

ТРУДЫ МЕЖДУНАРОДНОЙ КОНФЕРЕНЦИИ  
**Итоги Электронного Геофизического Года**  
3–6 июня 2009 • Переславль-Залесский, Россия

## Проект “Электронная Земля” и академическая инфраструктура пространственных научных данных

А. П. Афанасьев,<sup>1</sup> В. Б. Бритков,<sup>1</sup> И. Э. Дрибинская,<sup>2</sup> и В. Е. Пермитин<sup>2</sup>

Получено 1 июня 2009; принято 1 ноября 2009; опубликовано 22 января 2010.

В работе рассмотрены итоги проектирования информационно-аналитической среды сетевой инфраструктуры распределенной обработки данных в области наук о Земле по проекту “Электронная Земля”, а также значение выполненных междисциплинарных исследований для создания академической инфраструктуры пространственных научных данных и формирования механизмов обмена геоданными результатами фундаментальных и прикладных исследований. **КЛЮЧЕВЫЕ СЛОВА:** проект “Электронная Земля”; информационные ресурсы; информационно-коммуникационные технологии; информационное обслуживание; распределенная информационно-аналитическая среда.

**Ссылка:** Афанасьев, А. П., В. Б. Бритков, И. Э. Дрибинская, и В. Е. Пермитин (2010), Проект “Электронная Земля” и академическая инфраструктура пространственных научных данных, *Росс. ж. наук о Земле*, 11, RE3004, doi:10.2205/2009ES000431.

В современных условиях решение сложного комплекса актуальных проблем экономического и социального развития общества невозможно без научного и информационного обеспечения, без аналитической базы фундаментальных наук о Земле, углубления и расширения их прикладных аспектов, тесно связанных с решением насущных практических задач и, в конечном счете, без комплексной переработки информации, в т.ч. пространственно-временных данных о состоянии окружающей среды и ее возможных неблагоприятных изменениях и воздействии на общество [Бритков и др., 2008].

В рамках одной из программ фундаментальных исследований Президиума РАН в 2004–2008 гг. по направлению “Электронная Земля: научные информационные ресурсы, информационно-коммуникационные технологии, информационное обслуживание, взаимодействие с национальными и международными системами” (далее – проект “Электронная Земля”) выполнен комплекс междисциплинарных исследований, направленных на создание многопользовательского информационно-аналитического пространства по наукам о Земле.

Созданная структурно-содержательная модель пространства потенциального пользователя (Рис. 1) опирается на ядро сетевой инфраструктуры распределен-

ной обработки данных – ИАС GeoSINet (Geographic Science Information Network), включающее ресурсы сети проблемно-тематических научных порталов, академические разработки в области технологий и инструментариев ГИС-анализа и Грид-вычислений, средства навигации, классификации и индексации.

Широкий спектр выполненных междисциплинарных научных исследований в области организации информационной деятельности, информационно-поисковых массивов и баз данных, манипулирования данными и файлами, изучения информационных потребностей и запросов, разработки средств обработки геоданных позволил получить ряд новых фундаментальных и прикладных результатов в области информатизации наук о Земле, в т.ч.:

- новые аналитические методы, принципы и программно-технологические средства комплексного пространственно-временного научного анализа и прогноза (ИППИ, ИСА, ИФЗ, МИТП РАН);
- сценарии решения ряда комплексных научных задач и экспериментальной проверки аналитических возможностей распределенной сети порталов научных учреждений с использованием созданной системы кластеров компьютеров и мощностей Межведомственного суперкомпьютерного центра РАН (ВИНИТИ, ГГМ, ГЦ, ИФЗ и НС РАН в г. Бишкеке).

<sup>1</sup>Институт системного анализа РАН, Москва, Россия

<sup>2</sup>Институт физики Земли им. О. Ю. Шмидта РАН, Москва, Россия



Рис. 1. Структурно-содержательная модель пространства потенциального пользователя.

