

Министерство образования и науки Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г.ЧЕРНЫШЕВСКОГО»

Кафедра радиофизики и нелинейной динамики

Влияние потери данных на диагностику сложных режимов динамики

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

студентки 2 курса 241 группы
направления 03.04.03 «Радиофизика»
физического факультета
Дрозденко Татьяны Александровны

Научный руководитель
г.н.с., д.ф.-м.н., профессор
доцент, к.ф.-м.н., доцент

А.Н. Павлов

О.Н. Павлова

Зав. кафедрой
д.ф.-м.н., профессор

В.С. Анищенко

Саратов 2018 год

ВВЕДЕНИЕ

Сигналы, записанные в ходе экспериментальных исследований, часто содержат различные сбойные участки (артефакты), появление которых может быть связано как с анализируемым объектом, так и с регистрирующим оборудованием. Наличие таких фрагментов данных усложняет анализ динамики исследуемой системы и может приводить к ошибочной диагностике режима ее функционирования. Характерным примером служат нейрофизиологические исследования, где различного рода артефакты присутствуют на участках, составляющих до 90% записи электроэнцефалограммы [1].

Стандартным приемом очистки экспериментальных данных от помех является их фильтрация, и за последние годы в этой области был предложен ряд новых решений, использующих, в частности, вейвлет-преобразование [2, 3]. Соответствующие фильтры способны устранять локализованные помехи, удаление которых в рамках классических методов фильтрации является чрезвычайно сложной задачей [4–6]. Тем не менее, этот прием далеко не всегда оказывается эффективным – если частотные диапазоны сигнала и помехи перекрываются, то фильтрация приводит к существенным искажениям сигнала. По этой причине широко используется подход, состоящий в удалении сбойных участков из экспериментальной записи, и последующий анализ проводится для сигнала, составленного из фрагментов, не содержащих артефактов. Если при этом исключается большое число сравнительно малых по длительности участков, то такая процедура может существенно влиять на корреляционные характеристики сигнала. В частности, в работе [7] отмечалось, что потеря данных особенно сильно сказывается при анализе анти-коррелированных процессов.

Тем не менее, во многих областях науки и техники невозможно избежать появления артефактов, поэтому необходимо иметь представление о степени влияния эффектов потери данных на диагностику режима функционирования системы.

Цель данной выпускной квалификационной работы – установить возможность диагностики режима динамики в случае экстремальной потери данных, составляющей до 90% от исходного объема выборки. Дополнительная сложность решаемой задачи связана с удалением большого числа сегментов малой длительности.

Материалы исследования. Учитывая, что для экспериментальных данных часто характерна нестационарность, отражающая изменение характеристик системы во времени, в качестве инструментов исследования были выбраны **методы**, которые относятся к числу универсальных приемов цифровой обработки сигналов, применимых как в случае стационарных процессов, так и динамики систем с меняющимися во времени характеристиками. Такими методами, рассмотренными в данной работе, являются флуктуационный анализ относительно тренда (“detrended fluctuation analysis”, DFA) [8], метод мультифрактального анализа, применяющий непрерывное вейвлет-преобразование [9, 10], метод эмпирических мод и многомасштабный анализ, использующий дискретное вейвлет-преобразование [11, 12].

Выпускная квалификационная работа содержит введение, три главы (1. Методы исследования; 2. Основные результаты ранее проводившихся исследований; 3. Результаты исследований), заключение и список использованных источников. Общий объем работы 60 стр.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Методы исследования. Одним из методов, применяемых в данной работе, является метод флуктуационного анализа (метод DFA) [8]. В отличие от спектральной плотности, которую можно вычислить непосредственно по временному ряду, метод DFA предусматривает построение одномерного случайного блуждания и его последующий среднеквадратичный анализ. В соответствии с работами, где этот метод был впервые предложен и протестирован, случайное блуждание аппроксимируется кусочно-линейной

функцией (после процедуры деления на сегменты, аппроксимация проводится в пределах каждого сегмента), и стандартная ошибка аппроксимации вычисляется в зависимости от длины сегмента.

