

**МЕТОД МОДЕЛИРОВАНИЯ И ПРОГНОЗИРОВАНИЯ ДАННЫХ КРИТИЧЕСКОЙ
ЧАСТОТЫ f_0F_2 НА ОСНОВЕ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ**
**METHOD OF MODELLING AND PROGNOSIS IN THE DATA OF CRITICAL FREQUENCY
 f_0F_2 ON THE BASIS OF WAVELET-TRANSFORMATION**

О.В. Мандрикова, В.В. Богданов

Институт космических исследований и распространения радиоволн ДВО РАН, ikir@ikir.ru

As the critical frequency f_0F_2 has complicated structure, application of traditional methods of data processing and analysis is inefficient. For the study we used new classes of models, methods and numerical algorithms which are based on a mathematical structure so called multicomponent model of time series. Identification of multicomponent model is based on the development of wavelet-transform methods and parametrical models of time series. Within the framework of this theory a method of approximation of critical frequency f_0F_2 data and fillings of omissions was developed.

Современные методы аппроксимации сигналов, основанные на разложении функции по вейвлет-базисам, дают возможность на основе небольшого числа параметров получить аппроксимации требуемой точности [2], и, ввиду большого разнообразия базисных функций, могут быть использованы для широкого спектра сигналов с различной структурой. Они играют важную роль в теории обработки сигналов и позволяют решать такие важные задачи анализа природных временных рядов как выделение тонких структур, локальных особенностей и их классификация [1,3]. На их основе авторами данной работы разработан метод многокомпонентного моделирования и прогнозирования природных временных рядов, включающий операцию подавления шума и выделения информативных компонент. Предложенный метод позволяет адаптивно определить аппроксимирующие функции и обеспечивает полное сохранение информации об исследуемом процессе. Он основан на разложении временного ряда на компоненты в пространстве вейвлет-пакетов, оценка параметров компонент ряда в процессе идентификации его модели производится на основе методов авторегрессии-проинтегрированного скользящего среднего (АР-модели) [4]. Эти модели получили широкое распространение во многих областях науки и техники. Получаемые на их основе оценки имеют простую форму и легко интерпретируются физически, что представляет наилучшую основу для построения физической модели. Важным преимуществом этого класса моделей также является наличие эффективных методов автоматического обнаружения изменения параметров модели, на основе которых может быть реализована процедура выделения аномальных эффектов в данных.

Описание метода

Временной ряд сложной структуры может быть представлен в виде суммы разномасштабных ортогональных компонент f_j :

$$f(t) = x_1 f_1(t) + x_2 f_2(t) + \dots + x_n f_n(t), \quad (1)$$

$$\sum_i g_i f_j(t_i) f_k(t_i) = \delta_{jk}, \quad g_i = 1/\sigma_i^2. \quad (2)$$

Процедуру идентификации компонент f_j построим на применении методов аппроксимации, основанных на разложении функции по базису в пространстве Лебега $L^2(R)$. Учитывая локальный характер анализируемых особенностей, их разномасштабность и разнообразие по форме, наиболее подходящим базисом для представления данных является вейвлет-базис $f_j(t) = \sum_k c_{j,k} \psi_{j,k}(t)$, где, $\psi_{j,k}$ – базисные вейвлет-функции пространства $L^2(R)$. Вейвлет-коэффициенты $c_{j,k} = \langle f, \psi_{j,k} \rangle$ рассматриваются здесь как результат отображения f в пространство с разрешением j . Будем использовать нелинейные аппроксимирующие вейвлет-схемы, тогда приближение f выполняется M векторами, зависящими от структуры сигнала $f_M = \sum_{m \in I_M} \langle f, \psi_m \rangle \psi_m$, где

I_M – множество индексов, определяемое структурой сигнала f .

Рассмотрим в качестве пространства исходных дискретных последовательностей данных пространство с разрешением $j=0$

$$V_0 = \text{clos}_{L^2(R)}(\phi(2^0 t - k)), k \in Z.$$

В этом случае нулевой уровень детальности f_0 соответствует интервалу снятия показаний.

Применяя к данным конструкцию разложения в вейвлет-пакеты, получим для пространства V_0 схему разложения $W_0 = \bigoplus_{i=1}^{2^m-1} W_{-j_i} \oplus V_{-n}$, где $W_0 = V_0$; W_{-j_i} – пространства вейвлет-пакета.

