

WINABD

ПАКЕТ ПРОГРАММ ДЛЯ СОПРОВОЖДЕНИЯ И АНАЛИЗА ДАННЫХ ГЕОФИЗИЧЕСКОГО МОНИТОРИНГА

А.В. Дещеревский, В.И. Журавлев, А.Н. Никольский

1. Программы для обработки данных мониторинга

Ряды данных, получаемые при геофизическом мониторинге, содержат не только полезный сигнал, но и непредсказуемо меняющиеся помехи, пропуски, разные нестационарные эффекты. Для обработки таких данных используются разные среды:

- В универсальных пакетах статистического анализа часто отсутствуют важные для мониторинга инструменты – сложно организовать сопровождение базы данных, обработку пропусков или календарную синхронизацию данных.
- В полуоткрытой среде можно использовать готовые функции и/или разрабатывать новые алгоритмы для решения нестандартных задач. Это требует хороших навыков программирования и глубокого освоения среды, но все равно не избавляет от всех проблем. Так, в среде матлаба (одной из лучших на сегодняшний день) нет встроенных инструментов для организации базы данных, большинство функций не работают с пропусками, а инструменты для календарной синхронизации рядов данных надо программировать самому, так как имеющиеся в пакете средства крайне ресурсоемки.
- Многие научные коллективы в России ориентируются на программы собственной разработки. Но в этом случае сложно реализовать большой спектр методов усилиями компактного коллектива. Такие программы часто недостаточно универсальны, плохо документированы и требуют постоянно авторского сопровождения.

Анализ показывает, что все эти варианты имеют серьезные недостатки. Поэтому наш коллектив разработал программу WinABD, реализующую весь спектр ключевых технологий, необходимых при исследовательской работе с экспериментальными временными рядами.

2. Структура пакета WinABD

- система управления базой данных временных рядов,
- мощный исследовательский комплекс
- интерактивная среда визуализации данных.

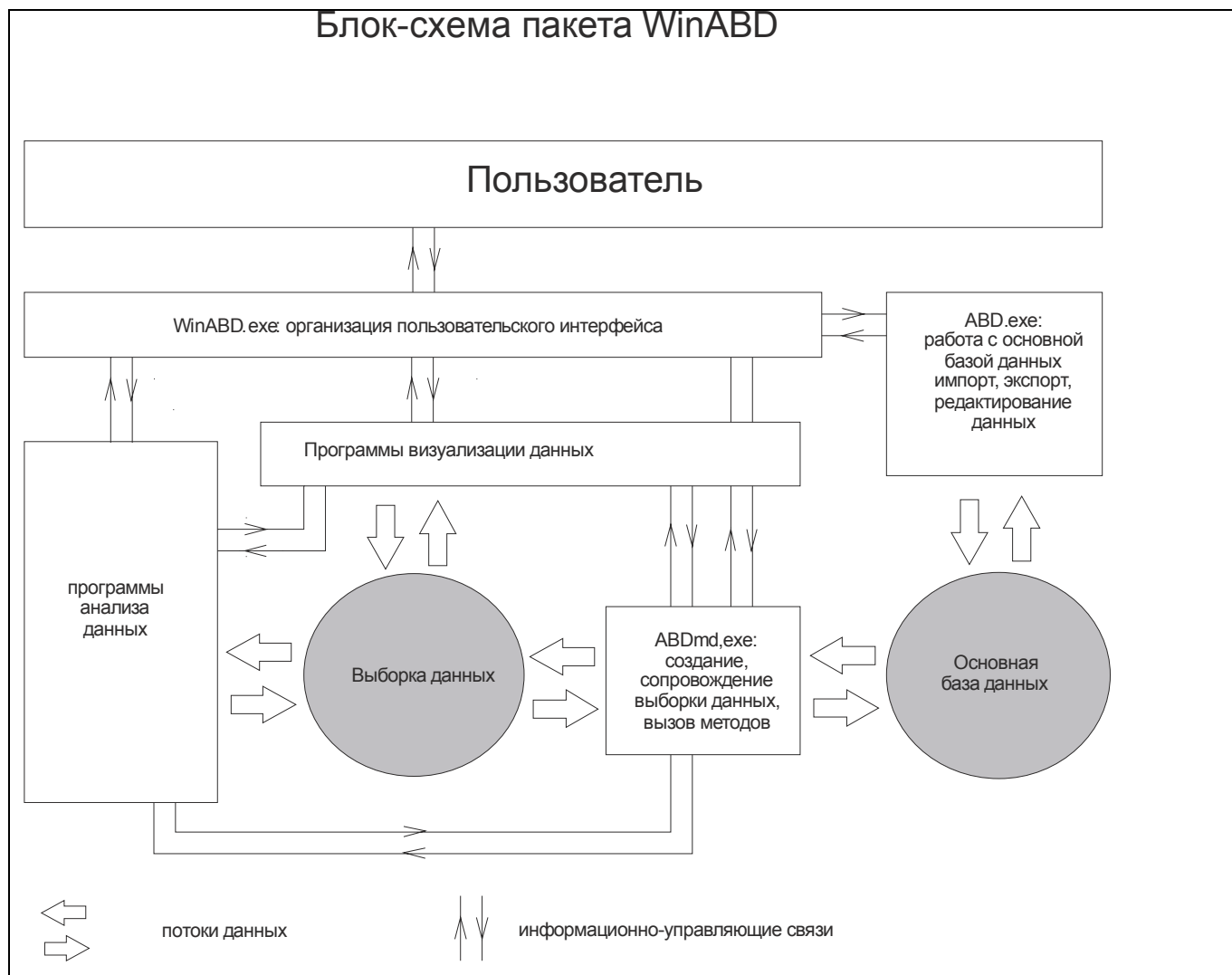


Рис.1 Блок-схема пакета WinABD

Особенности пакета:

