

А. Д. Кузнецов, О. С. Сероухова, Д. С. Ефременко

**ОБ ОДНОМ АЛГОРИТМЕ ВЫЯВЛЕНИЯ ЛОКАЛЬНЫХ ТРЕНДОВ
ПРИ АНАЛИЗЕ МЕТЕОРОЛОГИЧЕСКИХ ВРЕМЕННЫХ РЯДОВ**

A. D. Kuznetsov, O. S. Serouhova, D. S. Efremenko

**AN ALGORITHM IDENTIFY LOCAL TRENDS IN THE PROCESS
OF ANALYSIS OF TIME SERIES OF METEOROLOGICAL VARIABLES**

В данной работе анализируется методика автоматизации процесса выявления во временных рядах метеорологических величин квазиоднородных участков с устойчивой тенденцией для последующего расчета параметров временного тренда. Для апробации предложенной методики используются временные ряды температуры воздуха и атмосферного давления.

Ключевые слова: *временной ряд, временной тренд, приземная температура воздуха, атмосферное давление.*

In this paper we analyze the technique of automating the process of identifying in time series of meteorological variables quasi-homogeneous regions with a stable trend for the subsequent calculation of the parameters of a time trend. To test the proposed method uses time series of air temperature and atmospheric pressure.

Keywords: *time series, time trend, surface air temperature and atmosphere pressure.*

Выявление тренда во временных рядах метеорологических величин является важным и весьма распространенным этапом их анализа. Его реализация при рассмотрении априори заданного временного отрезка ряда является тривиальной алгоритмической задачей и требует лишь задания вида функции, описывающей такой «глобальный» временной тренд. Однако при необходимости выделения нескольких временных отрезков, отражающих локальные изменения тенденций, определение границ таких локальных участков с целью последующего расчета в них временного тренда уже не является тривиальной задачей [1, 2].

Одним из подходов к решению этой задачи является визуализация временного ряда для выделения в нем локальных отрезков с устойчивыми тенденциями. Однако такой подход весьма субъективен, а при значительной длительности временного ряда не всегда приемлем.

В данной статье исследуются возможности известного и весьма простого подхода к автоматизации решения задачи определения параметров локальных трендов внутри временного ряда [3].

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

$$T_i = (\tau_i), \quad (1)$$

где i — порядковый номер члена временного ряда: $i = 1, 2, \dots, N$; $\tau_i = \tau_0 + \Delta\tau \cdot i$; 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, состоящих из nk членов. Значение параметра nk 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 $n_2 = nk$ вычисляется среднее значение: $s(n_1, n_2, nk) = s_j, j = 1$.
3. Задается параметр np — протяженность 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.
4. Проверяется выполнение одного из следующих трех условий для всех значений временного ряда, начиная со значения от n_2 до значения $n_2 + np$, для определения на этом промежутке значения индикатора характера протекающего процесса r :
 - а) все значения временного ряда с номерами $i = (n_2 \div n_2 + np)$ больше или равны $s(n_1, n_2, nk), j = n_2 + 1$: параметр $r_j = +1$, значения членов временного ряда в среднем растут;
 - б) все значения временного ряда с номерами $i = (n_2 \div n_2 + np)$ меньше значению $s(n_1, n_2, nk), j = n_2 + 1$: параметр $r_j = -1$, значения членов временного ряда в среднем уменьшаются;
 - в) все значения временного ряда с номерами $i = (n_2 \div n_2 + np)$ не удовлетворяют ни одному из условий а) или б), $j = n_2 + 1$: параметр $r_j = 0$, 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

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

то повторяется выполнение пунктов 4 и 5.

Таким образом, в процессе реализации данного алгоритма определяются все значения вспомогательного ряда R с номерами от $i = nk + 1$ до $i = N$. При этом для удобства дальнейшего анализа значениям вспомогательного ряда $R(\tau_i)$ с номерами от $i = 1$ до $i = nk$ можно присвоить значения 1.5, что будет индикатором того, что для этих членов временного ряда $T(\tau_i)$ индикатор характера протекающего процесса не определялся.

