

ИНСТРУМЕНТЫ ВИЗУАЛИЗАЦИИ ВРЕМЕННЫХ РЯДОВ В КОСМИЧЕСКИХ ИССЛЕДОВАНИЯХ

В. С. Мараев

Сибирский федеральный университет,
г. Красноярск, Российская Федерация

Анализ временных рядов является основным шагом в построении модели прогнозирования, но очень важно рассмотреть данные временных рядов с различных сторон. Особенно важно это бывает при анализе данных космических исследований, так как зачастую их анализ не приводит к достаточно чёткой закономерности. Также анализ помогает выявить различные признаки и параметры рассматриваемого временного ряда. Именно поэтому визуализация данных является сильным инструментом на этапе формирования модели прогнозирования. Очень часто для визуализации временных рядов в космических исследованиях используются только линейные графики, которые не могут отобразить всей специфики ряда, поэтому стоит рассматривать другие способы и инструменты визуализации данных космических исследований. В данной статье проводится анализ материалов по инструментам визуализации временных рядов. Рассмотрены основные инструменты визуализации, такие как гистограммы, графики плотности распределения, диаграммы размаха и рассеяния, тепловые карты и графики автокорреляции. Данные инструменты продемонстрированы на примерах данных, полученных из космических исследований. Выявлены достоинства и недостатки различных инструментов визуализации временных рядов в космических исследованиях. Изложены рекомендации по целесообразности применения данных инструментов визуализации в различных ситуациях. Сделаны соответствующие выводы на основе проведённых исследований по анализу инструментов визуализации временных рядов в космических исследованиях.

Ключевые слова: визуализация, временные ряды, космические исследования, гистограмма, диаграмма размаха, тепловая карта, график автокорреляции.

Введение

Большинство данных космических исследований, представленных в виде временных рядов, отлично визуализируются на обычном графике. Линейные графики наблюдений относительно времени очень популярны, но есть набор и других методов визуализации, при использовании которых можно больше узнать об изучаемых данных, что особенно необходимо при точном изучении временных рядов в космических исследованиях.

Чем глубже изучается набор данных, тем более точную модель прогнозирования можно реализовать [1].

Визуализация временных рядов

Визуализация играет очень важную роль в анализе временных рядов и прогнозировании.

С помощью графика исходных данных можно выявить временные структуры, такие как: тренды, циклы и сезонности, которые могут влиять на выбор модели [2, 3]. Проблема заключается в том, что не всегда используют полный спектр инструментов визуализации, останавливаясь только на линейном графике.

На самом деле инструментов визуализации данных достаточно много, но в нашем исследовании рассмотрено шесть основных видов визуализации:

- линейный график;
- гистограммы и график плотности распределения;
- ящик с усами (диаграмма размаха);
- тепловая карта;
- диаграмма рассеяния;
- график автокорреляции.

Основное внимание уделено одномерным временным рядам космических исследований, но описываемые инструменты можно использовать и на многомерных рядах.

Визуализация рядов реализована с помощью языка программирования Python [4, 5]. Благодаря его специальным библиотекам использование различных инструментов визуализации при построении рядов многофункционально, эффективно и просто [5, 6].

Визуализируемые примеры временных рядов космических исследований

Для демонстрации инструментов визуализации выбрано несколько наборов данных, отличающихся сложностью, массивностью, распределением и другими характеристиками.

Ряд номер 1 описывает число нумерованных малых планет, открытых в разные годы шестью основными обсерваториями. Данный ряд является наименьшим из визуализируемых рядов. Источником данного набора является «Международный центр малых планет» [7].

Ряд номер 2 описывает число открытых больших околоземных астероидов (километровый размер и больше) в период 1980–2017 гг. Данный ряд превосходит по размерности первый, однако также является достаточно небольшим. Источником данного набора является «Центр динамики околоземных объектов» [8].

Ряд номер 3 описывает число открытых в год метеоритов в период 1995–2015 гг. Данный ряд также не является большим, поэтому отнесём его к разряду средних. Источником данного набора является «Международная метеорная организация» [9].

Ряд номер 4 описывает число орбитальных запусков (включая неудачные) в период 1957–2017 гг. Источником данного набора является «Хронология пилотируемых космических полётов» [10].

Ряд номер 5 описывает данные, переданные космическим телескопом «Хаббл», отражает зависимость рецессивной скорости от дистанции. Этот ряд содержит большое количество данных, поэтому отнесём его к разряду больших. Источником данного набора является «Официальное открытое хранилище данных космического телескопа “Хаббл”» [11].

Ряд номер 6 описывает изменение уровня космического шума каждую миллисекунду в течение десяти секунд. Данный ряд является наибольшим из визуализируемых рядов, так как содержит большое число значений, однако примерно половина из них нулевая. Источником данного набора является «Официальное открытое хранилище данных космического телескопа “Хаббл”» [11].

