

В работе рассмотрено применение комплексного подхода при исследовании бимолекулярных соединений с помощью метода флуоресцентной спектроскопии с временным разрешением [6].

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Cheng M., Cao L., Ning K. Microbiome Big-Data Mining and Applications Using Single-Cell Technologies and Metagenomics Approaches Toward Precision Medicine // Front. Genet. 2019. V. 10:972.
2. Яцков Н. Н., Скакун В. В., Апанасович В. В. Метод исследования флуоресценции молекулярных соединений с использованием алгоритмов интеллектуального анализа данных // Прикладные проблемы оптики, информатики, радиофизики и физики конденсированного состояния: Матер. 5-ой Междунар. науч.-практ. конф. НИУ «Ин-т прикл. физ. проблем им. А.Н. Севченко» БГУ, 2019. С. 125–127.
3. Яцков Н. Н., Апанасович В. В. Комплексный анализ данных при исследовании сложных биомолекулярных систем // Квантовая электроника: Матер. XII Междунар. конф. Мн.: БГУ, 2019. С. 282–283.
4. Яцков Н. Н. Интеллектуальный анализ данных: пособие. Мн.: БГУ, 2014. 151 с.
5. Yatskou M. Computer simulation of energy relaxation and -transport in organized porphyrin systemsю Ponsen & Looijen Printing Establishment. Wageningen. The Netherlands, 2001. 176 p.
6. Яцков Н. Н., Скакун В. В., Апанасович В. В. Метод обработки кинетических кривых затухания флуоресценции с использованием алгоритмов интеллектуального анализа данных // ЖПС. 2020. Т.87, №2. С. 311–322.

ОБРАБОТКА ГИСТОГРАММ ЧИСЛА ФОТООТСЧЕТОВ В ЭКСПЕРИМЕНТАХ ФЛУОРЕСЦЕНТНОЙ ФЛУКТУАЦИОННОЙ СПЕКТРОСКОПИИ

Н. Н. Яцков, В. В. Скакун, В. В. Апанасович

*Белорусский государственный университет, Минск, Беларусь
E-mail: yatskou@bsu.by*

Перспективным направлением исследований в флуоресцентной спектроскопии является анализ флуктуаций интенсивности флуоресценции во времени с использованием гистограмм числа фотоотсчетов. Ключевым элементом подхода является применение метода главных компонент для проверки гипотезы о разделимости кластеров многомерных экспериментальных данных. В работе исследуются возможности использования нормировки локального взвешенного сглаживания данных в пространстве первых двух главных компонент и метода независимых компонент для интерпретации независимых факторов, объясняющих вариацию в данных.

Ключевые слова: флуоресцентная флуктуационная спектроскопия; гистограмма числа фотоотсчетов; имитационное моделирование; метод главных компонент; нормировка локального взвешенного сглаживания; метод независимых компонент.

Методы флуоресцентной спектроскопии широко используются для изучения диффузии белков и их взаимодействий в растворах и живых клетках [1]. Перспективным направлением исследований является анализ флуктуаций интенсивности флуоресценции во времени с использованием гистограммы числа фотоотсчетов (ГЧФ) [2, 3]. Для обработки наборов ГЧФ экспериментальных измерений предложен комплексный подход на основе применения методов интеллектуального анализа данных [4]. Ключевым элементом подхода является использование метода главных компонент (МГК) для проверки гипотезы о разделимости набора многомерных экспериментальных данных на несколько поднаборов (соответствующих различным видам молекул) в пространстве первых главных компонент. Предполагается, что главные компоненты флуктуаций интенсивности флуоресценции обусловлены влиянием двух независимых факторов в исследуемых молекулярных системах – яркостью и концентрацией молекулярных соединений [2]. Вследствие чего ожидается сферообразный или эллипсоидальный кластер рассматриваемых данных в пространстве первых главных компонент. Однако, для ряда экспериментальных систем, как, например, для белков различных концентраций в растворах и клетках, наблюдается дугообразное облако, не объяснимое факторами яркости и числа молекул (рис. 1).

