

Временной тренд

Профессор Кузнецов Анатолий Дмитриевич
Доцент Сероухова Ольга Станиславовна
Доцент Восканян Карина Левановна

Временной тренд временного ряда – это гладкая функция, описывающая его долгосрочное поведение.

Выявление тренда во временных рядах метеорологических величин является важным и весьма распространенным этапом их анализа.

Нахождение временного тренда – это задание вида этой функции и определение ее параметров (коэффициентов) по имеющимся значениям выборки исследуемого временного ряда.

Вид функции, описывающей временной тренд, не определяется однозначно самим рядом и является некоторым условным объектом, используемым для более полного понимания особенностей рассматриваемого процесса.

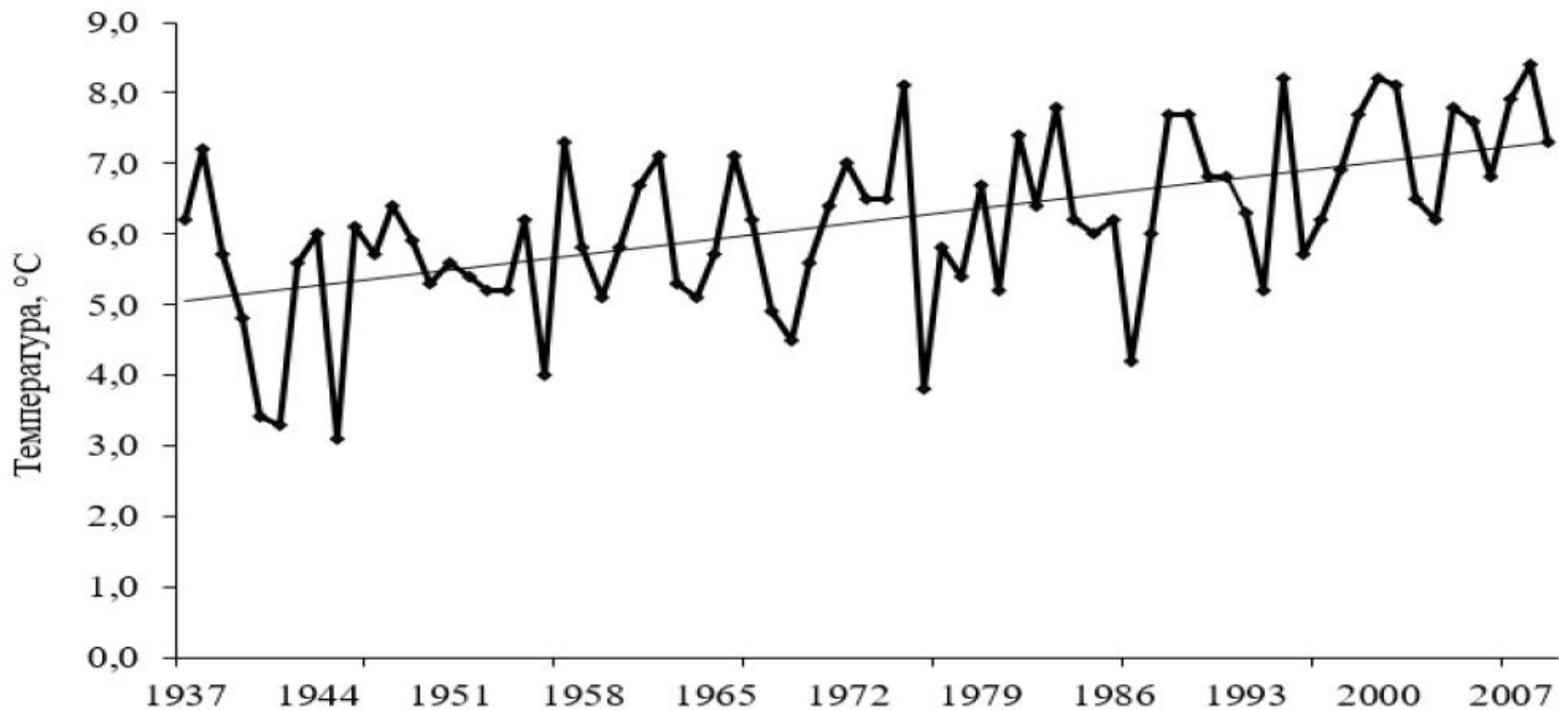


Рис. 1. Изменение средней годовой температуры воздуха
(по данным метеостанции Саратов Юго-Восток).

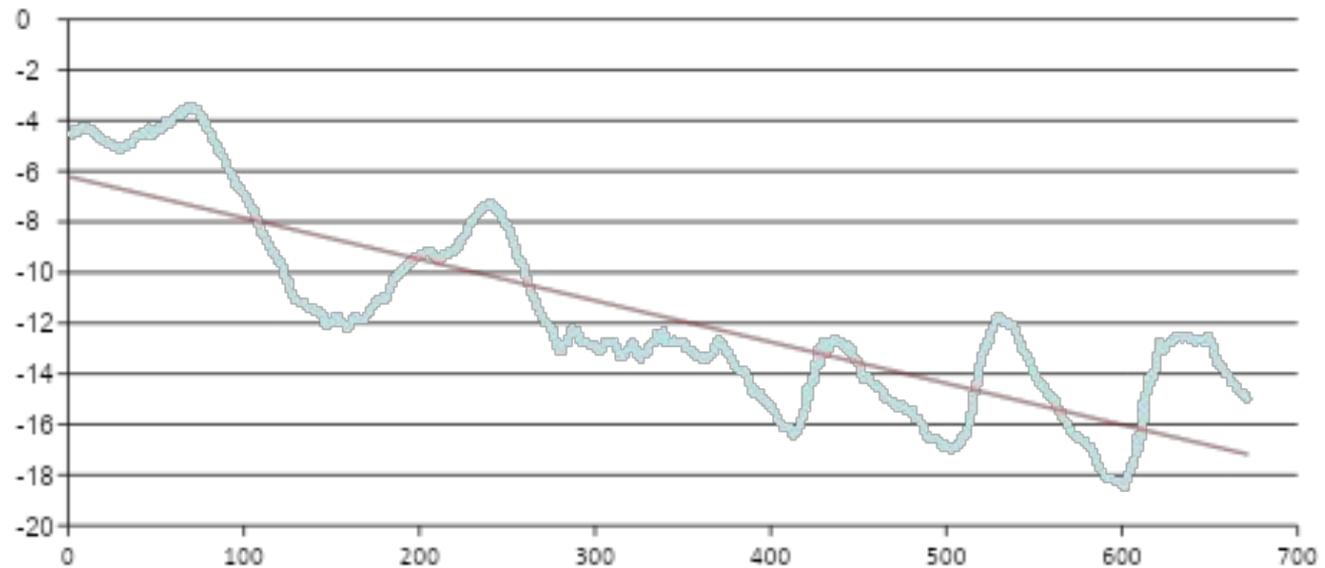
$$y = 0,0313x - 55,583$$

Достаточно легко определить временной тренд, если члены временного ряда монотонно изменяются во времени. (значения временного ряда устойчиво возрастают или устойчиво убывают). В этом случае наличие тренда часто хорошо видно на графике.

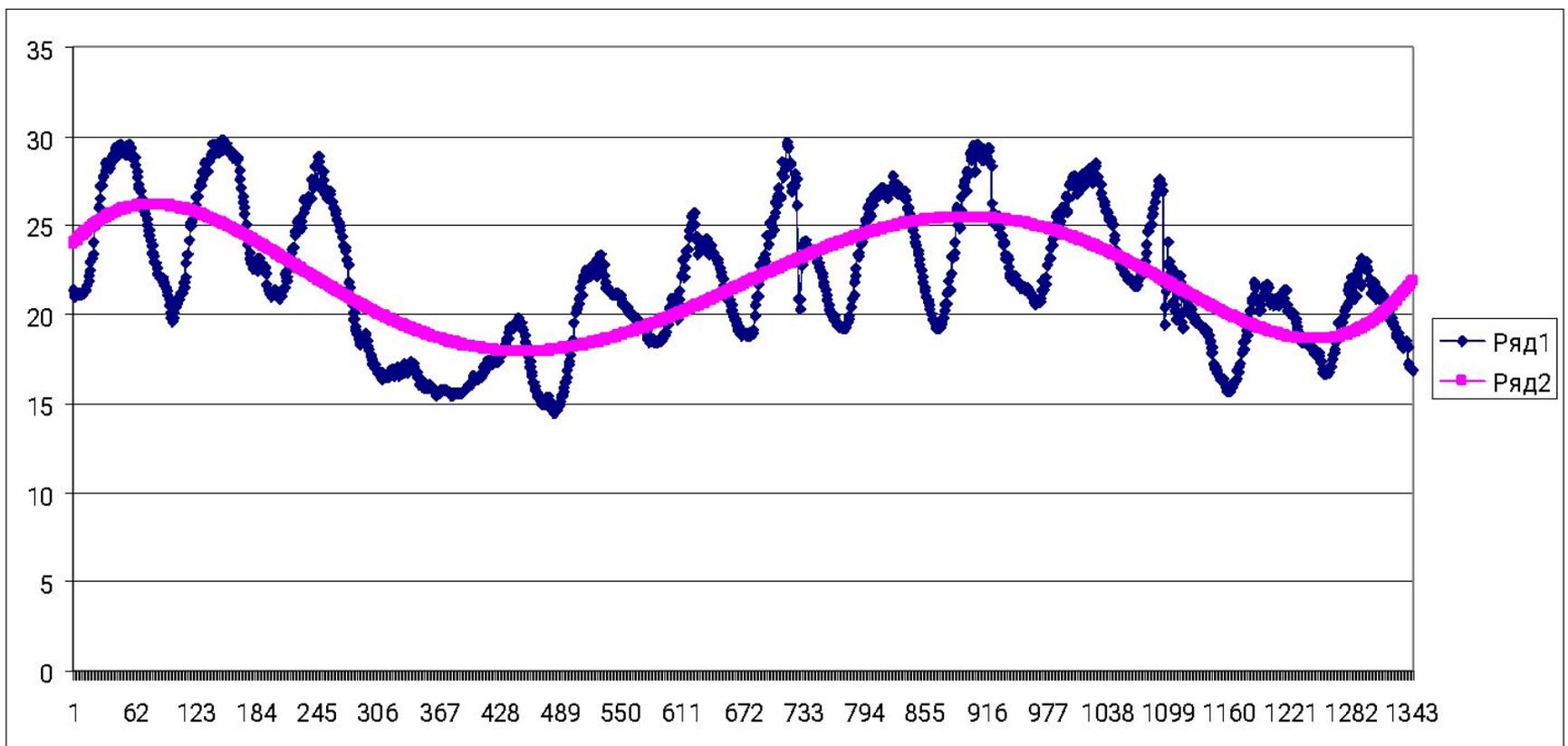
Распределение значений температуры, измеренных с 22.01.2012 по 28.01.2012г.

