Ю. И. Молодродов<sup>1</sup>, А. М. Федотов<sup>2</sup>

Институт вычислительных технологий СО РАН пр. Акад. Лаврентьева, 6, Новосибирск, 630090, Россия E-mail: 1yumo@ict.nsc.ru, 2fedotov@ict.nsc.ru

# МЕТОДОЛОГИЯ ИНТЕГРАЦИИ ДАННЫХ ДЛЯ АТЛАСА «АТМОСФЕРНЫЕ АЭРОЗОЛИ СИБИРИ» <sup>1</sup>

#### Введение

В последние три десятилетия в связи с интенсивным изучением проблемы глобальных изменений окружающей среды выполнялись крупные региональные проекты [Гордов и др., 2002; 2003] по комплексному сбору необходимой информации о характеристиках приземного слоя земной атмосферы и конкретно по атмосферным аэрозолям (АА). Они – неотъемлемая компонента атмосферы, и объяснение всех сложных процессов, происходящих в ней, невозможно без подробной информации о химическом составе АА.

Начиная с 1991 г. специалисты ИХКиГ СО РАН, ИВЭП СО РАН, ИНХ и ЛИН СО РАН проводили регулярные измерения характеристик приземного слоя атмосферы. Мониторинг проводился на территории Западной и Восточной Сибири, Алтайского и Красноярского краев. Важной компонентой исследований являлось получение сведений об ионном и элементном составе, спектре размеров, массовой и счетной концентрации, химическом и фазовом составе, структуре аэрозольных частиц, содержании органического и неорганического углерода в АА в различных регионах Сибири, а также пространственно-временная изменчивость этих характеристик. Огромный объем экспериментальных данных неизмеримо расширил и во многом изменил существующие представления о техногенном влиянии индустриальных центров на окружающую среду региона.

Создание региональной базы данных по AA является актуальной задачей, поскольку это позволит повысить эффективность комплексного мониторинга окружающей среды Сибирского региона. Анализ эмпирического материала даст возможность понять механизмы образования AA естественного и антропогенного происхождения, что позволит разработать математические модели для описания различных процессов в атмосфере и оценки влияния AA на естественные атмосферные процессы, изменчивость погоды и эволюцию климатических изменений качества окружающей среды. Появляется возможность оценить влияние AA на растительные и животные биоценозы, биохимические циклы в различных почвенно-климатических зонах Сибири и биосферы в целом, состояние здоровья населения.

Важным преимуществом предлагаемого решения является значительное расширение научной общественности, которая может ознакомиться с результатами уникальных исследований и огромным эмпирическим материалом, виртуально присутствующим в сети Интернет. Хранение этого наследия, всех описаний и взаимосвязей экспонатов в единой базе данных, позволит сохранить огромный объем научных результатов для следующего поколения исследователей.

### Постановка задачи

В настоящее время существуют порталы и информационные системы, решающие вопросы интеграции, обработки и публикации информации о состоянии окружающей среды, методов

 $<sup>^1</sup>$  Работа выполнена при частичной финансовой поддержке Российского фонда фундаментальных исследований (гранты № 06-07-89060, 06-07-89038, 06-04-48971, 07-07-00271), при поддержке президентской программы «Ведущие научные школы РФ» (грант № НШ 9886.2006.9) и интеграционных проектов СО РАН.

ее анализа. Продуманная многоуровневая система сбора, обработки и хранения атмосферных данных построена в США. Основной объем данных сосредоточен в центрах активных распределенных архивов (Distributed Active Archive Center – DAAC) [Гордов, Фазлиев, 2004], среди них:

- 1. Alaska Synthetic Aperture Radar (SAR) Facility (ASF) содержит данные о полярных процессах и радарные данные (http://www.asf.alaska.edu/).
- 2. EROS Data Center (EDC) Land Processes описывают поверхностные процессы (http://landcover.usgs.gov/).
- 3. Goddard Space Flight Center (GSFC) сведения о верхней атмосфере, глобальной биосфере, атмосферной динамике и геофизики (http://www.gsfc.nasa.gov/) и др.
- 4. Распределенная информационная система «Наблюдение за Землей» (Earth Observation System/Data Information System (EOS/DIS)) (National Center for Atmospheric Research http://www.ncar.ucar.edu/) строится для сбора и обработки данных и их связи с данными центров активных распределенных архивов.