Линейный график временного ряда

Первым и наиболее популярным способом визуализировать временной ряд является линейный график. Линейный график – это график, на котором через определенные промежутки времени отмечаются и соединяются линией точки, изображающие данные [3, 12]. На графиках (рис. 1) представлена линейная визуализация зависимостей для всех шести анализируемых примеров временных рядов космических исследований.

На графике 6 мы наблюдаем резкие скачки уровня космического шума, это объясняется частым полным отсутствием шума. Как результат, данный линейный график получается достаточно плотным, поэтому иногда очень полезно изменить его стиль на пунктир или точки. На рис. 2 представлен измененный график ряда номер 6.

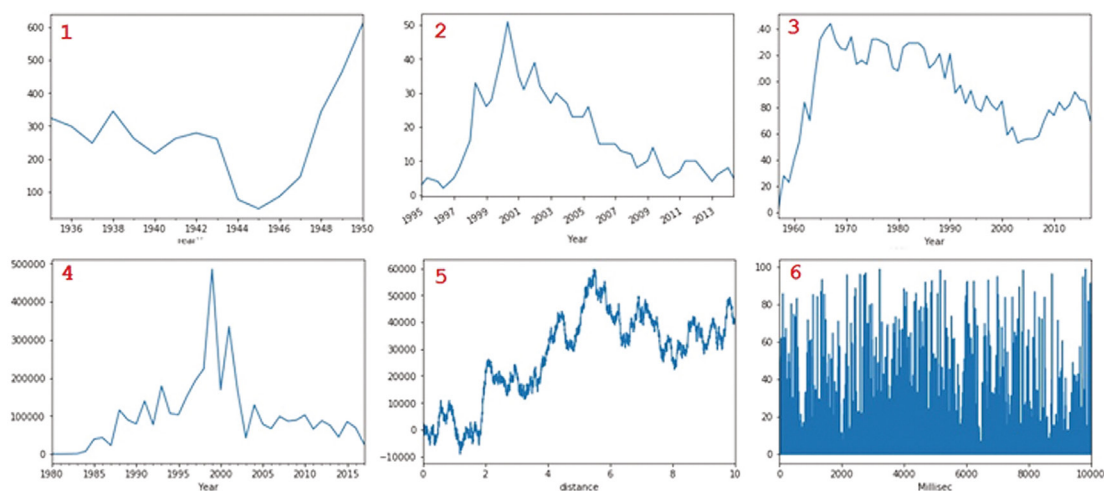


Рис. 1. Примеры линейных графиков (каждый ряд обозначен красной цифрой, соответствующей номеру визуализируемого ряда)

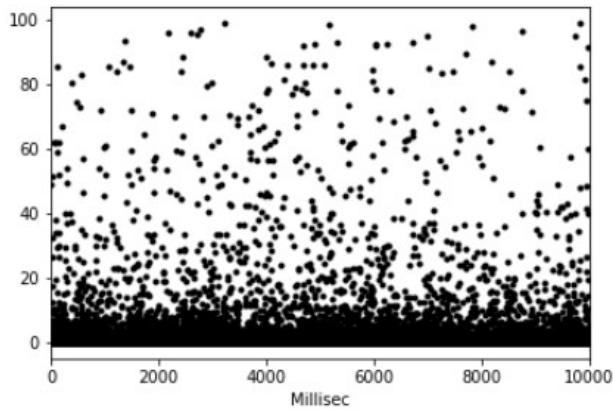


Рис. 2. Линейный график ряда 6 с измененным стилем линий на точки

Гистограммы временных рядов и графики плотности распределения

Другим очень важным способом визуализации является график распределения данных временных рядов. Это график средних значений без упорядочивания по времени [3, 13].

Некоторые методы прогнозирования временных рядов предполагают хорошую производительность на данных нормального распределения. Эти гипотезу можно проверить очень быстро, используя график распределения. Также преимуществом данного инструмента анализа является то, что проверку можно произвести как на необработанных данных, так и на преобразованных [14].

На рис. 3 представлены гистограммы для всех шести анализируемых примеров временных рядов космических исследований.

Однако можно получить более полное представление о форме распределения наблюдений с использованием графика плотности. Сам график подобен гистограмме, за исключением того, что используется функция сглаживания результатов для этого распределения в общем [15].

На рис. 4 представлены графики плотности распределения для всех шести анализируемых примеров временных рядов космических исследований.

Диаграмма размаха

Ещё один способ визуализации — это диаграмма размаха.

Диаграмма размаха, также известная как ящик с усами, — это график, компактно изображающий одномерное распределение вероятностей. Такой вид диаграммы в удобной форме показывает медиану (или, если нужно, среднее), нижний и верхний квартили, минимальное и максимальное значения выборки и выбросы [3, 16]. Прямоугольник фиксирует среднее 50 % наблюдений. Внутренняя линия — медиана, а усы, расположенные выше и ниже прямоугольника, демонстрируют минимальное и максимальное значения выборки.

