

# ОБРАБОТКА СИГНАЛОВ

УДК 519.651; 523.24:521.1

*Ю.В. ЧЕРКАС, Ю.И. ВОЛОЩУК, д-р техн. наук*

## ОЦЕНИВАНИЕ ТРЕНДОВ ПРИМЕНИТЕЛЬНО К РАСПРЕДЕЛЕНИЯМ ОКОЛОЗЕМНЫХ АСТЕРОИДОВ ПО БОЛЬШОЙ ПОЛУОСИ

### Введение

Тренд – основная тенденция изменения значений ряда. В статистическом анализе в широком смысле под трендом подразумевают длительное изменение среднего значения случайного процесса.

Задача нахождения тренда может быть поставлена в двух формулировках:

1) Ряд интерпретируется как сумма систематической составляющей  $u(t)$ , или тренда, и случайной составляющей  $e(t)$ , или ошибки:

$$z(t) = u(t) + e(t). \quad (1)$$

2) Ряд рассматривается как сумма систематической составляющей  $u(t)$ , полезного сигнала  $x(t)$  и случайной составляющей  $e(t)$ :

$$z(t) = u(t) + x(t) + e(t). \quad (2)$$

В первом случае нас интересует сам тренд для построения или описания модели порождающего его процесса, либо для прогнозирования следующих значений ряда. Задачи такого типа часто возникают в экономике. Во втором случае нас интересует полезный сигнал, и здесь необходимо избавиться от тренда для последующего анализа остаточного сигнала.

Существуют следующие методы оценки тренда:

а) Тренд является известной функцией времени (или других наблюдаемых величин). В этом случае задача нахождения тренда заключается в оценке неизвестных параметров этой функции.

б) Параметрические – рассматривают ряд как гладкую функцию (т.е. функцию, имеющую непрерывную производную на всем множестве определения) от своего аргумента  $u(t) = f(t)$ ,  $t = t_1 \dots t_N$ . При этом сначала выявляют один или несколько допустимых типов функции  $f(t)$ , затем производят оценку параметров этих функций, после чего на основе проверки критериев адекватности выбирают окончательную модель тренда.

в) Непараметрические – методы сглаживания исходного ряда – скользящие средние, экспоненциальное сглаживание и др. Эти методы полезны в случае, когда не удастся подобрать подходящую функцию для оценки тренда.

Применительно к распределениям астероидов NEA (Near Earth Asteroids – околоземные астероиды, т.е. астероиды, орбиты которых сближаются с орбитой Земли) по большой полуоси, в работах [1 – 7] рассматриваются теории и гипотезы, согласно которым это распределение, помимо систематического компонента, должно содержать периодическую составляющую и, следовательно, ему соответствует модель описываемая выражением (2). В работах [8, 9] проведен анализ распределений астероидов NEA из базы орбит [10], одним из этапов которого было устранение тренда в ряде распределения орбит для последующего спектрального анализа остаточного ряда на предмет наличия периодической составляющей. Задача выявления и устранения тренда не является тривиальной, поскольку кроме выбора самого метода выделения тренда, необходимо задать и обосновать его параметры, что с другой стороны усложняется наличием мощной шумовой составляющей. Неудачное устранение тренда приводит как к искажениям выделяемого сигнала, так и вовсе делает невозможным оценивания его параметров в остаточном ряде.

Данная статья может рассматриваться как краткий обзор популярных методов устранения тренда в общем, а также их сравнение для случая конкретного типа тренда в частности.

### Аппроксимация полиномом

В основе полиномиального метода выделения тренда лежит аппроксимационная теорема Вейерштрасса, согласно которой любая непрерывная функция  $f(t)$ , определенная на закрытом интервале, может быть сколь угодно близко  $\varepsilon > 0$  аппроксимирована многочленом  $p(t)$ :

$$|f(t) - p(t)| < \varepsilon, \quad (3)$$

где  $p(t) = \alpha_0 + \alpha_1 t + \dots + \alpha_n t^n$  – полином степени  $n$ .