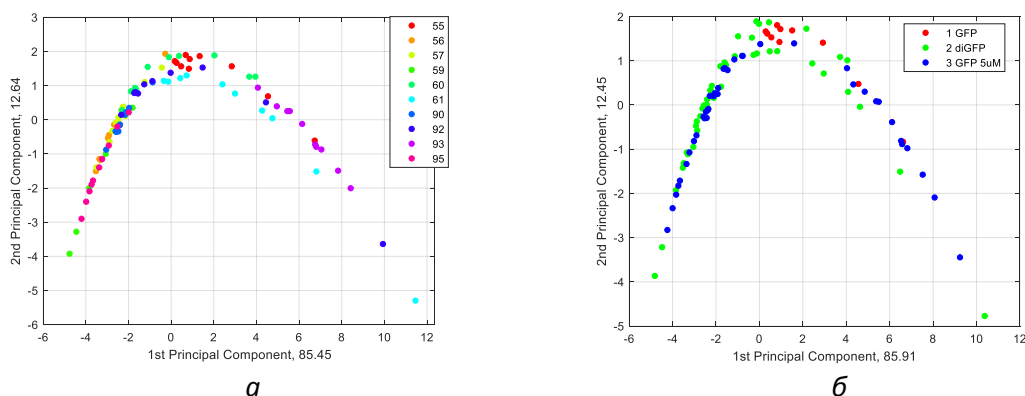


Рис. 1. Результаты анализа наборов экспериментальных данных с использованием метода главных компонент. *a* – мономер и димер зеленого флуоресцирующего белка (GFP и diGFP) в живых клетках, *б* – белки GFP и diGFP в лизатах клеток

Дугообразность облака данных позволяет предположить о наличии большего, чем два числа факторов или об их статистической взаимосвязи. Кроме того, дугообразность облака приводит к общему сжатию кластеров данных, поэтому, для эффективного применения МГК требуется устранение нелинейности данных путем проведения нормировки. Усовершенствованным подходом к анализу подобных данных является применение метода независимых компонент (Independent component analysis) [5], позволяющего более точно определить независимые компоненты или скрытые факторы, объясняющие вариацию в данных.

В работе, на примерах смоделированных наборов данных, исследуются возможности использования нормировки локального взвешенного сглаживания данных (LOWESS – Locally Weighted Scatterplot Smoothing) [6] в пространстве главных компонент и метода независимых компонент для интерпретации независимых факторов, объясняющих вариацию в экспериментальных данных.

Методы. *Нормировка локального взвешенного сглаживания данных.* Идея метода состоит в нахождении некоторой сглаживающей функции или нелинейной регрессионной модели, аппроксимирующей дугообразное облако экспериментальных данных, и последующего преобразования данных с использованием восстановленной регрессионной модели. *Метод независимых компонент.* Суть метода заключается в разложении наблюдаемого сигнала на аддитивные статистически независимые негауссовы подкомпоненты. *Имитационная модель.* Модель потока фотоотчетов представлена в [3] и позволяет смоделировать наборы ГЧФ флуктуаций флуоресценции молекулярных соединений, различающиеся средним числом молекул и их яркостью.

Результаты. Смоделированы наборы ГЧФ флуктуаций флуоресценции, представляющие три вида идеализированных систем молекул, характеризующиеся средним числом молекул $\langle N \rangle = 1$ в объеме регистрации и различной яркостью $\langle q \rangle = 5, 10$ и 30×10^4 . Внефокусное излучение характеризуется параметром $F = 0.8$. Интервал наблюдения для построения ГЧФ – 5×10^{-5} с. Проводилось моделирование со среднеквадратическим отклонением 0.1 и 0.2 от абсолютных значений параметров.

Вычислительные алгоритмы реализованы на языке программирования R с использованием функций `eig`, `lowess` и `fastICA` открытых библиотек R-функций проекта CRAN. Анализ данных выполнен на вычислительном сервере, основные характеристики которого – 12 ядерный процессор Intel i9 (3.9 GHz), 64 Gb RAM, 8 Tb HDD.

