

**Л. Ф. Вьюненко** (Санкт-Петербург, СПбГУПТД). **Фазовый сплайн-анализ как инструмент разведочного анализа данных.**

УДК 51-77

DOI [https://doi.org/10.52513/08698325\\_2024\\_31\\_1\\_1](https://doi.org/10.52513/08698325_2024_31_1_1)

*Резюме:* В работе фазовый сплайн-анализ рассматривается с точки зрения целей разведочного анализа данных. На примере данных о числе преступлений, совершенных с использованием информационно-телекоммуникационных технологий или в сфере компьютерной информации, показано, что фазовый сплайн-анализ может дополнить арсенал методов и инструментов разведочного анализа данных.

*Ключевые слова:* разведочный анализ данных, сглаживающий сплайн, фазовый анализ, ИТ-преступления.

Под разведочным анализом данных (англ. Exploratory Data Analysis, EDA) обычно понимается предварительное изучение основных свойств данных, нахождение в них наиболее общих зависимостей и тенденций, распределений и аномалий. Результат EDA — построение начальных моделей, часто с использованием инструментов визуализации.

Цели такого анализа сформулированы в классической работе [1]:

- максимальное «проникновение» в данные,
- выявление основных структур,
- выбор наиболее важных переменных,
- обнаружение отклонений и аномалий,
- проверка основных гипотез,
- разработка начальных моделей.

В практике EDA основными средствами являются изучение вероятностных распределений переменных, построение и анализ корреляционных матриц, факторный анализ, дискриминантный анализ, многомерное шкалирование. Арсенал методов и инструментов EDA постоянно расширяется и дополняется. В работе фазовый сплайн-анализ [2, 3] рассматривается как средство EDA, что иллюстрируется на примере данных о числе преступлений, совершенных с использованием информационно-телекоммуникационных технологий или в сфере компьютерной информации. «Цифровизация» преступности, то есть использование информационных технологий в качестве средства и орудия совершения преступления, является общемировой тенденцией последних лет [4]. Количественные данные о преступлениях такого вида стали доступны сравнительно недавно. В РФ в открытом доступе, в виде ежемесячных кратких характеристик состояния преступности, такие данные представлены с сентября 2019 г. со следующей детализацией: «... в том числе совершенных с использованием или применением: расчетных (пластиковых) карт, компьютерной техники, программных средств, фиктивных электронных платежей, сети «Интернет», средств мобильной связи» [5].

Данные, аккумулированные из ежемесячных характеристик в низкочастотные временные ряды, обработаны по технологии фазового сплайн-анализа. Основные этапы этой технологии — аппроксимация исходного ряда интерполирующим сплайн-

ном, построение фазовой траектории и фазового портрета. На первом этапе сплайн-аппроксимация шести перечисленных выше видов данных позволила обнаружить выбросы, обусловленные ошибками в данных. Построенные на втором этапе фазовые траектории и фазовые портреты выявили некоторые закономерности в динамике числа анализируемых преступлений. На рис. 1 в качестве примера представлен результат построения фазовых портретов процесса изменения числа преступлений, совершенных с использованием пластиковых карт, в период 2020–2023 гг.