При использовании данного метода для выделения тренда предпочтение отдается полиномам меньшей степени, что в свою очередь накладывает ограничения на интенсивность изменения характера ряда (чередования возрастаний и убываний значений тренда). Использование полиномов высокой степени может привести к таким нежелательным явлениям как выделению вместе с трендом части полезного сигнала, а также к феномену Рунге, т.е. нежелательной осцилляции в начале и в конце интервала аппроксимации.

Выбор коэффициентов при степенях полинома происходит по методу наименьших квадратов (МНК) путем минимизации целевой функции:

$$Q = \sum_{i=1}^N (z(t_i) - p(t_i))^2 \rightarrow \min. \quad (4)$$

Отметим, что данная целевая функция отвечает только за выбор коэффициентов при степенях полинома и не имеет отношения к выбору степени полинома.

Полиномиальный тренд является в первую очередь средством описания, содержащим в сжатой форме общие характеристики ряда. В большинстве случаев коэффициентам полинома нельзя придать никакого реального смысла.

От выбора степени полинома зависит степень приближенности полученного тренда к имеющимся значениям ряда. В терминах статистики недостатком выбора слишком низкой степени является наличие смещения при оценивании тренда, а недостатком выбора слишком высокой степени – большая вариабельность при оценивании тренда [11].

### Сглаживание сплайном

Сплайн – функция, область определения которой разбита на конечное число отрезков, на каждом из которых сплайн совпадает с некоторой функцией одного вида. Наиболее распространенными являются сплайны, состоящие из алгебраических полиномов. Как правило, это полиномы нечетных степеней, степени выше пяти используются редко. Максимальная степень из использованных полиномов называется степенью сплайна. Использование сплайнов на основе многочленов для выделения трендов является логическим продолжением полиномиального метода для случая, когда тренд не удается удовлетворительно аппроксимировать полиномом малой степени. Кроме алгебраических многочленов за основу сплайнов берут также экспоненты, тригонометрические полиномы и другие функции.

Другим управляемым параметром сплайна, кроме степени, является число узлов, соединяющих фрагменты, на которые разбивается его область определения. Использование сплайнов с равной шириной фрагментов позволяет упростить расчеты. Ключевым свойством сплайнов является условие стыковки фрагментов. Как правило, считают, что фрагменты стыкуются гладко, то есть обеспечивается непрерывность значений и первых производных в узлах. Понятие дефекта сплайна связано с числом непрерывных производных, имеющих функцией лежащей в основе сплайна и числом непрерывных производных в узлах. Визуально разрыв второй и высших производных не заметен и, как правило, не фигурирует в условиях стыковки фрагментов.

Увеличивая количество узлов сплайна, можно сколь угодно точно подгонять сплайн под имеющийся ряд, вплоть до того, что аппроксимационный сплайн превратится в интерполяционный. Выделяя тренд путем сглаживания исходного ряда  $z(t)$  сплайном  $\mu(t)$ , руководствуются выбором тех параметров, использование которых минимизирует следующую целевую функцию:

$$Q = \sum_{i=1}^N (z(t_i) - \mu(t_i))^2 + \lambda \int_{x_0}^{x_N} \frac{d^2\mu(t)}{dt^2} dx \rightarrow \min. \quad (5)$$

где  $\lambda \geq 0$  – коэффициент гладкости. Первая составляющая целевой функции отвечает за степень приближенности сплайна к значениям исходного ряда, вторая – за степень его гладкости.

### Методы скользящего среднего

Сглаживающие процедуры базируются на предположении, что ряд являет собой сумму гладкого тренда и случайных, флуктуирующих составляющих, взвешенная сумма которых в пределах окна сглаживания становится малой величиной (подразумевают, что флуктуирующие составляющие независимы и имеют нулевое математическое ожидание) [11]. При наличии в ряде полезного сигнала необходимо брать ширину окна сглаживания больше интервала корреляции сигнала. Если же полезный сигнал является периодическим, то выбор ширины окна сглаживания кратным его периоду позволит предотвратить смещение тренда и, как результат, искажение остаточного ряда.

