

Санкт-Петербургский государственный университет

ЗУБАКОВА Дарья Викторовна

Выпускная квалификационная работа

Анализ данных от систем пространственно распределённых датчиков

Уровень образования: магистратура

Направление: 01.04.02 «Прикладная математика и информатика»

Основная образовательная программа: ВМ.5504 «Исследование операций и системный анализ»

Научный руководитель:
доктор физ.-мат.наук,
профессор кафедры МТИСР,
Громова Е. В.

Рецензент: канд. физ.-мат.наук,
руководитель направления
исследований ООО Сименс,
Вениаминов Н. А.

Санкт-Петербург

2020 год

Содержание

Введение.....	3
Постановка задачи.....	5
Обзор литературы.....	6
Глава 1. Детектирование событий.....	8
1.1. Алгоритм кумулятивных сумм.....	8
1.2. SVD-разложение.....	13
Глава 2. Альтернативные методы детектирования событий и локализация событий.....	16
2.1. Спектральный анализ.....	16
2.1. Кепстральный анализ.....	17
2.2. Корреляционный метод локализации событий.....	21
Глава 3. Валидация данных.....	29
3.1. Виды валидаторов и их задачи.....	29
3.2. Многомерный валидатор.....	31
Выводы.....	36
Заключение.....	38
Список литературы.....	39
Приложение.....	43

Введение

Анализ данных — широкая и непрерывно развивающаяся область математики и информатики. Вычислительные алгоритмы и математические методы анализа данных позволяют извлекать полезные очевидные и неочевидные закономерности из большого объема сырых данных. Это помогает в решении классов задач предаварийного контроля, поддержки принятия решений и различных оптимизационных задач.

Системы пространственно распределённых датчиков необходимы для мониторинга состояния различных крупных объектов (например, систем водоснабжения, теплоснабжения, продуктопроводов, ГЭС, ГАЭС). В системах датчиков используются сенсорные элементы разного типа: акустические, электрические, пьезоэлектрические, тепловые, электромагнитные [10][18].

Например, для организации системы предаварийного контроля на основе систем пожарной автоматики следует устанавливать тепловизионные камеры, датчики температуры, давления, виброакустические датчики, интегральные датчики контроля высокодисперсных аэрозолей и некоторые другие [11].

Для любого из таких объектов важно предотвратить или, в крайнем случае, моментально обнаружить возникновение неполадок в работе оборудования.

Актуальность исследования данной темы обусловлена высокой значимостью поддержания работоспособности систем пространственно распределённых датчиков.

Новизна работы состоит в предложенных комплексных подходах к обнаружению неполадок работы систем.

Целью работы является реализация методов детектирования и локализации неполадок на системах пространственно распределённых датчиков, а также валидации данных.

Для достижения данной цели были поставлены следующие **задачи**:

- 1) Анализ методов детектирования и локализации событий на системах пространственно распределённых датчиков.
- 2) Анализ методов валидации данных от систем сенсоров.
- 3) Выбор конкретных методов детектирования и локализации событий, валидации данных.
- 4) Разработка программной реализации выбранных методов.
- 5) Оценка результатов применения методов на тестовых и реальных данных.

В рамках сотрудничества с компанией Сименс было рассмотрено применение исследуемых методов для решения реальных задач.

Работа состоит из введения, постановки задачи, обзора литературы, трёх глав, выводов и заключения.

Глава 1 данной работы посвящена детектированию неполадок в работе систем датчиков. В главе рассматривается применение модифицированного алгоритма кумулятивных сумм и SVD-разложения. Демонстрируются результаты применения методов детектирования на реальных данных.

В главе 2 рассматриваются альтернативные методы детектирования (спектральный и кепстральный анализ) и метод локализации неполадок (корреляционный метод). Приведены результаты применения методов на тестовых и реальных данных.

Глава 3 посвящена валидации данных. В главе приводятся краткий обзор задач валидаторов и подробное описание реализации многомерного валидатора. Демонстрируются результаты применения многомерного валидатора на реальных данных.

Выражаю благодарность компании Сименс за сотрудничество, команде аналитиков за консультации по исследуемым темам. Особо благодарю инженеров исследования и разработки Сазыкина Дмитрия и Бакутеева Антона, руководителя направления исследований Вениаминова Николая за консультации и общее руководство практической частью работы.

Постановка задачи

В рамках сотрудничества с компанией Сименс была предложена задача программной реализации методов детектирования и локализации событий на продуктопроводах.

Под событиями на трубопроводе будем понимать возникновение прорывов, врезок, работу запорной арматуры и переключение режимов работы насосов.

Наличие прорывов и врезок в продуктопроводе приводит к потерям продукта, что невыгодно в финансовом плане и экологическом, и к временным затратам на обнаружение мест прорывов и их устранение. Появляется необходимость применения методов анализа данных для обнаружения таких критических событий.

Реализация методов детектирования и локализации событий осуществлена на языке MATLAB.

Работоспособность и адекватность методов проверена на данных датчиков при эмуляции возникновения разного рода событий на реальном продуктопроводе. В главах 1-2 приведены результаты работы методов.

Также в рамках сотрудничества была предложена задача валидации данных для систем сенсоров.

Валидация данных является важным этапом технологического процесса, позволяющим принимать решение о достоверности и надёжности данных. Обработка сигналов, поступающих с сенсоров, позволяет выявить неустойчивость системы, аварийные ситуации, сбой в работе оборудования или его непригодность.

Реализация валидации данных осуществлена на языке Python.

Работоспособность метода валидации данных проверена на реальных данных системы сенсоров. В главе 3 также приведены результаты работы метода.

Обзор литературы

Классические методы детектирования событий на продуктопроводах представлены в обзорных статьях [12] и [13]. В [13] одним из простых методов, не требующих наличия дополнительного оборудования, кроме датчиков давления, является метод обработки кривой падения давления. Изменение уровня давления с течением времени будет свидетельствовать о наличии события на продуктопроводе. Тем же автором в [12] приведено описание корреляционного метода локализации событий. Для определения местоположения события необходимо найти задержку по времени прихода события на датчики. Задержка определяется по максимуму кросс-корреляционной функции сигналов.

В [1] описано исследование задачи обнаружения утечек и несанкционированных воздействий на частном примере продуктопроводов – магистральных нефтепроводах. Предлагаемый метод идентификации событий заключается в применении преобразования Фурье (FT) наряду с вейвлет-преобразованием (WT). Метод FT хорошо выделяет периодические компоненты сигнала, в то время как WT позволяет находить низкочастотные события (преобладающие гармоники) в частотной области, а высокочастотные (например, скачки амплитуды сигнала, пики) – во временной.

Сравнение алгоритмов детектирования утечек в системах водоснабжения, имеющих простую топологию, с помощью быстрого преобразования Фурье (FFT) и оконного преобразования Фурье (STFT) приведено в [19]. Исследование показывает, что в целом алгоритм FFT имеет преимущество перед STFT.

В [24] рассматривается применение кепстрального анализа к задаче обнаружения утечек на трубопроводах. Методика основана на анализе периодических компонент волн давления. При наличии подходящей

фильтрации шумов, кепстральный анализ помогает идентифицировать события на простой Т-образной сети трубопровода. Для подтверждения работоспособности алгоритма топология сети менялась. Было выявлено, что при увеличении размера утечки отношение амплитуд пиков кепстра также увеличивалось.

Применение кепстров к задаче детектирования утечек также хорошо описано в статье [25]. Предлагается комплексный метод обнаружения утечек, основанный на кепстральном анализе кросс-корреляционной функции сигналов. Результаты такого подхода будут менее точными при сложной топологии сети продуктопроводов (количество пиков кепстров возрастает, так что сложнее определить их особенности).

Классические методы валидации данных описаны в обзорных статьях [22] и [28]. Рассмотрены разные способы реализации валидаторов для простых задач (обнаружение выбросов, пропусков данных, константных значений), так и для более сложных (обнаружение отличающегося сигнала в группе схожих сигналов).

Глава 1. Детектирование событий

Была рассмотрена простая (без ответвлений и дополнительных соединений) топология продуктопровода, состоящая из продуктопровода, трёх датчиков давления, расположенных на разных участках продуктопровода, и насосной станции, состоящей из двух насосов. Насосы могут быть включены одновременно или же независимо друг от друга. На некотором промежутке времени было проведено 19 эмуляций прорывов/врезок разных диаметров. Вместе с включением/выключением насосов – 21 событие.

При высокой скорости перекачки продукта необходимо максимально точно определить время возникновения события, иначе последующая его локализация будет иметь большие погрешности, что приведет и к большей потере продукта, и к большим временным затратам на устранение утечки.

В связи с этим, имеет смысл рассматривать последовательные алгоритмы обнаружения событий, которые на каждом шаге (отсчёте времени) принимают решение о наличии события, используя данные, полученные на предыдущих шагах [4]. Оптимальной группой алгоритмов обнаружения событий являются алгоритмы, основанные на вычислении кумулятивных сумм [2][3][4][5].

1.1. Алгоритм кумулятивных сумм

Существует несколько модификаций алгоритма кумулятивных сумм (например, алгоритм Сегена-Сандерсона, алгоритм Надлера-Роббинса, алгоритм Гиршика-Рубина-Ширяева, алгоритм Керра, алгоритм Шьюхарта).

Для решения задачи детектирования событий рассмотрена реализация алгоритма Сегена-Сандерсона. Такой выбор обоснован более низкими

значениями вычислительной сложности алгоритма и занимаемой памяти при сравнении модификаций алгоритмов [9][14][15].

Модификация Сегена-Сандерсона алгоритма кумулятивных сумм:

$$G(n) = \max_{i \leq n} \left\{ S_i - \min_{k \leq i} S_k \right\},$$

где $S_i = \sum_{j=1}^i z(j)$, G – решающая функция, z – значения сигнала.

Измерения, получаемые с датчиков давления, имеют неотрицательные значения. Для такого частного случая решающую функцию можно переписать в следующем виде:

$$G(n) = S_n - S_1, S_n = \sum_{i=1}^n (z(i)).$$

Пусть $h > 0$ – порог допустимого изменения решающей функции. Тогда для наступления какого-либо события на продуктопроводе справедливо следующее:

- $G(n) - G(n-1) \geq h$, – произошло событие,
- $G(n) - G(n-1) < h$, – событие не произошло.

Во избежание ложного обнаружения событий на участках сигналов, имеющих одиночные большие выбросы, применяется сглаживание сигналов с помощью фильтра Савицкого-Голея (полиномом второго порядка).

В данной фильтрации вместо линейного приближения в окрестности каждой точки измерения строится аппроксимирующий полином n -го порядка по методу наименьших квадратов [16]:

$$\hat{y}_j = \sum_{i=\frac{1-m}{2}}^{\frac{m-1}{2}} C_i y_{j+i}, j = 1, \dots, n,$$

$$\frac{m-1}{2} \leq j \leq n - \frac{m-1}{2},$$

где C_i – коэффициенты свёртки, m – окно сглаживания.

Перейдем к результатам работы метода детектирования. Ниже приведены графики исходных данных (рис. 1) с вертикальными линиями чёрного цвета (момент проведения эмуляции события).

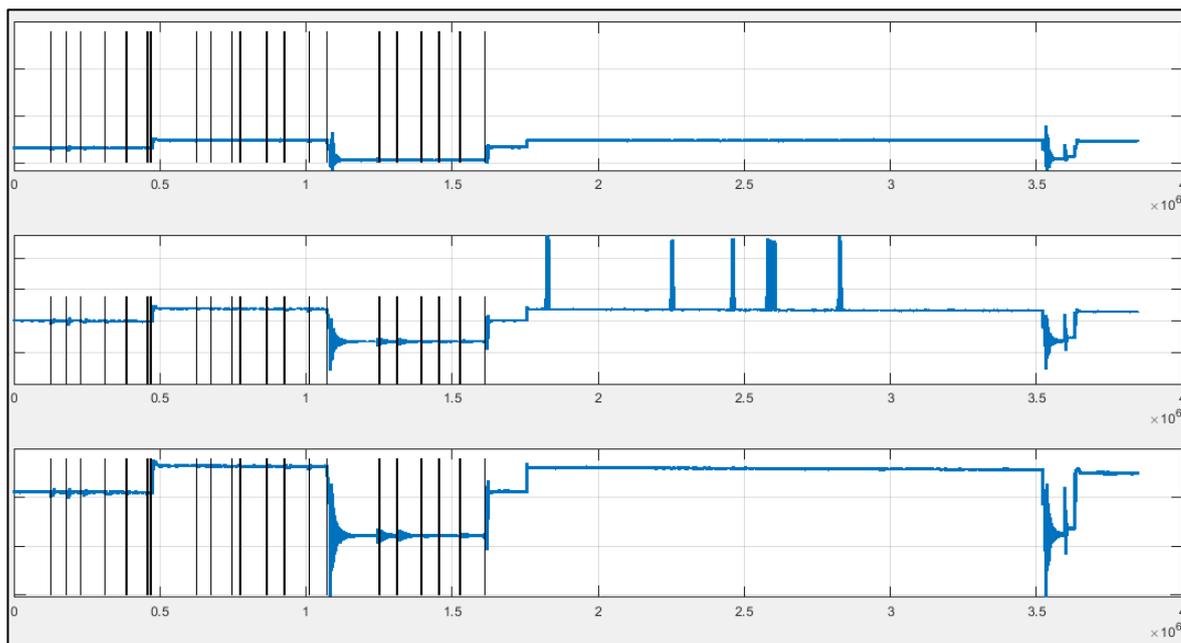


Рис. 1. Исходные данные

Приведены графики статистики кумулятивной суммы (значений решающей функции) для каждого датчика давления (рис. 2 – 4) с вертикальными линиями чёрного цвета (момент проведения эмуляции события) и закрашенными прямоугольниками синего цвета (обнаружение алгоритмом события). На рис.2 также размечены участки включения и выключения насосов. Такая разметка справедлива и для датчиков №2, 3 (рис.3-4).