Следовательно, после реализации приведенного алгоритма значения вспомогательного ряда могут принимать четыре 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).

Проиллюстрируем работу данного алгоритма на двух примерах: временных рядах температуры воздуха, полученных с помощью АМС с дискретностью 15 мин, и среднесуточной температуры воздуха в Санкт-Петербурге.

На рис. 1 представлен график временного ряда температуры воздуха, состоящий из 205 значений, измеренных автоматической метеорологической станцией (АМС) РГГМУ и приведенных к дискретности 15 мин. На этом же графике представлены график и уравнение, определяющие «глобальный» линейный временной тренд, рассчитанный по всем 205 значениям ряда. Хорошо видно, что полученное уравнение не отражает временные изменения тенденций временного хода температуры воздуха — ничтожно мал уровень достоверности линейной аппроксимации: $R^2 = 0,0008$.

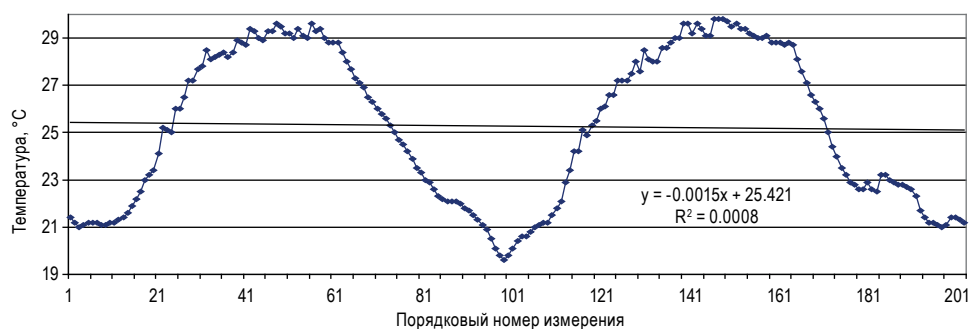


Рис. 1. Графики временного ряда температуры воздуха, измеренной АМС РГГМУ с 01.07.2011 по 02.07.2011 (дискретность 15 мин) и «глобального» линейного временного тренда, рассчитанного по 205 значениям ряда

На рис. 2 представлен блок графиков того же временного ряда температуры воздуха (на всех графиках — левая ось ординат) и вспомогательного ряда R (правая ось ординат), значения которых были рассчитаны по приведенному выше алгоритму при следующих параметрах: $nk = 25$, значения np меняются от 9 до 3.

Проанализируем поведение графика для ряда R при значении параметра np , равного 7 (см. рис. 2, б). 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 $26 \div 54$, имеющие значения +1, т.е. на этом участке алгоритм выявил тенденцию к росту температуры воздуха. Здесь следует отметить, что в этой части ряда алгоритм не совсем точно определил положения правой границы интервала: падение температуры заканчивается на 100-ом члене ряда, а алгоритм определил границу на 102-ом члене.