- Большое количество инструментов обработки и анализа данных
- Никаких ограничений при анализе данных с пропусками
- Обработка в скользящем окне
- При всех операциях с данными используется календарное время
- Корректная совместная обработка рядов, имеющих неодинаковые даты начала и несовпадающую периодичность наблюдений

3. База данных WinABD

- Данные хранятся в компактном формате (16 или 32 бита на значение). На типовой персоналке можно работать с рядами длиной миллиарды точек.
- Каждый ряд – это бесконечная последовательность ячеек, имеющих точную календарную привязку. Интервал между наблюдениями от 0.000001с (1 мегагерц) до 999 лет.
- При импорте данные могут добавляться к уже существующему ряду с соблюдением хронологии, что необходимо при продолжающихся наблюдениях. Есть инструменты для автоматизации импорта.
- Предусмотрена возможность работы с каталогами землетрясений.

Импорт данных в базу требует определенных усилий, но позволяет впоследствии работать не с именами файлов (столбцов), а с хорошо паспортизированными выборками, что существенно повышает удобство работы.

4. Два уровня доступа к данным

- В WinABD реализована двухуровневая система доступа к данным. Первичная информация хранится в основной базе данных (рис.1). Все алгоритмы анализа оперируют не с первичными данными, а со специально создаваемыми копиями этих рядов – так называемой «выборкой данных». Это гарантирует, что первичные данные ни при каких условиях не будут повреждены или затерты.
- При создании выборки из базы извлекаются данные, нужные для текущего сеанса работы. Ряды автоматически пересчитываются к согласованной частоте опроса. Это позволяет строить простые и эффективные алгоритмы для совместного анализа данных с неодинаковым графиком наблюдений
- Ряды имеют специальные паспорта, куда в сжатой форме заносится информация о выполненной обработке. Параметры счета и результаты вычислений заносятся в протокол обработки данных.
- Рабочее пространство WinABD хранится не в оперативной памяти, а на диске. Это позволяет работать с массивами данных неограниченного размера. Созданная выборка не теряется даже при аварийном окончании сеанса работы, а вновь автоматически открывается при следующем запуске программы.

5. Концепция пропусков данных WinABD

- Экспериментальные ряды почти неизбежно содержат пропуски данных, но многие алгоритмы обработки данных требуют, чтобы интервал между опросами был постоянным. Можно заполнять пропуски на этапе предварительной подготовки данных. Но явное заполнение пропусков данных всегда основано на какой-то модели сигнала. Обычно можно предложить много таких моделей. Обязательный выбор одной из них еще до начала анализа данных – это очень негибкое решение.
- WinABD использует принципиально иной подход. Модель данных WinABD допускает наличие пропусков и перерывов в наблюдениях на любом этапе анализа. Те методы, которые алгоритмически допускают неравномерный шаг между опросами данных (например, при оценке среднего или ВКФ), используют в таких случаях специально поправленные формулы, по возможности обеспечивающие получение несмещенных оценок. Если же алгоритм не допускает наличия пропусков (например, БПФ), то метод сам заполнит пропуски данных перед проведением вычислений. При желании пользователь может сам заполнить пропуски данных перед вызовом метода, для чего в WinABD имеется множество инструментов.
- Благодаря описанной технологии исследователь получает инструменты, которые очень сложно реализовать при стандартных подходах. Для поиска и выбраковки дефектных наблюдений можно использовать любые имеющиеся в WinABD алгоритмы. Но главное, обеспечивается получение гораздо более обоснованных статистических заключений, чем в случае обязательного заполнения пропусков уже на стадии подготовки данных.