Значения температуры, °С

Порядковый номер

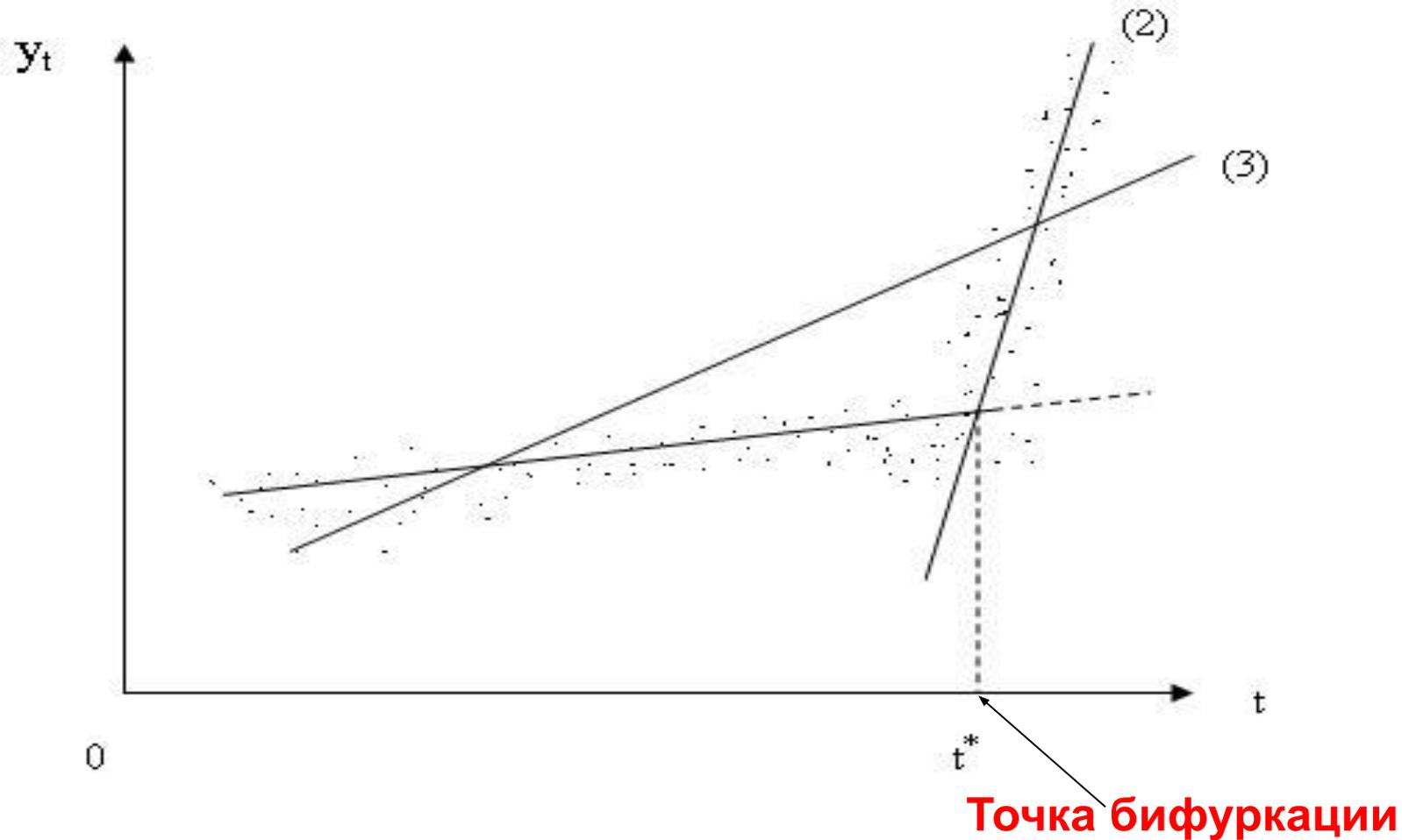


Пример линейного временного тренда



C	D	E	F	G	H
КОЭФФИЦИЕНТЫ ПОЛИНОМА					Программа Тренд полиномиальный
0	23.900				В колонке A - исходный временной ряд длиной N, $T(i) = a_0 + a_1 * i + a_2 * i^2 + \dots + a(stp) * i^{stp}$
1	0.06285				Общая длина временного ряда N =
2	-0.000520869				Степень полинома stp =
3	0.00000				
4	-1.0371E-09				
5	0.0000				Дата начала расчета
					01.03.2016
					Среднее (от 1 до n) =
					СКО (от 1 до n) =

Пример нелинейного временного тренда (полином степени 5)



Сложный случай - когда в поведении временного ряда не прослеживается длительная монотонность. В этом случае необходимо осуществить проверку гипотезы о существовании **устойчивого** временного тренда на всем временном интервале с использованием критериев или тестов.

Наиболее простым видом функции, описывающей временной тренд, является линейная функция.

Уравнение линейного временного тренда может быть представлено следующей формулой:

$$y(t) = a_1 t + a_0 + \varepsilon ,$$

где t – время; a_0 и a_1 – коэффициенты тренда; ε – ошибка трендовых составляющих.

Уравнение нелинейного (квадратичного) тренда:

$$y = a_2 t^2 + a_1 t + a_0 + \varepsilon$$

Коэффициенты уравнения тренда находятся методом наименьших квадратов.

Сущность данного метода заключается в нахождении параметров модели, при которых **минимизируется** сумма квадратов отклонений фактических значений временного ряда от теоретических (по выбранному уравнению временного тренда), то есть:

$$S = \sum_{i=1}^n (y_i^{\delta} - y_i)^2 = \sum_{i=1}^n (y_i^p - a_0 - a_1 x)^2 \rightarrow \min$$

где y_i – значение, вычисленное по уравнению тренда;

$(y_i^p - y_i)$ – отклонение ε для каждого значения временного ряда (см рис.);

n – количество данных.

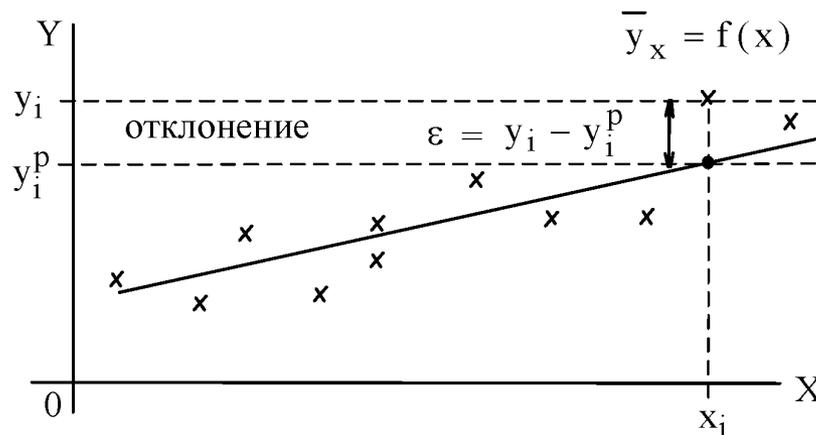


Рис. Понятие отклонения ε

Проведя необходимые преобразования, получим для определения коэффициентов систему двух линейных уравнений с двумя неизвестными a_0 и a_1 . Решение этой системы имеет следующий вид:

$$a_1 = \frac{n(\sum y_i x_i) - \sum y_i \sum x_i}{n(\sum x_i^2) - (\sum x_i)^2};$$

$$a_0 = \frac{1}{n}(\sum y_i - a_1 \sum x_i)$$

Оценка

«СООТВЕТСТВИЯ»

**временного тренда -
коэффициент детерминации R^2 .**

Коэффициент детерминации R^2 позволяет оценить качество аппроксимации табличных данных выбранным видом временного тренда, а величина $(1 - R^2)$ характеризует долю необъяснённой временным трендом дисперсии (дисперсию случайной ошибки модели временного тренда).

