

# ОБРАБОТКА МЕТЕОРОЛОГИЧЕСКИХ ДАННЫХ МЕТОДОМ СИНГУЛЯРНОГО СПЕКТРАЛЬНОГО АНАЛИЗА

К.В. Новицкая, И.А. Ботыгин  
Томский политехнический университет  
E-mail: kvn8@tpu.ru

## Введение

Одна из распространенных задач статистического анализа – нахождение скрытых закономерностей, исследуемых данных и их реконструкция из выделенного базиса собственных векторов. Для многомерных данных обычно используют метод главных компонент (Principal Components Analysis, PCA). Заметим, что на одномерном наборе данных удобнее использовать сингулярный спектральный анализ (Singular spectrum analysis, SSA) [1].

Сингулярный спектральный анализ относится к непараметрическим методам анализа временных рядов. Целью метода является разложение временного ряда на интерпретируемые аддитивные составляющие. Метод не требует стационарности ряда, знания модели тренда, а также сведений о наличии в ряде периодических составляющих [2]. При таких слабых предположениях метод может решать различные задачи, например, выделение тренда, обнаружение периодической составляющей, сглаживание ряда.

Теория SSA отвечает на вопросы [1]: какие составляющие временного ряда могут быть выделены, как выбрать длину окна сингулярного разложения и провести правильную группировку собственных векторов для выделения нужных компонентов.

В настоящем докладе проведен анализ данных, полученных с ультразвуковой метеостанции №167 с координатами (56.48° с.ш., 85.05° в.д.). Данные содержат значения о средней температуре в градусах Цельсия за период с 2018-10-01 по 2019-10-15. В программных экспериментах были использованы температурные ряды наблюдений с различным шагом дискретизации.

## Этапы обработки

Обработка данных методом SSA включала два взаимодополняющих этапа [3]: декомпозиция и реконструкция.

На этапе декомпозиции одномерный ряд данных преобразуется в многомерный путем образования траекторной матрицы, и матрица раскладывается на сингулярные векторы.

На этапе реконструкции формируются (группируются) тройки сингулярных векторов. Далее каждая матрица сгруппированного разложения ганкелизуется (усредняется по анти-диагоналям) и затем полученная ганкелева матрица трансформируется в новый временной ряд на основе взаимно-однозначного соответствия между ганкелевыми матрицами и временными рядами.

Диагональное усреднение, применённое к каждой результирующей матрице, производит восстановление ряда. Таким образом, исходный ряд раскладывается в сумму восстановленных рядов. Данное разложение является главным результатом алгоритма SSA для анализа временного ряда. Это разложение имеет смысл, если каждый из его компонентов может быть интерпретируем как-либо тренд, либо колебания (периодики), либо шум.

## Анализ полученных данных

Практическая реализация указанных выше этапов осуществлялась с использованием пакета Rssa языка статистического моделирования R. На этапе декомпозиции из исходного температурного ряда были получены пары собственных векторов (рис. 1). Графики пар собственных векторов могут указывать на то, какие векторы отвечают за сезонность. Обычно, эти графики образуют полигон или окружность [4]. Проведенные вычисления и визуализация исследуемого ряда показали, что пары векторов 4-5 и 6-7 образует круг. Это значит, что составляющие температурного ряда – гармонические, т.е. свидетельствуют о сезонности. Остальные векторы, предположительно, относятся к шуму.

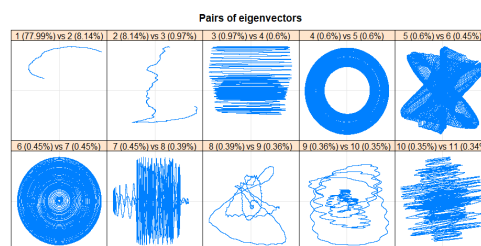


Рис. 1. Пары собственных векторов

Проведя реконструкцию компонентов, можно четко выделить тренд, однако сезонность имеет сложную форму. На рисунке 2 представлены восстановленные векторы, где тренд соответствует векторам 1-3, а сезонность – векторам 4-7.