Скользящее среднее – общее название семейства функций, значения которых в каждой точке равны взвешенной сумме значений ряда в окрестности данной точки:

$$\hat{y}(t) = \sum_{s=-m}^m c_s y(t+s), \quad t = m+1, \dots, N-m, \quad (6)$$

где  $c_s$  – веса сглаживания,  $2m+1$  – ширина окна сглаживания. В ряде задач процедуру скользящего среднего используют также для предсказания последующих значений ряда. В таких случаях значение в каждой точке ряда рассчитывается не по значениям окрестности данной точки, а по значениям, предшествующим ей.

Веса  $c_s$  предполагаются нормированными:

$$\sum_{s=-m}^m c_s = 1. \quad (7)$$

Простейшим случаем скользящего среднего является арифметическое скользящее среднее, численно равное среднему арифметическому значений исходной функции за интервал сглаживания. В некоторых ситуациях веса сглаживания выбирают, беря за основу такие функции как экспонента, алгебраический полином и др.

Математически скользящее среднее является одним из видов свертки, что позволяет рассматривать его как аналог фильтра нижних частот используемого в обработке сигналов.

### Метод переменных разностей

Пусть  $\Theta$  – оператор, воздействие которого на последовательность  $f(t_i)$ , сдвигает все индексы на один шаг (элемент запаздывания в цифровой фильтрации):

$$\Theta f(t_i) = f(t_{i+1}). \quad (8)$$

Тогда разностный оператор  $\Delta$  определяется как:

$$\begin{aligned}
\Delta f(t_i) &= (\Theta - 1)f(t_i) = f(t_{i+1}) - f(t_i), \\
\Delta^2 f(t_i) &= (\Theta - 1)^2 f(t_i) = f(t_{i+2}) - 2f(t_{i+1}) + f(t_i), \\
&\dots \\
\Delta^r f(t_i) &= (\Theta - 1)^r f(t_i) = \sum_{j=0}^r (-1)^{r-j} \binom{r}{j} f(t_{i+j}).
\end{aligned} \tag{9}$$

Важным свойством операции вычисления разностей является то, что применение данного оператора к полиному понижает его степень на единицу. Из этого следует, что если  $p(t)$  – полином степени  $k$ , то [11]

$$\Delta^k p(t) = 0. \tag{10}$$

Метод переменных разностей базируется на уже упомянутом свойстве гладких функций быть аппроксимированными в последовательных интервалах полиномами небольших степеней. Следовательно, применение разностного оператора к ряду, тренд которого является гладкой функцией, редуцирует последний к значениям, стремящимся к нулю. Используя метод переменных разностей для избавления от систематической составляющей ряда, следует также учитывать влияние разностного оператора на полезный сигнал.

### Цифровая фильтрация

Исходя из определения тренда как основной тенденции изменений значений ряда, можно утверждать, что в частотном представлении он будет состоять из постоянной составляющей и низкочастотных компонент, а для его выделения можно воспользоваться хорошо разработанными методами цифровой фильтрации. Пропускание ряда через фильтр нижних частот является процедурой сглаживания или выделения тренда. Если нас интересуют остаточные флуктуации ряда избавленного от тренда, то следует использовать фильтр верхних частот. Выбор частоты среза фильтра следует производить на основе априорной информации о тренде или, в случае наличия, полезном сигнале.

Когда цифровая фильтрация производится для устранения тренда, выбор типа и конкретных параметров фильтра базируется на свойствах и цели выделения полезного сигнала. Если речь идет, например, о выявлении периодических составляющих в ряде, то можно воспользоваться фильтрами с бесконечной импульсной характеристикой (БИХ), которые в отличие от фильтров с конечной импульсной характеристикой (КИХ) обеспечивают лучшую амплитудно-частотную характеристику, хотя и не обладают линейной фазовой характеристикой.