Термин R^2 кроме коэффициента детерминации также носит следующие названия: достоверность аппроксимации, квадрат смешанной корреляции.

Коэффициент детерминации R^2 для табличных данных:

$$\{t_i, y_i, i = 1, \dots, N\}$$

и уравнения временного тренда

$$y = w(t)$$

определяется следующим соотношением:

$$R^2 = S_{\text{рег}} / S_{\text{общ}}$$

Здесь

$$y_{\text{сум}} = \sum y_i \quad S_{\text{общ}} = \sum (y_i - y_{\text{сум}})^2 \quad S_{\text{рег}} = \sum [w(t_i) - y_{\text{сум}}]^2$$

где суммирование по индексу i ведется от 1 до N .

В зависимости от уровня коэффициента детерминации, принято разделять модели на **три группы**:

- $0,8 - 1$ — модель хорошего качества;
- $0,5 - 0,8$ — модель приемлемого качества;
- $0 - 0,5$ — модель плохого качества.

В последнем случае качество модели говорит о невозможности её использования для прогноза.

Основная проблема применения R^2 заключается в том, что его значение увеличивается (не уменьшается) от добавления в модель новых переменных (в нашем случае – при повышении степени аппроксимирующего полинома), даже если эти переменные никакого отношения к объясняемой переменной не имеют!

Поэтому сравнение моделей с разным количеством факторов с помощью коэффициента детерминации, вообще говоря, некорректно.

Для этих целей можно использовать альтернативные показатели.

Например, использовать при полиномиальной аппроксимации **скорректированным коэффициентом детерминации**:

$$r^2 = 1 - (1 - R^2) * [n / (n - k)] ,$$

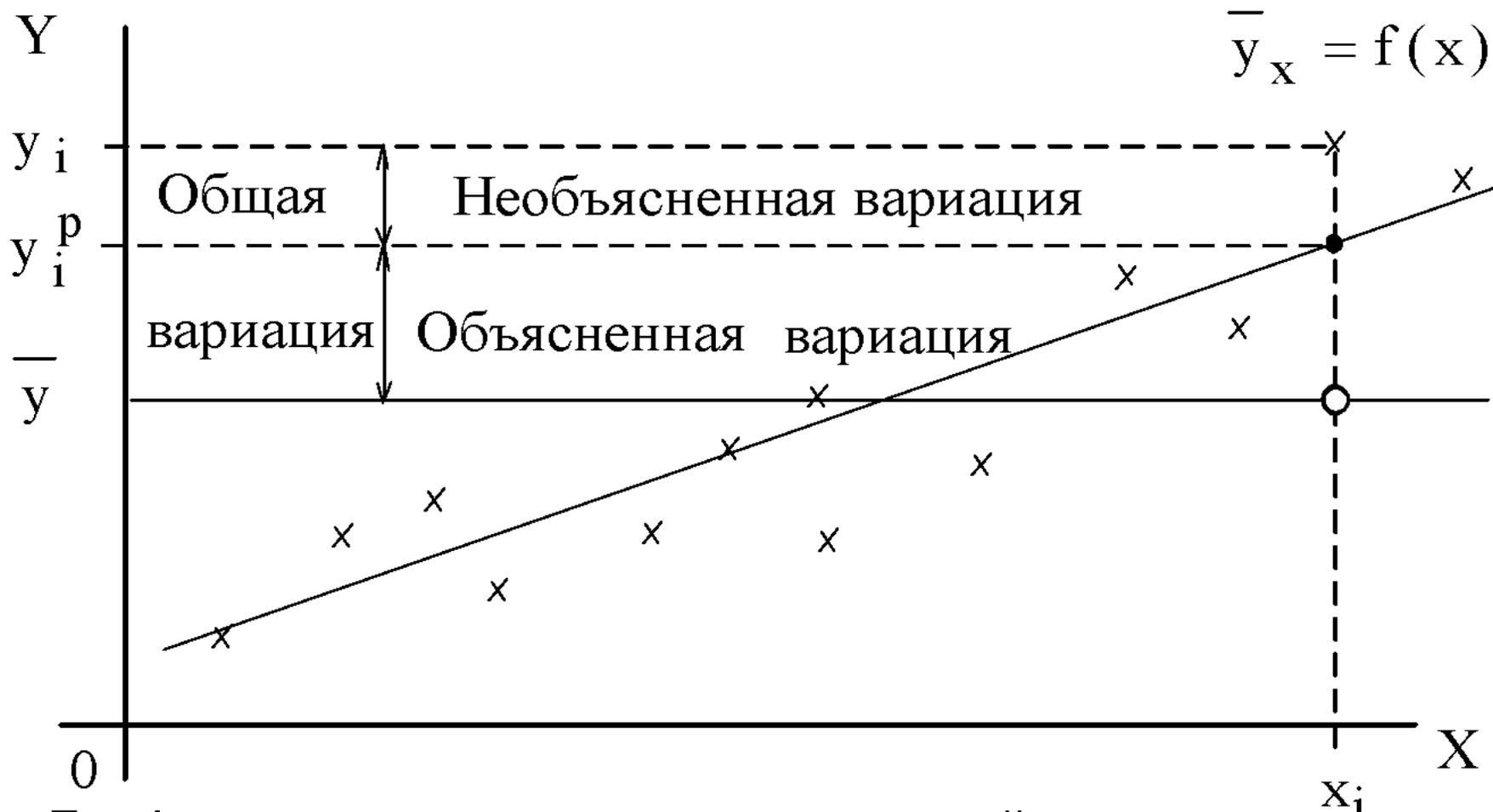
где n - число наблюдений; k - число оцениваемых параметров (коэффициентов) аппроксимирующего временной тренд полинома.

Интерпретация составляющих коэффициента детерминации R^2

Коэффициент детерминации R^2 показывает, на сколько процентов ($R^2 * 100$) найденная функция тренда описывает связь между исходными значениями факторов x и y (в нашем случае времени и метеорологического параметра)

$$R^2 = \frac{\sum_{i=1}^n (y_i^{\delta} - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

где числитель – объясненная временная изменчивость метеорологической величины за рассматриваемый промежуток времени (ее вариация); знаменатель – общая временная изменчивость (наблюдаемая вариация).



Графическая интерпретация значений слагаемых в числителе и знаменателе в формуле для коэффициента детерминации R^2 для случая линейного временного тренда (x это время t): $f(x) = a_0 + a_1 \cdot x$

**Оценка
значимости
коэффициентов
линейного временного тренда**

При определении временного линейного тренда важно оценить его значимость, т.е. насколько обоснован его вклад в описание изменчивости временного процесса, поскольку вклад тренда в общую дисперсию временного ряда может быть как значимым, так и незначимым.

Алгоритм оценки значимости линейного временного тренда:

1. Рассчитывается коэффициент детерминации R^2 , по величине которого определяется коэффициент корреляции r между линейным трендом и фактическим рядом:

$$r = \sqrt{R^2}$$

2. Рассчитывается эмпирическое значение критерия Стьюдента:

$$t = \frac{r \sqrt{N-1}}{\sqrt{1-r^2}} = \sqrt{\frac{R^2 (N-1)}{1-R^2}}$$

3. Определяется критическое значение критерия Стьюдента на уровне значимости α : $t_{\text{кр}}(\alpha, N-1) = \text{СТЮДРАСПОБР}(1-\alpha, N-1)$

4. Если выполняется неравенство

$$t > t_{\text{кр}}(\alpha, v = N-1)$$

то линейный тренд значим.