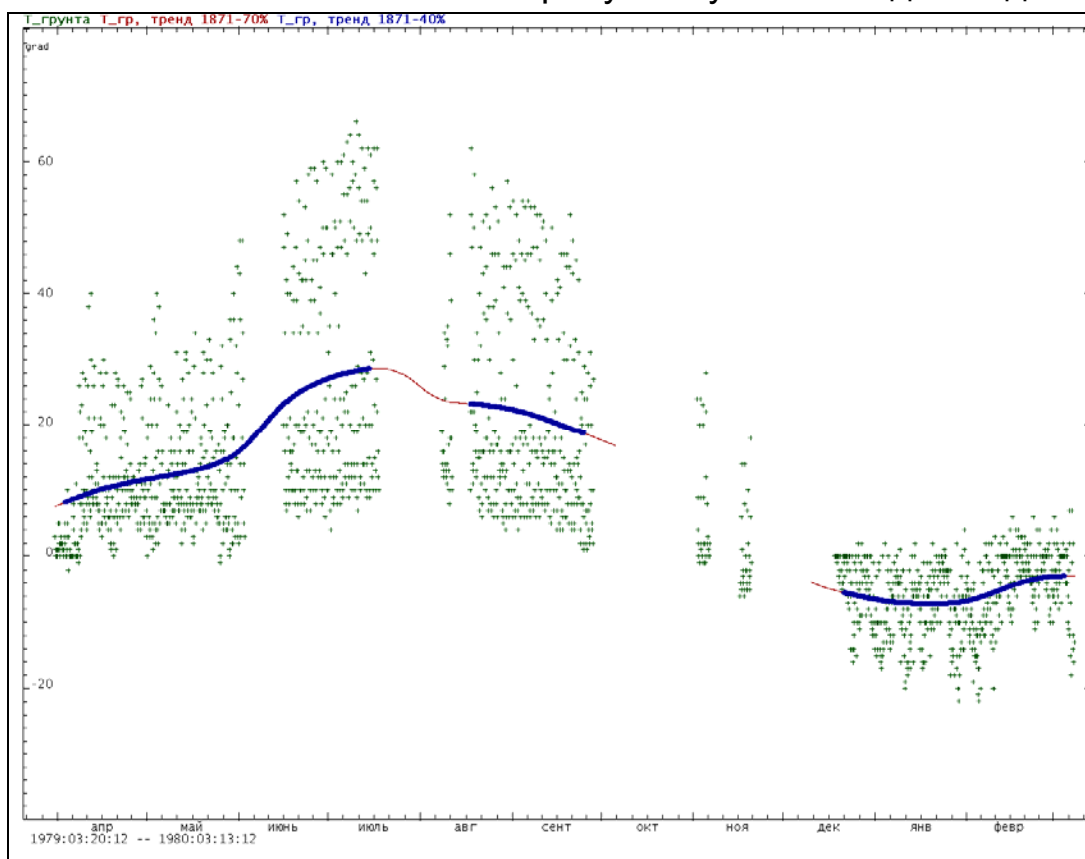


Рис.2. Данные температуры грунта на ст.Лайрон (точки) с интервалами пропусков данных и тренд, выделенный методом ядерного скользящего сглаживания в окне шириной 1871ч с гауссовой весовой функцией окна при разрешенной доле пропусков 70% (тонкая линия) и 40% (жирная линия).

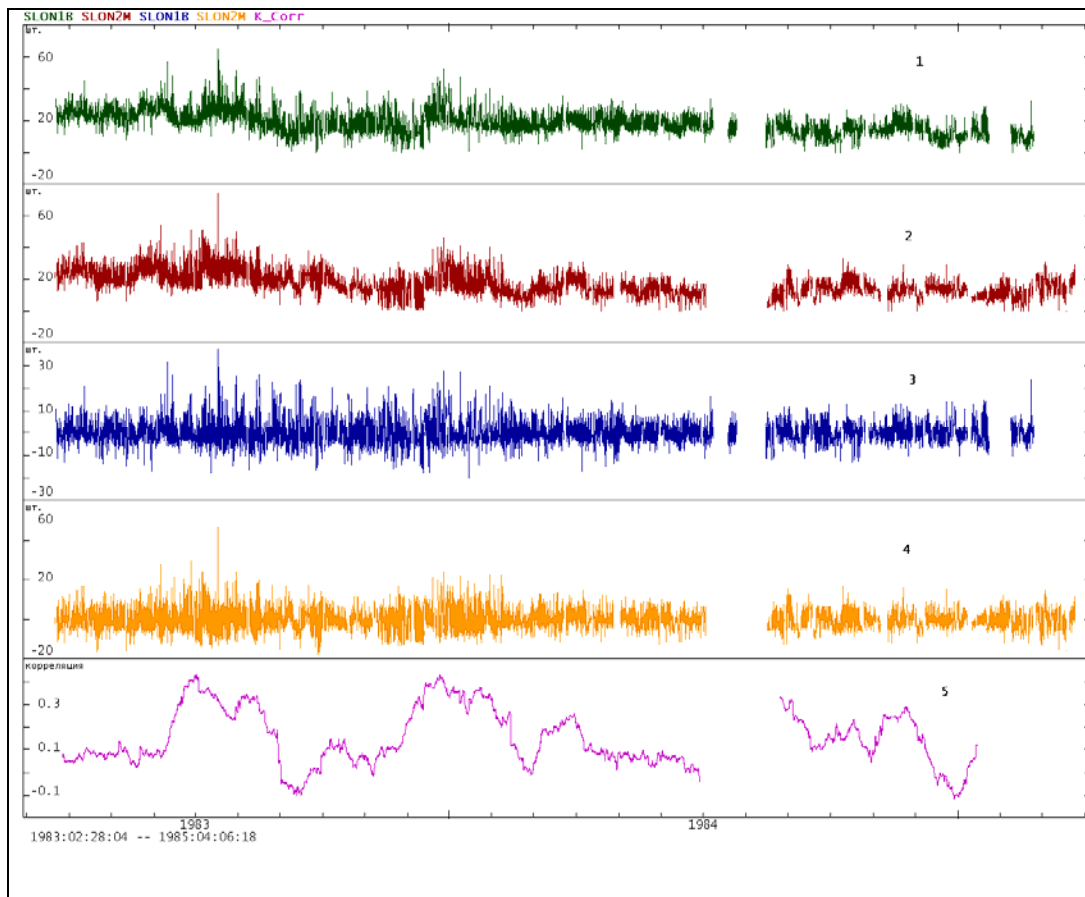


Рис.3. Ряды двигательной активности биоиндикаторов на ст.Гарм (графики 1 и 2, видны интервалы пропусков данных), те же ряды с удаленным трендом (графики 3 и 4) и коэффициент корреляции между ними в скользящем окне шириной 1 мес (график 5), рассчитанный при разрешенной доле пропусков 40%.

6. Правило синхронизации наблюдений

- Любые операции с данными выполняются с учетом времени наблюдений. При совместной обработке сигналов исследователь не должен думать о датах начала-конца каждого ряда, а может сосредоточиться на содержательных аспектах анализа данных.

Для примера рассмотрим сложение двух рядов $A+B=C$. Допустим, что параметр A измерялся ежедневно с 1980 по 2000 год, а параметр B – с 1990 по 2010 год. В этом случае ряд C будет рассчитан для интервала времени с 1990 по 2000 год. Информацию о дате начала и окончания каждого ряда, участвующего в вычислениях, WinABD берет из паспорта ряда.

