

Школа: ИШИТР Инженерная школа информационных технологий и робототехники
 Направление подготовки 09.03.01 «Информатика и вычислительная техника»
 Отделение школы (НОЦ): Отделение информационных технологий

БАКАЛАВРСКАЯ РАБОТА

Тема работы
Исследование применимости метода SSA для обработки временных рядов в нефтегазовой отрасли

УДК 519.246.8:622.32

Студент

Группа	ФИО	Подпись	Дата
3-8В41	Белуха Евгений Евгеньевич		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Буркатовская Ю.Б.	к.ф.-м.н.		

КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН	Криницына З.В.	к.т.н., Доцент		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ассистент ООД	Мезенцева И.Л.			

ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
09.03.01 Информатика и ВТ	Погребной А. В.	к.т.н., Доцент		

**Планируемые результаты обучения по основной образовательной
программе подготовки бакалавров 09.03.01 «Информатика и
вычислительная техника»**

Код результата	Результат обучения (выпускник должен быть готов)
<i>Профессиональные компетенции</i>	
P1	Применять <i>глубокие</i> естественнонаучные, математические и инженерные <i>знания</i> для создания и обработки новых материалов
P2	Применять <i>глубокие</i> знания в области современных технологий машиностроительного производства для решения <i>междисциплинарных</i> инженерных задач
P3	Ставить и решать <i>инновационные</i> задачи <i>инженерного анализа</i> , связанные с созданием и обработкой материалов и изделий, с использованием системного анализа и моделирования объектов и процессов машиностроения
P4	Разрабатывать технологические процессы, <i>проектировать</i> и использовать <i>новое</i> оборудование и инструменты для обработки материалов и изделий, конкурентоспособных на <i>мировом</i> рынке машиностроительного производства
P5	Проводить теоретические и экспериментальные <i>исследования</i> в области современных технологий обработки материалов, нанотехнологий, создания <i>новых</i> материалов в <i>сложных</i> и неопределенных условиях
P6	Внедрять, <i>эксплуатировать</i> и обслуживать современные высокотехнологичные линии автоматизированного производства, обеспечивать их <i>высокую эффективность</i> , соблюдать правила охраны здоровья и безопасности труда на машиностроительном производстве, выполнять требования по защите окружающей среды
<i>Универсальные компетенции</i>	
P7	Использовать <i>глубокие знания</i> по <i>проектному менеджменту</i> для ведения <i>инновационной</i> инженерной деятельности с учетом юридических аспектов защиты интеллектуальной собственности
P8	<i>Активно</i> владеть <i>иностранным языком</i> на уровне, позволяющем работать в иноязычной среде, разрабатывать документацию, презентовать и защищать результаты инновационной инженерной деятельности
P9	Эффективно работать индивидуально, в качестве <i>члена и руководителя группы</i> , состоящей из специалистов различных направлений и квалификации, демонстрировать ответственность за результаты работы и готовность <i>следовать корпоративной культуре</i> организации
P10	Демонстрировать <i>глубокие знания социальных, этических и культурных аспектов</i> инновационной инженерной деятельности, компетентность в вопросах <i>устойчивого развития</i>
P11	<i>Самостоятельно учиться</i> и непрерывно <i>повышать квалификацию</i> в течение всего периода профессиональной деятельности

Министерство науки и высшего образования Российской Федерации
 федеральное государственное автономное
 образовательное учреждение высшего образования
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Школа ИШИТР Инженерная школа информационных технологий и робототехники
 Направление подготовки (специальность) 09.03.01 «Информатика и вычислительная техника»
 Отделение школы (НОЦ) Отделение информационных технологий

УТВЕРЖДАЮ:
 Руководитель ООП

 (Подпись) (Дата) (Ф.И.О.)

ЗАДАНИЕ на выполнение выпускной квалификационной работы

В форме:

бакалаврской работы

Студенту:

Группа	ФИО
38В41	Белуха Евгений Евгеньевич

Тема работы:

Исследование применимости метода SSA для обработки временных рядов в нефтегазовой отрасли	
Утверждена приказом директора (дата, номер)	От 11.03.2019 №1799/с

Срок сдачи студентом выполненной работы:

--	--

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

Исходные данные к работе	Техническое задание к исследованию применимости метода SSA для набора данных технологических процессов нефтегазовой отрасли; Набор данных.
Перечень подлежащих исследованию, проектированию и разработке вопросов	Изучение и представление метода; предварительный анализ набора данных и выбор временных рядов для анализа; отработка методики анализа и создание метрики выбора модели; построение моделей и их описание; обнаружение разладки временного ряда; прогнозирование по полученной модели.

Перечень графического материала	Результаты проведенного исследования
Консультанты по разделам выпускной квалификационной работы	
Раздел	Консультант
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Доцент отделения социально-гуманитарных наук Криницына З. В.
Социальная ответственность	Ассистент отделения общетехнических дисциплин Мезенцева И.Л.

Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику	
---	--

Задание выдал руководитель):

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ	Буркатовская Юлия Борисовна	к. ф.-м. н., Доцент		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
38В41	Белуха Евгений Евгеньевич		

**ЗАДАНИЕ ДЛЯ РАЗДЕЛА
«ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И
РЕСУРСОСБЕРЕЖЕНИЕ»**

Студенту:

Группа	ФИО
3-8В41	Белуха Евгений Евгеньевич

Школа	ИШИТР	Отделение школы (НОЦ)	ОИТ
Уровень образования	Бакалавриат	Направление/специальность	09.03.01 Информатика и ВТ

Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:

1. Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих	Затраты на выполнение НИР включают в себя материальные затраты, основную и дополнительную заработную плату исполнителей, отчисления на социальные нужды, накладные расходы
2. Нормы и нормативы расходования ресурсов	Премияльный коэффициент 30%; Коэффициент доплат и надбавок 15%; Коэффициент, учитывающий накладные расходы 16%; Районный коэффициент 30%.
3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования	Отчисления по страховым взносам – 27,1% от ФОТ

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. Оценка коммерческого потенциала, перспективности проведения НИ с позиции ресурсоэффективности и ресурсосбережения	- Потенциальные потребители результатов исследования; - Стратегии развития НИ как проекта
2. Планирование и формирование бюджета научных исследований	Расчет бюджета НИ
3. Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования	Определение показателя ресурсоэффективности, сравнительной эффективности вариантов исполнения

Перечень графического материала (с точным указанием обязательных чертежей):

1. Оценка конкурентоспособности технических решений
2. Матрица SWOT
3. График проведения и бюджет НИ
4. Оценка ресурсоэффективности НИ

Дата выдачи задания для раздела по линейному графику	27.03.19
--	----------

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН ШБИП	Криницына Зоя Васильевна	к.т.н., доцент		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
3-8В41	Белуха Евгений Евгеньевич		

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа	ФИО
3-8В41	Белуха Евгений Евгеньевич

Школа	ИШИТР	Отделение (НОЦ)	ОИТ
Уровень образования	Бакалавриат	Направление/специальность	09.03.01 Информатика и ВТ

Тема ВКР:

Исследование применимости метода SSA для обработки временных рядов в нефтегазовой отрасли	
Исходные данные к разделу «Социальная ответственность»:	
1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	Статистический метод SSA для анализа данных в нефтегазовой отрасли.
Перечень вопросов, подлежащих исследованию, проектированию и разработке:	
1. Правовые и организационные вопросы обеспечения безопасности: – специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства; – организационные мероприятия при компоновке рабочей зоны.	ТК РФ от 30.12.2001 N 197-ФЗ СанПиН 2.2.2/2.4.1340-03 ТОИ Р-45-084-01
2. Производственная безопасность: 2.1. Анализ выявленных вредных и опасных факторов 2.2. Обоснование мероприятий по снижению воздействия	Отклонение показателей микроклимата Электромагнитное излучение экрана монитора ПК Недостаточная освещенность рабочей зоны Перенапряжение зрительного анализатора Опасность поражения электрическим током
3. Экологическая безопасность:	Риск загрязнения литосферы, гидросферы, атмосферы при ненадлежащей утилизации ламп освещения рабочего места
4. Безопасность в чрезвычайных ситуациях:	Среди возможных ЧС типичной может быть возникновение пожара в здании

Дата выдачи задания для раздела по линейному графику	27.03.19 г.
--	-------------

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
ассистент ООТД	Мезенцева И.Л.			

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
3-8В41	Белуха Евгений Евгеньевич		

РЕФЕРАТ

Выпускная квалификационная работа 102 с., 57 рис., 14 табл., 26 источников, 2 прил.

Ключевые слова: метод SSA, сингулярный спектральный анализ, «Гусеница», Python, анализ данных, временные ряды, разладка.

Объектом исследования являются исторические данные о технологических процессах скважины томского нефтегазового месторождения.