Чувствительность метода зависит от выставления порога допустимого изменения решающей функции. Таким образом, метод позволяет детектировать и небольшие события (рис. 5). Отметим, что значительное снижение порога допустимого изменения решающей функции приведёт к

ложным срабатываниям алгоритма, поэтому целесообразно выбирать пороговое значение большее, чем уровень колебания давления при нормальном функционировании продуктопровода. Для определения нормального уровня колебания давления потребуется провести эксперимент с включением и выключением насосов на заполненном продуктом заведомо неповреждённом продуктопроводе.

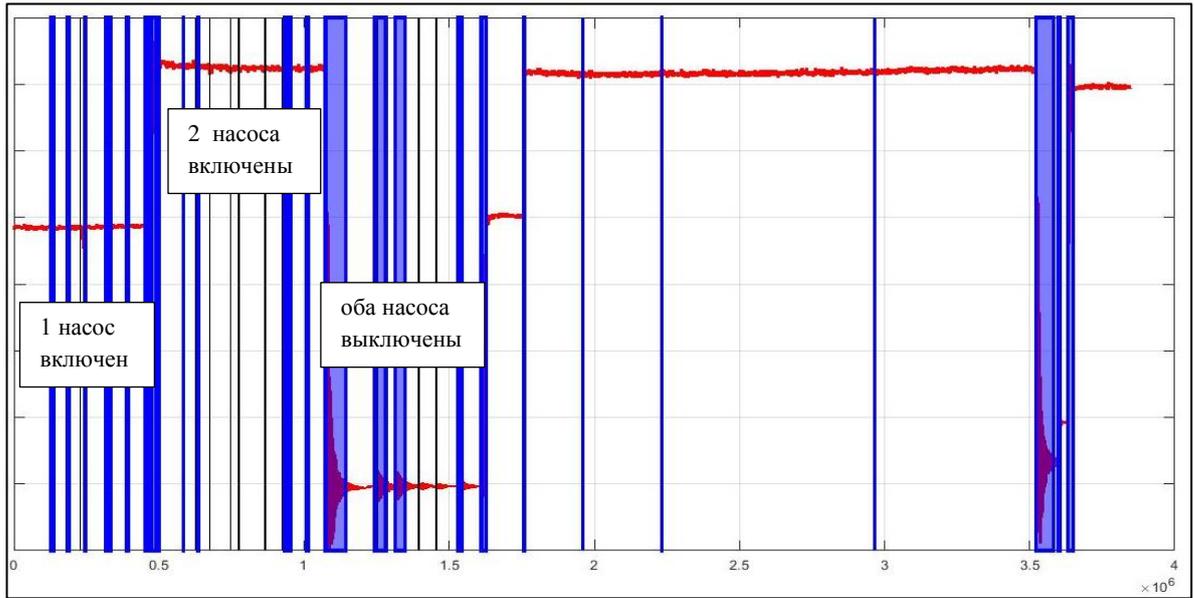


Рис. 2. Детектирование событий на датчике №1

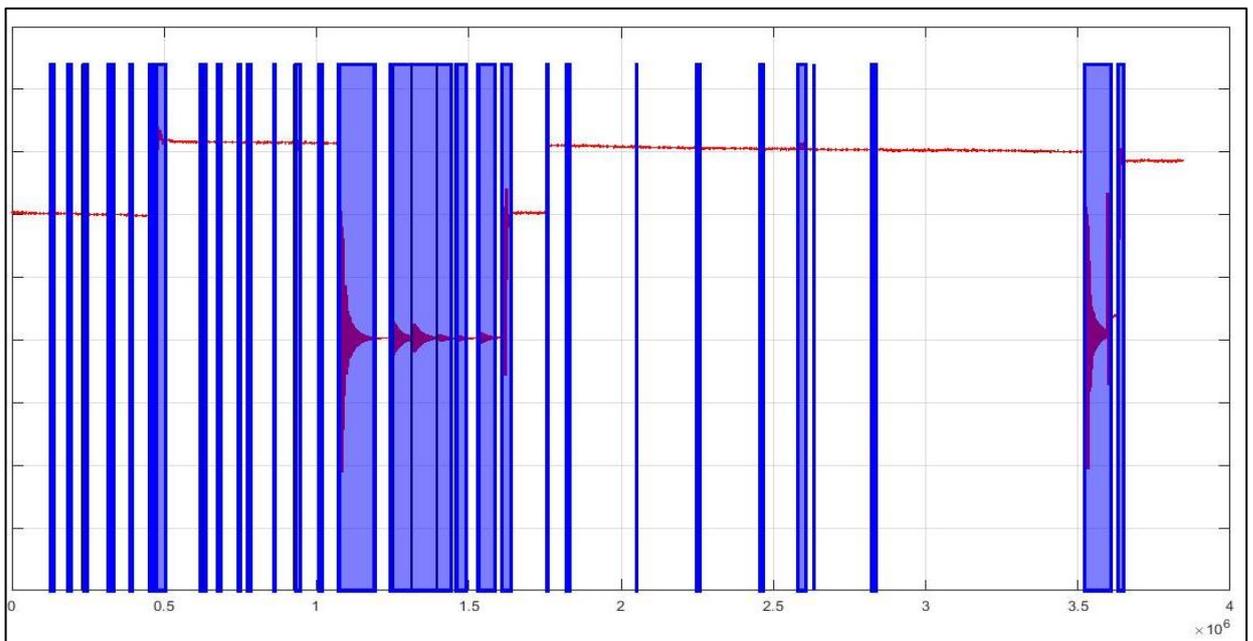


Рис. 3. Детектирование событий на датчике №2

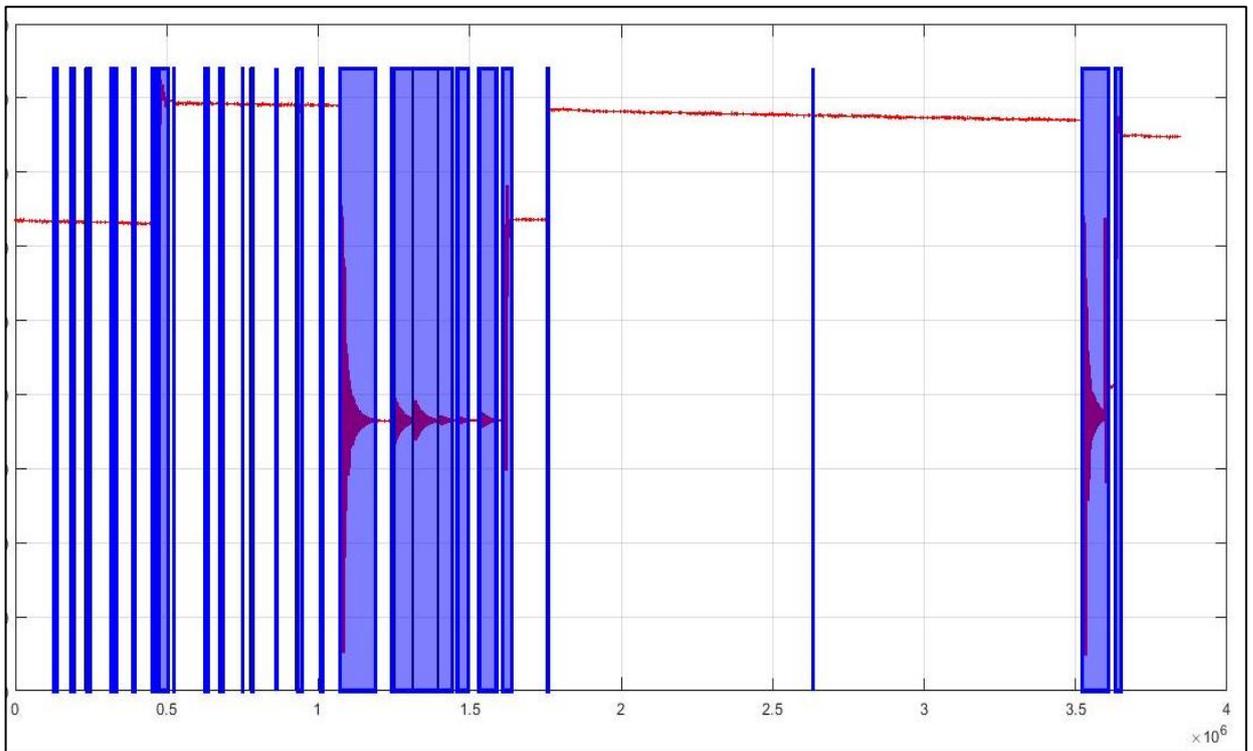


Рис. 4. Детектирование событий на датчике №3

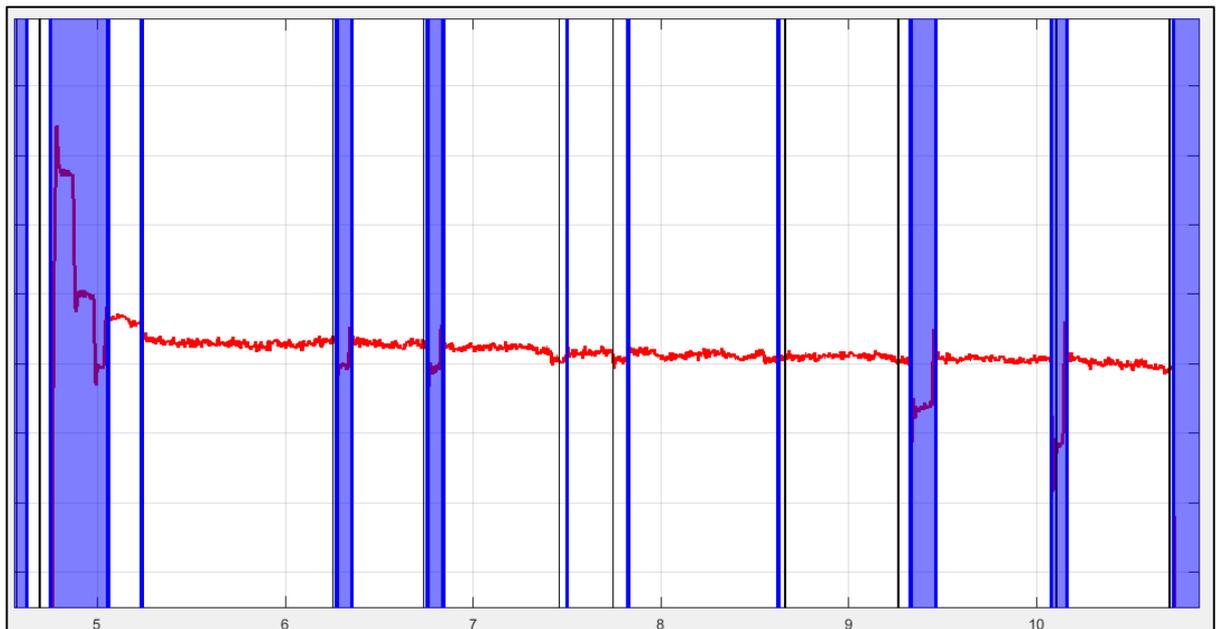


Рис. 5. Детектирование мелких событий

На датчиках №2 и №3 события детектировались лучше, чем на датчике №1 (обнаруживались все события при одном и при двух включенных насосах). Нужно учитывать, что точность детектирования также зависит от дальности размещения датчика от точки эмульсии.

1.2. SVD-разложение

Предположим, у нас имеется группа датчиков, регистрирующих одно и то же событие. По причине сильного зашумления сигнала или дальности установки датчика возможна ситуация, когда часть события будет «видна» на сигнале от одного датчика, другая часть – на сигнале второго, или же событие будет только на одном датчике.

Для выявления такого события представим сигналы от нескольких датчиков как один сигнал с помощью метода понижения размерности (SVD-разложения).

Перейдём к формальному описанию. Пусть имеется исходная матрица $X_{\langle n \times r \rangle}$, $n > r$. Рассмотрим экстремальную задачу: найти матрицу \tilde{X} той же размерности из условия:

$$\sum_{i,j} (x_{ij} - \tilde{x}_{ij})^2 \rightarrow \min \quad (1)$$

при ограничении $\text{rank}(\tilde{X}) < \min(n, r)$.

Тогда под SVD-разложением будем понимать следующее:

$$X = L S R^T, \quad (2)$$

где S – диагональная матрица с неотрицательными элементами; элементы матрицы S являются сингулярными числами исходной матрицы X ; столбцы матрицы $L_{n \times n}$ являются левыми сингулярными векторами исходной матрицы X (собственные векторы матрицы $X^T X$); столбцы матрицы $R_{m \times m}$ являются правыми сингулярными векторами исходной матрицы X (собственные векторы матрицы $X^T X$).

Разложение (2) можно переписать в виде суммы:

$$X = \sum_{i=1}^r s_i L_i R_i^T = s_1 L_1 R_1^T + \dots + s_r L_r R_r^T. \quad (3)$$

Согласно теореме Эккардта-Янга [21], решением экстремальной задачи (1) является сумма первых k слагаемых в (3):

$$\tilde{X} \cong \sum_{i=1}^k s_i L_i R_i^T = s_1 L_1 R_1^T + \dots + s_k L_k R_k^T.$$

Перейдем к результатам работы метода SVD-разложения. На рис.6-7 представлены графики сигналов сенсоров №1, 2, 3 и результирующего SVD-сигнала в случаях, когда событие можно обнаружить только на одном и на двух датчиках. Чёрной вертикальной линией отмечено время начала события.

