

К.Л. Восканян, А.Д. Кузнецов, О.С. Сероухова, Т.Е. Симакина

**К ВОПРОСУ О ВЫДЕЛЕНИИ АДДИТИВНЫХ СОСТАВЛЯЮЩИХ
ВРЕМЕННОГО РЯДА ПРИЗЕМНОЙ ТЕМПЕРАТУРЫ ВОЗДУХА**

K.L. Voskonyan, A.D. Kuznetsov, O.S. Serouhova, T.E. Simakina

**ON THE ALLOCATION OF TIME SERIES ADDITIVE COMPONENT OF
SURFACE AIR TEMPERATURE**

В данной работе рассматривается задача выделения из временного ряда приземной температуры воздуха, полученного с помощью автоматической метеорологической станции РГГМУ, временного тренда и периодических составляющих.

Ключевые слова: временной ряд, периодическая составляющая, временной тренд, приземная температура воздуха.

This paper considers the problem of extracting a time series of surface air temperature, obtained using an automatic weather station RSHU, a time trend and periodic components

Key words: time series, a periodic component, time trend, surface air temperature.

При решении различных задач, связанных с анализом временных рядов метеорологический величин (например, текущего прогнозирования [1]), неявно предполагается, что рассматриваемый временной ряд является стационарным [2, 3]. Стационарным же, как известно, является тот временной ряд, статистические характеристики и функции распределения которого не меняются во времени (в том числе среднее значение, дисперсия, автокорреляции). При наличии временного тренда и (или) периодических составляющих временной ряд не является стационарным. Однако если осуществить выделение трех его составляющих: временного тренда R , периодических составляющих P и шума N , то временной ряд только шумовой составляющей может быть уже близок к стационарному. Для выделения шумовой составляющей представим временной ряд в следующем виде:

$$T_i = R_i + P_i + W_i, \quad (1)$$

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

В случае регулярного временного ряда порядковый номер связан с моментом проведения измерения следующим соотношением:

$$\tau_i = \tau_0 + \Delta_\tau \cdot i, \quad (2)$$

где τ_0 – начальный момент проведения измерений; Δ_τ – дискретность измерений (временной интервал между измерениями), $i = 1, 2, \dots, N$.

Задача представления временного ряда в форме (1) может решаться в широком временном интервале: от тысяч и миллионов лет при рассмотрении изменения климата Земли до нескольких часов и даже минут при текущем прогнозировании атмосферных процессов и явлений. В данной работе мы ограничимся рассмотрением весьма коротких временных периодов: от одного месяца до нескольких десятков минут.

Представление временного ряда в виде (1) может быть использовано как при создании модельных временных рядов метеорологических величин, так и в процессе выделения аддитивных составляющих в процессе анализа реальных временных рядов. И в том, и в другом случаях соответствующая задача может быть сведена к нахождению аналитического вида функций, описывающих временную изменчивость соответствующих слагаемых, т.е. задача представления временного ряда в виде (1) может быть сведена к нахождению трех функций: $F_R(\tau)$, $F_P(\tau)$ и $F_W(\tau)$, которые на заданном временном промежутке $[a, b]$ описывали бы поведения трех составляющих ряда:

$$R(\tau) = F_R(\tau), \quad P(\tau) = F_P(\tau), \quad W(\tau) = F_W(\tau), \quad (3)$$

где

$$a < \tau < b. \quad (4)$$

В случае анализа реальных временных рядов временной ряд для шумовой составляющей $W(\tau)$ может быть получен после определения функций $F_R(\tau)$ и $F_P(\tau)$ и использования их для решения уравнения (1) относительно $W(\tau)$.

Для моделирования временной составляющей могут быть использованы различные предположения о статистических свойствах ее составляющих. В случае предположения о том, что члены этой части временного ряда имеют нормальное распределения с известным значением среднего LW (LW можно интерпретировать как систематическая погрешность измерений) и среднеквадратичного отклонения σ_w , то для расчета значений W_i могут быть использованы псевдослучайные числа, моделируемые на ПЭВМ. При использовании языка программирования «Бейсик» псевдослучайные числа, равномерно распределенные на промежутке $[0, 1]$, задаются оператором *rnd*, что позволяет использовать следующую расчетную схему для задания функции $F_W(\tau)$:

$$F_W(\tau_i) = L_w + \sigma_w \cdot [2 \cdot \ln(1 / rnd_i)]^{1/2} \cdot \cos(2\pi \cdot rnd_{i+1}), \quad (5)$$

где время измерения τ_i и порядковый номер измерения i связаны соотношением (2), rnd_i и rnd_{i+1} являются двумя последовательными псевдослучайными числами, генерируемыми ПЭВМ [4].