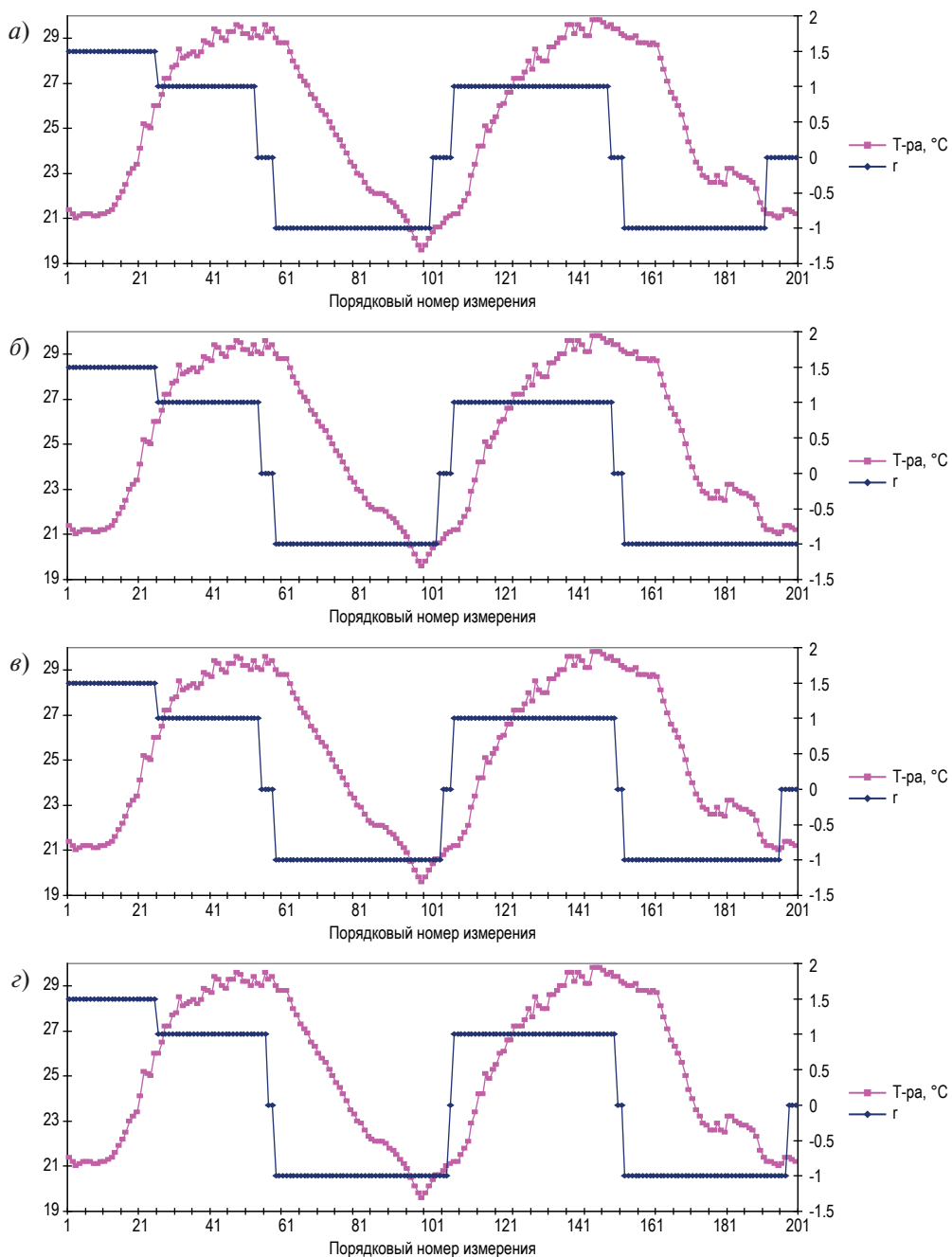


Рис. 2. Графики временного ряда температуры воздуха (левая ось ординат) и временного ряда функции R (правая ось ординат) при значениях:
а — $nk = 25, np = 9$; б — $nk = 25, np = 7$; в — $nk = 25, np = 5$; г — $nk = 25, np = 3$