Применение нормировки локального взвешенного сглаживания позволило устранить эффект дуги и сжатия кластеров ГЧФ трех видов молекул, что расширяет использование результатов МГК и повышает эффективность последующего анализа и интерпретации кластеров данных (рис. 2). Достоинством МГК является сохранение геометрических расстояний между объектами в пространстве главных компонент.

Применение МНК с числом компонент 2 не позволило устранить эффект дуги и сжатия данных, что допускает наличие более чем 2 независимых компонентах в рассматриваемой системе (рис. 3а). Использование МНК с числом компонент 3 и более устраняет эффект дуги, однако требует интерпретации дополнительных независимых компонент (рис. 3б-г). Для объяснения независимых компонент рассмотрен набор дан-

ных, состоящий из 2 видов молекул. Применение МНК с числом компонент 2 также не позволило устранить эффект дуги и сжатия, что объясняется наличием более 2 компонент в системе (рис. 4а). Использование МНК с числом компонент 3 устраняет эффект дуги и позволяет логически объяснить наблюдаемые компоненты (рис. 4б-г).

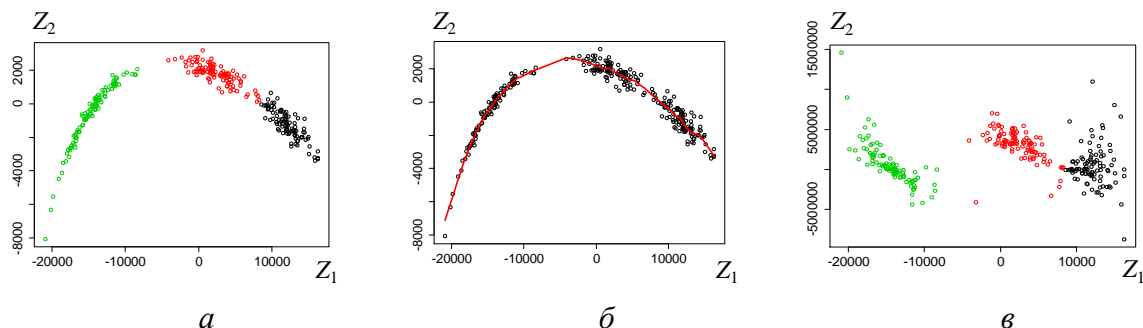


Рис. 2. Результаты применения метода главных компонент (а) и нормировки локального сглаживания (б и в) к набору данных трех видов молекул

Компоненты 1 и 2 действительно являются независимыми – кластеры данных образуют единое сферическое не разделяющееся облако (рис. 4б). Компонента 3 в комбинации с компонентами 1 и 2 позволяет различить кластеры, причем для отдельных кластеров видов молекул наблюдается линейная зависимость компонент (рис. 4в и г).

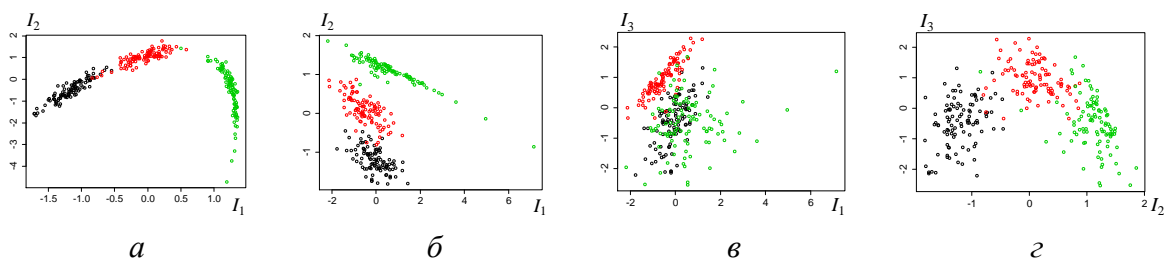


Рис. 3. Результаты применения метода независимых компонент с числом компонент 2 (а) и 3 (б-г) к набору данных трех видов молекул