Другим применяемым методом является мультифрактальный анализ на основе вейвлет-преобразования [9, 10]. Практическая реализация этого метода предусматривает следующие вычисления. Сначала вычисляется непрерывное вейвлет-преобразование. Затем производится выделение всех линий скелетона, которые представляют собой линии локальных максимумов и минимумов коэффициентов вейвлет-преобразования, вычисленных при фиксированных значениях параметра масштаба. Несмотря на возможность провести расчеты показателя Гельдера непосредственно по отдельным линиям, этот подход является очень нестабильным и обычно приводит к неправильным результатам, вызванным, например, присутствием соседних сингулярностей. По этой причине проводится статистический анализ сингулярностей с использованием статистических функций. После их вычисления далее проводится расчет показателей Гельдера $h(q)$ и спектра сингулярностей $D(h)$.

Другими методами, применяемыми в данной работе, являлись метод эмпирических мод и метод многомасштабного анализа на основе дискретного вейвлет-преобразования.

Основные результаты ранее проводившихся исследований. Проблема потери данных известна давно, однако в литературе отсутствует систематический анализ влияния сбойных участков на результаты диагностики режимов динамики. Одной из работ, где подобный анализ был выполнен, является статья [7]. Ее авторы рассмотрели стохастические процессы с различными значениями скейлингового показателя α метода DFA. При этом они проанализировали, как объем потерянных данных влияет на процессы с разными корреляционными характеристиками. Был разработан алгоритм, который случайным образом выделяет в сигнале участки (которые считаются сбойными и удаляются), при этом размер исключаемого участка и расстояние между остающимися «чистыми» сегментами распределены по случайному

закону. Более того, авторы работы [7] рассмотрели разные варианты случайных распределений, чтобы сделать обобщающие выводы.

В частности, был приведен пример анализа анти-коррелированных сигналов, характеризующихся значениями $\alpha < 0.5$ при наличии большого объема выборки. Средний размер удаляемого участка составлял 10 отсчетов. Были рассмотрены 4 значения показателя скейлинга, для которых проводился подробный анализ. Чем меньше α , тем сильнее сказывается удаление сегментов. При $\alpha = 0.4$ возможность диагностировать режим (то есть определить величину α) сохраняется не только при удалении сравнительно небольшого объема выборки (около 10% от исходного временного ряда), но и при экстремальной потере данных (95% от объема выборки). Таким образом, располагая сигналом, собранным из малых сегментов, суммарный размер которых не превышает 5% исходного объема выборки, можно правильно оценить α , то есть правильно охарактеризовать поведение спектральной плотности мощности и корреляционной функции. Чем меньше величина α , тем сложнее провести правильную диагностику. Например, если $\alpha = 0.1$, то даже потеряв 10% данных, можно получить вместо 0.1 значение 0.5, то есть погрешность в 400%.

В случае коррелированной динамики ситуация меняется кардинально – как при малом объеме потерянных данных, так и при экстремальной их потере (90% от объема выборки) результаты вычисления показателя скейлинга α почти не меняются. Таким образом, рассмотренные сигналы могут быть использованы для правильного определения их корреляционных характеристик.

Аналогичные выводы можно сделать для данных сравнительно малого объема (4000 отсчетов). По-прежнему, в случае коррелированной динамики эффект потери данных выражен слабее – он почти не виден при выборе показателя скейлинга, равного 0.9, даже при удалении 50% от исходной выборки. В случае экстремальной потери данных результаты начинают отклоняться от ожидаемых значений при удалении участков средней

длительностью 10 отсчетов. Однако, если рассматривать уменьшение значений показателя скейлинга α , то влияние потери данных возрастает.

Результаты исследований. Вначале решалась задача проверки результатов ранее выполненных работ на примерах, рассмотренных в статье [7]. С этой целью был выбран стохастический сигнал (1/ f -шум) и применен метод мультифрактального анализа. Показатели Гёльдера $h(q)$ в этом случае не зависят от q , и справедливо равенство $h(q) = H = 1$. Ошибки вычислений зависят от объема выборки. Если анализировать сигналы большой длительности, то ошибка вычисления будет мала (менее 1%). Для малого объема данных ошибка возрастает, но не сильно. В частности, при анализе временных рядов, содержащих порядка 3000 отсчетов, величины $h(q)$ оценивались с ошибкой менее 1% в диапазоне больших флуктуаций ($q > 0$), а в диапазоне малых флуктуаций (то есть для $q < 0$) максимальная ошибка составляла 3% - 4%.

