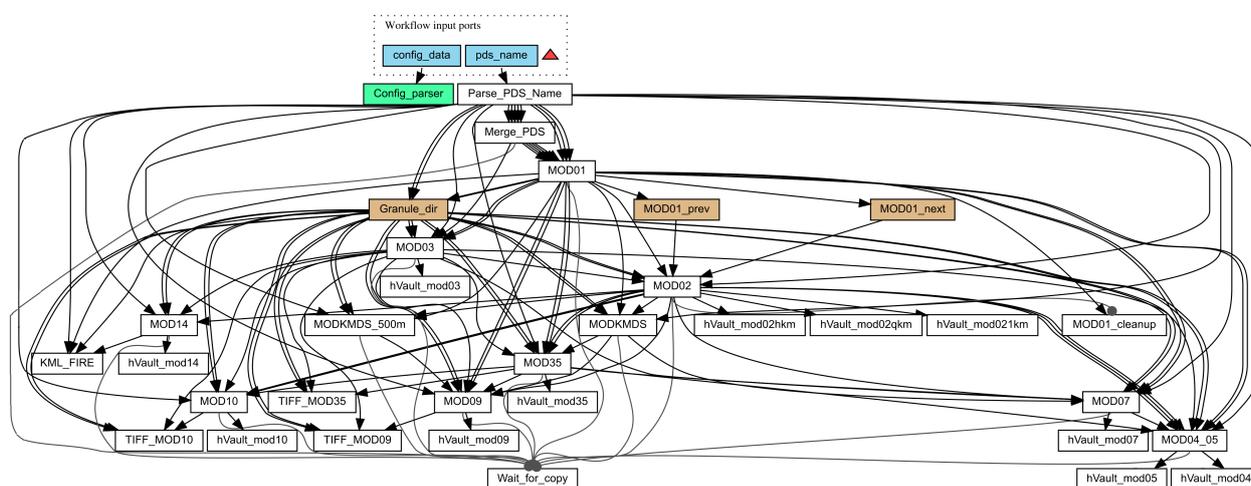


Рис. 2. Схема процесса обработки данных в оперативном режиме

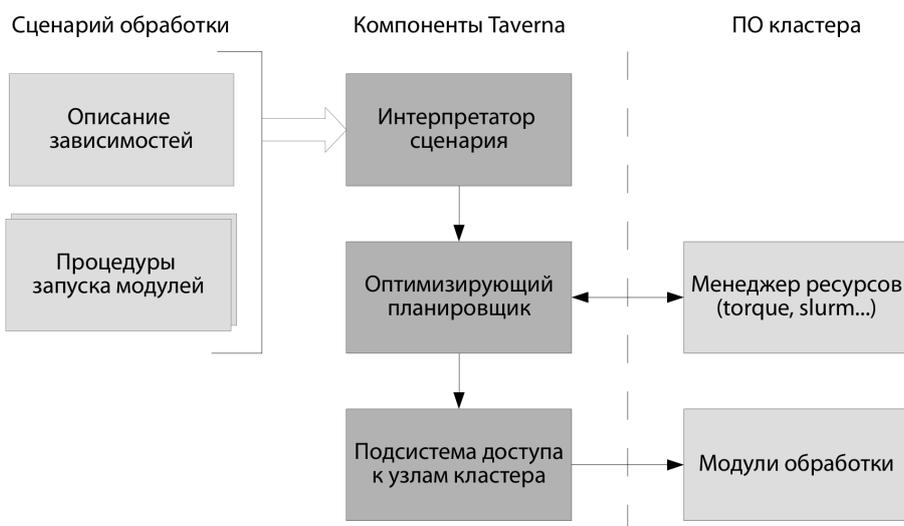


Рис. 3. Направление потоков управления между компонентами, обеспечивающими выполнение процесса потоковой обработки данных на вычислительном кластере

или выходному порту. Входным портам соответствуют вершины, не имеющие входящих дуг, а вершины, не имеющие исходящих дуг, — выходным портам.

Разбиение алгоритма обработки на отдельные элементы обеспечивает его автоматическое распараллеливание и выполнение в распределенном режиме на узлах кластера (рис. 3). При этом осуществляется балансировка нагрузки, учитывающая обмен данными между модулями. Помимо ускорения обработки за счет параллельного выполнения наличие информации о передаче данных позволяет снизить нагрузку на систему хранения благодаря размещению задач, работающих с одним и тем же набором данных, в пределах одного узла, избегая загрузки и выгрузки данных при запуске каждой задачи. Такая оптимизация процесса существенно сказывается на общем времени выполнения задач, так как именно операции ввода-вывода вносят основные задержки при получении и архивации продуктов.

Внедрение комплекса позволило в несколько раз сократить время между поступлением данных в систему и получением продуктов их обработки. Для данных, поступающих с платформ Terra, Aqua, система производит информационные продукты как в оперативном (сразу после поступления информации с приемного комплекса), так и в архивном (после поступления всей дополнительной информации, аккумулируемой в течение суток) режимах [1].

1.3. Технология извлечения данных

Традиционная схема работы с данными дистанционного зондирования Земли подразумевает выгрузку отдельных сцен из архивов и их подключение к геоинформационной системе (ГИС) в виде отдельных слоев. Такая схема обеспечивает возможности фильтрации и последующей визуализации данных в виде наборов тематических слоев, но ограниченные ресурсы рабочих станций затрудняют применение более сложных функций, в особенности при одновременном использовании большого числа спутниковых снимков, их трансформации и агрегации. В результате работа с данными, имеющими не только пространственное, но и временное распределение, затруднена, а к их числу, в частности, относятся все материалы, производимые системами спутникового мониторинга.

Для преодоления этого ограничения предложена технология hVault, обеспечивающая виртуальную интеграцию представленных в архиве данных в реляционной СУБД. В основе технологии лежит принцип представления данных в виде реляционной структуры — набора таблиц, содержащих данные спутниковых изображений или построенных на их основе информационных продуктов. Отображение этих данных в таблицы производится следующим образом: каждому пикселю ставится в соответствие кортеж, содержащий его географические координаты, временной интервал наблюдений и последовательность значений измеренных величин. В таком представлении извлечение и преобразование данных можно реализовать с помощью SQL-запросов, не заботясь об исходном формате хранения данных и их разделении на отдельные файлы. Программный модуль, реализующий этот принцип, разработан как расширение свободно распространяемой СУБД PostgreSQL.