Цель работы: исследование базового алгоритма метода SSA и его применение для построения моделей ВР, обнаружение разладки и прогнозирование с помощью полученной модели

В процессе исследования был освоен базовый метод SSA, его реализация на ЯП Python, представлено наглядное описание метода, определены оптимальные параметры для анализа и построения моделей ВР, выработана методика их анализа, используемое ПО было дополнено функциями и методами для проведения анализа. С помощью SSA были проанализированы данные нефтегазовой отрасли, определена и продемонстрирована применимость метода: построены модели, описана структура, определена разладка, сделан прогноз.

В результате исследования были получены модели, демонстрирующие применимость метода SSA для обработки временных рядов в нефтегазовой отрасли.

Результаты работы могут быть применены в экспертных системах в нефтегазовой отрасли.

Область применения: анализ данных в нефтегазовой отрасли.

Экономическая эффективность и значимость работы, себестоимость исследуемой системы представлена в главе финансовый менеджмент, ресурсоэффективность и ресурсосбережения.

СПИСОК СОКРАЩЕНИЙ

SSA – сингулярный спектральный анализ;
SVD – сингулярное разложение;
lmin – локальный минимум;
АКФ – автокорреляционная функция;
ВР – временной ряд;
ВЧ – высокая частота, высокочастотный;
ГМ – ганкелева матрица;
НЧ – низкая частота, низкочастотный;
ПО – программное обеспечение;
ТМ – траекторная матрица;
ТП – технологический процесс, техпроцесс;
ЯП – язык программирования.

ОГЛАВЛЕНИЕ

РЕФЕРАТ.....	7
СПИСОК СОКРАЩЕНИЙ.....	8
ВВЕДЕНИЕ.....	11
1 ОБЗОР ЛИТЕРАТУРЫ.....	13
2 ОБЪЕКТЫ И МЕТОДЫ ИССЛЕДОВАНИЯ.....	14
2.1 Объект, метод и постановка задач.....	14
2.2 Границы и условия.....	14
2.3 Выбор средств реализации исследования.....	16
3 РАСЧЕТЫ И АНАЛИТИКА.....	17
3.1.1 Представление метода SSA.....	17
3.1.2 Разложение траекторной матрицы.....	19
3.1.3 Обобщение SVD-разложения.....	23
3.1.4 Восстановление ВР.....	25
3.1.5 Выделение компонентов ВР и их группировка.....	31
3.2 Предварительный анализ и подготовка данных.....	34
3.3.1 Анализ и моделирование.....	36
3.3.2 ВР активной мощности, $L = 72$	38
3.3.3 ВР активной мощности, $L = 9$	44
3.3.4 ВР активной мощности, $L = 18$	47
3.3.5 ВР буферного давления.....	49
3.3.6 ВР тока двигателя.....	55
3.3.7 Определение разладки.....	58
3.3.8 Построение прогноза.....	61
4 РЕЗУЛЬТАТЫ ПРОВЕДЕННОГО ИССЛЕДОВАНИЯ.....	62
5.1 Потенциальные потребители результатов исследования.....	63
5.2 SWOT-анализ.....	65
5.3 Планирование научно-исследовательских работ.....	66
5.3.1 Структура в рамках научного исследования.....	66
5.3.2 Определение трудоемкости выполнения работ.....	67
5.3.3 Разработка графика проведения научного исследования.....	67
5.4 Бюджет научно технического исследования (НТИ).....	71
5.4.1 Расчет материальных затрат.....	71
5.4.2 Основная заработная плата исполнителей темы.....	72
5.4.3 Дополнительная заработная плата исполнителей темы.....	74
5.4.4 Отчисления во внебюджетные фонды (страховые отчисления).....	74
5.4.5 Накладные расходы.....	75
5.4.6 Формирование бюджета затрат научно-исследовательского проекта.....	75
5.5 Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования.....	75
6 СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ.....	78
6.1 Специальные правовые требования для рабочей зоны оператора ПК.....	78
6.2 Организационные мероприятия при компоновке рабочей зоны.....	78
6.3 Профессиональная социальная безопасность.....	79

6.5 Электромагнитное излучение экрана монитора	82
6.6 Недостаточная освещенность рабочей зоны	83
6.7 Перенапряжение зрительного анализатора	84
6.8 Опасность поражения электрическим током	85
6.9 Экологическая безопасность	87
6.10 Безопасность в чрезвычайных ситуациях	88
6.11 Общий порядок действий при возникновении пожара	88
ЗАКЛЮЧЕНИЕ	91
СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ.....	92
ПРИЛОЖЕНИЕ А.....	94
ПРИЛОЖЕНИЕ Б.....	99

ВВЕДЕНИЕ

С 80-х годов прошлого века в различных исследованиях, как и в СССР, так и за рубежом стал независимо формироваться метод, позже получивший название сингулярного спектрального анализа (Singular Spectrum Analysis, SSA) или «Гусеница». Метод изначально применялся в климатологии, океанологии и геофизике. Но на сегодняшний день насчитывается несколько десятков прикладных статей и методик, рассматривающих метод. Метод применяется в астрофизике, электроэнцефалографии, анализе вулканической активности и экономических процессов.

Среди особенностей метода есть такие как возможность работы с нестационарными временными рядами (последовательности наблюдений, описывающих протекающий во времени процесс, измеренных в последовательные моменты времени, обычно через равные промежутки), проведение анализа ряда без гипотезы и предполагаемых моделей. Что является преимуществом по сравнению с классическими методами прикладной статистики.

Прогнозирование технологических процессов и обнаружение моментов разладки (изменения таких характеристик ряда как среднее значение, дисперсия, автоковариация на новые, присущие другому ряду) являются важными задачами в нефтегазовой отрасли. Так своевременное обнаружение разладки процесса может предотвратить аварийную ситуацию, например, газонефтеводопроявления, поломку оборудования и, вследствие чего, нарушение производственного процесса.

Целью ВКР было выбрано исследование базового алгоритма метода SSA и его применение для построения моделей ВР, обнаружение разладки и прогнозирование с помощью полученной модели.

В представленной работе была отлажена методика анализа нестационарных временных рядов, определена метрика для подбора оптимальной модели и наглядно продемонстрировано влияние входных

параметров метода на конечный результат. Дополнена существующая программная реализация алгоритма методами и функциями. Полученные результаты могут быть использованы для развития экспертной системы.

Среди перспектив работы можно обозначить исследование и реализацию метода заполнения пропусков в данных и прогнозирования, реализацию алгоритма определения коэффициентов для линейной рекуррентной формулы, применяемой для прогнозирования, реализацию построения фазовых диаграмм собственных векторов и периодограммы, что позволит получить более полную характеристику об исследуемом процессе.

1 ОБЗОР ЛИТЕРАТУРЫ

Для исследования потребовалось провести большой подбор информационных источников и анализ существующих решений.

Для более детального понимания работы алгоритма метода потребовалось освежить знания из программы по высшей математике [1, 2]. Разобраться в разделах прикладной статистики и ознакомиться с теорией динамических процессов [3, 4, 5, 6].

Основным трудом по методу SSA является монография [7], и методическое пособие [8]. Также были использованы источники [6, 9, 10].

По реализации в ПО метода были использованы источники [10, 11, 12, 13].

Прикладной аспект метода SSA отражен в статьях [14, 15, 16].

Предметная область и автоматизация технологических процессов в нефтегазовой отрасли отражены в монографии [17].

Исходя из проведенного анализа источников, можно сделать вывод, что метод SSA является достаточно известным, однако не столь распространенным и классическим, как метод авторегрессии и проинтегрированного скользящего среднего (АРПСС, ARIMA) [3]. Возможно из-за теоретической сложности [7, 8].

Алгоритм SSA реализован в таких ПО как Matlab, прикладной библиотеке RSSA для ЯП R, есть реализация для Excel, не было найдено реализации для Statistica, хотя там реализовано сингулярное разложение – основная часть алгоритма. В итоге была выбрана реализация на ЯП Python.

2 ОБЪЕКТЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

2.1 Объект, метод и постановка задач

Объектом исследования являются исторические данные о технологических процессах скважины томского нефтегазового месторождения. Методом исследования является базовый метод сингулярного спектрального анализа «Гусеница»-SSA.

Для выполнения поставленных целей будет проведен предварительный анализ данных с целью выбора фрагментов ВР с наглядной динамикой для исследования и демонстрации применимости SSA. Метод имеет достаточно сложный теоретический аппарат и поэтому наглядное представление с подробным описанием так же необходимо для понимания метода. Учитывая эту сложность, необходима методика для подбора входных параметров, таких как ширина окна вложения траекторной матрицы L и группировка компонент для построения модели.

Все вышеописанные задачи подводят к решению главных – получение моделей ВР для применения в обнаружении разладки ТП и прогнозировании будущих значений.

