

ПРИМЕНЕНИЕ РОБАСТНОГО КРИТЕРИЯ КАЧЕСТВА ДЛЯ ФИЛЬТРАЦИИ РЯДОВ КРИТИЧЕСКИХ ЧАСТОТ СЛОЯ F2

И.А. Белашова, В.В. Бочкарев

Казанский (Приволжский) федеральный университет, Казань, Россия
inkin91-91@mail.ru

APPLICATION OF A ROBUST CRITERION FOR FILTERING THE SERIES OF CRITICAL FREQUENCIES OF THE F2 LAYER

I.A. Belashova, V.V. Bochkarev

Kazan (Volga region) Federal University, Kazan, Russia

Аннотация. В данной работе рассматривается метод нелинейной фильтрации рядов критических частот слоя f_0F2 . В качестве альтернативы критерию квадратичной ошибки, применяемому в стандартных методах фильтрации, к которым относится и вейвлет-трешолдинг, рассмотрен робастный критерий Рамачандрана—Ранганатана, позволяющий фильтровать данные, распределение которых неизвестно. Статистика данного критерия представляет собой сумму квадрата длин серий — последовательности точек, для которых ошибка имеет одинаковый знак, умноженных на число серий данной длины. Минимизируя данную сумму, мы уменьшаем число длинных серий, предполагая, что длинные серии свидетельствуют о неслучайном характере флуктуаций. Для минимизации данной статистики применялись такие методы оптимизации, как генетические алгоритмы. Показано, что применение критерия серий позволяет добиться лучших результатов при фильтрации рядов критических частот слоя f_0F2 по сравнению с вейвлет-трешолдингом.

Abstract. This paper describes a method of non-linear filtering of critical frequencies series of the F2 layer. The robust Ramachandran—Ranganathan criterion is considered as an alternative to the criterion of minimum least-square errors applied in standard methods of filtering (such as wavelet-thresholding). It allows us to filter data which distribution is unknown. Statistics of this criterion is a sum of squared length of series multiplied by numbers of long series, assuming that the long series indicate a nonrandom behavior of fluctuations. Such optimization methods as genetic algorithms were applied to minimize this statistics. It was shown that using the runs criterion in filtering of critical frequencies series of the F2 layer has an advantage over wavelet-thresholding.

1 Вейвлет-фильтрация рядов

В настоящее время все большее распространение получил такой метод анализа и обработки сигналов, как вейвлет-анализ, применяемый в различных областях науки, техники, медицины. Популярность вейвлет-анализа обусловлена частотно-временной локализацией, возможностью исследования фрактальных функций, высокой эффективностью.

Одним из важнейших приложений дискретного вейвлет-анализа является фильтрация данных, основной принцип которой заключен в фильтрации коэффициентов, содержащих наименьший объем информации о фильтруемом сигнале, что в большинстве стандартных методов приводит к оставлению самых больших коэффициентов.

Использование критерия квадратичной ошибки как функции качества дает очень хорошие результаты в случае рядов с гауссовским законом распределения вероятности. Однако часто приходится обрабатывать ряды, флуктуации которых описываются отличным от гауссовского законом, либо с неизвестным законом распределения. Для временных рядов с известным распределением можно использовать критерий максимального правдоподобия, который применим для рядов с любым распределением вероятностей. В случаях, когда распределение рассматриваемого ряда неизвестно, предлагается использовать робастные критерии оценки (критерий серий).

Задача состоит в том, чтобы найти представление анализируемых данных в виде вейвлет-ряда с небольшим числом слагаемых (по сравнению с количеством отсчетов) числом слагаемых. Как и для стандартного вейвлет-трешолдинга, необходимо выбрать, какие термины вейвлет-ряда будут использованы для аппроксимации, и оценить оптимальные значения соответствующих вейвлет-коэффициентов.

2 Критерии тренда и случайности

Такие критерии предназначены для проверки гипотез о случайности расположения выборочных данных. Иными словами, проверяется отсутствие взаимосвязи между значениями реализаций случайной величины и их номерами в выборочной последовательности [Кобзарь 2006]. Данные критерии находят свое применение в промышленности при статистическом контроле и урегулировании технологических процессов, медицине и других областях деятельности человека.

