

А.В. ПОНОМАРЕВА,
А.В. МОСКОВЦОВА
**Использование методов
предварительной обработки
данных при анализе временных
рядов**

УДК 004.9

Харьковский
Национальный
Университет
Радиоэлектроники,
г.Харьков,
Украина

В статье проведен анализ предварительной обработки данных. Были экспериментально исследованы различные методы сглаживания временных рядов. Изучено влияние возмущающих воздействий (шум) на корректность предварительной обработки.

The article analyzes the data pre-processing. We investigated various methods of smoothing time series. The effect of disturbances (noise) on the correct pre-treatment.

В современных инфокоммуникационных системах, временные ряды являются широко распространенным типом данных. Временной ряд содержит информацию об особенностях и закономерностях протекания процессов.

Временной ряд представляет собой набор чисел, который привязан к последовательным моментам времени. Уровни временного ряда – числа, полученные путем наблюдения за протеканием определенного процесса, составляющие непосредственно временной ряд.

При обработке временных рядов к исходным данным предъявляется ряд требований. Таких как однородность данных, сопоставимость и полнота данных, устойчивость, предположения о типе их распределения и т. д.

В задачах обработки любого сигнала выделяют этапы предварительной (первичной) и вторичной обработки сигналов. Это связано с тем, что в общем случае на входе системы наблюдается смесь полезного сигнала и некоторого шума. Для повышения вероятности

правильного приема сигнала должна быть произведена предварительная обработка принятого сигнала. Кроме того, предварительный анализ дает возможность определять и использовать данные закономерности для оценивания характеристик процесса в будущем (прогнозирования) при решении задач идентификации моделей процессов.

Однако перед началом процедуры прогнозирования процессов, необходим предварительный анализ данных. В процессе предварительного анализа выявляют соответствие имеющихся данных требованиям, предъявляемым к ним математическими методами.

Процедура предварительной обработки сигнала в основном включает в себя проверку наличия тренда во временном ряде, выявление аномальных наблюдений, сглаживание временных рядов. Главной задачей при этом является получение полных, однородных, корректных данных.

В работе проведено исследование различных алгоритмов сглаживания временных рядов на базе цифровой обработки сигналов, выбор наиболее подходящих алгоритмов для различных типов входной информации, анализ влияния внешних возмущений на корректность проводимой предварительной обработки.

Предложено исследование наиболее распространенных методов сглаживания временных рядов как:

- метод скользящего среднего с различным значением апертур;
- метод взвешенного скользящего среднего;
- метод экспоненциального сглаживания.

Метод скользящего среднего позволяет, как сглаживать мелкие беспорядочные колебания, так и сохранять мелкие волны. В зависимости от необходимых действий апертюра (интервал сглаживания), увеличивается и уменьшается, соответственно.

Математически данный метод можно описать следующим выражением:

$$y_t = \frac{\sum_{i=t-p}^{t+p} y_i}{m},$$

где y_i – исходный временной ряд;

y_t – сглаженный временной ряд;

m – апертура;

p – количество наблюдений, стоящих по разные стороны от сглаживаемого элемента.

Алгоритм использования и программной реализации данного метода заключается в следующем:

- определяется величина апертуры (обязательно нечетное число);

- на первой итерации рассчитывается среднее значение наблюдений вошедших в апертуру, эта величина и является первым значением сглаженного ряда;

- затем интервал сглаживания сдвигается на один элемент вправо, и процедура повторяется;

- процедура продолжается до тех пор, пока в интервал сглаживания не войдет последнее наблюдение временного ряда.

При необходимости сохранить исходный размер временного ряда, те наблюдения, которые не вошли в процедуру сглаживания приравниваются соответственным значениям исходного ряда.

Метод взвешенной скользящей средней отличается тем, что сглаживание внутри интервала производится не по прямой, а по кривой более высокого порядка.

Элементы ряда входящие в интервал сглаживания суммируются с определенными весами, рассчитанными по методу наименьших квадратов. В том случае, если сглаживание производится с помощью полинома второго или третьего порядка, берутся веса следующих порядков:

- при $m=5$ - веса $1/35$ (-3; 12; 17; 12; -3);

- при $m=7$ - веса $1/21$ (-2; 3; 6; 7; 6; 3; -2).

Алгоритм работы метода является аналогичным методу скользящего среднего.

Алгоритм работы метода экспоненциального сглаживания состоит в том, что при процедуре сглаживания каждого наблюдения используются только значения предыдущих уровней, взятых с определенным весом. Сглаженное значение наблюдения ряда y_t на момент времени t определяется как:

$$y_t = \alpha \cdot y_i + (1 - \alpha) \cdot y_{t-1},$$

где α - вес выравниваемого наблюдения ($0 < \alpha < 1$).