Модуль взаимодействует с PostgreSQL с помощью технологии Foreign Data Wrapper (FDW), которая реализует стандарт SQL/MED для подключения внешних источников данных к СУБД. Преобразование исходных данных в формат строк СУБД производится только в момент выполнения запроса, а сами они по-прежнему хранятся в исходном

формате в файловом архиве. Использование специализированного формата обеспечивает их компактное хранение и уменьшает нагрузку на систему хранения данных, связанную с чтением их с диска. Кроме того, использование исходного формата данных позволяет обойтись без их преобразования, упрощая развертывание технологии для существующих больших архивов.

Обеспечивая для конечного пользователя абстракцию исходных данных в виде единого многомерного информационного пространства, разработанная система дает возможность производить поиск информации, последующее ее извлечение, фильтрацию, агрегацию и сохранение результатов для последующей обработки. Используя представления СУБД, исследователь может многократно создавать динамические наборы данных с различными вариациями без необходимости их сохранения. Такой подход позволяет в рамках одного SQL-запроса к системе интегрировать большие объемы разнородных данных, что в традиционных ГИС является сложной и ресурсоемкой задачей. Технология hVault преодолевает эти ограничения за счет переноса вычислений на сервер и оптимизации ядром СУБД планов выполнения пространственно-временных запросов. В частности, заранее отбрасываются те файлы продуктов, которые не содержат данных, необходимых для конкретного запроса.

Использование технологии hVault возможно как при создании специализированного программного обеспечения для организации доступа к данным по стандартным протоколам, как это реализовано в веб-сервере hVault-WMS, с последующим подключением традиционных ГИС, так и непосредственно, например с помощью специализированной системы для автоматизации научных расчетов IPython. В IPython задания формулируются непосредственно в виде предложений на языке SQL, которые направляются СУБД для исполнения. Результаты запросов преобразуются в числовые массивы, пригодные для дальнейшей обработки и визуализации, а также экспорта в виде растровых изображений уже с помощью средств интерактивной оболочки.

2. Применение технологии в задачах мониторинга окружающей среды

Представленные в предыдущем разделе архитектурные и инфраструктурные решения открывают новые возможности для мониторинга окружающей среды с использованием многолетних спутниковых наблюдений. В качестве примеров применения технологии далее рассматриваются задачи, связанные с анализом температурных режимов земной поверхности и выявлением тепловых аномалий на территории Семипалатинского испытательного полигона и в районе Чуйского землетрясения (Горный Алтай), а также с изучением пожарной обстановки на больших территориях. Преимуществами спутникового мониторинга для решения подобных задач являются регулярность измерений и доступность данных.

2.1. Источники данных

В решении перечисленных выше задач использовались продукты, получаемые на основе данных приборов MODIS, установленных на спутниках Terra и Aqua.

Восстановленные по данным MODIS значения температур содержатся в информационных продуктах MOD11. Эти продукты также включают информацию о точности

восстановления и условиях наблюдения. В архиве эти продукты накапливаются начиная с 2000 г. и покрывают значительную часть территории Северной Евразии. Значения температуры восстанавливались с помощью двухканального алгоритма Вана и Дозьера [8] на основе измерений интенсивности инфракрасного излучения, регистрируемого в каналах 31 (11 мкм) и 32 (12 мкм) сенсора MODIS. Содержащиеся в MOD11 данные представлены на регулярной сетке в синусоидальной проекции с размером ячейки около 1 км. При хороших атмосферных условиях наблюдения алгоритм обеспечивает точность восстановления температуры в пределах 1 К.

Продукты MOD14 группы Тепловые аномалии/Пожары дают информацию об областях на поверхности Земли с аномальной яркостной температурой [9]. Для выделения таких областей используются измерения яркости уходящего излучения земной поверхности в среднем и дальнем инфракрасных диапазонах, выполняемые сенсором MODIS. Локальные тепловые аномалии могут соответствовать как очагам природных пожаров, так и промышленным объектам. Алгоритм их выделения выполняется на втором уровне обработки данных MODIS после каждого сеанса связи с космическим аппаратом. Выделяемые алгоритмом области представляют собой наборы пикселей, размер которых на поверхности Земли определяется разрешением прибора и составляет около 1 км. Центроиды пикселей называются *точками тепловых аномалий*. Их списки обновляются в оперативном режиме и архивируются.

Выявление точек тепловых аномалий по данным MODIS производится на основе данных, получаемых в каналах 21, 22 (4 мкм) и 31 (11 мкм). Для отделения облачности используется вспомогательный диапазон. Стратегия определения пожара основана на абсолютном детектировании пожара (если его сила достаточна для детектирования) и на детектировании относительно фоновых значений (для учета изменчивости температуры поверхности и отражения солнечного света). Для определения ложных срабатываний, таких как солнечные блики или незамаскированные участки береговой линии, а также для маскирования облаков применяются дополнительные каналы MODIS: 1 и 2 (250-метровые, огрубленные до 1 км разрешения, диапазоны 0.65 и 0.68 мкм), также 500-метровый канал 7 (диапазон 2.1 мкм) и канал 32 (разрешение 1 км, диапазон 12 мкм).

Имеющийся на данный момент в информационной системе массив термоточек за 15-летний период содержит свыше 3 млн записей. Каждая запись характеризуется областью, соответствующей одному пикселю спутникового изображения, а также оценкой интенсивности пожара и временем наблюдения.