Данные с шумом будут иметь представление

$$y_0(t) = \sum_{-j_i} (g[2^{-j_i}] + e[2^{-j_i}]) + f[2^{-n}] = f_1(t) + f_2(t) + e(t),$$

где компоненты $f_1(t) = \sum_{(-j_i, k) \in I_M} d_n^{-j_i} \psi_{-j_i, k}(t)$, $f_2(t) = \sum_k c_k^{-n} \phi_{-n, k}(t)$, единственным образом

определяются последовательностями коэффициентов $\bar{d}^{-j_i} = \{d_k^{-j_i}\}_{k \in \mathbb{Z}}$, $\bar{c}^{-n} = \{c_k^{-n}\}_{k \in \mathbb{Z}}$:

$d_k^{-j_i} = \langle f, \psi_{-j_i, k} \rangle$, $c_k^{-n} = \langle f, \phi_{-n, k} \rangle$, $e(t)$, определяемая коэффициентами $e_k^{-j_i} = \langle e, \psi_{-j_i, k} \rangle$, является

шумовой компонентой. Выделенная составляющая $f_1(t) = \sum g_{-j_i}$ является детализирующей компонентой и определяет локальные особенности структуры данных, сглаженная составляющая $f_2(t) = f_{-n}$ содержит устойчивые характеристики структуры. Компоненты g_{-j_i} и f_{-n} имеют более простую структуру, чем исходный временной ряд f , и, в силу условия (2), не коррелируют между собой. Эти свойства позволяют выполнить идентификацию моделей для каждой из полученных компонент и объединить полученные представления на основе соотношения (1) в общую модель временного ряда.

Численный метод идентификации компонент g_{-j_i} и f_{-n} подробно описан в работе [3], он включает следующие три операции: подавление шумовой компоненты $e(t)$; выбор аппроксимирующего вейвлет-базиса ψ_m ; выделение компонент ряда. Выделенные компоненты g_{-j_i} и

f_{-n} аппроксимируются АР-моделью. Для идентификации АР-модели, может быть использован итеративный подход [4]: выбирается полезный класс моделей; получают предварительные оценки параметров модели; диагностические проверки позволяют выявить возможные погрешности полученной модели, если погрешности удовлетворяют требованиям, модель готова к использованию.

При подгонке многокомпонентной модели идентификация АР-моделей выполняется для различных компонент разложения и на основе результатов диагностических проверок должна быть идентифицирована окончательная модель, вид которой:

$$f(t) = \sum_{j, k} s_k^j(t) \psi_{j, k}(t), \quad s_k^j(t) = \sum_{l=1}^{p_j} \gamma_l^j \omega_{k-l}^j(t), \quad (3)$$

где γ_l^j – коэффициенты авторегрессии компоненты модели масштаба j , $\omega_k^j(t) = \nabla^d r_k^j(t)$, $r_k^j = \{d_k^{-j_i}, c_k^{-n}\}_{(-j_i, -n) \in I_M}$, p_j – порядок АР-модели компоненты масштаба j , ∇^d – оператор взятия разности назад порядка d .

Прогнозирование значения s_{k+q}^j , $q \geq 1$ определяет прогноз s_k^j в момент $t = k$ с упреждением q .

Значение s_{k+q}^j на основе модели (3) можно выразить следующим образом: $s_{k+q}^j(t) = \sum_{l=1}^{p_j} \gamma_l^j \omega_{k+q-l}^j(t)$.

Моделирование и прогнозирование данных критической частоты f_oF2

Регистрация параметров ионосферы методом вертикального радиозондирования в ИКИР ДВО РАН (с. Паратунка, Камчатский край) проводится с 1968г. На фоне регулярных изменений, обусловленных суточным и сезонным ходом, в данных регистрации ионосферных параметров могут возникать аномальные эффекты длительностью от несколько десятков минут до нескольких часов. Эти аномалии возникают на фоне мощных ионосферных возмущений, которые обусловлены активностью Солнца. Проводимые в последнее время исследования процессов в средней и верхней атмосфере позволяют обнаружить связи в развитии некоторых явлений в литосфере и магнитосфере, которые формируются в ионосфере на фоне солнечной активности [1]. Задача выделения и классификации таких особенностей является крайне важной. Вариации критической частоты f_oF2 имеют сложную многомасштабную структуру, содержат пропуски в данных, возникающие по различным причинам, включая технические. Это существенно затрудняет процесс построения моделей и, в свою очередь, делает неэффективным применение к ним традиционных методов

моделирования временных рядов. С целью заполнения пропусков и выполнения анализа данных, на основе описанного выше метода, были построены многокомпонентные модели временных рядов f_oF2 и выполнено прогнозирование их значений.