пользована гибридная топология распределенных систем – в виде децентрализованной архитектуры с элементами централизации. С одной стороны, комбинирование децентрализации с централизацией имеет большие преимущества: децентрализация – это возможность дальнейшей расширяемости, независимость и отказоустойчивость, а централизация – это возможность использования более эффективных методов поиска данных в корпоративной среде. С другой стороны, такая схема архитектуры является закрытой для свободного доступа – требуется централизованная регистрация на главном портале и авторизация, но повышается устойчивость к постороннему вмешательству.

Узлы распределенной системы представлены проблемно-тематическими и технологическими порталами научных учреждений, по сути являющимися Веб-серверами, ресурсы которых представляют в настоящий момент колоссальнейший объем распределенных документографических и фактографических данных, а также приложений (Рис. 2). Они связаны с результатами исследования глубинного строения и геодинамики Земли, взаимодействия внутренних и внешних геосфер и их влиянием на окружающую среду, сейсмической

безопасности, водных ресурсов и водообеспечения, минеральных ресурсов и полезных ископаемых, мониторингом и прогнозом состояния криосферы и многим другим актуальными проблемами.

Активное использование метаданных, извлекаемых через форму импорта файла, позволяет достичь скорости поиска, сравнимой с показателями централизованной архитектуры, хотя нередко возникают проблемы со сбоями загрузки файлов. Совокупность метаданных распределенных ресурсов узлов образует информацию о всей информационно-аналитической среде и представляет собой систему баз метаданных – локальных хранилищ ресурсов.

В целом по проекту “Электронная Земля” решен ряд научных и технологических проблем в отношении используемых типов геопространственных данных, применяемых аналитических методов и алгоритмов на основе конвергенции современных информационных технологий Веб, ГИС и Грид.

Это можно проиллюстрировать следующими примерами:

1. На центральном портале (в ВИНТИ РАН) для расширения функциональности созданного инте-

