

УДК 519.65

## МЕТОД ИДЕНТИФИКАЦИИ СТРУКТУРНЫХ КОМПОНЕНТОВ СЛОЖНОГО ПРИРОДНОГО СИГНАЛА НА ОСНОВЕ ВЕЙВЛЕТ-ПАКЕТОВ

*Мандрикова О.В., д.т.н., в.н.с., Института космических исследований и распространения радиоволн ДВО РАН, e-mail: oksanam1@mail.kamchatka.ru*

*Горева Т.С., аспирантка Института космических исследований и распространения радиоволн ДВО РАН.*

**Ключевые слова:** вейвлет-преобразование, природный сигнал, численный алгоритм, идентификация.

### Постановка задачи

Настоящая работа посвящена разработке методов и алгоритмов обработки и анализа сложных природных сигналов. Наряду с другими методами исследований обработка и анализ регистрируемых природных сигналов имеет важное значение в физике, геофизике и геоэкологии [1-5]. На основе модельных построений и соответствующих алгоритмических решений этот подход позволяет получить количественную оценку процессам, которые формируются в той или иной природной среде, выявить сложные внутренние связи и механизмы взаимодействия. Первая задача, которая стоит перед исследователем, найти способ уменьшить размерность системы и выявить компоненты, наиболее полно описывающие исследуемый природный процесс. Не случайно на протяжении последних 30 лет в центре внимания ученых, связанных с обработкой статистических данных, находится проблема выделения полезных сигналов на фоне помех. Сложности решения задач анализа природных данных связаны со сложной внутренней структурой регистрируемых сигналов [3-7]. Геофизические сигналы содержат переходные процессы, аномальные эффекты, которые связаны с активностью Солнца, возникают накануне сейсмических явлений и др. процессами различной физической природы. В первую очередь, в рамках данной работы, рассматривается класс задач, связанных с выделением и классификацией локальных особенностей и аномальных эффектов в природных сигналах. Эти особенности содержат полезную информацию об изучаемых процессах и при обработке не должны быть отфильтрованы как помеховый сигнал или шум из-за неизбежной в этом случае потери информации. Локальные особенности, как правило, имеют форму резких всплесков и пиков, могут иметь ступенеобразный вид.

При регистрации и анализе природных данных существенные погрешности в их оценку вносит ряд факторов. Исключая приборные погрешности, можно наметить неравнозначные группы помех:

- наличие в сигналах пропущенных значений, обусловленных сбоями в работе приборной и регистрирующей аппаратуры;
- космогенные помехи, связанные с возникновением аномальных потоков космических лучей, их резкими изменениями в верхних слоях атмосферы и за ее пределами;

*На основе конструкции вейвлет-преобразования предложен метод анализа природных сигналов со сложной структурой. В основе метода лежит численный алгоритм идентификации структурных компонентов сложного сигнала. На примере обработки данных критической частоты показана эффективность метода.*

- помехи, связанные с атмосферно-погодными явлениями;
- помехи сезонные, обусловленные действием переменных сезонных колебаний температуры;
- помехи, обусловленные вариабельностью природных процессов в связи с их нормальной самоорганизацией (типа пульсирующих гейзеров);
- случайные помехи неизвестной природы.

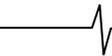
Методика обработки природных сигналов включает в себя как независимую обработку данных, так и совместный анализ всей совокупности компонентов. Процедуру анализа природного сигнала можно разбить на следующие взаимозависимые этапы:

1. Оценка погрешностей, вносимых рядом факторов различной природы. С целью уменьшения влияния этих факторов производится фильтрация полученных экспериментальных данных.

2. Оценка периодических изменений компонент сигнала с использованием классических методов спектрального анализа на основе Фурье-преобразования. После определения информативных частотных интервалов проводится фильтрация и анализ наблюдений в выбранных диапазонах частот. В настоящее время для анализа геофизических сигналов получают распространение такие современные методы обработки как вейвлет-преобразование и нейронные сети [2-4].

3. Анализ устойчивых изменений на основе проведения процедуры сглаживания и использования традиционных моделей временных рядов: в виде линейной комбинации трендов, сезонных компонент, циклов – колебаний относительно тренда и случайного шума [1].

