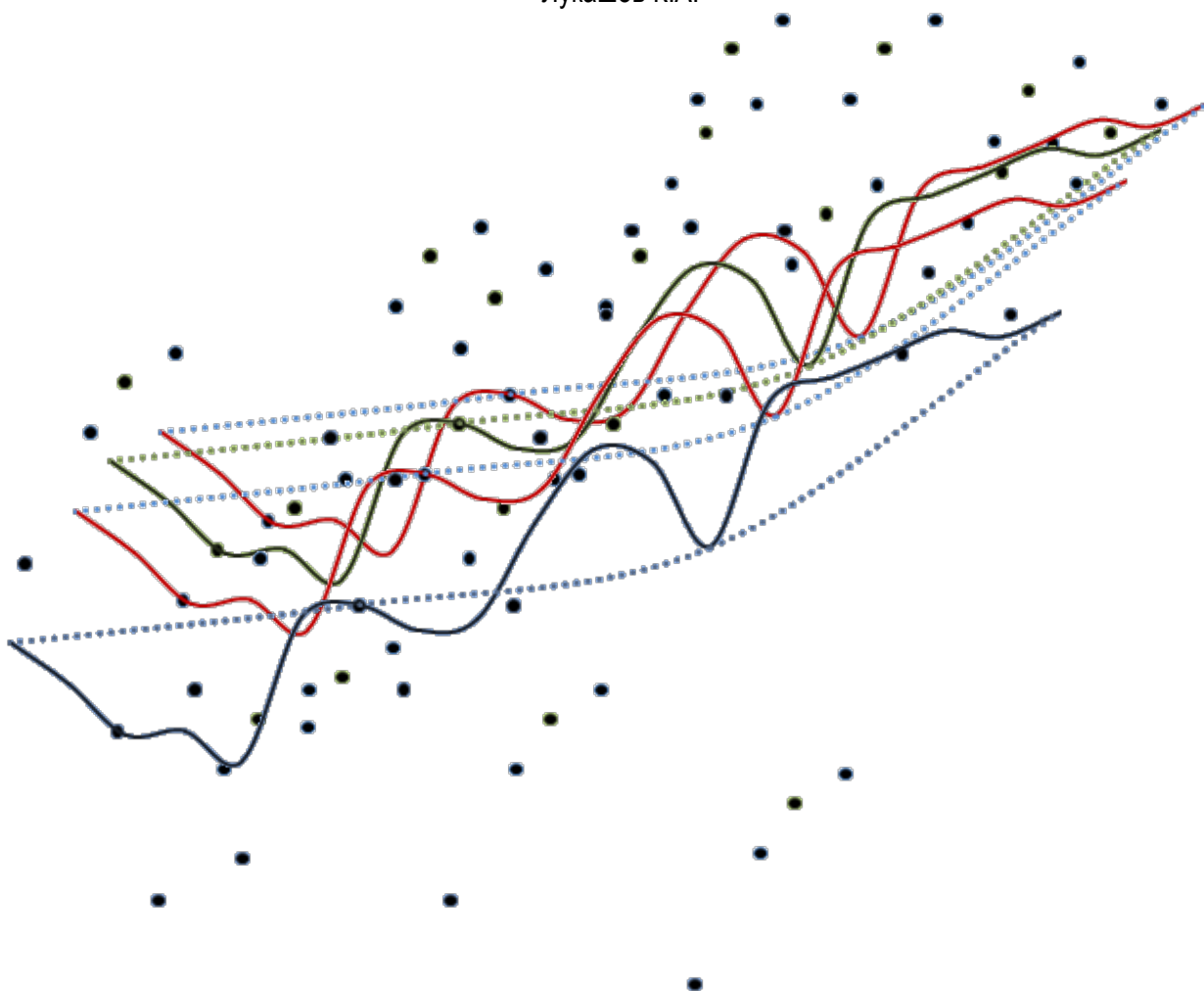


Построение линий тренда в декартовой системе координат

Методом линейной полосовой интерполяции переменной точности по
переменным весовым коэффициентам.