Для решения задачи фильтрации временных рядов предлагается использовать сериальный критерий (критерий Рамачандрана—Ранганатана) для выбора ненулевых коэффициентов и их значений. В данном случае под серией понимается последовательность точек, для которых ошибка аппроксимации имеет одинаковый знак. Основная идея применения такого критерия заключается в предположении о

корреляции структуры и числа серий со степенью случайности поведения временного ряда: малое число серий свидетельствует об объединении элементов одной выборки в группы, то есть говорит о неслучайном характере расположения значений числовой последовательности [Кобзарь, 2006]. Критерий Рамачандрана–Ранганатана учитывает не только количество серий, но также их длину. Статистика данного критерия имеет вид:

$$R = \sum_j j^2 n_j, \quad (1)$$

где j — длина серии, n — объем выборки, n_j — количество серий длины j [Кобзарь, 2006].

Применение данного критерия подразумевает выполнение следующих шагов:

1. Исходный ряд сравнивается с оценкой (отфильтрованным рядом) — находится разность между ними;

2. Для применения данного критерия к непрерывным функциям данная разность преобразуется к следующему виду $\tanh((x - \bar{x}) / \alpha)$, где α — масштабирующий коэффициент;

3. Определяются серии, которые больше и меньше нуля;

4. Вычисляется статистика по формуле (1).

Для минимизации статистики применяются такие методы оптимизации как генетические алгоритмы.

3 Применение генетических алгоритмов при фильтрации временных рядов

Использование генетических алгоритмов в качестве методов оптимизации обусловлено их применимостью как к непрерывным, так и к дискретным функциям, а также возможностью нахождения глобального минимума ошибки путем подбора таких параметров, как: численность популяции, число родителей в начальной популяции, число поколений, вероятность кроссинговера и мутации, число «элитных» потомков в новой популяции и так далее.

Метод фильтрация при помощи генетических алгоритмов можно разделить на два ключевых блока:

- Определение нового набора ненулевых коэффициентов;

- Варьирование значений ненулевых коэффициентов с целью минимизации статистики критерия серий Рамачандрана—Ранганатана R

Для формирования нового набора ненулевых коэффициентов составляется хромосома из коэффициентов вейвлет-разложения, подвергшихся вейвлет-фильтрации. Она представляет строку из нулей и единиц, где «1» соответствует оставленному при фильтрации коэффициенту, а «0» — обнуленному. При помощи генетических алгоритмов происходит варьирование положения «1» в хромосоме, и находится такой набор, при котором статистика R минимальна.

Важнейшим условием оптимизации является постоянство числа ненулевых коэффициентов. Его необходимо учитывать при разработке функции кроссинговера и мутации, а также при создании начальной популяции.

В данной работе применялся однородный кроссинговер. Функция мутации разрабатывалась с учетом сохранения числа ненулевых коэффициентов: если в хромосоме «0» изменился на «1», то в каком-то другом локусе происходит обратная операция.

5 Сравнение результатов фильтрации модельных рядов и эмпирических данных критических частот f_0F2

Для проверки преимущества сериального критерия был смоделирован временной ряд, закон изменения которого представляет собой функцию из примера «Bumps» математического пакета MATLAB. На данный сигнал накладывался шум с гауссовским, равномерным и пуассоновским распределениями вероятностей. Результаты анализа представлены на рис. 1 и в таблице.

Сравнительные величины СКО отфильтрованного сигнала от исходного для двух критериев качества

Метод фильтрации	распределение Гаусса	равномерное распределение	распределение Пуассона
вейвлет-трешолдинг	0.96	0.79	1.04
Критерий Рамачандрана–Ранганатана	0.55	0.43	0.6

Из данного рис. 1 можно сделать следующий вывод: несмотря на некоторые несоответствия отфильтрованного сигнала истинному (интервал отсчетов 50–150), фильтрация на основе критерия серий точнее совпадает с истинным сигналом. Подчеркнем, что обе сравниваемые аппроксимации представляют собой вейвлет-ряды с одинаковым количеством ненулевых коэффициентов.

По результатам моделирования для 50 случайных рядов (для каждого из распределений) было рассчитано среднеквадратичное отклонение отфильтрованного сигнала от исходного сигнала «Bumps». В таблице представлено среднее значение СКО для трех распределений.

Таким образом, результаты статистического моделирования демонстрирует преимущество робастного критерия перед критерием минимума среднеквадратичной ошибки.