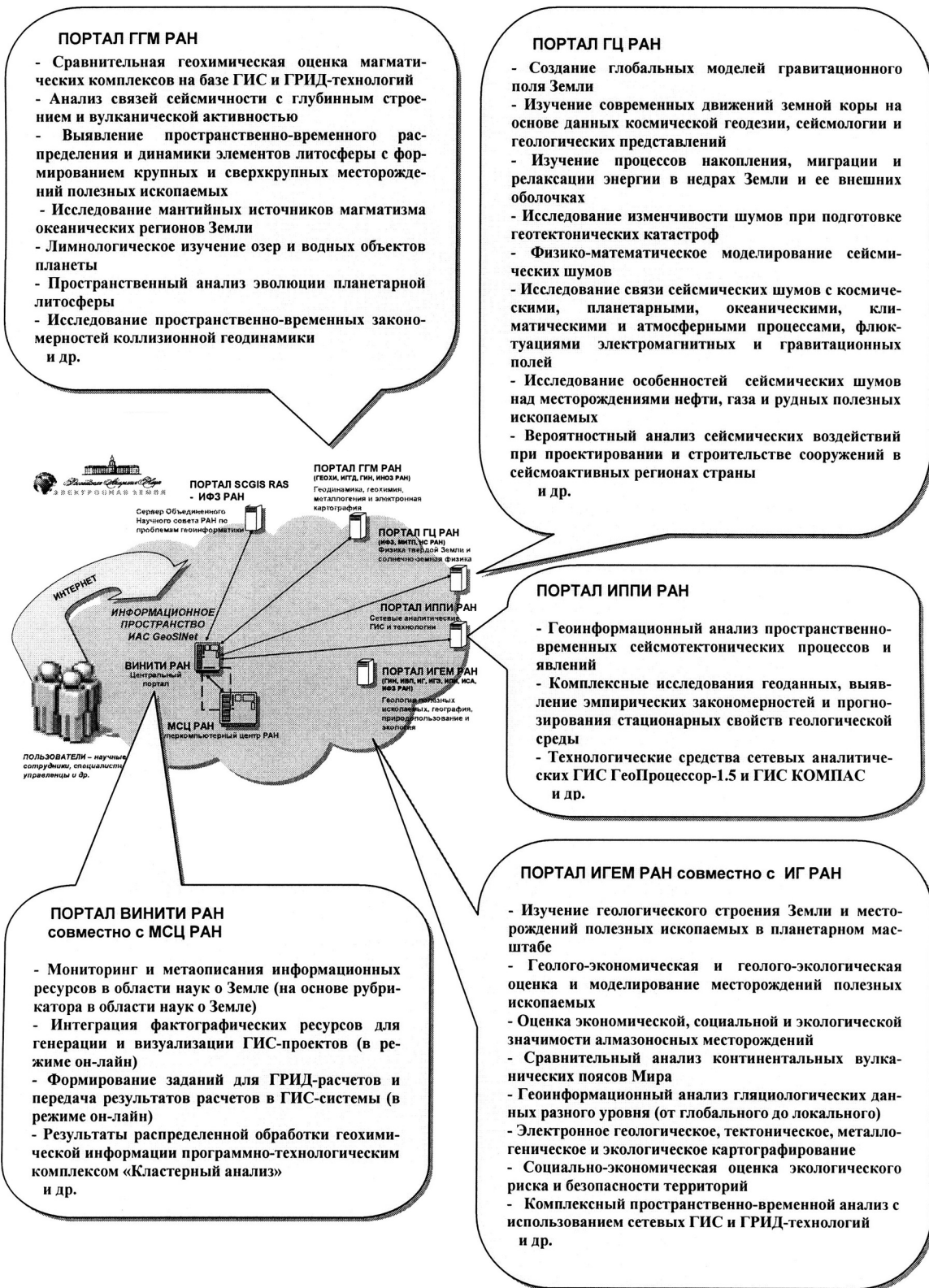


Рис. 2. Корпоративная информационно-аналитическая среда проекта “Электронная Земля”.

грального информационно-аналитического поля потенциального пользователя установлен вычислительный кластер на базе Linux Grid Engine, параллельно с работами по установке аналогичных кластеров в ИФЗ РАН и ГЦ РАН и по объединению их в высокопроизводительный Грид-сегмент Отделения наук о Земле РАН с последующим его включением в российский и европейские сегменты. Разработаны распараллеливаемые методы нейронного анализа массивов метаданных для построения информационной картины результатов поиска. Выполнено подключение к системе онлайн-клиентской ГИС ГеоТайм, базируемой на технологии “Java Web Start” и с возможностями пространственно-временного анализа геоданных. Этим существенно повышены ее возможности по работе с локальным компьютером и к сетевому взаимодействию, использованию многомерных картографических слоев. Наиболее важными изменениями, внесенными в информационно-поисковую систему, установленную на центральном и ряде других порталов, были введение прямых индексов метаданных и организация кэширования результатов поиска. Это позволяет использовать элементы фактографического поиска на базе языка SQL вместе с обычным контекстным поиском и резко сокращает время на переход между блоками результатов поиска и на переход к полному формату выдачи результатов. Решение указанных задач позволило значительно приблизить интегральную среду к пользователям и начать реальную эксплуатацию [Гитис, Шогин, 2009].

2. Разработаны новые методы обработки и анализа пространственных и пространственно-временных геоданных, а также созданы аналитические сетевые ГИС КОМПАС V, ГеоПроцессор 2.0 и ГеоТайм 2.0 (ИППИ РАН), функционирующие в распределенной геоинформационной среде. В ГИС ГеоПроцессор 2.0 (ИППИ РАН совместно с ВИНТИ РАН) существенно развиты аналитическая и сервисная функциональность подсистем ввода/вывода данных, визуального исследования растровых и векторных данных, преобразования данных и пространственного правдоподобного вывода: расширен формат входных данных (SHP, WMS, растровые изображения), изменение проекции карты, обработка векторных объектов, построение буферных зон и оценивание в них произвольных функций, автоматическое построение продукционного правила и др. В ГИС ГеоТайм II реализованы разрабатываемые методы комплексной обработки и анализа пространственно-временных природных процессов: пространственно-временного моделирования поверхностного стока и переноса загрязнений, обнаружения кластеров в потоке сейсмических событий. Выполнен первый этап внедрения разработок в российские и зарубежные научно-исследовательские, производственные и учебные центры. Эти технологии предоставляют авторизо-

ванному пользователю средства создания из распределенных сетевых и локальных информационных и аналитических ресурсов персонализированного интегрального поля решаемой задачи.