2.2 Границы и условия

В данной работе представлен базовый метод SSA, который не включает в себя прогнозирование, потому в качестве оценки того, что метод возможно применять в задаче прогнозирования, будет использование полученной модели с методом экспоненциального сглаживания – популярным методом прогнозирования ВР, обеспечивающим достаточную применимость и точность при своей простоте.

В проводимом анализе не учитываются единицы измерения значений ТП и даты наблюдений, т.к. эта информация в данном случае не несет информативности, поэтому для простоты заменена на метки наблюдений.

Актуальность информации, регистрируемой по ТП с целью обнаружения разладки ВР возможна только при ведении частой записи значений ключевых параметров. Иначе уже аварийное состояние, не говоря о предаварийном, может быть даже не отражено в истории из-за редкой регистрации наблюдений.

Исходя из имеющихся данных с интервалом измерения в 15 минут и малой информацией по производственным ТП, внесем некоторые допущения на этапе анализа ВР с целью обнаружения разладки (п.3.3.7). Допустим, что замеры производятся за более короткие интервалы времени. Задача состоит в том, чтобы по временному ряду определить: случится разладка или нет. Разумеется, что невозможно заранее предугадать появление разладки без наличия каких-либо предварительных признаков.

Представленная в работе методика выработана в связи с тем, что все рассматриваемые ВР показывали значение относительного вклада более 95% для первой компоненты, что связано вероятно с частотой записи данных, вследствие чего это выразилось в отражении только трендовой разделимости, а влияние периодических и шумовых компонент безусловно есть, но ввиду временного масштаба оно сведено до минимума.

Для проверки ряда на стационарность используются диаграммы распределения плотности и ящичная (диаграмма размаха, «ящик с усами») и АКФ. Возможно, что данных средств недостаточно для определения стационарности ряда.

В данной работе понятие «шум», «остатки» и «стационарный ряд» [3] имеют один контекст: большая часть элементарных компонент (собственных троек SVD), с незначительным вкладом в ВР относительно первых компонент и большим количеством ковариаций с другими компонентами ряда, минимальными коэффициентами автоковариации. Хотя это допущение в целом некорректно, но для решения задачи выбора оптимальных компонентов модели достаточное.

Исчисление порядков множеств, используемых в работе (в основном, нумерация элементарных компонент ВР) представлены с 0, т.е. первый элемент множества имеет нулевой порядок. Такой подход удобен с точки зрения математического описание – соответствует монографии [7] и с точки зрения программной реализации.

2.3 Выбор средств реализации исследования

Под средствами реализации исследования подразумевается выбор ПО для проведения исследования.

В качестве такого инструмента был выбран дистрибутив Anaconda для ЯП Python, имеющий вычислительную среду с веб-интерфейсом Jupyter и набор библиотек, необходимым для анализа данных.

Особенностями данного решения являются бесплатность и доступность дистрибутива, возможность выполнять задачи встроенными средствами, поддержка языка верстки LaTeX. Гибкость, быстрота реализации идей, доступность – все это, присущее Python, интерпретаторному ЯП с динамической типизацией, получило еще большее развитие в проекте IPython Jupyter: не нужно специально настраивать виртуальное окружение, для работы достаточно веб-браузера, программный код можно реализовывать блоками – что очень удобно в задачах исследования и прототипирования. Имеется возможность в один клик создать копию проекта, откатить изменения до сохраненной контрольной точки, при этом автосохранение проекта позволяет не потерять написанный код. В основу программной реализации был положен класс SSA [10]. Данный класс был самостоятельно дополнен функциями и методами для методики анализа в данном ВКР. Используются прикладные библиотеки numpy, matplotlib и pandas.

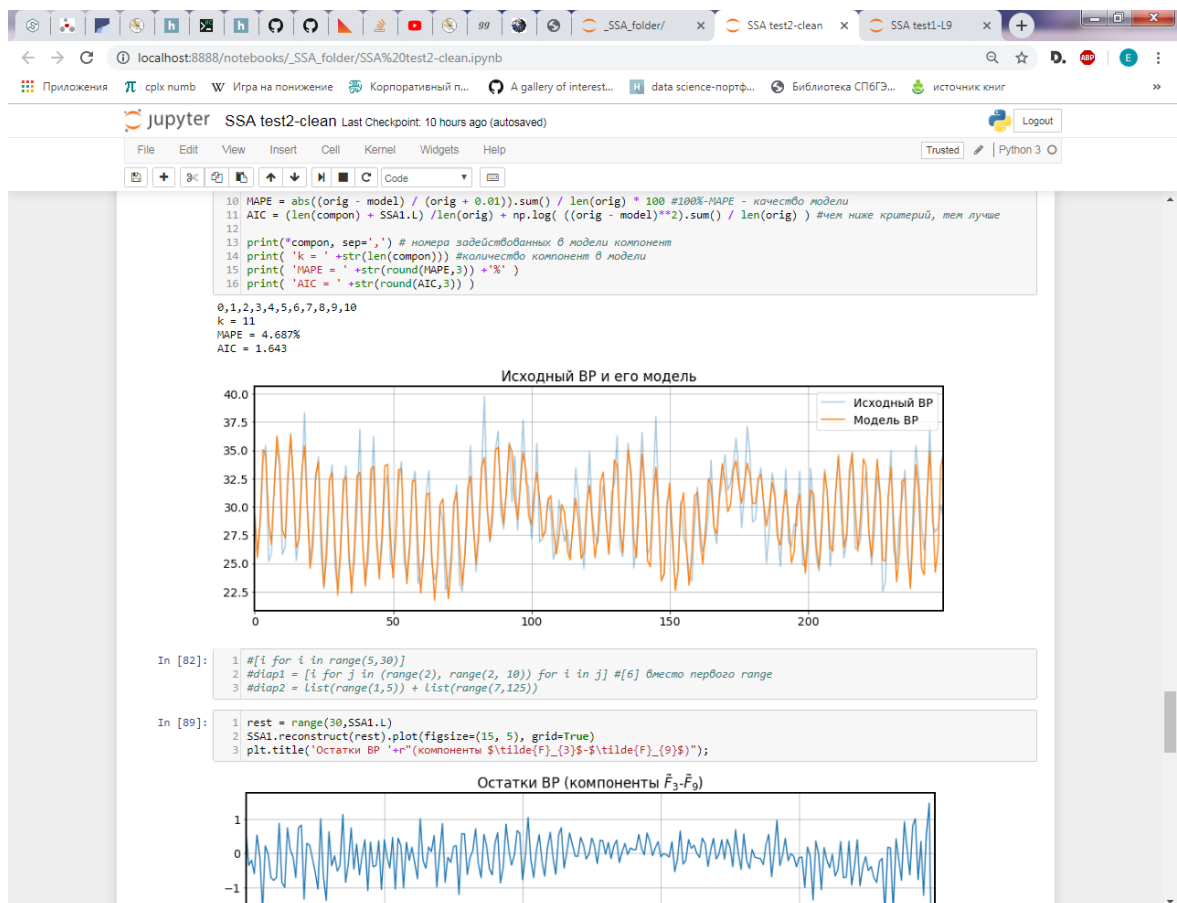


Рис.1 Интерфейс среды Jupyter Notebook

Вторым инструментом для анализа является Microsoft Office Excel, который в современной версии позволяет отображать графики, проводить прогнозирование очень быстро и просто. Благодаря табличному интерфейсу удобно проводить обработку данных. Этот инструмент будет использован для подготовки данных и прогнозирования.

3 РАСЧЕТЫ И АНАЛИТИКА

3.1.1 Представление метода SSA

Первый шаг SSA – преобразование ВР F в последовательность многомерных векторов. Пусть натуральное число L будет шириной окна, $2 \leq L \leq N/2$. Сформируем окно из отрезка $\{f_i, f_{i+1}, \dots, f_{i+L-1}\}$ для $i = 0, \dots, N - L$. Переместим это окно по ВР, формируя вектор-столбец X_i для каждого отрезка. Получим:

$$X_0 = (f_0, f_1, f_2, \dots, f_{L-1})^T$$

$$X_1 = (f_1, f_2, f_3, \dots, f_L)^T$$

$$X_2 = (f_2, f_3, f_4, \dots, f_{L+1})^T$$

$$\vdots$$

$$X_{N-L} = (f_{N-L}, f_{N-L+1}, f_{N-L+2}, \dots, f_{N-1})^T$$

Эти векторы-столбцы формируют L -траекторную матрицу X ВР F :

$$X = \begin{bmatrix} f_0 & f_1 & f_2 & \cdots & f_{N-L} \\ f_1 & f_2 & f_3 & \cdots & f_{N-L+1} \\ f_2 & f_3 & f_4 & \cdots & f_{N-L+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_{L-1} & f_L & f_{L+1} & \cdots & f_{N-1} \end{bmatrix}$$

Из вышеприведенной матрицы видно, что $x_{ij} = f_{i+j-2}$ и значения элементов f в побочных диагоналей матрицы равны $i + j = \text{const}$. Этот вид матрицы известен как матрица Ганкеля.