Современные автоматические метеостанции позволяют чрезвычайно детально описать поведения временных рядов регистрируемых ими метеорологических величин за счет высокой дискретности Δ проведения измерений. Используемые в данной работе данные были получены с дискретностью 10 с. Однако, при такой высокой частоте измерений появляется большое количество избыточных данных. Проведенная оценка оптимальной величины дискретности измерений Δ показала, что для временных рядов приземной температуры воздуха она составляет 10–15 мин. При такой дискретности

основные статистические характеристики временного ряда практически совпадают с аналогичными значениями исходного ряда и, в тоже время, существенно сокращается объем используемых выборок.

В данной статье рассматриваются результаты решения задачи выделения аддитивных составляющих для временных рядов приземной температуры воздуха, полученных с помощью автоматической метеорологической станции РГГМУ.

Выделение временного тренда. Для аппроксимации временного тренда наиболее простым и чаще всего используемым подходом является его полиномиальное представление, когда соответствующая составляющая временного ряда – функция $F_R(\tau)$, является полиномом, степень которого k задается эмпирически:

$$F_R(\tau_i) = b_0 + b_1 \cdot \tau_i + \dots + b_n \cdot \tau_i^k. \quad (6)$$

Входящие в соотношение (6) коэффициенты аппроксимационного полинома степени k : b_i , могут быть найдены, например, методом наименьших квадратов по всему используемому ансамблю данных, содержащимся в заданном временном промежутке $[a, b]$, [2, 5]. Рассмотрим результаты выделения временного тренда с использованием модельных и реальных данных

Модельные данные. В качестве примера рассмотрим временной ряд приземной температуры воздуха с дискретностью 0,25 ч (15 мин), моделируемый следующим соотношением:

$$T_i = 10 + 0,06 \cdot i \cdot 0,25 + 5 \cdot \sin(2\pi \cdot i \cdot 0,25 / 24). \quad (7)$$

где i – порядковый номер измерения; второе слагаемое – линейный временной тренд, имитирующий погодные изменения температуры (прогрев воздушной массы) (температура повышается на 0,06 °C за каждые 0,25 ч: $\Delta_t = 0,25$ ч); третье слагаемое – периодическая составляющая с амплитудой в 5 °C и периодом в 24 ч.

Результаты аппроксимации временного тренда полиномами первой, второй и третьей степеней для модельного ряда, определяемого формулой (7), иллюстрирует рис. 1.

Из анализа характера поведения представленных здесь кривых, описывающих временной тренд, видно, что начиная с полинома третьей степени характер поведения аппроксимационной кривой в этом простейшем случае уже меняется мало. Однако анализ показывает, что в случае экстраполяции тренда использование полиномов степени выше первого будет наблюдаться заметное искажение временного тренда.

Реальные данные. Для реальных временных рядов приземной температуры воздуха, как это следует из анализа представленных в качестве типичного примера на рис. 2 кривых, с увеличением степени аппроксимационного полинома применение полиномов степени выше первой начинает описывать не только сам тренд, но и периодические составляющие. Поскольку анализ и исключение периодических составляющих в нашем случае рассматривается как отдельная задача, то представляется не целесообразно для описания временного тренда реальных временных рядов приземных значений температуры воздуха использовать полиномы выше первой степени.