**АЛГОРИТМ
ВЫЯВЛЕНИЯ
ЛОКАЛЬНЫХ ТРЕНДОВ
ПРИ АНАЛИЗЕ
МЕТЕОРОЛОГИЧЕСКИХ
ВРЕМЕННЫХ РЯДОВ**

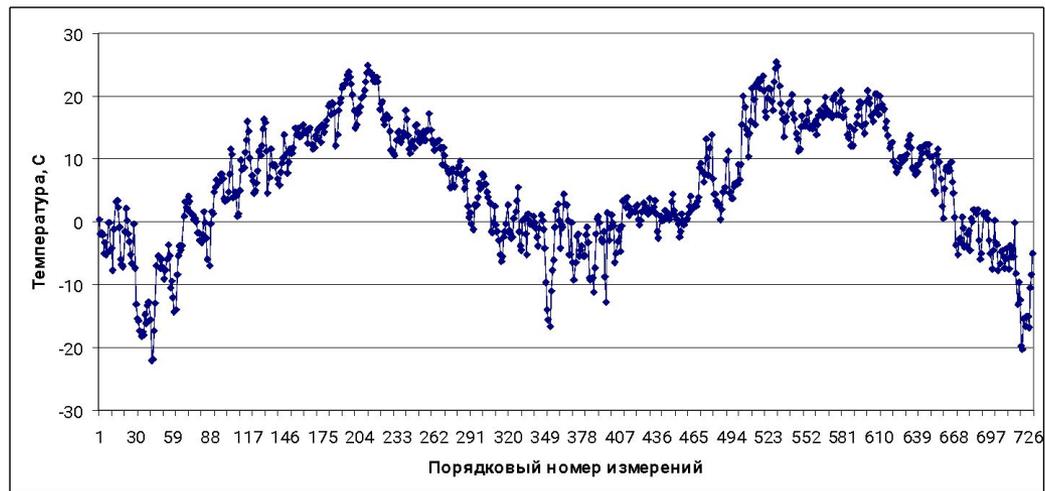


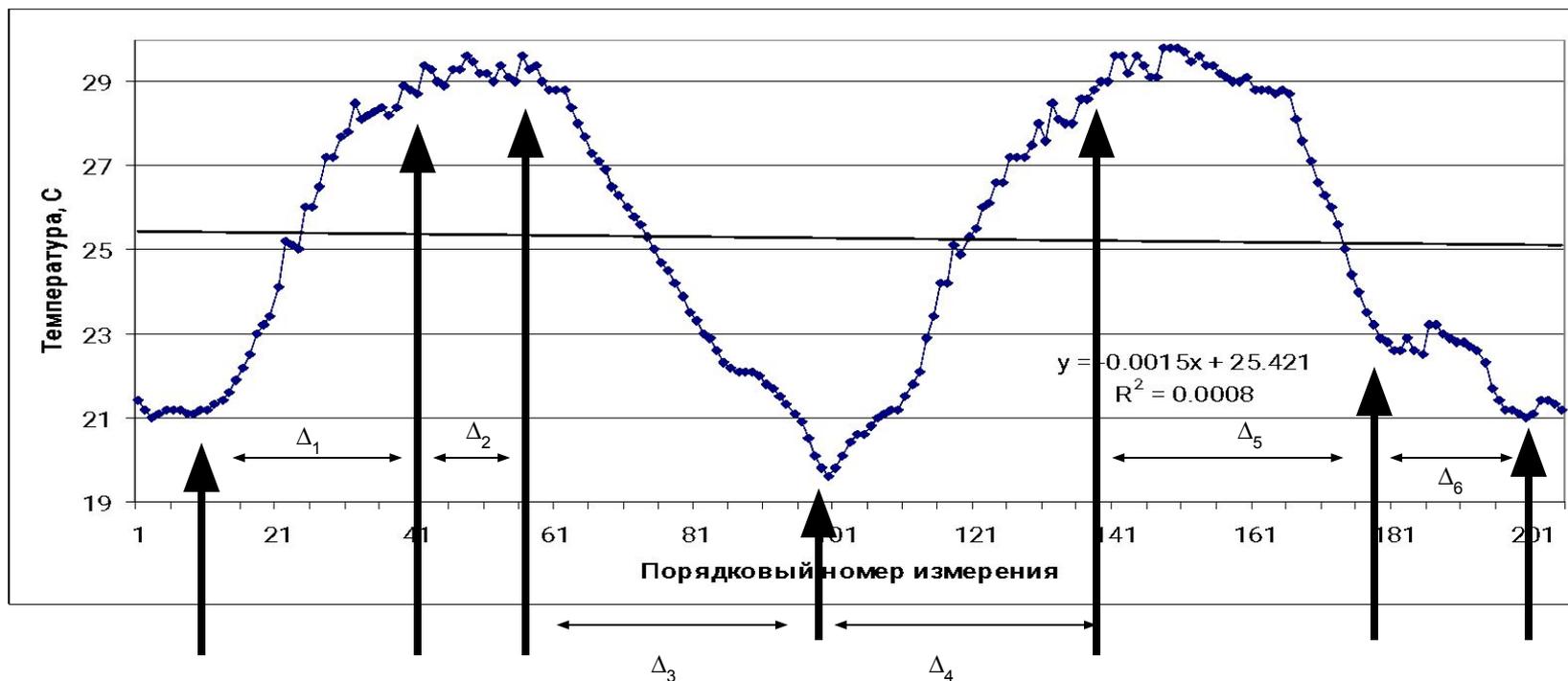
График среднесуточной температуры воздуха в СПб за 1994-1995 гг



График температуры воздуха в СПб с дискретностью 15 мин за 2 суток: $R^2 = 0.008$

Определение границ локальных участков временного ряда, имеющие устойчивые тенденции к изменению с целью последующего расчета в них временного тренда не является тривиальной задачей.

Одним из подходов к решению задачи выделения во временном ряде локальных временных отрезков Δ_i с устойчивыми тенденциями является его визуализация (построения графиков). Однако такой подход весьма субъективен, а при значительной длительности временного ряда - весьма трудоемок.



Рассмотрим возможности использования известного и весьма простого подхода к автоматизации решения задачи определения параметров локальных трендов внутри временного ряда.

Пусть имеется регулярный временной ряд, который представим в следующем виде:

$$P_i = P(\tau_i), \quad (1)$$

где

i – порядковый номер члена временного ряда: $i = 1, 2, \dots, N$,

$$\tau_i = \tau_0 + \Delta \tau * i,$$

τ_0 – начальный момент проведения измерений,

$\Delta \tau$ - дискретность измерений (временной интервал между измерениями).

Для выделения «локальных» по своей тенденции участков временного ряда необходимо сформировать вспомогательный временной ряд

$$r_j = R(\tau_j), \quad j = 1, 2, \dots, N, \quad (2)$$

значения которого будут индикаторами наблюдающейся тенденции.

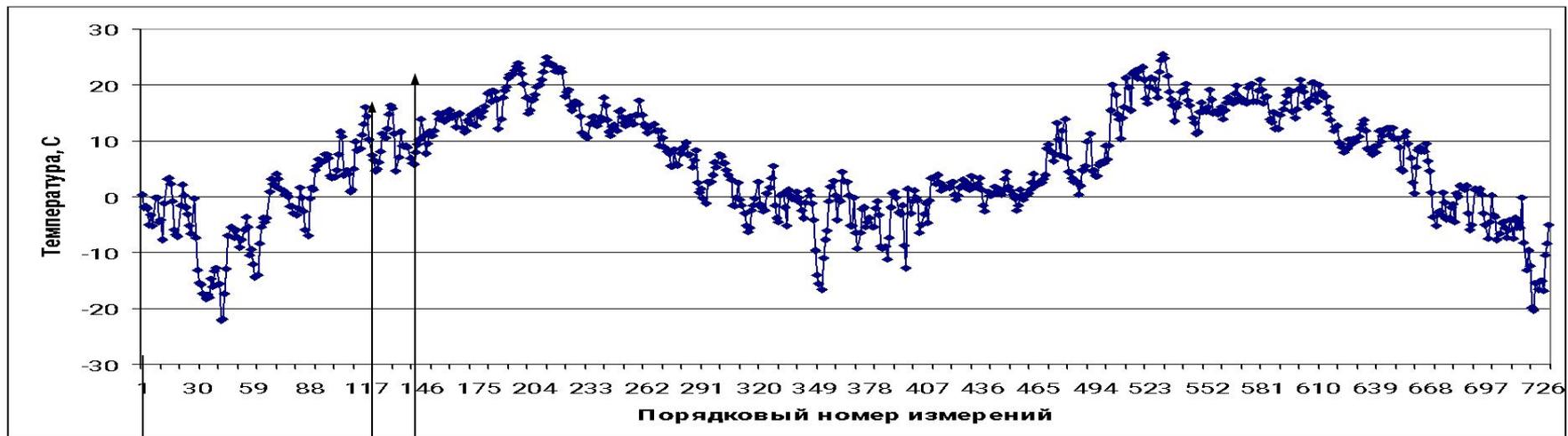
Алгоритм определения значения индикатора, позволяющего выделить «монотонные» по своей тенденции участки временного ряда, заключается в следующем.

1. Задается протяженность **контрольных участков** временного ряда, каждый из которых состоит из n_k членов: параметр **d** .
Значение параметра $n_k = d$ достаточно просто подобрать в процессе проведения нескольких численных экспериментов при апробации предлагаемого алгоритма для рассматриваемого временного ряда.
2. Для первого ($j = 1$) контрольного участка с использованием членов временного ряда с номерами от $n_1 = 1$ до $n_2 = n_k$ вычисляется **среднее значение**:

$$s(n_1, n_2, n_k = d) = s_j, \quad j = 1 .$$

3. задается параметр g - протяженность последовательно анализируемых участков временного ряда, т.е. анализируемых на наличие устойчивой тенденции участков временного ряда.

Именно эти отрезки временного ряда будут использоваться для оценки наличия или отсутствия на этом участке временного тренда по данным для контрольного участка временного ряда.



Контрольный
участок

Контролируемый
на наличие
устойчивой
тенденции к
изменению
участок

Начало работы алгоритма: $n_1 = 1$, $n_2 = 117$, $d = 117$, $g = 30$; определяется значение r_{118}

4. Проверяется выполнение одного из следующих трех условий для всех значений временного ряда, начиная со значения от n_2 до значения $n_2 + g$, для определения на этом промежутке значения индикатора характера протекающего процесса r :

а) **вариант 1** - все значения временного ряда с номерами $i = (n_2 \div n_2 + g)$ больше или равны ранее вычисленного для контрольного участка среднего значения $s(n_1, n_2, d)$: в этом случае параметру r_j с номером $j = n_2 + 1$ присваивается значение $r_j = +1$, что является индикатором того, что на анализируемом участке протяженностью в g значения члены временного ряда в среднем **возрастают**;

б) **вариант 2** - все значения временного ряда с номерами $i = (n_2 \div n_2 + g)$ меньше значению $s(n_1, n_2, d)$: параметр $r_j = -1$ для $j = n_2 + 1$, значения членов временного ряда в среднем **уменьшаются**;

в) **вариант 3** - все значения временного ряда с номерами $i = (n_2 \div n_2 + g)$ не удовлетворяют ни одному из условий а) или б) : параметр $r_j = 0$ для номера $j = n_2 + 1$, значения членов временного ряда в среднем **не меняются**.

