УДК 004.896

А.В. Замятин, П.М. Острасть, Д.А. Мурзагулов

# Подход к обнаружению аномалий технологических сигналов c использованием классификации и вейвлет-преобразований

**Замятин Александр Владимирович**, докт. техн. наук, доцент, заведующий кафедрой теоретических основ информатики ТГУ, Томск, тел: 89069576374, [zamyatin@mail.tsu.ru](mailto:zamyatin@mail.tsu.ru).

**Острасть Павел Михайлович**, к.т.н., начальник отдела корпоративной сети – заместитель начальника управления ИТ ОАО «Востокгазпром», Томск, [op@vostokgazprom.ru](mailto:op@vostokgazprom.ru)

**Мурзагулов Дамир Альбертович**, ведущий инженер автоматизированных систем управления технологическими процессами группы компаний Imperial Energy Ltd, аспирант кафедры теоретических основ информатики ТГУ, Томск, тел: 89138177056, [damir.murzagulov@imperialenergy.ru](mailto:damir.murzagulov@imperialenergy.ru).

*Работа выполнена в ходе реализации комплексного проекта при финансовой поддержке Министерства образования и науки Российской Федерации (договор от 1 декабря 2015 г. № 02.G25.31.0130) в рамках постановления Правительства Российской Федерации № 218 от 9 апреля 2010 года.*

**Аннотация**

Уровень развития современной ИТ-инфраструктуры промышленных предприятий позволяет осуществлять сбор и хранение технологической информации, тем самым открывая возможности применения интеллектуальных систем анализа данных. В работе рассматривается задача обнаружения аномалий в технологических сигналах в целях повышения безопасности и экономической эффективности промышленных предприятий. Для обнаружения аномалий предлагается комбинация методов машинного обучения и вейвлет-преобразований. Рассмотрена специфика технологических сигналов, которую необходимо учесть, а также преимущества вейвлет анализа перед известными методами обработки цифровых сигналов. Разработан подход к обнаружению аномалий и проведена его предварительная апробация на реальных технологических сигналах.

**Ключевые слова**: *технологические сигналы, машинное обучение, вейвлет-преобразование*

## Введение

Современная промышленная индустрия характеризуется активным переходом к этапу цифровой трансформации - помимо классических автоматизированных систем управления технологическими процессами (АСУ ТП), появляются и внедряются системы интернета вещей (IoT), больших данных (Big Data), машинного обучения [1-5]. Внедрение подобных автоматизированных систем поддержки принятия решения в основном направлено на две группы задач. К первой относятся вопросы повышения безопасности производства, включающие анализ и предупреждение аварий. Ко второй группе следует отнести экономические эффекты от оптимизации режимов потребления ресурсов, а также прогнозирование возможных поломок. Для решения обеих групп задач широко применяются методы интеллектуального анализа к массивам производственных данных [6]. При этом особое место в целях обеспечения надежности производства занимает задача обнаружения аномалий в данных, так как именно они зачастую являются предвестниками аварий, сбоев оборудования и нештатных ситуаций. Применительно к системам мониторинга техногенных объектов под аномалиями могут восприниматься любые события, нарушающие регламентное протекание контролируемого технологического процесса [7].

Основным источником производственных данных являются технологические сигналы, поступающие в АСУ ТП с различных измерительных датчиков (датчиков температуры, давления, вибрации и т.п.) [8]. Как правило, технологические сигналы имеют сложную структуру, они включают большое количество компонент, содержат локальные особенности различной формы и временной протяженности, а также шумовые факторы различной природы. Именно с этими факторами связана основанная сложность автоматического обнаружения аномалий в технологических сигналах традиционными статистическими методами. При использовании статистического анализа [9] требуется исследование характеристик процесса, построение его профиля (модели), который затем сравнивается с реальным поведением. Такой подход не обеспечивает устойчивости к большому разнообразию технологических сигналов и быстрой динамике процессов, что затрудняет эффективное применение статистических методов на практике.