В процессе экспериментов были обработаны файлы с часовыми данными f_oF2 , содержащими результаты измерений за период с 1999 по 2005гг. Учитывая сезонный характер ионосферного процесса, разложение данных производилось отдельно для каждого времени года.

В процессе выделения вейвлет-компонент на каждом следующем масштабном уровне мы получаем сглаженную компоненту f_{-n} все более простой структуры, что упрощает как процесс идентификации AP-модели, так и понижает ее порядок, при этом часть информации переходит в детализирующие компоненты. При обработке временные ряды f_oF2 были разложены до третьего масштабного уровня, выполнено подавление шумовых компонент и выделены структуры g_{-j_i} и f_{-n} , формирующие его в разные периоды времени. Для полученных компонент были идентифицированы AP-модели. Характеристики моделей компонент f_{-n} для нескольких лет представлены в таблице 1. Результаты прогнозирования значений компонент f_{-n} с шагом упреждения, равным шести отсчетам для компоненты третьего уровня вейвлет-разложения сигнала (что соответствует 70-ти отсчетам восстановленного сигнала, один отсчёт соответствует 1 часу), представлены на рис.1.

Таблица 1. Характеристики моделей

| Год | Длина компоненты разложения | Время упреждения | Дисперсия ошибки прогноза |
|------|-----------------------------|------------------|---------------------------|
| 2000 | 85 | 6 | 10,96 |
| 2001 | 52 | 6 | 6,57 |
| 2002 | 266 | 6 | 15,14 |
| 2004 | 222 | 6 | 2,89 |
| 1999 | 66 | 6 | 47,74 |
| 2005 | 111 | 6 | 27,06 |



Рис. 1. Прогноз данных критической частоты f_oF2 за сейсмически спокойный зимний период: а – 2004 г.; б – 2000 г.

Далее были рассмотрены годы, имеющие повышенную сейсмическую активность в зимний период (1999 и 2005гг). Для них также были идентифицированы AP-модели, но они при диагностике показали плохие результаты (1999 и 2005гг., таблица 1). Результаты прогнозирования компонент f_{-n} для этих лет показаны на рис.2. Анализ этих периодов времени показал, что данные содержат аномальные эффекты, которые предшествуют сильным сейсмическим событиям и возникают в данных за период от 5-ти до 15-ти дней до момента землетрясения (рассматривались события энергетического класса с $k \geq 12,5$).

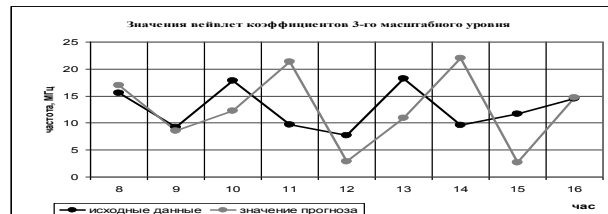


Рис.2. Прогноз данных критической частоты за сейсмически активный зимний период: а – 1999 г.; б – 2005 г.

Также было выполнено сравнение эффективности предложенного метода с традиционными подходами. С этой целью была произведена процедура сглаживания временного ряда f_oF2 на основе метода скользящего среднего с последующей попыткой идентификации AP-модели. Процесс идентификации AP-модели оказался безуспешным: наблюдались большие значения и существенная автокорреляция остаточных ошибок моделей, что свидетельствовало об их

неадекватности. На рис.3 показан исходный временной ряд f_0F2 и результаты его моделирования традиционным и предлагаемым методами. Анализ графиков показывает, что применение предложенного метода позволило значительно уменьшить остаточные ошибки модели (в 8 раз), что обеспечило при диагностике ее адекватность.