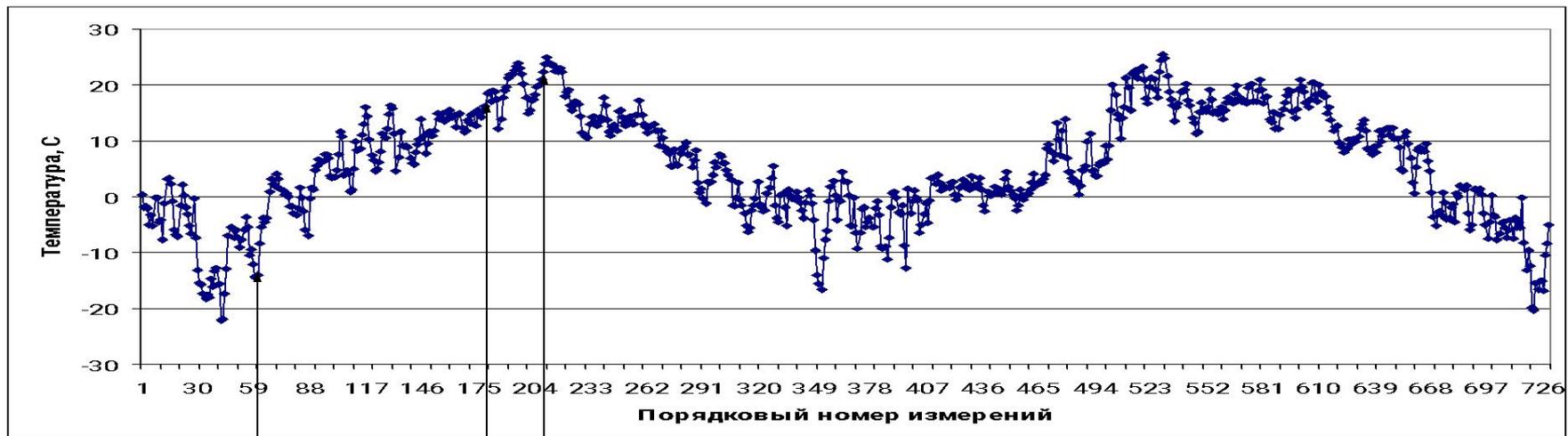
5. Полученное на этапе 4 значение r_j присваивается $n_2 + 1$ элементу вспомогательного ряда R :

$$R(\tau_j) = r_j \quad j = n_2 + 1. \quad (3)$$

6. Значения n_1 и n_2 увеличиваются на 1 и если выполняется условие

$$(n_2 + g) < N, \quad (4)$$

то повторяется выполнение пунктов 4 и 5.



Контрольный
участок

Контролируемый
на наличие
устойчивой
тенденции к
изменению
участок

Иллюстрация работы алгоритма: $n_1 = 59$, $n_2 = 175$, $d = 117$, $g = 30$;
определяется значение r_{176}

Таким образом, в процессе реализации данного алгоритма определяются все значения вспомогательного ряда R с номерами от $i = d + 1$ до $i = N$.

При этом для удобства дальнейшего анализа значениям вспомогательного ряда $R(\tau_i)$ с номерами от $i = 1$ до $i = d$ можно присвоить значения 1.5, что будет индикатором того, что для этих членов временного ряда $P(\tau_i)$ индикатор характера протекающего процесса не определялся.

Следовательно, после реализации приведенного алгоритма значения вспомогательного ряда могут принимать 4 значения:

+1, -1, 0 и 1.5.

Значения +1 соответствуют наличию устойчивой тенденции к увеличению значений временного ряда с того номера, где это значение записано.

Значение -1 будет указывать на наличие тенденции к уменьшению значений временного ряда/

Значение 0 – на относительно стабильное поведение членов временного ряда, а значение 1.5 в начале вспомогательного ряда будет показывать протяженность контрольных участков (на этом участке тенденции не определяются).

В заключении отметим, что рассмотренный алгоритм можно рекомендовать к использованию при анализе временных рядов метеорологических величин различной природы и имеющих различную дискретность. Однако при этом следует учитывать тот факт, что значения входящих в алгоритм параметров d и g (особенно d) не являются универсальными и требуют «настройки» в зависимости от решаемой задачи.

Этот недостаток не является принципиальным, поскольку легко преодолевается путем проведения численных экспериментов с использованием небольшого отрезка исследуемого ряда, в результате которых варьируются указанные параметры и на основе полученных результатов определяются оптимальные значения.

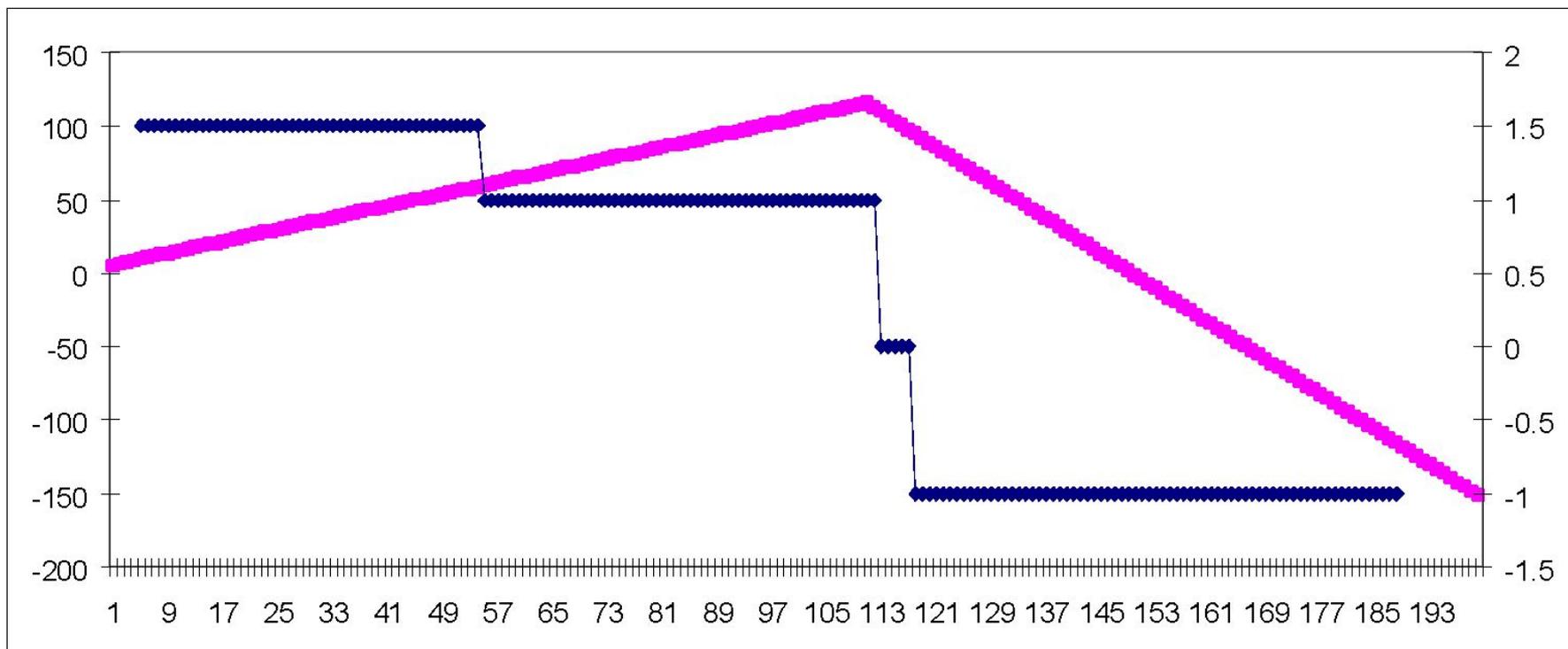
Программа «Процесс-#.xls» .

G	H	I
Перед запуском введите		
в колонку A - исследуемый временной ряд.		
По отрезкам длины ndl вычисляется статистика:ndl =>	50	
по следующим glu = 7 значениям проверяется их расположение относительно среднего значения.		
В колонку B записывается индикатор смены процесса: r		
В колонку C записывается разность двух соседних значений ряда		
$dT(i)=T(i)-T(i+1)$		