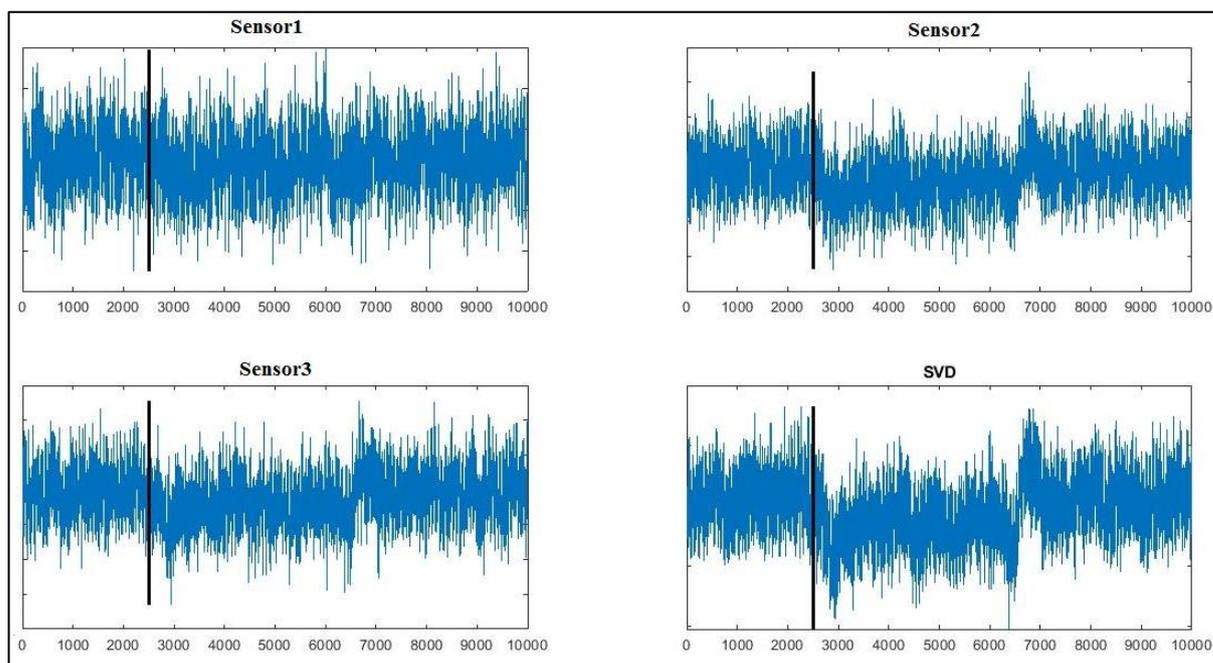


Рис. 6. Сигналы сенсоров №1, 2, 3 и результирующий сигнал SVD (случай с одним событием на двух сенсорах)

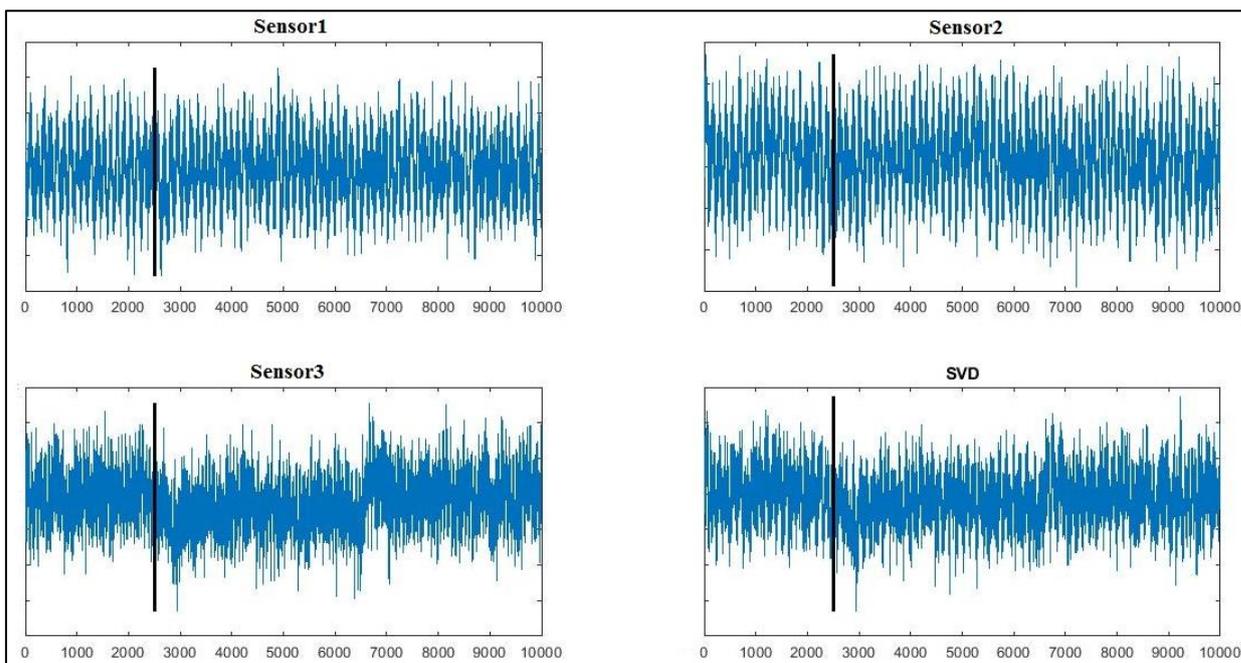


Рис. 7. Сигналы сенсоров №1, 2, 3 и результирующий сигнал SVD (случай с одним событием на одном сенсоре)

После выделения результирующего SVD-сигнала применяется алгоритм кумулятивных сумм. На рис. 8 приведен пример детектирования события для SVD-сигнала (само выделение SVD-сигнала – на рис. 6).

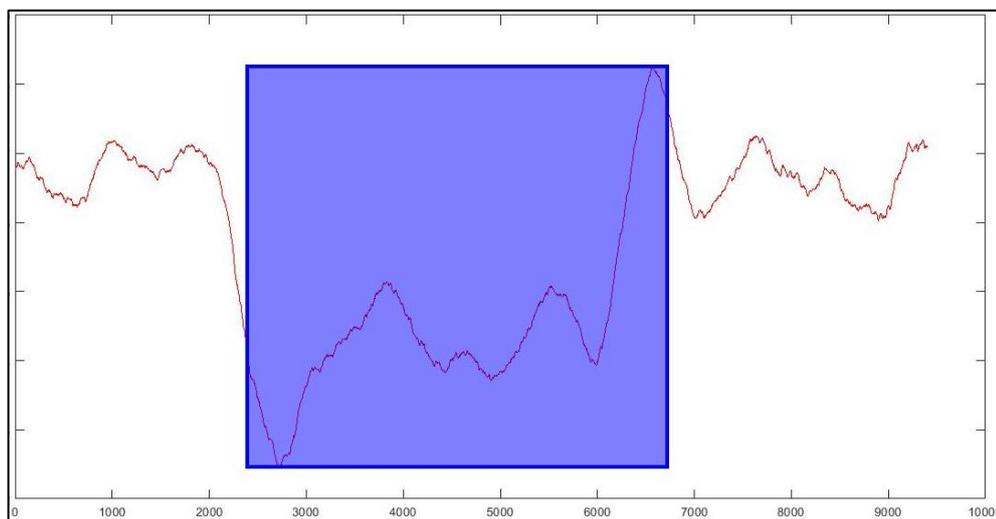


Рис. 8. Детектирование события на результирующем SVD-сигнале

Использование метода SVD-разложения полезно в случае сильного зашумления сигналов и слабой выраженности события.

Глава 2. Альтернативные методы детектирования событий и локализация событий

Появление событий на продуктопроводе вызывает образование волн переходного процесса по давлению. По мере распространения волны отражаются и затухают.

Возвращаясь к рассматриваемой системе продуктопровода, рассмотрим ситуацию, когда при выключенных насосах возникает некоторое событие. Тогда график сигнала датчика будет похож на медленно затухающую синусоиду.

Существуют методы, отражающие изменение видов колебаний – спектральный и кепстральный анализ [25][7].

2.1. Спектральный анализ

Спектр сигнала представляет собой набор гармонических колебаний, в сумме составляющих исходный сигнал [8][17]. Спектр можно записать через преобразование Фурье в следующем виде:

$$S(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} s(t) e^{-i\omega t} dt,$$

где ω – угловая частота, $s(t)$ – исходный сигнал.

На нижнем графике рисунка 9 представлен спектр исходного сигнала при выключенных насосах и некотором событии. Сам исходный сигнал изображён на верхнем графике. На рис. 10 представлен увеличенный фрагмент спектра сигнала.

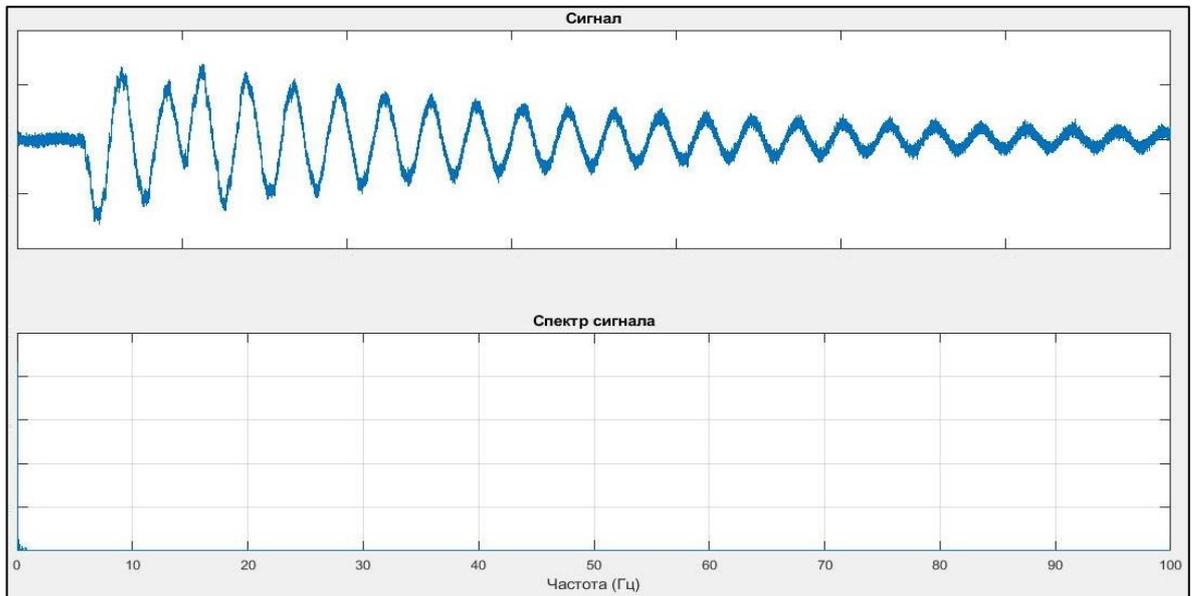


Рис. 9. Спектр сигнала

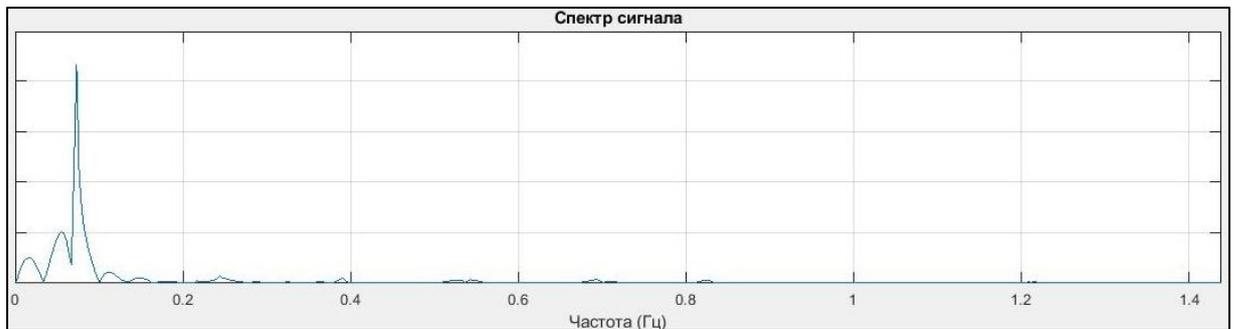


Рис. 10. Увеличенный фрагмент спектра сигнала

При возвращении во временную область из частотной области, найденное по пику спектра время не совпало со временем наступления события.

2.1. Кепстральный анализ

Понятие «кепстр» определено следующим выражением:

$$C_s(q) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \ln |S(\omega)|^2 e^{i\omega q} d\omega,$$

где $S(\omega)$ – амплитудный спектр сигнала $s(t)$, аргумент q имеет размерность времени (это особое, кепстральное время, поскольку $C_s(q)$ в любой момент q зависит от $s(t)$).

Богерт, Хили и Тьюки в статье [20] отметили, что логарифм спектра мощности колебания, содержащего отраженный сигнал, имеет аддитивную периодическую компоненту, созданную этим сигналом, и поэтому преобразование Фурье от логарифма спектра мощности имеет пик на месте, соответствующем задержке отраженного сигнала.

Существует несколько видов кепстральных коэффициентов: комплексный кепстр, энергетический кепстр, реальный кепстр, фазовый кепстр, кепстр автокорреляционной функции, мел-частотные кепстральные коэффициенты. Приведём некоторые из них.

Пусть F – функция преобразования Фурье, F^{-1} – обратное преобразование Фурье. Тогда энергетический кепстр можно выразить следующим образом:

$$C_p = |F^{-1}\{\log(|F\{f(t)\}|^2)\}|^2.$$

Комплексный кепстр можно выразить следующим образом:

$$C_c = F^{-1}\{\log(F\{f(t)\})\}.$$

Реальный кепстр:

$$C_r = F^{-1}\{\log(|F\{f(t)\}|)\}.$$

Вычисление кепстра автокорреляционной функции отличается, соответственно, взятием автокорреляционной функции перед спектральным преобразованием. Автокорреляционную функцию запишем следующим образом:

$$\rho_{XX}(t_1, t_2) = \frac{E[(X_{t_1} - \mu_{t_1})(X_{t_2} - \mu_{t_2})]}{\sigma_{t_1} \sigma_{t_2}},$$

где μ – среднее, σ – среднеквадратическое отклонение, E – математическое ожидание.

