

Математические методы анализа формы временных рядов и их приложения к дистанционным исследованиям Земли и атмосферы

А.Ф. Идрисов*

*Московский государственный университет имени М.В. Ломоносова,
физический факультет, кафедра математического моделирования и информатики
Россия, 119991, Москва, Ленинские горы, д. 1, стр. 2*

(Поступила в редакцию 09.06.2024; подписана в печать 16.07.2024)

Работа посвящена методам анализа временных рядов на основе теорий морфологического анализа и измерительно-вычислительных систем, разрабатываемых на кафедре математического моделирования и информатики физического факультета МГУ имени М. В. Ломоносова. Изучены современные актуальные методы анализа временных рядов. Предложен метод морфологической фильтрации для выделения трендовой составляющей временного ряда. Проведен компьютерный эксперимент на реальных экспериментальных данных (измерения температуры в городе Тверь) с применением вышеуказанного метода. Результаты продемонстрировали успешное выделение трендовой составляющей временного ряда. Планируется разработка и развитие новых методов анализа временных рядов для решения более широкого класса задач и для более широкого класса входных данных.

PACS: 05.45.Tr. УДК: 519.246.8.

Ключевые слова: временные ряды, морфологический анализ, форма сигнала, методы анализа, выделение тренда.

ВВЕДЕНИЕ

В настоящее время во многих направлениях научной деятельности возникает необходимость анализа и обработки данных, полученных путем регистрации в течение длительного периода времени через заданные временные промежутки (так называемые временные ряды). Кроме того, различные природные процессы обладают свойством периодичности, обусловленным вращением Земли вокруг Солнца и вокруг своей оси. Подобные процессы возникают в таких областях, как, например, физика Земли и физика атмосферы. При работе с данными такого рода вызывает интерес возможность выделения из них различных компонент, что позволяет в дальнейшем более эффективно проводить исследования в соответствующих областях.

Для решения этих задач существует множество математических методов [1, 2], таких как, например, Фурье-анализ [3] и вейвлет-преобразования. В данной работе предлагается использовать метод морфологической фильтрации, являющийся более гибким по сравнению с вышеуказанными методами. Он позволяет выделить желаемые компоненты временного ряда для проведения с ними последующих исследований.

Суть метода заключается в поиске проекции временного ряда на форму желаемой компоненты. Проекцией на форму является наилучшее в среднеквадратичном приближение временного ряда, которое удовлетворяет заданным условиям (принадлежит форме). Так, например, для выделения трендовой составляющей в качестве формы можно выбрать множество неубывающих последовательностей точек. В таком случае проекцию можно найти, применив подходящий алгоритм построения искомой последовательности.

Работа метода морфологической фильтрации продемонстрирована на температурных данных, регистрировавшихся ежемесячно в городе Тверь в период с 1974 по 2017 гг. Результаты демонстрируют успешное выделение трендовой составляющей временного ряда, которую затем можно применить для анализа периодической компоненты. В дальнейшем планируется развитие данного метода, его улучшение, обобщение и проверка применения для анализа различных видов данных и выделения различных компонент временных рядов.

1. МЕТОД МОРФОЛОГИЧЕСКОЙ ФИЛЬТРАЦИИ

При проведении исследований временных рядов, описывающих изменения различных метеорологических параметров, таких как температура, концентрация газов и др., анализируемый сигнал зачастую состоит из трендовой и (квази)периодической компонент [4]. Трендовая составляющая может быть объяснена определенными условиями и обстоятельствами, как антропогенного, так и природного характера, при которых происходила регистрация данных. Периодическая составляющая, как нетрудно заметить, является следствием движения Земли вокруг Солнца и вокруг своей оси (сезонность явлений). Во время анализа таких временных рядов может возникнуть задача разделения ряда на эти компоненты, в частности — задача выделения трендовой составляющей временного ряда.

Одними из возможных методов решения данной задачи являются метод Фурье-анализа, осуществляющий разложение ряда на гармонические компоненты, и вейвлет-анализ, осуществляющий разложение на масштабируемые базисные компоненты. Однако данные методы не всегда удобны, поскольку не обладают достаточной гибкостью при работе с различными входными данными. Более практичным методом явля-

* idrisov.af17@physics.msu.ru

ется метод морфологической фильтрации. Он основан на понятиях теории морфологического анализа [5–8], созданной и развиваемой Ю.П. Пытьевым и его учениками.

