

Сервис для обработки удаленных рядов данных на основе Wavelet-функций*

© Е.В. Корсаков, Ю.И. Молородов

Институт вычислительных технологий СО РАН
yumo@ict.nsc.ru

Аннотация

Описан сервис для математической обработки временных рядов данных наблюдений, хранящихся на сервере-хранилище информации. Доступ к данным и программам обработки организован с помощью Веб-интерфейса. Сервис обеспечивает доступ к данным и их обработку с помощью вейвлетов Морле и Добеши. Цель работы – создание доступной по сети Интернет распределенной информационно-вычислительной среды для обработки результатов измерений.

1 Введение

В последние годы возрос интерес к изучению проблемы глобальных изменений окружающей среды. Особую роль при этом играют атмосферные аэрозоли (АА). Они – неотъемлемая компонента атмосферы и объяснение всех сложных процессов, происходящих в ней, невозможно без подробной информации об их химическом составе и поведении. Специалистам, занимающихся прогнозированием, моделированием и анализом процессов, происходящих в окружающей среде необходимо своевременно получать информацию о характеристиках, химическом составе и поведении атмосферного аэрозоля.

Для этого необходимо разрабатывать информационные системы, в которых информация будет структурирована и собрана в виде, доступном для понимания пользователю любой степени подготовки. Эту проблему, как и задачу выявления регулярных сезонных и суточных колебаний в поведении аэрозолей можно решить, если создать информационно-вычислительные системы на основе архитектуры клиент-сервер. Они решают задачи сбора, обработки и публикации сведений об атмосферных аэрозолях, в том числе:

- Добавление, хранение, редактирование исходных данных.

Труды 10-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2008, Дубна, Россия, 2008.

- Обработку временных рядов данных современными алгоритмами.
- Обеспечение авторизации пользователей и ограничение доступа (при необходимости) к информационным структурам и данным.

Данные могут храниться как на стороне клиента, так и на серверах владельцев информации.

Разработанные интерфейсы должны обеспечить представление исходных данных или результат их обработки в пригодном для анализа виде. Современные математические алгоритмы обработки временных рядов данных наблюдений позволяют проводить их частотно-временной анализ, в том числе и с помощью Вейвлет-анализа [1–4].

2 Алгоритмы обработки и анализ

2.1 Расчет функции спектр размеров

Ранее [5] мы описывали метод измерения характеристик аэрозолей и методик обработки на основе регуляризации по Тихонову. Эти алгоритмы позволяют решить обратную задачу восстановления спектра размеров из разнообразных данных, полученных при измерениях. Спектр размеров является предпочтительным описанием для включения АА в какие-либо модельные расчеты. Дифференциальные характеристики атмосферного аэрозоля, как правило, представляются в форме, имеющей аналитическую зависимость от приведенного размера частиц, называемого радиусом или диаметром. Соответственно, для получения интегральных характеристик атмосферного аэрозоля, имеющих дальнейшее прикладное значение, необходимо проводить интегрирование дифференциальных характеристик с учетом спектра размеров. Программный модуль позволял привести многочисленные и разрозненные данные по физическим свойствам атмосферного аэрозоля к единообразной форме. В качестве такой формы мы использовали счетное распределение частиц по размерам $dN / d(\log(r))$. Для полидисперсного аэрозоля, значение измеренной концентрации $P(m)$ ($m = 1, \dots, M$), M – число экспериментов) может быть выражено через нормированный спектр размеров аэрозоля:

$$f(x) = dN / (2.303 * N * d(\log(r))). \quad (1)$$

В нашем случае $P(1) = N$ – полная счетная концентрация. Зная $f(x)$, можно восстановить искомую функцию «спектр размеров». Тогда спектр размеров

по счетной концентрации представляется в виде числа частиц dN в интервале размеров $d(\log(r))$.

2.2 Вейвлет-анализ: основные определения

Вейвлет-анализ – ориентирован на локализацию и классификацию особых точек сигнала, вычисление его различных фрактальных характеристик, частотно-временной анализ нестационарных сигналов.