После вычисления кепстральных коэффициентов и нахождения максимальных пиков кепстра последующая фильтрация сигнала заключалась в удалении аддитивной периодической компоненты.

Были рассмотрены вариации кепстрального анализа с вычислением кепстра автокорреляционной функции, энергетического кепстра и реального кепстра. Пик на месте, соответствующем задержке отраженного сигнала, чаще находился с помощью вычисления реального кепстра.

На рис.11 представлен результат работы кепстрального анализа с использованием реального кепстра на незашумлённом сигнале, имеющем форму затухающей синусоиды. Чёрные вертикальные линии соответствуют найденным позициям пиков кепстра, красная вертикальная линия – максимальный пик.

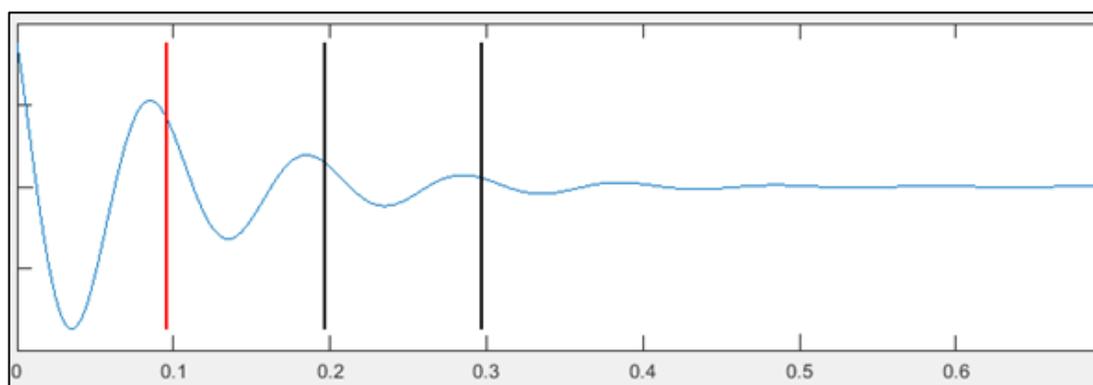


Рис. 11. Результат работы кепстрального анализа (реальный кепстр)

На верхнем графике рисунка 12 изображён исходный сигнал, а на нижнем – отфильтрованный. Фильтрация сигнала осуществлялась на отрезке сигнала, начиная с максимального пика кепстра.

На верхнем графике рисунка 13 изображён исходный сигнал, а на нижнем – отфильтрованный. Заметно, что нахождение пиков кепстра ухудшилось.

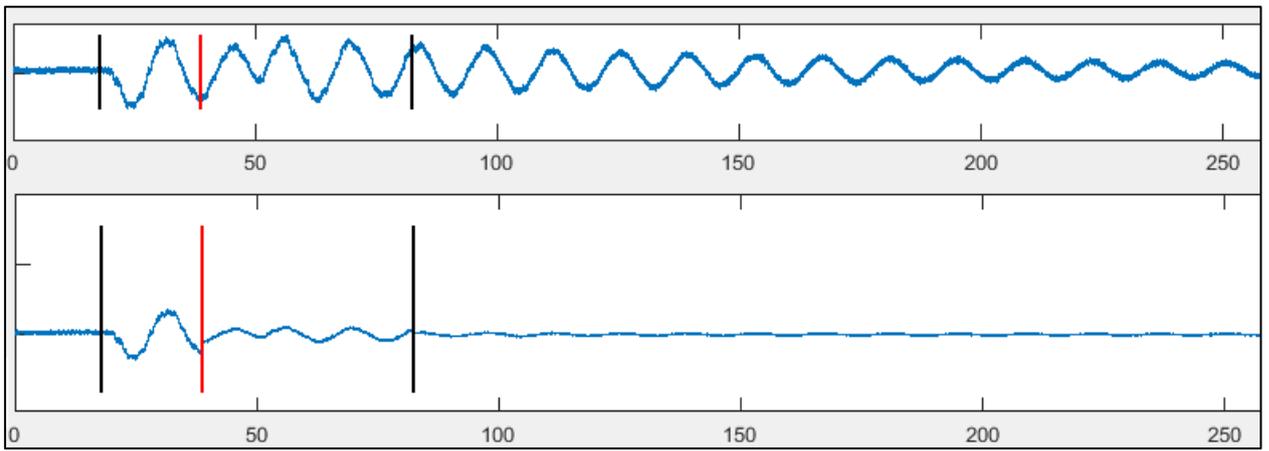


Рис. 12. Результат работы кепстрального анализа (реальный кепстр) и последующая фильтрация сигнала

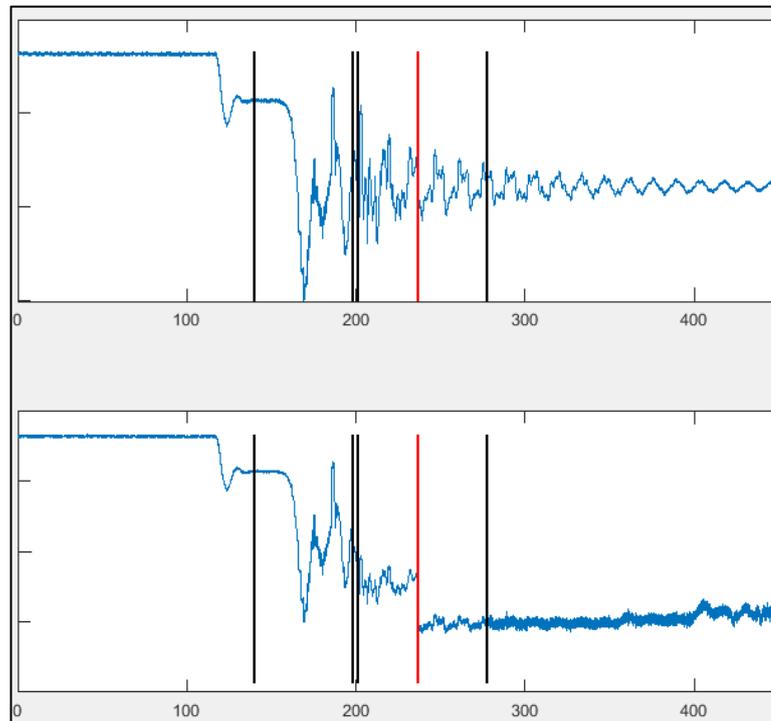


Рис. 13. Результат работы кепстрального анализа (реальный кепстр) и последующая фильтрация сигнала

На точность нахождения мест, соответствующих задержке отраженного сигнала, влияет форма сигнала, зашумленность сигнала и выставяемые границы амплитуды пиков кепстров.

2.2. Корреляционный метод локализации событий

Для локализации событий необходимо как минимум два сенсора. Зная разницу времён обнаружения событий на обоих датчиках и скорость распространения волны в продуктопроводе, можно определить место прорыва/врезки.

Пусть L – расстояние между двумя сенсорами. Найдём l – расстояние от первого сенсора до места прорыва/врезки:

$$l = \frac{a \Delta t + L}{2},$$

где a – скорость распространения волны в продуктопроводе, Δt – разница времён обнаружения событий на датчиках.

Корреляционный метод нахождения временного сдвига между сигналами заключается в поиске наибольшего коэффициента корреляции между выделенным событием [26][27][12] (например, фронтом падения давления) одного сигнала и участками второго сигнала (длина участков второго сигнала равна длине выделенного события) при движении вдоль всего второго сигнала с шагом в один отсчёт. Временной сдвиг можно записать следующим образом:

$$t_{delay} = t_1 - \underset{t \in \mathbb{R}}{\operatorname{argmax}}(\operatorname{corr}(s'_1, s_2)(t)),$$

где t_1 – время начала выделенного события, s'_1 – выделенный участок первого сигнала с событием, s_2 – второй сигнал, corr – корреляционная функция.

Было рассмотрено использование разных коэффициентов корреляции (Пирсона, Кендалла, Спирмена) в корреляционном методе.

Коэффициент корреляции Пирсона рассчитывается по следующей формуле:

$$r_{xy} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}},$$

где n – количество наблюдений.

Ранговый коэффициент корреляции Кендалла:

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)},$$

где n – количество наблюдений, n_c – число совпадений, n_d – число инверсий.

Ранговый коэффициент корреляции Спирмена:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)},$$

где n – количество наблюдений, d – разница между рангами сравниваемых переменных.

На верхнем графике рисунка 14 представлены два смоделированных сигнала с временным сдвигом равным -176 отсчетов. Фрагмент синего сигнала на нижнем графике (этот же фрагмент выделен на верхнем графике прямоугольником синего цвета) – фронт падения давления с некоторым захватом сигнала до него. Прямоугольником красного цвета на верхнем графике выделен найденный участок красного сигнала, имеющий наибольший коэффициент корреляции с фронтом падения давления синего сигнала. На рис.15 изображена динамика изменения значений коэффициентов корреляции при движении вдоль красного сигнала.

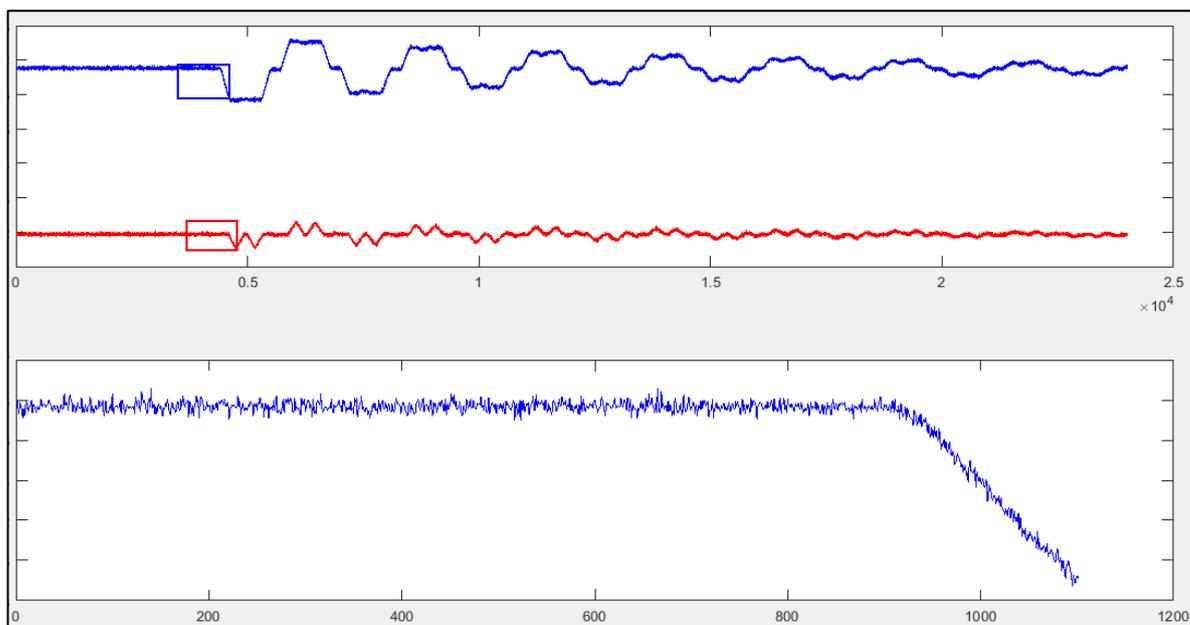


Рис. 14. Нахождение сдвига с помощью коэффициента корреляции Спирмена

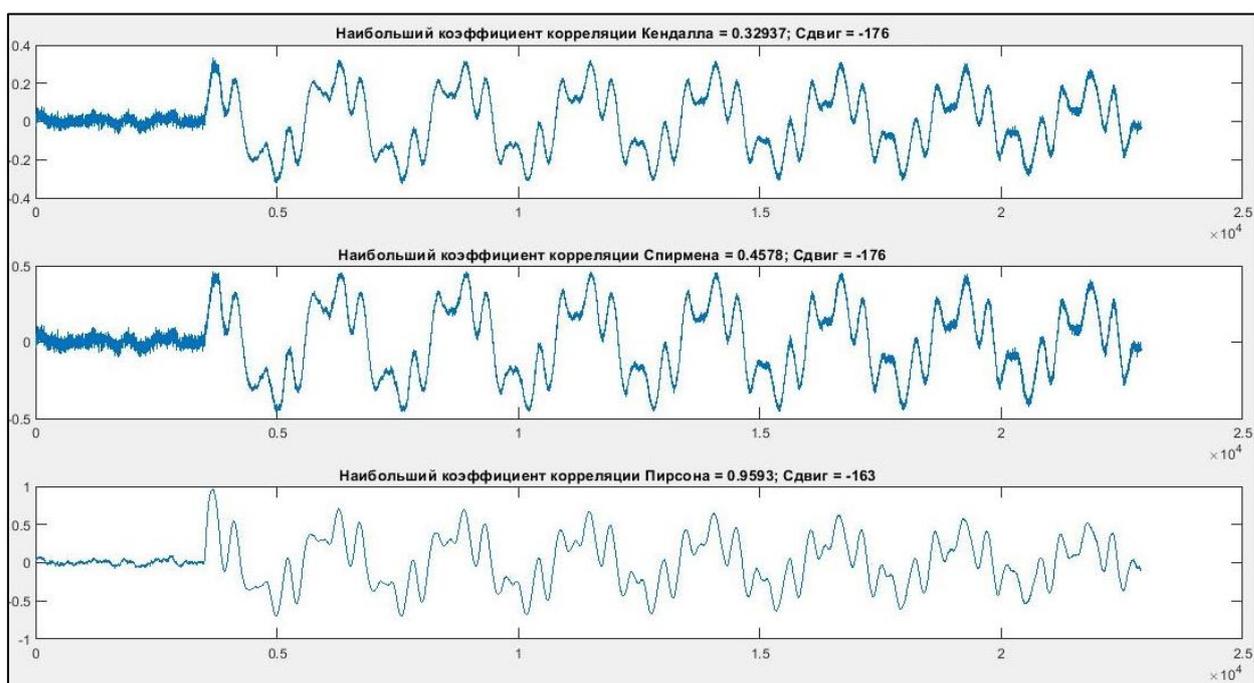


Рис. 15. Найденные сдвиги

Проверка работы метода нахождения временных сдвигов осуществлялась на тестовых данных с имитацией врезок, определенными временными сдвигами и разной ошибкой моделирования.