Лукашов К.А.



Содержание.

1.	Постановка задачи.....	3
2.	Методология.	4
a.	Определение количества и позиций баз. Разрешение.	5
b.	Определение уравнения весов.	5
c.	Влияние коэффициента трансформации на значения весов.	6
d.	Принцип полосовой интерполяции.....	7
3.	Примеры метода для рядов с тенденцией и облаков.....	7
4.	Заключение.	9

1. Постановка задачи.

Любой числовой, управленческий, коммерческий, финансовый или производственный процесс можно рассматривать в контексте поступающей и выходной информации. Процесс интерпретирует исходные данные согласно неким формулам преобразования поступающих на вход величин в величины выходные. Правильность обработки информации целиком зависит от степени автоматизации процесса преобразования данных, наглядности этого процесса и возможности доступа к функциям, описывающим конкретный процесс. Перечисленные возможности практически целиком отсутствуют в широко известных и доступных коммерческих автоматизированных вычислительных системах, таких как MS Excel, где можно лишь наблюдать характер тенденции, зачастую совершенно не удовлетворяющий аналитиков.

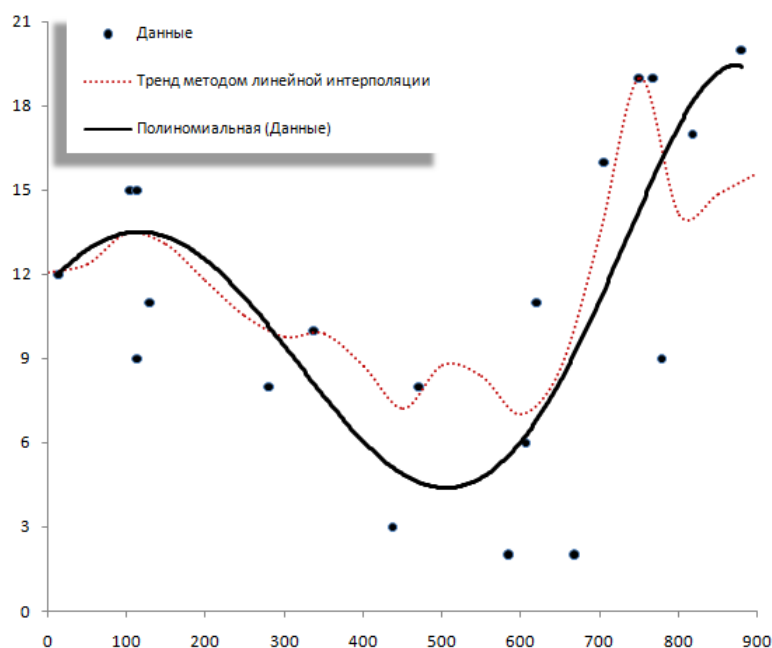


Рисунок 1. Встроенный полином 6-го порядка Excel не обеспечивает ни точности, ни параметрической регулировки при рассмотрении случайного ряда $x_{\max}=900$; $y_{\max}=20$

Практически всегда поступающая информация имеет на выходе разброс значений функции при одних и тех же значениях аргумента, как, например, в случае с разной температурой воды в водоеме при одинаковой температуре воздуха. Обусловлено это наличием внешних факторов, дисбалансирующих систему. Таких факторов настолько много, что, даже для самых простейших систем, анализ их влияния на результат является непосильно тяжелой задачей. На практике в большинстве случаев прибегают к анализу тенденции (тренда) или усредненному значению для всех возможных значений выхода по единичному входному значению. Основной проблемой данного подхода является тот аспект, что в ряде случаев влияние внешних факторов слишком высоко. Ряды данных несут беспорядочный характер без тенденции (облако). Данная проблема ставит задачу создания общей методики обработки рядов беспорядочных данных для получения усредненного значения функции для каждого значения аргумента, которая бы соотносилась со следующими задачами:

- Линии тренда должны быть описаны линейными уравнениями для простоты анализа и исключения вмешательства искажений криволинейных функций высших порядков;
- Линии тренда должны наиболее точно отображать процессы с тенденцией и без нее;
- Тренд должен содержать инструменты регулировки процесса обработки информации с целью настройки методики на каждый конкретный процесс. Данный пункт трансформируется в

создание общей формулы для любого процесса с параметрическим заданием коэффициентов при обработке информации;

- Методика должна быть приемлема для автоматического решения любой задачи подобного рода на современном предприятии.