Среди множества разномасштабных проектов, выполняемых в настоящее время и направленных на решение упомянутых проблем, стоит выделить четыре: Program for Integrated Earth System Modelling (PRISM) (http://www.enes.org), Earth Frontier Project (http://www.es.jamstec.go.jp/), Earth System Modelling Framework (ESMF) (http://www.esmf.ucar.edu/) и Earth System Grid (ESG) (http://www.earthsystemgrid.org/).

В Сибирском отделении РАН, также созданы распределенные базы данных по разным разделам наук об окружающей среде.

База данных для ГИС «Аэрозоли Алтая» http://iwep.asu.ru/altai/index.php [Дмитриев, Суторихин, 2002] сформирована на основе многолетних наблюдений за характеристиками атмосферных аэрозолей в приземном слое воздуха, проведенных на территории Алтайского края и Республике Алтай в период 1991–2002 гг. Ее основой служит материал, полученный в ходе разовых и многосуточных наблюдений в экспедициях с помощью передвижной лаборатории (ПЛ). База данных содержит информацию о микрофизических параметрах аэрозолей: спектрах размеров частиц, элементном и ионном составах, массовой концентрации, а также данные о метеорологических параметрах в пунктах отбора проб. В базу данных вошли, также, результаты исследований качества атмосферного воздуха и заболеваемости населения г. Барнаула Алтайского края.

«Электронный атлас биоразнообразия животного и растительного мира Сибири» является информационной системой (http://www.sbras.ru/win/elbib/atlas/) представляющей набор связанных между собой электронных коллекций. Каждая коллекция описывается набором элементарных информационных объектов, из значений которых формируются электронные документы. К настоящему моменту в системе существует более двадцати актуальных информационных коллекций [Федотов и др., 2006].

Двуязычный научный веб-портал ATMOS (http://atmos.iao.ru/) был разработан в Институте оптики атмосферы СО РАН как интегрированное множество распределенных, но согласованных тематических веб-сайтов, комбинирующих стандартную мультимедийную информацию с исследовательскими базами данных, моделями и аналитическими инструментами для использования и визуализации online [Гордов и др., 2004]. Основной темой является физика и химия атмосферы, включая спектроскопию атмосферы, аэрозоли, радиацию, а также прикладные области оценки и управления качеством воздуха, моделирования и оценки воздействия на окружающую среду. Портал представляет собой интегрированный набор множества распределенных, но координируемых предметных сайтов, содержащих типовую мультимедийную информацию с исследовательскими базами данных, моделями и аналитическим инструментарием для прямого использования и визуализации.

Информационно-вычислительный атлас (рис. 1), представленный на сайте Института вычислительных технологий СО РАН (http://web.ict.nsc.ru/aerosol/), посвящен атмосферным аэрозолям Сибири [Молородов, Куценогий, 2005].

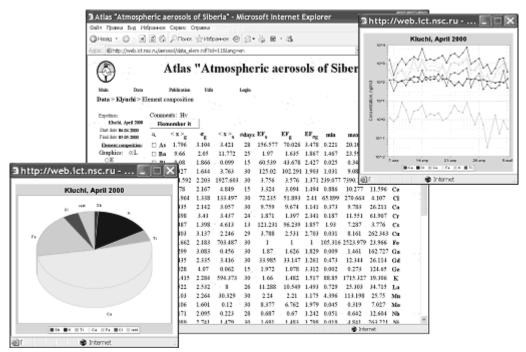


Рис. 1. Общая схема работы Атласа

В него вошли результаты измерений временных характеристик атмосферных аэрозолей, полученные специалистами Новосибирского, Красноярского и Томского научных центров в течение 1995—2006 гг. Мониторинг экологического состояния атмосферы проводился на территории Западной и Восточной Сибири в различных почвенно-климатических зонах (лесостепная, степная, лесотундровая, тундровая и зона Арктического бассейна России).