4. Многокомпонентная обработка, заключающаяся в выделении важных признаков, подлежащих классификации и их обработки на основе современных методов, основанных, например, на дискриминантном анализе совокупности наблюдений и элементах корреляционного анализа. Аномальные эффекты в данном подходе могут быть выделены, например, на основе анализа изменяющейся автокорреляционной структуры многомерной геофизической системы, а также различных показателей взаимного спектра. В последнее время для проведения



совместного анализа геофизических данных также предложены методы вейвлет-преобразования, позволяющие декомпозировать исходный сигнал на различные частотные диапазоны, выделять короткопериодные вариации параметров, что невозможно было реализовать спектральными методами [5].

Несмотря на разнообразие существующих методов и подходов, задача эффективной обработки природных сигналов и, в частности, выделения аномалий и их классификации остается пока нерешенной. Это связано со следующими обстоятельствами:

- В отдельных случаях, с редкой сетью наблюдений: существующие базы данных содержат очень мало длинных рядов разных параметров, что облегчало бы выделение аномалий на фоне шумов.
- С отсутствием адекватных математических моделей: традиционная процедура сглаживания, которую принято использовать для облегчения процесса построения модели, влечет определенную потерю и искажение важной информации. Результаты исследований в своем большинстве относятся к низкочастотным составляющим колебаний геофизических компонентов и свидетельствуют о возможности выделения длительных (3-5 месяцев) аномальных периодов.
- С большим разнообразием и достаточно сложной формой аномалий: хотя методы спектрального анализа формально применимы для оценки периодических изменений компонент ряда, но в данном случае наличие в сигналах резких пикообразных вариаций и ступенеобразных особенностей ставят под сомнение эффективность этих методов.

Естественным и наиболее эффективным способом представления природных сигналов является построение нелинейных адаптивных аппроксимирующих схем на основе экстраполирующих фильтров. Инструментом, позволяющим реализовать такую процедуру для сигналов с подобными особенностями, является вейвлет-преобразование [6]. Вейвлет-преобразование имеет широкий класс базисов различной формы и временной протяженности, имеющих компактный носитель (в том числе пикообразные и ступенеобразные по форме базисы). Это позволяет исследовать сигнал с достаточной степенью локализации и выявить и классифицировать локальные особенности его структуры. В настоящее время ведутся интенсивные исследования по применению этого метода в различных прикладных задачах, связанных с анализом сигналов сложной внутренней структуры [3, 4, 6, 8]. Но качество работы конечных систем, основанных на данном аппарате, определяется технологией его применения: способ определения аппроксимирующего базиса, метод подавления шума, выбор конструкции преобразования, выбор уровня разложения и др. Данные аспекты позволяют оптимизировать время реализации алгоритмов и определяют точность работы систем, что имеет важное значение при решении задач анализа природных сигналов. Авторами данной работы разработан метод идентификации структурных компонентов сложного природного сигнала, включающий следующие операции:

- удаление шума;
- выбор «наилучшего» базиса;
- выделение компонент сигнала.

Используя основы современной теории обработки сигналов, авторами выстроена наглядная цепочка рассуждений от задач моделирования до эффективных вычислительных решений. Предложенный метод позволяет:

- отобразить в модели сигнала, как сглаженную составляющую, так и локальные особенности его структуры, что важно при решении задач выявления и классификации аномалий;
- адаптивно определить аппроксимирующие функции в зависимости от структуры исходных данных;
- путем минимизации погрешности аппроксимации построить *наилучшую аппроксимирующую схему* сигнала, что обеспечивает точность и достоверность получаемой информации.

### Представление сложного природного сигнала на основе вейвлет-пакетов

Представим сигнал в виде линейной комбинации разномасштабных компонент  $f_j$  с различной структурой:

$$f(t) = x_1 f_1(t) + x_2 f_2(t) + \dots + x_r f_r(t). \quad (1)$$

Когда коэффициенты  $x_1, x_2, \dots, x_r$  коррелируют между собой, вывод о том, какие аппроксимирующие функции использовать, сделать достаточно трудно, поэтому на функции  $f_j, j = \overline{1, r}$  наложим выполнение условия ортогональности относительно величин  $t_i$  с весовыми коэффициентами  $g_i = 1/\sigma_i^2$ :

$$\sum_i g_i f_j(t_i) f_k(t_i) = \delta_{jk}.$$

Поскольку функции  $f_j$  имеют разную структуру, подверженную изменению в случайные моменты времени, наиболее эффективным способом их идентификации является применение методов аппроксимации, основанных на разложении функции по базису. Учитывая локальный характер анализируемых особенностей и их разномасштабность, наиболее подходящим пространством для их представления является пространство, натянутое на базис смещенных функций или вейвлет – базис.

