

***Построение инструментов онлайн-анализа
временных серий данных сверхбольших
распределенных архивов***

***Руткевич Б.П., Балашов И.В., Лупян Е.А.,
Марченков В.В., Прошин А.А., Уваров И.А.***

***Институт космических исследований
Российской академии наук***

**XX Двадцатая международная конференция «СОВРЕМЕННЫЕ ПРОБЛЕМЫ
ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ ИЗ КОСМОСА »
Москва, 2022**

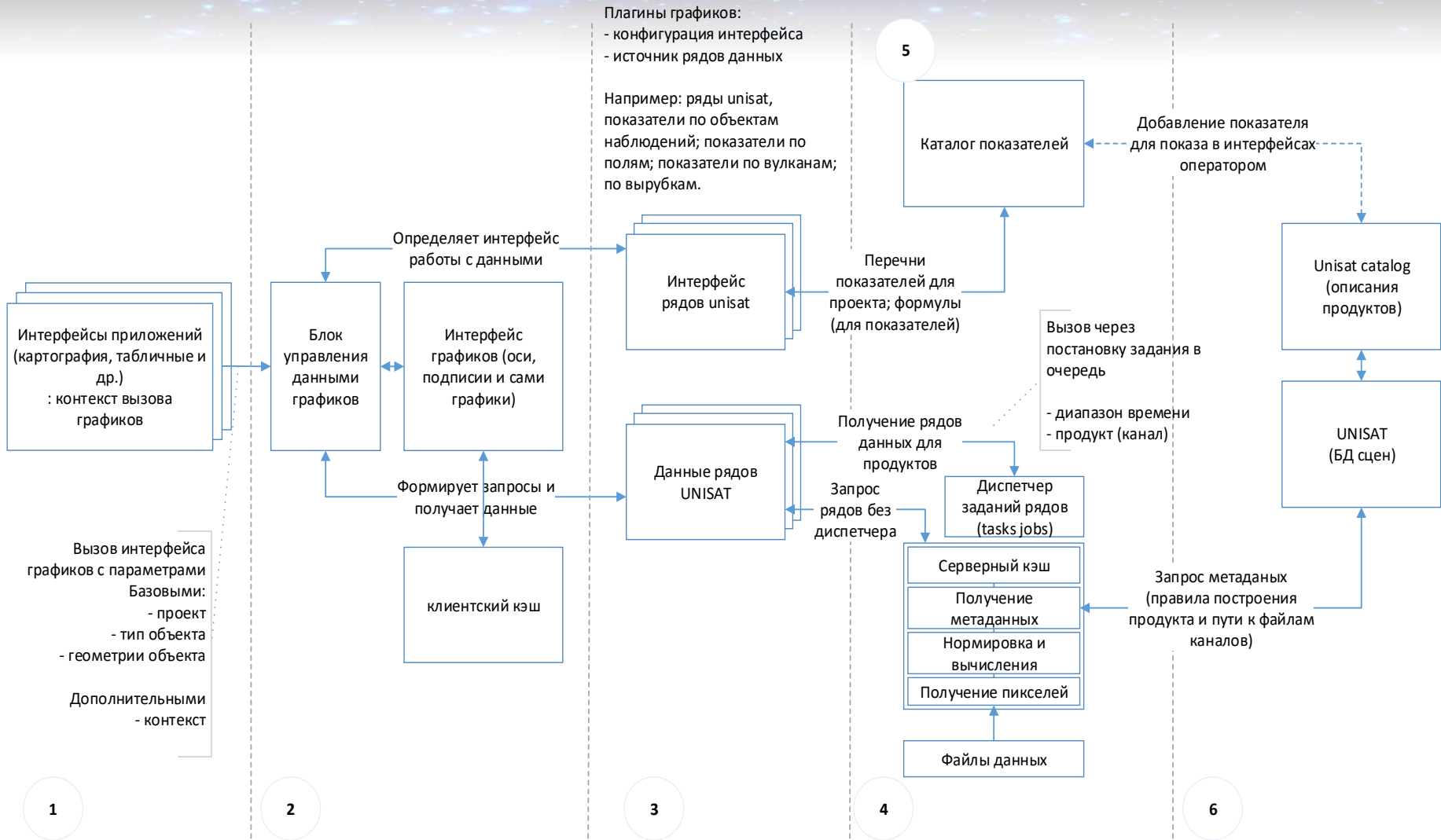
История вопроса

- Развитие космических средств ДЗЗ привело к увеличению объемов данных и наблюдаемых информационных продуктах.
 - Накопленные архивы данных имеют большую территорию покрытия и протяженные временную историю.
 - Эти архивы распределены на различных серверах хранения.
 - Ранее уже был создан сервис для анализа протяженных временных данных в наперед заданных точках, но время работы было значительно больше.
 - Однако формирование протяженных временных серий с использованием существующих подходов занимает время, неприемлемое для онлайн инструментов работы с данными.
- Появилась необходимость «на лету» анализировать протяженные временные данные в произвольной точке.
- Для этого необходимо:
 - оптимизировать получение данных
 - анализировать «на лету»
 - визуализировать или возвращать результат

Задачи

- В числе основных задач, возникающих при построении схемы эффективного доступа к данным временных серий, необходимо решить следующие:
 - Нормализацию, калибровку и вычисления виртуальных продуктов по большим рядам исходных данных временных серий «на лету» как в точке, так и в ее окрестности;
 - Быстрое получение комплекта метаданных для всей протяженной серии данных (включая информацию об исходных файлах, способах построения продуктов и формулах их вычислений);
 - Формирование единого ряда, состоящего из данных, расположенных на разных узлах распределенного архива;
 - Постановка заданий, запуск процедур получения временных серий и получение результатов (диспетчеризация) на узлах системы предоставления данных распределенных архивов;
 - Кэширование результатов чтения временных рядов из исходных файлов для минимизации повторных обращений к данным.