### Схема анализа

Сравнительный анализ рассмотренных методов выделения тренда применительно к распределениям астероидов NEA по большой полуоси будем проводить по следующей схеме:

1) формирование ряда  $z(a)$ , состоящего из эталонного тренда  $u(a)$ , наперед заданного гармонического сигнала  $x(a)$  и случайной составляющей  $e(a)$  ( $a$  – большая полуось орбиты,  $[a] = 1 \text{ а.е.}$ , а.е. – астрономическая единица);

2) проведение процедуры, направленной на отыскание гармонической составляющей в ряде  $z(a)$ , с помощью устранения тренда выбранным методом и последующим вычислением периодограммы [12, 13] остаточного ряда;

3) шаги 1 и 2 повторяются для разных значений мощности шума; для обеспечения статистической устойчивости результатов каждая конфигурация параметров прогоняется для 1000 разных реализаций шума с одними и теми же характеристиками.

Эталонный тренд  $u(a)$  получим путем построения распределения астероидов NEA [10] по большой полуоси  $a$ , с последующим сглаживанием полученной гистограммы численности полиномом 9-й степени (рис.1). Данные действия также являются процедурой выделения тренда, но в этом случае нам неважно, насколько точно сглаженный полином соответствует «истинному» тренду, необходимо лишь, чтобы их общие характеры совпадали. Гистограмма численности рис.1 была построена для  $N=200$  интервалов разбиения и, следовательно, ряд тренда  $u(a)$ , полезного сигнала  $x(a)$  и все реализации шума  $e(a)$  также будут содержать по  $N=200$  отсчетов.

Полезным сигналом была выбрана синусоида с амплитудой  $A=5$  и частотой  $f=10 \text{ а.е.}^{-1}$  умноженная на соответствующий коэффициент  $K_i$  (рис.2):

$$x_i(a) = K_i A \sin(2\pi f a), \quad (11)$$

$$K_i = \frac{u_i(a)}{u_{\max}(a)} \quad (12)$$

где  $u_i(a)$  –  $i$ -е значение ряда тренда,  $u_{\max}(a)$  – максимальное значение тренда. Сума эталонного тренда  $u(a)$  и полезного сигнала  $x(a)$  приведена на рис.3.

В качестве случайной составляющей  $e(a)$  был выбран аддитивный гауссовский белый шум. Его мощность изменялась в границах, обеспечивающих изменение отношения мощности полезного сигнала  $x(a)$  к мощности шума  $e(a)$  в пределах от  $\text{SNR}=0 \text{ дБ}$  до  $\text{SNR}=-20 \text{ дБ}$ , а сами значения, как и значения полезного сигнала, были взвешены коэффициентом  $K_i$  (12). Мощность шума изменялась с шагом, обеспечивающим прирост отношения сигнал-шум равным  $\Delta \text{SNR}=0.5 \text{ дБ}$ . Для каждого значения мощности шума были использованы 1000 его разных реализаций. После устранения тренда выбранным методом и вычисления периодограммы, полезный сигнал  $x(a)$  считается успешно выявленным, если максимум периодограммы находится на частоте  $f = 10 \pm 0.2 \text{ а.е.}^{-1}$ . Подсчитав число выявления полезного сигнала для всей серии реализаций шума, находим вероятность выявления сигнала выбранным методом выделения тренда для данного  $\text{SNR}$ .

Прежде чем приступить к сравнению разных методов выделения тренда, найдем оптимальные параметры для каждого из них. С этой целью будем строить графики вероятности выявления полезного сигнала от величины  $\text{SNR}$  при разных значениях управляемых параметров.