На рис. 5 представлены диаграммы размаха для всех шести анализируемых примеров временных рядов космических исследований.

На графике, отображающем уровень космического шума, мы наблюдаем узкий прямоугольник вблизи нуля, это объясняется частым полным отсутствием шума.

Такой график можно построить для любого временного интервала, что является преимуществом при анализе временных рядов.

В отличие от гистограмм и графиков плотности, дающих представление о распределении всех данных в общем, диаграмма размаха может

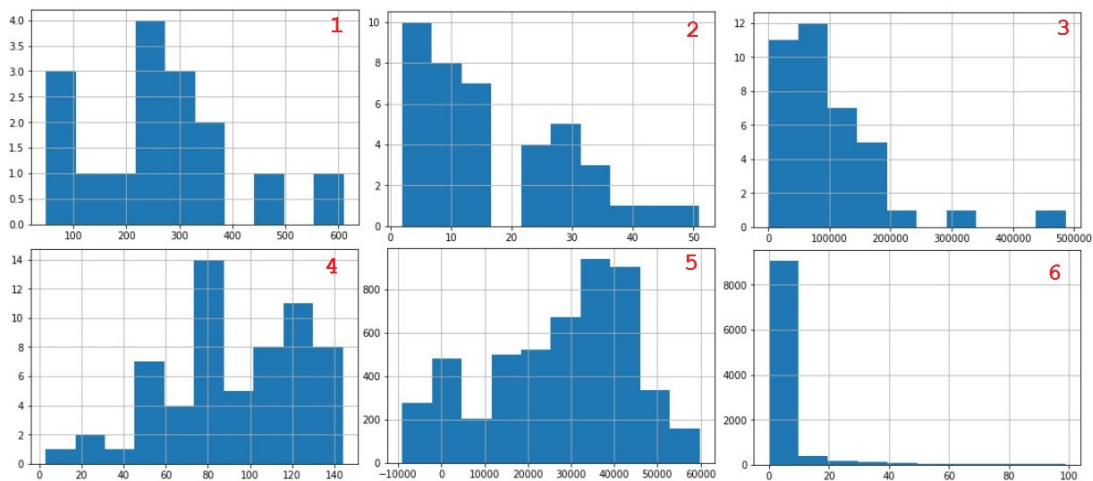


Рис. 3. Примеры гистограмм (каждый ряд обозначен красной цифрой, соответствующей номеру визуализируемого ряда)

отобразить наглядный график распределения значения по интервалам.

Нет смысла разбивать маленькие ряды на интервалы, поэтому разобьем только наиболее массивные ряды. Ряд, отображающий число орбитальных запусков (ряд 1 на рис. 6), разобьем по десятилетиям. Ряд, отображающий рецессивную скорость (ряд 2 на рис. 6), разобьем по дистанции. Ряд, отображающий уровень космического шума (ряд 3 на рис. 6), разобьем по секундам.

Тепловая карта временных рядов

Матрица значений температур может быть сформирована в качестве представления, где значение в каждой ячейке матрицы может присваивать определенный цвет [17]. Данный график называется тепловой картой, где чем меньше значение в наблюдении, тем теплее цвета и наоборот.

На рис. 7 представлены тепловые карты для всех шести анализируемых примеров временных рядов космических исследований.

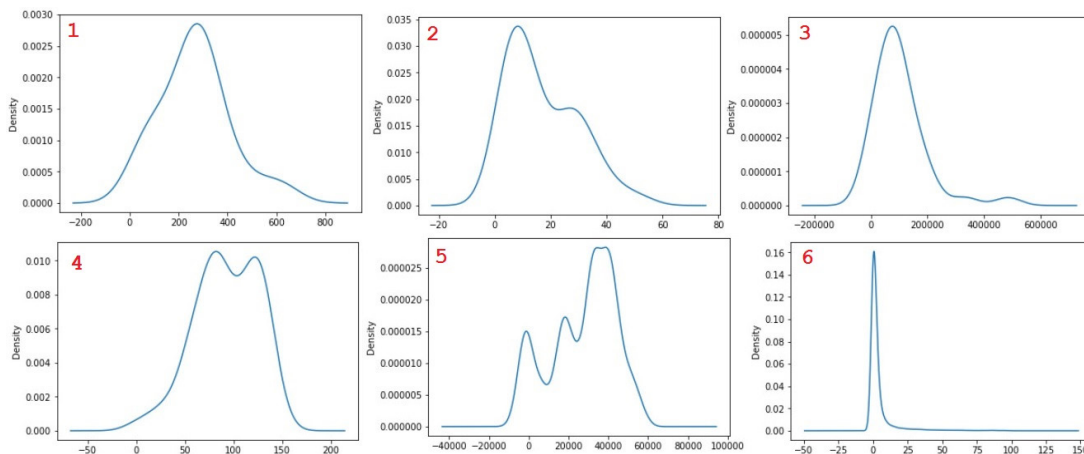


Рис. 4. Примеры графиков плотности распределения (каждый ряд обозначен красной цифрой, соответствующей номеру визуализируемого ряда)