Начальным значением принимается первый уровень исходного временного ряда. При выборе величины сглаживающего параметра α необходимо учитывать то, что чем выше значение параметра α , тем меньше сказывается влияние предшествующих уровней и, следовательно, меньшим оказывается сглаживающее воздействие экспоненциальной средней.

При исследовании влияния внешнего возмущающего воздействия (белый шум) на входной синусоидальный сигнал использовались средства цифрового USB-осциллографа (рисунок. 1).



Рис. 1. Исходный сигнал

Данный сигнал в виде числовой последовательности искусственно подвергался добавлению шумовой составляющей с различными уровнями сигнал/шум от 5 до 30 дБ. Результат такой операции показан на рисунке 2.

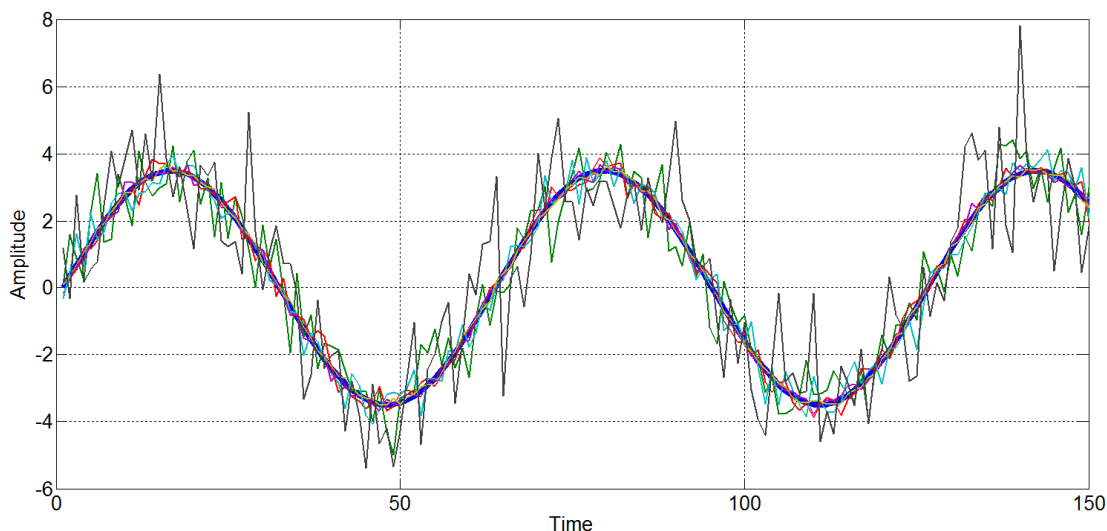


Рис. 2 . Зашумленный сигнал

Проведено сглаживание по выбранным методам. Так как необходимо сглаживать мелкие беспорядочные колебания для методов скользящего среднего была выбрана широкая апертура ($m=7$), для экспоненциального сглаживания уравновесим воздействие экспоненциальной средней выбором соответствующего значения $\alpha=0,5$.

Оценку степени сглаживания предложено производить по рассчитанной средней ошибке аппроксимации (количество отклонений по каждому наблюдению, в процентном выражении).

$$\bar{A} = \frac{1}{n} \sum \left| \frac{y_i - \hat{y}}{y_i} \right| \cdot 100\%,$$

где y_i – фактическое значение временного ряда;

\hat{y} – результат аппроксимации;

n – количество элементов временного ряда.

Полученные данные показаны на рисунке 3.