3. На основе использования сервис-ориентированного подхода и созданного инструментария IARnet (ИСА РАН) разработаны высокоуровневые средства интеграции ресурсов в распределенную вычислительную среду типа Грид, включающую сервер метаданных, описывающих различные прикладные и фундаментальные вычислительно сложные алгоритмы в области наук о Земле. Созданный сервис GeoDataProcessor (GDPS) производит загрузку данных, находит требуемый алгоритм и осуществляет его запуск или передает на высокопроизводительный узел Грид (Рис. 3). Пользователь может самостоятельно выбрать алгоритм обработки, отслеживать прохождение задания в Грид (его статус) и получать результат в одном из форматов геоданных, которые могут быть использованы при пространственно-временном ГИС-анализе на ГИС КОМПАС V. Реализован вариант распределенной информационно-вычислительной среды для решения интеграционной задачи в области геоэкологии [Афанасьев и др., 2009].

Программно-технологические средства и технологии, разработанные в завершившейся 1-й фазе мегапроекта “Электронная Земля”, позволили организовать совершенно различные по информационному содержанию, по формату хранимых данных в массивах информации, по применяемому составу методологических средств информационной поддержки и сопровождения решения сложных научных задач, интегрирующие решения при использовании персональных рабочих станций и многопроцессорных вычислительных комплексов, в т.ч.:

- комплексного анализа золото-серебрянных месторождений, а также природных опасностей на территории РФ;
- исследования сейсмического потенциала регионов и анализа пространственно-временных предвестников землетрясений;
- оценки сейсмической опасности, сейсмического риска и уязвимости, прогноза месторождений полезных ископаемых;
- построения геодинамических реконструкций основных континентальных блоков;
- оценки временной изменчивости гидрологических, гидрохимических и гидробиологических характеристик;
- расчета обобщенных геохимических и петрохимических характеристик внутриплитного магматизма, отдельного регионального прогноза нефтегазовых провинций.