Структура разложения  $L^2(R)$ , порожденная ортогональным вейвлетом  $\Psi \in L^2(R)$ , имеет вид:

$$L^2(R) = \sum_{j \in Z}^{\oplus} W_j := \dots \oplus W_{-1} \oplus W_0 \oplus W_1 \oplus \dots, \quad (2)$$

где  $W_j := \text{clos}_{L^2(R)}(\Psi_{j,n}; n \in Z)$

Функция  $f$  при этом представляется в виде суммы компонент:

$$\forall f \in L^2(R); f(t) = \dots + v_{-1}(t) + v_0(t) + v_1(t) + \dots, v_j \in W_j, j \in Z. \quad (3)$$

Каждая компонента  $v_j$  из (3) имеет единственное представление в виде вейвлет-ряда:

$$v_j = \sum_{n \in Z} c_{j,n} \Psi_{j,n}(t),$$

где  $\Psi_j = \{\Psi_{j,n}\}_{n \in Z}$  – ортонормированный базис пространства  $W_j$ . Коэффициенты  $c_{j,n}$  определяются из соотношения

$$c_{j,n} = \langle f, \Psi_{j,n} \rangle.$$

Определим функции  $f_j$  как  $f_j = v_j / x_j$ , в силу соотношения (3) получаем представление сигнала в виде

$$f(t) = \dots + x_{-1}f_{-1}(t) + x_0f_0(t) + x_1f_1(t) + \dots$$

В силу ортогональности базиса  $\Psi$  (см. (2)):

$$\sum_i \frac{1}{\sigma_i^2} f_j(t_i) f_l(t_i) = 0, \text{ если } j \neq l.$$

Случайный сигнал  $y$  представляет зависимость

$$y(t) = f(t) + e(t),$$

где  $f$  – истинное значение измеряемой величины,  $e$  – ошибки измерений.

Математическое ожидание

$$E\{e_\Psi(t_n)e_\Psi(t_k)\} = \sum_l \sum_m \Psi_n(t_l)\Psi_k(t_m)E\{e(t_l)e(t_m)\} = \sigma^2 \delta(n-k).$$

Тогда компоненты случайного сигнала  $y$  в пространстве вейвлет-образов имеют вид

$$y_j(t) = f_j(t) + e_j(t).$$

Для дискретного сигнала не нарушая общности примем  $j=0$ . В качестве базовой конструкции для построения отображения будем использовать конструкцию вейвлет-пакетов ВП, имеющую быстрые алгоритмы преобразования и позволяющую идентифицировать различные типы частотно-временных структур [6, 8]. Получим представление сигнала в виде:

$$y_0(t) = \sum_{j_i} (g_{j_i}(t) + e_{j_i}(t)) + f_{-m}(t), \quad (4)$$

где  $e_{j_i}, g_{j_i} \in W_{j_i}$  – детализирующие компоненты сигнала,  $W_{j_i}$  – пространства вейвлет-пакета,  $f_{-m} \in V_{-m}$  – аппроксимирующая компонента сигнала,  $V_{-m} = \dots \oplus W_{-m-2} \oplus W_{-m-1}, V_{-m} = \text{clos}_{L^2(R)}(\phi(2^{-m}t-n)), \phi$  – скейлинг-функция.

Каждая компонента (4) единственным образом определяется последовательностями коэффициентов

$$\bar{d}^{j_i} = \{d_n^{j_i}\}_{n \in Z}, \bar{e}^{j_i} = \{e_n^{j_i}\}_{n \in Z} \text{ и } \bar{c}^{-m} = \{c_n^{-m}\}_{n \in Z}:$$

$$d_n^{j_i} = \langle f, \Psi_{j_i, n} \rangle, e_n^{j_i} = \langle e, \Psi_{j_i, n} \rangle \text{ и } c_n^{-m} = \langle f, \phi_{-m, n} \rangle.$$

**Подавление шумовых составляющих  $e_{j_i}$  и выделение локальных особенностей сигнала**

Имея представление сигнала в виде (4), подавление шума может быть реализовано на основе применения пороговой функции для каждой детализирующей компоненты

$$P_T(x) = \begin{cases} x, & \text{если } |x| \geq T \\ 0, & \text{если } |x| < T \end{cases} \quad T - \text{порог.} \quad (5)$$