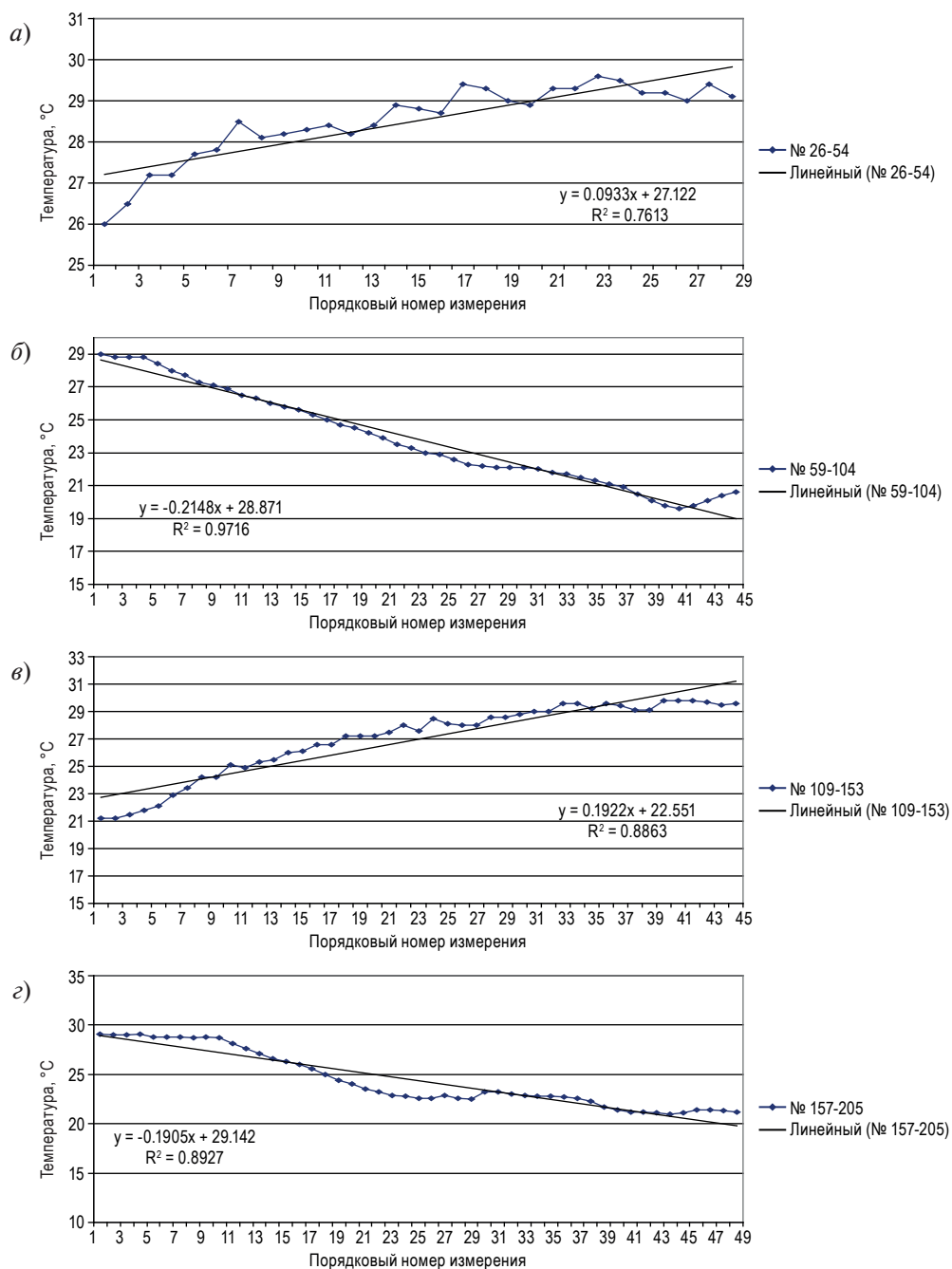


Рис. 3. Графики отрезков временного ряда, представленного на рис. 1, и локальных линейных временных трендов для этих же отрезков

Члены ряда с номерами $55 \div 58$ равны 0, т.е. алгоритм не выявил на этом участке никаких тенденций, а члены ряда с номерами $59 \div 104$ имеют значения -1 , что указывает на тенденцию к уменьшению значений температуры воздуха на этом временном промежутке.

Для следующих суток (номера $105 \div 205$) наблюдается похожая картина, отражающая суточный ход температуры воздуха.