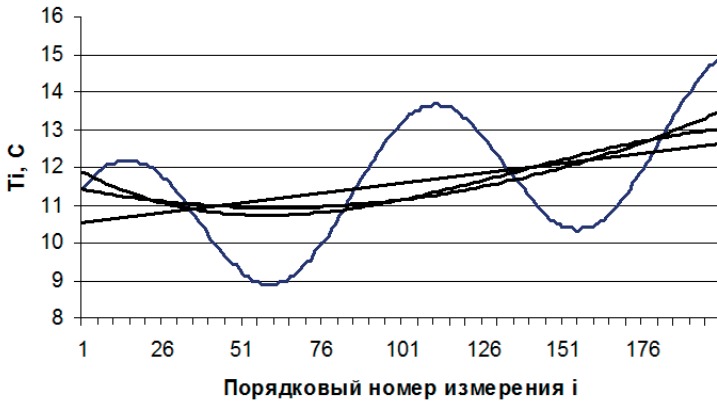


Рис. 1. Результаты аппроксимации временного тренда полиномами первой, второй и третьей степеней для фрагмента модельного временного ряда приземной температуры.

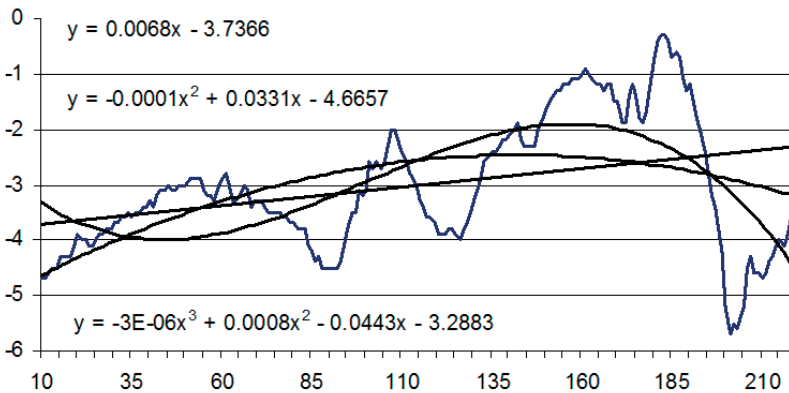


Рис. 2. Результаты аппроксимации временного тренда полиномами первой, второй и третьей степеней для фрагмента временного ряда приземной температуры, измеренной с дискретностью 15 мин.

Рис. 3 иллюстрирует изменение характера поведения временного ряда до и после исключения временного тренда. При этом величина среднеквадратического отклонения для обоих рядов практически не меняется, а величина среднего значения после исключения временного ряда становится близкой к нулю.

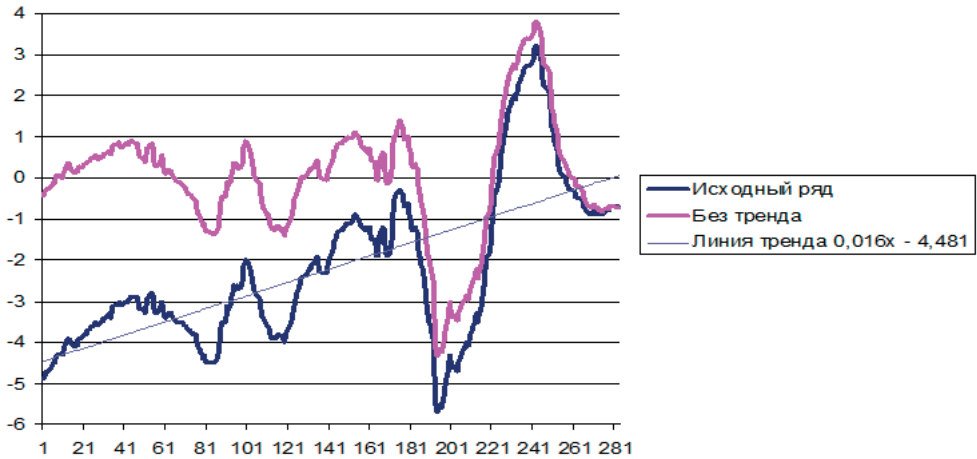


Рис. 3. Фрагмент временного ряда приземной температуры с дискретностью $\Delta t = 15$ мин (нижняя кривая) и тот же временной ряд после исключения линейного тренда (верхняя кривая); прямая линия — линия временного тренда.