Чтобы определить, как потеря различных участков сигнала влияет на проводимые оценки, осуществлялось удаление сегментов различной длины (от 1 до 10 отсчетов), при этом длительность удаляемых сегментов выбиралось случайно, а интервалы времени между исключаемыми сегментами выбирались в соответствии с распределением Гаусса. Было показано, что даже при условии экстремальной потери данных (до 90% от первоначального сигнала), величина ошибки E не превышает 10%. Рост ожидаемых значений $h(q)$ наблюдается на малых масштабах ($q < 0$), а на больших масштабах ($q > 0$) осуществляется недооценка значений $h(q)$, что приводит к уменьшению наклона зависимости скейлинговых показателей $\tau(q)$ с уменьшением величины μ - объема оставшихся данных после удаления сегментов (в % от исходного объема выборки). При рассмотрении абсолютных значений, ошибка будет меньше для положительных q . Тем не менее, в случае удаления около 60% от объема данных (то есть для $\mu=40\%$), отклонение от ожидаемых значений показателей Гёльдера сопоставимо с ошибками вычисления, которые наблюдаются для

исходных данных сравнительно малой длительности. Полученные результаты согласуются с выводами работы [7], которые были сделаны на основе применения метода DFA. Аналогичное соответствие достигалось и для других примеров стохастических процессов (шумов с разными корреляционными характеристиками). Например, шумы с показателем Гёльдера $H > 0.5$ не демонстрировали сильных изменений вычисляемых характеристик, если удалялось больше половины объема исходных данных. Сигналы с $H < 0.5$ проявляли более выраженные различия спектра сингулярностей даже при устранении небольшой части данных.

Помимо стохастических процессов осуществлялся анализ сложной динамики систем, демонстрирующих режимы динамического хаоса, для которых внутренняя структура хаотических аттракторов, отражаемая в последовательности времен возврата в секущую Пуанкаре, может быть разрушена удалением фрагментов данных. Чтобы сопоставить результаты для разных режимов динамики, в качестве модельной системы нами выбрана модель двух связанных систем Рёсслера, которая записывается в виде системы обыкновенных дифференциальных уравнений вида

$$\begin{aligned}\frac{dx_{1,2}}{dt} &= -\omega_{1,2}y_{1,2} - z_{1,2} + \gamma(x_{2,1} - x_{1,2}), \\ \frac{dy_{1,2}}{dt} &= \omega_{1,2}x_{1,2} + ay_{1,2}, \\ \frac{dz_{1,2}}{dt} &= b + z_{1,2}(x_{1,2} - c),\end{aligned}$$

где управляющие параметры a , b , c определяют режим динамики, γ характеризует силу связи. Частоты $\omega_{1,2}$ выбраны разными ($\omega_{1,2} = \omega_0 \pm \delta$) за счет введения расстройки по частоте. Были рассмотрены следующие значения параметров: $a = 0.15$, $b = 0.20$, $\gamma = 0.02$, $\omega_0 = 1.0$. Остальные параметры менялись, и на плоскости этих параметров рассматривались переходы между различными режимами динамики.

Взаимодействие приводит к разнообразию наблюдаемой динамики. Для синфазных аттракторов, которые характеризуются разностью фаз колебаний $x_1(t)$ и $x_2(t)$ близкой к нулю при расстройке $\delta = 0$, наблюдается довольно простая зависимость $D(h)$, близкая к монофрактальной структуре времен возврата в секущую Пуанкаре хаотического аттрактора CA_0 , который возникает в результате перехода к хаосу через последовательные бифуркации удвоения периода. Хаотический аттрактор несинфазного семейства CA_1 , который появляется в ходе последовательности бифуркаций удвоения периода колебаний, характеризуется разностью фаз для субгармоник, равной 2π . Для него характерен мультифрактальный спектр сингулярностей. Но этот бифуркационный сценарий (последовательность удвоений периода) приводит к антикоррелированной последовательности времен возврата.

С ростом параметра c , происходит объединение аттракторов CA_0 и CA_1 , и появляющийся в результате этого режим CA_{hyper} характеризуется двумя положительными показателями Ляпунова. При изменении расстройки наблюдается переход к квазипериодическим колебаниям $4T^2$ в результате бифуркации предельного цикла $4C_0$ или (с ростом c) к асинхронным хаотическим колебаниям (CA_i).