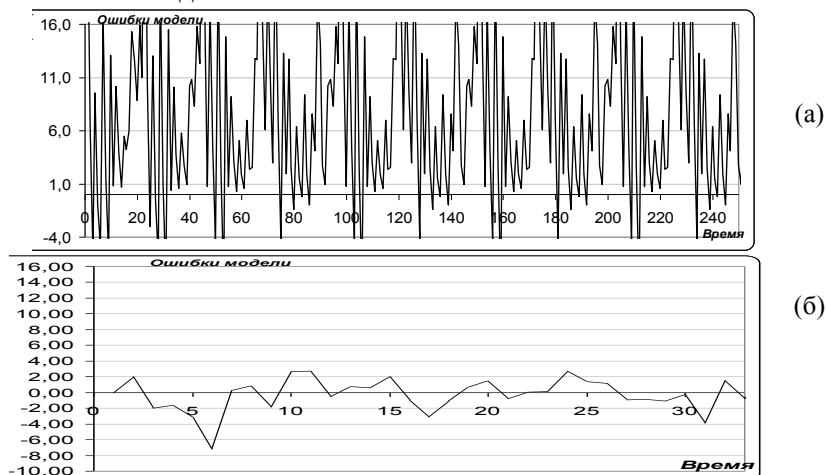


Рис. 3. Результаты обработки частоты f_0F2 за 2002 год; остаточные ошибки АР-модели, полученной в результате традиционного подхода (а); остаточные ошибки многокомпонентной модели f_0F2 (б).

Результат прогноза данных критической частоты f_0F2 и заполнения пропусков в них на основе полученной многокомпонентной модели показан на рис.4.а. Из рисунка видно, что прогноз выполнен практически на трое суток вперед (70 часов). На рис.4.б, для сравнения, показан результат заполнения пропусков в этих же данных способом, используемым в настоящее время в ИКИР ДВО РАН, на основе медианных значений.

Выводы

Предложен метод многокомпонентного моделирования и прогнозирования сложного природного сигнала, в основе которого лежит нелинейная аппроксимирующая вейвлет-схема. На примере обработки данных критической частоты f_0F2 показана эффективность метода: выполнен прогноз значений данных с шагом упреждения 70 отсчетов, что невозможно было реализовать существующими методами; выявлен характер связи между процессами, протекающими в ионосфере и сильными сейсмическими явлениями.

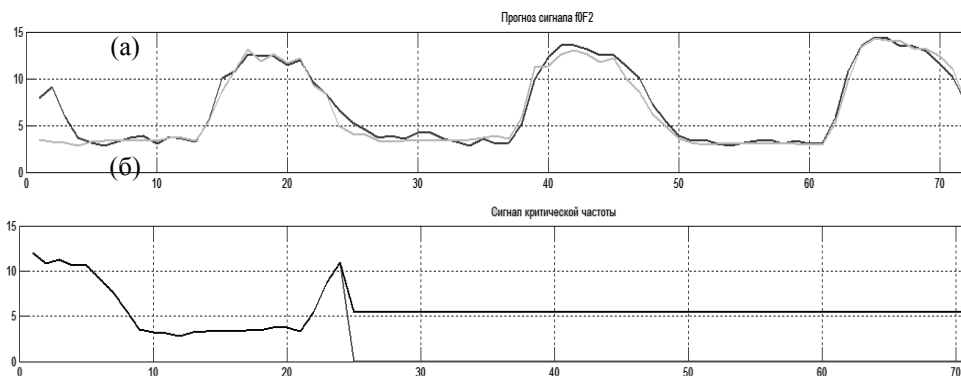


Рис. 4. Заполнение пропусков в данных критической частоты f_0F2 (шаг упреждения 70 отсчетов, 1 отсчет – 1 час) за 2002 год: (а) – предлагаемый метод; (б) – используемый метод.

Авторы работы выражают благодарность аспирантке ИКИР ДВО РАН Горевой Т.С. за помощь в проведении экспериментальных исследований.

Литература

- [1] Богданов В.В., Геппенер В.В., Мандрикова О.В. Моделирование нестационарных временных рядов геофизических параметров со сложной структурой. – С.-Петербург: ЛЭТИ. 2006. – 107с.
- [2] Stephane Mallat, A Wavelet tour of signal processing. Пер. с английского, – М.: Мир, 2005.
- [3] Мандрикова О.В., Горева Т.С. Метод идентификации структурных компонентов сложного природного сигнала на основе вейвлет-пакетов // Цифровая обработка сигналов. 2010 №1, с.45-50.
- [4] Марпл.-мл. С.Л. Цифровой спектральный анализ и его приложения: Пер. с англ. -М.: Мир, 1990.