Выделение периодических составляющих. Стандартным методом выделения периодических составляющей временного ряда является разложение его в ряд Фурье (анализ Фурье). Для этого могут быть использованы различные статистические пакеты, например, пакет *Statistica* [5]. Однако в том случае, когда процесс выделения периодических составляющих является промежуточным этапом вычислений при решении других задач удобнее воспользоваться упрощенным алгоритмом, который было бы легко реализовать в процессе написания основной программы. При этом следует учесть и простоту реализации прямого и обратного «переходов»: исключения периодических составляющих из временного ряда и, в случае необходимости, добавления этих же периодических составляющих. С учетом этих требований вместо полного преобразования Фурье нами был реализован следующий упрощенный алгоритм:

1. Задавалось общее количество периодических составляющих: M .
2. Каждая m -я периодическая составляющая задавалась следующим соотношением:

$$F_p^m(\tau_i) = A_m + K_m \sin(2 \cdot \pi \cdot \tau_i / D_m + F_m). \quad (8)$$

где для порядкового номера периодической составляющей $m = 1, 2, \dots, M$; а для порядкового номера члена временного ряда $i = 1, 2, \dots, N$.

Тогда для каждого момента времени τ можно записать следующее соотношение для периодической составляющей временного ряда

$$F_p(\tau) = \sum F_p^m(\tau), \quad (9)$$

где суммирование идет по всем значениям m : $m = 1, 2, \dots, M$.

3. Коэффициенты, входящие в выражение (8), последовательно для каждой периодической составляющей с номером m находились методом прямого перебора их значений. При реализации такого перебора для каждого коэффициента K_m , D_m и F_m задается множество значений, лежащих в заданном диапазоне (между заданными максимальным и минимальным значениями с дискретностью Δ):

$$K_m^j = K_m^{\min} + \Delta_K \cdot j, \quad j = 0, 1, \dots, (K_m^{\max} - K_m^{\min}) / \Delta_K, \quad (10)$$

$$D_m^r = D_m^{\min} + \Delta_D \cdot r, \quad r = 0, 1, \dots, (D_m^{\max} - D_m^{\min}) / \Delta_D, \quad (11)$$

$$F_m^l = F_m^{\min} + \Delta_F \cdot l, \quad l = 0, 1, \dots, (F_m^{\max} - F_m^{\min}) / \Delta_F, \quad (12)$$

где $m = 1, 2, \dots, M$.

Для реализации расчетов с использованием соотношений (10)–(12) необходимо задать общее число периодических составляющих: M . При задании числа M мы исходили из предположения, что с учетом относительно малой длины отрезков исследуемых рядов у них имеется не более 3 значимых гармоник ($M < 4$).

Поскольку нахождение периодических составляющих производилось после исключения временного тренда, то значения всех коэффициентов A_m в формуле (8) задавались равным нулю. В противном случае, как и для параметров K и F , можно осуществить перебор значений коэффициентов A в заданном диапазоне их значений

$$A_m^g = A_m^{\min} + \Delta_A \cdot g, \quad g = 0, 1, \dots, (A_m^{\max} - A_m^{\min}) / \Delta_A, \quad (13)$$

Из большого числа генерируем на основе соотношений (10)–(12) коэффициентов для каждой периодической составляющей выбирались те из них, которые минимизировали значение суммарной невязки S :

$$S = \sum \left[(T_i - R_i) - K_1^j \sin(2 \cdot \pi \cdot i / D_1^l + F_1^l) \right]^2, \quad (14)$$

где значения R_i рассчитываются по формуле (6).

В формуле (14) при определении коэффициентов суммирование производилось по всем значениям временного ряда: $i = 1, 2, \dots, N$, и учтено условие $A_1 = 0$. Таким образом, из всего многообразия рассмотренных коэффициентов для первой периодической составляющей определялись три значения: $K_1^{\text{опт}}$, $D_1^{\text{опт}}$, $F_1^{\text{опт}}$, на основе минимизации величины S после перебора всех значений коэффициентов в диапазонах, указанных в формулах (10)–(12).