2. Методология.

Идея метода возникла при рассмотрении графика производства заказов на мебельной фабрике. В силу определенных обстоятельств, связанных с логистикой, объем каждого из заказов мог принимать одно из трех - четырех значений с несущественными отклонениями. Срок исполнения заказа (выходная информация) не состоял в логической зависимости от количества единиц продукции по заказу. Следовательно, для определения сроков исполнения типового заказа можно было рассчитать средний показатель по всем предыдущим заказам данной количественной группы. Для обобщенной модели (облака) такой подход также вполне применим. Однако при стохастическом распределении параметров нельзя выделить строгие группы данных.

Различные научные теории используют понятие «веса» того или иного элемента данных при рассмотрении информационной совокупности с разными, а иногда противоположными, значениями. Примером может служить базовая дисциплина технико-экономической направленности Multicriteria decision making (Многокритерный выбор решения), где зачастую приходится разрешать компромисс между ценой и качеством.

В случае с облаком данных можно применить весовой метод. При этом каждый предыдущий заказ (в случае с мебельной фабрикой) несет информацию о производственных возможностях предприятия безотносительно к своему объему. Следовательно, вся без исключения история исполнения заказов должна стать информационной основой расчета сроков исполнения нового отвлеченного заказа. Объединенные по количественному признаку, различные заказы будут иметь различный вес в данном расчете. Близкие по количеству к новому старые заказы сходны в исполнении. Заказы, кардинально отличающиеся количеством от нового, имеют свои принципы исполнения и, соответственно, сроки. В противном случае, совокупно для всех заказов, выход (срок исполнения) был приблизительно пропорционален входу (количество единиц), а это совсем не так.

Таким образом, общий усредненный выход для данного значения входа - **базы** рассчитывается как среднее арифметическое значений функции всех точек ряда с учетом веса каждой из точек для каждой базы k :

$$y_{\text{ср}} = \frac{\sum_{i=1}^n y_i * w_i}{\sum_{i=1}^n w_i}; (2.1)$$

где i – порядковый номер точки от первой до последней (n);
 y_i – значение выхода для точки i ;
 w_i – вес выхода точки для данного значения аргумента.

Остается определить количество баз и формулу, однозначно определяющую вес каждой точки ряда для каждой из баз.

a. Определение количества и позиций баз. Разрешение.

Базовые значения данных могут быть определены непосредственно на каждой точке ряда. Однако при большом количестве исходных данных такой способ может оказаться (а) слишком трудоемким, (б) с неконтролируемой абсолютной точностью (в) как следствие, - с непредсказуемой и ненужной для анализа осцилляцией. Вместо такого, предлагается подход с фиксированными значениями баз. Переменное количество баз может обеспечить требуемую гладкость тренда, удобную для анализа. Трудоемкость составления таблицы значений весов для каждой из баз/ точки ряда зависит от технических возможностей автоматизированной системы расчета. Для примера с мебельной фабрикой, имеет смысл определить значения баз соответствующие среднему количеству корпусов, отправляемых автомобилем, контейнером и вагоном. Возможный разброс данных между этими тремя базами, вероятно, будет носить непостоянный характер, несущественный для детального анализа. Метод позволяет не сбрасывать со счетов даже самые незначительные данные, определив базы более часто.