Данные о спектре размеров АА, концентрации, химическому составу и морфологии частиц, которые необходимы для описания физико-химических и радиационных процессов, определяющих состояние и эволюцию Земной атмосферы, были получены по различным методикам измерений и обработки.

Массовую концентрацию для последующего анализа элементного и химического состава определяли с помощью фильтров и импакторов. Аэрозольное светорассеяние устанавливали с помощью нефелометрических измерений. Концентрацию особо малых частиц измеряли счетчиками ядер конденсации. Разделение на размерные фракции особо малых частиц определялось диффузионными батареями и электростатическими сепараторами. Концентрацию субмикронных и около микронных оптических частиц измеряли оптическими счетчиками аэрозоля.

Наполнение информацией Атласа происходило с учетом изначальных данных измерений: концентрации аэрозолей, время, место и длительность отбора конкретного элемента, его массовой доли, привеса фильтра (масса осадка на фильтре). Учитывались информация о молярной массе каждого элемента, коэффициент Рана (environment factor) — отношение. На основании этих данных определялись среднее геометрическое значение и отклонение концентрации элемента, среднее арифметическое значения массы элемента, минимальное и максимальное значения. При необходимости строились таблицы корреляции между выбранными элементами и проводился факторный анализ.

На текущий момент в рамках Атласа реализованы следующие сервисы:

- визуализация исходных данных: таблицы, графики, диаграммы;
- математическая обработка данных о химическом составе АА, в том числе:
  - корреляционный анализ;
  - факторный анализ;
  - кластерный анализ;
  - дискриминантный анализ;

- вейвлет-анализ одномерных массивов, представляющих собой оцифровку показаний прибора нефелометра ФАН-А;
  - ГИС-приложение, отображающее станции отбора проб;
  - нахождение коэффициента Альбедо по материалам космической съемки;
  - определение направления потока воздушных масс по данным NOAA.

## Общая схема работы Атласа

Основным инструментом системы является программный модуль SMART [Guskov et al., 2004]. Схема его работы представлена на рис. 2. Каждый документ в системе изначально представлен в виде RDF-описания и, возможно, прикрепленных стилевых шаблонов. Генерация запрашиваемой страницы осуществляется в два этапа. На первом этапе по запросу пользователя SMART определяет: требуется ли, например, обработка математическими алгоритмами, в случае необходимости вызывает методы определенных библиотек. После обработки данных или их выборки из базы, модуль SMART генерирует RDF-описание содержания конкретного документа. Второй этап генерации состоит в получении конечного представления страницы посредством применения к этому RDF-описанию стилевых шаблонов. Стилевые шаблоны в системе представлены XSL и CSS файлами.

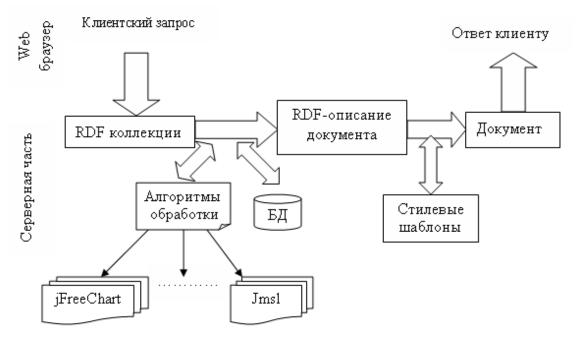


Рис. 2. Схема информационной системы

Программный модуль SMART является «гибким» Java-приложением. Благодаря этому имеется возможность расширения системы за счет добавления Java библиотек, таких как, например: jFreeChart² и Jmsl³, отвечающих за построение графиков и диаграмм и обработку математическими алгоритмами статистического анализа соответственно. Это дает быстрый и удобный способ увеличения возможностей системы по работе с данными, не затрагивая другие рабочие модули.