После этого из ряда исключалась первая периодическая составляющая

$$T_i^1 = T_i - K_1^{\text{опт}} \sin(2 \cdot \pi \cdot i / D_1^{\text{опт}} + F_1^{\text{опт}}). \quad (15)$$

Полученный таким образом новый ряд температуры T_i^1 использовался для расчета величины S по формуле (14) (ряд T_i заменялся на ряд T_i^1), определялось, при каких коэффициентах эта величина минимальна и эти коэффициенты $K_2^{\text{опт}}$, $D_2^{\text{опт}}$, $F_2^{\text{опт}}$ определяли вторую периодическую составляющую. Эта вторая периодическая составляющая исключалась из ряда T_i^1 для формирования ряда T_i^2 . Затем тот же процесс, но уже с использованием ряда T_i^2 повторялся для нахождения третьей периодической составляющей.

4. Для реализации описанного в пункте 3 алгоритма, необходимо задать диапазон изменения коэффициентов в формулах (10)–(12) и дискретность такого изменения.

Для определения значений K^{max} и K^{min} на каждом шаге реализации описанного в пункте 3 алгоритма вычислялись среднее значение ряда и среднеквадратичное отклонение ряда от этого среднего значения: W . На основе вычисленного таким образом значения W задавались $K^{\text{max}} = W$, $K^{\text{min}} = -W$. Величина дискретности ΔK принималась равной $0,1$ °C.

Для остальных значений диапазона изменений коэффициентов использовались следующие значения:

$$D^{\text{min}} = 0,5 \text{ ч}, \quad D^{\text{max}} = 72 \text{ ч}, \quad \Delta_D = 0,5 \text{ ч}, \\ F^{\text{min}} = -\pi, \quad F^{\text{max}} = \pi, \quad \Delta_F = 0,1 \text{ рад}.$$

Для тестирования данного алгоритма также использовались модельные и реальные данные.

Модельные данные. Для тестирования алгоритма выделения временного тренда использовались различные модельные ряды, содержащий как тренд, так и периодически составляющие. Здесь в качестве примера рассмотрим модельный ряд, заданный следующим соотношением (шумовая составляющая отсутствует):

$$T_i = F_R(i) + F_P^1(i) + F_P^2(i) + F_P^3(i), \quad (16)$$

где линейный временной тренд определяется как

$$F_R(i) = 10 + 0,06 \cdot i, \quad (17)$$

а три периодических составляющих определялись соответственно следующими уравнениями

$$F_P^1(i) = 3 \cdot \sin(2 \cdot \pi \cdot i \cdot 0,25 / 24 + 1,6), \\ F_P^2(i) = -2 \cdot \sin(2 \cdot \pi \cdot i \cdot 0,25 / 6 + 0,5), \\ F_P^3(i) = 0,5 \cdot \sin(2 \cdot \pi \cdot i \cdot 0,25 / 2 - 1,2). \quad (18)$$

В формулах (16)–(18) индекс i (порядковый номер измерения) менялся от 1 до 200. Фрагмент этого ряда ($i = 1 - 50$) представлен на рис. 4.

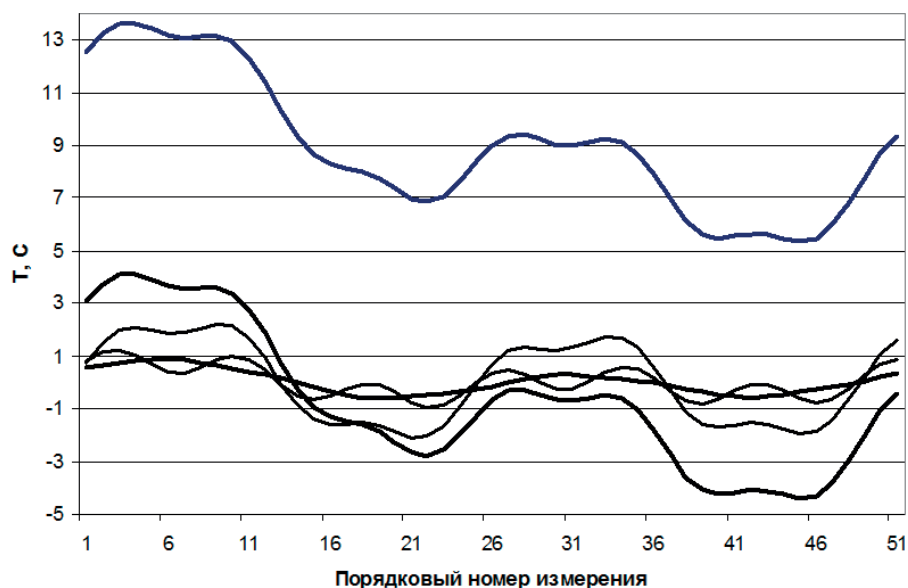


Рис. 4. Фрагмент модельного временного ряда (верхняя кривая), заданного соотношениями (16)–(18), и временные ряды после исключения линейного тренда и первой, второй и третьей периодических составляющих.