Переменное положение базовых значений аргумента дает возможность изменять точность расчетов. Такую точность назовем разрешением по аналогии с механикой или цифровыми технологиями. Разрешение рассчитаем по формуле:

$$A = \frac{D}{d};$$

где D – ожидаемое или действительное максимальное значение аргумента - x_{\max} ;
 d – дискретность диапазона базы (или расстояние между базами) в тех же единицах.

Величина дискретности не может быть выше чем весь диапазон значений аргумента, следовательно, разрешение находится в пределах от 0 до 1. Значение 1 получится при определении двух баз, - в минимальной и максимальной точках и невыгодно тем, что линией тренда всегда будет прямая. Значение 0 получится при бесконечно большом количестве баз и невыгодно чрезмерной чувствительностью к разбросу точек ряда.

b. Определение уравнения весов.

Совершенно очевидно, что чем дальше находится точка от базы, тем вероятнее всего она в наименьшей степени соотносится со значением в базе. Данный принцип позволяет предположить, что вес значения точки ряда находится в обратной пропорциональной зависимости от расстояния до базы. Если определить возможные значения весов в пределах от 0 до 1 то можно записать значение веса точки i на любую из баз b общим уравнением:

$$w_{ib} = \frac{1}{|x_i - x_b| + 1};$$

где x_i – значение входного аргумента точки ряда;
 x_b – значение аргумента в базе.

Согласно уравнению, при бесконечно большом удалении от базы, вес точки в определении общего среднего арифметического ничтожен. Если точка максимально приближена к базе, то ее вес приближается к единице. Однако запишем конечное уравнение в более обобщенном виде, позволяющим ввести параметр контроля за величиной веса по усмотрению аналитика, оставляя действительными общие принципы вышеизложенного и пределы значений весов:

$$w_{ib} = \frac{1}{|x_i - x_b|/K + 1}; (2.2)$$

где K – коэффициент весовой трансформации.

с. Влияние коэффициента трансформации на значения весов.

Переменный коэффициент весовой трансформации позволяет аналитикам самим задавать относительное значение веса для каждой точки ряда исходных данных. При малых значениях коэффициента, веса соседних точек отличаются более разительно, чем при бо́льших.

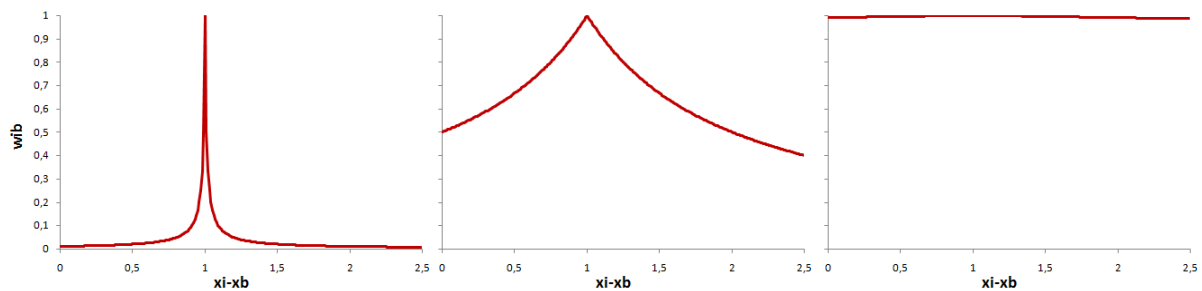


Рисунок 2. Влияние коэффициента трансформации на вес точки числового ряда.
Графики при $K=0,01$; $K=1$; $K=100$

Данный принцип позволяет аналитикам регулировать процесс с точки зрения распределения весов различных точек на каждую из баз. Если существует необходимость выделить только принципиальные значения (значения баз), то надо снизить значение коэффициента трансформации ниже единицы. Если есть желание задействовать все данные в равной степени безотносительно от их расстояния до конкретной базы, то нужно увеличить значение коэффициента до десятков или даже сотен. При бесконечно большом значении коэффициента трансформации, значения весов всех точек на все базы будут равны единице, производя горизонтальную линию тренда со значением выхода, равным среднему арифметическому значению выходов всех точек ряда данных.