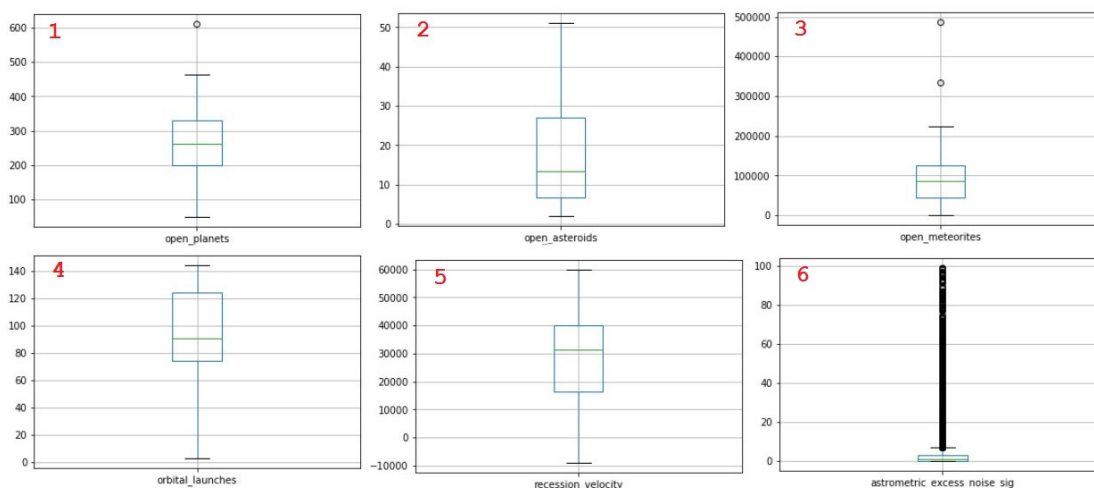


Рис. 5. Примеры диаграмм размаха (каждый ряд обозначен красной цифрой, соответствующей номеру визуализируемого ряда)