1.1. Понятия морфологического анализа

Морфологический анализ сигналов подразумевает нахождение некой структуры, *формы* сигнала, которая сохраняется при заданном классе преобразований. Такой подход позволяет достичь устойчивости к внешним условиям регистрации сигналов.

В теории морфологического анализа одним из ключевых понятий является форма сигнала $f(t)$, $t \in T$ — множество V_f всех сигналов, полученных из f всевозможными преобразованиями из заданного класса. При подходящем выборе модели и класса преобразований становится разрешимой задача наилучшего приближения произвольного заданного сигнала $g(t)$, $t \in T$ элементами множества V_f . Данный процесс называется *операцией проецирования* или *проецированием* сигнала на форму, а само решение этой задачи называется *проекцией* на форму.

В настоящей работе описанные понятия и инструменты морфологического анализа применяются для создания так называемого метода морфологической фильтрации временного ряда, который позволяет выделить трендовую составляющую ряда, удовлетворяющую заданным условиям. В качестве формы берется множество всевозможных элементов (временных рядов), которые удовлетворяют условиям, заданным исследователем. Далее выполняется проецирование заданного экспериментального временного ряда на вышеуказанную форму. В проведенном в работе эксперименте математический метод нахождения проекции зависит от заданных условий, определяющих форму искомого временного ряда. Найденная проекция и будет являться выделенной компонентой временного ряда.

1.2. Математическая модель и постановка задачи

Рассмотрим применение метода морфологической фильтрации для решения задачи выделения трендовой составляющей временного ряда. Пусть экспериментальный ряд $\xi(t)$, $t \in T$ состоит из тренда и квазипериодической компоненты, и мы хотим выделить неубывающий тренд. Для решения этой задачи нужно найти проекцию временного ряда на форму желаемой компоненты, т.е. наилучшее в среднеквадратичном приближение временного ряда набором точек, который удовлетворяет некоторым заданным условиям, т.е. принадлежит форме:

$$\Pi_V \xi = \arg \min_{f_i \in V} \sum_{i=1}^N |\xi(t_i) - f_i|^2.$$

Для выделения неубывающего тренда в качестве формы можно выбрать множество неубывающих числовых последовательностей:

$$V = \{f(t_1) \leq \dots \leq f(t_N)\} = \{f_1 \leq \dots \leq f_N\},$$

тогда проекцию можно найти, применив подходящий алгоритм построения искомой последовательности.

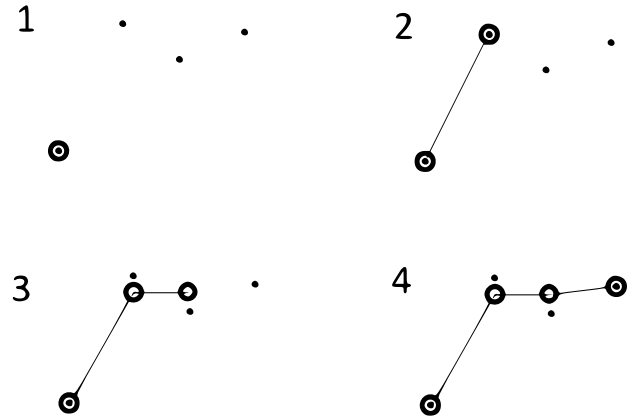


Рис. 1. Иллюстрация работы алгоритма построения неубывающей последовательности

На рис. 1 представлен алгоритм, позволяющий построить неубывающую последовательность, аппроксимирующую тренд временного ряда. Данный алгоритм проходит по всем точкам экспериментальных данных и строит последовательность в зависимости от их взаимного расположения, обеспечивая при этом монотонность уже пройденной части данных.

Схема алгоритма:

1. Первая точка берется равной соответствующей точке экспериментальных данных, $f_1 = \xi_1$.
2. Если положение следующей точки ξ_2 сохраняет неубывание ряда, то она так же берется без изменений, $f_2 = \xi_2$. Если же следующая точка находится ниже предыдущей, то две последние точки аппроксимируются их средним значением: $f_{1,2} = \frac{1}{2}(\xi_1 + \xi_2)$. Если такая аппроксимация сохраняет неубывание всей строящейся последовательности, то движемся к следующей точке (пункт 2). Если сохраняется убывание, то аппроксимируются три последние точки, и так далее, пока не будет выполняться неубывание или пока не будет достигнуто начало временного ряда.

Таким образом, применяя данный алгоритм, можно получить неубывающую последовательность, аппроксимирующую исходные данные.