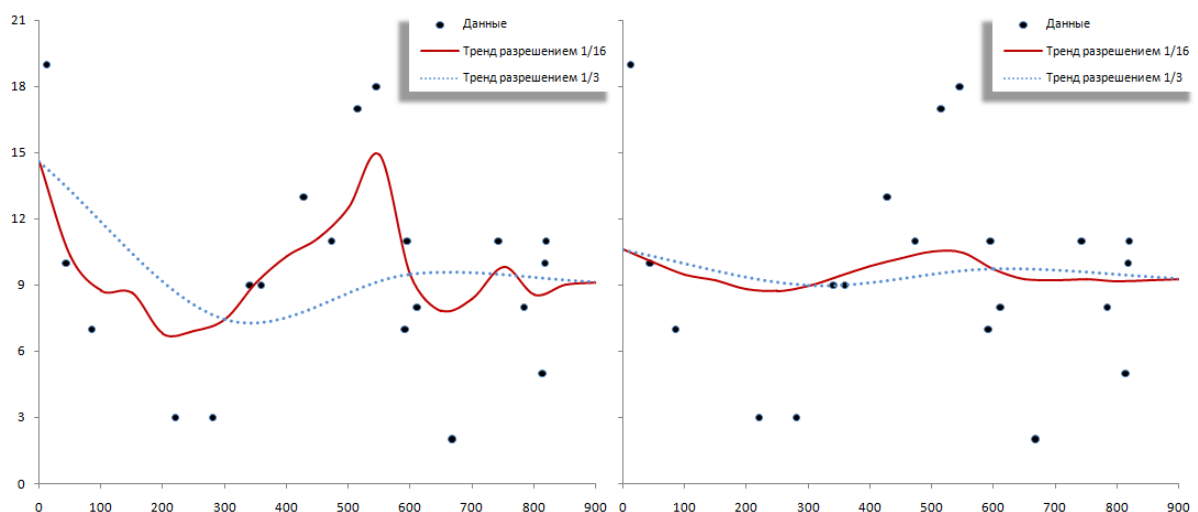


Рисунок 3. Линии тренда стохастического случайного ряда с диапазоном $x_{\max}=900$, $y_{\max}=20$ с коэффициентами трансформации $K=0,01$ и $K=100$ (расположение баз определяется разрешением)

d. Принцип полосовой интерполяции.

До сих пор рассматривалось влияние всех без исключения точек ряда данных на результат. Каждая точка имела свой вес в определении среднего арифметического значений функции для каждого из принципиальных значений – баз. Существуют процессы, где необходимо исключить некоторые значения ряда при рассмотрении какой либо базы. Примером может служить сезонность в потреблении тех или иных видов продукции, влияющая на продажи. Для извлечения конкретного диапазона данных при рассмотрении базы воспользуемся принципом