Рассмотрим, как потеря данных влияет на характеристики различных аттракторов, вычисляемые по последовательностям времен возврата в секущую Пуанкаре. Проанализируем вначале «синфазный» хаотический режим CA_0 . Пропуск части фрагментов данных сильно меняет значения показателей Гельдера. При удалении большого объема данных, положение спектра сингулярностей вдоль оси h приближается к значению $h = 0.5$, которое соответствует некоррелированной динамике времен возврата. Учитывая, что исходный спектр сингулярностей располагается вблизи $h = 0$, описание эффекта пропуска данных целесообразнее проводить в терминах абсолютной ошибки, а не относительной.

Даже потеря 10% данных смещает показатели Гёльдера на величину, примерно равную 0.2. Это согласуется с основными выводами работы [7], где продемонстрирован более сильный эффект потери данных на анти-коррелированную динамику. Изменения для несинфазного хаотического режима динамики (аттрактора CA_1) являются похожими на случай синфазного режима колебаний CA_0 за исключением более выраженной степени мультифрактальности аттрактора CA_1 . В рассмотренных случаях, то есть для аттракторов CA_0 и CA_1 , потеря данных приводила к похожим изменениям в случаях слабых и сильных сингулярностей, которые ассоциируются с отрицательными и положительными значениями q , соответственно. Анализ гиперхаотических колебаний CA_{hyper} , которые появляются после объединения аттракторов CA_0 и CA_1 , выявил отличия эффектов потери данных. Изменения показателей Гёльдера при положительных q похожи на изменения, которые были ранее. Но относительное смещение значений h является довольно малым в случае слабых сингулярностей, ассоциирующихся с отрицательными q . Абсолютная ошибка расчета характеристик составляет около 0.2, что превышает ошибку для $1/f$ -шума с удалением 90% исходных данных. Таким образом, мы видим, что разные сложные режимы динамики характеризуются разными эффектами потери данных. Однако общий вывод о более сильных изменениях статистических характеристик для анти-коррелированных последовательностей времен возврата справедлив независимо от анализируемого типа сложных колебаний.

Несинхронные режимы динамики, которые возникают при больших расстройках по частоте, демонстрируют другое поведение, которое характеризуется большей шириной спектра сингулярностей и показателями Гёльдера $h(q) > 1$. Следует выделить два обстоятельства: 1) Значение ошибки E меньше, чем в случае анти-коррелированной динамики синхронных колебаний, если происходит небольшая потеря данных. 2) Анализируемый динамический режим характеризуется шириной спектра сингулярностей, которая что

примерно в 5 раз больше по сравнению с аттрактором CA_1 и примерно в 30 раз больше по сравнению с аттрактором CA_0 . По этой причине относительные изменения статистических характеристик значительно меньше выражены для рассмотренных последовательностей времен возврата с положительными корреляциями, которые ассоциируются с квазипериодическим режимом $4T^2$. Такой же вывод справедлив для хаотических несинхронных колебаний CA_1 , которые возникают при разрушении квазипериодических колебаний $4T^2$ с увеличением параметра s . Эффекты потери данных в этих случаях менее выражены, если анализировать относительное смещение спектра сингулярностей. Это показано на рисунке 1, где смещения спектра сингулярностей нормированы на их ширину. В случае синхронных колебаний, которые ассоциируются с анти-коррелированной статистикой времен возврата, относительные изменения спектра сингулярностей превосходят диапазон изменения показателей Гельдера даже при потере около 10% данных. Для несинхронных колебаний, соответствующих последовательностям времен возврата с положительными корреляциями, такие смещения наблюдаются только в случае экстремальной потери данных (80%–90% от исходной выборки).