Сравнительный анализ всего блока графиков на рис. 2 показывает, что изменение значения параметра nr хотя и оказывают некоторое влияние на положения границ выделяемых участков, но это влияние не носит монотонного характера и не приводит к заметному изменению результатов.

Представленные на рис. 3 графики иллюстрируют использование «локальных» по своей тенденции участков временного ряда для построения линий линейного тренда. В отличие от рис. 1, теперь построенные линии тренда значительно лучше описывают временные изменения ряда температуры, учитывают как возрастающие (рис. 3, а: $n = 26 \div 54$ и рис. 3, в: $n = 109 \div 153$), так и понижающиеся ветви (рис. 3, б: $n = 59 \div 104$ и рис. 3, г: $n = 157 \div 205$). Обращает на себя внимание и резкое увеличение уровня достоверности линейной аппроксимации: величина параметра R^2 колеблется от 0,76 до 0,97.

Рис. 4 иллюстрирует работу алгоритма для временного ряда приземной температуры воздуха, но уже длительностью 14 суток: с 01.07.2011 по 14.07.2011 (дискретность 15 мин). На всем протяжении временного ряда устойчиво выделяются восходящие и нисходящие ветви временного хода температуры.

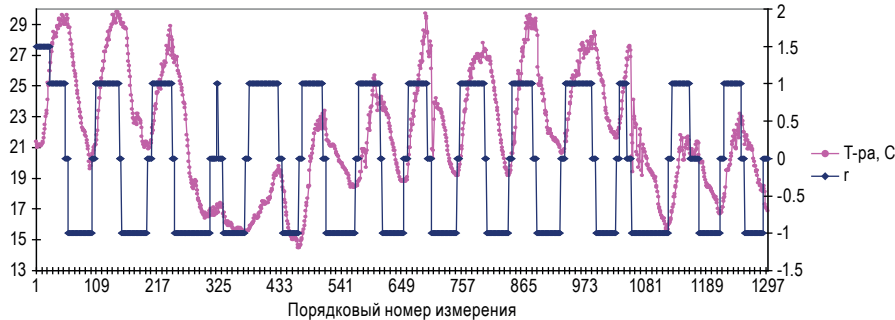


Рис. 4. Графики временного ряда температуры воздуха (левая ось ординат) и временного ряда функции R (правая ось ординат) при $nk = 25$, $nr = 9$

На рис. 5 представлены графики временного хода атмосферного давления и функции R в качестве примера работы алгоритма с временным рядом, имеющим значительно меньшую, по сравнению с температурой воздуха, изменчивость.

В заключение рассмотрим применение описанного ранее алгоритма выделения монотонных отрезков на примере временного ряда среднесуточных значений температуры воздуха в Санкт-Петербурге за 1994–1995 гг. (дискретность 24 часа). Причем на этом примере проиллюстрируем влияние величины контрольного участка (значения параметра nk) на положение и величину выделяемых участков.

На рис. 6 представлен график среднемесячной температуры воздуха в Санкт-Петербурге в 1994–1995 гг.

Блок графиков на рис. 7 демонстрирует результаты, полученные при расчете временного ряда функции R по данным о среднемесячной температуре воздуха, представленным на рис. 6.

Данные рис. 7 получены при одном и том же значении параметра np , тогда как величина параметра nk варьировалась от 10 (рис. 7, а) до 200 (рис. 7, з). Иными словами, при $nk = 10$ скользящее среднее определялось по 10 суткам, а при $nk = 200$ — по 200 суткам. Вследствие этого в первом случае значения ряда испытывали большие колебания около среднего значения, что приводило к формированию относительно коротких отрезков с устойчивой тенденцией изменения, а во втором — наоборот, к формированию длинных участков, имеющих устойчивую тенденцию относительно среднего значения, рассчитанному по гораздо большему участку временного ряда. Таким образом, можно сделать следующий вывод о выборе параметра nk при анализе временных рядов. В том случае, когда требуется выявить «тонкую» временную структуру ряда, выделяя большое число локальных изменений, необходимо брать небольшое значение параметра nk .