Широкое применение в задачах обнаружения аномалий нашли методы машинного обучения [10]. К достоинствам методов машинного обучения относится их простая параметризуемость, возможность задавать различные метрики и количество степеней свободы модели. К алгоритмам машинного обучения, применяющихся для обнаружения аномалий, относится алгоритм k-means [11], изолирующий лес [12], метод опорных векторов [13]. В зависимости от наличия априорной информации об аномалиях могут быть выбраны разные тактики обучения - контролируемая или неконтролируемая классификация. Контролируемая классификация основана на предположении о том, что штатное поведение системы может определяться одним или несколькими известными классами сигналов, а фрагменты сигнала, не принадлежащие ни к одному из классов, являются аномалией. Не контролируемая классификация (кластеризация) предполагает группировку похожих фрагментов сигнала в кластеры и не требует знаний о свойствах возможных отклонений. При этом выявление аномалий строится на следующем предположении - регулярные данные образуют большие плотные кластеры, а аномальные – маленькие и разрозненные. В работе рассматривается задача обнаружения аномалий в технологических сигналах с использованием неконтролируемой классификацией.

**Предлагаемый** **подход**

Процедура неконтролируемой классификации сигналов требует построения информативного признакового пространства, в котором будет возможно разделение аномальных участков от участков со штатным протеканием процесса. Сложность построения признакового пространства для технологических сигналов заключается в том, что аномальные эффекты могут характеризоваться не только явными амплитудными изменениями, но и изменением частотно-временных свойств сигнала.

Наибольшую сложность для частотно-временного (спектрального) анализа представляют нестационарные во времени сигналы, которые наиболее часто встречаются на практике. Как правило, такие сигналы состоят из близких по времени, коротких высокочастотных компонентов и долговременных, близких по частоте низкочастотных компонентов. Для анализа таких сигналов необходим метод, способный обеспечить хорошее разрешение и по частоте и по времени. Первое необходимо для локализации низкочастотных составляющих, второе - для выделения компонентов высокой частоты. Одним из наиболее эффективных методов спектрального преставления сигналов является вейвлет-преобразование [14]. Основная область применения вейвлет преобразований – анализ и обработка сигналов, нестационарных во времени или неоднородных в пространстве сигналов. На сегодняшний день применение вейвлет анализа нашло широкое прикладное применение в распознавании человеческой речи, в задачах классификации биомедицинских данных, для фильтрации и сжатия данных [15-21].

Предлагается для обнаружения аномалий формировать признаковое пространство на основе вейвлет-преобразований сигнала, совместно с показания сигнала. Данный подход обоснован двумя факторами. Во-первых, вейвлет-преобразование предоставляет хорошее разрешение в частотно-временной области и способно локализовать аномалии в этой области. Во-вторых, исходные показания характеризуют аномалии связанные с амплитудными изменениями. Другими словами исходный анализируемый сигнал, представленный в виде дискретного ряда  где  - показания технологического параметра определенные в моменты времени  описывается признаковым множеством:

, (1)

где,  - вейвлет-преобразование сигнала .

Цель работы состоит в построении модели, которая разбивает множество  на подмножества  (кластеры) и для каждого кластера ставит в соответствие показатель аномальности . При этом для каждого показания  присваивается метка -1, если значение относится к аномалиям, и 1 – к регулярным значениям.

## Вейвлет-преобразование в задачах анализа сигналов

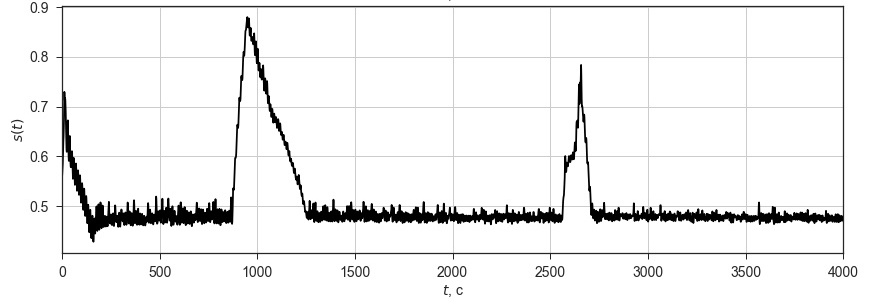
Технология вейвлет-преобразования является обобщением спектрального анализа и представляет собой альтернативу широко используемому классическому преобразованию Фурье [22]. Преобразование Фурье (ПФ) представляет сигнал, заданный во временной области, в виде разложения по ортогональным базисным функциям (синусоидам) с выделением частотных компонентов. Недостаток ПФ заключается в том, что частотные компоненты не могут быть локализованы во времени, обусловливая его применимость только к анализу стационарных сигналов, в то время как многие реальные сигналы имеют сложные частотно-временные характеристики. Основное отличие вейвлет-преобразования от ПФ заключается в том, что для преобразования исходного сигнала используется в качестве базисных функций вейвлеты (от англ. «wavlet» - короткая волна). Вейвлет это обобщенное название семейства математических функций определённой формы, которые локализованы по времени и частоте, и в которых все функции получаются из одной базовой (порождающей) функции путем ее сдвигов и растяжения по оси времени [23]. Существует два вида реализации вейвлет-преобразования - дискретное вейвлет преобразование (ДВП) и непрерывное вейвлет преобразование (НВП). В задачах фильтрации и сжатия данных преимущественно применяется ДВП, для проведения анализа сигналов интерес представляет НВП.