Технология SMART, в отличие от большинства других серверных языков программирования, представляет собой средство описания содержательной части документа, а его конечное представление пользователю остается за стилевыми шаблонами.

Атлас построен на основе классической трехкомпонентной архитектуры веб-систем. Ее первым компонентом выступает тонкий клиент, которым является программа-навигатор – браузер. Вторым, промежуточным компонентом, является язык программирования PHP (PHP Hypertext Processor), используемый на стороне веб-сервера для динамической генерации

<sup>&</sup>lt;sup>2</sup> Java chart library: http://www.jfree.org/jfreechart/.

<sup>&</sup>lt;sup>3</sup> Visual Numerics' library: http://www.visual-numerics.de/.

HTML-страниц. При этом в роли сервера приложений выступают PHP и Web-сервер. Третий компонент — это сервер базы данных MySQL, обеспечивающий хранение и извлечение данных. Обработка данных производится программами написанными на PHP. Они включаются в состав HTML-страниц и разбираются на стороне сервера, в результате чего программенавигатору (браузеру) передается сформированная HTML-страница. Таким образом, вся обработка будет происходить на стороне сервера (и базы данных), а клиенту предоставляется только результат обработки.

#### Вейвлет-анализ

В пригородной зоне Новосибирска (п. Ключи) проводится непрерывный мониторинг пространственно-временной изменчивости массовой концентрации субмикронной фракции атмосферного аэрозоля с использованием компьютеризированного варианта нефелометра ФАН-А. Это позволяет оценивать уровень загрязнения пограничного слоя атмосферы в пригородной зоне промышленного центра юга Западной Сибири.

Измерения проводятся в течение суток с частотой одно измерение в минуту, всего 1440 замеров в сутки. Они вносятся в массив размерности  $2 \times N$ , где N – количество произведенных измерений. Первый столбец – время произведения замера, второй – значения массовой концентрации. Данные записываются в файл и хранятся на сервере Института химической кинетики и горения (ИХКиГ) СО РАН.

Для выявления регулярных сезонных и суточных колебаний необходимо проводить анализ полученных числовых рядов, в том числе и частотно-временной анализ на основе вейвлет-функций [Малла, 2005]. Целью вейвлет-анализа является выявление циклических закономерностей сигнала, таких как, например, суточные, месячные, сезонные локальные максимумы.

Сервис, реализующий вейвлет-анализ, состоит из пяти функциональных модулей. Структура сервиса и связь модулей представлена на рис. 3.

По запросу клиента *Модуль I* выполняет подключение к удаленному серверу по протоколу FTP, просматривает категории, содержимое категорий, выбирает интересующие клиента файлы и копирует их на сервер-обработчик (сервер Атласа). *Файл-менеджер* просматривает имена содержащихся на сервере-обработчике файлов, удаляет ненужные файлы, а нужные отправляет модулю подготовки данных к анализу. Модуль подготовки данных к анализу последовательно извлекает информацию из файлов, имена которых переданы *модулем 2*, проводя ее подготовку к анализу. Это включает в себя следующие действия: удаление служебных строк, выделение из строк информации о времени и значении сигнала в каждый момент времени, исследование информации о времени на предмет отсутствия замеров в какой-то момент, заполнение значения сигнала, соответствующего «выпавшему» времени величиной «—1». После выполнения этих этапов значения сигнала построчно записываются в файл temporary. txt. Дальнейшее обращение обрабатывающих классов *модуля 4* происходит к содержимому этого файла. Математические алгоритмы: быстрое преобразование Фурье, вейвлет Хаара, вейвлеты Добеши и вейвлеты Морле составляют библиотеку обрабатывающих классов.

Результатом работы сервиса является графическое или табличное представление суточных колебаний массовой концентрации субмикронной фракции атмосферного аэрозоля, которое получается после работы *модуля 5* или *модуля 6*.