В результате расчета параметров временной линейной регрессии для модельного ряда были получены следующие значения коэффициентов полинома первой степени (с скобок указаны точные значения)

$$b_0 = 9,4(10,0), \quad b_1 = 0,06(0,06).$$

Полученные значения практически полностью совпадают с коэффициентами в формуле (17).

Значения коэффициентов периодических составляющих, вычисленные по указанной выше методика для рассматриваемого ряда представлены в табл. 1 (в скобках указаны точные значения).

Таблица 1

Параметры трех периодических составляющих модельного временного ряда приземной температуры воздуха

Номер периодической составляющей	Период, час	Амплитуда, °C	Фаза, рад
1	23,5 (24)	2,6 (3)	1,4 (1.6)
2	6,0 (6)	1,5 (–2)	5,7 (–2,6)
3	2,0 (2)	0,6 (0,5)	–1,1 (–1.2)

Анализ данных табл. 1 показывает, что предложенный алгоритм практически точно позволяет определить коэффициенты периодических составляющих, содержащиеся в модельном временном ряде (с учетом фазового сдвига знак амплитуды второй составляющей соответствует знаку точного значения).

Реальные данные. Рассмотрим пример, иллюстрирующий использование рассмотренных выше алгоритмов нахождения периодических составляющих для фрагмента временного ряда приземной температуры воздуха. Используемая для этого компьютерная программа позволяет после исключения линейного временного тренда для заданного фрагмента временного ряда последовательно находить синусоидальные составляющие в форме (8) и рассчитывает статистические характеристики остатков. Пример работы этой программы иллюстрируют рис. 5–6 и данные табл. 2. Причем на рис. 5–6 приводится лишь начальный фрагмент ряда ($i = 10 - 110$), тогда как общая длина рассматриваемого ряда $N = 2960$.

На рис. 5а представлен фрагмент временного ряда приземной температуры (дискретность 15 мин), а на рис. 5б – тот же фрагмент после исключения линейного тренда. Здесь, прежде всего, необходимо обратить внимание на изменении характера колебания температуры: в первом случае она колеблется около среднего значения в $-1,4\text{ }^{\circ}\text{C}$, а во втором около $0\text{ }^{\circ}\text{C}$. Кажущееся несоответствие указанного среднего значения и поведения кривой на рис. 5а связано с тем, что приведенные значения соответствуют всему ряду, тогда как при рассмотрении только значений $i = 1\div 110$ температура колеблется около среднего значения $-3,5\text{ }^{\circ}\text{C}$.

На рис. 5б этот же фрагмент представлен после исключения линейного тренда. Теперь амплитуда изменения приземной температуры в рассмотренном фрагменте уменьшилась от $-4,7\div -2,6$ до $-1,3\div 0,9$, однако, остались периодические составляющие. Причем хорошо видно, что в данном фрагменте периодичность в диапазоне $i = 10\div 90$ и диапазоне $i = 90\div 110$ различна.

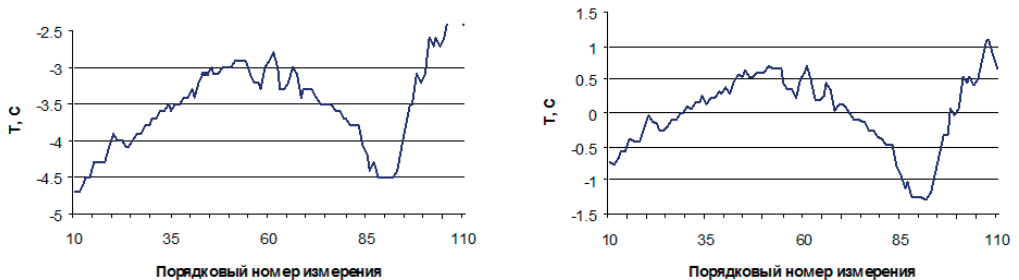


Рис. 5. Фрагмент временного ряда (дискретность 15 мин) и тот же фрагмент после исключения линейного тренда.