Непрерывное вейвлет-преобразование определяется как сумма по всему времени сигнала, умноженного на масштабируемые, сдвинутые версии вейвлет-функции:

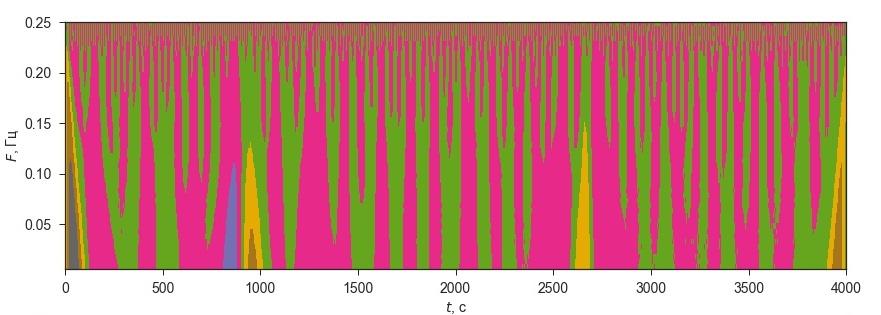
, (2)

где  - вейвлет функция, - сигнал, параметр  отвечает за расположение во времени, а  - за масштаб. Большие значения параметра  соответствуют низким частотам, меньшие - высоким. Посредством уменьшения (увеличения) параметра  вейвлет-преобразование позволяет фокусироваться на локальных (глобальных) структурах сигнала и выявлять особенности и сегментировать сигнал по интенсивности.

В результате преобразования (2), вследствие зависимости вейвлетов от двух параметров и локализации вейвлет-функции, получается 2-мерный образ анализируемого сигнала в координатах плоскости . На рис. 1. изображен пример технологического сигнала и его НВП спектрограмма. По оси абсцисс отложено время , по оси ординат частота (является обратной ), в качестве базового вейвлета применён вейвлет типа «мексиканская шляпа» [13]. Градацией цвета показаны значения вейвлет-коэффициентов . С ростом масштаба базисная вейвлет-функция увеличивается, захватывая все больший диапазон значений , коэффициенты преобразования в соответствующей точке будут зависеть от значений части составляющих . Диапазон этот тем больше, чем больше масштаб, поэтому высокочастотная информация вычисляется на основе малых отрезков значений функции , а низкочастотная информация — на основе больших. Таким образом, локальные изменения сигнала, которые могут являться аномалиями, представляют собой высокочастотную информацию и вычисляются на малых масштабах преобразования.



а)



б)

Рис.1. Пример НВП технологического сигнала: а) пример исходного технологического сигнала, б) вейвлет спектрограмма технологического сигнала.

С учётом соотношения (2) дополним определение признакового описания сигнала (1). Для сигнала  c  количеством отсчётов признаковое описание сигнала, принимает следующий вид:

, (3)

где  - коэффициенты НВП, индекс - масштаб разложения,  - номер отсчёта сигнала,  - показания сигнала при - ом отсчёте.

## Описание численного эксперимента

В работе предлагается подход обнаружения аномалий технологических сигналов посредством неконтролируемой классификации в пространстве признаков (3). Для разработки модели классификации использовался алгоритм «изолирующий лес» [11], широко применяемый в задачах обнаружения аномалий. Алгоритм основан на построении разделяющих деревьев на основе случайного признака, описывающих данные. Деревья строятся, пока каждая точка рассматриваемого пространства не окажется в листовом узле, после чего критерием нормальности служить глубина (или среднее значение глубин для леса) этого узла. Основное отличие алгоритма от других алгоритмов, использующих деревья решений, заключаются в том, что в описанном «случайном» способе построения деревьев аномалии будут попадать в листья на ранних этапах (на небольшой глубине дерева), т.е. аномалии легче изолировать. В работе используется реализация алгоритма изолирующий лес из библиотеки scikit-learn [24] для Python с параметрами по умолчанию.