2.2. Исследование температурного режима Семипалатинского испытательного полигона

Рассмотрим задачу, решаемую с использованием продуктов MOD11 и связанную с изучением температурного режима в районе Семипалатинского испытательного полигона, где весной 1997 г. в процессе мониторинга снежного покрова территории Казахстана были обнаружены температурные аномалии. До сих пор неизвестны точные причины возникновения аномалии. В течение десяти с лишним лет несколькими исследовательскими группами проводились работы и обсуждались две гипотезы, объясняющие происхождение тепловой аномалии — это природные тектонические процессы и последствия ядерных взрывов. Результаты исследований можно свести к следующим выводам:

- Бесснежное пятно существует с середины января до середины марта с вариациями площади и температуры в различные годы. Аномальная зона вытянута с се-

веро-запада на юго-восток, а ее ось совпадает с Главным Чингизским разломом. Выделены два устойчивых очага повышенных температур, не совпадающие с площадками подземных ядерных взрывов. Эти наблюдения позволили сделать вывод в пользу гипотезы о естественной природе этой “глобальной” температурной аномалии [10–12].

- Источником восходящих газообразных тепловых потоков являются продукты радиолиза и “тлеющих” реакций деления радиоактивных частиц в сильно увлажненном почвенном слое [13].
- Выявлены устойчивые “локальные” тепловые аномалии на дневной поверхности в области откольных зон подземных ядерных взрывов, имеющих “возраст” до 20 лет и более [14].

В ходе проведенных исследований остались неподтвержденными предположения о существовании “глобальной” тепловой аномалии, ее связи с “локальными” аномалиями и их природе. Ранее полученные результаты основывались на анализе отдельных сцен спутниковых снимков и не включали исследование динамики температурного режима полигона. Технология hVault позволила в течение нескольких дней разработать программное обеспечение, выполнить обработку многолетней серии сцен и получить результаты в формате GeoTIFF, удобном для дальнейшего использования в геоинформационных системах. Результаты анализа динамики среднемесячных температур за 13 лет в окрестности Семипалатинского испытательного полигона приведены на рис. 4, где показаны средние значения температур земной поверхности в марте. Хорошо видна

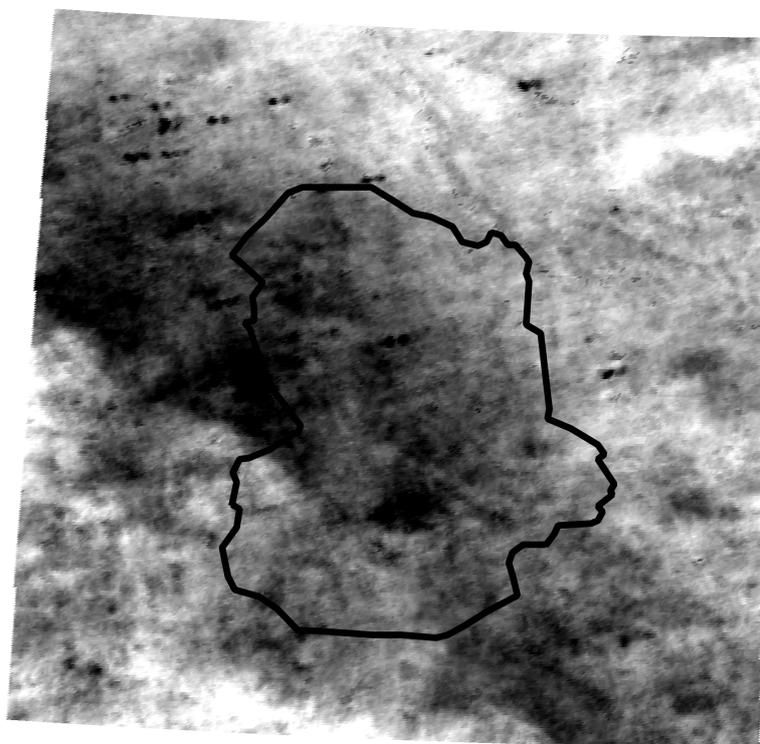


Рис. 4. Средняя ночная температура (MOD11) в районе Семипалатинского полигона за 13 лет (март 2001–2013 гг.). Черный цвет соответствует областям, где средняя температура превышает 264 К (-9°C), белый — 259 К (-14°C)

область, прилегающая к границе Главного Чингизского разлома, где регистрируется статистически значимое превышение температуры над фоном [15].

2.3. Выявление аномалий температуры

Пространственно-временные флуктуации теплового излучения Земли, которые возникают незадолго до сейсмических событий, изучались в работах отечественных и зарубежных исследователей [16–19]. Все они указывали на возможность использования спутниковых данных для выявления аномалий теплового излучения Земли, фиксируемого сенсором в тепловом инфракрасном диапазоне (TIR) за несколько дней и даже недель до сейсмического события.

Коллективом под руководством итальянского геофизика В. Трамутоли выполнены исследования серии сейсмических событий, произошедших в разные годы в Греции, Италии, Турции, Индии [18, 20–26]. Анализ температурных режимов для выявления тепловых аномалий выполнялся с помощью метода RST (Robust Satellite Techniques). В основе метода лежит статистический анализ наборов спутниковых данных для выделенной территории, полученных в условиях, обеспечивающих малость изменений изучаемой величины вследствие суточного и сезонного хода температур. Для этого фиксируется временной интервал продолжительностью несколько десятков дней, содержащий исследуемое событие. Затем извлекаются значения температуры в эти дни года за несколько лет. Для выявления тепловых аномалий, сопряженных с землетрясением, рассматривается область на поверхности Земли, охватывающая эпицентр и основные геологические структуры, вовлеченные в тектонические процессы, связанные с сейсмическим событием. В этой области по данным дистанционного зондирования вычисляется индекс

$$\eta_t(r, t) = \frac{\Delta T(r, t) - \langle \Delta T(r, t) \rangle_t}{\sigma_{\Delta T}(r)},$$

где $\langle T(r, t) \rangle_r$ — среднее по области значение температуры в момент времени t ; $\Delta T(r, t) = T(r, t) - \langle T(r, t) \rangle_r$ — отклонение температуры в пикселе r в момент времени t ; $\langle \Delta T(r, t) \rangle_t$ — усредненное по времени отклонение температуры в пикселе r ; $\sigma_{\Delta T}(r)$ — среднеквадратичное отклонение $T(r, t)$ от $\langle T(r, t) \rangle_t$.