Первый рассматриваемый метод выделения тренда – полиномиальный. Единственным управляемым параметром является степень полинома. На рис.4 приведены графики зависимости вероятностей обнаружения полезного сигнала от  $\text{SNR}$ , цифры возле линий обознача-

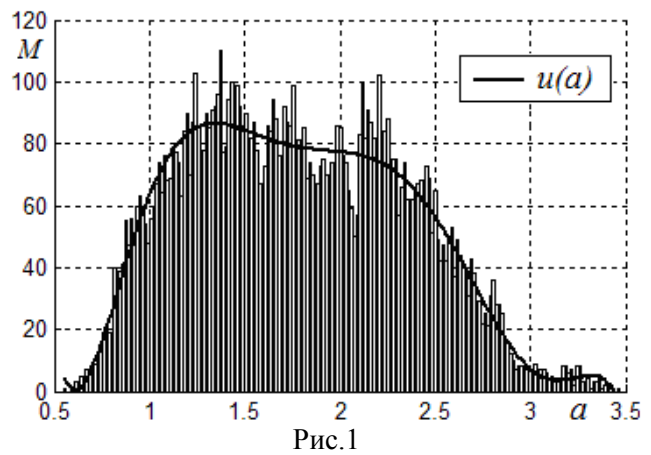


Рис.1

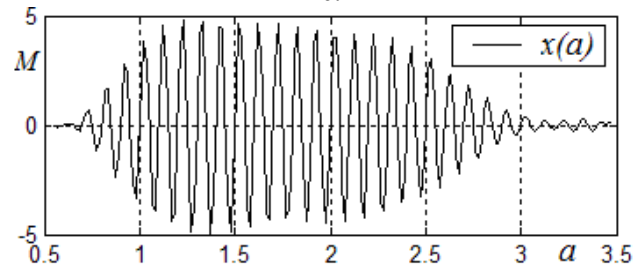


Рис.2

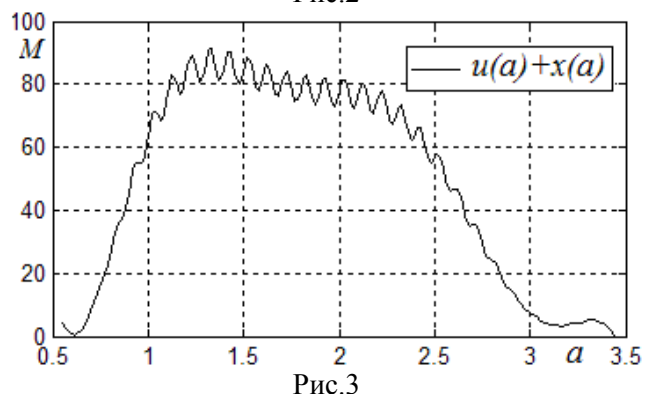


Рис.3

ют максимальную степень используемого полинома. Как видим, все кривые кроме  $k=5$  и  $k=6$  практически совпадают. Поэтому в качестве оптимального значения выберем степень полинома, использованную при формировании эталонного тренда –  $k=9$ .

Метод выделения тренда, основанный на сглаживании сплайном, имеет два управляемых параметра – максимальную степень и количество узлов, соединяющих фрагменты сплайна. В качестве степени сплайна нами было выбрано значение  $k=5$ . Число узлов изменялось от  $l=5$  до  $l=50$  с шагом  $\Delta l=5$ . На рис. 5 приведена зависимость вероятности выявления сигнала от SNR. Кривые для  $l=5$  до  $l=50$  имеют заметно худшие характеристики обнаружения сигнала во всем диапазоне изменения SNR. Остальные кривые практически совпадают в своих значениях, но зависимость для  $l=35$  имеет незначительно лучшие характеристики – это значение и было выбрано в качестве оптимального для числа узлов сплайна.

В качестве метода скользящего среднего был выбран самый простой из них – арифметическое скользящее среднее. Ширина окна изменялась в пределах от 3 до 21 отсчетов ряда. По рис.6 видим, что лучшей характеристикой выявления сигнала в заданном диапазоне изменения SNR обладает арифметическое скользящее среднее с шириной окна 9, которое и было выбрано в качестве оптимального.