### Гис-приложение

Использование картографической привязки данных не только повышает уровень визуализации данных, но в некоторых случаях способствует получению нового качественного представления о рассматриваемой ситуации. Наглядное отображение данных на географической карте чрезвычайно полезно при решении задач мониторинга.

В качестве ГИС-приложения было выбрано программное обеспечение jMapper<sup>4</sup>. Это Javaапплет, предназначенный для визуализации картографических данных в формате ArcView. Он не является полнофункциональной ГИС, что связано с тем, что основной целью его созда-

<sup>&</sup>lt;sup>4</sup> ГИС-приложение Jmapper: http://www.uiggm.nsc.ru/uiggm/geology/.

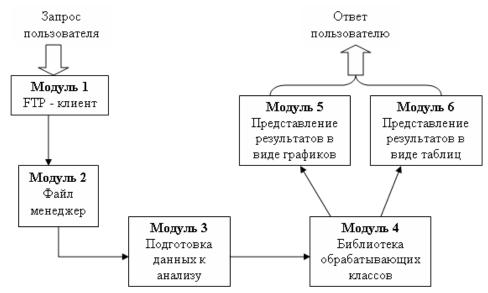


Рис. 3. Структура сервиса, реализующего вейвлет-анализ

ния была разработка программы способной отображать картографическую информацию и производить простейшие манипуляции с ней в сети Интернет. Апплет јМаррег позволяет обращаться к базе данных. Пример работы ГИС-приложения Атласа представлен на рис. 4.

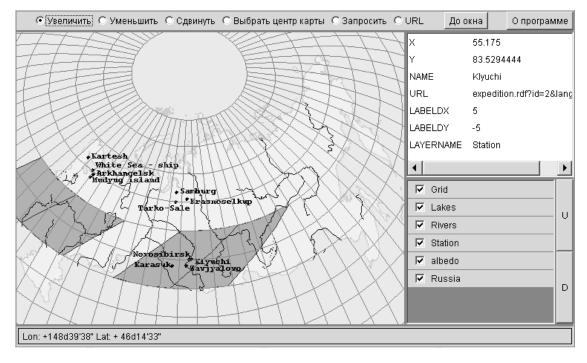


Рис. 4. ГИС-приложение

На карте отображаются все станции отбора проб AA, данные с которых поступают в систему. В окне представления метаинформации (правая верхняя часть карты) отображаются координаты каждой станции. Выбрав опцию URL в верхней части карты и нажав на название станции, откроется список экспедиций, проводимых на данной станции.

### Нахождение коэффициента Альбедо

Институтом космических исследований РАН (Москва) проводится сбор данных, представляющих собой спутниковые снимки различных участков Земли. По этим снимкам можно вычислить характеристику отражения света от земной поверхности – anb6edo, равную отно-

шению отраженного потока к падающему. Далее выявляется влияние АА на способность поверхности отражать падающий на нее поток электромагнитного излучения или частиц.

Результаты космической съемки хранятся на сервере ИХКиГ СО РАН. Каждый файл альбедо содержит массив значений альбедо точек земной поверхности, принадлежащих одной грануле определенного размера. Гранула (tile) — это квадратная область на карте, полученная синусоидальной проекцией Земли. Размер одного файла составляет почти двести мегабайт. Передача данных таких объемов между серверами является трудоемкой и неуместной задачей. С целью извлечения из файлов альбедо необходимой информации реализован сервис альбедо (http://194.85.127.206:8080/albedo/).

Сервис [Dubrov, 2005] предоставляет следующую функциональность:

- получение списка файлов альбедо для заданного промежутка времени либо для заданной точки и промежутка времени;
  - получение значения альбедо для заданной точки и промежутка времени;
- получение гранул с известным значением альбедо (гранул, для которых есть данные хотя бы для одного момента времени) в виде ГИС-данных (которые впоследствии визуализируются программой јМаррег области на карте, выделенные зеленым цветом).

На рис. 5 представлена схема взаимодействия *сервиса альбедо* с атласом «Атмосферные аэрозоли Сибири».