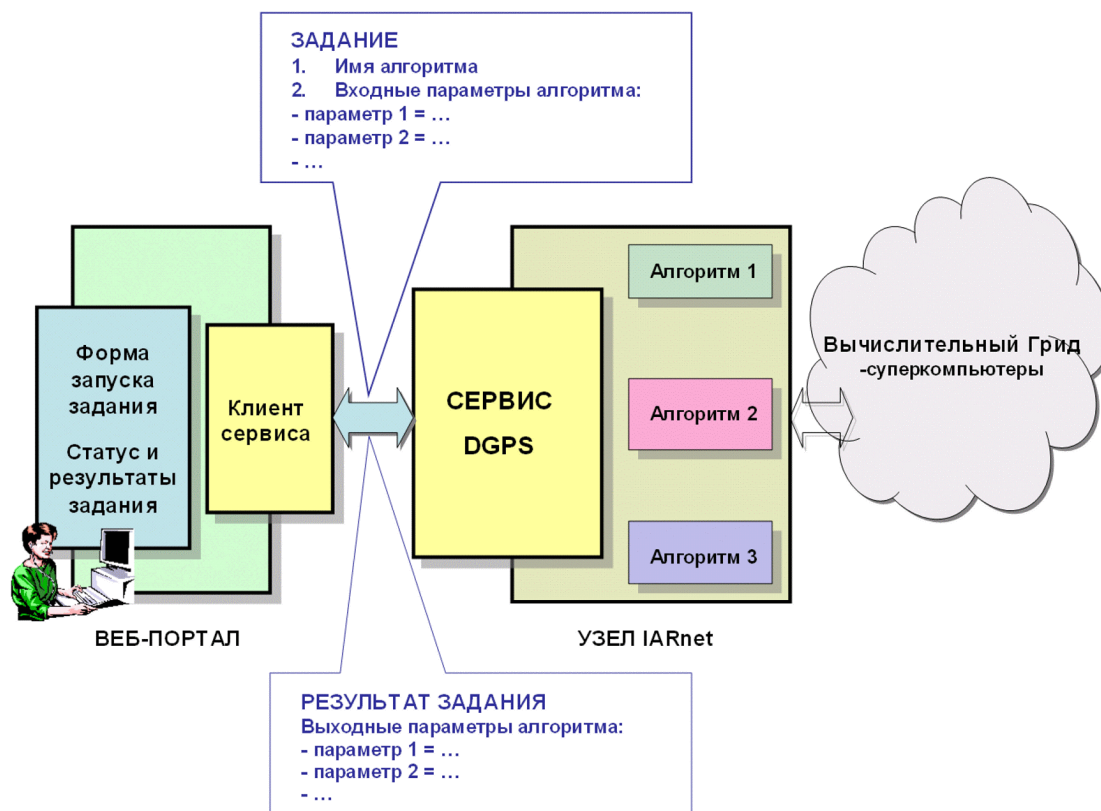


Рис. 3. Пример реализации сервис-ориентированного подхода.

И чем сложнее решаемые задачи, тем больше ощущается интеллектуальная потребность в результатах функционирования динамических информационных систем для научного прогноза на основе определенным образом сформированной и доступной пространственно-временной информации, а также потребность в сложных моделях обработки и анализа. Ведь от степени применения геоданных (в картографической и иной форме пространственной информации) и соответствующих технологий зависит качество, достоверность и востребованность результатов исследований, связанные с изучением развития различных процессов и явлений [Кошкарёв и др., 2008].

Это в свою очередь вызывает необходимость более тщательного подхода к формированию целевых задач следующей фазы проекта “Электронная Земля”, ориентированной на информационно-аналитическое обеспечение утвержденных 27 февраля 2008 г. Правительством РФ направлений научных исследований на перспективу до 2012 г.:

- создания инфраструктуры пространственных научных данных (ИПНД) в виде общедоступного каталога пространственной информации (системы баз метаданных геопорталов), цифровых геобиблиотек

и хранилищ интегрированных геоинформационных ресурсов академических программ и проектов по разным областям наук о Земле и социально-экономическим дисциплинам;

- широкого использования стандартов ISO (Международной организации по стандартизации) и OGC (Консорциума открытых ГИС) для геоинформатики, как целостной системы методов и знаний в рамках информационной технологии;
- предметной и проблемной ориентации создаваемой инфраструктуры пространственных данных, а также поддержки командной работы групп пользователей в реальном времени над приоритетными научными и прикладными задачами;
- обеспечения доступа к данным, информации, услугам и моделям, а также к сценариям и прогнозам – от простых задач к сложным анализам в экологической и социально-экономической областях, для принятия обоснованных управляющих решений и использования результатов разработок в учебно-образовательных целях;
- объединения результатов геоинформационного мо-



делирования с расчетами социально-экономического воздействия и прогноза рисков;

- развития сервис-ориентированной поддержки многопараметрического анализа, моделирования и визуализации больших информационных наборов и временных рядов данных (на примере сценариев решения комплексных прогнозных задач с использованием научных ГИС и Грид-приложений);
- использования оперативной информации от систем дистанционного зондирования Земли.