Проводилось тестирование разных отрезков исходных данных на нормальное распределение шума [6] на окнах шириной в 200 отсчётов и с

шагом в 1 отсчёт. Тест Лиллиефорса и тест Хи-Квадрат в подавляющем большинстве случаев дают нам принятие нулевой гипотезы о нормальном распределении на уровне значимости 0,05. Исходя из этого, для тестовых данных шум формировался также из нормального распределения.

На рис. 16-17 приведены результаты тестирования отрезков исходных данных на нормальное распределение. Каждому отсчету на приведенных графиках соответствует результат тестов на окне в 200 отсчетов (0 – нулевая гипотеза о нормальности распределения не отвергается на уровне значимости 0.05, 1 – отвергается).

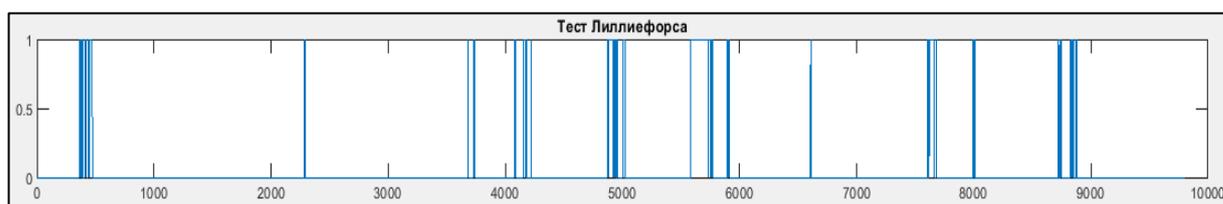


Рис. 16. Результаты тестирования шума исходных данных на нормальное распределение.

Тест Лиллиефорса

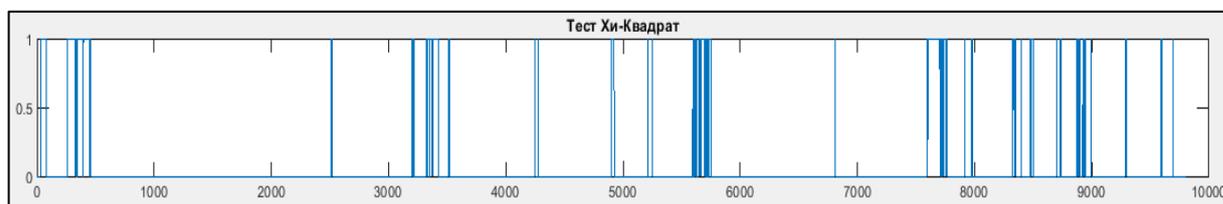


Рис. 17. Результаты тестирования шума исходных данных на нормальное распределение.

Тест Хи-Квадрат

Важным параметром для корреляционного метода является длина захватываемого сигнала до и после обнаружения события. При недостаточном или избыточном захвате сигнала, и с учётом отличия кривых сигналов, может оказаться, что выделенное событие имеет больший коэффициент корреляции с участком другого сигнала, где события нет/оно уже произошло. На рис. 18 приведён пример выделения данных с разными длинами захвата.

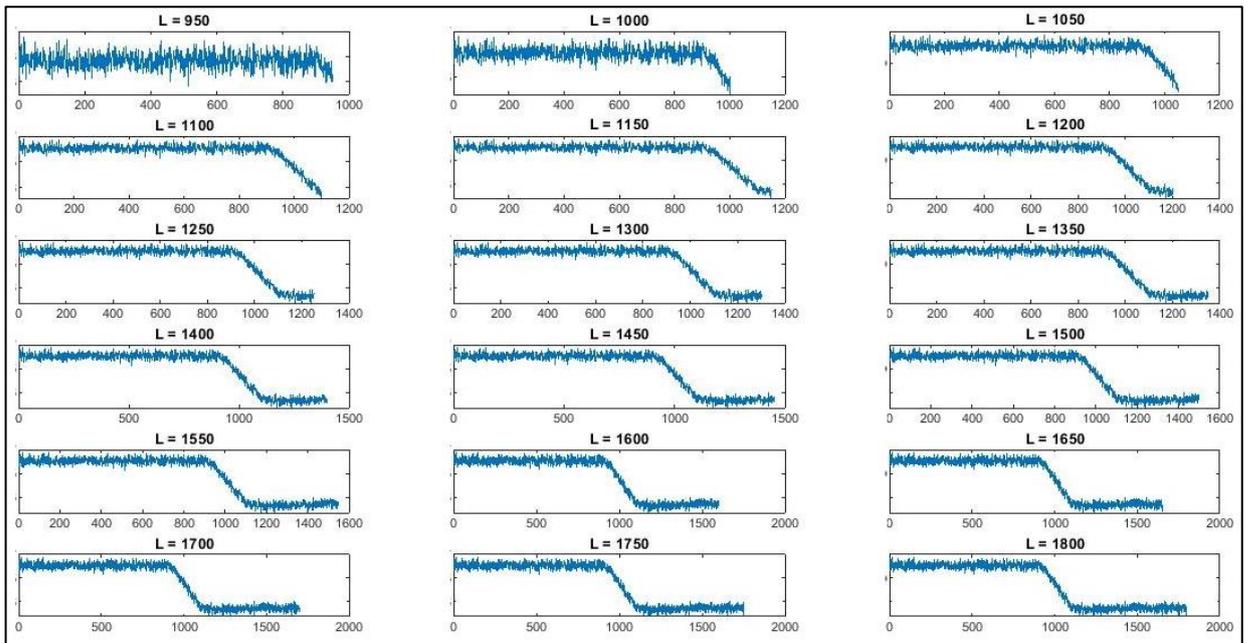


Рис. 18. Выделение события с разными длинами захвата сигнала до и после фронта падения давления

На рис. 19-20 отображена зависимость найденных временных сдвигов от выбранной длины захвата фронта падения давления.

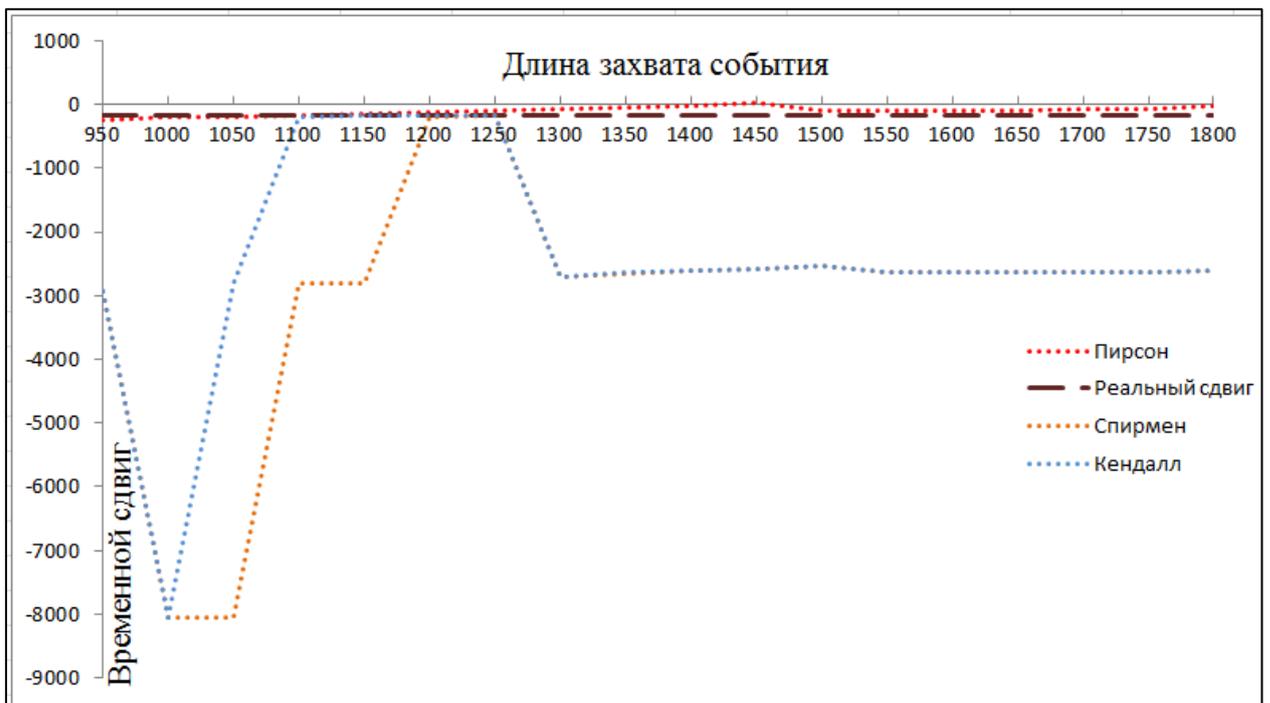


Рис. 19. Нахождение оптимальной длины захвата события для зашумленного сигнала

Отдельной штрихпунктирной линией чёрного цвета на график вынесено реальное значение сдвига. Оптимальной длиной захвата для данного моделирования можно считать то значение, в котором пунктирные линии результатов работы алгоритма с разными коэффициентами корреляции совпадают с линией реального сдвига.

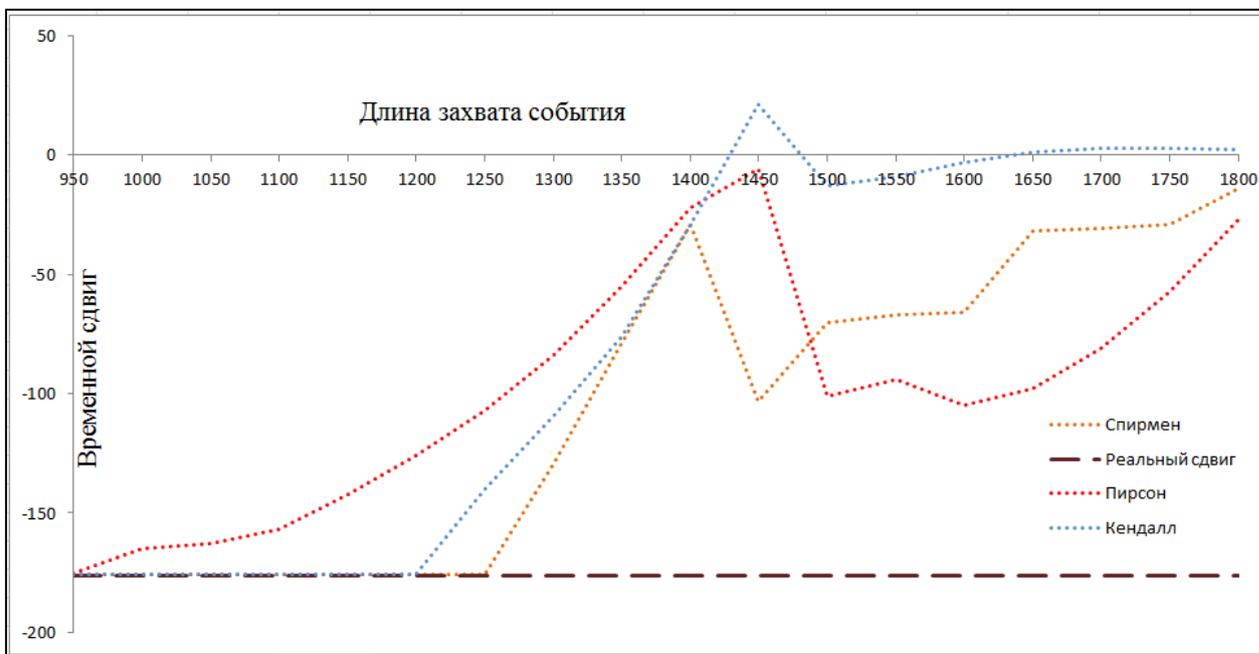


Рис. 20. Нахождение оптимальной длины захвата события для незашумленного сигнала

Метод нахождения сдвигов был проверен на трёх моделированиях двух датчиков (то есть, по 3 пары временных рядов) с зашумлением данных и без добавления шума, с разной длиной захвата данных по обе стороны от фронта падения давления, и с разным стандартным отклонением шума.

Результаты тестирования приведены в таблице 1.

Таблица 1 – Результаты корреляционного алгоритма нахождения сдвигов

№ Моделирования (L – длина фронта)	MAE Кендалл, сек	MAE Спирмен, сек	MAE Пирсон, сек
751 (2, 1) без шума, L=1261	0	0	0,09
751(1, 2) без шума, L=1261	0,32	0,04	0,365
7(1, 2) без шума, L=1200	0,08	0	0,55

7(2, 1) без шума, L=900	0	0	0,235
1(1, 2) без шума, L=1200	89,38	33,52	0,855
1(1, 2) без шума, L=1600	89,38	0	0,92
1(2, 1) без шума, L=1600	0	0	0,64
1(2, 1) без шума, L=1000	0	0	0,035
751(2,1) std=0,025, L=1000	7,22	7,236	4,14
751(2,1) std=0,025, L=1200	0,564	0,437	0,167
751(2,1) std=0,05, L=1200	11,36	11,36	10,24
751(2,1) std=0,0125, L=1200	0,04	0,04	0,19
751(2,1) std(2)=0,025, std(1)=0,0125, L=1200	10,87	10,10	0,417
751(2,1) std(2)=0,0125, std(1)=0,025, L=1200	1,98	1,86	0,279
751(2,1) std(2)=0,05, std(1)=0,025, L=1200	1,077	1,33	0,145
751(2,1) std(2)=0,025, std(1)=0,02, L=1200	0,433	0,563	0,173

Для каждой строки с шумом результаты были получены в ходе 100 прогонов, в качестве оценочной характеристики приведена средняя абсолютная ошибка (mae).