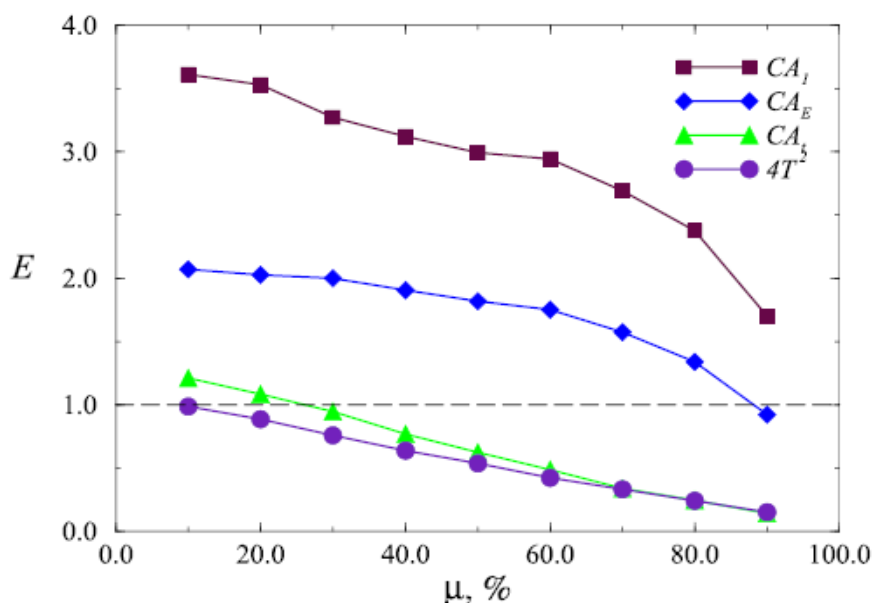


Рисунок 1 – Смещения спектров сингулярностей, нормированные на их ширину, в зависимости от значения μ

Таким образом, мы можем сделать вывод о том, что количественное описание сложных динамических режимов с удаленными фрагментами данных сильно зависит от статистики времен возврата. Сделанные выводы были подтверждены с помощью других методов исследования, которые позволили получить аналогичные результаты.

ЗАКЛЮЧЕНИЕ

В работе проведен анализ, каким образом потеря фрагментов экспериментальных данных, вызванная наличием различных артефактов и сбойных участков, ошибками записи или внезапными изменениями условий эксперимента влияет на возможность количественного описания режима функционирования. Были проанализированы различные типы сигналов, включая стохастические процессы (например, $1/f$ -шум), представляющий собой сравнительно простой пример, поскольку его спектр сингулярностей состоит из одной точки, и особенности скейлинга описываются одним показателем Гельдера. При этом были подтверждены выводы работы [7].

Проведенный анализ сложных динамических режимов, генерируемых моделью связанных хаотических осцилляторов, подтвердил значительно более высокую чувствительность процессов с анти-корреляциями к удалению фрагментов данных. Отметим, что для хаотической динамики эффекты потери данных являются более выраженными, чем для стохастических процессов.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- [1] Tuckwell, H. C. Introduction to theoretical neurobiology / H. C. Tuckwell. – Cambridge: Cambridge University Press, 1988.
- [2] Meyer, Y. Wavelets: Algorithms and applications / Y. Meyer. – Philadelphia: S.I.A.M., 1993.
- [3] Vetterli, M. Wavelets and subband coding / M. Vetterli, J. Kovacevic. – NJ: Prentice Hall, 1995.

- [4] Akansu, A. N. Multiresolution signal decomposition: transforms, subbands and wavelets / A. N. Akansu, R. A. Haddad. – San Diego: Academic Press, 2001.
- [5] Астафьева, Н. М. Вейвлет–анализ: основы теории и примеры применения / Н. М. Астафьева // Успехи физических наук. – 1996. – Т. 166, №11. – С. 1145–1170.
- [6] Cohen, A. Wavelets and multiscale signal processing / A. Cohen, R. Ryan. – London: Chapman and Hall, 1995.
- [7] Ma, Q. D. Y. Effect of extreme data loss on long–range correlated and anticorrelated signals quantified by detrended fluctuation analysis / Q. D. Y. Ma, R. P. Bartsch, P. Bernaola–Galvan, M. Yoneyama, P. Ch. Ivanov // Phys. Rev. E. – 2010. – Vol. 81. – P. 031101.
- [8] Peng C.–K. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series / C.–K. Peng, S. Havlin, H. Stanley, A. Goldberger // Chaos. – 1995. – Vol. 5,– P. 82–87.
- [9] Muzy, J. F. Multifractal formalism for fractal signals: the structure– function approach versus the wavelet–transform modulus–maxima method / J. F. Muzy, E. Bacry, A. Arneodo // Phys. Rev. E. – 1993. – Vol. 47. – P. 875–884.
- [10] Muzy, J. F. The multifractal formalism revisited with wavelets / J. F. Muzy, E. Bacry, A. Arneodo // Int. J. Bifurcation Chaos. – 1994. – Vol. 4, №2. – P. 245–302.
- [11] Mallat, S. G. A wavelet tour of signal processing / S. G. Mallat. – New York: Academic Press, 1998.
- [12] Addison, P. S. The illustrated wavelet transform handbook: applications in science, engineering, medicine and finance / P. S. Addison. – Bristol ; Philadelphia: IOP Publishing, 2002.