Следует отметить, что инфраструктура пространственных данных ориентирована на сетевое взаимодействие (главным образом через Интернет) и использование картографических Веб-служб, в частности, WMS, WFS, WMC и др. Инфраструктура предусматривает создание целой сети геопорталов (как специализированных Веб-сайтов), в т.ч. на базе Веб-порталов – некоторых узлов существующей сети ИАС GeoSINet. Для создания академической инфраструктуры пространственных научных данных и формирования механизмов обмена геоданными результатов фундаментальных и прикладных научных исследований необходимо внедрение стандартов (методик), обеспечивающих общий язык и правила согласования тематических геопространственных данных (ГИС-данных), привлечение соответствующей нормативно-методической и правовой базы, развитие сети академических геопорталов и вычислительных ресурсов. Кроме того, проблема ИПНД должна решаться в общем комплексе работ создания и развития инфраструктуры пространственных данных на территории Российской Федерации, концепция которой была одобрена Распоряжением Правительства РФ в 2006 г. Априори ИПНД неразрывно связана с информационным обеспечением исследований в науках о Земле и в смежных областях знания, а также с научно-организационной и управленческой деятельностью. В этом отношении наиболее перспективным является использование новых инструментов и технологий интеграции информационных ресурсов, в том числе программно-технических средств поддержки аудио- и видеоконференций для повышения эффективности совместной работы и обсуждения результатов в реальном времени большими аудиториями ученых научных учреждений ОНЗ РАН.

Иными словами, вторая фаза проекта “Электронная Земля” видится в создании информационно-аналитической среды, объединяющей пространственные, временные и предметные данные и ресурсы в области наук о Земле, разрабатываемых по проектам программ

фундаментальных исследований центральных и региональных отделений Российской академии наук, для решения крупных проблем национальной и мировой значимости в области информатики и информатизации общества.

В заключение следует отметить, что задачи интеграции ГИС-данных и приложений, архитектура геопорталов и другие актуальные вопросы организации пространственных данных были освещены в сообщениях представителей ИГ РАН, ВЦ РАН, ООО “Дата+”, а также в сообщении “Электронная Земля” на пути к инфраструктуре пространственных научных данных” (одного из авторов этой работы) на семинаре Московского совета Русского географического общества “Пространственные метаданные и геопорталы как средства интеграции геоинформационных ресурсов и сервисов” в 2008 г.

## Литература

- Афанасьев, А. П., В. Б. Бритков, В. В. Волошинов, О. В. Сухорослов, В. Е. Пермитин (2009), Инструментарий IARnet для создания GRID-ресурсов, в кн. *Электронная Земля: использование информационных ресурсов и современных технологий для повышения достоверности научного прогноза на основе моделирования решений в интегральных информационных полях*, Отв. редакторы: акад. Ю. М. Арский, акад. Е. П. Велихов, чл.-корр. А. Б. Жижченко, акад. Н. П. Лаверов, акад. Г. И. Савин, 48–53, ВИНТИ РАН, Москва.
- Бритков, В. Б., В. А. Геловани, И. Э. Дрибинская, В. Е. Пермитин (2008), Информационно-аналитическая и программно-технологическая поддержки принятия решений, связанных с безопасностью, *Труды XVI международной конференции “Проблемы управления безопасностью сложных систем”*, 392–396, РГГУ, Москва.
- Гитис, В. Г., А. Н. Шогин (2009), Анализ компонентов современных систем и сетевых средств распределенной обработки данных, в кн. *Электронная Земля: использование информационных ресурсов и современных технологий для повышения достоверности научного прогноза на основе моделирования решений в интегральных информационных полях*, Отв. редакторы: акад. Ю. М. Арский, акад. Е. П. Велихов, чл.-корр. А. Б. Жижченко, акад. Н. П. Лаверов, акад. Г. И. Савин, 1–11, ВИНТИ РАН, Москва.
- Кошкарев, А. В., А. Н. Антипов, А. Р. Батуев, В. В. Ермошин, В. П. Каракин (2008), Геопорталы в составе инфраструктур пространственных данных. Российские академические ресурсы и геосервисы, *География и природные ресурсы*, 21–32.

А. П. Афанасьев, В. Б. Бритков, Институт системного анализа РАН, Москва, Россия

И. Э. Дрибинская, В. Е. Пермитин, Институт физики Земли РАН, Москва, Россия (scgis@uipe-ras.scgis.ru, v.permitin@mail.ru)