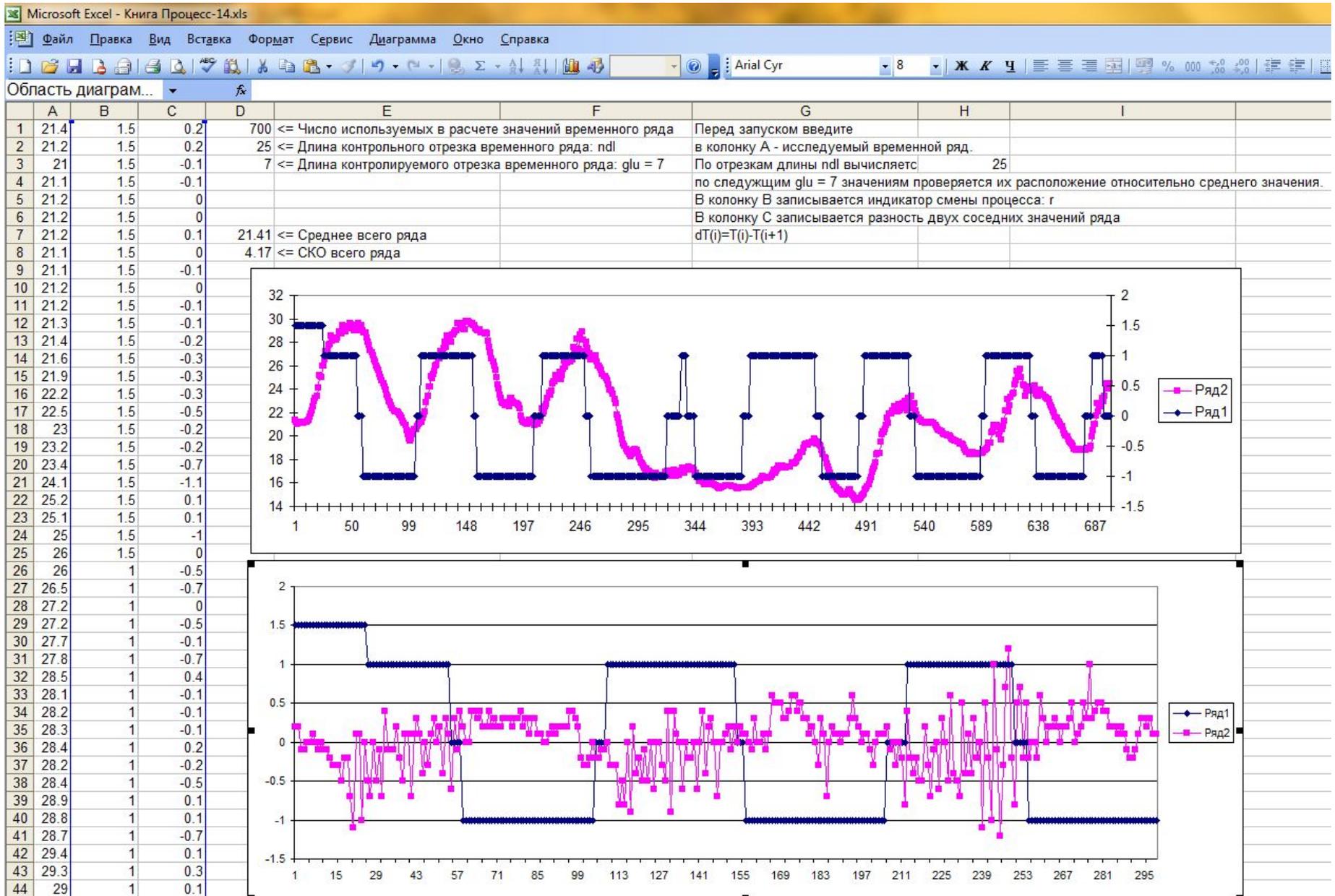
Информационные сообщения программы «Процесс-19»

	A	B	C	D	E	F
1	5			200	<= Общее число значений временного ряда в колонке A	
2	6			5	<= Начальный номер используемого отрезка временного ряда	
3	7			195	<= Конечный номер используемого отрезка временного ряда	
4	8			191	<= Общее число используемых в расчете значений временного ряда	
5	9	1.5	-1	55	<= Процесс определяется начиная с этого значения временного ряда	
6	10	1.5	-1	188	<= Процесс определения заканчивается этим значением временного ряда	
7	11	1.5	-1	50	<= Длина контрольного отрезка временного ряда: ndl	
8	12	1.5	-1	7	<= Глубину контроля от контрольного отрезка ряда равна 7	
9	13	1.5	-1			
10	14	1.5	-1	29.23	<= Среднее используемого отрезка ряда	
11	15	1.5	-1	65.03	<= СКО используемого отрезка ряда	
12	16	1.5	-1			
13	17	1.5	-1			
14	18	1.5	-1			

Результат работы программы «Процесс-19»

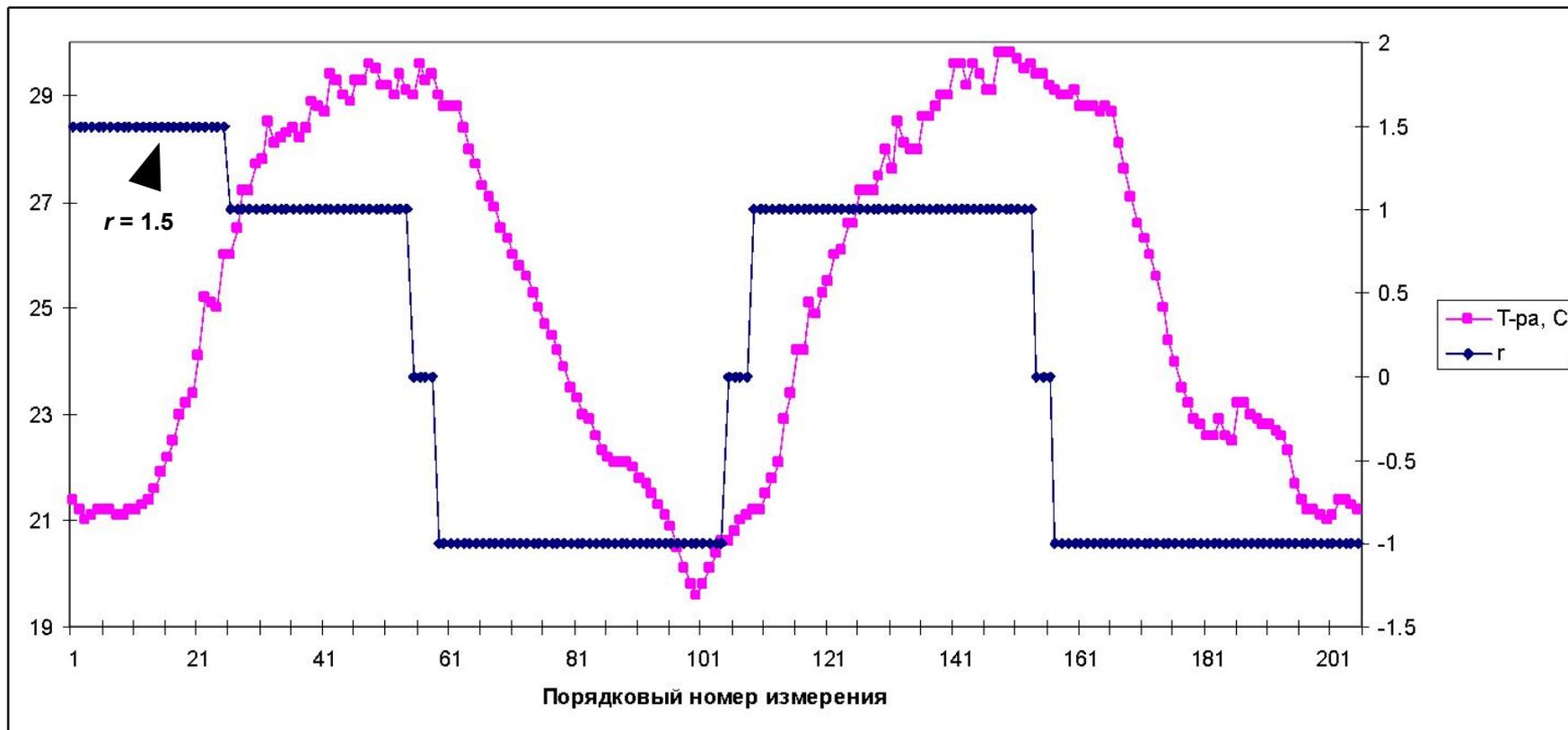


Тестовая проверка программы «Процесс-19»: графики с данными колонок А и В

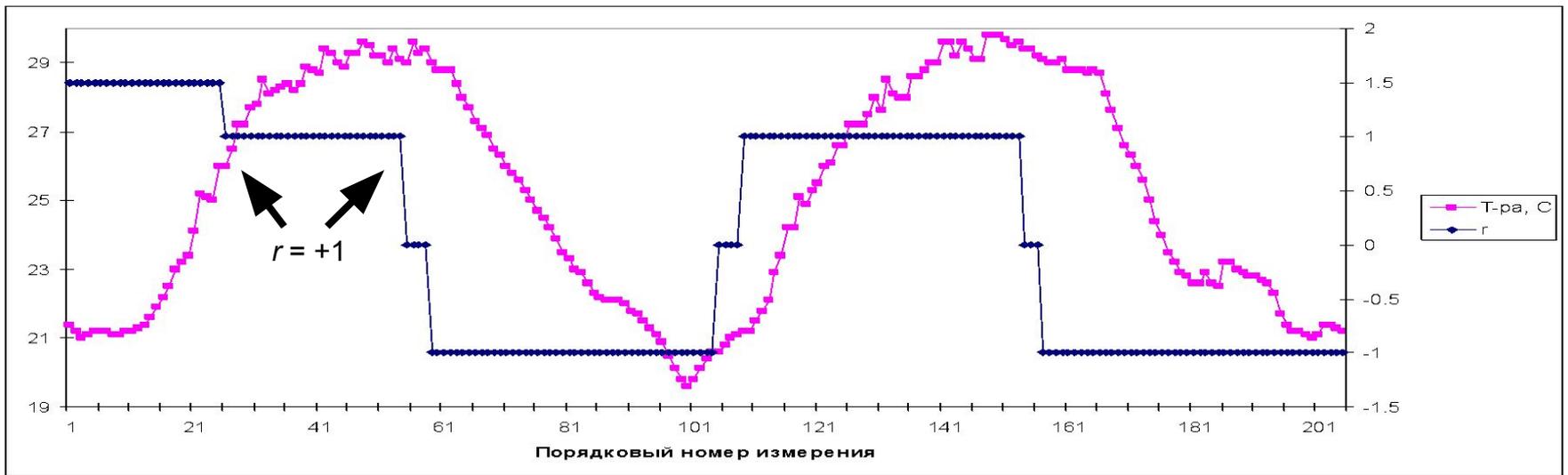


Программа «Процесс-14»: графики с данными колонок A и B, колонок B и C.