Для базового ВР и наглядного примера будет установлена ширина окна $L = 70$. Пусть $K = N - L + 1$ – количество столбцов в траекторной матрице. Назовем столбцы траекторной матрицы (в дальнейшем ТМ) X как L -вложенные векторы, а ее строки – K -вложенные векторы (рис.2).

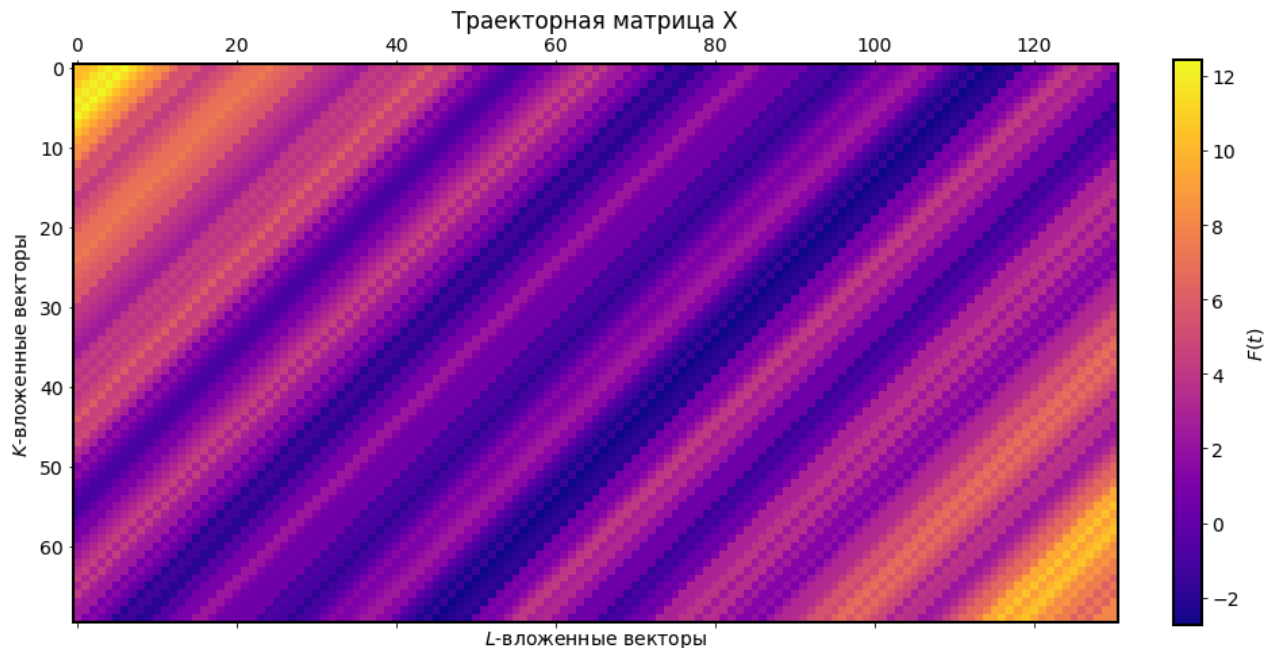


Рис.2 Траекторная матрица базового ВР

3.1.2 Разложение траекторной матрицы

Второй шаг – разложение ТМ с помощью SVD:

$$X = U\Sigma V^T$$

где:

$U - L \times L$ унитарная матрица, столбцы которой есть ортонормированный базис левых сингулярных векторов ТМ X .

$\Sigma - L \times K$ прямоугольная диагональная матрица, содержащая L сингулярных чисел от X , упорядоченных от большего к меньшему значениям;

$V - K \times K$ унитарная матрица, столбцы которой есть ортонормированный базис правых сингулярных векторов ТМ X .

Тогда SVD траекторной матрицы может быть записано так:

$$X = \sum_{i=0}^{d-1} \sigma_i U_i V_i^T \equiv \sum_{i=0}^{d-1} X_i$$

где:

$\sigma_i - i$ -е сингулярное число;

U_i и V_i – векторы, выражающие i -е столбцы U и V соответственно;

$d \leq L$ – ранг ТМ X ;

$X_i = \sigma_i U_i V_i^T$ – i -я элементарная матрица от ТМ X . Набор $\{\sigma_i, U_i, V_i^T\}$ есть i -я собственная тройка SVD.

Чтобы составить представление о том, что все это означает, давайте рассмотрим матрицы U , V и Σ .

$U - L \times L$ матрица, столбцы которой ортонормированы, то есть:

$$U_i \cdot U_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Это означает $UU^T = U^T U = 1$ делая матрицу U унитарной.

Чтобы выяснить роль, которую U играет в разложении для ТМ X выше, пусть $Z_i = \sigma_i V_i$ будет вектором-столбцом, тогда:

$$X = \sum_{i=0}^{d-1} U_i Z_i^T$$

И каждый L -вложенный вектор-столбец X_j выражен:

$$X_j = \sum_{i=0}^{d-1} z_{j,i} U_i$$

где $z_{j,i}$ – j -й элемент вектора Z_i .

Выражение для X_j предполагает, что $U = \{U_0, \dots, U_{d-1}\}$ является базисом, охватывающим пространство столбцов ТМ, и $z_{j,i}$ – является j -м коэффициентом L -вложенного вектора X_j , представленного в базисе U . Другими словами, столбцы матрицы U образуют ортонормированный базис, который описывает отрезки времени $\{f_i, \dots, f_{i+L-1}\}_{i=0}^{N-L}$ в столбцах ТМ. Линейная комбинация базисных векторов, базис U ниже по ссылке получаем что допустим $j=1$, то вектор X_1 будет выражен базисом (линейная комбинация базисных векторов): $X_1 = z_{10}U_0 + z_{11}U_1 + \dots + z_{1d-1}U_{d-1}$.

Матрица V , представляется в выражении SVD по условию транспонированной, и представляет собой матрицу $K \times K$ с ортонормированными столбцами, которая, как и матрица U , является унитарной.

Чтобы интерпретировать столбцы матрицы V в SVD-разложении ТМ, сначала отметим, что для любых матриц подходящей формы A и B будет $(AB)^T = B^T A^T$. Транспонируя матрицу X , получим:

$$X^T = V \Sigma^T U^T = \sum_{i=0}^{d-1} V_i Y_i^T$$

где $Y_i = \sigma_i U_i$.

Тогда:

$$X_j^{(T)} = \sum_{i=0}^{d-1} y_{j,i} V_i$$

где: $X_j^{(T)}$ – j -й столбец X^T и $y_{j,i}$ – j -й компонент вектора Y_i . Это выражение полагает, что $V = \{V_0, \dots, V_{d-1}\}$ является базисом, охватывающим пространство столбцов матрицы X^T , а $y_{j,i}$ – i -й коэффициент вложенного вектора $X_j^{(T)}$, представленного в базисе V . Эквивалентно, V – базис, охватывающий пространство строк матрицы X . То есть, столбцы матрицы V образуют ортонормированный базисный набор, описывающий отрезки времени $\{f_i, \dots, f_{i+N-L}\}_{i=0}^{L-1}$ в строках ТМ.

Матрица Σ представляет собой прямоугольную диагональную матрицу $L \times K$, содержащую сингулярные числа ТМ X . Сингулярные числа упорядочены от большего к меньшему, т.е. $\sigma_0 \geq \sigma_1 \geq \dots \geq \sigma_{L-1} \geq 0$. Можно интерпретировать σ_i как скалярный множитель, который определяет относительный вклад собственной тройки (σ_i, U_i, V_i) в выражении:

$$X = \sum_{i=0}^{d-1} \sigma_i U_i V_i^T$$

Определим норму Фробениуса [3,5] $\|X_F\|$ (нормы используются для определения расстояния между векторами; норма вектора/матрицы – обобщает длину вектора и абсолютное значение числа) для ТМ X :

$$\|X\|_F = \sqrt{\sum_{j=0}^{L-1} \sum_{k=0}^{K-1} |x_{j,k}|^2}$$

где $x_{j,k}$ обозначает элемент в j -й строке и k -й столбце матрицы X .

Давайте обратим внимание на элементарные матрицы $X_i = \sigma_i U_i V_i^T$. Теперь для внешнего произведения $U_i V_i^T$, имеем $\|U_i V_i^T\|_F = \|U_i\|_F \|V_i\|_F$, равное 1 благодаря нормализации U_i и V_i . Из чего ясно, что $\|X_i\| = \sigma_i$.

$$\|X\|_F^2 = \sum_{i=0}^{d-1} \sigma_i^2$$

Например, квадрат нормы Фробениуса ТМ. равен сумме квадратов сингулярных чисел. Здесь предполагается, что мы можем взять частное $\sigma_i^2 / \|X\|_F^2$ как меру вклада элементарной матрицы X_i при разложении ТМ.