Примем порог  $T = \sigma^2$ , где  $\sigma^2$  – дисперсия шума. Следуя работе [6], дисперсию шума  $\sigma^2$  оценим следующим образом

$$\sigma^2 \approx \text{Med}(\langle y, \Psi_{j,k} \rangle)_{0 \leq k < \frac{n}{2}},$$

где  $\text{Med}$  – медиана,  $j$  – наименьший масштаб,  $n$  – длина сигнала. В работах Donoho D. [7] по минимаксным оценкам сигнала в смеси с шумом показано, что данный способ подавления шума позволяет получить почти оптимальные минимаксные оценки.

Выполнив процедуру (5), получим аппроксимирующую схему сигнала в виде:

$$y_0(t) = \sum_{j_i \in I_M} g_{j_i}(t) + f_{-m}(t),$$

где  $I_M$  – множество индексов, определяемое свойствами функции  $f$ ,  $g_{j_i}$  – детализирующие составляющие сигнала,  $f_{-m}$  – аппроксимирующая компонента.

Погрешность аппроксимации есть:

$$\varepsilon[M] = \|f - f_{I_M}\|^2 = \sum_{j_i \notin I_M} |\langle f, \Psi_{j_i} \rangle|^2.$$

Из теоремы Жаффара [6] следует, что когда масштаб убывает, амплитуды вейвлет-коэффициентов имеют быстрое убывание до нуля в областях, где сигнал гладкий. Тогда операция выделения локальных особенностей функции  $f$  в виде пиков, перегибов и т.п. может быть основана на анализе детализирующих компонентов модели (4) путем определения наибольших значений  $|d_n^{j_i}|$  на малых масштабах.

**Выбор «наилучшего» базиса**

«Наилучшим» базисом будем считать базис, погрешности аппроксимации в котором наименьшая. Выбор «наилучшего» базиса выполним путем реализации следующего алгоритма:

1. Построение полного дерева разложения:  $W_j^0: W_j^0 = \bigoplus_{i=1}^l W_{j_i}^{p_i}, \{\Psi_{j_i}^{p_i}(2^{j_i}t-n)\}_{n \in Z}$  есть базис пространства  $W_{j_i}^{p_i}$ ;
2. Определение ветвей дерева путем минимизации погрешности аппроксимации: наилучший базис  $O_j^p$  пространства  $W_j^p$  есть базис

$$O_j^p = \begin{cases} \{\Psi_j^p(2^j t-n)\}_{n \in Z}, & \text{если } \sum_{n \in I_M^p} |\langle y, \Psi_{j,n}^p \rangle|^2 \geq \sum_{n \in I_M^{2p}} |\langle y, \Psi_{j+1,n}^{2p} \rangle|^2 + \sum_{n \in I_M^{2p+1}} |\langle y, \Psi_{j+1,n}^{2p+1} \rangle|^2 \\ \{\Psi_{j+1}^{2p}\}_{n \in Z} \cup \{\Psi_{j+1}^{2p+1}\}_{n \in Z}, & \text{если } \sum_{n \in I_M^p} |\langle y, \Psi_{j,n}^p \rangle|^2 < \sum_{n \in I_M^{2p}} |\langle y, \Psi_{j+1,n}^{2p} \rangle|^2 + \sum_{n \in I_M^{2p+1}} |\langle y, \Psi_{j+1,n}^{2p+1} \rangle|^2 \end{cases}$$

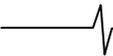
где множества индексов  $I_M^l, l = P, 2P, 2P+1$  определяются следующим образом:

$$\text{индекс } n \in I_M^l, \text{ если } |\langle y, \Psi_{j,n}^l \rangle| \geq T.$$

В процессе выполнения данного алгоритма будут подавлены шумовые составляющие сигнала и идентифицированы его структурные компоненты. Выполняя операции (1), (2) для различных видов базисных функций и определяя для каждой из таких функций погрешность в «наилучшем» базисе, мы определим «наилучшую» вейвлет-функцию для данного сигнала. Аппроксимирующую схему сигнала в этом базисе назовем *наилучшей аппроксимирующей схемой* для данного сигнала.