В качестве эмпирического примера рассматривались временные ряды критических частот, полученные на ионозонде «Циклон» Казанского федерального университета. Был выбран временной интервал 11.08.2013–20.08.2013 г. Основной задачей при обработке данных временных рядов являлось выделение волновых процессов, происходящих в ионосфере на фоне суточной составляющей. На рис. 2 представлен пример обработки рядов критических частот f_0F2 .

Как можно видеть, в результате применения вейвлет-трешолдинга неудовлетворительно описывается даже общий вид тренда данного ряда с выделением одного высокочастотного всплеска небольшой величины. Данная процедура не позволила выделить никаких волновых процессов, наблюдающихся

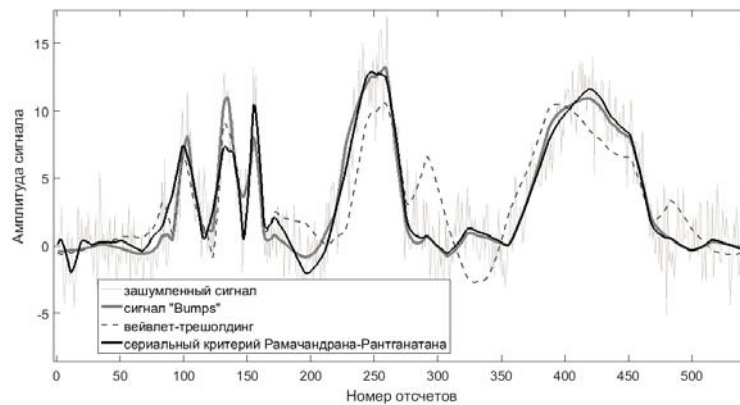


Рис. 1. фильтрация модельного ряда «Bumps» с нормально распределенной шумовой составляющей

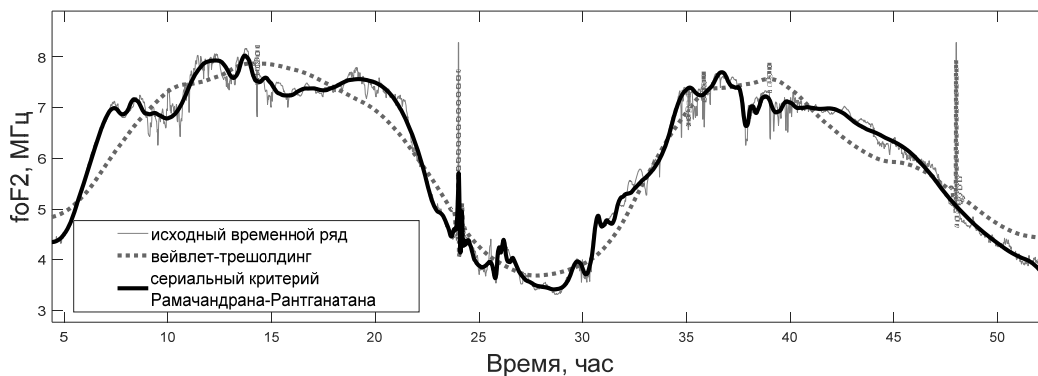


Рис. 2. Фильтрация рядов критически частот слоя f_oF_2 (за ноль взято 00 ч : 00 мин : 00 с 15.08.2013 г.)

в исходном ряде. Применение сериального критерия Рамачандрана—Ранганатана позволило не только довольно точно описать тренд данного ряда, но также выделить волновые процессы в интервале 5–36 ч (рис. 2).

Данный пример показывает, что применение альтернативного критерию минимума среднеквадратичной ошибки сериальный критерия Рамачандрана—Ранганатана для фильтрации временных рядов позволяет не только избавиться от шумовой составляющей, распределение вероятностей которой может быть неизвестно, но также выделить проявления волновых процессов, происходящих в ионосфере.

Рассмотренные примеры позволяют убедиться в эффективности робастных критериев при фильтрации временных рядов с неизвестным распределением, а также в возможности решения практической задачи выявления волновых процессов, происходящих в ионосфере. Предложенный метод фильтрации дает большой выигрыш по сравнению с широко известным методом вейвлет-трешолдинга.

Список литературы

- Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. М.: Физматлит, 2006. 816 с.
- Малла С. Вейвлеты в обработке сигналов // Пер. с англ. М.: Мир, 2005. 671 с.
- Рутковская Д., Пилиньский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы // Пер. с польск. И.Д. Рудинского. М.: Горячая линия — телеком, 2006. 452 с.