

Скользящее усреднение

Простой и очень эффективный метод сглаживания – это скользящее усреднение. Его суть состоит в расчете для каждого значения аргумента среднего значения по нескольким соседним данным. Число точек w , участвующих в расчете такого среднего, называют *окном* скользящего усреднения: чем оно больше, тем больше данных участвуют в расчете среднего и тем более сглаженная кривая получается. Формула для расчета скользящего среднего такова:

$$Y_j(w) = \frac{\sum_{i=j}^{j+w} y_i}{w} \quad (41)$$

На рис. 72 показан результат скользящего усреднения Y одних и тех же данных y (кружки) с разным окном:

$w=3$ (сплошная кривая) и $w=8$ (пунктир).

Видно, что при малых w сглаженные кривые практически повторяют ход изменения данных, а при больших w отражают лишь закономерность их медленных вариаций. Соответствующие результаты хорошо видны на графике Фурье-спектров (рис. 73) исходных данных (точки) и данных после скользящего усреднения с окном $w=3$ (сплошная кривая) и $w=8$ (пунктир). Видно, что в низкочастотной области спектр практически не меняется, а в высокочастотной – значительно уменьшается (тем заметнее, чем больше окно w).

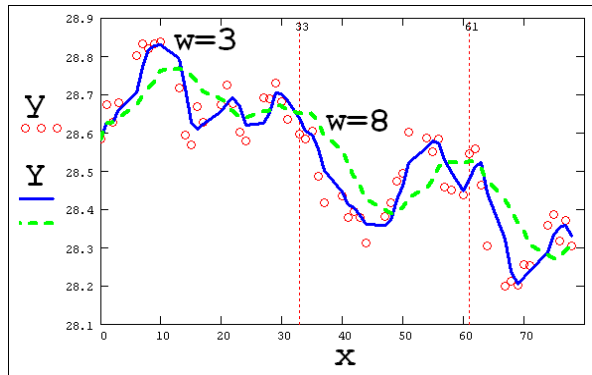


Рис. 72. Скользящие средние

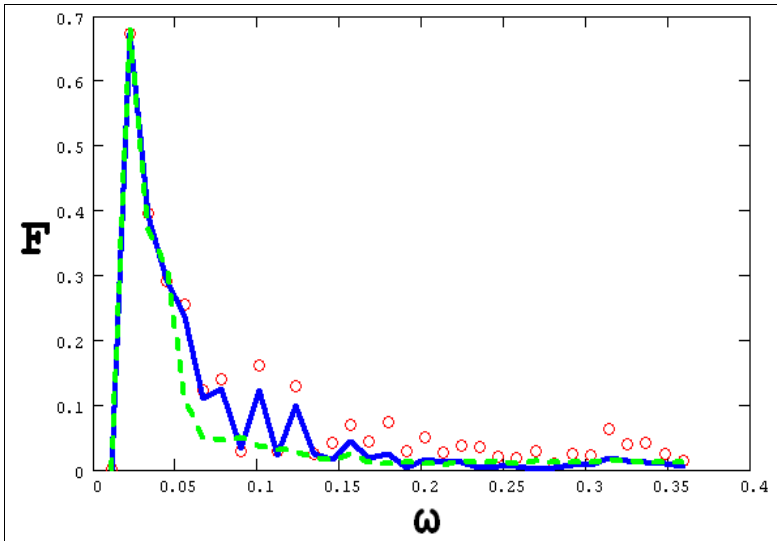


Рис. 73. Спектр исходных и скользяще-усредненных данных

Приведенная формула (41) скользящего усреднения самая простая, но не самая лучшая. Возможно, вы обратили внимание, что все кривые скользящего среднего на рис. 72 слегка «обгоняют» исходные данные. Почему так происходит, понятно: согласно (41), скользящее среднее для каждой точки вычисляется путем усреднения значений предыдущих w точек. Чтобы результат скользящего усреднения был более адекватным, лучше применить центрированный алгоритм расчета по $w/2$ предыдущим и $w/2$ последующим значениям. Он будет немного сложнее, поскольку придется учитывать недостаток точек не только в начале, но и в конце массива исходных данных.