Далее матрицу X в SVD-виде умножим справа на X^T :

$$\begin{aligned} XX^T &= U \Sigma V^T X^T \\ &= U \Sigma V^T V \Sigma^T U^T \\ &= U \Sigma \Sigma^T U^T \end{aligned}$$

Допуская, что квадрат диагональной матрицы $\Sigma^2 = \Sigma \Sigma^T$, и умножая справа на матрицу U , получим:

$$(XX^T)U = U\Sigma^2$$

где: Σ^2 – диагональная матрица из элементов σ_i^2 , показывает, что столбцы U – это собственные векторы матрицы XX^T с собственными значениями $\{\sigma_0^2, \dots, \sigma_{L-1}^2\}$. Следуя той же логике, умножая матрицу X слева на X^T увидим, что столбцы матрицы V – собственные векторы матрицы $X^T X$ так же с собственными значениями $\{\sigma_0^2, \dots, \sigma_{L-1}^2\}$.

Теперь следует упомянуть о ранге d ТМ. Столбцы ТМ – это последовательность L -вложенных векторов, которые охватывают траекторное пространство ВР. Это пространство не более, чем L -мерное, однако если столбцы матрицы X линейно зависимы, то траекторное пространство будет иметь $d < L$ измерений. Это проявляется наличием

нулевых сингулярных чисел в матрице Σ . Ранг матрицы X – это максимальное значение i при $\sigma_i > 0$. Другими словами, $d = \text{rank}\{X\}$ можно рассматривать как внутреннюю размерность пространства ВР.

Примечание: для зашумленных данных временных рядов реального мира пространство траекторий, скорее всего, будет иметь $d = L$ измерений.

Важно отметить, что каждая элементарная матрица X_i имеет ранг 1, и что матрица:

$$X^{(r)} = \sum_{i=0}^r X_i$$

для $r < d$ – наилучшее приближение ранга r к ТМ X , такое что $\|X - X^{(r)}\|_F$ минимизируется. То есть, можно суммировать первые r элементарных матриц, чтобы получить оптимальную аппроксимацию меньшей размерности для матрицы X .

3.1.3 Обобщение SVD-разложения

Проведем обобщение описанных выше элементов сингулярного разложения: была сопоставлена ВР $F = \{f_0, \dots, f_{N-1}\}$ с набором многомерных вложенных векторов $X_i = (f_i, f_{i+1}, \dots, f_{i+L-1})^T$, $i = 0, \dots, N - L$, которые составляют столбцы ТМ X . Затем выполнено SVD-разложение матрицы, сделав это мы нашли два ортонормированных базиса U и V , которые охватывают пространство столбцов и строк ТМ соответственно.

SVD-разложение матрицы X может быть записано так:

$$X = \sum_{i=0}^{d-1} \sigma_i U_i V_i^T \equiv \sum_{i=0}^{d-1} X_i$$

где X_i – i -я элементарная матрица от X , определяемая по собственной тройке $\{\sigma_i, U_i, V_i\}$. i -е сингулярное число σ_i определяет относительный вклад X_i в разложение матрицы X выше. Целое число $d \leq L$ – это внутренняя

размерность пространства траекторий ВР, и мы можем выбирать, какой меньшей размерности получить аппроксимацию от X посредством суммы только первых $r < d$ элементарных матриц.

Ниже представлены первые 12 элементарных матриц ВР:

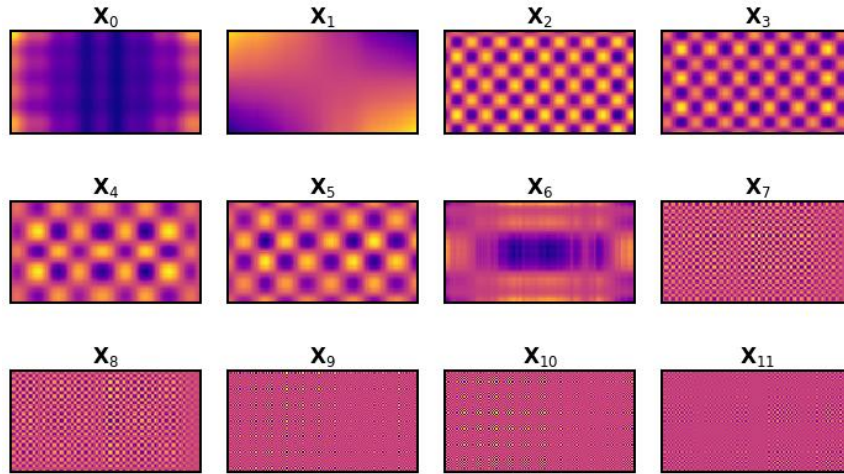


Рис.3 Элементарные матрицы (1-12) базового ВР.

По визуальному осмотру X_i выше очевидно, что элементарные матрицы не имеют антидиагональной структуры ТМ. Даже без проверки вектора U_i , связанного с каждой X_i , или восстановления составляющих ВР, появление X_i намекает на природу каждой составляющей, будь то тренд, периодичность или шум.

Например, L - и K -вложенные векторы в X_0 и X_1 изменяются относительно медленно по всей матрице, предполагая, что X_0 и X_1 могут быть связаны с общим трендом во ВР. Матрицы от X_2 до X_5 имеют вид шахматной доски, предполагая периодичность. X_6 может быть отнесена либо к периодичности, либо к тренду. Матрицы X_7 и далее (вплоть до X_{69}), быстро чередуются между несколькими значениями – эти элементарные матрицы могут быть связаны с шумом в исходном ВР.

Отобразим относительные вклады $\frac{\sigma_i^2}{\sum_{k=0}^{d-1} \sigma_k^2}$ и суммарные вклады $\frac{\sum_{j=0}^l \sigma_i^2}{\sum_{k=0}^{d-1} \sigma_k^2}$ этих 12 элементарных матриц в ТМ базового ВР:

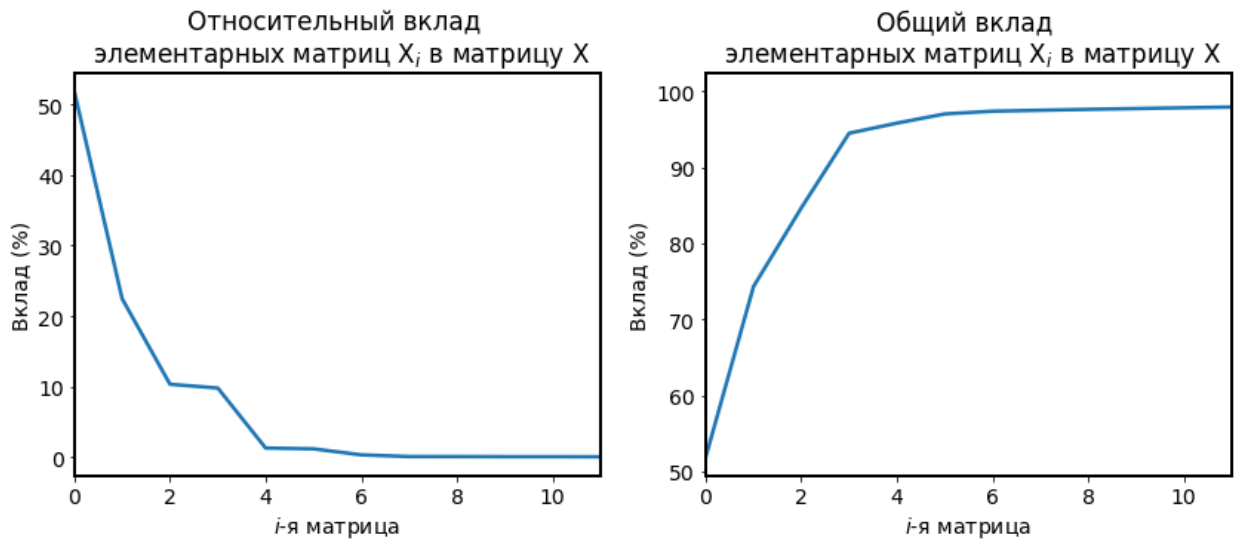


Рис.4 Вклад относительный и суммарный элементарных матриц в базовый ВР.

Графики выше показывают относительный и общий вклад первых 12-ти X_i в разложении $X = \sum_{k=0}^{d-1} X_i$. Элементарные матрицы X_0 и X_1 вносят в разложение матрицы вклад X 52% и 22% соответственно. Первые семь элементарных матриц вместе дают 97%. Элементарные матрицы, которые вносят равные вклады в разложение (т.е. $\sigma_i \approx \sigma_{i+1}$), вероятно, следует сгруппировать вместе при восстановлении ВР. Они отображаются как перегибы на графике относительного вклада. Например, перегибы на графике выше предполагают, что X_2 и X_3 , а также X_4 и X_5 должны быть сгруппированы вместе.