**Результаты экспериментов**

Для оценки эффективности предложенного метода проведены эксперименты по обработке и анализу данных регистрации критической частоты  $f_0F2$  Камчатского геодинамического полигона, полученные в ходе плановой работы в лаборатории Института космических исследований и распространения радиоволн ДВО РАН. Вариации критической частоты  $f_0F2$  имеют сложную структуру. На фоне регулярных изменений, обусловлен-



ных суточным и сезонным ходом, могут возникать резкие одиночные «пики» длительностью от нескольких десятков минут до нескольких часов [4]. Эти аномалии возникают на фоне мощных ионосферных возмущений.

При обработке использовались файлы с часовыми данными  $f_oF2$ , содержащими результаты измерений за период с 1968 по 2006гг. В качестве языка для написания программ, выполняющих поставленные задачи, был выбран язык программирования системы Matlab.

Нулевой уровень детальности  $j=0$  соответствует интервалу взятия отсчетов, равному 1 час. Тогда уровень детальности 4 включает колебания от 8 до 24-ех часов. Учитывая особенности динамики ионосферных процессов [4], этим уровнем разложения можно ограничиться. Учитывая сезонный характер поведения параметра  $f_oF2$ , процедура его разложения выполнялась отдельно для каждого времени года. Для исключения влияния краевого эффекта на результаты обработки, сигналы были дополнены значениями, захватывающими конец прошлого сезона и начало следующего.

На этапе построения аппроксимирующей схемы из сигнала была выделена шумовая компонента  $e(t)$  и произведена оценка погрешностей для различных базисных функций. Использовался класс ортогональных вейвлетов Добеши. Анализ погрешностей показал, что вейвлет Добеши 2-го порядка является «наилучшей» аппроксимирующей функцией для исследуемого сигнала. На рис. 1, в качестве примера, показаны сигналы  $f_oF2$  с шумом и без шума за период 15.11.02-15.03.03 гг.. В таблице 1 показаны результаты экспериментов.

Операция идентификации структурных компонентов сигнала, выполненная для нескольких лет, позволила

выделить составляющие критической частоты  $f_oF2$ , формирующие этот процесс в различные периоды времени. На рис. 2, в качестве примера, показаны составляющие сигнала  $f_oF2$  зимнего периода времени для нескольких лет. Анализ результатов обработки позволил выделить в структуре сигнала  $f_oF2$  характерные для него составляющие. Одна из таких характерных составляющих выделена на рис. 2 пунктирной линией.

Детальный анализ полученных компонент сигнала показал наличие в нем локальных особенностей, возникающих в определенные моменты времени и проявляющих себя в виде значительного увеличения абсолютных значений вейвлет-коэффициентов. Сопоставление выделенных особенностей с данными индекса геомагнитной активности  $K$  и с данными сейсмического каталога п-ова Камчатка показало, что в большинстве случаев значительное увеличение абсолютных значений вейвлет-коэффициентов наблюдается в периоды повышенной Солнечной активности, а также наблюдается накануне сильных землетрясений на Камчатке. На рис. 3, в качестве примера, показан результат анализа одной из характерных компонент сигнала  $f_oF2$  (данная компонента выделена на рис.2 пунктирной линией). Стрелками со сплошной линией отмечены моменты возникновения сейсмических событий (рассматривались сейсмические события энергетического класса с  $k \geq 12.5$ ), стрелками с пунктирной линией отмечены моменты возрастания индекса геомагнитной активности  $K$ . Анализ данной компоненты сигнала  $f_oF2$  показал, что выявленные аномальные особенности наблюдаются в сигнале в интервале времени от нескольких часов до 7 суток.

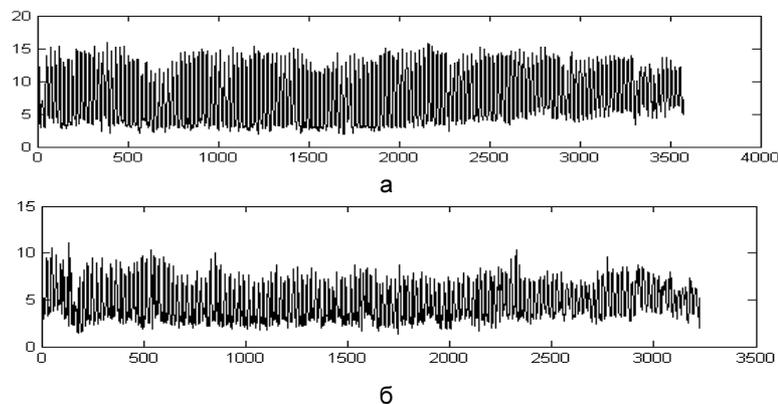


Рис. 1. Сигналы  $f_oF2$  с шумом (а) и без шума (б) за период 15.11.02-15.03.03 гг.