7. Методы скользящего окна

- Одна из основных задач мониторинга – это отслеживание изменений, происходящих в контролируемой системе. Для этого свойства сигнала оцениваются в пределах некоторого временного окна. Затем окно сдвигается вправо и все вычисления повторяются.
- В WinABD имеется более 20 методов скользящего окна (рис.2-5). Скользящее окно WinABD обычно сдвигается на одну точку, поэтому шаг дискретизации обработанного сигнала равен шагу дискретизации исходного ряда (а не размеру окна). Это позволяет наиболее точно отслеживать момент изменения контролируемых параметров. Комбинируя имеющиеся в WinABD методы скользящего окна, можно реализовать практически любые алгоритмы, которые могут понадобиться при анализе данных.

8. Обработка без уменьшения длины ряда

- Общий недостаток методов скользящего окна – их трудно комбинировать друг с другом. Такие методы требуют, чтобы окно целиком помещалось «внутри» ряда, поэтому полный «пробег» окна всегда меньше, чем длина исходного ряда. Обработанный ряд всегда короче исходного.
- В WinABD все методы скользящего окна могут работать без уменьшения длины отфильтрованного сигнала. В начальный момент времени центр окна совмещается с первой точкой сигнала. Левая половина окна заполняется пропусками. Специальная настройка «разрешенное количество пропусков в скользящем окне» позволяет регулировать величину «выезда» окна за пределы ряда. Аналогично обрабатываются и пропуски внутри ряда (рис.2).
- Описанный механизм весьма гибок и имеет неоспоримые преимущества перед подходом, предполагающим либо обязательное предварительное заполнение пропусков, либо выбраковку всего массива данных при наличии даже незначительного количества «испорченных» наблюдений.

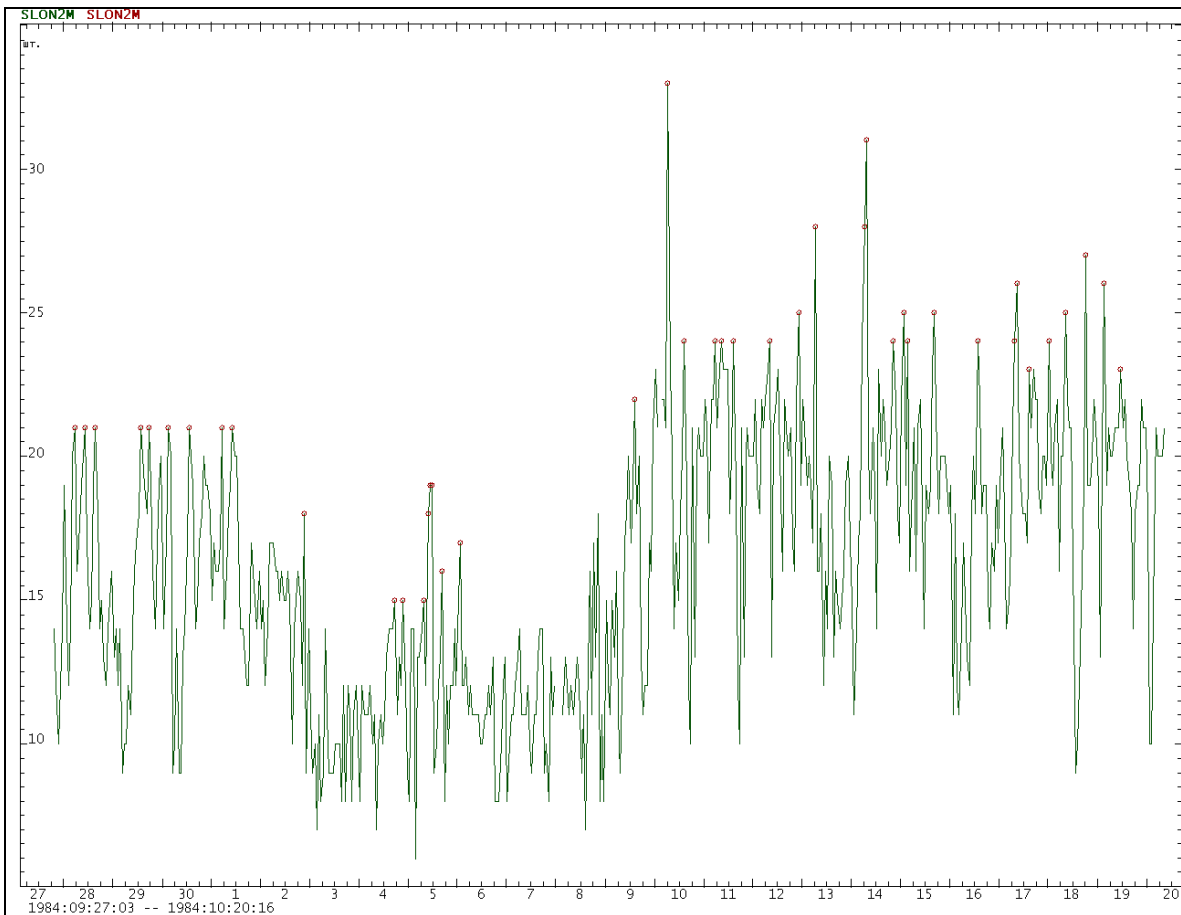


Рис.4. Выделение всплесков двигательной активности биоиндикатора методом квантильного анализа. Ширина окна 73ч, уровень 90%.