Важно отметить, что элементарные матрицы представляют оптимальное (хотя, возможно, не единственное) разделение компонентов в пространстве ТМ: по определению, строки и столбцы одной элементарной матрицы ортогональны строкам и столбцам других элементарных матриц.

3.1.4 Восстановление ВР

Пока что мы сопоставили ВР F к серии L -вложенных векторов, образуя ТМ X . Затем мы разложили эту матрицу с помощью SVD и построили набор элементарных матриц, которые в сумме и являются матрицей X . Затем мы дали небольшое пояснение, чтобы классифицировать эти элементарные матрицы как тренд, периодичность и шум.

В идеальном мире все компоненты ВР $F = \sum_j F^{(j)}$ были бы разделимы, и можно было бы соответствующим образом сгруппировать результирующие элементарные матрицы X_i так, чтобы:

$$X = \sum_{k \in \mathcal{S}} X_k + \sum_{l \in \mathcal{T}} X_l + \dots = \sum_j X^{(j)}$$

где $\mathcal{S}$ и $\mathcal{T}$ – непересекающиеся наборы индексов, а $X^{(j)}$ – ТМ ВР компоненты $F^{(j)}$. В этом случае каждая $X^{(j)}$ будет иметь ганкелеву структуру, подобную исходной ТМ, тогда построение каждого $F^{(j)}$ будет простым. Однако в реальном мире ни одна из составляющих ТМ не будет иметь равных значений на своих антидиагоналях (это определяющее свойство ганкелевой матрицы, в дальнейшем – ГМ). Поэтому мы ищем способ преобразования элементарной матрицы в ГМ, а затем во ВР.

Чтобы извлечь ВР из элементарных матриц, мы будем использовать диагональное усреднение, которое определяет значения восстановленного ВР $\tilde{F}^{(j)}$ как средние значения соответствующих антидиагоналей матриц $X^{(j)}$. Формально это представляется введением оператора ганкелизации ($\hat{\mathcal{H}}$), который воздействует на $L \times K$ матрицу $X^{(j)}$, чтобы получить ГМ $\tilde{X}^{(j)}$; то есть:

$$\tilde{X}^{(j)} = \hat{\mathcal{H}} X^{(j)}$$

Элемент $\tilde{x}_{m,n}$ в $\tilde{X}^{(j)}$ для $s = m + n$, получен по формуле:

$$\tilde{x}_{m,n} = \begin{cases} \frac{1}{s+1} \sum_{l=0}^s x_{l,s-l} & 0 \leq s \leq L-1 \\ \frac{1}{L-1} \sum_{l=0}^{L-1} x_{l,s-l} & L \leq s \leq K-1 \\ \frac{1}{K+L-s-1} \sum_{l=s-K+1}^L x_{l,s-l} & K \leq s \leq K+L-2 \end{cases}$$

На первый взгляд, вышенаписанное выглядит как непробиваемая завеса из матричных индексов. Однако все, что данный оператор делает – это вычисляет заданное значение $\tilde{x}_{m,n}$ усреднением элементов антидиагонали. Их количество зависит от положения антидиагонали по m и n . На практике не важна полная ГМ $\tilde{X}^{(j)}$, и можно перейти прямо к построению ВР $\tilde{F}^{(j)}$. Тем не

менее, определение диагонального усреднения $\hat{\mathcal{H}}X^{(j)}$ выше было включено, чтобы завершить математическое разложение SSA.

Важно отметить, что $\hat{\mathcal{H}}$ – линейный оператор, к примеру, $\hat{\mathcal{H}}(A + B) = \hat{\mathcal{H}}A + \hat{\mathcal{H}}B$. Тогда для ТМ X:

$$\hat{\mathcal{H}}X = \hat{\mathcal{H}}\left(\sum_{i=0}^{d-1} X_i\right) = \sum_{i=0}^{d-1} \hat{\mathcal{H}}X_i \equiv \sum_{i=0}^{d-1} \tilde{X}_i$$

Поскольку ТМ X изначально является ГМ, то по определению $\hat{\mathcal{H}}X = X$. Следовательно, ТМ может быть выражена через составляющие ее ганкелизированные элементарные матрицы:

$$X = \sum_{i=0}^{d-1} \tilde{X}_i$$

Поскольку ВР определяется однозначно по ГМ, приведенное выше выражение также определяет ВР F как сумму его компонентов $\tilde{F}_i$. Мы должны сгруппировать эти компоненты вместе и классифицировать их как тренд, периодичность или шум, и тогда мы можем сами решать, как их использовать.

Отобразим (первые) 12 элементарных матриц ганкелизованными:

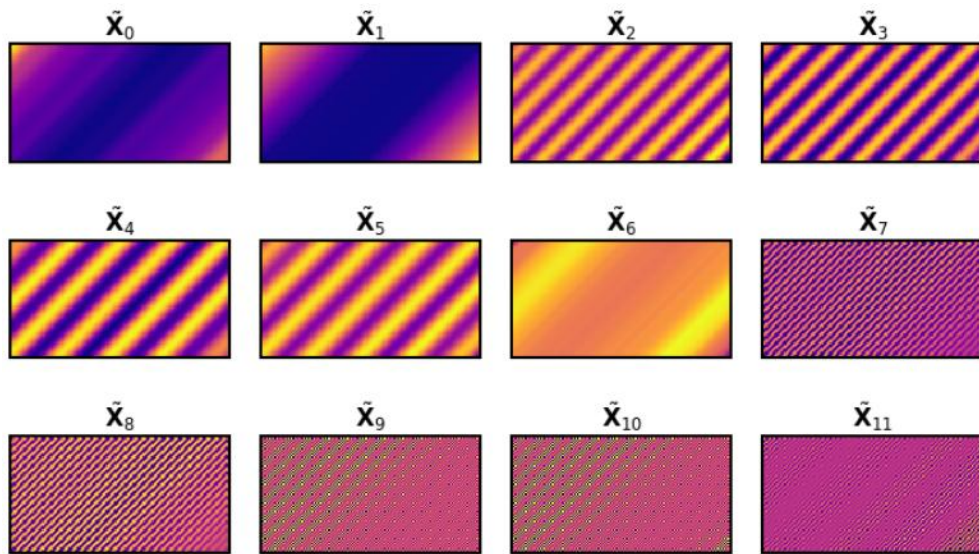


Рис.5 Первые 12 элементарных матриц базового примера после ганкелизации

Осмотр ганкелизованных элементарных матриц базового примера подтверждает наше подозрение об элементарных матрицах: $\tilde{X}_0$ и $\tilde{X}_1$ изменяются медленно на протяжении всего ВР и могут быть сгруппированы вместе как компонента тренда. $\tilde{X}_2$ и $\tilde{X}_3$ обе являются периодическими, с одинаковой частотой и могут быть сгруппированы как первая периодическая компонента ВР F .

$\tilde{X}_4$ и $\tilde{X}_5$ также являются периодическими, с частотой, но отличной от $\tilde{X}_2$ и $\tilde{X}_3$, и будут сгруппированы как вторая периодическая компонента. $\tilde{X}_6$, лишенная очевидной периодичности, но медленно меняющаяся, будет сгруппирована с составляющими тренда. Мы соберем вместе все составляющие от $\tilde{X}_7$ и выше – как шум. Подведем итог:

$$\begin{aligned}\tilde{X}^{(\text{тренд})} &= \tilde{X}_0 + \tilde{X}_1 + \tilde{X}_6 &\Rightarrow \tilde{F}^{(\text{тренд})} &= \tilde{F}_0 + \tilde{F}_1 + \tilde{F}_6 \\ \tilde{X}^{(\text{период1})} &= \tilde{X}_2 + \tilde{X}_3 &\Rightarrow \tilde{F}^{(\text{период1})} &= \tilde{F}_2 + \tilde{F}_3 \\ \tilde{X}^{(\text{период2})} &= \tilde{X}_4 + \tilde{X}_5 &\Rightarrow \tilde{F}^{(\text{период2})} &= \tilde{F}_4 + \tilde{F}_5 \\ \tilde{X}^{(\text{шум})} &= \tilde{X}_7 + \dots + \tilde{X}_{69} &\Rightarrow \tilde{F}^{(\text{шум})} &= \tilde{F}_7 + \dots + \tilde{F}_{69}\end{aligned}$$

Отобразим первые 12 элементарных компонент $\tilde{F}_i$ для базового примера:

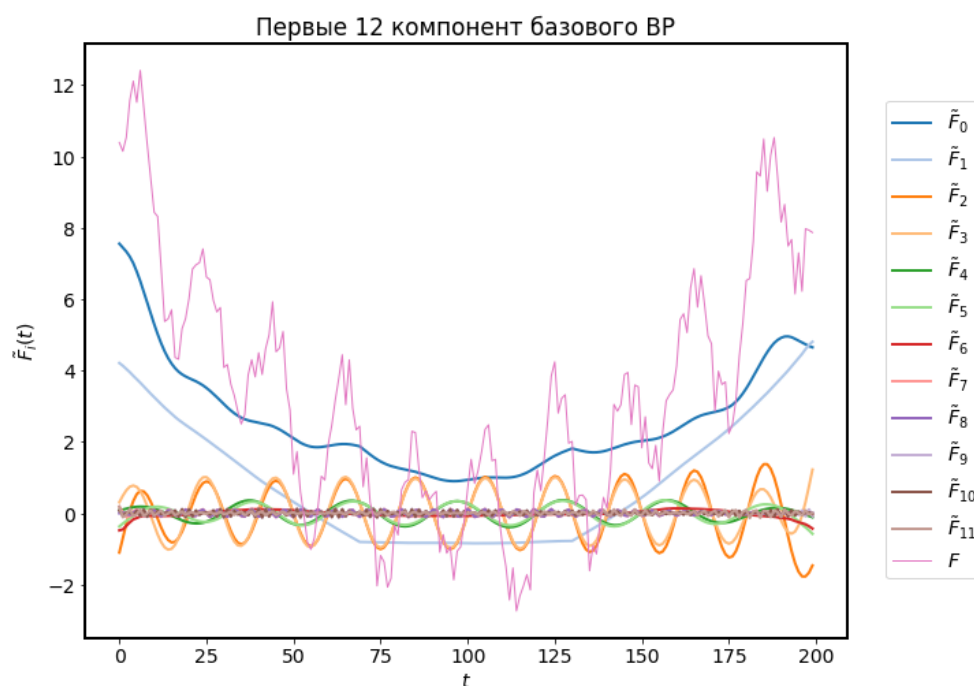


Рис. 6 Первые 12 компонент базового ВР (как уже ясно из названия выше ☺)

Как упоминалось ранее, элементарные составляющие, выделяемые из траекторного пространства ВР, могут не совпадать с конкретной интерпретируемой компонентой самого ВР. Например, $\tilde{F}_0$ и $\tilde{F}_1$ оба выглядят неявно как тренд — действительно ли они являются отдельными компонентами? Точно так же $\tilde{F}_2$ и $\tilde{F}_3$ почти идентичны, за исключением возле границ ВР.

Отобразим группировку для $\tilde{F}^{(\text{тренд})}$, $\tilde{F}^{(\text{период1})}$, $\tilde{F}^{(\text{период2})}$ и $\tilde{F}^{(\text{шум})}$, и посмотрим, как компоненты, выделенные SSA-методом, сравниваются с исходными компонентами, составляющими базовый ВР:

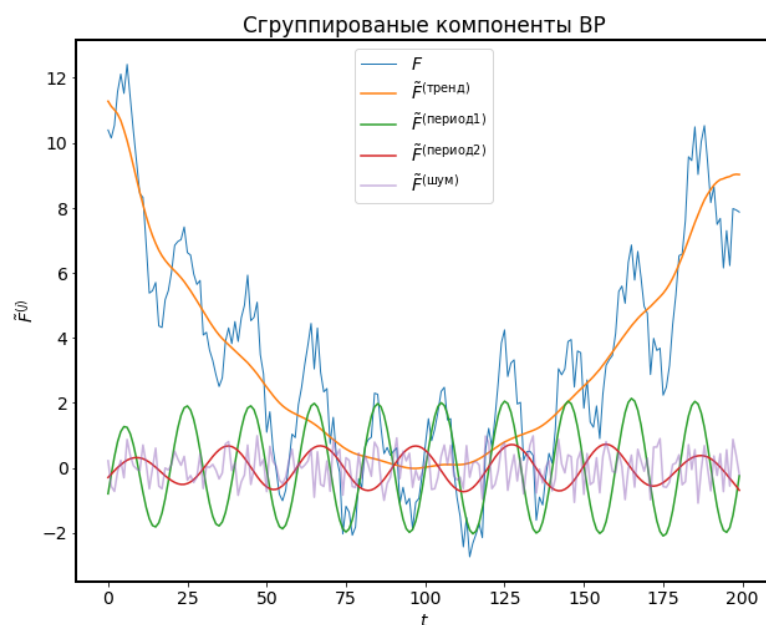


Рис. 7 Сгруппированные компоненты исходного ВР

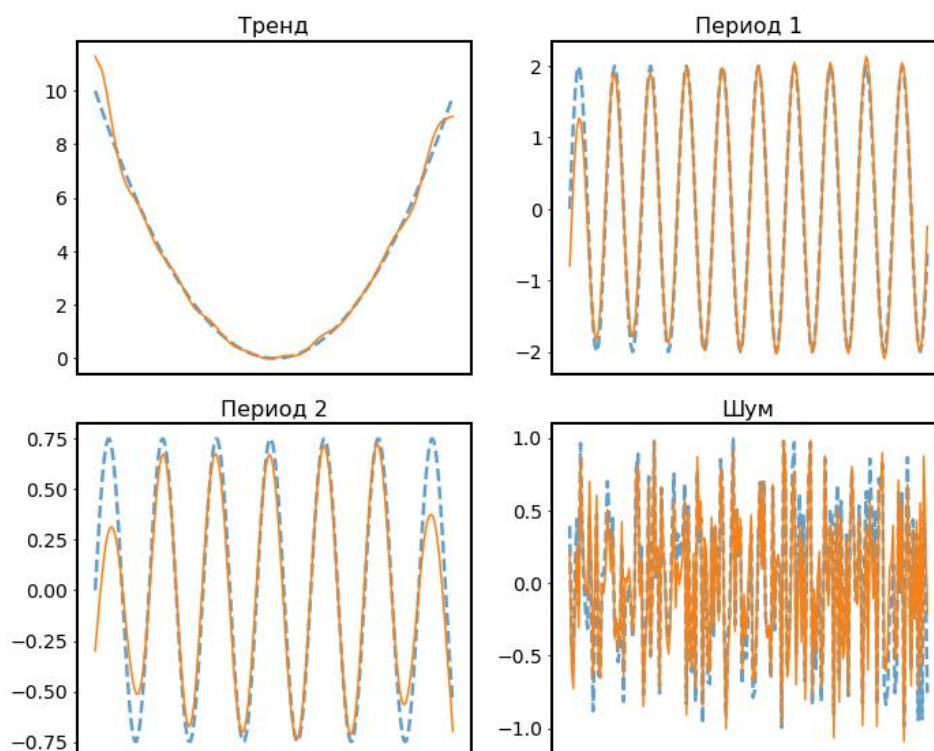


Рис. 8 Восстановленные компоненты и их идеальные версии

После пройденных этапов видно, что метод SSA демонстрирует хорошие результаты, однако следует обратить внимание на ухудшение аппроксимации вблизи границ ряда.

Это распространенный признак для SSA вследствие того, что в соответствии с формализмом SSA большинство типов рядов (т.е.

полиномиальных, синусоидальных, экспоненциальных и т.д.) не являются в точности разделимыми. Поэтому наша попытка восстановить точный параболический тренд и периодические компоненты из базового примера обречена на провал с самого начала.

Однако, это не означает, что компоненты ВР не могут быть аппроксимативно разделены, как было видно выше (существует также концепция асимптотической разделимости, когда длина ВР приближается к бесконечности, однако такой случай выходит за пределы данной работы).

3.1.5 Выделение компонентов ВР и их группировка

Ранее были сгруппированы собственные тройки/составляющие базового ВР по визуальному осмотру; т.е. было определено, какие компоненты принадлежат друг другу по их внешнему виду. Это хорошо для коротких и простых ВР, однако для более длинных и более сложных ВР предложим метод, который количественно определяет, можно ли рассматривать восстановленную компоненту $\tilde{F}_i$ отдельно от другой компоненты $\tilde{F}_j$, поэтому не нужно группировать решения через визуальный осмотр каждого $\tilde{F}_i$.

Для двух восстановленных ВР $\tilde{F}_i$ и $\tilde{F}_j$ длины N и ширины окна L мы определяем взвешенное внутреннее (скалярное) произведение $(\tilde{F}_i, \tilde{F}_j)_w$ как:

$$(\tilde{F}_i, \tilde{F}_j)_w = \sum_{k=0}^{N-1} w_k \tilde{f}_{i,k} \tilde{f}_{j,k}$$

где $\tilde{f}_{i,k}$ и $\tilde{f}_{j,k}$ – k -е значения $\tilde{F}_i$ и $\tilde{F}_j$ соответственно, а вес w_k определяется как:

$$w_k = \begin{cases} k+1 & 0 \leq k \leq L-1 \\ L & L \leq k \leq K-1 \\ N-k & K \leq k \leq N-1 \end{cases}$$

помнив, что $K = N - L + 1$. Вес w_k просто отражает число раз, когда $\tilde{f}_{i,k}$ и $\tilde{f}_{j,k}$ появляются в ГМ $\tilde{X}_i$ и $\tilde{X}_j$, из которых получены ВР $\tilde{F}_i$ и $\tilde{F}_j$.