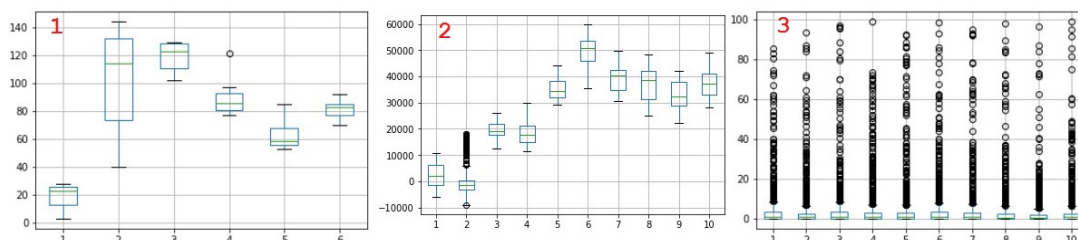


Рис. 6. Примеры диаграмм размаха для интервалов

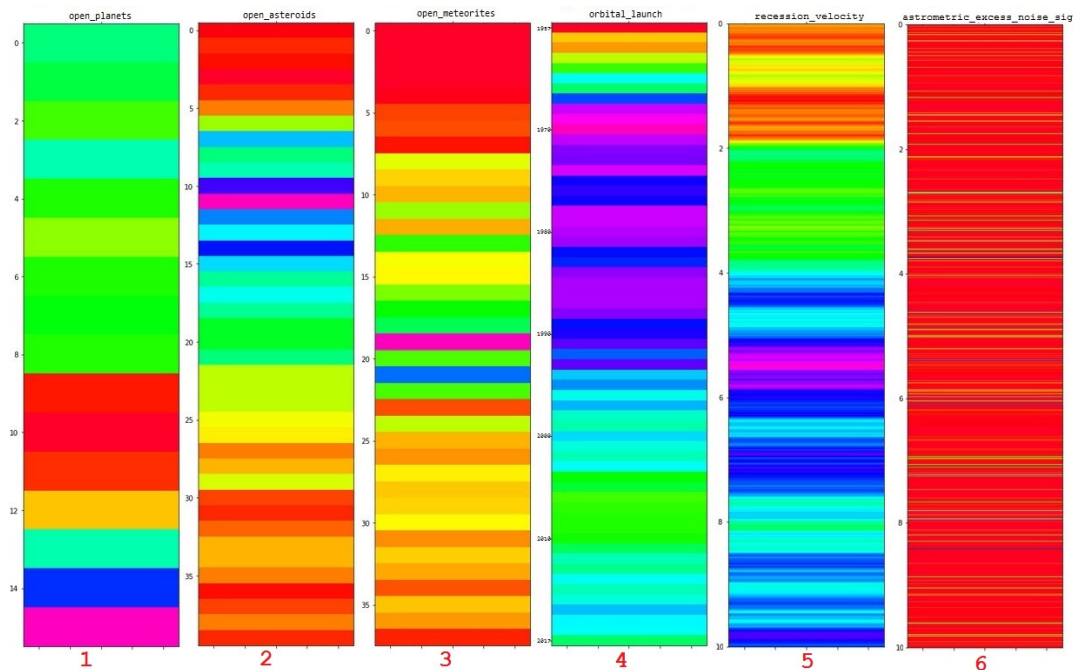


Рис. 7. Примеры тепловых карт (каждый ряд обозначен красной цифрой, соответствующей номеру визуализируемого ряда)

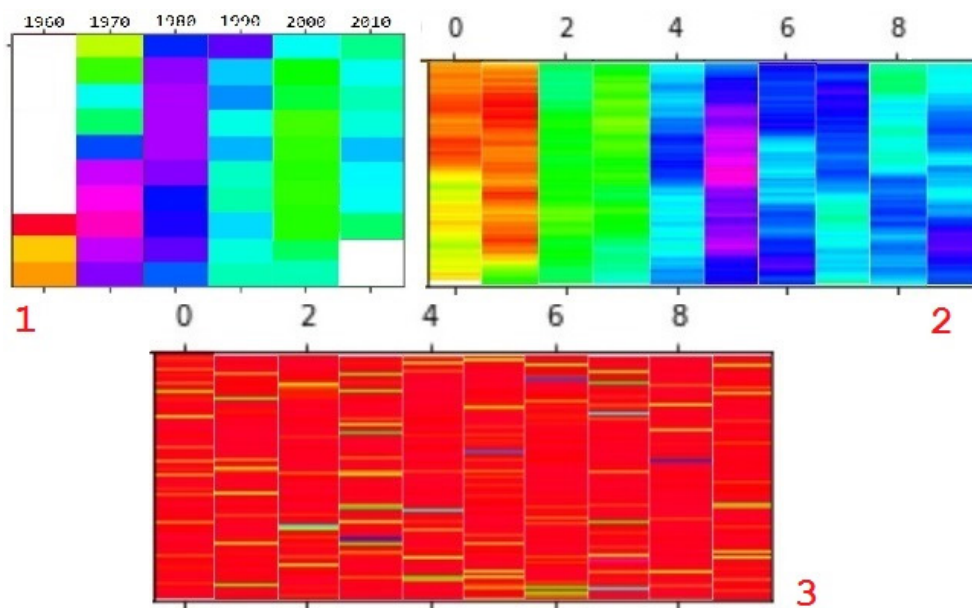


Рис. 8. Примеры тепловых карт для интервалов

Не все графики являются удачно-визуализируемыми. Например, график 6, визуализируемый при помощи тепловой карты, является практически полностью одноцветным. Среди красных (маленьких) значений проявляется немного больше холодных значений, отображающих резкие случайные скачки уровня космического шума. Схожую проблему можно увидеть на других визуализациях этого ряда.

Как и на диаграмме размаха, на тепловой карте возможно сравнить наблюдения между ин-

тервалами. Построим те же три наибольших графика, разбивая их на интервалы, так же как при построении диаграмм размаха (рис. 8).

Диаграмма рассеяния

Модель временных рядов предполагает существование взаимосвязей между текущим и предыдущим наблюдениями. Предыдущим наблюдением временного ряда называется лаг. Именно поэтому полезным инструментом визуализации

в данном случае является диаграмма рассеивания, в которой демонстрируются отношения между наблюдением и лагом.

Чаще всего диаграмма рассеивания строится со временными интервалами по абсциссе и лагами ($t - l$) по ординате. Расположение точек на графике имеет различный характер.

Так, если точки расположены вдоль диагонали с левого нижнего угла до правого верхнего, то у набора данных положительная корреляция, и если наоборот, то отрицательная [3, 18].

Таким же образом установлено, что чем меньше расстояние между точками, тем сильнее взаимосвязь между наблюдениями [19].

На рис. 9 представлены графики рассеивания для всех шести анализируемых примеров временных рядов космических исследований.

На графике 5 хорошо демонстрируется положительная корреляция и сильная связь между наблюдениями.

Необходимо заметить, что данный график можно построить относительно каждого предыдущего наблюдения. К примеру, возможно построить график отношения с предыдущей неделей, месяцем, годом и т.д.