Для верификации предложенного подхода, в качестве анализируемых сигналов, рассмотрены показания датчиков перепада давления  между жидкостной и газовой линии сепаратора нефти устанавливаемых в автоматизированных замерных установках. Сигналы представлены трёхдневными показаниями с 9-ти замерных установок с количеством отсчётов. Для каждого сигнала предварительно вычислялись коэффициенты НВП в диапазоне масштабов .

Дополнительно для понижения размерности вейвлет-коэффициентов с матрицы размерностью [x] до вектора столбца  и наглядной визуализации результатов применялся методов главных компонент, основанный на поиске подпространства меньшей размерности, в [ортогональной проекции](https://ru.wikipedia.org/wiki/%D0%9E%D1%80%D1%82%D0%BE%D0%B3%D0%BE%D0%BD%D0%B0%D0%BB%D1%8C%D0%BD%D0%B0%D1%8F_%D0%BF%D1%80%D0%BE%D0%B5%D0%BA%D1%86%D0%B8%D1%8F) на которые разброс данных максимален [25]. Таким образом, каждый сигнал описывался следующим множеством:

,

где  - показания датчика, - главная компонента вейвлет-коэффициентов, - номер отсчёта,  - порядковый номер датчика. Множество разделялось на обучающую (75%) и тестовую выборку (25%).

Введём метрику , позволяющей оценить отклонение между главной компонентой в заданной точке и всем множеством для дополнительной интерпретации результатов модели:

, (4)

где - среднее значение множества ,  - среднеквадратичное отклонение множества .

## Обсуждение результатов

На рис 2. представлены результаты работы модели на примере показаний датчика №3. По оси ординат отложены показания датчика, по оси абсцисс значения главной компоненты коэффициентов НВП. Градацией цвета ранжированы области с различными показателями аномальности, красными точками обозначены показания из тестовой выборки, которые модель распознала как аномалии. Анализ рисунка показывает, что большая часть регулярных показаний датчика и с обучающей и с тестовой выборки образуют плотное облако с низкими показателями аномальности. Аномальные показания оказались вытянуты вдоль оси абсцисс, заметим, что часть показаний из тестовой выборки в той же области модель распознала как регулярные. Стоит также обратить внимание, что часть показаний из обучающей выборки оказалась в области с высоким показателем аномальности.

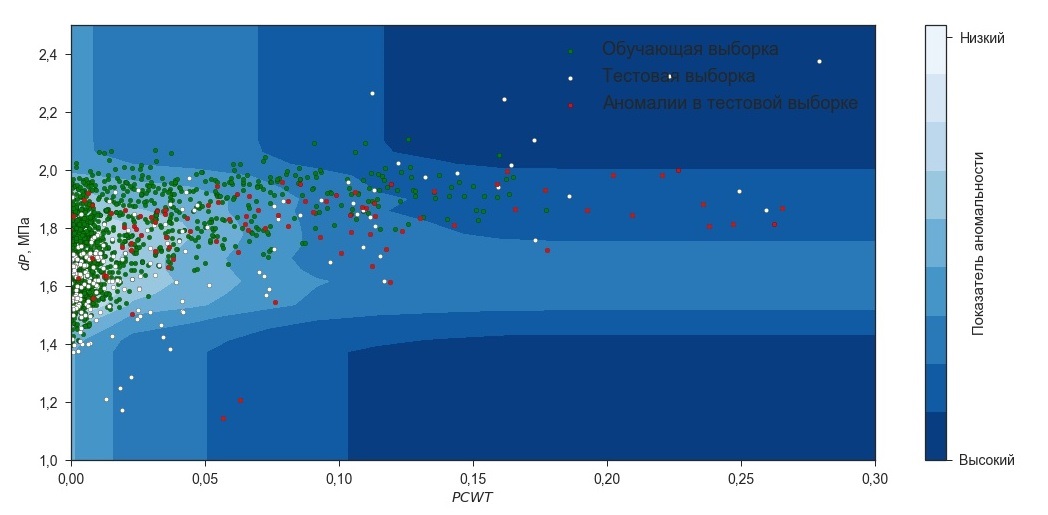


Рис. 2. Визуализация результатов работы модели.