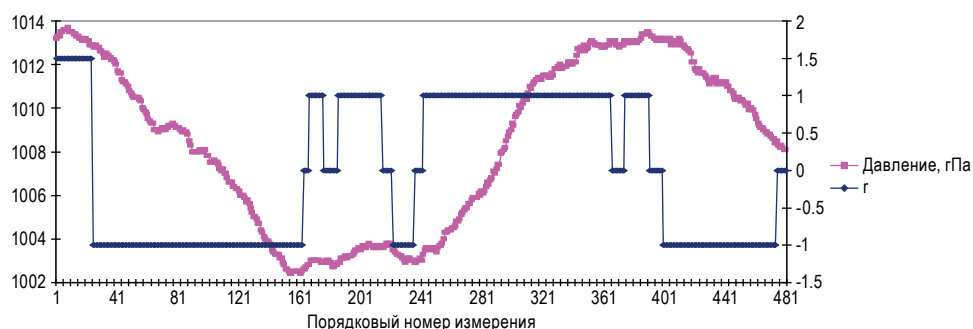


Рис. 5 Графики временного ряда атмосферного давления (левая ось ординат) и временного ряда функции R (правая ось ординат) при $nk = 25$, $np = 7$

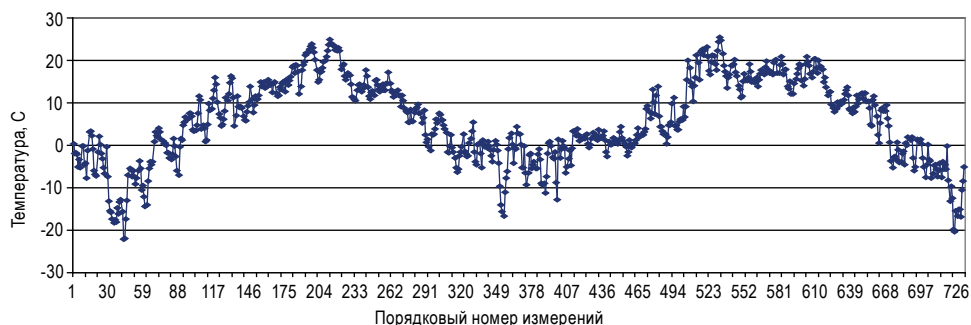


Рис. 6. Среднесуточная температура воздуха в Санкт-Петербурге в 1994–1995 гг.