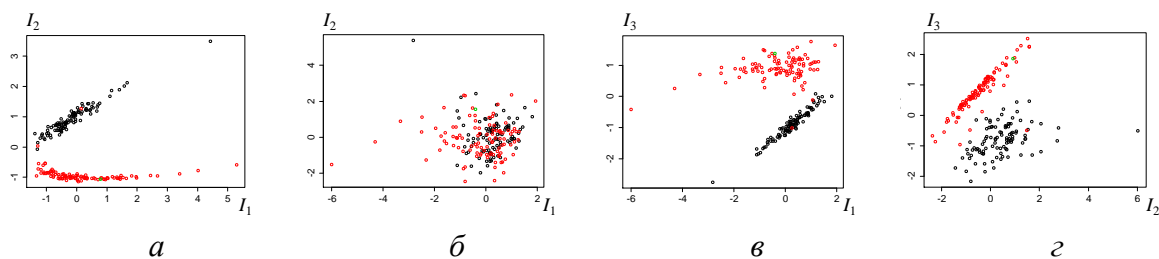


Рис. 4. Результаты применения метода независимых компонент с числом компонент 2 (а) и 3 (б-г) к набору данных двух видов молекул

Последнее позволяет предположить о наличии трех факторов/компонент, связанных с числом и яркостью молекул: 1 – яркость молекул первого вида, 2 – яркость молекул второго вида, 3 – число молекул первого или второго вида. Таким образом, диаграмма компонент

1-2 подтверждает независимость яркостей молекул двух систем, а диаграммы компонент 1-3 и 2-3 свидетельствует о линейной связи яркости определённого типа молекул и числа молекул данного вида.

Выводы. Разработаны и успешно подтверждены методы локального взвешенного сглаживания и независимых компонент для нормировки данных и определения независимых факторов, объясняющих флуктуации флуоресценции молекулярных соединений. Полученные результаты позволяют сделать следующие выводы:

1) Для отображения данных в пространстве главных компонент можно использовать нормирование локального взвешенного сглаживания.

2) Для молекулярных систем различной яркости независимыми факторами являются яркости отдельных видов молекул. Факторы числа молекул и их яркость являются линейно-зависимыми.

3) Дугообразность облака данных нескольких видов молекул в пространстве главных компонент объясняется эффектом суммирования яркостей отдельных видов молекул и их линейной-зависимостью от числа молекул.

Предложенные алгоритмы обработки данных могут использоваться в методе комплексного анализа флуктуаций интенсивности флуоресценции молекулярных соединений с применением алгоритмов интеллектуального анализа данных для определения структурного состава олигомеров белков, а также в качестве дополняющих вычислительных процедур к классическим методам анализа РСН [2] и FIDA [3].

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Kitamura A., Kinjo M. State-of-the-Art Fluorescence Fluctuation-Based Spectroscopic Techniques for the Study of Protein Aggregation // *Int. J. Mol. Sci.* 2018. Vol. 19. N4. P. 1–18.
2. Chen Y., Müller J. D., So P. T., Gratton E. The Photon Counting Histogram in Fluorescence Fluctuation Spectroscopy // *Biophys. J.* 1999. Vol. 77. P. 553–567.
3. Skakun V. V., Digris A. V., Borst J.W., Visser A.J. Global Analysis of Autocorrelation Functions and Photon Counting Distributions // *Front. Biosci.* 2011. Vol. 3, P. 489-505.
4. Яцков Н. Н., Скакун В. В., Апанасович В. В. Метод комплексного анализа спектров флуктуации флуоресценции молекулярных соединений // Прикладные проблемы оптики, информатики, радиофизики и физики конденсированного состояния: Матер. 5-ой Междунар. науч.-практ. конф. НИУ «Ин-т прикл. физ. проблем им. А.Н. Севченко» БГУ, 2019. С. 122–124.
5. Hyvarinen A., E. Oja E. Independent Component Analysis: Algorithms and Applications // *Neural Networks*, 2000. Vol. 13. №4-5. P. 41–430.
6. Cleveland W. S. Robust locally weighted regression and smoothing scatterplots. // *J. Am. Stat. Ass.* 1979. Vol. 74. P. 829–836.