Общая схема эффективного доступа к данным временных серий

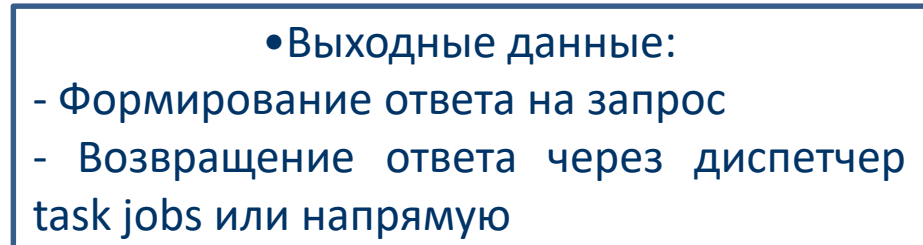
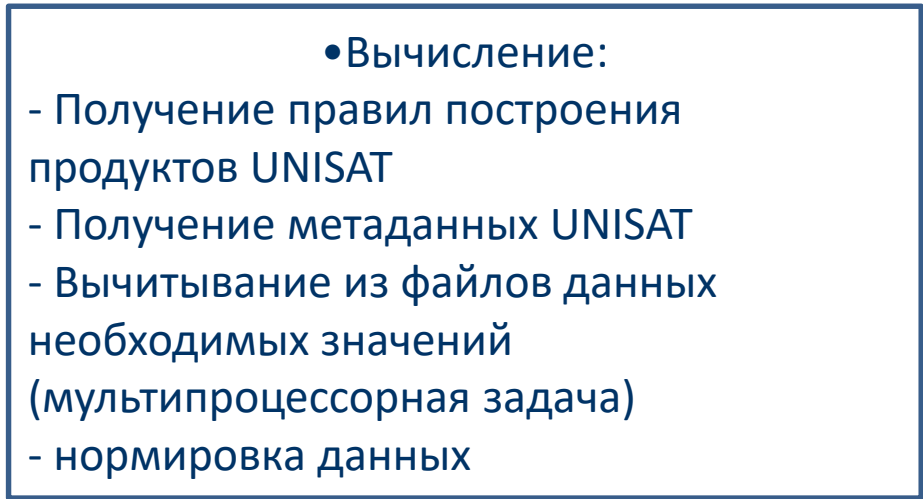
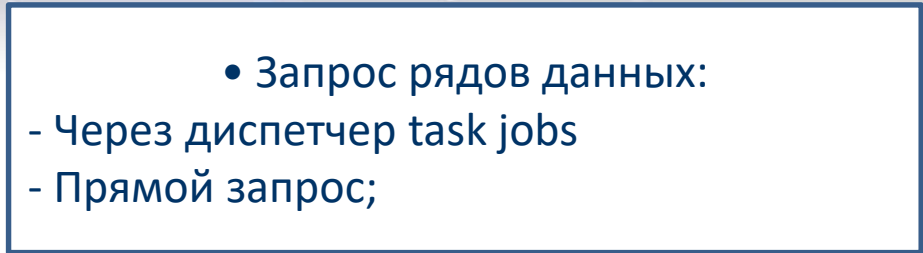
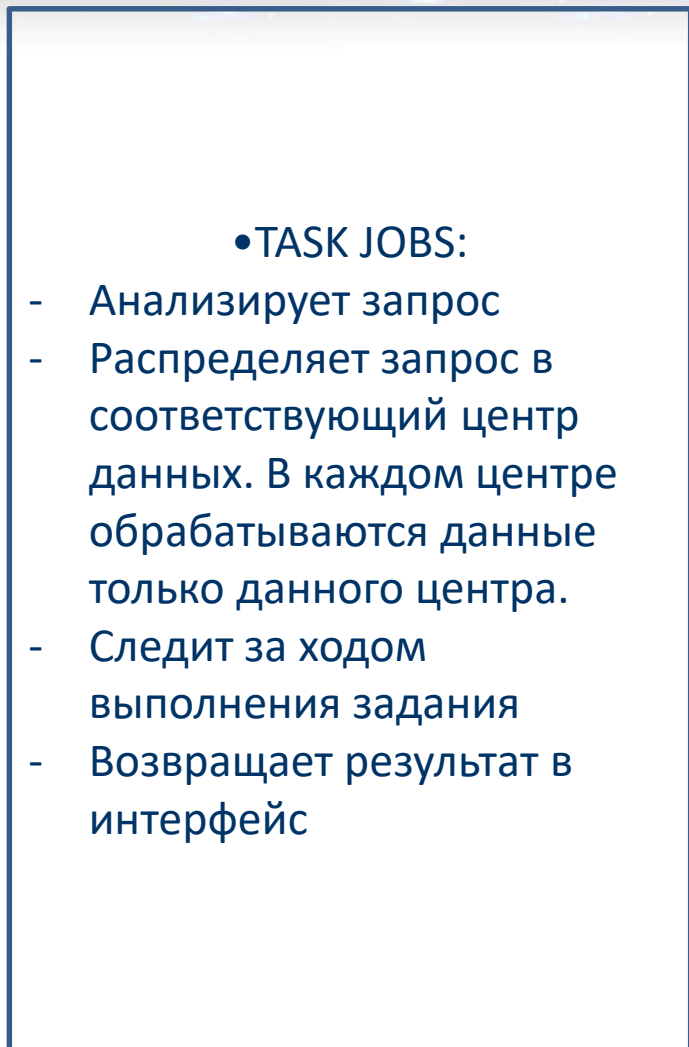


Реализация

Созданы новые блоки и компоненты для эффективного доступа к временным данным:

- Блок диспетчеризации заданий TASK JOBS
- Блок быстрого получения наборов метаданных для временных серий
- Блок быстрого получения значений данных и продуктов
- Блок кеширования данных
- Специализированные интерфейсы и плагины для построения графиков по данным UNISAT