Рис. 6 иллюстрирует те изменения, к которым приводит исключение как линейного временного тренда, так и трех периодических составляющих с периодами соответственно 15, 8 и 6 ч. С учетом величины амплитуды изменения получившегося «остатка» ряда, он может рассматриваться как шумовая составляющая, обусловленная, прежде всего, погрешностью измерений.



Рис. 6. Фрагмент временного ряда после исключения линейного тренда, первой, второй и третьей периодических составляющих.

Данные табл. 2 показывают изменение статистических характеристик всего временного ряда в процессе его обработки: исключения тренда и периодических составляющих. Показательным является уменьшение величины среднеквадратического отклонения с $3.05\text{ }^{\circ}\text{C}$ до $0.17\text{ }^{\circ}\text{C}$.

Таблица 2

Изменение статистических характеристик временного ряда ($N = 2960$) в процессе исключения временного тренда и периодических составляющих

Параметры	Среднее значение, $^{\circ}\text{C}$	Среднеквадратическое отклонение, $^{\circ}\text{C}$
Исходный фрагмент ряда	-1,4	3,0
После исключения тренда	$5,4 \cdot 10^{-2}$	0,6
После исключения тренда и первой периодической составляющей	$6,8 \cdot 10^{-2}$	0,3
После исключения тренда и двух периодических составляющих	$4,8 \cdot 10^{-2}$	0,2
После исключения тренда и трех периодических составляющих	$2,8 \cdot 10^{-2}$	0,1

Эффективность исключения тренда и периодических составляющих можно оценить, рассматривая изменения внутрирядового коэффициента корреляции значения приземной температуры. На рис. 7 представлены значения коэффициентов корреляции при временном сдвиге от 15 до 120 мин для временного ряда до (пунктирная линия) и после (сплошная линия) исключения линейного тренда и периодических составляющих. Выполненное исключение резко снизило статистическую зависимость значений временного ряда, т.е. после обработки в основном осталась лишь случайная составляющая.

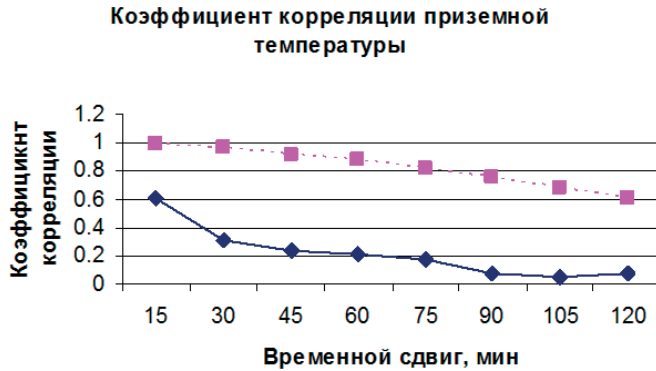


Рис. 7. График коэффициентов корреляции приземных значений температуры в зависимости от временного сдвига до (пунктирная линия) и после (сплошная линия) исключения линейного тренда и периодических составляющих.

Литература

1. *Алдошкина Е.С., Кузнецов А.Д., Пугачев А.А., Сероухова О.С., Симакина Т.Е., Чукин В.В.* Опыт использования аппарата нейронных сетей для анализа и прогноза временного ряда температуры воздуха. // Ученые записки РГГМУ, 2009, № 11, с. 91–110.
2. *Васильев А.В., Мельникова И.Н.* Методы прикладного анализа результатов натурных измерений в окружающей среде. Учебное пособие. — СПб.: Изд-во БГТУ «Военмех», 2009. — 370 с.
3. *Малинин В.Н.* Статистические методы анализа гидрометеорологической информации. — СПб., 2008. — 407 с.
4. *Дьяконов В.П.* Справочник по алгоритмам и программам на языке бейсик для персональных ЭВМ. — М.: «Наука», главная редакция физико-математической литературы, 1987. — 246 с.
5. STATISTICA. Искусство анализа данных на компьютере. 2-е издание. — М.: Питер, 2003. — 63 с.