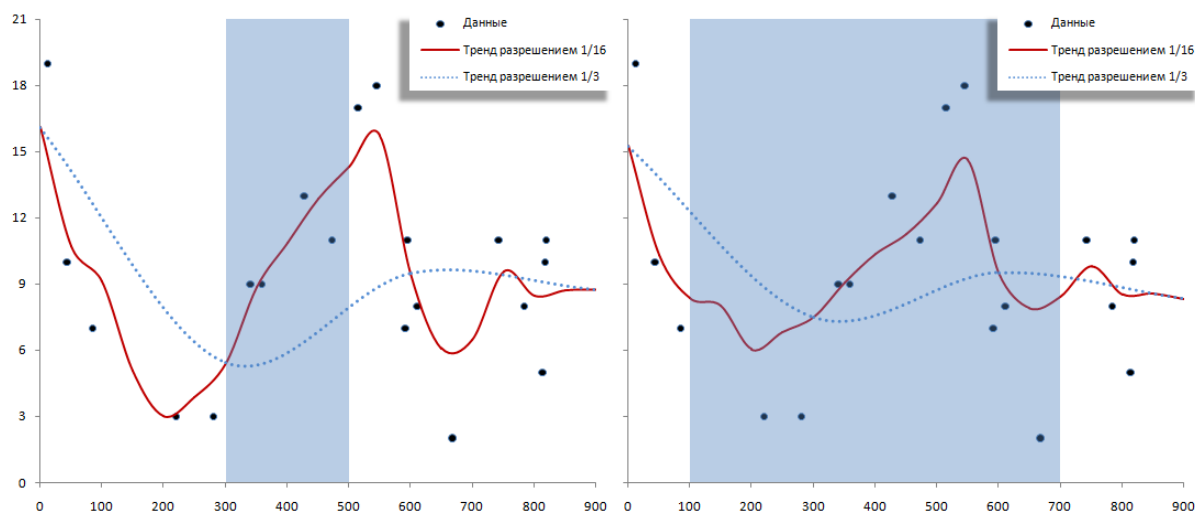


Рисунок 4. Примеры построения тенденций одного и того же ряда с шириной полосы 100 и 300. $K=1$. Полоса на графиках указывает задействованные точки для расчета весов для базы со значением аргумента 400

полосового фильтра информации. Только те данные, которые входят в полосу определенной ширины вокруг конкретной базы приобретут вес в соответствии с формулой 2.2, остальные же будут иметь нулевой вес для данной базы. Отметим, что полосы различных баз могут пересекать друг друга без какой-либо нежелательной интерференции. Данный метод дает еще больше возможностей аналитикам в определении линий тренда для процессов любого характера.

Чем шире полоса, тем больше точек она охватывает, тем больше влияние на конечное значение среднего арифметического в базе. Данный инструмент служит для настройки модели на конкретный процесс.

Перебором значения ширины полосы по какому-либо критерию можно создать автоматизированную адаптивную расчетную систему. Такая система внедрена на мебельной фабрике Россиббалт в Калининградской области для автоматического построения графика заказов.

3. Примеры метода для рядов с тенденцией и облаков.

Тренды рядов с тенденцией являются простейшими и желаемыми рядами данных, но и при их анализе требуется понимание фактора случайности. Процессы с заведомой тенденцией происходят с минимумом разброса значений. Следовательно, тенденция уже очевидна. В большинстве случаев такого типа входной информации аналитикам будет необходима высокоточная кривая, отражающая самые ничтожные неконтролируемые возмущения, как в случаях с температурами плавления легкоплавких металлов. Необходимо применять очень низкие коэффициенты трансформации с самым малым разрешением для получения прецизионных линий тренда. В этом случае осцилляции не должны чрезмерно проявлять себя.

Значения ширины полосы интерполяции можно оставлять на самом минимальном уровне. Бо́льшие значения ширины полосовой интерполяции приведут к сглаживанию кривой.