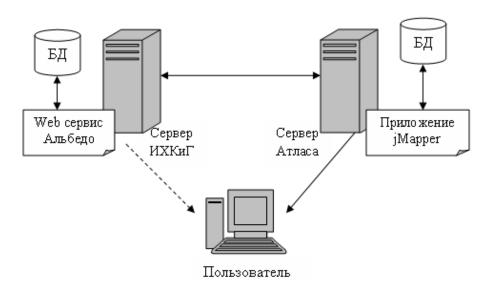


Рис. 5. Взаимодействие сервиса альбедо с Атласом

Ответ пользователю может происходить напрямую с сервера ИХКи $\Gamma$ , где расположена база данных космической съемки. Если работать с географической  $\Gamma$ ИС картой системы – jMapper, то результат запроса клиента поступит на сервер Атласа, а затем, соответственно, пользователю системы.

Сервис альбедо предоставляет возможность интеграции существующей базы данных альбедо, представленной в виде набора файлов, с атласом «Атмосферные аэрозоли Сибири».

При обращении к сервису альбедо необходимо задать географические координаты (долготу и широту) интересующей нас местности, временной интервал, в течение которого производилась съемка. После отработки запроса мы получаем таблицу с нужными нам коэффициентами альбедо (рис. 6).

### Определение направления потока воздушных масс по данным NOAA

Портал HYSPLIT Model расположен на сервере Лаборатории Воздушных Ресурсов ARL (The Air Resources Laboratory) под эгидой Национальной Океанической и Атмосферной Администрации NOAA (National Oceanic and Atmospheric Administration)<sup>5</sup>. Модель смешан-

<sup>&</sup>lt;sup>5</sup> The Air Resources Laboratory: http://www.arl.noaa.gov/.



Рис. 6. Итог работы сервиса альбедо

ной однородной лагранжевой интегрированной траектории HYSPLIT Model (Hybrid Single-Particle Lagrangian Integrated Trajectory Model) — это новейшая версия комплексной системы для вычислений, начиная с простых частичных воздушных траекторий до моделирования сложного рассеивания и перемещения частиц. Эта модель может быть запущена интерактивно в Интернет через систему READY на сайте <sup>6</sup>.

Для успешной работы этой модели необходимо непременно на портале заполнять множество форм, задающих входные данные для этой модели. С целью упрощения работы с сервером портала HYSPLIT Model был создан сервис, основной задачей которого является организация быстрого доступа к программам портала HYSPLIT Model через сайт атласа «Атмосферные аэрозоли Сибири». Основная цель данного сервиса заключается в построении графика облака рассеивания частиц, получаемого по моделям переноса и рассеяния аэрозольных частиц, заложенных в алгоритмах климатологии портала HYSPLIT. Структура работы сервиса представлена на рис. 6.

Пользователь Атласа заполняет форму с входными данными, указывая следующее:

- год и месяц, первую или вторую половину указанного месяца, за который необходимо получить траекторию перемещения воздушных масс;
  - высоту движения воздушных масс, задается в метрах;
- географические координаты (широта и долгота) точки, к которой (или из которой) двигались воздушные массы;
- направление траектории движения (прямая к указанной точке или обратная из точки).

После отправки формы с данными подготавливаются НТТР заголовки<sup>7</sup>. В соответствии со спецификацией НТТР, этот протокол поддерживает передачу служебной информации от сервера к браузеру, оформленной в виде специальных заголовков. После GET или POST-строки могут следовать и другие строки с информацией, разделенные символом перевода строки. Их обычно формирует браузер. Такие строки называются заголовками (headers), и их может быть сколько угодно. Протокол НТТР как раз и задает правила формирования и интерпретации этих заголовков. Он представляет собой ни что иное, как просто набор заголовков, которыми обмениваются сервер и браузер. Не все заголовки обрабатываются сервером — некоторые просто пересылаются запускаемому сценарию с помощью переменных окружения. Переменные окружения представляют собой именованные значения параметров, которые операционная система (точнее, процесс-родитель) передает запущенной программе. Про-

<sup>&</sup>lt;sup>6</sup> Система READY портала HYSPLIT Model: http://www.arl.noaa.gov/ready.html.