Для интерпретации результатов работы модели рассмотрим график исходного анализируемого сигнала (рис. 3.). Выделенный красной окружностью участок был распознан моделью как аномальный, за ним следует практически мгновенный всплеск. Экспертный анализ реальной производственной ситуации показал, что данный всплеск возник вследствие недостоверности показаний датчика. При сопоставлении результатов работы модели, с журналом регистрации работ эксплуатирующей организации, выявлено, что причиной недостоверных показаний датчика послужили пропуски жидкостного клапана и попадание жидкости в газовую линию.

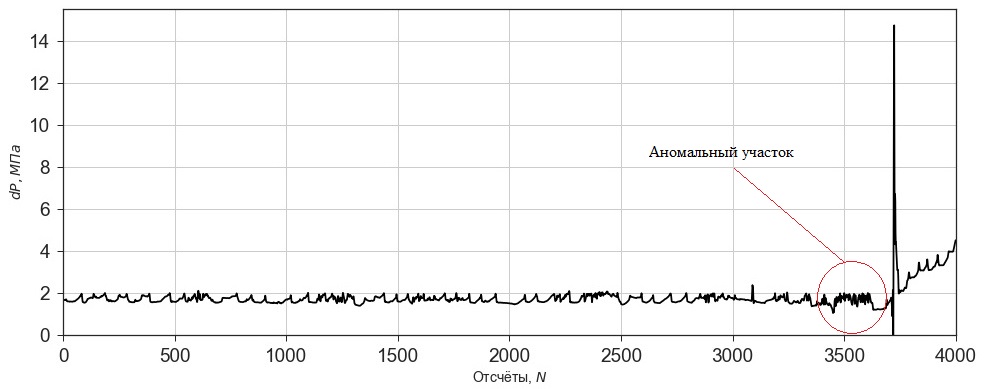


Рис. 3. Исходный анализируемый сигнал

Данное нарушение технологического процесса отразилось в структуре сигнала в виде локальных изменений частоты сигнала, однако система автоматики не обнаружила этого, так как амплитуда сигнала оставалась в пределах норм технологического регламента. На рис. 4. приведена оценка отклонения  по формуле (4) для главной компоненты коэффициентов НВП. Данный график характеризует отклонение частоты сигнала на текущем отсчёте от среднего.

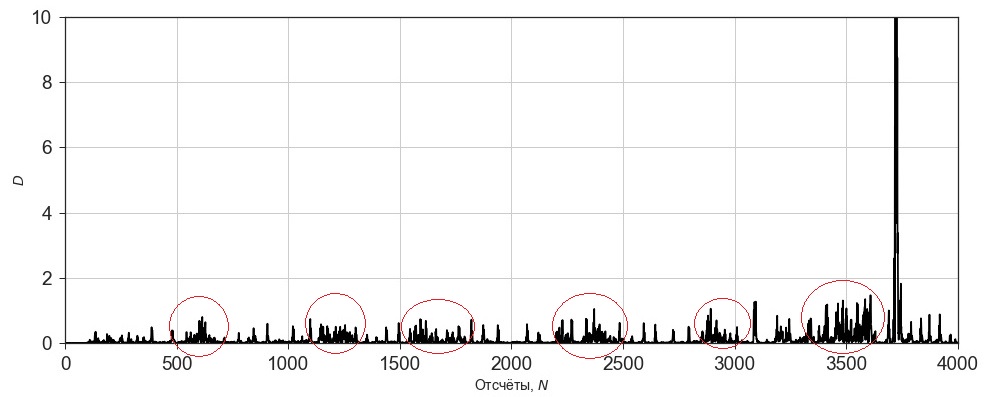


Рис. 4. Зависимость отклонений коэффициентов вейвлет-преобразования от среднего.

При анализе графика можно выделить локальные всплески, которые достигают максимума амплитуды в момент перед отказом датчика. Физически природа данного явления интерпретируется различным взаимодействием жидкости и газа с чувствительной мембраной датчика, следовательно, пропуски жидкости в газовую линию происходили систематически. Это предположение подтверждается тем, что часть обучающей выборки оказалась в области с высоким показателем аномальности (рис. 2.).

Для численной оценки результатов алгоритма экспертным путем произведена маркировка обучающих данных для датчиков №3,4,7,8 с аналогичными отказами. В таблице №1 приведена оценка работы алгоритма по метрикам точность (precision) и полнота (recall) [26].