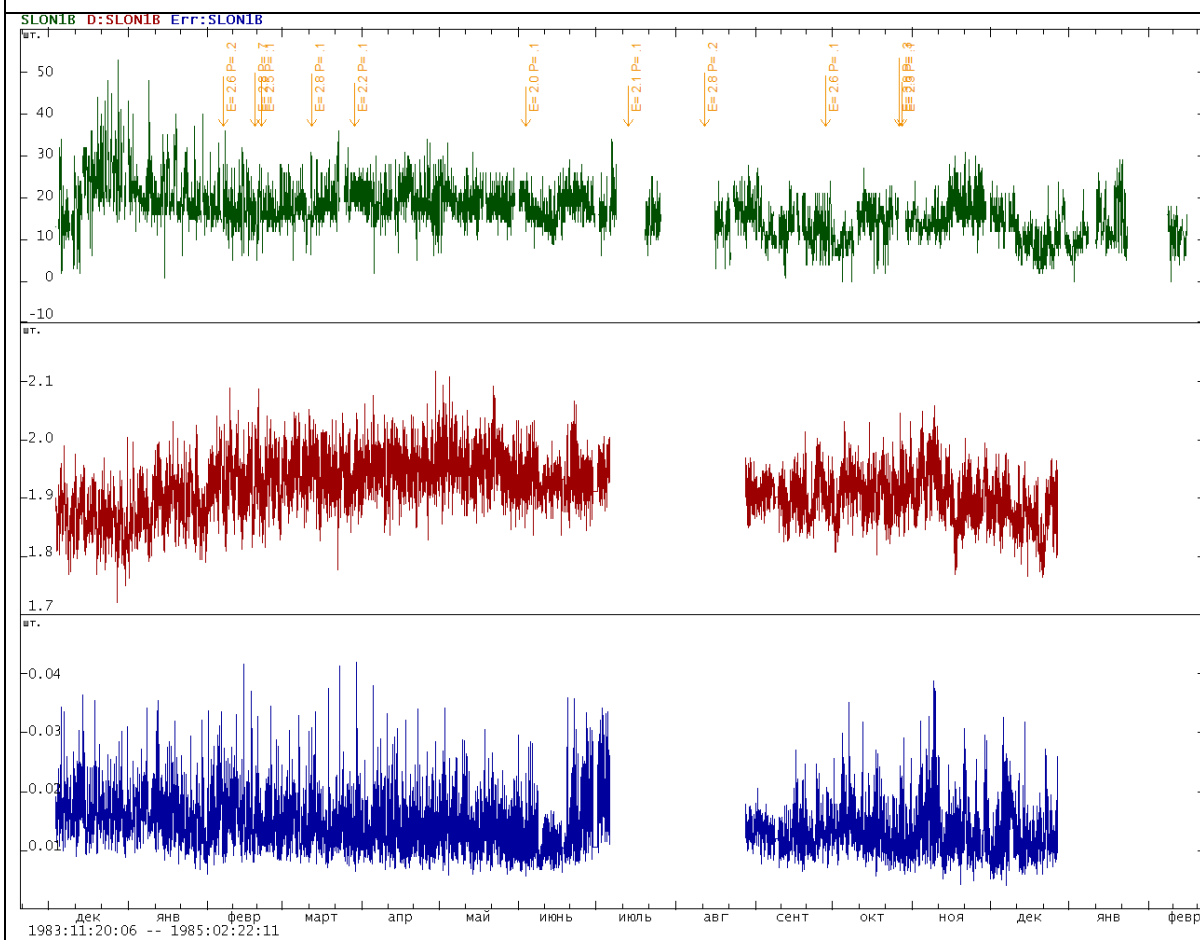


Рис.5. Ряд двигательной активности биоиндикатора (ст.Гарм) и изменения фрактальной размерности ряда, оцененные в окне шириной 1461сут при разрешенной доле пропусков 40%. Нижний график – погрешность оценки фрактальной размерности

9. Визуализация данных

- Основу современных пакетов статистического анализа обычно составляют методы обработки данных, а инструменты визуализации рассматриваются, как полезное дополнение к ним. При выводе графиков основное внимание уделяется оформлению, а не содержанию, что идеально подходит для презентаций, но не слишком удобно при повседневной работе.
- В WinABD средства визуализации данных – это концептуальный, а не вспомогательный инструмент. Работа с любым сигналом начинается с изучения его графика, выбора оптимальных методов обработки, их параметров и настроек. Результаты процессинга также выводятся в виде графиков, что позволяет немедленно оценить эффективность использованного алгоритма и внести необходимые коррективы.
- При поисковой работе наиболее интересны новые, неожиданные эффекты, которые проявляются в отклонении характеристик сигнала от предполагаемых. Автоматические алгоритмы всегда исходят из заранее заданной модели возмущения и сигнала, и далеко не всегда эффективны при обнаружении подобных явлений. Роль исследователя, обладающего необходимыми средствами визуализации данных, остается незаменимой.
- Инструменты визуализации данных WinABD позволяют работать с рядами длиной миллионы точек практически без задержки. Можно «на лету» развернуть любой фрагмент ряда сколь угодно детально. Оси координат размечаются в реальной календарной шкале. Все вспомогательные надписи и поля минимизированы, чтобы наиболее полно использовать площадь экрана для отображения данных.
- При работе с большими массивами данных WinABD выполняет настраиваемую генерализацию графиков. Это не только ускоряет отрисовку изображения, но и кардинально сокращает размер файла с рисунком при его сохранении в векторном формате.