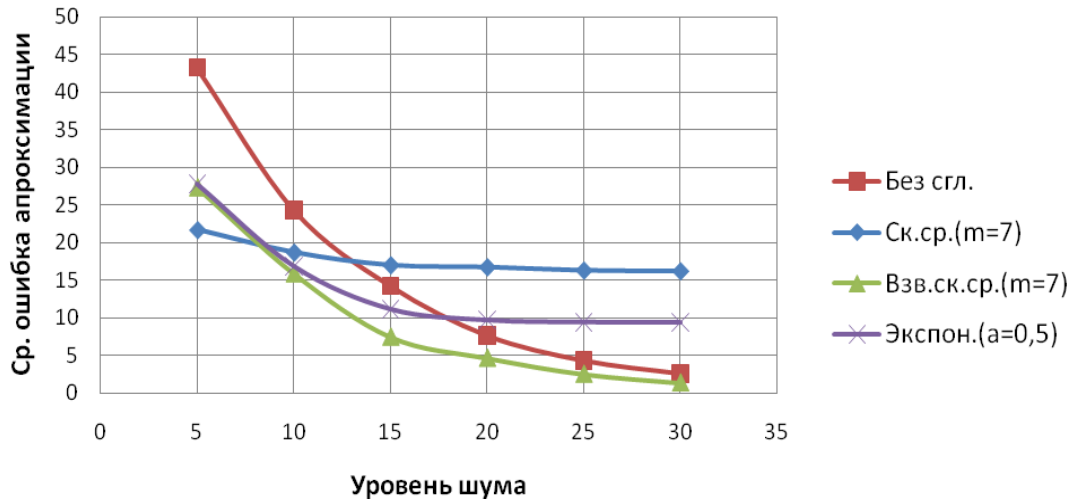


Рис.3. Оценка сглаживания

По графической зависимости был сделан вывод, что наиболее подходящим методом для синусоидального сигнала является метод взвешенной скользящей средней.

Предварительная обработка данных является также неотъемлемой составляющей процесса непараметрической идентификации линейных детерминированных объектов. Проведена идентификация апериодического звена 1-го порядка по переходной характеристике.

Для этого проведен натурный эксперимент с использованием генератора, измерительного комплекса ИМ-1 и лабораторного макета, реализующего типовые динамические звенья (рисунок 4).

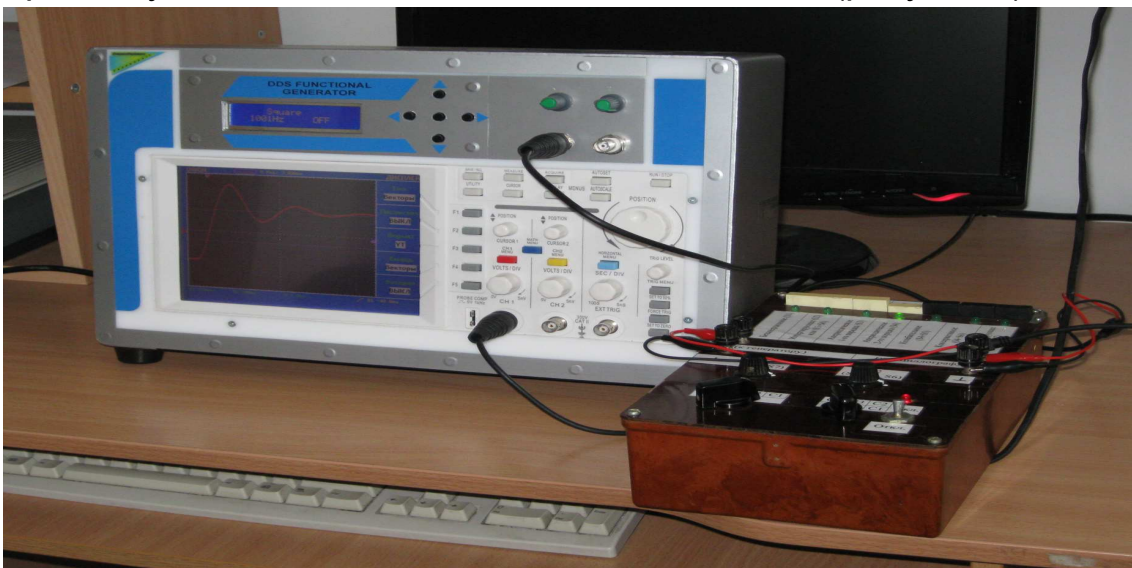


Рис. 4 . Экспериментальная установка

Проведено измерение ряда значение переходной характеристики исследуемого звена (рисунок 5).