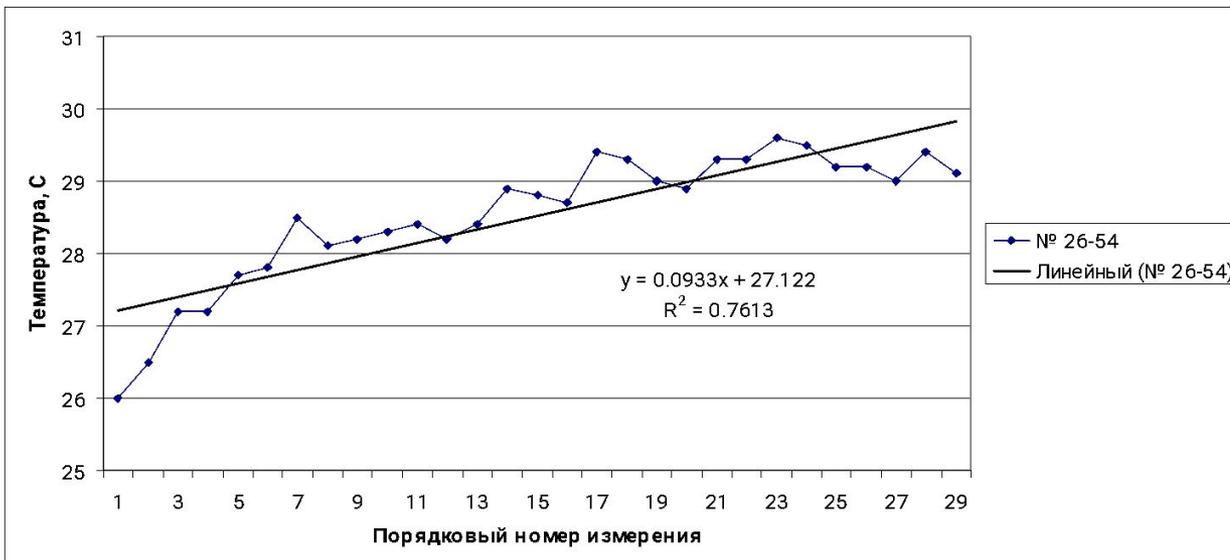
Проиллюстрируем работу данного алгоритма на временном ряде температуры воздуха, полученных с помощью АМС с дискретностью 15 мин в Санкт-Петербурге (*Excel*, «Книга Процесс-14»).



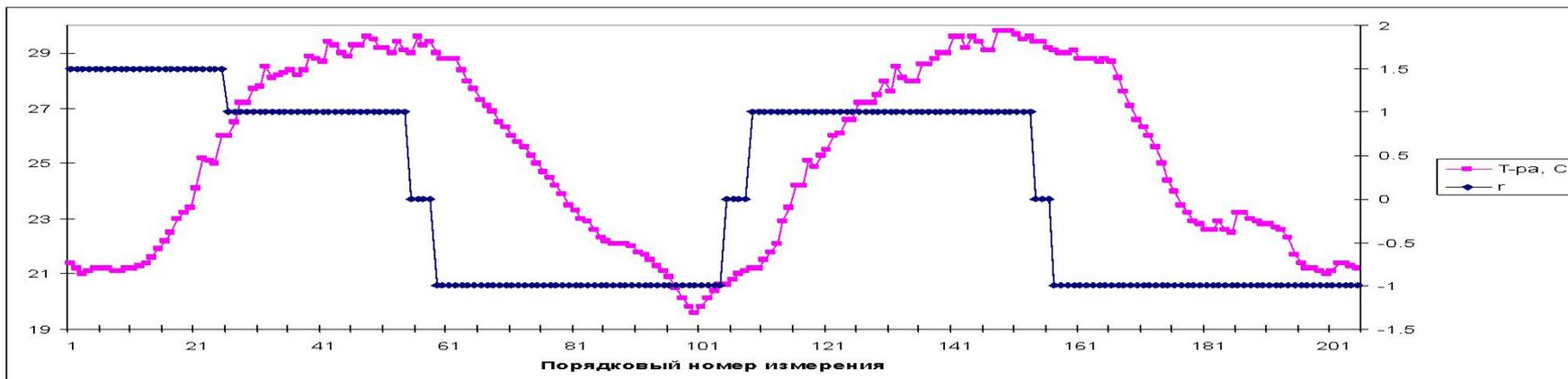
Первые 25 членов этого вспомогательного ряда имеют значения, равные 1.5, т.е. это те значения, для которых по приведенному алгоритму не определялись значения параметра r .



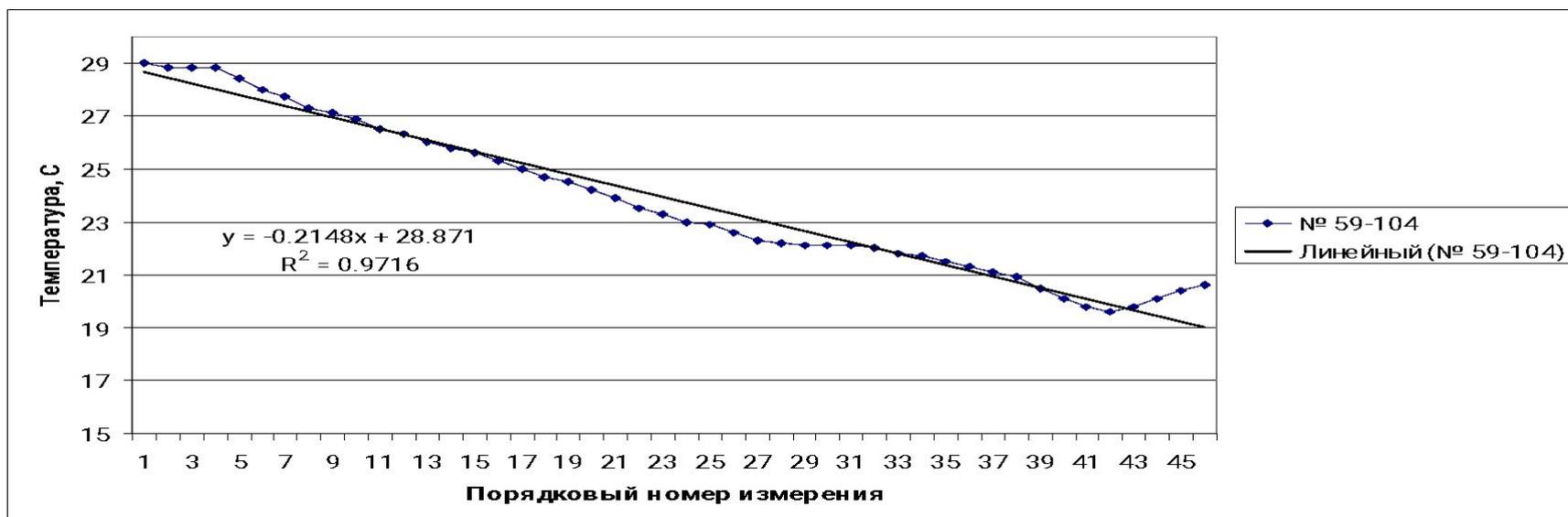
Далее следуют члены ряда с номерами 26 ÷ 54, имеющие значения +1, т.е. на этом участке алгоритм выявил тенденцию к росту температуры воздуха. Считая этот участок квазиоднородным, строим для него линейный тренд, для которого $R^2 = 0.76$ (ранее для всего ряда $R^2 = 0.0008$).