Экран **функций распределения** позволяет не только оценивать параметры распределения для ряда в целом или любого его фрагмента, но и удалять выбросы, выделяя их прямо на графике. Адаптивные методы идентификации выбросов и коррекции функции распределения работают в скользящем окне.

Взаимная диаграмма – важнейший инструмент визуального анализа данных (рис.6). Можно не только оценивать параметры регрессии, но и визуализировать траекторию процесса в режиме «мультфильма». Временная координата может отображаться цветом (рис.7).

Структурные и корреляционные функции полезны при разведочном анализе (рис.8).

Спектры и периодограммы. В геофизике применяется множество спектральных методов, но почти все они предназначены для исследования гармонических колебаний. Реальные циклические процессы – сезонные колебания электропроводности и осадков, суточные колебания влажности и давления; биологические и экономические ритмы – очень редко можно аппроксимировать синусоидой. В такой ситуации может пригодиться метод периодограмм наложенных эпох, существенно развитый авторами.

Фрактальные свойства рядов. Строго говоря, временной ряд не является обычным фракталом, поскольку временная и физическая переменные (координаты) в данном случае существенно неравноправны, что не позволяет ввести метрику в соответствующем двумерном пространстве. Тем не менее, с помощью определенных приемов фрактальные свойства таких рядов поддаются исследованию. В WinABD для этого есть R/S-анализ, метод показателя Херста, а также метод фрактальных длин (рис.9).

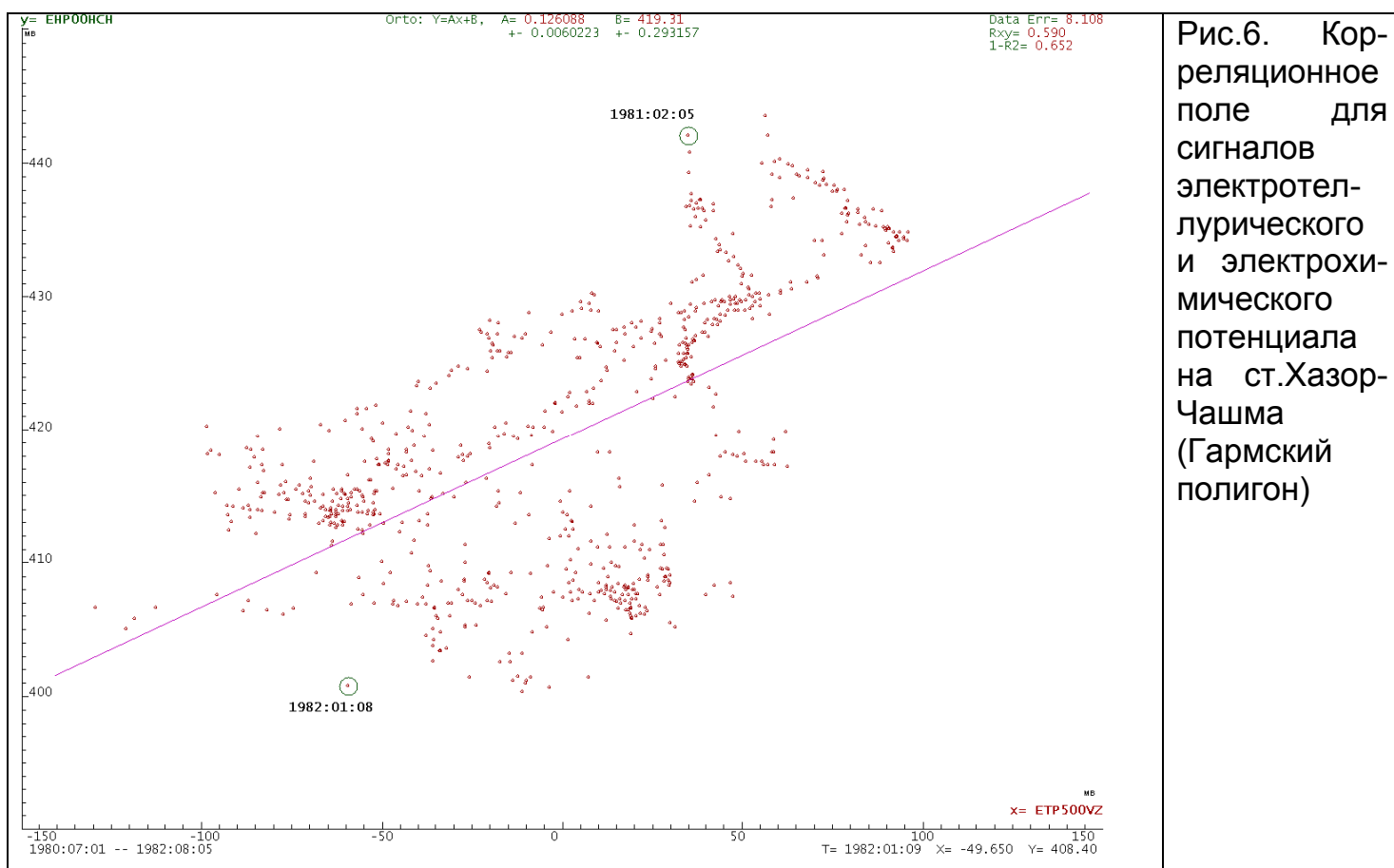


Рис.6. Корреляционное поле для сигналов электротеллурического и электрохимического потенциала на ст.Хазор-Чашма (Гармский полигон)