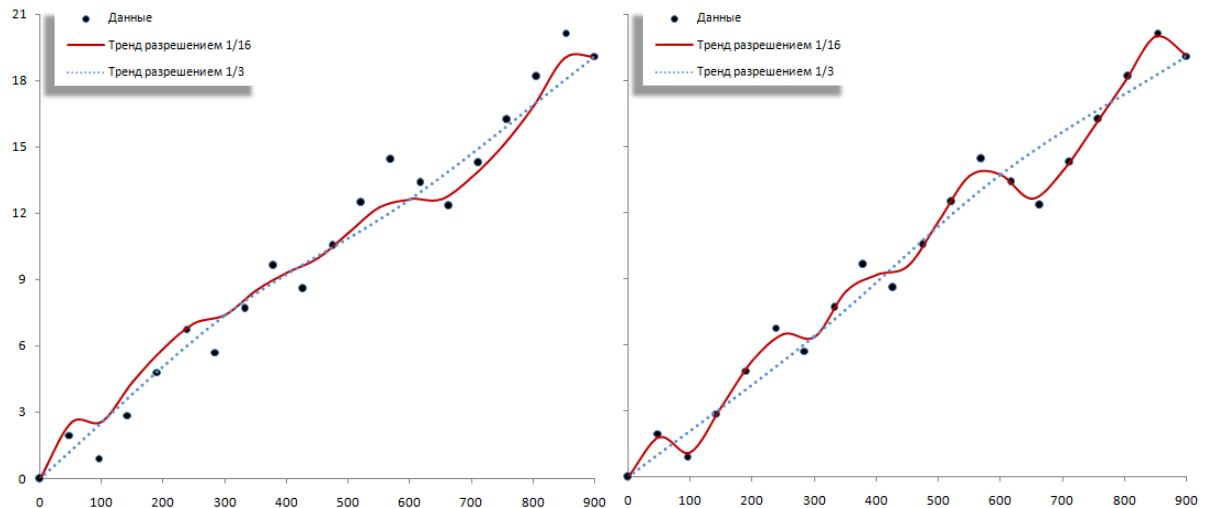


Рисунок 5. Линии тренда для случайного ряда с тенденцией. $K=0,0001$. Ширина полосы интерполяции 900 и 50.

Особенностью метода для рядов без тенденции (облаков) является расширенная возможность контролировать в самых широких пределах характер линейной кривой тренда. Подбором параметров ширины полосы интерполяции, коэффициента трансформации и разрешением можно получить желаемую форму кривой от самой чувствительной до характеристической. Форма линии меняется кардинально при изменении параметров.

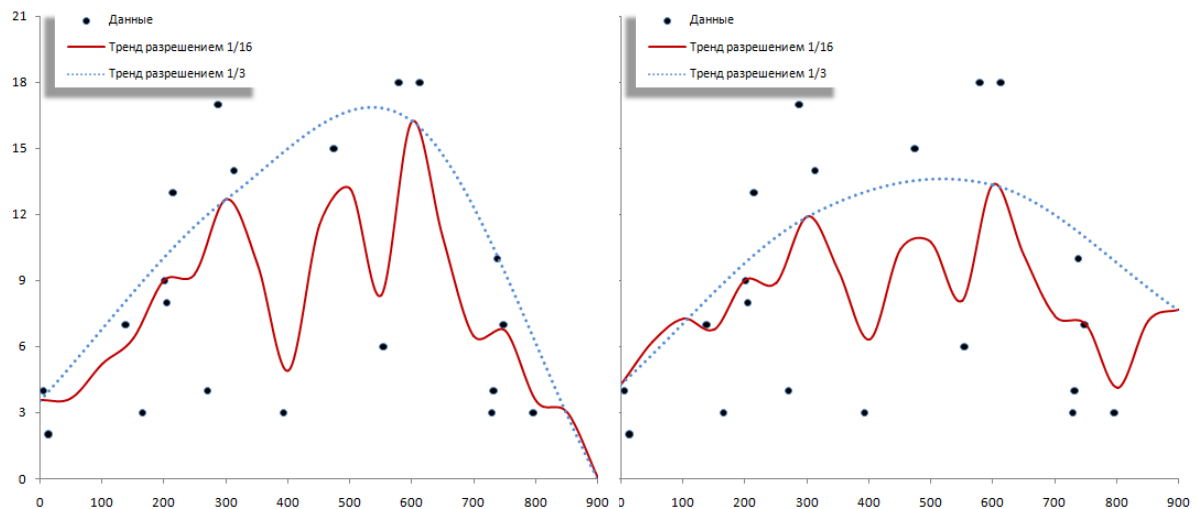


Рисунок 6. Линии тренда облака с шириной полосы интерполяции 80 и 900. $K=0,0001$