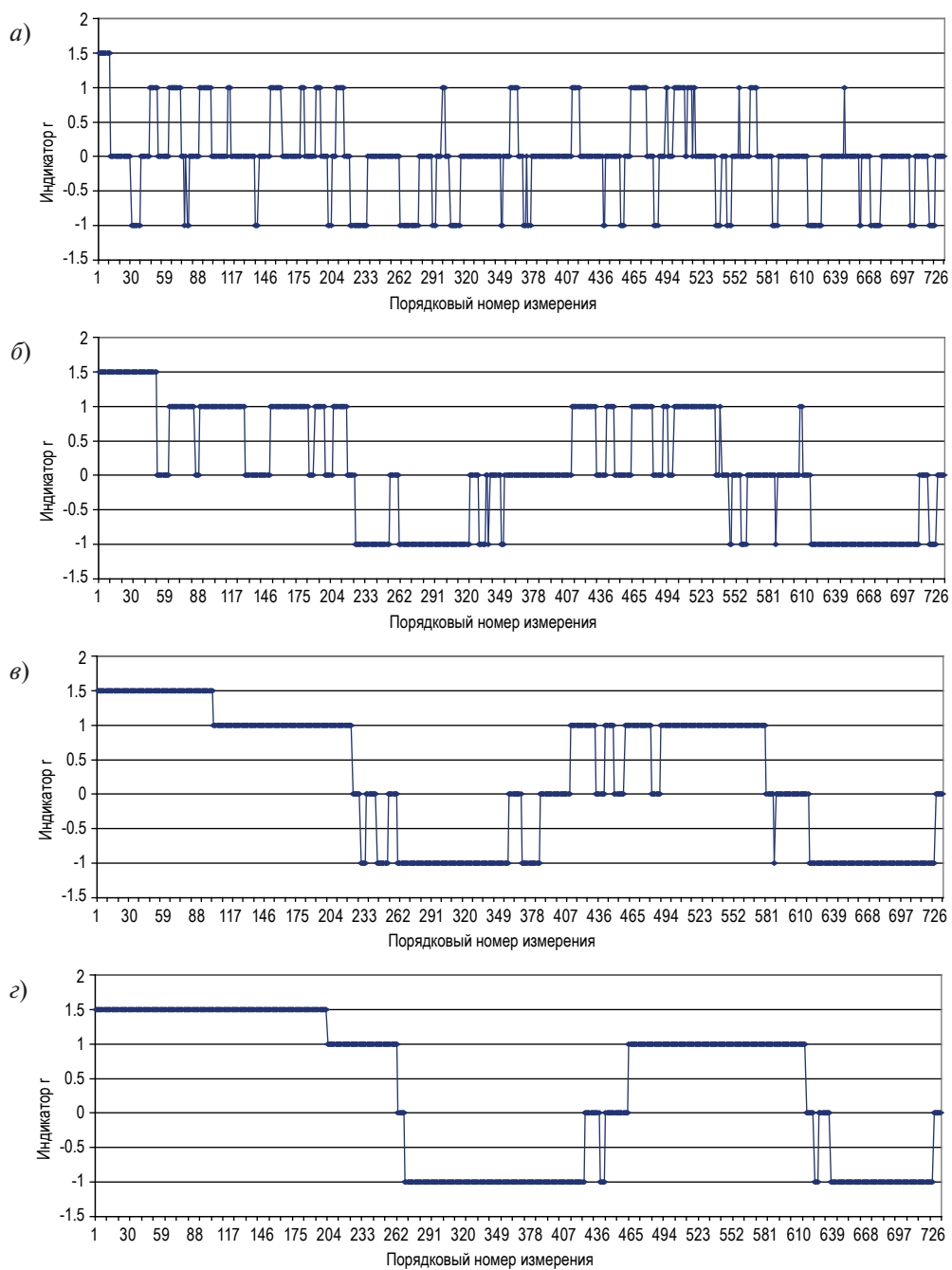


Рис. 7. Значения временного ряда функции W при значениях:
 а — $nk = 10, np = 7$; б — $nk = 50, np = 7$; в — $nk = 100, np = 7$; г — $nk = 200, np = 7$

И наоборот, для выявления продолжительных тенденций значение параметра nk надо увеличивать. Этот вывод иллюстрирует рис. 8. На нем два графика слева (рис. 8, *а*) позволяют сопоставить изменения временного ряда и функции R при малом значении параметра nk ($nk = 10$), а два графика справа (рис. 8, *б*) — при большом значении параметра nk ($nk = 200$). Это значение позволяет на графике отразить величину параметра nk и во всех значениях ряда от 1 до nk индикатор r не определяется.