Метод переменных разностей имеет один управляемый параметр  $k$ , который задает количество раз применения разностного оператора  $\Delta$  к ряду. Без приведения соответствующих графиков, отметим, что для выбранного эталонного тренда  $u(a)$  любые значения  $k > 1$  не приводили к удовлетворительным вероятностям выявления сигнала (сказывается низкая частота сигнала  $x(a)$   $f=10 \text{ а.е.}^{-1}$ , что влекло к устранению вместе с трендом и искомого сигнала).

Для цифровой фильтрации ряда был выбран эллиптический фильтр верхних частот с затуханием в полосе подавления  $R_s=80 \text{ дБ}$ , неравномерностью в полосе пропускания  $R_p=0.1 \text{ дБ}$  и разностью между границами полос пропускания и затухания  $\Delta f=0.005 \text{ а.е.}^{-1}$ . Порядок фильтра, а также его параметры перерасчитывались для каждого нового значения границы полосы затуха-

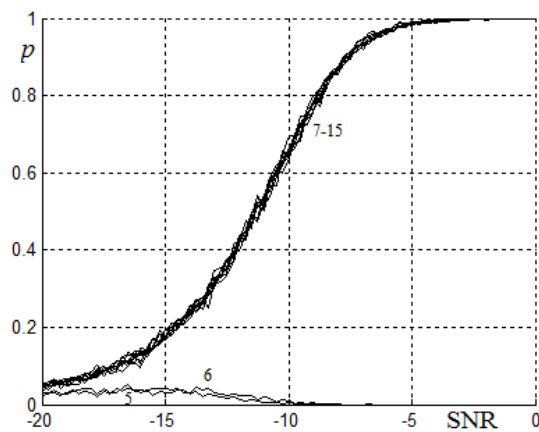


Рис.4

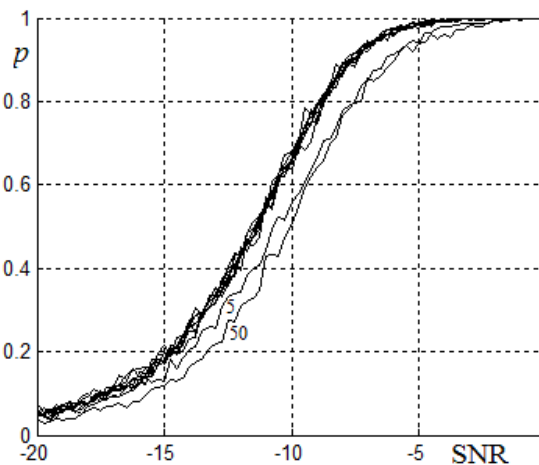


Рис.5

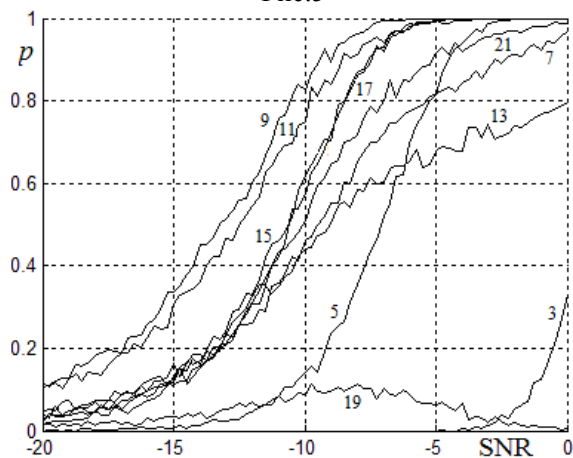


Рис.6

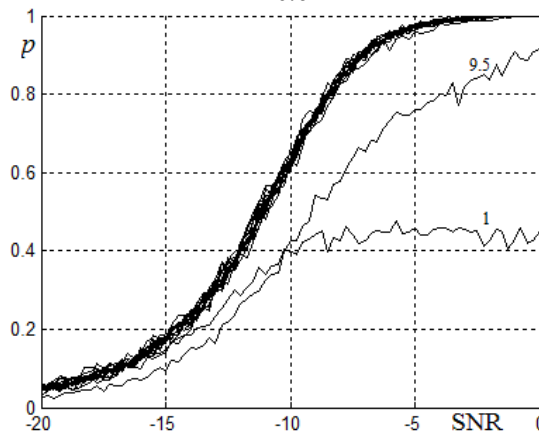


Рис.7

ния, которая изменялась в пределах  $f_s = 1 \div 9.5 \text{ а.е.}^{-1}$  с шагом  $f = 0.5 \text{ а.е.}^{-1}$ . Как видим из рис. 7, кривые для всех значений границ полосы затухания кроме  $f_s = 1 \text{ а.е.}^{-1}$  и  $f_s = 9.5 \text{ а.е.}^{-1}$  имеют практически одинаковые значения во всем диапазоне изменения SNR. По этой причине в качестве оптимальной частоты среза фильтра верхних частот было выбрано среднее значение –  $f_s = 5 \text{ а.е.}^{-1}$

К данному этапу провели процедуру выбора оптимальных параметров для всех пяти рассматриваемых методов выделения тренда и имеем возможность на их основании построить обобщенный график зависимости вероятности выявления полезного сигнала (11) от величины SNR (рис.8).