2. КОМПЬЮТЕРНЫЙ ЭКСПЕРИМЕНТ

Работу вышеописанного метода морфологической фильтрации можно продемонстрировать на примере

температурных данных, регистрировавшихся ежемесячно в городе Тверь в период с 1974 по 2017 гг. Такие данные содержат в себе сезонную и трендовую составляющую, что позволяет использовать компьютерную реализацию метода для извлечения трендовой составляющей временного ряда.

Результаты (рис. 2) демонстрируют успешное выделение неубывающей трендовой составляющей временного ряда. Данную составляющую можно вычлест из исходного временного ряда и получить периодическую компоненту, которую затем можно проанализировать отдельно (рис. 3).

Для сравнения можно рассмотреть результат применения метода скользящего среднего [9, 10] к временному ряду (рис. 4). Хотя этот метод может давать меньшее значение ошибки, он имеет большую зависимость от размера сглаживающего окна, а также не позволяет получить монотонное приближение, которое и представляло интерес изначально.

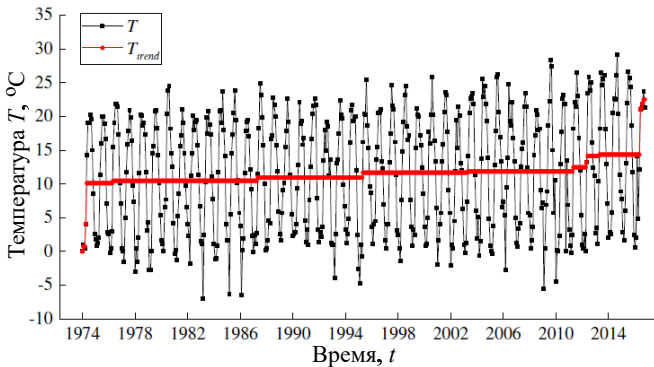


Рис. 2. Экспериментальный временной ряд с неубывающим трендом и построенная линия тренда

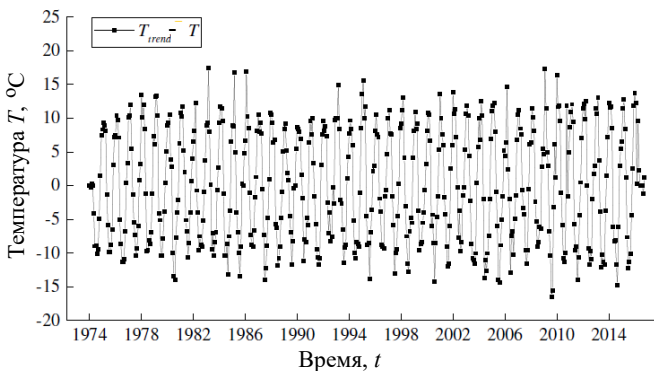


Рис. 3. Разность временного ряда и линии тренда

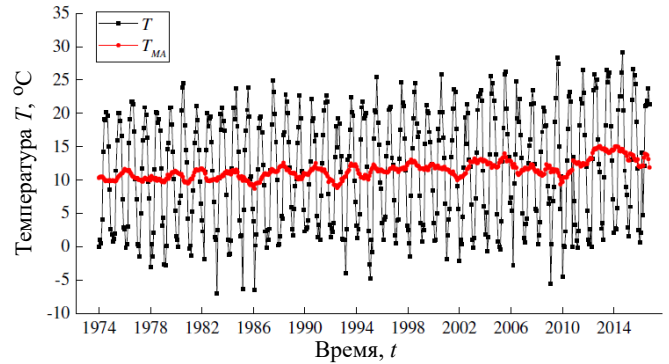


Рис. 4. Результаты применения метода скользящего среднего к временному ряду

3. СОГЛАСИЕ ИЗМЕРЕНИЯ С ПРЕДПОЛОЖЕНИЕМ О ФОРМЕ СИГНАЛА

Для проверки гипотезы о форме сигнала можно использовать следующее предположение. Пусть $\xi = g + \nu$, где $g \in V$, а ν — вектор той же размерности n , что и ξ , $M\nu = 0$, $\Sigma_\nu = \sigma^2 I$. Отсюда также следует утверждение $M\xi = g$. Противоречит ли ξ этому предположению?

Расстояние от ξ до V можно объяснить наличием флуктуаций/шума, и, очевидно, чем оно больше, тем больше возникает сомнений в адекватности применения данной модели. В таком случае какова вероятность получения реализации шума длиной $\|\xi - P_V \xi\|$, где $P_V \xi$ — проекция исходного сигнала на заданную форму?