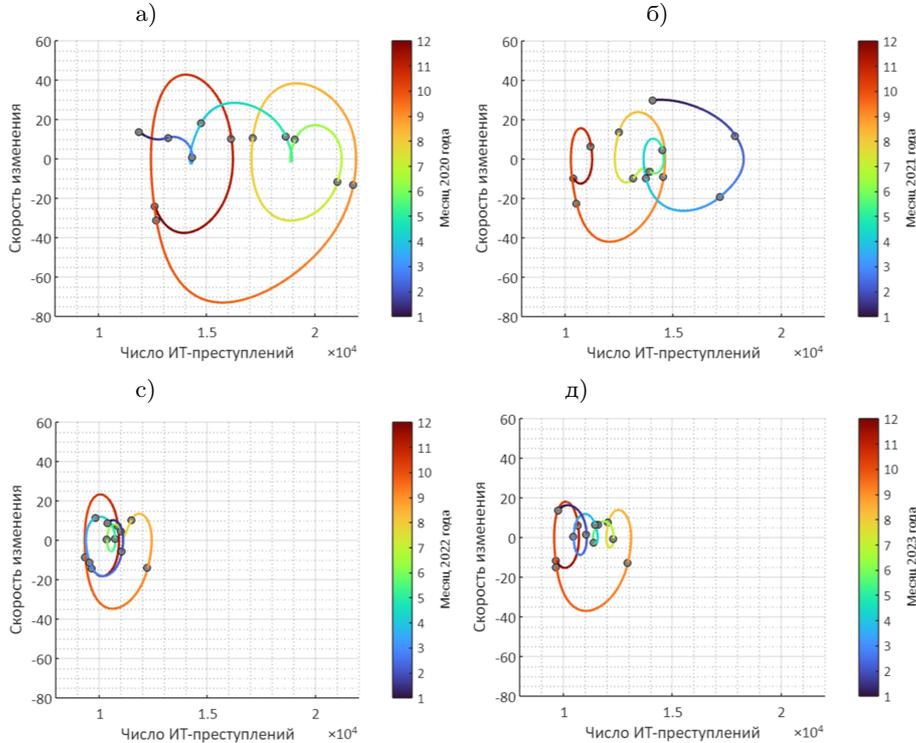


Рис. 1. Фазовые портреты процесса изменения числа преступлений, совершенных с использованием пластиковых карт в 2020 г. (а), 2021 г. (б), 2022 г. (в), 2023 г. (г).

Из рис. 1 видно, что число преступлений, совершенных с использованием пластиковых карт, имеет тенденцию к снижению. Фазовый портрет отражает некоторую «сезонность» в анализируемом процессе: число преступлений увеличивается в июле-августе и снижается к концу года. Нетипичный вид фазового портрета (б) можно объяснить тем, что в начале 2021 года было объявлено о принятии мер усиления контроля операций по пластиковым картам и изменении правил приема безналичных платежей в торговых точках, что спровоцировало всплеск активности мошенников.

Полученные в работе результаты показывают, что фазовый сплайн-анализ может дополнить арсенал методов и инструментов разведочного анализа данных.

#### СПИСОК ЛИТЕРАТУРЫ

1. Джон Тьюки. Анализ результатов наблюдений. Разведочный анализ. М.: Мир, 1981, 696 с. // *J.W. Tukey. Exploratory data analysis.* — Addison-Wesley Publishing Company, Reading, Mass., 1977, 704 p.
2. Gadasina L., Vyunenko L. Applying spline-based phase analysis to macroeconomic dynamics. — *Dependence Modeling*, 2022, 10(1), p. 207–214.

3. Вьюненко Л. Ф., Гадасина Л. В. Исследование динамики временных рядов сред-ствами фазового сплайн-анализа. — Обозрение прикл. и промышл. матем., 2022, т. 29, в. 3, с. 282–283. // *Vyunenko L. F., Gadasina L. V. Applying Phase Spline Analysis to Time Series Dynamics.* — OPPM Surveys Appl. Industr. Math., 2022, v. 29, is. 3, p. 282–283. (In Russian.)
4. Ляпин А. Е. Киберпреступность как новый объект статистического анализа. — Статистика и Экономика, 2021, т. 18(6), с. 4–16. // *Lyapin A. E. Cybercrime as a New Object of Statistical Analysis.* — Statistics and Economics. 2021, v. 18(6), p. 4–16. (In Russian.)
5. <https://мвд.рф/reports> — официальный сайт министерства внутренних дел российской федерации (дата обращения 01.07.2024). // <https://en.mvd.ru/> — The Ministry of Internal Affairs of the Russian Federation (Accessed 01.07.2024)

Поступила в редакцию  
07.VIII.2024

UDC 51-77

DOI [https://doi.org/10.52513/08698325\\_2024\\_31\\_1\\_1](https://doi.org/10.52513/08698325_2024_31_1_1)

*Vyunenko L. F.* (Saint Petersburg, SPbSUITD). **Phase Spline Analysis as a Tool for Exploratory Data Analysis.**

*Abstract:* In this paper, phase spline analysis is considered from the point of view of the exploratory data analysis goals. Using the example of data on the number of crimes committed using information and telecommunication technologies or in the field of computer information, it is shown that phase spline analysis can supplement the arsenal of methods and tools for exploratory data analysis.

*Keywords:* exploratory data analysis, smoothing spline, phase analysis, IT crimes.