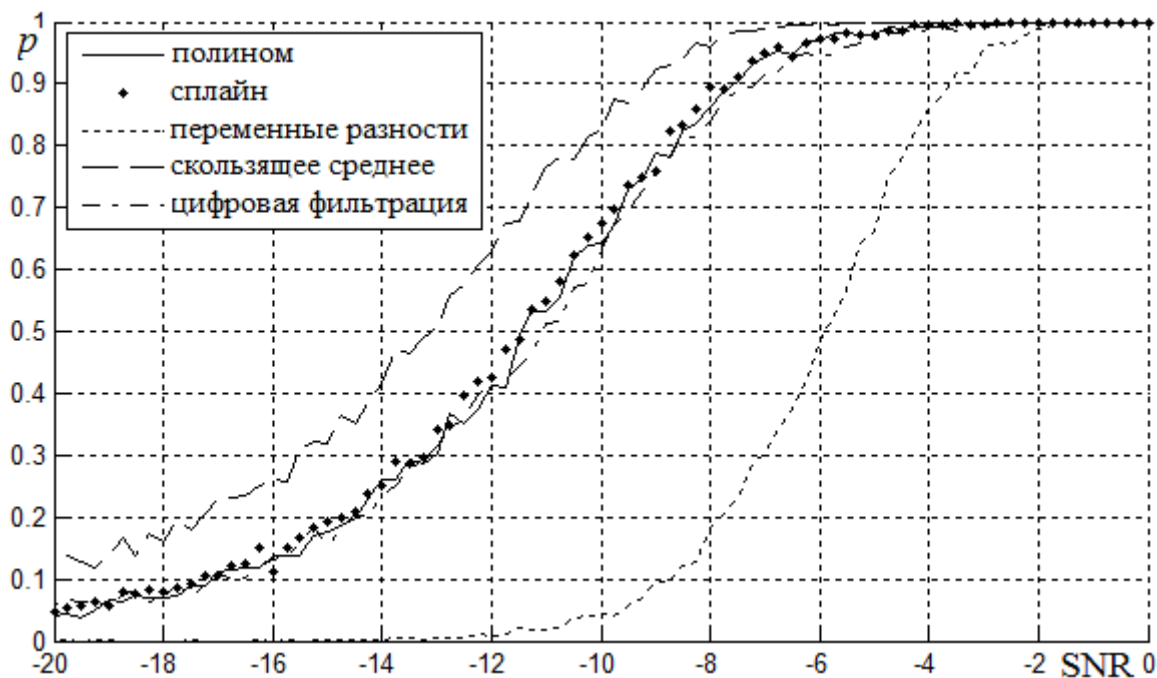


Рис.8

## Выводы

Проведенное моделирование и сравнительный анализ разных методов выделения трендов позволяют утверждать, что в рамках рассмотренного тренда  $u(a)$  (рис.1) и гармонического полезного сигнала  $x(a)$  наилучшими характеристиками (в терминах вероятности обнаружения сигнала в зависимости от SNR) обладает метод скользящего среднего с шириной окна сглаживания, найденной по приведенной методике. Следует отметить, что для рядов небольшой длины эта ситуация изменится, поскольку применения метода скользящего среднего приводит к укорочению ряда, а следовательно, и ухудшению спектрального разрешения. Наихудшими характеристиками обладает метод переменных разностей, единственным плюсом которого является простота реализации. Остальные три метода (сглаживания сплайном, аппроксимация полиномом, цифровая фильтрация) обладают практически одинаковыми характеристиками. В случае, если бы характер тренда (чередование областей возрастания и убывания) изменялся более интенсивно, то полиномиальный тренд был бы непригодным для его выделения (увеличения степени привело бы к появлению феномену Рунге).

Приведенные результаты справедливы для трендов, характер которых совпадает с  $u(a)$  (рис.1). Для большинства других трендов, встречающихся в радиотехнике, экономике, естественных системах, результаты будут сопоставимыми с полученными нами, а приведенная здесь методика оценки эффективности выделения тренда может быть использована для их более точного оценивания.