График автокорреляции

Для анализа временных рядов также полезно количественно оценить силу и тип взаимоотношений между наблюдениями и их лагами. В статистике это называется корреляцией, и при расчете зависимости от значений лагов во вре-

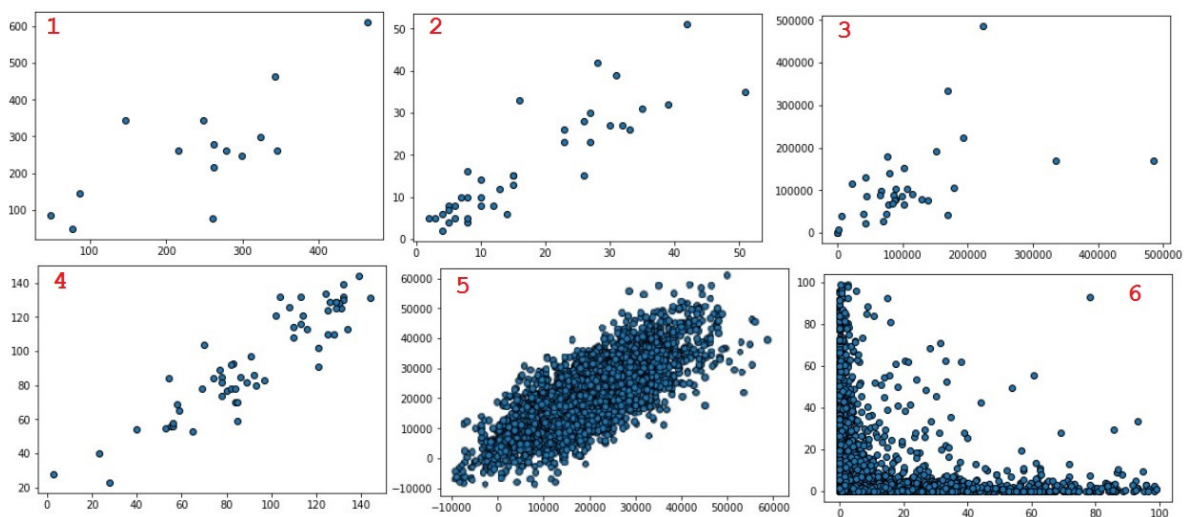


Рис. 9. Примеры графиков рассеивания (каждый ряд обозначен красной цифрой, соответствующей номеру визуализируемого ряда)

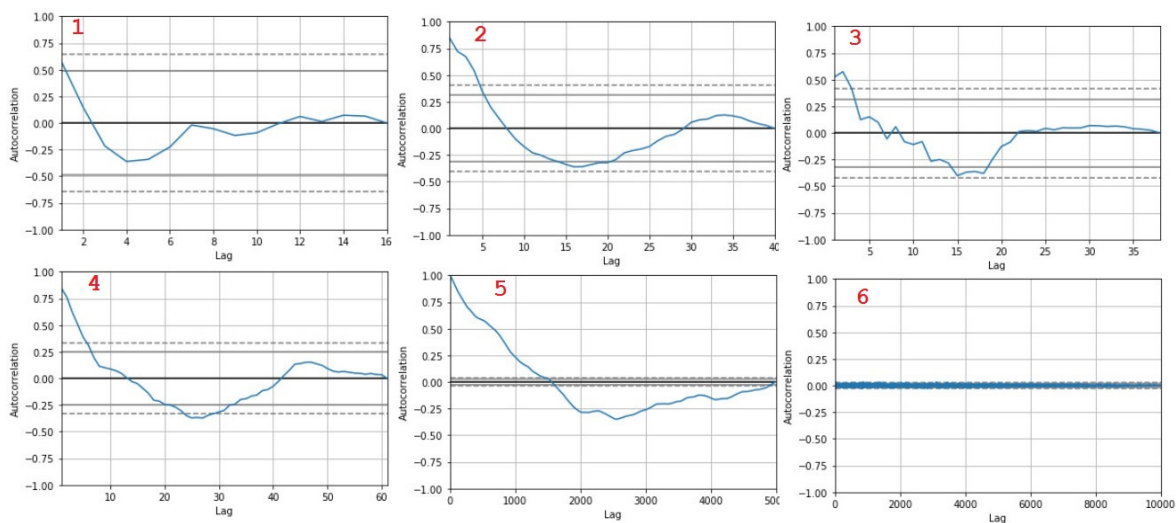


Рис. 10. Примеры графиков автокорреляции (каждый ряд обозначен красной цифрой, соответствующей номеру визуализируемого ряда)

менных рядах также известна как автокорреляция [20].

Значение корреляции рассчитывается между двумя группами чисел, таких как наблюдения и лаги, результатом является число между -1 и 1 . Знак результата показывает на отрицательную или положительную корреляцию, а значение, близкое к нулю, указывает на слабую корреляцию и наоборот [3].