Таблица №1. Оценка качества работы модели

|  |  |  |  |
| --- | --- | --- | --- |
| Номер датчика, | Точность | Полнота | Длина аномального участка (в отсчётах) |
| 3 | 0,77 | 0,63 | 2200 |
| 4 | 0,75 | 0,61 | 1800 |
| 7 | 0,78 | 0,65 | 2500 |
| 8 | 0,64 | 0,55 | 1200 |

Среднее значение точности для всех датчиков составило 0,73, полноты – 0,61. Наиболее высокая точность алгоритма получена для датчика  с наибольшей длинной аномального участка. Полученный результат может быть улучшен оптимизацией параметров модели и улучшением качества разметки данных. В целом, результаты эксперимента подтверждают работоспособность и применимость на практике предложенного подхода.

## Заключение

Предложен подход обнаружения аномалий в технологических сигналах путём неконтролируемой классификации. В основе подхода лежит применение коэффициентов непрерывного вейвлет преобразования, для формирования информативного признакового пространства. Рассмотрен математический аппарат и особенности вейвлет преобразования, позволяющие локализовать аномалии в частотно-временной области. В соответствие с предложенным подходом разработана модель обнаружения аномалий с применением алгоритма изолирующий лес. Проведен вычислительный эксперимент с реальными технологическими сигналами и произведена оценка результатов. Анализ полученных результатов подтвердил теоретические выводы о вейвлет-преобразовании и работоспособность предложенного подхода. В анализируемых сигналах обнаружены аномалии отражающие нарушения технологичного процесса и неисправность оборудования. Полученный результат имеет прикладное значение в целях повышения безопасности и техноэкономических показателей промышленных объектов. На основе предложенного подхода возможна разработка системы контроля и предиктивной диагностики технологического оборудования, идентификации стадий технологических процессов и производств.