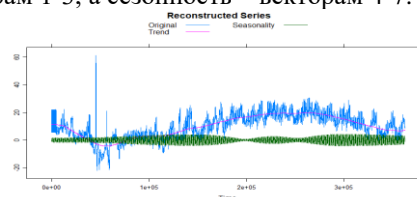


Рис. 2. Восстановленные компоненты ряда с большим количеством данных

Выполненный рекуррентный и векторный прогнозы на данных большого объема (дискретность наблюдений – 1 мин.) близки по результатам (рис.3),

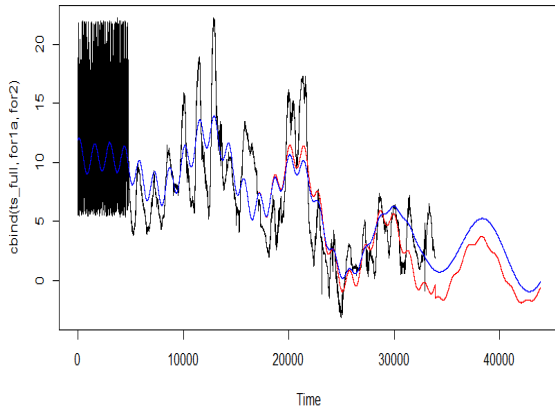


Рис. 3. Прогноз временного ряда

Для временного ряда с наблюдениями, взятыми с дискретностью раз в два дня, тренд, выделяется также хорошо и можно увидеть, что форма сезонности стала проще и отчетливо видно, что только одна пара векторов теперь отвечает за сезонность. На рисунке 4 выделенные тренд и сезонность наложены на исходный график временного ряда. Прогнозирование проходит чуть менее точно, чем на полном объеме данных.

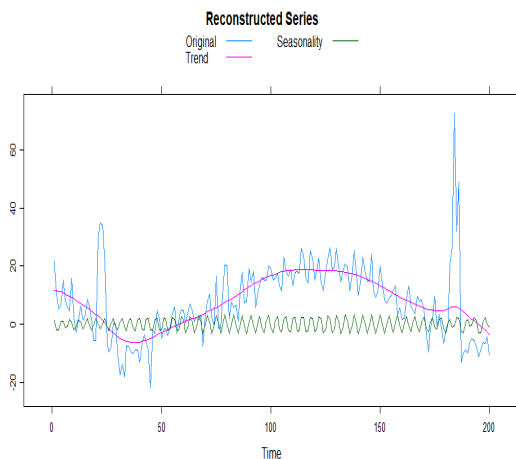


Рис.4. Восстановленные компоненты

На данных, взятых с дискретностью 5 раз в месяц, форма графика сезонности сильно упростилась, прогнозирование затруднительно из-за колебаний, но тренд выделяется хорошо (рис. 5).

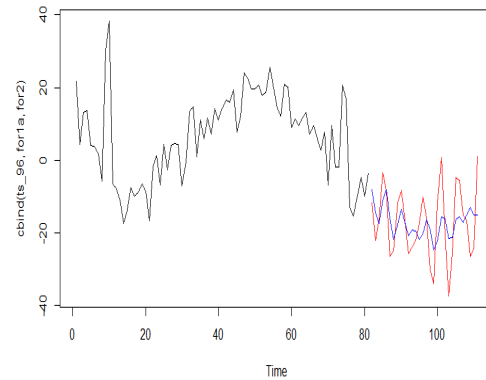


Рис. 5. Прогноз для данных наблюдений 5 раз в месяц

### Заключение

В результате проведения сингулярного спектрального анализа на данных разного размера можно сделать вывод о том, что малый объем данных не дает почти никакой полезной информации. Чем больше объем данных, тем сложнее выделяется сезонность или она может иметь сложную форму, однако прогнозирование – более точное. Кроме того, анализ данных слишком большого объема требует больших вычислительных ресурсов, что существенно замедляет проведение анализа.

### Список использованных источников

1. Unsupervised learning for time series data: Singular spectrum versus principal components analysis. [Электронный ресурс] – URL: <http://karthur.org/2017/learning-for-time-series-ssa-vs-pca.html>
2. Golyandina, Nina & Korobeynikov, Anton. (2014). Basic Singular Spectrum Analysis and Forecasting with R. Computational Statistics & Data Analysis. [Электронный ресурс] / ResearchGate – URL: [https://www.researchgate.net/publication/228092069\\_Basic\\_Singular\\_Spectrum\\_Analysis\\_and\\_Forecasting\\_with\\_R](https://www.researchgate.net/publication/228092069_Basic_Singular_Spectrum_Analysis_and_Forecasting_with_R)
3. SSA (метод). [Электронный ресурс] / Википедия. – URL: [https://ru.wikipedia.org/wiki/SSA\\_\(%D0%BC%D0%B5%D1%82%D0%BE%D0%B4\)](https://ru.wikipedia.org/wiki/SSA_(%D0%BC%D0%B5%D1%82%D0%BE%D0%B4))
4. Analysis of time series structure: SSA and related techniques / N. Golyandina, V. Nekrutkin, and A. Zhigljavsky.
5. Документация пакета Rssa. [Электронный ресурс] / CRAN. – URL: <https://cran.r-project.org/web/packages/Rssa/Rssa.pdf>
6. Elsner, J., and A. Tsonis. 1996. Singular spectrum analysis: a new tool in time series analysis. New York and London: Plenum Press.