Следуя [1] опишем основные принципы анализа временных рядов. Вейвлет-преобразование $Wf(x, a)$ одномерного сигнала $f(t)$ заключается в его разложении по базису, сконструированному из некоторой специальной материнской базисной функции (вейвлета) – $\Psi(t, x, a)$ посредством масштабных изменений a и переносов x :

$$(W_{\Psi}f)(x, a) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} \Psi\left(\frac{t-x}{a}\right) f(t) dt \quad (2)$$

При каждом масштабе a , согласно (2) базисный вейвлет растягивается по горизонтали и сжимается по вертикали.

Т.о., в отличие от преобразования Фурье, вейвлет-преобразование обеспечивает двухмерную развертку исследуемого одномерного сигнала, при этом масштаб и координата рассматриваются как независимые переменные. В результате этого появляется возможность анализировать свойства сигнала одновременно во временном и в частотном пространствах.

Для анализа вариаций мы будем использовать вейвлет Морле. Он хорошо приспособлен для анализа квазипериодических процессов. Он позволяет наглядно представить результаты анализа и легко интерпретировать их в терминах анализа Фурье.

$$\Psi(t) = \frac{1}{2\pi} \exp\left(i\omega_0 t - \frac{t^2}{2}\right) \quad (3)$$

В общем случае выбор анализирующего вейвлета является неоднозначным и зависит от конкретной задачи. Вейвлет Морле был выбран потому, что он хорошо приспособлен для анализа квазипериодических процессов (атмосферные вариации, являющиеся собственными колебаниями атмосферы [3], принято считать квазипериодическими), также он позволяет наглядно представить результаты анализа и легко интерпретировать их в терминах анализа Фурье.

2.3 Структура сервиса обработки данных

Сервис обработки, реализующий возможности Вейвлет-анализа, состоит из пяти функциональных модулей (рис. 1).

По запросу клиента выполняется авторизация и подключение к удалённому серверу по протоколу FTP. Затем клиент просматривает содержимое каталогов, выбирает интересующие его файлы и копирует их на сервер-обработчик. *Модуль подготовки данных* к анализу последовательно извлекает информацию из файлов, и готовит их к анализу. При этом удаляется служебная информация, из строк выделяются сведения о времени и амплитуде сигнала

ла в каждый момент времени, выделяется «сбойные» показания (сбои аппаратуры регистрации), сигнал, соответствующий «выпавшему» времени заменяется на «-1». После этого значения сигнала построчно записываются в файл temporgay.txt. Далее модуль подготовки передает файл на кластер. Обращение к кластеру вызвано достаточно большими затратами времени на обработку. На кластере хранятся алгоритмы: быстрого преобразования Фурье, а также необходимые для работы вейвлет-функции: вейвлет Хаара, вейвлеты Добеши и вейвлеты Морле. По окончании обработки посылается сообщение клиенту. После этого клиент сможет посмотреть таблицы обработки и нарисовать графики.

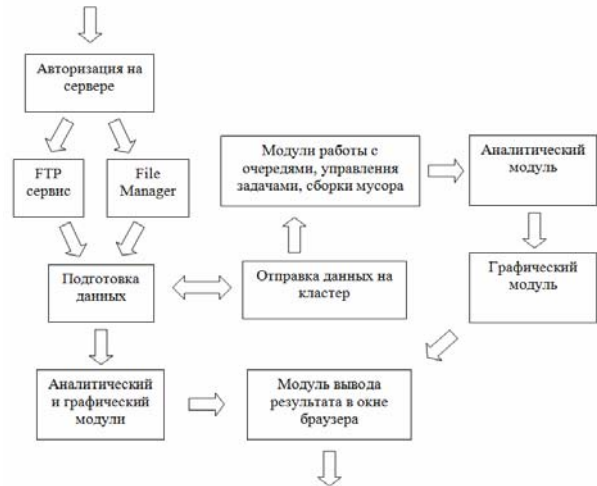


Рис. 1. Общая схема сервиса

2.4 Схема измерений

В пригородной зоне г. Новосибирска в п. Ключи (9 км от Академгородка) находится станция наблюдения за состоянием пограничного слоя атмосферы. Здесь проводятся измерения массовой концентрации субмикронной фракции атмосферного аэрозоля. Для этого используется нефелометр ФАН-А, связанный с компьютером IBM-PC класса Pentium IV. Измерения проводятся в течение одного месяца каждого сезона с частотой одно измерение в минуту (1440 замеров/сут.).