Главное достоинство индекса заключается в том, что при выборе области и временного промежутка подходящего размера он позволяет исключить влияние вариаций температуры, вызванных климатическими процессами, неоднородностью рельефа и погодными условиями. Метод не требует дополнительной информации, например данных наземных наблюдений, поэтому вычисление индекса может быть автоматизировано.

В настоящей работе метод RST реализован для данных спектрорадиометра MODIS, он применялся для поиска аномалий теплового режима, связанных с Чуйским землетрясением ($M = 7.3$), произошедшим на территории Горного Алтая 27.09.2003. Анализируемая область ($48 - 52^\circ$ с.ш. и $86 - 90^\circ$ в.д.) включала в себя зону эпицентра землетрясения, которая располагается в южной части Алтае-Саянской складчатой области в районе Северо-Чуйского хребта и характеризуется повышенной сейсмической активностью. В качестве исходных данных использованы значения ночных температур поверхности за период с 28 августа по 31 октября в течение 2001–2003 гг. Как уже было сказано, при хороших атмосферных условиях наблюдения алгоритм восстановления температуры (MOD11) обеспечивает точность в пределах 1 К. Для выбранной области исследования число ячеек, в которых ошибка оказалась менее 1 К, составило в среднем

около 30 % в день, но число ячеек, где ошибка не превышала 2 К, значительно больше. Поэтому для вычисления индексов использовались только пиксели, в которых ошибка не превышала 2 К (далее мы будем называть их *валидными*).

При вычислении индекса, следуя В. Трамутоли, выбраны пороговые значения, превышение которых рассматривалось как признак наличия аномалии. Управление вычислениями осуществлялось в среде Jupyter с модулем IPython, что позволило построить последовательность изображений для отслеживания динамики локализации аномалий.

На рис. 5 показано изменение числа аномальных пикселей на протяжении исследуемого периода. Оно возрастает в те дни, когда доля валидных данных опускается ниже 40 %, однако анализ зависимости значений индексов от количества валидных пикселей не показал значимой корреляционной связи между этими величинами. На рис. 6 отображено распределение значений индекса в разные дни как до, так и после сейсмического события. Почти все обнаруженные аномалии на следующий день на том же месте уже не наблюдались. Тем не менее некоторые из них все же были локализованы в одном и том же месте на протяжении нескольких дней.

Выявленные аномалии обнаруживались в распределении значений индекса и при изменении размера области исследований. Это подтверждает, что результаты, полученные с использованием метода RST, мало зависят от размера области.

Вполне возможно, что достаточно устойчивые аномалии, появившиеся через месяц после самого сильного толчка, не имеют непосредственно к нему отношения. Ответ на вопрос о связи выявленных тепловых аномалий и сейсмических процессов, вызвавших Чуйское землетрясение, можно дать только с привлечением результатов наземных измерений и геофизических исследований.

Таким образом, показана принципиальная возможность применения метода RST для выявления тепловых аномалий на территориях с выраженными сезонными колебаниями температуры и со сложным рельефом, таких как территория Алтайских гор.

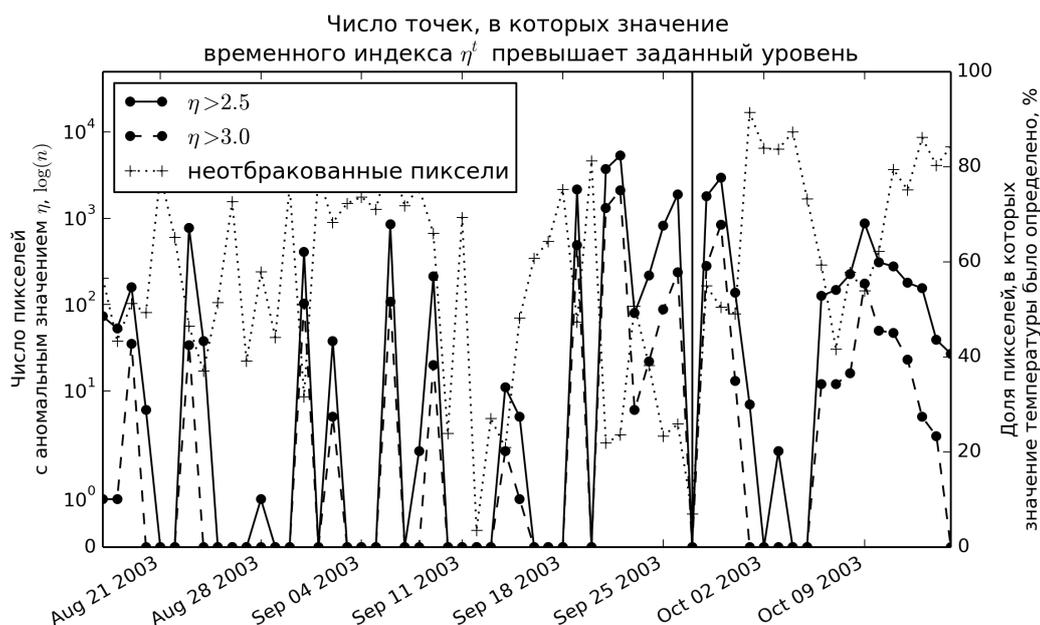


Рис. 5. Количество аномальных и доля валидных пикселей. Вертикальная линия соответствует дню землетрясения — 27 сентября 2003 г.