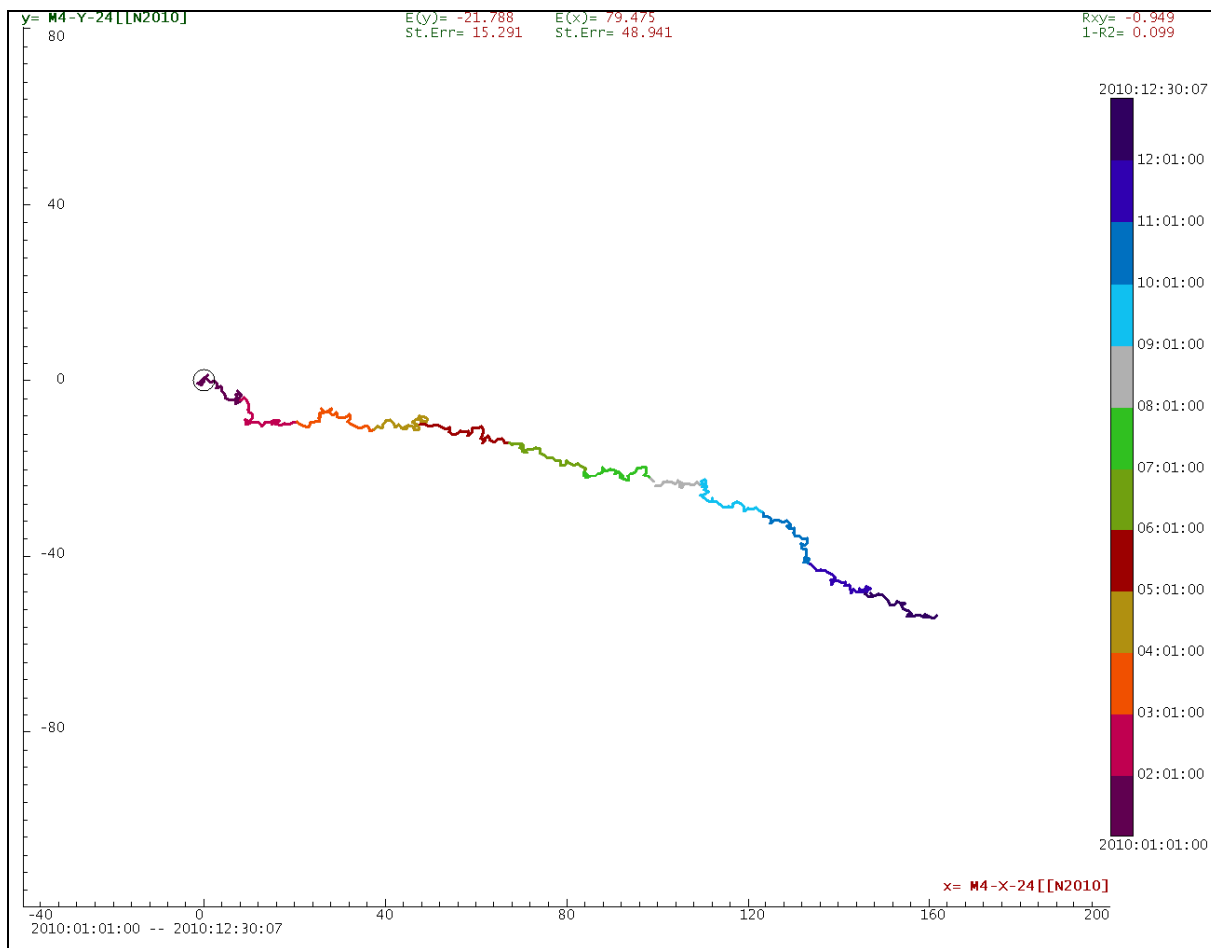


Рис.7. Годограф Шустера для ряда чисел землетрясений Греции, 2010г.