Файлы данных переносятся для постоянного хранения на сервере организации, хозяина данных.

2.5 Результаты обработки

Для работы с данными мы использовали ортогональный вейвлет-анализ на основе вейвлетов Добеши. Они позволяют нам исключить шум сигнала, сжать его, а также выделить его низкочастотную и высокочастотную компоненты. Анализ данных на основе вейвлетов Морле ориентирован на локализацию и классификацию особых точек сигнала, вычисление его различных фрактальных характеристик, частотно-временной анализ.

На рис. 2–3 представлены результаты обработки ряда данных измерений, полученных в течение суток 5 ноября 2006 г. в пункте наблюдений Ключи.

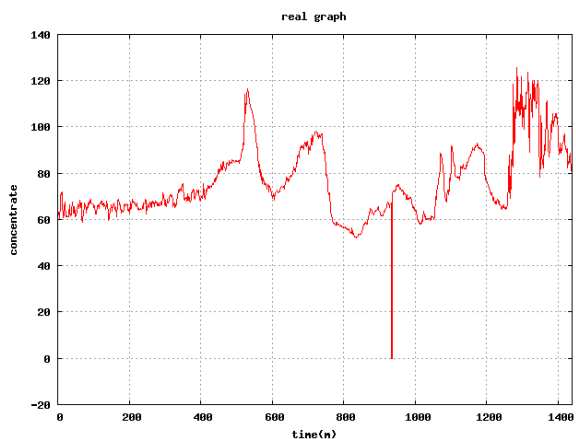


Рис. 2. Исходный сигнал

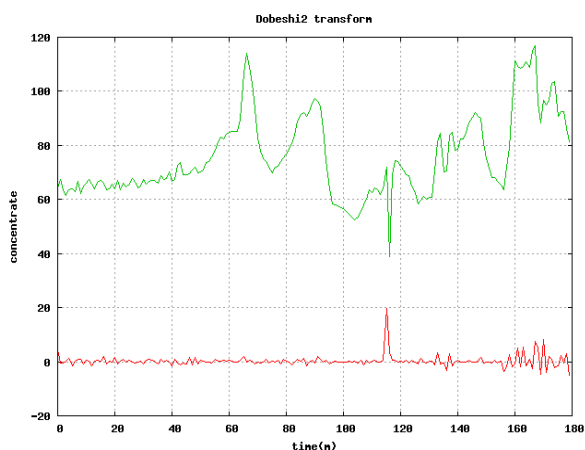


Рис. 3. Обработка Вейвлетом Добеши



Рис. 4

На рис. 3 показана низкочастотная составляющая сигнала (верхний график) и его высокочастотная компонента (нижний график). На рис. 4 показан результат обработки среднечасового сигнала полученного за 1 месяц в январе/феврале 2006 г. На верхнем графике показано его среднее значение, а на нижнем – среднее квадратичное отклонение от среднего. Поскольку на нем явно не видны какие-либо периодические структуры, то это сигнал был обработан с помощью вейвлетов Морле. Рис. 5 показывает результат обработки. Более явно показаны

периодические структуры на рис. 6 (увеличена нижний фрагмент рис. 5).