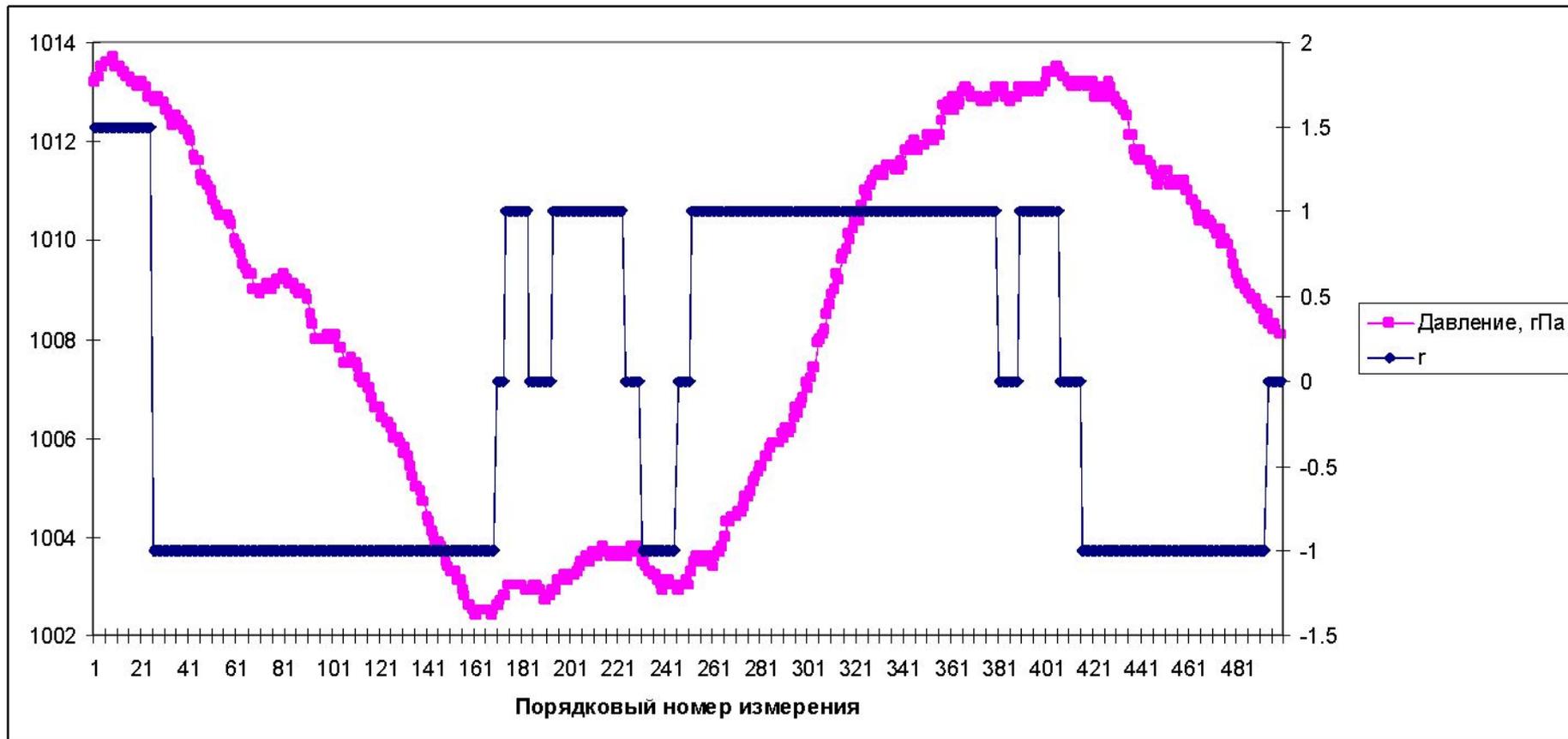


$R^2 = 0.76$



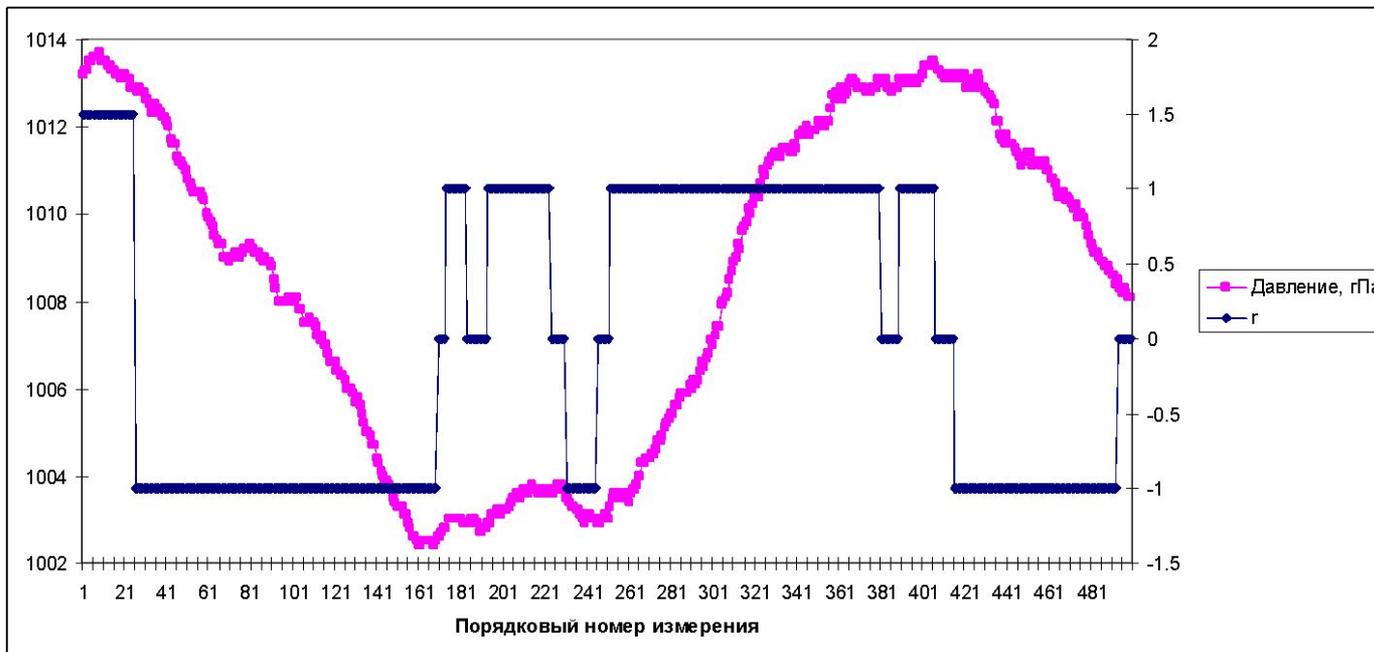
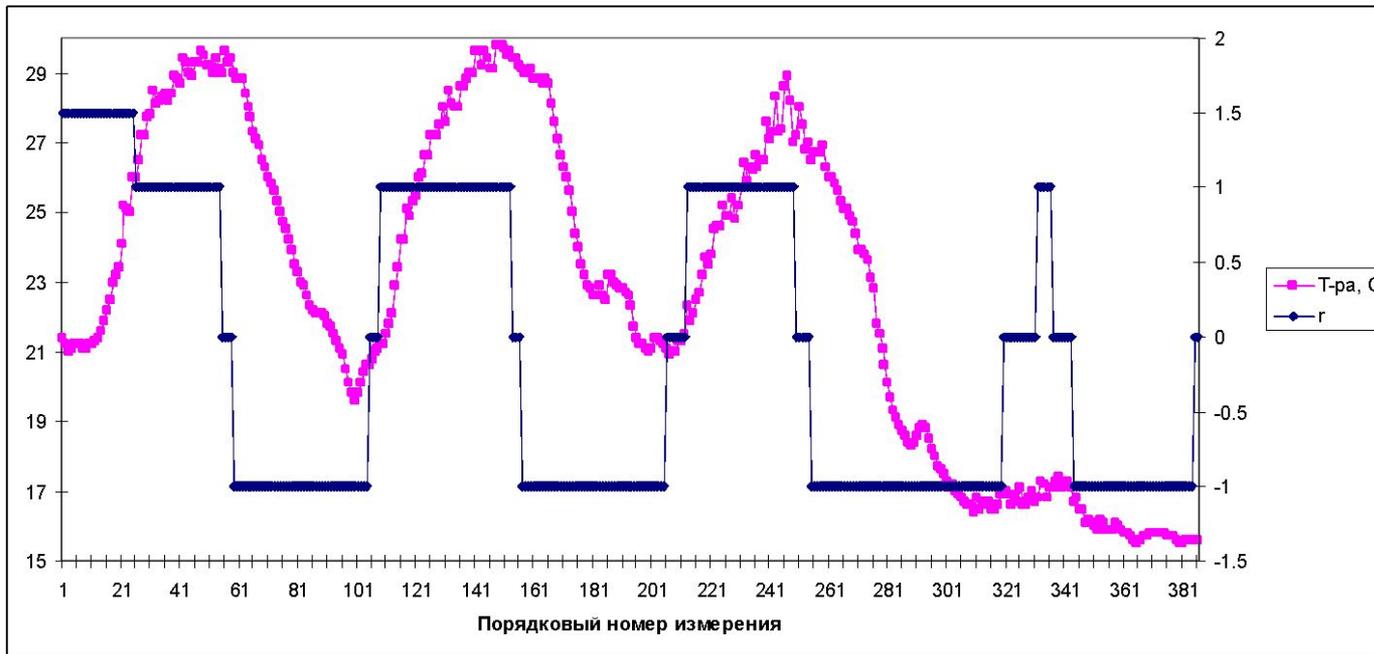
Члены ряда с номерами 55 ÷ 58 равны 0, т.е. алгоритм не выявил на этом участке никаких тенденций, а члены ряда с номерами 59 ÷ 104 имеют значения -1 , что указывает на тенденцию к уменьшению значений температуры воздуха на этом временном промежутке ($R^2 = 0.97$).





Графики временного ряда температуры (красная кривая) и вспомогательного временного ряда r (синие линии).

Для наглядности не произведено объединение **коротких** отрезков значений параметра $r = 0$: такие короткие отрезки при объединении делятся на две равные части и каждой из них приписываются значения индекса r справа и слева от этих частей короткого отрезка.



Этапы анализа временного ряда:

1. Анализ качества временного ряда – отсутствие временных разрывов и «выбросов».
2. Анализ основных статистических характеристик временного ряда.
3. Проверка временного ряда на стационарность.
4. Проверка наличия тренда (глобального и локальных) и определение их параметров.
5. Спектральный анализ.
6. Выделение случайной стационарной составляющей на основе исключения временного тренда и периодических составляющих.
7. Анализ вида распределения на основе построения эмпирической функции распределения (гистограммы).
8. Анализ автокорреляционной функции временного ряда.
9. Построения и анализ математической модели текущего прогнозирования.



Какие будут вопросы?

Ну спросите меня, пожалуйста!