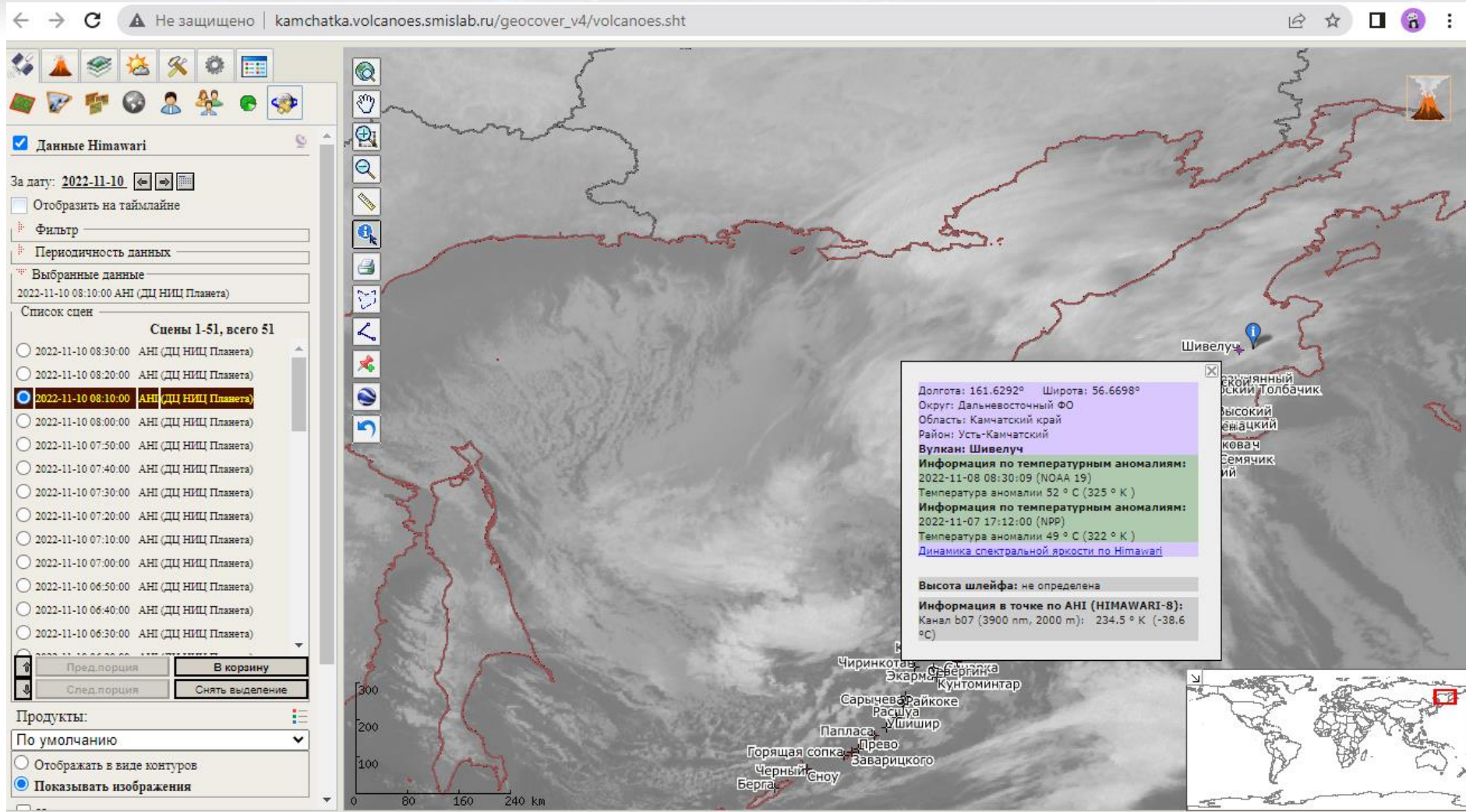
Блок быстрого получения значений данных и продуктов



Особенности блока быстрого получения значений данных и продуктов

- В связи с тем, что объемы данных очень большие с большой временной протяженностью, например только за один день более 100 временных точек и время обращения к данным становится существенным. Для ускорения нами применено параллельное вычисление.
- Реализована возможность получать значение не только в точке, но и анализ (минимальное значение, максимальное значение, среднее значение, среднеквадратичное) по полигону вокруг заданной точки. Это, например, актуально для вулканов, так как ветер может смещать горячие выбросы в сторону от вершины вулкана, а такой анализ показывает наличие активных процессов в окрестности выбранной точки.
- Кеширование позволяет не посылать один и тот же запрос повторно, что может существенно сократить время на получение временного ряда.