Скользящие медианы

Похожий алгоритм, называемый *скользящими* (или, иначе, *бегающими*) *медианами*, основан на вычислении по w соседним данным не среднего, а медианного значения. Это – прекрасный инструмент для устранения локальных выбросов исходных данных. На рис. 74 приведено применение сглаживания тех же

данных (кружки) скользящими медианами с окном $w=3$ (сплошная кривая) и $w=7$ (пунктир).

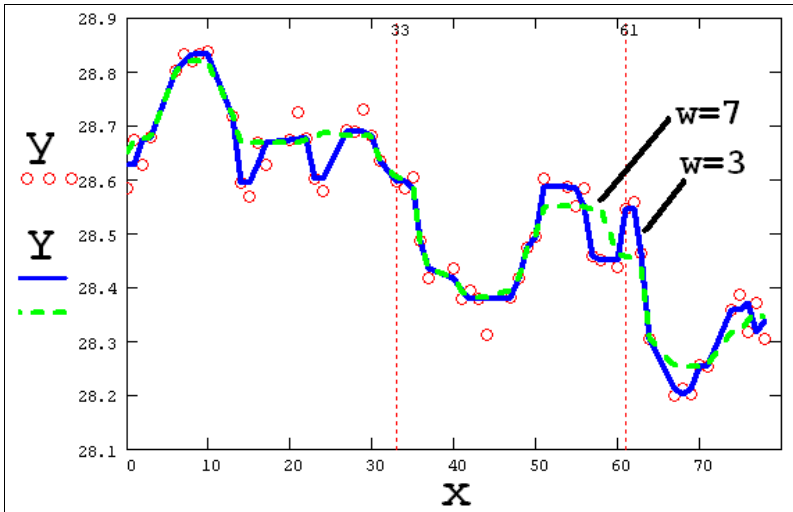


Рис. 74. Скользящие медианы

Устранение тренда

Еще одна типичная задача возникает, когда интерес исследований заключается не в анализе медленных (или низкочастотных) вариаций сигнала $y(x)$ (для чего применяется сглаживание данных), а в анализе быстрых его изменений. Часто бывает, что быстрые (высокочастотные) вариации накладываются определенным образом на медленные, которые обычно называют *трендом*. Часто тренд имеет заранее предсказуемый вид, например, линейный.

Чтобы устранить тренд, можно предложить следующую последовательность действий. Во-первых, вычислить регрессию $f(x)$, например, линейную, исходя из априорной информации о тренде. Во-вторых, вычесть из данных $y(x)$ тренд $f(x)$.

На рис. 75 показаны исходные данные (кружками), выделенный с помощью регрессии линейный тренд (сплошной прямой линией)

и результат устранения тренда (пунктир, соединяющий крестики). Напомним также, что пример устранения тренда для данных о курсах валют был приведен в самом начале главы, на рис. 5 и 6.

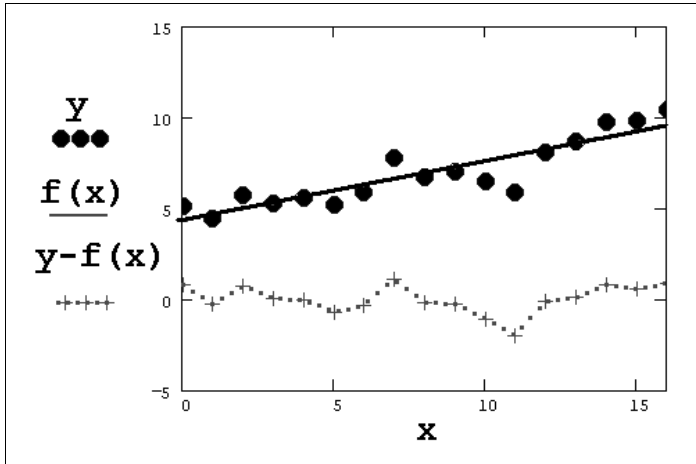


Рис. 75. Устранение тренда