 $<sup>^7</sup>$  Оптимизация HTTP заголовков страницы: http://webkomora.com.ua/ru/articles/web/raskrutka/httpoptimisation. html.

грамма может с помощью специальных функций получить значение любой установленной переменной окружения, указав ее имя.

Далее HTTP заголовки отправляются на сервер портала HYSPLIT Model, где производится расчет данных по моделям NOAA с учетом входных данных, отправленных через форму информационной системы «Атлас атмосферных аэрозолей Сибири». По расчетам модели строится график переноса воздушных масс.

#### Заключение

Создана информационно-вычислительная среда для хранения и обработки эмпирического материала по атмосферным аэрозолям Сибири и арктического побережья России. Организовано хранение огромного эмпирического материала, всех описаний и взаимосвязей экспонатов в единой базе данных, что позволит сохранить огромный объем научных результатов для следующего поколения исследователей.

Применение для создания модели информационных ресурсов региона той же методологической базы, что и для описания задач, при решении которых используются эти ресурсы, облегчает организацию взаимодействия между моделями задач и средств их решения. В результате может быть создана единая распределенная многофункциональная среда информационной поддержки принятия решений в области экологии окружающей среды и управления сохранением природных ресурсов. Такая среда предоставит лицам, принимающим решения на различных уровнях управления, средства анализа и оценки принимаемых решений, что обеспечит интеграцию имеющихся в регионе интеллектуальных и информационных ресурсов и вычислительных мощностей.

### Список литературы

*Гордов Е. П., Родимова О. Б., Фазлиев А. 3.* Атмосферно-оптические процессы: простые нелинейные модели. Томск: ИОА СО РАН. 2002. 251 с.

*Гордов Е. П., Фазлиев А. 3.* Научные информационные ресурсы для поддержки исследований об атмосфере в сети Интернет // Выч. технологии. 2003. Т. 9. Спец. выпуск, ч. 1. С. 123–136.

Гордов Е. П., De Rudder А., Лыкосов В. Н. и др. Веб-портал АТМОС как основа для выполнения интегрированных исследований по окружающей среде Сибири // Выч. технологии. 2004. Т. 9. Спец. выпуск, ч. 2. Тр. междунар. конф. «Вычислительно-информационные технологии для наук об окружающей среде». С. 3–13.

Дмитриев Б. Н., Суторихин И. А. Базы данных для ГИС «Аэрозоли Сибири» // Тр. междунар. конф. «ENVIRONMIS-2002» / Под ред. проф. Е. П. Гордова. Томск: Изд. ГУ «Томский ЦНТИ». 2002. Т. 1. С. 70–77.

Малла С. Вейвлеты в обработке сигналов. М.: Мир. 2005. 671 с.

Молородов Ю. И., Куценогий К. П. Атлас по атмосферным аэрозолям Сибири как основа обеспечения мониторинга Сибирского региона: Сб. науч. тр. 2005. Т. 5: «Мониторинг окружающей среды, геоэкология, дистанционные методы зондирования земли». Новосибирск: ГОУ ВПО «Сибирская государственная геодезическая академия». С. 142–146.

Федотов А. М., Барахнин В. Б., Гуськов А. Е. и др. Распределенная информационно-аналитическая среда для исследований экологических систем // Выч. технологии. 2006. Т. 11. Спец. выпуск, ч. 1.

*Dubrov I. S.* Visualization albedo using data from cosmix observations (Novosibirsk State University) // Proceedings of SPIE. 2005. Vol. 6160OL, pt. I. P. 21–25.

*Guskov A., Shraibman V., Molorodov Y.* SMART – an approach for information systems development on the basis of RDF-technology // Proc. of the VIII International Conference on Electronic Publishing. Brasilia, 2004. P. 293–305.