## Список использованной литературы

1. **Замятин А.В.** Введение в интеллектуальный анализ данных: учеб. пособие. Томск: Изд. Дом гос. Университета, 2016. 26 с.
2. [**Gian A Susto**](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Gian%20Antonio%20Susto.QT.&newsearch=true), [Simone P](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Simone%20Pampuri.QT.&newsearch=true) and other. Machine Learning for Predictive Maintenance: A Multiple Classifier Approach// [IEEE Transactions on Industrial Informatics](https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=9424).  V.11, [Issue: 3](https://ieeexplore.ieee.org/xpl/tocresult.jsp?isnumber=7116638), 2015.
3. **Liu Y.,** Ding W. A KNNS based anomaly detection method applied for UAV flight data stream // Prognostics and System Health Management Conference, Beijing, 2015. doi: 10.1109/PHM.2015.7380051.
4. **Brad C.,** [Radu S.N](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Radu%20Stefan%20Niculescu.QT.&newsearch=true)**.** Predictive maintenance applications for machine learning// Reliability and Maintainability Symposium (RAMS), Saint Petersburg,2017, doi: 10.1109/RAM.2017.7889679
5. **Tibco** Statistica Inc [Электронный ресурс]. Режим доступа : http://statistica.io/wordpress/wp-content/uploads/Striim-Partner-Solution-Brief.pdf. (дата обращения 24.03.2018 г.)
6. **Мурзагулов Д.А**., Замятин А.В. Адаптивные алгоритмы машинного обучения в упарвдении технологическими процессами //Автоматизация и современные технологии. №7. 2018.
7. **Хуссейн X.Ш.,** А. Г. Якунин. Методы выявления аномалий при контроле динамических процессов природных и техногенных объектов// Вестник ИжГТУ имени М. Т. Калашникова. № 1(65).2015.
8. **Селевцов Л.И.** Автоматизация технологических процессов. / М.: Академия, 2014. 345 c.
9. **Заварзин Д.В.** К вопросу поиска аномалий во временных рядах//Инновации в науке. № 29. 2014. С. 59-64.
10. **Шкодырев В.П.,** Ягафаров К.И. Обзор методов обнаружения аномалий в потоках данных// Proceedings of the Second Conference on Software Engineering and Information Management, V.1864, 2017.
11. [**Kanungo**](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.T.%20Kanungo.QT.&newsearch=true) **T.,** [Mount](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.D.M.%20Mount.QT.&newsearch=true) D.M. An efficient k-means clustering algorithm: analysis and implementation//[IEEE Transactions on Pattern Analysis and Machine Intelligence](https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=34). V.: 24, [Issue: 7](https://ieeexplore.ieee.org/xpl/tocresult.jsp?isnumber=21893), 2002.
12. **Fei T. L.,** Kai Ming Ting. Isolation Forest // [Data Mining, Eighth IEEE International Conference on](https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=4781077), 2008.
13. **Shengfeng T.,** Jian Y., Chuanhuan Y. Anomaly Detection Using Support Vector Machines// Advances in Neural Networks. Volume 3173, pp 592-597, 2004.
14. **Мисриханов А.М.** Применение методов вейвлет-преобразования в электроэнергетике, Автоматика и телемеханика, 2006, выпуск 5, C. 5–23
15. **Балакирев Н.Е.**, Гуснин С.Ю., Малков М.А., Червяков Л.М. [Фильтрация речевого сигнала с помощью вейвлет-преобразования при решении задач распознавания](https://elibrary.ru/item.asp?id=18963117) речи// [Известия Юго-Западного государственного университета](https://elibrary.ru/contents.asp?issueid=1116534). 2012. [№ 5-2 (44)](https://elibrary.ru/contents.asp?issueid=1116534&selid=18963117). С. 044-050.
16. **Ручай А.Н.** [Модифицированный метод сегментации речевого сигнала на основе непрерывного вейвлет-преобразования](https://elibrary.ru/item.asp?id=18814310)// [Доклады Томского государственного университета систем управления и радиоэлектроники](https://elibrary.ru/contents.asp?issueid=1108353). 2012. Т. 2. [№ 1](https://elibrary.ru/contents.asp?issueid=1108353&selid=18814310). С. 189-192.
17. **Туровский Я.А**., Кургалин С.Д., Семёнов А.Г., Максимов А.В. [Классификация динамики локальных максимумов в вейвлет-спектрах электроэнцефалограмм](https://elibrary.ru/item.asp?id=18937103) // [Вестник Тамбовского государственного технического университета](https://elibrary.ru/contents.asp?issueid=1115048). 2013. Т. 19. [№ 1](https://elibrary.ru/contents.asp?issueid=1115048&selid=18937103). С. 90-96.
18. **Григорьев Д.С.**, Спицын В.Г. [Применение нейронной сети и дискретного вейвлет-преобразования для анализа и классификации электрокардиограмм](https://elibrary.ru/item.asp?id=18783661) // [Известия Томского политехнического университета](https://elibrary.ru/contents.asp?issueid=1106789). 2012. Т. 321. [№ 5](https://elibrary.ru/contents.asp?issueid=1106789&selid=18783661). С. 57-61.
19. **Александрин В.В.,** Лузянин Б.П., Иванов А.В., Кубатиев А.А. [Влияние гипергомоцистеинемии на мозговой кровоток, по данным вейвлет-анализа](https://elibrary.ru/item.asp?id=18762449)// [Патологическая физиология и экспериментальная терапия](https://elibrary.ru/contents.asp?issueid=1105557). 2011. [№ 2](https://elibrary.ru/contents.asp?issueid=1105557&selid=18762449). С. 13-18.
20. **Хаща О.Ю.** [Обзор методов сжатия при помощи вейвлет-преобразований](https://elibrary.ru/item.asp?id=23946433) // [68 -я научно-техническая конференция профессорско-преподавательского состава, учёных, аспирантов и студентов](https://elibrary.ru/item.asp?id=23834330) Одесская национальная академия связи им. А.С. Попова. 2013. С. 29-31.
21. **Жизняков А.Л.**, Фомин А.А. [Устройство фильтрации особенностей изображения на основе непрерывного вейвлет-преобразования](https://elibrary.ru/item.asp?id=18681180). Патент на полезную модель RUS 61441 24.10.2006.
22. **Айфичер Э.,** Джервис Б. Цифровая обработка сигналов. Практический подход. 2-е изд. М.: Вильямс, 2004. 992 с.
23. **Малла С.** Вейвлеты в обработки сигналов. М.:Мир, 2005.
24. **Documentation of scikit-learn.** [Электронный ресурс]. Режим доступа : <http://scikit-learn.org/stable/documentation.html>. (дата обращения 24.05.2018 г.)
25. **Maćkiewicz1 A.,** Ratajczak W. Principal components analysis (PCA)// Computers & Geosciences. Volume 19, Issue 3, 1993, pp 303-342.
26. **Михайлов Г.** Метрики качества [Электронный ресурс]. – Режим доступа: http://glebmikhaylov.com. (дата обращения 20.05.2018 г.)