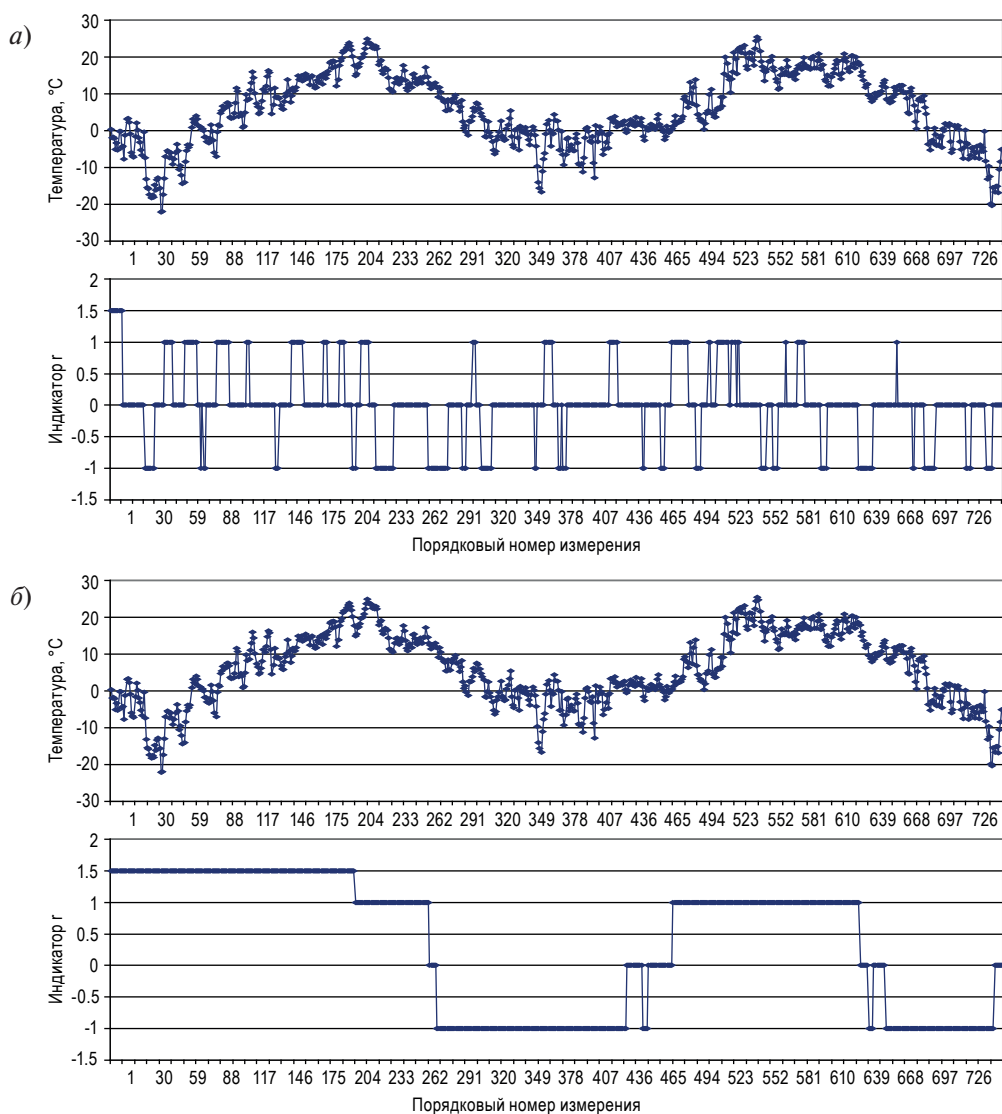


Рис. 8 Совместное представление временных рядов среднесуточной температуры воздуха и функции R при различных значениях параметра nk : *а* — $nk = 10$; *б* — $nk = 200$

В заключение 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 nk и nr (особенно nk) 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.

Литература

1. *Восканян К.Л., Кузнецов А.Д., Сероухова О.С.* Автоматические метеорологические станции. Ч. 2. Цифровая обработка данных автоматических метеорологических станций: учеб. пос. — СПб.: РГГМУ, 2015. — 80 с.
2. *Восканян К.Л., Кузнецов А.Д., Сероухова О.С., Симакина Т.Е.* К вопросу о выделении аддитивных составляющих временного ряда приземной температуры воздуха // Учен. зап. РГГМУ. 2013. № 32. — С. 55–65.
3. http://profbeckman.narod.ru/Doerffel.files/Doerffel_g12.pdf