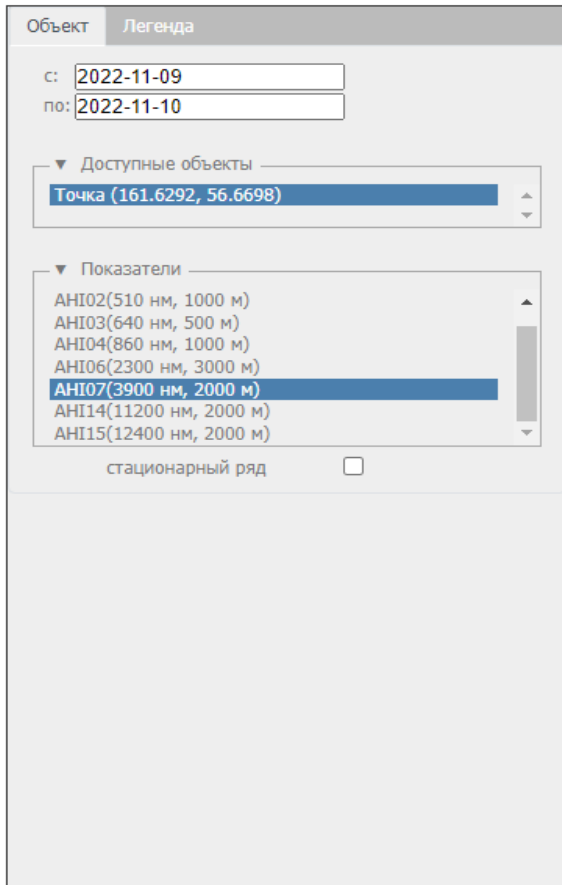
Пример картографического интерфейса «Дистанционный мониторинг активности вулканов Камчатки и Курил»



Яркостная температура

Пример графика яркостной температуры для Himawari в районе вулкана Шивелуч для одного дня (~140 временных точек)

Показано значение в точке и обработка (avr, min, max, std) по полигону вокруг точки



Выводы

- Рост количества спутниковых данных и увеличение временной глубины приводит к повышению требований в архитектуре хранения этих данных. Распределенный архив UNISAT хорошо справляется с этими требованиями.
- В работе проведен обзор требований к инструментам для проведения онлайн-анализа временных серий.
- Описаны вновь созданные сервисы и инструменты для быстрой обработки запросов пользователей.
- Показаны некоторые элементы оптимизации, которые приводят к получению графиков временных рядов «на лету».
- Созданные инструменты применимы как для небольших, так и для протяженных временных серий, таких как, например, геостационаров - более 50 тыс сеансов в год
- На примере одного из центров данных показан пример работы этой системы: на картографическом интерфейсе задается искомая координата, прибор, спутник, затем нужно уточнить интервал времени за который интересуют данные и сделать запрос на построение временного ряда.

Литература

- Лупян Е.А., Прошин А.А., Бурцев М.А., Кашницкий 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.
- Константинова А.М., Балашов И.В., Кашницкий А.В., Лупян Е.А. Унифицированная технология дистанционного мониторинга природных и антропогенных объектов // Современные проблемы дистанционного зондирования Земли из космоса. 2021. Т. 18. № 4. С. 41-52. DOI: 10.21046/2070-7401-2021-18-4-41-52.
- Работы по созданию инструментов выполняются в рамках темы Минобрнауки РФ «Большие данные в космических исследованиях: астрофизика, солнечная система, геосфера» (№122042500019-6) с использованием возможностей ЦКП «ИКИ-Мониторинг» (<http://ckp.geosmis.ru/>).



***Спасибо за
внимание!***