Таблица 1.

Результаты экспериментов: определение порога  $T$ ; оценка погрешности  $\varepsilon$ .  
В качестве базисных функций использовались ортогональные вейвлеты Добеши 2-го порядка.

Временной интервал	Длина сигнала $N$	Значение порога $T$	Погрешность $\varepsilon$
15.11.01-15.03.02	3227	0,340207	143,9
15.11.02-15.03.03	3120	0,337452	122,66
15.11.03-15.03.04	3116	0,233995	87,52
15.11.04-15.03.05	3225	0,324784	121,52
15.11.05-15.03.06	3316	0,281272	118,9

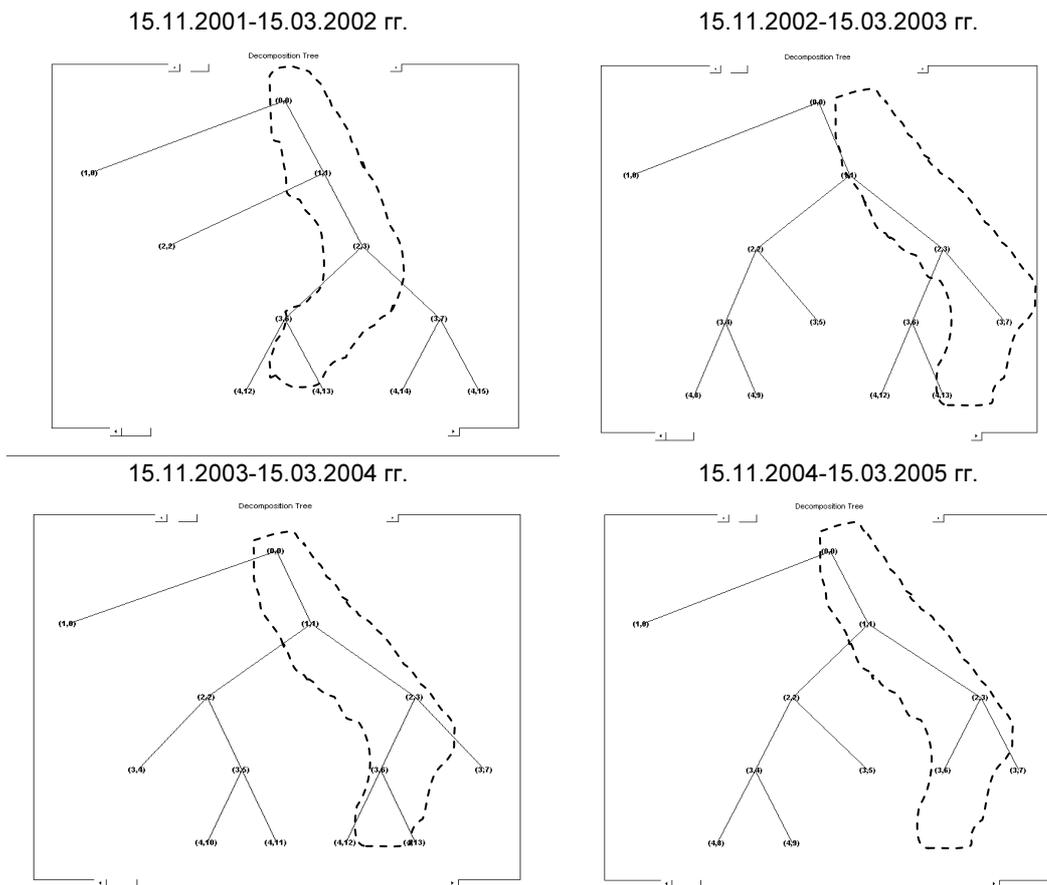


Рис. 2 Структурные компоненты сигнала  $f_oF2$  зимнего периода времени.