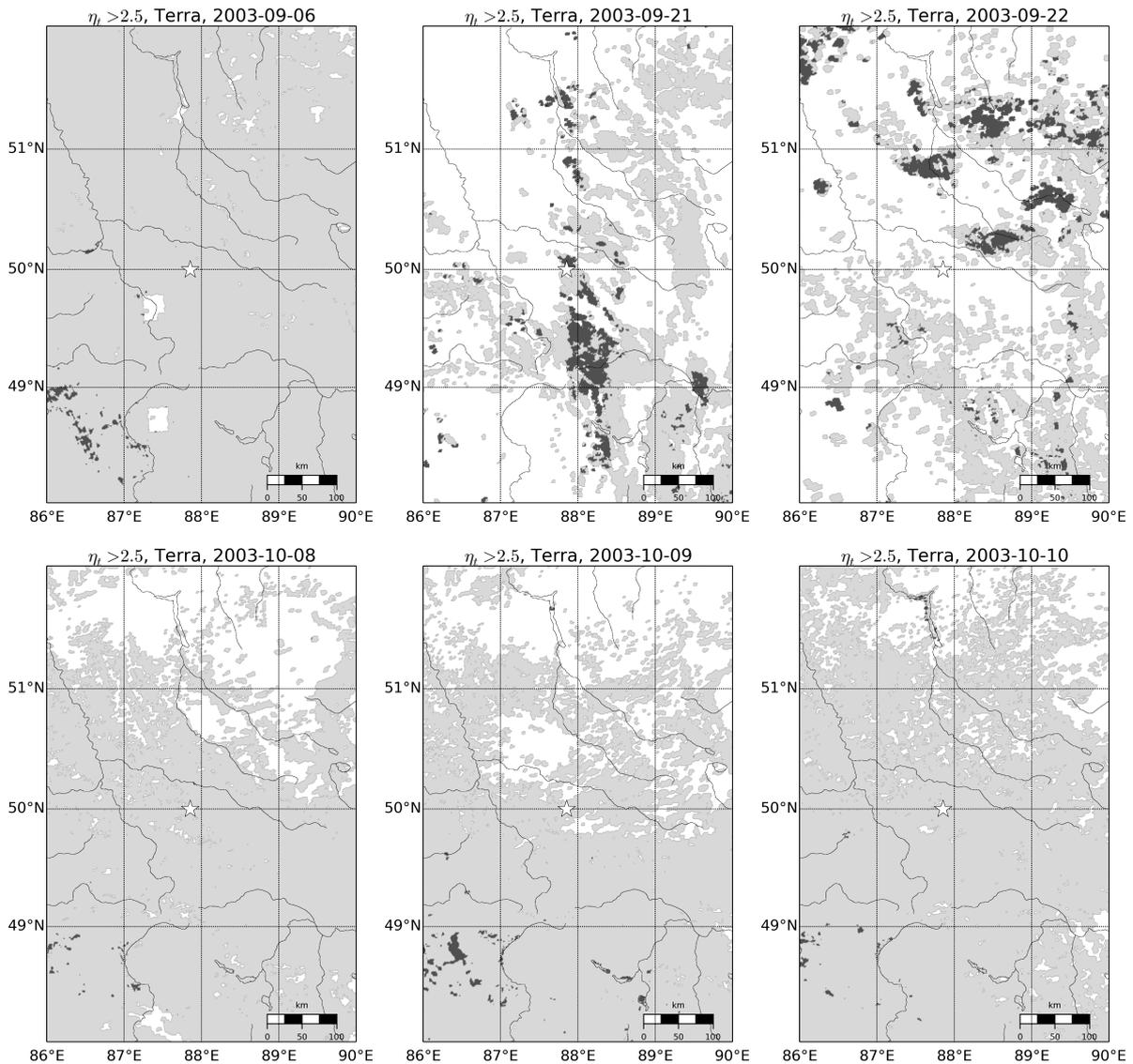


Рис. 6. Пространственное распределение значений индекса η_t . Темным цветом окрашены области $\eta_t > 2.5$, белым — области, где валидные данные отсутствуют

При этом механизм возникновения аномалий требует дальнейшего исследования. Практика опытной эксплуатации разработанной информационной системы показала высокую эффективность используемых технологий извлечения, обработки и анализа информации при решении конкретной тематической задачи.

2.4. Исследование пространственного распределения точек тепловых аномалий

Данные дистанционного зондирования Земли из космоса позволяют оперативно выявлять очаги природных пожаров, а также анализировать статистические характеристики распределения областей с аномальной яркостной температурой на больших территориях. В настоящее время накоплен значительный опыт использования спутниковой информации для прогноза пожарной опасности в региональном и глобальном масшта-

бах [27 – 29]. Однако решение этих задач связано с необходимостью обработки больших объемов данных мониторинга аномалий яркостной температуры.

Разработанная информационная система позволяет эффективно решать задачи выявления пространственно-временных закономерностей распределения детектируемых сенсором MODIS точек тепловых аномалий (продукт MOD14). Эти точки соответствуют как очагам природных пожаров, так и промышленным источникам тепла, а также могут быть вызваны ложными срабатываниями алгоритма. Учет пространственного распределения тепловых аномалий позволяет установить их природу средствами статистического анализа их характеристик. Сгущения точек могут указывать на наличие постоянно действующих источников тепла, к числу которых относятся промышленные объекты. Однако такие сгущения нередко соответствуют районам с повышенной частотой возгораний. Для выделения сгущений в рамках информационной системы реализован алгоритм автоматической кластеризации DBSCAN [30]. Выбор этого метода обосновывается его применимостью в задачах, требующих выделения большого количества классов. В дальнейшем планируются реализация методов автоматической классификации для определения типа сгущения и исследование динамики пространственного распределения точек тепловых аномалий.

Перспективным направлением анализа распределения тепловых аномалий является проведение исследований с привлечением данных из различных областей знаний. Представляет интерес изучение связи медико-экологических показателей и многолетних рядов метеорологических наблюдений с пространственным и временным распределением пожарных ситуаций.

Заключение

Информационная система для работы с данными дистанционного зондирования Земли предоставляет исследователю широкий ассортимент инструментов. Поддержка стандартных протоколов взаимодействия между компонентами позволяет использовать существующий инструментарий для визуализации и обработки данных, в том числе такие свободно распространяемые пакеты программ, как QGIS и GRASS GIS. Этот подход позволил отказаться от разработки полнофункциональных компонентов ГИС и сосредоточиться на расширении ее функциональных возможностей за счет сопряжения с сервисами системы извлечения данных hVault. Также возможно использование такого популярного среди научных работников интерактивного инструмента для совместной работы с данными, как Jupyter, основанного на интерпретаторе IPython.