Можно заметить, что $\|\nu\| \geq \|\xi - P_V \xi\|$, а из предположения $g = P_V \xi$ следует $\|\nu\| = \|\xi - P_V \xi\|$. Таким образом, для наилучшего согласования зарегистрированного сигнала ξ с предположением $M\xi = g$ нужно взять $g = P_V \xi$. Тогда, используя неравенство Чебышева, можно получить оценку вероятности получения результата, согласующегося с предположением $g \in V$ не лучше, чем ξ :

$$P(\|\nu\| \geq \|\xi - P_V \xi\|) \leq \frac{M\|\nu\|^2}{\|\xi - P_V \xi\|^2} = \frac{\sigma^2 n}{\|\xi - P_V \xi\|^2}.$$

Если при этом допустить нормальность распределения ν , то оценка приобретает вид:

$$P(\|\nu\|^2 \geq \|\xi - P_V \xi\|^2) = 1 - F_n(\|\xi - P_V \xi\|^2),$$

где F_n — функция распределения хи-квадрат с n степенями свободы.

Для получения численной оценки вышеописанной вероятности стоит заметить, что выражение $\frac{\|\xi - P_V \xi\|^2}{n}$ представляет собой среднеквадратичную ошибку MSE . Данная величина может быть рассчитана в компьютерном эксперименте ($n = 514$, $\|\xi - P_V \xi\|^2 = 35847.902$), следовательно, $MSE = 69.743$ — это также приведено на рис. 2).

Тогда численная оценка вероятности:

$$P(\|\nu\| \geq \|\xi - \Pi_V \xi\|) \leq \frac{\sigma^2}{MSE} \approx 0.014\sigma^2.$$

ЗАКЛЮЧЕНИЕ

На данный момент в настоящей работе был предложен и описан метод морфологической фильтрации для

выделения компонент временного ряда, а также был успешно проведен компьютерный эксперимент по выделению тренда на реальных данных.

В дальнейшем планируется развитие данного метода, его улучшение, обобщение и проверка применения для анализа различных видов данных и выделения различных компонент временных рядов.

Автор выражает благодарность профессору А. И. Чуличкову за научное руководство и помощь в работе.

-
- [1] *Авилов В.К., Алешновский В.С., Безрукова А.В.* и др. // Журнал вычислительной математики и математической физики. **61**, № 7. 1113 (2021).
[2] *Aleshnovskii V., Avilov V., Chulichkov A.* et al. // Pure and Applied Geophysics. **179**. 1 Oct. (2022).
[3] *Bloomfield P.* Fourier Analysis of Time Series: An Introduction. 2004.
[4] *Harvey A.* Forecasting with Unobserved Components Time Series Models. 2004.
[5] *Пытьев Ю.П.* // Докл. АН СССР. **224**, № 6. 1283 (1975).
[6] *Пытьев Ю.П.* // Докл. АН СССР **269**, № 5. 1061 (1983).
[7] *Пытьев Ю.П.* // Математические методы исследования природных ресурсов Земли из космоса. 1984. С. 41–82.
[8] *Пытьев Ю.П., Чуличков А.И.* Методы морфологического анализа изображений. М., 2010.
[9] *Кричевский М.* Временные ряды в менеджменте. 2016.
[10] *Gazaryan V., Kurbatova J., Ovsyannikov T.* et al. // Moscow Univ. Phys. Bull. **70**, 346 (2015).

Mathematical methods for analyzing the shape of time series and their applications to remote sensing of the Earth and atmosphere

A.F. Idrisov

*Department of Mathematical Modeling and Computer Science, Faculty of Physics, Lomonosov Moscow State University,
Moscow 119991, Russia
E-mail: idrisov.af17@physics.msu.ru*

The work is devoted to methods of time series analysis based on the theory of morphological analysis and measuring and computing systems developed at the Department of Mathematical Modeling and Computer Science of the Faculty of Physics of Lomonosov Moscow State University. Modern relevant methods of time series analysis have been studied. A morphological filtering method is proposed to detect a trend in a time series. A computer experiment was conducted on real experimental data (temperature measurements in the city of Tver) using the above method. The results demonstrated the successful identification of the trend component of the time series. It is planned to develop new methods of time series analysis to solve a wider class of problems for a wider class of input data.

PACS: 05.45.Tr.

Keywords: time series, morphological analysis, signal form, methods of analysis, trend detection.

Received 09 June 2024.

Сведения об авторах

Идрисов Александр Фаритович — аспирант; e-mail: idrisov.af17@physics.msu.ru.