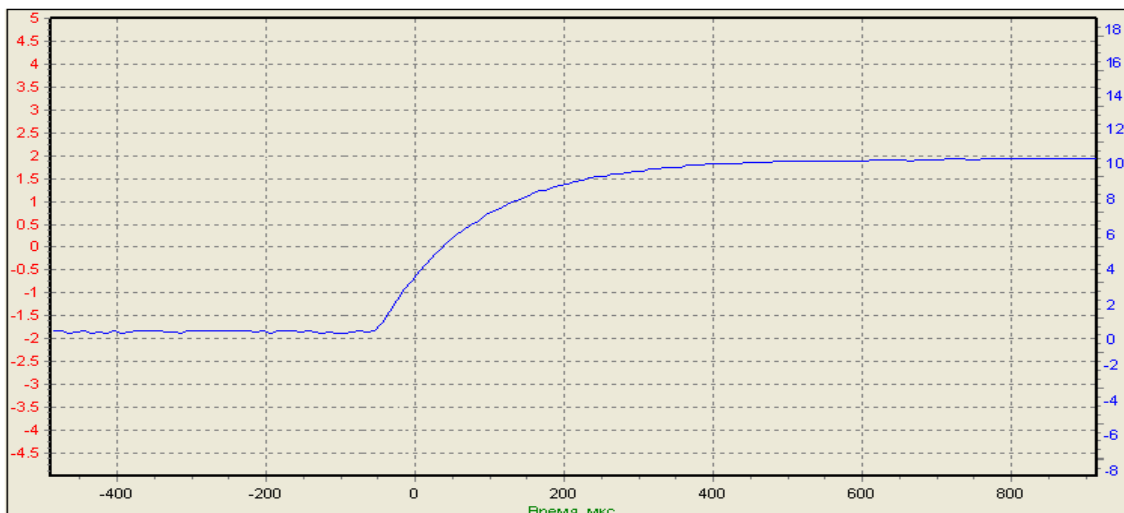


Рис. 5 .Исходный сигнал

Произведено добавление шумовой составляющей с различными уровнями сигнал/шум от 5 до 35 дБ, к исходным данным.

Проведено сглаживание, для методов скользящего среднего, берем среднее значение апертуры ($m=5$), для экспоненциального сглаживания также уравновешено воздействие экспоненциальной средней ($\alpha=0,5$). Затем рассчитана средняя ошибка аппроксимации. Полученные данные показаны на рисунке 6.

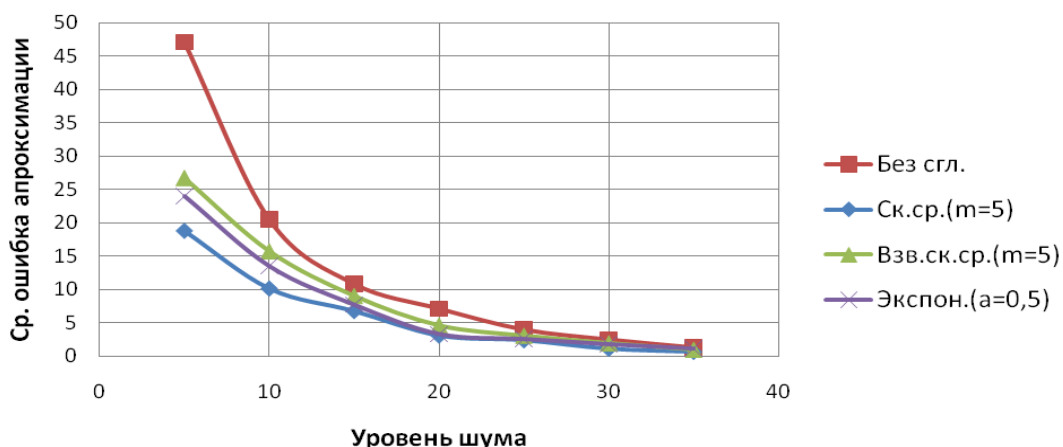


Рис. 6.Оценка сглаживания

Для оценивания корректности сглаживания построены автокорреляционные функции (АКФ) сигналов. На рисунке 7 показаны АКФ, до процедуры предварительной обработки временных рядов. На рисунке 8 показаны АКФ обработанных рядов.