Отметим, что в таблице 1 в первом столбце после номера моделирования в скобках указана последовательность обращения к сигналам. Например, запись 751 (2, 1) означает, что рассматриваются два сигнала моделирования №751, и фронт падения давления выделяется у сигнала 2. Затем находится наибольший коэффициент корреляции между выделенным фронтом и сигналом 1 при движении вдоль сигнала 1 с шагом в один отсчёт и шириной окна равной длине выделенного фронта. Запись 751 (1, 2) означает, что фронт падения давления выделяется у сигнала 1.

Из таблицы 1 и рис. 19-20 заметно, что метод нахождения временных сдвигов, использующий коэффициент корреляции Пирсона, имеет более высокую точность результатов работы на зашумлённых данных, а при

использовании ранговых коэффициентов корреляции точность выше на незашумлённых данных.

В таблице 2 приведены результаты тестирования метода нахождения сдвигов на моделированиях 100-104 с тридцатью разными генерациями шума ($std=0,025$) без предобработки данных и с предобработкой (фильтрация шумов Савицкого-Голея). Средняя абсолютная ошибка (mae) в данной таблице выражается в секундах.

Таблица 2 – MAE

№ Моделирования	MAE Kendall, с	MAE Spearman, с	MAE Pearson, с
100	2,129833	2,129833	2,0485
101	2,906833	2,906833	2,022
102	0,901833	0,928333	0,259667
103	18,902	18,902	11,7625
104	1,365167	1,365167	0,913667
№ Моделирования	MAE Kendall+sgol, с	MAE Spearman+sgol, с	MAE Pearson+sgol, с
100	5,306333	5,473167	0,334
101	76,32167	79,90267	1,826833
102	56,70517	53,65183	0,079667
103	19,15517	18,65417	13,06133
104	4,409167	4,4135	0,089

Предобработка данных не даёт общего улучшения результатов.

Заметим, что точность определения места события будет зависеть не только от точности нахождения временных сдвигов между сигналами, но и от точности измерения скорости распространения волны в продуктопроводе.

Глава 3. Валидация данных

Измеряемые датчиками физические величины представляются в виде сигналов, обработка которых позволяет выявить неустойчивость системы, аварийные ситуации, сбой в работе оборудования или его непригодность. Поступающие с датчика необработанные данные могут содержать ошибки (большие отклонения от средних значений, пропущенные значения, нечисловые или нулевые значения, если таковые являются неприемлемыми), быть сильно зашумленными или вовсе отсутствовать.

Не существует универсального метода валидации, позволяющего выявить ошибочные данные всех типов. Надёжность каждого метода зависит от многих факторов. Следует учитывать тип измеряемых данных, величину измерений, группу рассматриваемых сенсоров и величину рассматриваемого окна данных.

3.1. Виды валидаторов и их задачи

Каждый реализуемый метод валидации данных можно определить как валидатор – инструмент для проверки корректности и достоверности данных.

Можно выделить две группы валидаторов: одномерные и многомерные валидаторы. Одномерные валидаторы обрабатывают сигнал каждого отдельного сенсора и принимают решение о корректности или неисправности сенсора. Многомерные валидаторы обрабатывают сигналы одной группы сенсоров, исходя из предположения о схожести сигналов внутри группы.

Основные задачи, решаемые одномерными валидаторами:

- обнаружение падения измеряемой физической величины в нулевое значение;

- обнаружение неизменяемого во времени (константного) значения;
- обнаружение выбросов (применяется выставление пороговых значений, основанных на предыдущих показаниях, или статистические методы, позволяющие идентифицировать текущую точку данных как выброс относительно рассматриваемого окна данных);
- обнаружение пропусков данных;
- обнаружение изменения уровня шума данных;
- обнаружение изменения статистических характеристик.

Задача многомерной валидации – определить сенсор, кривая сигнала которого сильно отличается от кривых сигналов остальной группы сенсоров. Для решения такой задачи применяются следующие методы:

- моделирование данных одного сенсора на основе данных остальных сенсоров (построение регрессионных моделей типа LPC, ARMA, AR, MA, EMD);
- оценка коэффициентов корреляции одного сигнала с другими;
- моделирование данных с помощью описания физической модели (что, однако, повышает сложность задачи и увеличивает время её выполнения);
- метод пространственной согласованности, основанный на согласованности наблюдений, измеренных на некотором расстоянии. Использование данных двух датчиков приводит к сомнительной достоверности данных, поэтому более эффективными системами для анализа будут являться системы, состоящие из трёх или более датчиков. Такой метод позволяет идентифицировать расходящиеся значения, сомнительные тренды, пробелы данных.

Суть методов многомерной валидации состоит в оценке согласованности реальных данных с модельными данными.

3.2. Многомерный валидатор

Каждый сенсор выражается через остальные сенсоры группы с помощью построения множественной линейной регрессии.

Уравнение множественной линейной регрессии имеет вид:

$$Y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n ,$$

где Y – моделируемый сигнал, b_0, \dots, b_n – коэффициенты уравнения регрессии, x_1, \dots, x_n – остальные сигналы в группе сенсоров.

После чего для смоделированного сенсора подсчитываются остатки модели (разница с настоящим сигналом). Введённая характеристика z_score отвечает за сравнение остатков конкретного сенсора с остатками остальных сенсоров группы. Характеристика z_score в общем виде:

$$z_score_{\{i\}} = \frac{\sigma(e_{\{i\}})}{\bar{\sigma}(e_{N \setminus \{i\}})},$$

где σ – среднеквадратическое отклонение, $\bar{\sigma}$ – среднее среднеквадратическое отклонение, e_i – остатки i -го смоделированного сигнала, N – количество сигналов.

Для удобства принятия решения о неисправности сенсора значения характеристики z_score нормализуются.

$$s = \left| 2 \left(\frac{1}{1 + e^{-z_score/a}} - 0.5 \right) \right|$$

Результаты работы валидатора неоднозначны: в некоторых случаях неисправный сенсор определён верно (рис. 21) – поведение сенсора заметно отличается от поведения остальных сенсоров в группе.

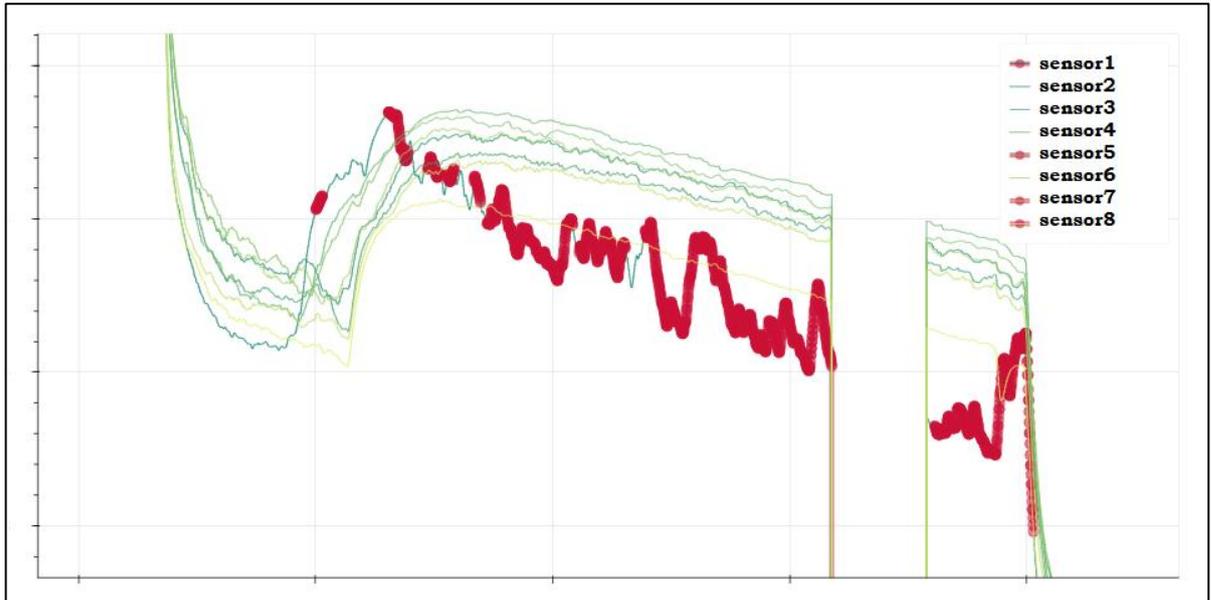


Рис. 21. Результат работы многомерного валидатора

Но также не редки случаи ложного срабатывания валидатора (рис. 22).

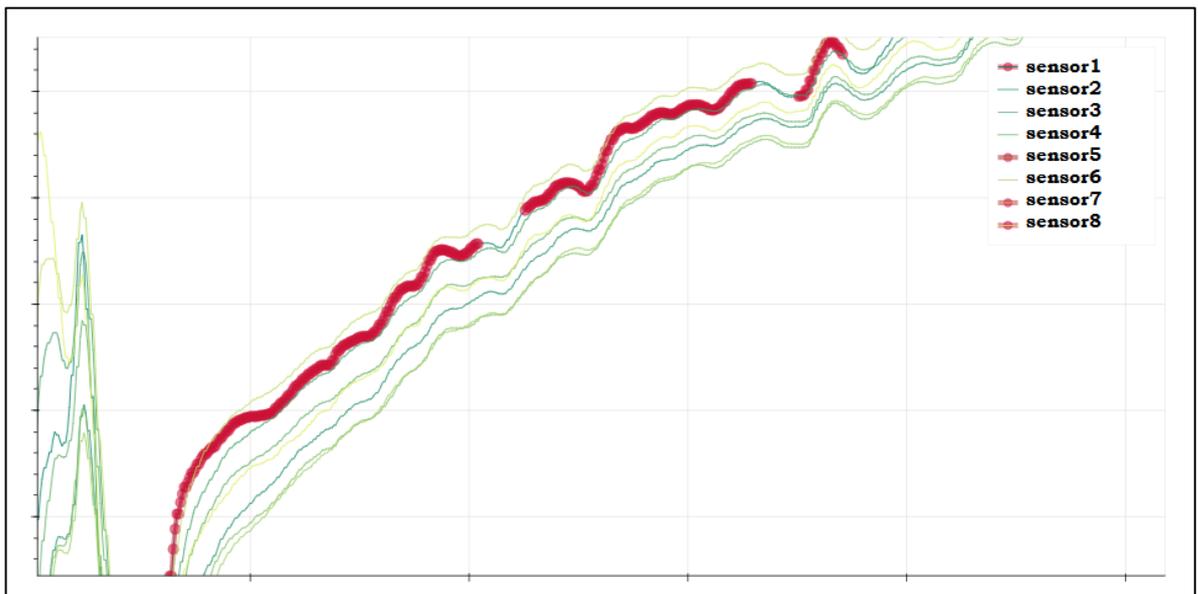


Рис. 22. Результат работы многомерного валидатора

Для решения проблемы ложного срабатывания валидатора были рассмотрены следующие методы и изменения параметров валидатора:

- Добавление регуляризации (L1 или L2, или L1 и L2):

$$L1 = \sum_{i=1}^N (y_i - \hat{y}_i)^2 + \lambda \sum_{i=1}^N |\omega_i| ,$$

$$L2 = \sum_{i=1}^N (y_i - \hat{y}_i)^2 + \lambda \sum_{i=1}^N \omega_i^2,$$

где λ – штрафной коэффициент, ω_i – веса модели.

- Предобработка данных – сглаживание фильтром Савицкого-Голея:

$$\hat{y}_j = \sum_{i=\frac{1-m}{2}}^{\frac{m-1}{2}} C_i y_{j+i}, j = 1, \dots, n$$

где C_i – коэффициенты свёртки, m – окно сглаживания,

$$\frac{m-1}{2} \leq j \leq n - \frac{m-1}{2}$$

- Предобработка данных – сглаживание медианным фильтром:

$$\hat{y}(n) = \text{med}[y(n-k), y(n-k+1), \dots, y(n), \dots, y(n+k-1), y(n+k)],$$

где $2k+1$ – ширина окна сглаживания.

- Предобработка данных – сглаживание фильтром скользящего среднего:

$$\hat{y}(n) = \frac{1}{N} \sum_{i=0}^{N-1} y(n-i),$$

где N – ширина окна сглаживания.

- Нормализация данных до построения линейной регрессии:

$$\hat{X} = \frac{X - \mu}{\sigma},$$

где X – исходная матрица данных, μ – среднее, σ – среднеквадратическое отклонение.

- Наложение ограничений на коэффициенты регрессии (неотрицательные коэффициенты):

$$\arg \min_x \|Ax - y\|_2, \quad x \geq 0.$$

- Подбор параметра a функции нормализации s .