Опыт применения разработанной информационной системы для решения нескольких модельных задач подтвердил, что принятые технологические и инфраструктурные решения создают удобный механизм доступа к большим объемам разнородной пространственной информации для конечного потребителя. Так, применение системы в задаче изучения температурного режима Семипалатинского полигона показало ее эффективность для агрегирования большого объема данных, а также возможность сопряжения с ГИС. Задача выявления тепловых аномалий продемонстрировала возможность вычисления сложных статистик на большом наборе разновременных спутниковых продуктов, анализа и визуализации природных явлений в динамике. Наконец, исследование распределения точек тепловых аномалий, вызванных природными пожарами, показало возможности применения алгоритмов анализа данных: плотностного анализа и методов автоматической кластеризации.

Таким образом, практика опытной эксплуатации компонентов разработанной информационной системы продемонстрировала высокую эффективность используемых технологий извлечения, обработки и анализа информации при решении конкретных тематических задач.

Благодарности. Работа выполнена при финансовой поддержке РФФИ (гранты № 13-07-12105-офи_м, № 12-07-00545-а), Президиума РАН (программа № 43П), Президентской программы “Ведущие научные школы РФ” (грант № НШ-5006.2014.9).

Авторский коллектив выражает благодарность А.А. Лагутину, В.Н. Антонову, А.В. Калашникову, А.А. Сорокину, С.И. Смагину, Л.Б. Чубарову.

Список литературы / References

- [1] **Шокин Ю.И., Антонов В.Н., Добрецов Н.Н., Кихтенко В.А., Лагутин А.А., Смирнов В.В., Чубаров Д.Л., Чубаров Л.Б.** Распределенная система приема и обработки спутниковых данных Сибири и Дальнего Востока // Современные проблемы дистанционного зондирования Земли из космоса. 2012. Т. 9, № 5. С. 45–54.
Shokin, Yu.I., Antonov, V.N., Dobretsov, N.N., Kikhtenko, V.A., Lagutin, A.A., Smirnov, V.V., Chubarov, D.L., Chubarov, L.B. The development and status of the distributed system for satellite data acquisition and processing for Siberia and the Far East // Current Problems in Remote Sensing of the Earth from Space. 2012. Vol. 9, No. 5. P. 45–54.
- [2] **Шокин Ю.И., Добрецов Н.Н., Кихтенко В.А., Смирнов В.В., Чубаров Д.Л., Чубаров Л.Б.** О распределенной инфраструктуре системы оперативного спутникового мониторинга ЦКП ДДЗ СО РАН // Вычисл. технологии. 2013. Т. 18. Специальный выпуск: Труды Всерос. конф. “Индустриальные информационные системы — 2013”. С. 86–94.
Shokin, Yu.I., Dobretsov, N.N., Kikhtenko, V.A., Smirnov, V.V., Chubarov, D.L., Chubarov, L.B. On a distributed infrastructure for the monitoring of satellite remote sensing data using the center for shared access // Computational Technologies. 2013. Vol. 18. Special issue: Proceedings of Russian Conference “Industrial Information Systems — 2013”. P. 86–94. (in Russ.)
- [3] **Кихтенко В.А., Мамаш Е.А., Смирнов В.В., Чубаров Д.Л.** Инфраструктура для выполнения научно-исследовательских работ с использованием данных дистанционного зондирования в ИВТ СО РАН // Труды Междунар. конф. “Дистанционные методы зондирования Земли и фотограмметрия, мониторинг окружающей среды, геоэкология”, Новосибирск, 20–22 апреля 2015. Т. 2. Новосибирск: СГУГиТ, 2015. С. 3–7.
Kikhtenko, V.A., Mamash, E.A., Smirnov, V.V., Chubarov, D.L. Infrastructure for scientific research involving remote sensing data at ICT SB RAS // Proc. Intern. Conf. “Earth Remote Sensing and Photogrammetry, Environmental Monitoring and Environmental Ecology”, Novosibirsk, April 20–22, 2015. Vol. 2. Novosibirsk: SGUGiT, 2015. P. 3–7. (in Russ.)
- [4] **Кихтенко В.А., Смирнов В.В., Чубаров Д.Л.** Использование технологий реляционных баз данных для ретроспективной обработки больших объемов спутниковых снимков // Тезисы докл. 11-й Всерос. открытой конф. “Современные проблемы дистанционного зондирования Земли из космоса”, Москва, 11–15 ноября 2013. М.: Ин-т космических исследований, 2013. С. 35.
Kikhtenko, V.A., Smirnov, V.V., Chubarov, D.L. Using relational database technology for processing satellite imagery in large amounts // Abstracts of 11th Russian Open Conference “On current problems in remote sensing of the Earth from space”, Moscow, November 11–15, 2013. Moscow: Space Research Institute, 2013. P. 35. (in Russ.)