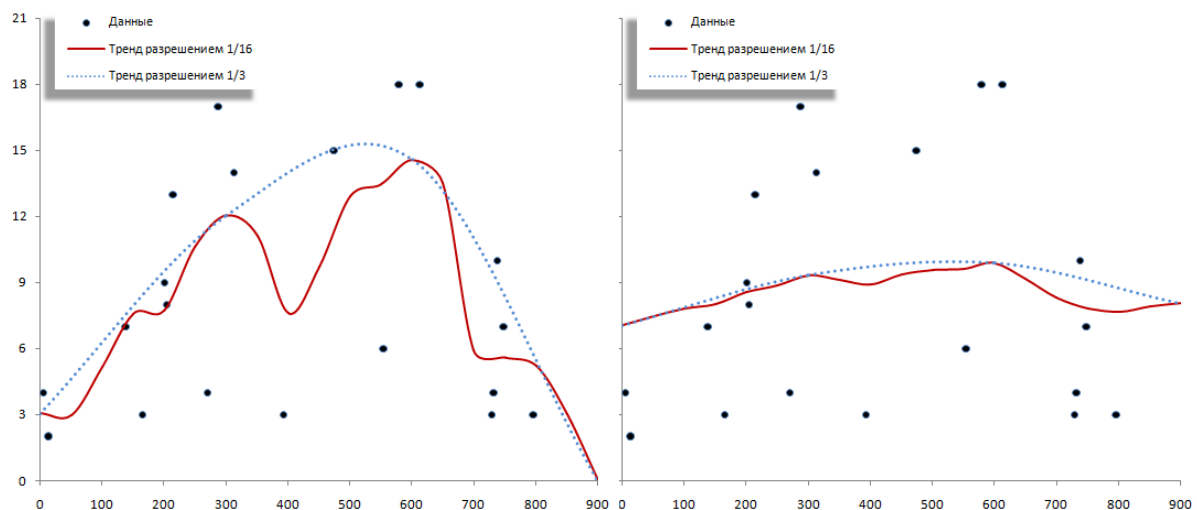


Рисунок 7. Линии тренда облака с шириной полосы интерполяции 80 и 900. $K=100$

4. Заключение.

Построение линий тренда с помощью данного метода сводится к выполнению следующих элементарных действий:

- Необходимо определить базовые значения ряда данных исходя из требований к разрешению кривой. Рекомендуемые значения разрешения: $1/16 \leq A \leq 1/3$, т.е. весь диапазон данных необходимо разделить на 3 – 16 (необязательно равных) промежутков;
- Линия тренда имеет только одну точку с координатой аргумента, соответствующей каждой из базовых координат. Координата функции точки линии тренда рассчитывается как среднее арифметическое координат функции всех точек, входящих в полосу интерполяции, с учетом веса каждой из точек по формуле 2.1.;
- Ширина полосы интерполяции определяет количество точек вокруг базы имеющих вес для данной базы, веса остальных данных считаются равными нулю. Вес точки имеет значение $0 \div 1$ и рассчитывается как обратное пропорциональное расстоянию до базы по формуле 2.2. Коэффициент трансформации в определении веса отвечает за чувствительность линии тренда к значениям близлежащих точек в пределах полосы интерполяции;
- После расчета координат функции точек тренда, они соединяются прямыми отрезками, образуя ломаную линию, каждый сегмент которой описывается аналитическим уравнением по двум точкам.

Таким образом, возвращаясь к техническому заданию метода, можно заключить следующее:

- Построение линий тренда элементарно и не требует знания основ полиномиальной алгебры. Построение происходит прямолинейными отрезками через вычисляемые точки на основе средних арифметических значений ряда данных;
- Линии тренда отображают процессы с тенденцией и без нее с точностью, задаваемой аналитиком или автоматически по адаптивным признакам;
- Точность построения регулируется тремя параметрами: общим разрешением графика, шириной полосы интерполяции и коэффициентом трансформации. Первые два параметра носят условный характер и выражают размерность входящей информации. Коэффициент трансформации определяет значимость значений точек внутри полосы интерполяции по отношению друг к другу и задается индивидуально для каждого процесса;
- Методика приемлема для автоматизированного усреднения любого процесса числового ракурса за счет простоты алгоритма, быстроты вычислений и широких возможностей для аналитика менять формы кривых без знания основ алгебры.