- Альтернативные способы вычисления z_score :

$$1. z_{score\{i\}} = \max \left\{ 0, \frac{e_{\{i\}} - e_{\{o\}}}{\sigma(e_{\{o\}})} \right\}, \quad e_{\{o\}} = \overline{e_{N \setminus \{i\}}},$$

$$2. z_{score\{i\}} = \max \left\{ 0, \frac{1 - \sigma(e_{\{o\}})}{\sigma(e_{\{i\}})} \right\},$$

$$3. z_{score\{i\}} = \left| \frac{\overline{e_{\{i\}}} - \overline{e_{\{o\}}}}{\sigma(e_{\{o\}})} \right|,$$

$$4. z_{score\{i\}} = \max \left\{ 0, \frac{\overline{e_{\{i\}}} - \overline{e_{\{o\}}}}{\sigma(e_{\{o\}})} \right\}.$$

Результаты работы многомерного валидатора при добавлении регуляризации и фильтрации данных стали хуже, а изменение параметра a функции нормализации наоборот принесло улучшение (снижение количества ложного срабатывания валидатора).

Изменение типа фильтрации (с медианного фильтра на фильтр Савицкого-Голея) не принесло значительных улучшений (рис. 23 – 24).

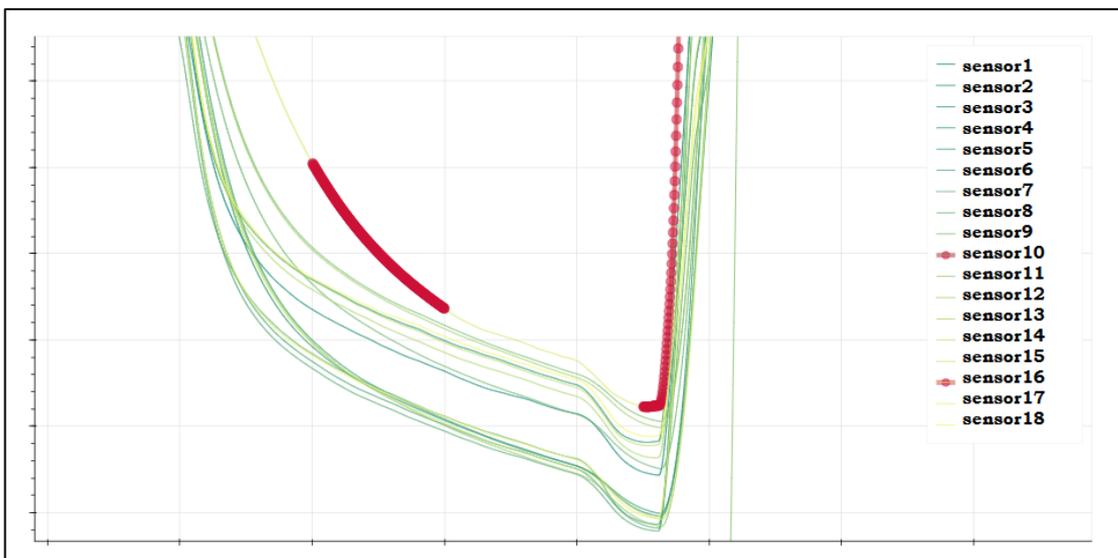


Рис. 23. Результат работы многомерного валидатора при изменении метода фильтрации

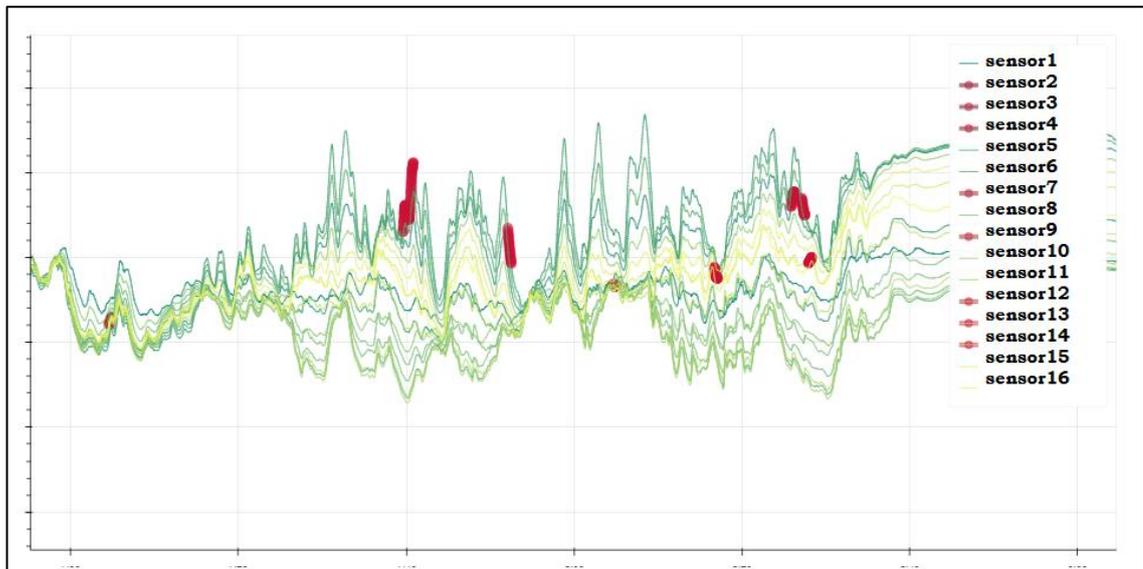


Рис. 24. Результат работы многомерного валидатора при изменении метода фильтрации

Метод построения линейной регрессии неэффективен в случае, когда несколько сенсоров в группе ведут себя одинаково, а неисправный сенсор является их зеркальным отражением. Тогда неисправный сенсор можно выразить через любой нормальный сенсор со знаком минус, что приведёт к точному предсказанию неисправного сенсора, и как следствие, небольшим остаткам построенной модели. Этого удаётся избежать при построении линейной регрессии с неотрицательными коэффициентами.

В результате сравнения работы валидатора при разных вариантах подсчёта z_score с нормализацией и без нормализации данных перед построением линейной регрессии предпочтение отдаётся методу с нормализацией данных. Отметим, что формальной метрики для сравнения модификаций валидатора нет, поэтому лучшая модификация определялась наименьшим количеством ложных срабатываний валидатора.

Модель линейной регрессии с неотрицательными коэффициентами решает проблему пропущенных неисправностей с «зеркальным» сенсором. Подобран оптимальный параметр функции нормализации s . Регуляризация и фильтрация данных не используются.

Выводы

Детектирование событий с помощью алгоритма кумулятивных сумм позволяет находить даже самые мелкие события. При необходимости, чувствительность алгоритма можно повысить с помощью уменьшения порога допустимого изменения решающей функции.

Представление всех сигналов в виде одного с помощью SVD-разложения может помочь в случае сильного зашумления сигналов и слабой выраженности события.

Для обнаружения момента первого отражения волны предпочтение отдается кепстральному анализу. Для улучшения точности кепстрального анализа следует предварительно использовать фильтрацию шумов.

Также в рамках кепстрального анализа можно изучить применение кепстрограммы, что позволит избежать неточностей ручного построения метода. На рис. 25 изображены два исходных сигнала (незашумленный и зашумленный). А на рис.26-27 соответствующие им кепстрограммы.

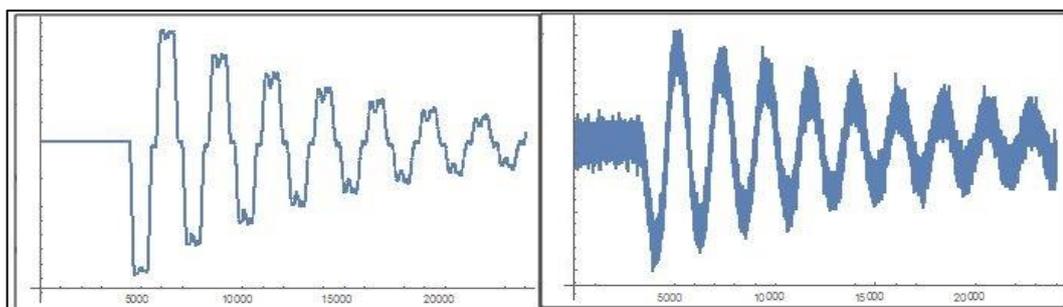


Рис. 25. Исходные сигналы

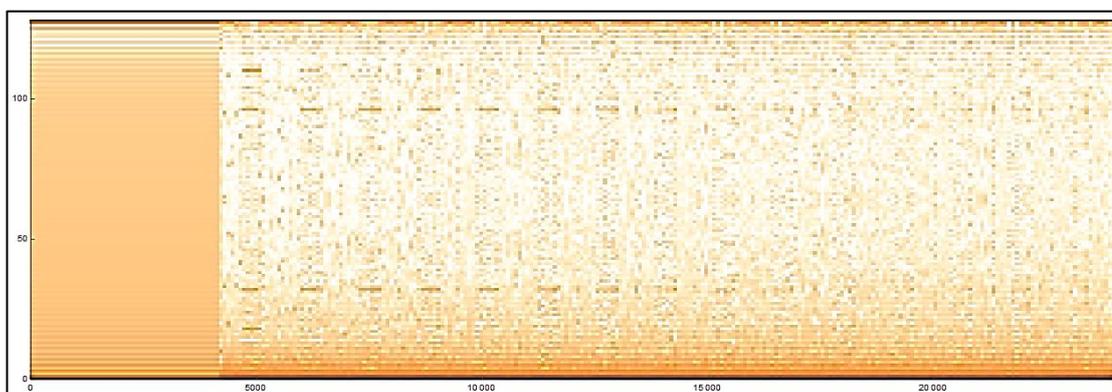


Рис. 26. Кепстрограмма незашумлённого сигнала

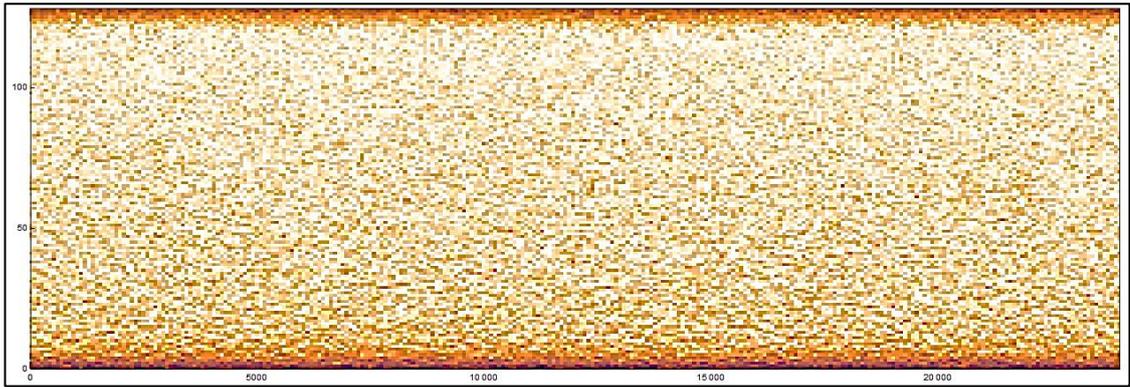


Рис. 27. Кепстрограмма зашумлённого сигнала

Корреляционный метод можно сделать адаптивным – основываясь на текущем уровне шума (а точнее, его std) выбирать коэффициент корреляции (Пирсона, Кендалла или Спирмена).

Для многомерного валидатора в качестве постобработки можно проводить анализ корреляционной матрицы группы сенсоров: если данные сенсора, отмеченного как неисправный, значительно коррелируют с данными другого сенсора в группе (коэффициент корреляции > 0.9), то отметка о неисправности снимается.

Заключение

Проведён обзор алгоритмов детектирования и локализации событий на продуктопроводе, методов валидации данных. Предложен новый подход к осуществлению многомерной валидации. Разработана программная реализация рассмотренных алгоритмов.

Приведены результаты работы методов на тестовых и реальных данных. На основании результатов можно заметить, что предложенный подход детектирования событий с помощью алгоритма кумулятивных сумм даёт достаточно точные результаты, а также позволяет сократить количество пропущенных событий. Метод SVD-разложения будет полезен при детектировании событий на многоканальных системах. Альтернативные методы детектирования (спектральный и кепстральный анализ) не дают высокой точности.

Результаты нахождения временной задержки между приходом события на разные датчики не так хороши. Предложения по адаптации метода для улучшения результатов вынесены в разделе «Выводы».

Реализация многомерного валидатора успешно прошла проверку на реальных данных. При предложенной модификации валидатора снижается количество ложных срабатываний валидатора.

Список литературы

[1] Анпилогов В.А., Веревкин А.П., Давлетшина З.Р., Кизина И.Д., Гайнуллин Д.В. Обнаружение утечек и несанкционированных воздействий в магистральных нефтепроводах с использованием Фурье- и вейвлет-преобразований в ПТК СОУ // Автоматизация, телемеханизация и связь в нефтяной промышленности. – 2012. – №4. – с. 28-31.

[2] Бродский Б.Е. Асимптотически оптимальные методы в задаче скорейшего обнаружения разладки. I. Характеристики методов скорейшего обнаружения разладки // Автоматика и телемеханика. – 1995. – №9. – с. 60-72.

[3] Бродский Б.Е. Асимптотически оптимальные методы в задаче скорейшего обнаружения разладки. II. Исследование методов скорейшего обнаружения // Автоматика и телемеханика. – 1995. – №10. – с. 50-59.

[4] Бродский Б.Е., Дарховский Б.С. Проблемы и методы вероятностной диагностики // Автоматика и телемеханика. – 1999. – №8. – с. 3-50.

[5] Бродский Б.Е., Дарховский Б.С. Сравнительный анализ некоторых непараметрических методов скорейшего обнаружения момента “разладки” случайной последовательности // Теория вероятностей и ее применения. – 1990. – т.35, №4. – с. 655-668.

[6] Буре В.М., Парилина Е.М. Теория вероятностей и математическая статистика. М.: Лань, 2013. 416 с.

[7] Васянович А.В., Грицунова А.В., Лебедев О.Г., Чурюмов Г.И. Эволюция спектров сигналов при переходных процессах в генераторах и усилителях М-типа // Радиоэлектроника и информатика. – 2005. – №1. – с. 22-28.