- [5] **Parkinson, C.L., Ward, A., King, M.D.** Earth science reference handbook: a guide to NASA's earth science program and earth observing satellite missions. Washington D.C.: National Aeronautics and Space Administration, 2006. 277 p.
- [6] **Patel, S.M., Mikesell, P.A.** System and methods for providing a distributed file system utilizing metadata to track information about data stored throughout the system. US Patent 7,685,126. 2010.
- [7] **Wolstencroft, K., Haines, R., Fellows, D., Williams, A., Withers, D., Owen, S., Soiland-Reyes, S., Dunlop, I., Nenadic, A., Fisher, P., Bhagat, J., Belhajjame, K., Bacall, F., Hardisty, A., Nieva de la Hidalga, A., Balcazar Vargas, M.P., Sufi, S., Goble, C.** The Taverna workflow suite: designing and executing workflows of Web Services on the desktop, web or in the cloud // *Nucleic Acids Research*. 2013. Vol. 41(W1). P. W557–W561.
- [8] **Wan, Z., Dozier, J.** A generalized split-window algorithm for retrieving land-surface temperature from space // *IEEE Trans. Geoscience and Remote Sensing*. 1996. Vol. 34, No. 4. P. 892–905.
- [9] **Justice, C.O., Giglio, L., Korontzi, S., Owens, J., Morisette, J.T., Roy, D., Descloitres, J., Alleaume, S., Petitcolin, F., Kaufman, Y.** The MODIS fire products // *Remote Sensing of Environment*. 2002. No. 83(1-2). P. 244–262.
- [10] **Sultangazin, U.M., Zakarin, E.A., Spivak, L.F., Arkhipkin, O.P., Muratova, N.R., Terehov, A.G.** Monitoring of temperature anomalies in the former Semipalatinsk Nuclear Test Site // *Comptes Rendus de l'Académie des Sciences. Series IIB: Mechanics, Physics, Chemistry, Astronomy*. 1998. Vol. 326. P. 135–140.
- [11] **Спивак Л.Ф., Витковская И.С., Батырбаева М.Ж., Сагатдинова Г.Н.** Космический мониторинг температурного режима района Семипалатинского испытательного полигона: 10 лет наблюдений // *Вест. НЯЦ РК*. 2006. No. 2. С. 97–103.
Spivak, L.F., Vitkovskaya, I.S., Baturbaeva, M.Z., Sagatdinova, G.N. Semipalatinsk Test Site temperature monitoring from space: 10 years of observation // *NNC RK Bulletin*. 2006. No. 2. P. 97–103. (in Russ.)
- [12] **Zakarin, E., Balakay, L., Mirkarimova, B., Tuseeva, N., Pak, K., Baklanov, A., Mahura, A., Sorensen, J.H.** Geoinformation modeling of radionuclide transfer from the territory of the Semipalatinsk Test Site. FP6 EC CA Enviro-RISKS; DMI Scientific Report 08-06. Copenhagen: DMI, 2008. 40 p.
- [13] **Мелентьев М.И., Великанов А.Е.** О природе региональной тепловой аномалии в районе Семипалатинского испытательного полигона, выявленной по данным дистанционного космического зондирования // *Вест. НЯЦ РК. Геофизика и проблемы нераспространения*. 2003. № 2. Адрес доступа: http://old.nnc.kz/fileadmin/nnc/downloads/bulletin/2003/NNC_RK_Bulletin_2_14_2003.pdf
Melent'ev, M.I., Velikanov, A.E. On regional thermal anomaly around Semipalatinsk Test Site, discovered via remote sensing data from space // *NNC RK Bulletin. Geophysics and non-proliferation problems*. 2003. No. 2. Available at: http://old.nnc.kz/fileadmin/nnc/downloads/bulletin/2003/NNC_RK_Bulletin_2_14_2003.pdf (in Russ.)
- [14] **Андреев А.И.** Идентификация мест выхода газообразных радиоактивных продуктов из горного массива на дневную поверхность по проявлениям тепловой аномалии в эпицентральной зоне подземного ядерного взрыва // *Вест. Тихоокеан. гос. ун-та*. 2008. № 1(8). С. 37–44.
Andreev, A.I. Identification of vents of gaseous radioactive products from the mountain range to the surface via manifestations of thermal anomaly in the epicentral area of an underground nuclear explosion // *Pacific National University Bulletin*. 2008. No. 1(8). P. 37–44. (in Russ.)

- [15] Мамаш Е.А., Аюнов Д.Е., Кихтенко В.А., Смирнов В.В., Чубаров Д.Л. Исследование температурного режима территории Семипалатинского полигона с использованием пространственно-временной агрегации длинных рядов спутниковых измерений // Тр. Междунар. конф. “Дистанционные методы зондирования Земли и фотограмметрия, мониторинг окружающей среды, геоэкология”, Новосибирск, 20–22 апреля 2015. Т. 2. Новосибирск: СГУГиТ, 2015. С. 39–44.
Mamash, E.A., Ayunov, D.E., Kikhtenko, V.A., Smirnov, V.V., Chubarov, D.L. Infrastructure for scientific research involving remote sensing data at ICT SB RAS // Proc. Intern. Conf. “Earth Remote Sensing and Photogrammetry, Environmental Monitoring and Environmental Ecology”, Novosibirsk, April 20–22, 2015. Vol. 2. Novosibirsk: SGUGiT, 2015. P. 39–44. (in Russ.)
- [16] Горный В.И., Сальман А.Г., Тронин А.А., Шилин Б.В. Уходящее инфракрасное излучение Земли — индикатор сейсмической активности // Докл. АН СССР. 1988. Т. 301, № 1. С. 67–69.
Gornyy, V.I., Salman, A.G., Tronin, A.A., Shilin, B.V. The outgoing infrared radiation of the Earth as an indicator of seismic activity // Doklady Akademiia Nauk SSSR. 1988. Vol. 301. P. 67–69.
- [17] Тронин А.А., Сараф А.К. Космическая тепловая съемка Передового надвига Гималаев // Соврем. пробл. дистанц. зондир. Земли из космоса. 2010. Т. 7, № 2. С. 350–353.
Tronin, A.A., Saraf, A.K. Satellite thermal survey of Himalayan Frontal Thrust // Current Problems in Remote Sensing of the Earth from Space. 2010. 2012. Vol. 7, No. 2. P. 350–353.
- [18] Tramutoli, V., Aliano, C., Corrado, R., Filizzola, C., Genzano, N., Lisi, M., Martinelli, G., Pergola, N. On the possible origin of thermal infrared radiation (TIR) anomalies in Earthquake-prone areas observed using robust satellite techniques (RST) // Chemical Geology. 2013. No. 339. P. 157–168.
- [19] Жуков Б.С., Халле В., Шлотцхауэр Г., Эртель Д. Пространственно-временной анализ тепловых аномалий как предвестников землетрясений // Соврем. пробл. дистанц. зондир. Земли из космоса. 2010. Т. 7, № 2. С. 333–343.
Zhukov, B.S., Khalle W., Shlotzhauer, G., Ertel, D. Spatial and temporal analysis of thermal anomalies as earthquake precursors // Current Problems in Remote Sensing of the Earth from Space. 2010. Vol. 7, No. 2. P. 333–343. (in Russ.)
- [20] Tramutoli, V., Di Bello, G., Pergola, N., Piscitelli, S. Robust satellite techniques for remote sensing of seismically active areas // Annals of Geophysics. 2001. No. 44(2). P. 295–312.
- [21] Filizzola, C., Pergola, N., Pietrapertosa, C., Tramutoli, V. Robust satellite techniques for seismically active areas monitoring: a sensitivity analysis on September 7, 1999 Athens’s Earthquake // Physics and Chemistry of the Earth. 2004. No. 29(4-9). P. 517–527.
- [22] Aliano, C., Corrado, R., Filizzola, C., Genzano, N., Pergola, N., Tramutoli, V. Robust TIR satellite techniques for monitoring Earthquake active regions: limits, main achievements and perspectives // Annals of Geophysics. 2008. No. 51(1). P. 303–318.
- [23] Genzano, N., Aliano, C., Corrado, R., Filizzola, C., Lisi, M., Mazzeo, G., Paciello, R., Pergola, N., Tramutoli, V. RST analysis of MSG-SEVIRI TIR radiances at the time of the Abruzzo 6 April 2009 Earthquake // Nat. Hazards Earth Syst. Sci. 2009. No. 9(6). P. 2073–2084.
- [24] Genzano, N., Aliano, C., Filizzola, C., Pergola, N., Tramutoli, V. Robust satellite technique for monitoring seismically active areas: The case of Bhuj-Gujarat Earthquake // Tectonophysics. 2007. No. 431(1-4). P. 197–210.