На рис. 10 представлены графики автокорреляции для всех шести анализируемых примеров временных рядов космических исследований.

Как видно на рис. 10, в примерах присутствуют циклы как положительных, так и отрицательных корреляций.

Выводы

Продемонстрировано, что такие инструменты визуализации временных рядов в космических исследованиях, как гистограммы, графики плотности распределения, тепловые карты, графики автокорреляции, диаграммы размаха и рассеяния, сокращают время анализа данных и позволяют намного глубже понимать проблему исследуемого временного ряда, что особенно необходимо при точном изучении данных космических исследований.

Установлено, что представленные инструменты визуализации временных рядов помогают выявить множество различных скрытых параметров космических исследований, которые необходимы для дальнейшей реализации модели прогнозирования.

Список литературы

- George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel. Time Series Analysis: Forecasting and Control. Wiley Series in Probability and Statistics. 2013. С. 92–161. ISBN 1118619064, 9781118619063.
- Кендалл М., Стьюарт А. Многомерный статистический анализ и временные ряды. М. : Наука, 1976. 736 с.
- Мишулина О. А. Статистический анализ и обработка временных рядов. М. : МИФИ, 2004. С. 38.
- Евдокимов И. В., Михалев А. С., Новиков О. С., Суханова А. В. Применение свободных лицензий для разработки программного обеспечения в России // Международный журнал прикладных и фундаментальных исследований. 2017. № 6-1. С. 33–36.
- Евдокимов И. В., Яценков К. Г., Телков А. Ю., Татауров В. А. Экспертные методы оценки трудоёмкости разработки программных проектов // Экономика и менеджмент систем управления. 2017. Т. 24, № 2.2. С. 272–276.
- Evdokimov I. V., Domantsevich V. S., Konyhov V. A. THE USE OF ONLINE APPLICATIONS FOR PROJECT MANAGEMENT FOR PLANNING IN IT-MANAGEMENT // Современные информационные технологии. 2017. № 25 (25). С. 44–47.
- Международный центр малых планет // Data Available from the Minor Planet Center [Электронный ресурс]. URL: <http://www.minorplanetcenter.net/> (дата обращения: 30.10.2017).
- Центр динамики околоземных объектов // The AstDyS database [Электронный ресурс]. URL: <http://hamilton.dm.unipi.it/astdys/index.php?pc=2.0&o=094> (дата обращения: 30.10.2017).
- Международная метеорная организация // Visual Meteor Database [Электронный ресурс]. URL: <https://www.imo.net> (дата обращения: 30.10.2017).
- Хронология пилотируемых космических полётов // LAUNCHLOG [Электронный ресурс]. URL: <http://planet4589.org/space/log/launch.html> (дата обращения: 30.10.2017).
- Официальное открытое хранилище данных космического телескопа “Хаббл” // Hubble Space Telescope database [Электронный ресурс]. URL: <http://www.nasa.gov/hubble/> (дата обращения: 30.10.2017).
- Айвазян С. А. Прикладная статистика. Основы эконометрики. Т. 2. М. : Юнити-Дана, 2001. С. 432. ISBN 5-238-00305-6.
- Магнус Я. Р., Катышев П. К., Пересецкий А. А. Эконометрика. Начальный курс. М. : Дело, 2007. С. 504. ISBN 978-5-7749-0473-0.
- Эконометрика : учебник / под ред. И. И. Елисеевой. 2-е изд. М. : Финансы и статистика, 2006. С. 576. ISBN 5-279-02786-3.
- Woodward W. A., Gray H. L. & Elliott A. C. Applied Time Series Analysis. CRC Press. 2012. С. 63–64.
- Hyndman Rob J., Athanasopoulos G. Forecasting: principles and practice. University of Western Australia, 2013. С. 14–32. ISBN-13: 978-0987507105.
- Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: Классификация и снижение размерности. М. : Финансы и статистика, 1989. 607 с.
- Shasha D. High Performance Discovery in Time Series // Springer. 2004. С. 3–7.
- Cowpertwait P. S. P., Metcalfe A. V. Introductory Time Series with R // Springer. 2009. С. 27–43.
- Елисеева И. И., Юзбашев М. М. Общая теория статистики : учебник / под ред. И. И. Елисеевой. 4-е изд., перераб. и доп. М. : Финансы и статистика, 2002. 480 с. ISBN 5-279-01956-9.

История статьи

Поступила в редакцию 31 октября 2017 г.

Принята к публикации 18 ноября 2017 г.