[8] Гоноровский И. С. Радиотехнические цепи и сигналы: Учебник для вузов.— 4-е изд., перераб. и доп. — М.: Радио и связь, 1986. – с. 20-25.

[9] Жиглявский А.А., Красковский А.Е. Обнаружение разладки случайных процессов в задачах радиотехники. – Л.: ЛГУ, 1988. 112 с.

[10] Каблов Е.Н. Инновационные разработки ФГУП «ВИАМ» ГНЦ РФ по реализации «Стратегических направлений и технологий их переработки на период до 2030 года» // Авиационные материалы и технологии. – 2015. – №1(34). – с. 3-33.

[11] Концепция предаварийного радиационного и газового контроля на кораблях ВМФ. – СПб.: 1 ЦНИИ МО РФ, 1997. – 11с.

[12] Мамонова Т.Е. Методы диагностики линейной части нефтепроводов для обнаружения утечек // Проблемы информатики. – 2012. №5(17). – с.103-112.

[13] Мамонова Т.Е. Учёт геометрического профиля нефтепровода при определении параметров утечки // Электронный научный журнал «Нефтегазовое дело». – 2012. – №2. – с. 85-102.

[14] Никифоров И.В. Последовательное обнаружение изменения свойств временных рядов / Академия наук СССР. Институт проблем управления. – М.: Наука, 1983. – 200 с.

[15] Никифоров И.В. Применение кумулятивных сумм для обнаружения изменения характеристик случайного процесса // Автоматика и телемеханика. – 1979. – №2. – с. 48-58.

[16] Никонов А. В., Давлетшин Р. В., Яковлева Н. И., Лазарев П. С. Фильтрация методом Савицкого-Голея спектральных характеристик чувствительности матричных фотоприемных устройств // Успехи прикладной физики. – 2016. –т.4, № 2. – с.198-205.

[17] Осипов Ю.М. Частотный и временной анализ стационарных и переходных характеристик линейных электрических цепей. Учебное пособие по курсам электротехники и ТОЭ. Ч. 2. – СПб: СПбГИТМО (ТУ), 2002. –99 с.

[18] Федотов М.Ю., Сорокин К.В., Гончаров В.А., Шиенок А.М., Зеленский П.В. Возможности сенсорных систем и интеллектуальных ПКМ на

их основе // Все материалы. Энциклопедический справочник. – 2013. – №2. – с.18-23.

[19] Aimé Lay-Ekuakille, Giuseppe Griffo, Paolo Visconti, Patrizio Primiceri, Ramiro Velazquez. Leak detection in waterworks: comparison between STFT and FFT with an overcoming of limitations // Metrology and Measurement Systems, 2017. 24(4). Pp. 631-644.

[20] Bogert B. P., Healy M. J. R., Tukey J. W. The Quefrency Analysis of Time Series for Echoes: Cepstrum, Pseudo Autocovariance, Cross-Cepstrum and Saphe Cracking // Symposium on Time Series Analysis (M. Rosenblatt, Ed) Chapter 15, – New York: Wiley, 1963. Pp. 209-243.

[21] Eckart C., Young G. The Approximation of One Matrix by Another of Lower Rank. Psychometrika, 1936. 4. Pp. 81-93.

[22] Ivan Miguel Pires, Nuno M. Garcia, Nuno Pombo, Francisco Flórez-Revuelta, and Natalia Díaz Rodríguez. Validation Techniques for Sensor Data in Mobile Health Applications // Journal of Sensors, 2016. 9 p.

[23] Lin Chen, Yongchun Liu, Fancheng Kong, Na He a. Acoustic Source Localization Based on Generalized Crosscorrelation Time-delay Estimation // Procedia Engineering, 2011. Vol. 15. Pp. 4912-4919.

[24] M. Taghvaei, S.B.M Beck and W.J. Staszewski. Leak detection in pipelines using cepstrum analysis // Measurement science and technology, 2006. 17. Pp. 367-372.

[25] Motazedi, N. and Beck, S. Leak detection using cepstrum of cross-correlation of transient pressure wave signals // Proc IMechE Part C: J Mechanical Engineering Science 0(0), 2017. 13 p.

[26] Rhudy, M., Bucci, B., Vipperman, J., Allanach, J., & Abraham, B. Microphone Array Analysis Methods Using Cross-Correlations. Volume 15: Sound, Vibration and Design. – 2009.

[27] Rhudy, Matthew. Real Time Implementation of a Military Impulse Classifier. University of Pittsburgh, Master's Thesis, 2009. 104 p

[28] Siao Sun; Jean-luc Bertrand-krajewski, Anders Lynggaard-Jensen, Joep van den Broeke, Florian Edthofer, Maria do Céu Almeida, Álvaro Silva Ribeiro, José Menaia. Literature review of data validation methods // Seventh framework programme, 2011. Vol. 019. 19 p

Приложение

Детектирование. Основной код:

```
load data1.txt; X = data1;

X1 = X(:,3:5);
T = 120;
dft =
[126002,180202,228402,312602,385002,457002,469202,625402,673602,
745802,774602,866202,926422,1010602,1070802,1251002,1311402,1395
402,1455602,1527802,1612002];
dft1 = dft1 - 600;

figure(1);

for k = 1:3
subplot(3,1,k);

    plot(X1(:,k),'LineWidth',2); grid;
    hold on;

    for i = 1:length(dft)

        plot([dft(i) dft(i)], [0 14]','black','LineWidth',1);
        hold on;

    end

end

end

for k = 1:3
    figure(k+1);

    ysgol=sgolayfilt(X1(:,k),2,601);
    g = CS_f(ysgol,T);
    d = DS_f(g);
    r = [d(1)];
    l = [d(1)];

    for i = 2:length(d);

        if d(i)-d(i-1)>10000

            r = [r d(i-1) d(i)];
            l = d(i);

        end

    end

end

end
```

```

r = [r d(length(d))];
plot(g, 'r', 'LineWidth', 2); grid;
hold on;

for i = 1:length(dft)

    plot([dft1(i) dft1(i)]', [0
max(g)]', 'black', 'LineWidth', 1);
    hold on;

end

for p = 1:2:length(r)

    block = rectangle('Position', [r(p) 0 (r(p+1)-r(p))
max(g)]')
    block.FaceColor = [0 0 1 0.5];
    alpha(0.5);
    block.EdgeColor = 'b';
    block.LineWidth = 2.5;
    hold on;

end

end

```

Детектирование. АКС:

```

function g = CS_f(x, T)

N = length(x);
g = zeros(1, N - T);

for k = 1:N - T
    y = x(k:k + T);
    z = cumsum(y);
    z0 = min(z);
    g(k) = z(T + 1) - z0;

end

end

```

Детектирование. Выставление порога суммы:

```

function d=DS_f(g)

d = [];
n = length(g);

for m = 201:n
    if abs(g(m)-g(m-600)) >= 20

```

```

        d = [d m];

    end

end

```

Спектральный анализ:

```

load data.mat;
y = data;
y = y-mean(y);
Signal = y';

Tm = 43.669;
Fd = 12000/60;
i = round(log2(length(y)));
FftL = 2^i;
T = 0:1/Fd:Tm;

FftS = abs(fft(Signal,FftL));
FftS = 2*FftS./FftL;
FftS(1) = FftS(1)/2;

figure(4)
subplot(2,1,1);
plot(T,Signal);
title('Сигнал');
xlabel('Время (с)');
ylabel('Амплитуда');

F = 0:Fd/FftL:Fd/2-Fd/FftL;
subplot(2,1,2);
plot(F,FftS(1:length(F))),grid;
title('Спектр сигнала');
xlabel('Частота (Гц)');
ylabel('Амплитуда');

```

Кепстральный анализ:

```

load data.mat; y = data;

%c = rceps(autocorr(y,'NumLags',length(y)-1)); autocepstrum
c = rceps(y); %real cepstrum

%y2 = dct(log(abs(spectrogram(y))));
% c = [y2(:,1)' y2(:,2)' y2(:,3)' y2(:,4)' y2(:,5)' y2(:,6)'
y2(:,7)' y2(:,8)'];
c = [y2(:,1)' y2(:,2)' y2(:,3)' y2(:,4)' y2(:,5)' y2(:,6)'
y2(:,7)'];
Fs = 12000/60;
t = (0:length(c)-1)/Fs;
t1 = (0:length(y)-1)/Fs;

```

```

[px,locs] = findpeaks(c,'Threshold',5,'MinPeakDistance',0.2);

clf;
plot(t,c,t(locs),px,'o');
xlabel('Time (s)');
%
maxpix = max(px);

dl = locs(1)-1;
alpha = 0.5;
yNew = filter(1,[1 zeros(1,dl-1) alpha],y);

for i=1:(length(locs)/2)

    if i==1

        yNew(locs(i):length(yNew)) = filter(1,[1
alpha],yNew(locs(i):length(yNew))*alpha+1.28*(yNew(dl)));

    else

        yNew(locs(i):length(yNew)) = filter(1,[1
alpha],yNew(locs(i):length(yNew))*alpha+(mean(yNew(locs(i-
1):locs(i)))));

    end

end

figure;
plot(t1,y);
hold on;

maxpix = max(px);
for i = 1:length(locs)

    block = rectangle('Position',[t(locs(i)) min(y) 0 (max(y)-
min(y))]');

    if px(i) == maxpix || abs(px(i)- maxpix) < 0.0000001

        block.EdgeColor = 'r';

    else

        block.EdgeColor = 'black';

    end

end

block.LineWidth = 1.5;
hold on;

```

```

end

plot(t1,yNew,'r');
hold on;

for i = 1:length(locs)

    block = rectangle('Position',[t(locs(i)) (min(yNew)) 0
(max(yNew)-min(yNew))]');

    if px(i) == maxpix || abs(px(i)- maxpix) < 0.0000001

        block.EdgeColor = 'r';

    else

        block.EdgeColor = 'black';

    end

    block.LineWidth = 1.5;
    hold on;

end

xlabel('Time (s)');

```

Корреляционный алгоритм на нескольких моделированиях с разным уровнем шума:

```

load data.mat;load data1.mat;

y = data; x1 = data1;

statK = zeros(1, 100); shiftK = zeros(1, 100);
diffK = -176.07485979489545*ones(1, 100);
diff1K = (-176.07485979489545/200)*ones(1, 100);

statS = statK; statP = statK;
shiftS = shiftK; shiftP = shiftK;
diffS = diffK; diffP = diffK;
diff1S = diff1K; diff1P = diff1K;

for m=1:100

    Noise1=0.025*randn(length(y),1);
    Noise2=0.025*randn(length(x1),1);
    y = y + Noise1;
    x1 = x1 + Noise2;
    t1 = 3500; t2 = 4700;
    x = x1(t1:t2);

```

```

RK = corr(x,y(1:length(x)),'Type','Kendall');
RS = corr(x,y(1:length(x)),'Type','Spearman');
RP = corr(x,y(1:length(x)),'Type','Pearson');

loc = 1;

for i=2:length(y)-length(x)

    RK1 = corr(x,y(i:i+length(x)-1),'Type','Kendall');
    RS1 = corr(x,y(i:i+length(x)-1),'Type','Spearman');
    RP1 = corr(x,y(i:i+length(x)-1),'Type','Pearson');

    if RK1>RK || isnan(RK)

        RK = RK1;
        locK = i;

    end

    if RS1>RS || isnan(RS)

        RS = RS1;
        locS = i;

    end

    if RP1>RP || isnan(RP)

        RP = RP1;
        locP = i;

    end

end

end

shiftK(m) = (t1 - locK); %сдвиг в отсчётах
statK(m) = (t1 - locK)/200; %сдвиг в секундах
diffK(m) = diffK(m) - shiftK(m); %разница с реальным сдвигом
в отсчётах
diff1K(m) = diff1K(m) - statK(m); %разница с реальным
сдвигом в секундах

shiftS(m) = (t1 - locS); %сдвиг в отсчётах
statS(m) = (t1 - locS)/200; %сдвиг в секундах
diffS(m) = diffS(m) - shiftS(m); %разница с реальным сдвигом
в отсчётах
diff1S(m) = diff1S(m) - statS(m); %разница с реальным
сдвигом в секундах

shiftP(m) = (t1 - locP); %сдвиг в отсчётах
statP(m) = (t1 - locP)/200; %сдвиг в секундах

```

```

    diffP(m) = diffP(m) - shiftP(m); %разница с реальным сдвигом
в отсчётах
    diff1P(m) = diff1P(m) - statP(m); %разница с реальным
сдвигом в секундах

end

```

Тестирование на нормальное распределение и на равенство средних

```

load data.mat;
y = data;
y1 = y - wdenoise(y);
h1 = zeros(1, length(y1)-201);
ptest1 = zeros(1, length(y1)-201);
L = zeros(1, length(y1)-201);
KS = zeros(1, length(y1)-201);
CH = zeros(1, length(y1)-201);

for I = 1:length(y1)-200

%     Лиллиефорса
    L(i) = lillietest(y1(i:i+200));
%     Колмогорова - Смирнова
    KS(i) = kstest(y1(i:i+200));
%     Хи-Квадрат
    CH(i) = chi2gof(y1(i:i+200));
%     Стьюдента
    [h1(i),ptest1(i)] = ttest(y1(i:i+200));

end

subplot(4,1,1);
plot(L);
title("Тест Лиллиефорса");
subplot(4,1,2);
plot(KS);
title("Тест Колмогорова - Смирнова");
subplot(4,1,3);
plot(CH);
title("Тест Хи-Квадрат");
subplot(4,1,4);
plot(h1);
title("Тест Стьюдента");

figure(2);
subplot(2,1,1);
plot(y); hold on; plot(wdenoise(y));
title("Исходный сигнал");
subplot(2,1,2);
plot(y1);
title("Шум");

```