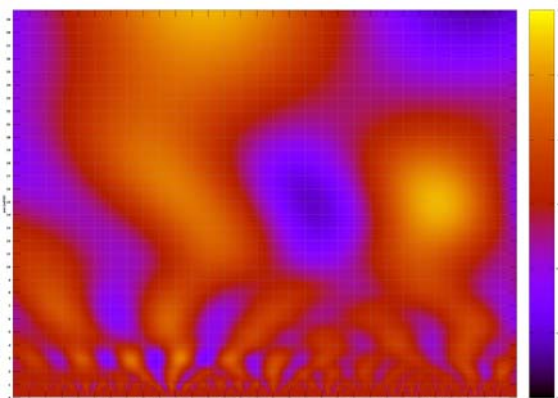


Рис. 5. Вейвлет-образ зима. Функции Морле

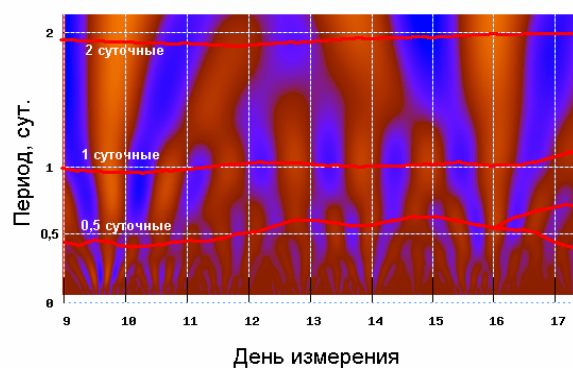


Рис. 6. Пространственно-временная изменчивость массовой концентрации субмикронной фракции атмосферного аэрозоля в пригородной зоне Новосибирска (зима 2006 г.)

Поскольку целью работы являлась только демонстрация возможностей метода обработки, то мы не показали обработку временных рядов данных полученных в весенний, летний и осенний периоды 2006 г.

3 Заключение

Ортогональное вейвлет преобразование Добеши используется для подавления шумов в сигнале. Как видно из полученных результатов быстрое вейвлет преобразование (БВП) Добеши с $M = 1$, не совсем подходит для подавления шумов в сигнале, что видно из рисунка раздела (3.2). Данное преобразование снизило среднее значение концентрации, хотя закономерности в поведении сигнала все же сохранились. Вейвлет преобразование с $M = 2$ (3.3), сохраняет среднее значение концентрации в сигнале в отличие от вейвлета Хаара.

Непрерывное вейвлет преобразование Морле может быть использовано для локализации и классификации особых точек, частотно-временного анализа нестационарных сигналов. Преобразование Морле мы также использовали для диапазона измерений полученных за один месяц.

Литература

- [1] Хуторова О.Г., Тептин Г.М. Исследование мезомасштабных вариаций в тропосфере по наблюдениям концентрации примесей // Известия РАН. Физика атмосферы и океана/ – 2001, т. 37, №6. – С. 853–856.
- [2] Torrence G. and Compo G.P. A practical guide to wavelet analysis. – Bull. Amer. Meteorol. Soc., 1998, vol.79, No 1, pp. 61–78.
- [3] Астафьева Н.М. Вейвлет-анализ: основы теории и примеры применения. – Т. 166, № 11. – С. 1145–1170.
- [4] Шитов А.Б. Разработка численных методов и программ, связанных с применением вейвлет-анализа для моделирования и обработки экспериментальных данных // Дис. на соискание ученой степени канд. физ.-мат. наук. – ИвГУ, 2001. – С. 125.
- [5] Молородов Ю.И., Куценогий П.К., Сударикова И.А. Содержательное наполнение Атласа «Атмосферные аэрозоли Сибири» // География и природные ресурсы. – 2004. Спец. выпуск. – С. 86–90.

Wavelet-function processing service for remote data rows

E. Korsakov, Yu. Molorodov

The technology and service was developed for remote processing of regular measurements experimental data files of submicronic fraction atmospheric an aerosol (AA), received by measurements of computerized nefelometer a FAN in a ground layer of the atmosphere, near the Novosibirsk (settlement Kluchi). Necessity of such data processing of measurements for the atlas is caused not only ordering of the saved up empirical numbers, but also carrying out of statistical processing together with other components AA, for example, organic and element carbon, elements etc. The purpose of offered processing is revealing of cyclic laws of a signal: daily, monthly, seasonal fluctuations and definition of local upper limit. Service provides connection to a remote data storage server, the primary analysis and preparation of the data on a server output agent and carrying out of the analysis on the basis of wavelet-functions of Naara, Dobeshi and Morlet. Results of the analysis are represented in the form convenient for the further analyze.

* Работа поддержана РФФИ (грант № 06-07-89060), президентской программой «Ведущие научные школы РФ» грант № НШ 931.2008.9 и Интеграционным проектом СО РАН № 34.