TOOLS FOR VISUALIZATION OF TIME SERIES IN SPACE RESEARCH

V. S. Maraev

Siberian Federal University, Krasnoyarsk, Russian Federation

Time series analysis is a key step in building a prediction model. That is why it is very important to consider the data from the various parties, because the analysis helps to identify the various features and options under consideration of the time series. This is especially important when analyzing the data of space research, as often, their analysis does not lead to a fairly clear pattern. Therefore, data visualization is a powerful tool in this stage of the prediction model. Very often, for the visualization of time series in space research, only linear graphs are used that can't represent the entire specifics of the series, so it is worthwhile to consider other methods and tools for visualizing space research data. This article analyzes the materials for time series visualization tools. The main visualization tools are considered, such as histograms, distribution density charts, box-and-whisker plots, heat maps, scatter and autocorrelation plots. These tools are demonstrated on examples of data obtained from space research. The advantages and disadvantages of various time series visualization tools in space research are revealed. The recommendations on the expediency of using these visualization tools in various situations are outlined. The corresponding conclusions are drawn on the basis of the conducted studies on the analysis of time series visualization tools in space research.

Keywords: visualization, time series, space research, histogram, box and whisker plots, heat map, autocorrelation plots.

References

1. George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel. Time Series Analysis: Forecasting and Control. Wiley Series in Probability and Statistics, 2013, pp. 92–161. ISBN 1118619064, 9781118619063.
2. Kendall M., Stewart A. Multivariate statistical analysis and time series. Moscow, Nauka Publ., 1976, 736 p.
3. Mishulina O. A. Statistical analysis and time series processing. Moscow, MIFI Publ., 2004, p. 38.
4. Evdokimov I. V., Mikhalev A. S., Novikov O. S., Sukhanov A. V. Application of free licenses for software development in Russia. International Journal of Applied and Fundamental Research, 2017, no. 6-1, pp. 33–36.
5. Evdokimov I. V., Yashchenkov K. G., Telkov A. Yu., Tataurov V. A. Expert methods for assessing the complexity of developing software projects. Economics and management systems management, 2017, vol. 24, № 2.2, pp. 272–276.
6. Evdokimov I. V., Domantsevich V.S., Konyhov V.A. THE USE OF ONLINE APPLICATIONS FOR PROJECT MANAGEMENT FOR PLANNING IN IT-MANAGEMENT. Modern Information Technologies, 2017, no. 25 (25), pp. 44–47.
7. International Center for Small Planets. Data Available from the Minor Planet Center. Available at: <http://www.minorplanetcenter.net/> (accessed 30.10.2017).
8. Center for Near-Earth Objects Dynamics. The AstDyS database [Electronic resource]. Available at: <http://hamilton.dm.unipi.it/astdys/index.php?pc=2.0&o=094> (accessed 30.10.2017).
9. International Meteor Organization. Visual Meteor Database [Electronic resource]. Available at: <https://www.imo.net> (reference date: 30.10.2017).
10. Chronology of manned spaceflight. LAUNCHLOG [Electronic resource]. Available at: <http://planet4589.org/space/log/launch.html> (accessed 30.10.2017).
11. The official open data storage of the Hubble Space Telescope. Hubble Space Telescope database [Electronic resource]. Available at: <http://www.nasa.gov/hubble/> (accessed 30.10.2017).
12. Ayvazyan S. A. Applied statistics. Fundamentals of Econometrics. Vol. 2. Moscow, Unity-Dana Publ., 2001, p. 432. ISBN 5-238-00305-6.
13. Magnus Ya. R., Katyshev P. K., Peresetsky A. A. Econometrics. The initial course. Moscow, Case Publ., 2007, p. 504. ISBN 978-5-7749-0473-0.
14. Econometrics. Ed. Eliseeva I. I. 2 nd ed. Moscow, Finances and Statistics Publ., 2006, p. 576. ISBN 5-279-02786-3.
15. Woodward W. A., Gray H. L. & Elliott A. C. Applied Time Series Analysis. CRC Press, 2012, pp. 63–64.
16. Hyndman Rob J., Athanasopoulos G. Forecasting: principles and practice. University of Western Australia, 2013, pp. 14–32. ISBN-13: 978-0987507105.
17. Ayvazyan S. A., Buchstaber V. M., Enyukov I. S., Meshalkin L. D. Applied Statistics: Classification and Dimension Reduction. Moscow, Finance and Statistics Publ., 1989, 607 p.
18. Shasha D. High Performance Discovery in Time Series. Springer, 2004, pp. 3–7.
19. Cowpertwait P. S. P., Metcalfe A. V. Introductory Time Series with R. Springer, 2009, pp. 27–43.
20. Eliseeva I. I., Yuzbashev M. M. General Theory of Statistics. Ed. I. I. Eliseeva. 4th edition, revised and enlarged. Moscow, Finance and Statistics Publ., 2002, 480 p. ISBN 5-279-01956-9.

Article history

Received 31 October 2017

Accepted 18 November 2017