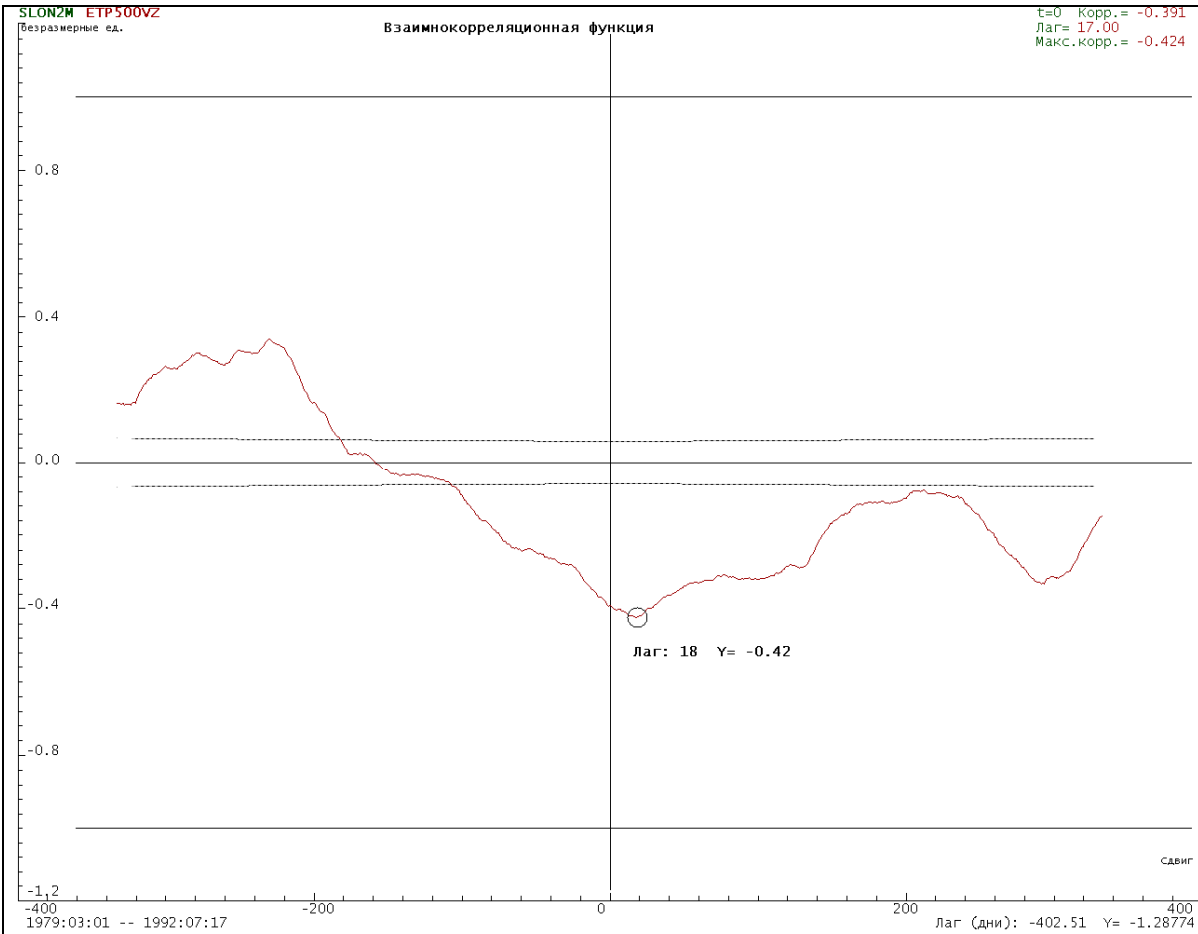


Рис.8. Взаимно-корреляционная функция для рядов активности биоиндикатора и потенциала естественного электрического поля на ст.Гарм

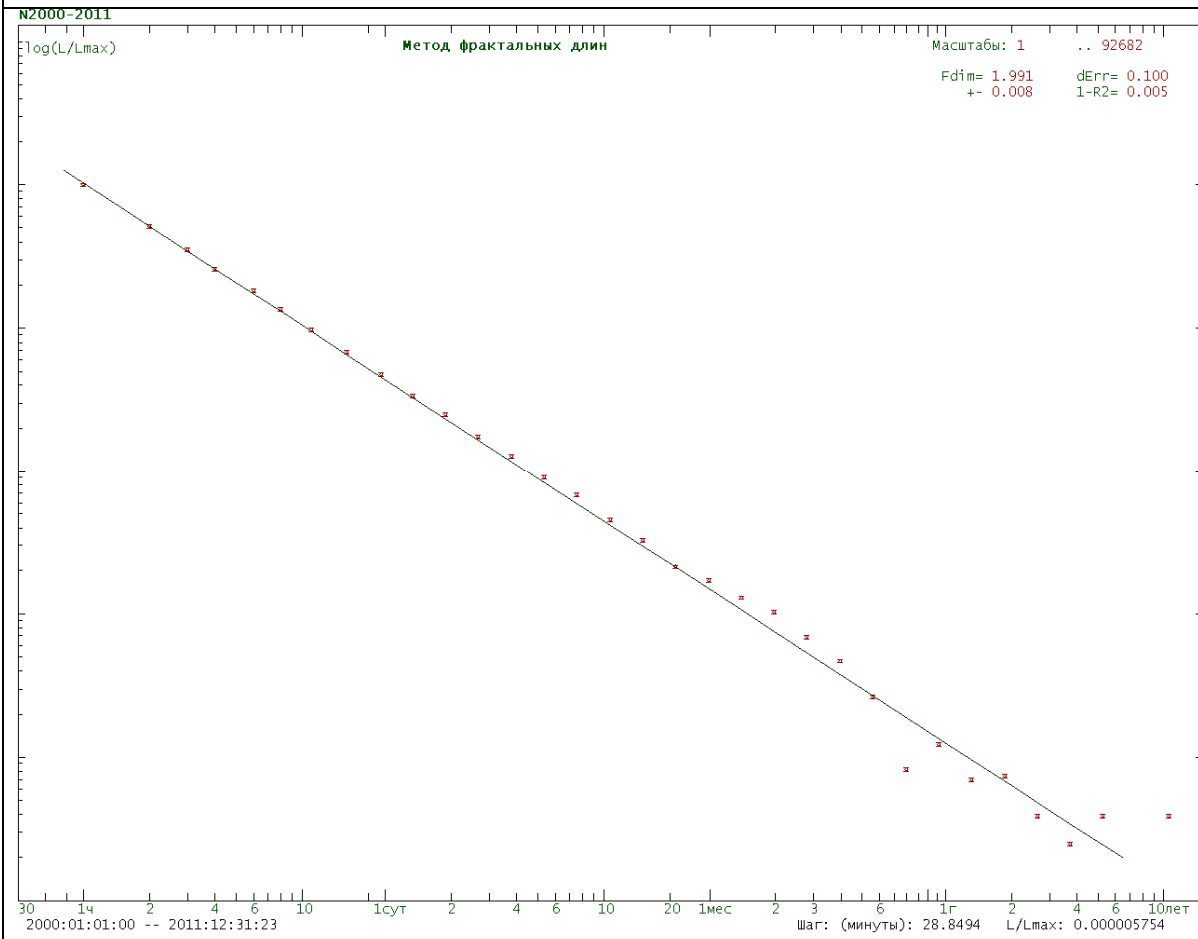


Рис.9. Зависимость накопленного интеграла приращений от масштаба (метод фрактальных длин) для ряда чисел землетрясений Греции за 2000-2011гг.