- [25] **Corrado, R., Caputo, R., Filizzola, C., Pergola, N., Pietrapertosa, C., Tramutoli, V.** Seismically active area monitoring by robust TIR satellite techniques: a sensitivity analysis on low magnitude Earthquakes in Greece and Turkey // *Nat. Hazards and Earth Syst. Sci.* 2005. No. 5(1). P. 101–108.
- [26] **Tramutoli, V., Cuomo, V., Filizzola, C., Pergola, N., Pietrapertosa, C.** Assessing the potential of thermal infrared satellite surveys for monitoring seismically active areas: The case of Kocaeli (Izmit) Earthquake, August 17, 1999 // *Remote Sensing of Environment.* 2005. No. 96(3-4). P. 409–426.
- [27] **Кузнецов Г.В., Барановский Н.В.** Прогноз возникновения лесных пожаров и их экологических последствий. Новосибирск: Изд-во СО РАН, 2009. 301 с.
Kuznetsov, G.V., Baranovskiy, N.V. Forecasting of wildfires and their environmental impact. Novosibirsk: Publishing House SB RAS, 2009. 301 p. (in Russ.)
- [28] Спутниковый мониторинг лесных пожаров в России. Итоги. Проблемы. Перспективы / Под. ред. В.В. Белова. Новосибирск: СО РАН, ИОА, ГПНТБ, 2003. 135 с.
Satellite Monitoring of Forrest Fires in Russia. Results. Problems. Perspectives / Ed. V.V. Belov. Novosibirsk: SO RAN, IOA, GBNTB, 2003. 135 p. (in Russ.)
- [29] **Le Page, Y., Morton, D., Bond-Lamberty, B., Pereira, J.M.C., Hurtt, G.** HESFIRE: a global fire model to explore the role of anthropogenic and weather drivers // *Biogeosciences.* 2015. Vol. 12, No. 3. P. 887–903.
- [30] **Sander, J., Ester, M., Kriegel, H.-P., Xu, X.** A density-based algorithm for discovering clusters in large spatial databases with noise // *Proc. of the Second Intern. Conf. on Knowledge Discovery and Data Mining (KDD-96)*, Portland, Oregon, USA. Menlo Park, California: AAAI Press, 1996. P. 226–231.

Поступила в редакцию 30 сентября 2015 г.

An information system for acquisition, processing and access to satellite data and its applications in environmental monitoring

SHOKIN, YURIY I.¹, DOBRETSOV, NIKOLAI N.^{1,3}, MAMASH, ELENA A.¹,
KIKHTENKO, VLADIMIR A.^{1,2}, VORONINA, POLINA V.^{1,2}, SMIRNOV, VALENTIN V.¹,
CHUBAROV, DMITRIY L.^{1,*}

¹Institute of Computational Technologies SB RAS, Novosibirsk, 630090, Russia

²Novosibirsk State University, Novosibirsk, 630090, Russia

³V.S. Sobolev Institute of Geology and Mineralogy SB RAS, Novosibirsk, 630090, Russia

*Corresponding author: Chubarov, Dmitriy L., e-mail: dchubarov@ict.nsc.ru

The paper presents an information system aimed to support scientific research projects that use satellite monitoring data. The system is based on an approach designed to represent a large archive of structured files which contain remote sensing data as a unified multidimensional information space, that is called as the hVault technology. The capabilities of the system are shown using several application problems as examples. The efficiency of the system is demonstrated on large volumes of spatial data.

The information system incorporates the infrastructure for data acquisition and storage, a workflow-based system for automated data processing and generation of

multiple information products, and an implementation of the hVault technology based on PostgreSQL that virtually integrates the data stored in the information products in a relational database. Users access the information system via multiple front ends, such as interactive web applications for data analysis or specialized web-protocol adapters that connect the system to the desktop GIS.

The applications presented in this paper show how the system can be used to perform both spatial and temporal analysis tasks on the archive of land surface temperature measurements done by satellite on a daily basis. The application is aimed to study the thermal properties of the Semipalatinsk Nuclear Test Site area and the epicenter zone of the 2003 Altai earthquake. The application is capable to analyze the distribution of thermal “hotspots” related to the wildfire hazards over a large swath of land.

The initial experience for the application of the presented information system shows that it can increase the efficiency of the use of satellite data in research.

Keywords: remote sensing of the Earth, MODIS, thermal anomalies, geospatial data, Robust Satellite Technique (RST).

Acknowledgements. This work was financially supported by the RFBR (grants No. 13-07-12105, 12-07-00545), Program of Presidium of the RAS No. 43P, President Programme of supporting leading schools of Russian Federation (grant NSH-5006.2014.9)

Received 30 September 2015