**Список литературы:** 1. Гулак Ю.К. Соизмеримости и макроквантовые явления в солнечной системе. I. Проблемы, принципы, модель. – К. : Институт теоретической физики АН УССР, 1986. – 27с. 2. Гулак Ю.К. Соизмеримости и макроквантовые явления в солнечной системе. II. Стабильные механические структуры. – К. : Институт теоретической физики АН УССР, 1986. – 28с. 3. Quiroga R.J. Angular momenta in the Solar System. – Earth, Moon and Planets. – 1984. – 30 N2. – P. 102-121. 4. Quiroga R.J. Angular momenta in the Solar System. II – Earth, Moon and Planets. – 1984. – 30 N2. – P. 102-121. 5. Чельницкий А.М. О квантовании Солнечной системы // *Астрономический циркуляр*. – 1983. – №1257. – С.5-7. 6. Hu Zhong-wei, Chen Zhi-xiong. Distance law and formation of satellite systems // *Astronomische Nachrichten*. – 1987. – №308. – P.359-362. 7. Волощук Ю.И. и др. Метеоры и метеорное вещество. – Киев : Наук. думка, 1989. – 296 с. 8. Черкас Ю.В. Применение спектрального анализа для поиска периодических составляющих в распределениях астероидов по большой полуоси / Ю.В.Черкас, Ю.И. Волощук // *Прикладная радиоэлектроника*. – 2012. – Т.11, №3. – С.335-341. 9. Черкас Ю.В. Анализ тонкой структуры распределений орбит малых тел Солнечной системы. Околосолнечные астероиды // Черкас Ю.В., Волощук Ю.И., Коломиец С.В. // *Материалы 22-й Междунар. Крымской конференции КрыМиКо*, 10-14 сентября 2012. – Т. 2. – С.1003-1004. 10. [http://neo.jpl.nasa.gov/cgi-bin/neo\\_elem](http://neo.jpl.nasa.gov/cgi-bin/neo_elem) 11. Андерсон Т. Статистический анализ временных рядов : пер. с англ. – М. : Мир, 1976. – 755с. 12. Марпл.-мл. С.Л. Цифровой спектральный анализ и его приложения : пер. с англ. – М. : Мир, 1990. – 584с. 13. Stoica P. Introduction to Spectral Analysis. – Prentice Hall, New Jersey, 1997. – 319p.

*Харьковский национальный  
университет радиоэлектроники*

*Поступила в редколлегию 05.07.2013*