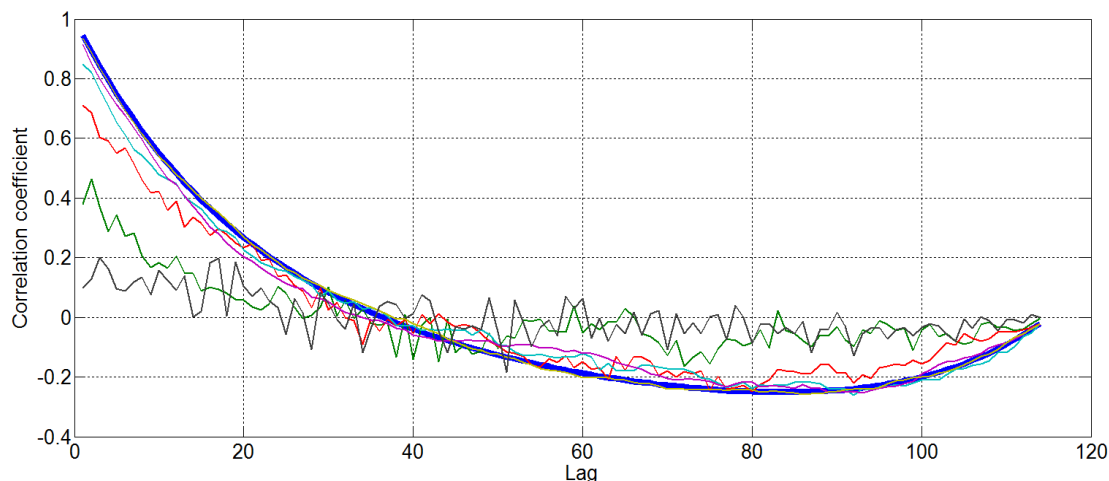


Рис. 7. АКФ до предварительной обработки

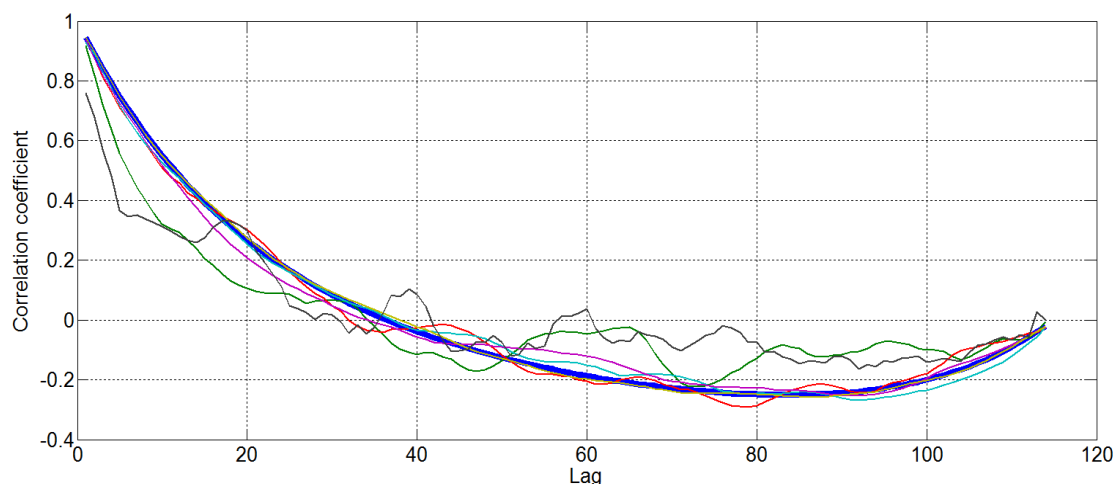


Рис. 8. АКФ после проведения процедуры предварительной обработки временных рядов

По графическим зависимостям был сделан вывод о том, что все использованные методы дают практически равновесные результаты, в данном случае наиболее предпочтительным является метод скользящей средней.

Стоит отметить, что при уровнях сигнал/шум менее 20дБ непараметрическая идентификация не обеспечивает заданную точность.

Для обеспечения заданной точности и корректности предварительной обработки выборка уровней временного ряда должна быть наиболее широкой.

Выводы.

После проведения исследований можно сделать вывод о неоспоримой необходимости предварительной обработки данных в задачах анализа входной информации в виде временных рядов. Выявлено, что для периодических сигналов или сигналов обладающих сезонностью наиболее предпочтительным методом является метод взвешенной скользящей средней. При сглаживании мелких беспорядочных колебаний, вызванных случайными помехами, апертуру стоит выбирать шире. Если возникает необходимость сохранять мелкие волны сигнала, апертуру фильтра необходимо уменьшать. При использовании фильтра, построенного на методе экспоненциального сглаживания необходимо учитывать то, что чем выше значение параметра α , тем меньше сказывается влияние предшествующих уровней сигнала.

Экспериментально установлено, что для задач непараметрической идентификации линейных детерминированных объектов, критическим является уровень сигнал/шум 20дБ.

С целью автоматизации процесса сглаживания временных рядов, рассмотренные методы реализованы в виде программных функций, что дает возможность непосредственно после снятия экспериментальных характеристик провести предобработку сигнала.

Литература

1. Бокс Дж., Дженкинс Г. Анализ временных рядов: прогноз и управление. Выпуск 1. М.: Мир, 1974. – 390 с.
2. Орлова И.В. «Экономико-математическое моделирование. Практическое пособие по решению задач» / И. В. Орлова; ВЗФЭИ. - М.: Вузовский учебник, 2004. - 144с.
3. Орлов А.Н. « Эконометрика: Учебное пособие» / А. Н. Орлов. - М.: Экзамен, 2002. - 576с.

E-MAIL: NURKA07@INBOX.RU
E-MAIL: TETYACYA@MAIL.RU