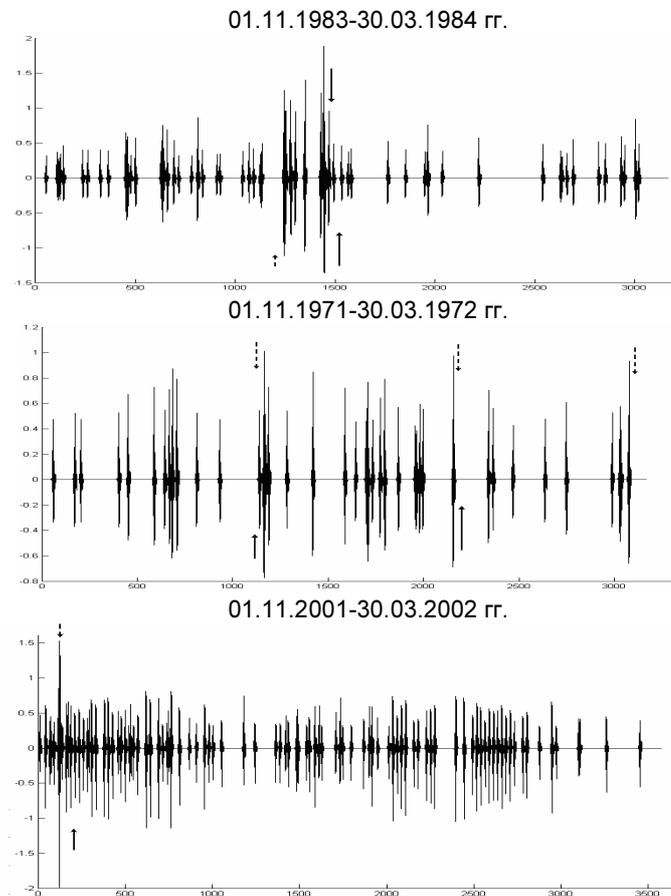


Рис.3. Компонента сигнала критической частоты  $f_oF2$ ; стрелками со сплошной линией отмечены моменты возникновения сейсмических событий; стрелками с пунктирной линией отмечены моменты возрастания индекса геомагнитной активности  $K$ .

---

## Заключение

На основе положений современной теории обработки сигналов предложен метод анализа природных сигналов со сложной внутренней структурой. В основе метода лежит численный алгоритм идентификации структурных компонентов сложного сигнала, включающий операции подавления шума и выбор «наилучшего» базиса.

На примере обработки данных критической частоты  $f_oF2$  показана эффективность метода: выявлены скрытые закономерности в структуре сигнала  $f_oF2$ ; обнаружена связь между процессами, протекающими в ионосфере и сильными сейсмическими явлениями, показан характер связи.

Данные сейсмического каталога любезно предоставлены Камчатским филиалом геофизической службы РАН (г. Петропавловск-Камчатский).

### Литература

1. Никифоров И.В. Последовательное обнаружение изменения свойств временных рядов. - М.: Наука, 1983
2. Филиппов Ю.А. Фирстов П.П. Применение нейронной сети для анализа динамики объемной активности подпочвенного радона за 1998-2002 г. на Паратунской геотермальной системе (Камчатка) // Нейроинформатика и ее приложения. Красноярск. 2002. С.33-37.
3. Мандрикова О.В. Моделирование геохимических сигналов на основе вейвлет-преобразования. – Владивосток: Дальнаука. 2007. – 123с.
4. Богданов В.В., Геппенер В.В., Мандрикова О.В. Моделирование нестационарных временных рядов геофизических параметров со сложной структурой. – С.-Петербург: ЛЭТИ. 2006. – 107с.
5. Любушин (мл.) А.А. Вейвлет-агрегированный сигнал и синхронные всплески в задачах геофизического мониторинга и прогноза землетрясений // Физика Земли, 2000, №3, с.20-30.
6. Stephane Mallat, A Wavelet tour of signal processing. Пер. с английского, – М.: Мир, 2005.
7. Donoho D., Johnstone I. Minimax estimation via wavelet shrinkage. Annals of Statistics, 1998.
8. Ingrid Daubechies, Ten Lectures on Wavelets. Пер. с английского, – Ижевск: НИЦ «Регулярная и хаотическая динамика».– 2001.

## THE METHOD OF INDIFICATION THE STRUCTURE COMPONENTS OF COMPLICATED NATURE SIGNAL ON THE BASIS OF WAVELET-PACKAGES

*Mandrikova O.V., Goreva T.S.*

The method of analyze of natural signals with complicated structure is offered on the base of structure wavelet - transformations. The algorithm of identification the structure components of complicated signal is the base of this method. The efficiency of this method is shown by the example of data processing of critical frequency  $f_oF2$ .