Проще говоря, если $(\tilde{F}_i, \tilde{F}_j)_w = 0$, то $\tilde{F}_i$ и $\tilde{F}_j$ – ортогональны, а компоненты ВР разделимы. Конечно, полная w-ортогональность не возникает в реальной жизни, поэтому вместо этого мы определяем $d \times d$ корреляционную матрицу весов W_{corr} , которая измеряет отклонение компонентов $\tilde{F}_i$ и $\tilde{F}_j$ от w-ортогональности. Коэффициенты W_{corr} задаются:

$$W_{i,j} = \frac{(\tilde{F}_i, \tilde{F}_j)_w}{\|\tilde{F}_i\|_w \|\tilde{F}_j\|_w}$$

где $\|\tilde{F}_k\|_w = \sqrt{(\tilde{F}_k, \tilde{F}_k)_w}$ для $k = i, j$. Интерпретация $W_{i,j}$ проста: если $\tilde{F}_i$ и $\tilde{F}_j$ сколь угодно близки друг к другу (но не идентичны), то $(\tilde{F}_i, \tilde{F}_j)_w \rightarrow \|\tilde{F}_i\|_w \|\tilde{F}_j\|_w$ и, следовательно, $W_{i,j} \rightarrow 1$. Конечно, если $\tilde{F}_i$ и $\tilde{F}_j$ ортогональны, то $W_{i,j} = 0$. Величина $W_{i,j}$ лежит в пределах от 0 до 1, допустим $W_{i,j} \geq 0,3$ указывает на компоненты, которые, возможно, стоит вместе сгруппировать.

Проведем расчет и отображение корреляционной матрицы.

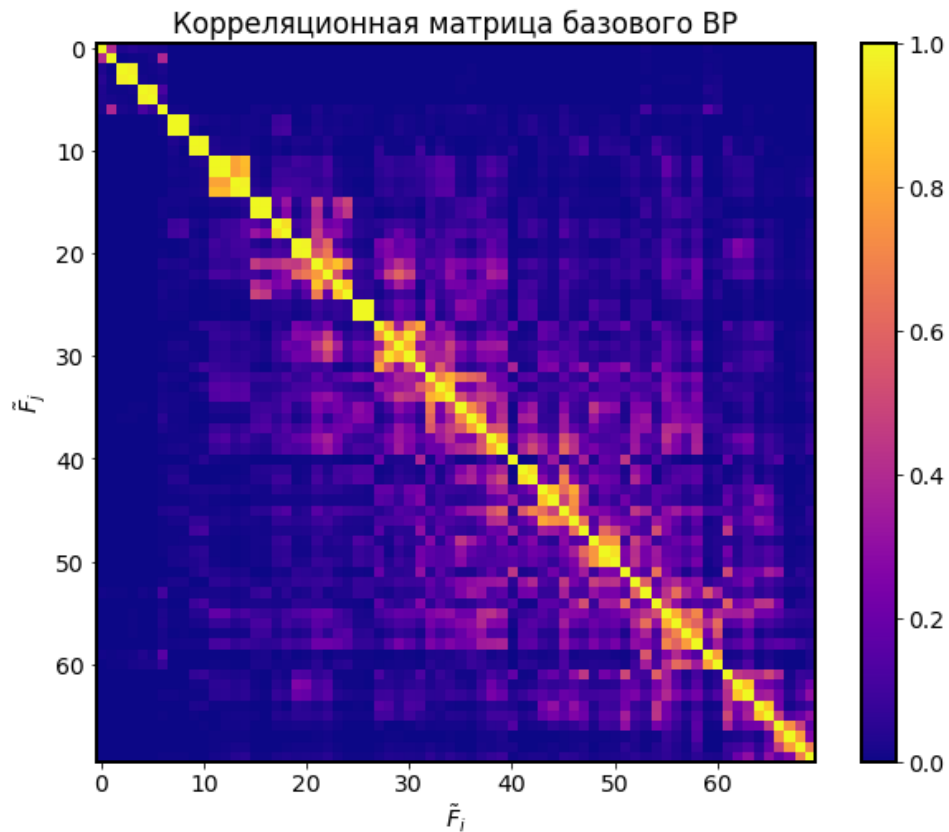


Рис. 9 Взвешенная корреляционная матрица базового примера

Структура W_{corr} показывает множество корреляций между составляющими ВР, в частности в диапазоне $7 \leq i, j \leq 69$. Поскольку это были компоненты, которые мы классифицировали как принадлежащие шуму во ВР, неудивительно, что между всеми ними есть ничтожные корреляции; это естественный результат того, что шум не имеет структурной компоненты, которая могла бы быть в дальнейшем отделена.

Важно отметить, что в матрице W_{corr} выделяется две области: $0 \leq i, j \leq 6$ и $7 \leq i, j \leq 69$. Они соответствуют двум основным группам: неслучайная группа ВР (т.е. тренд плюс две периодических компоненты) и шум. Посмотрим под увеличением на первые семь компонентов в W_{corr} :

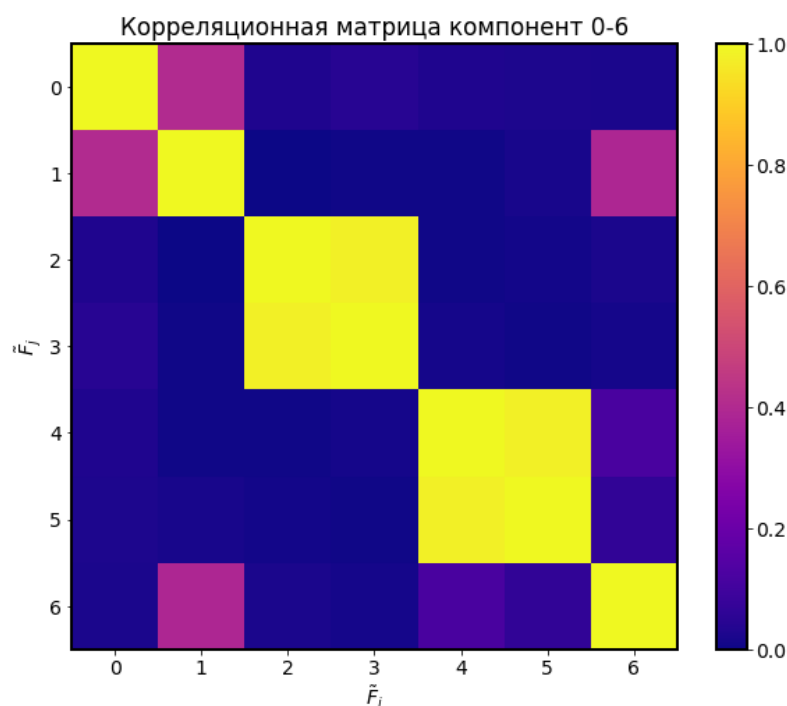


Рис. 10 Взвешенная корреляционная матрица первых 7 компонент базового ВР

Группировки, основанные на начальном появлении и сделанные для первых шести компонентов, поддерживаются соответствующими значениями w-корреляции. $\tilde{F}_0$ и $\tilde{F}_1$ имеют $W_{0,1} = 0,40$, предполагая, что они должны идти парой. $\tilde{F}_1$ и $\tilde{F}_6$ также имеют $W_{1,6} = 0,39$, предполагая, что $\tilde{F}_6$ также следует сгруппировать с $\tilde{F}_0$ и $\tilde{F}_1$ в качестве компоненты тренда. Тем не менее, $\tilde{F}_6$ также имеет небольшую w-корреляцию с $\tilde{F}_4$ и $\tilde{F}_5$, но поскольку $\tilde{F}_4$ и $\tilde{F}_5$ не имеют w-корреляции с $\tilde{F}_0$ и $\tilde{F}_1$, остановимся на связи $\tilde{F}_6$ с $\tilde{F}_0$ и $\tilde{F}_1$.

Сгруппированные $\tilde{F}^{(\text{период1})} = \tilde{F}_2 + \tilde{F}_3$ и $\tilde{F}^{(\text{период2})} = \tilde{F}_4 + \tilde{F}_5$ четко обоснованы матрицей w-корреляции, где $W_{2,3} = 0,98$ и $W_{4,5} = 0,98$.

3.2 Предварительный анализ и подготовка данных

Имеющийся набор данных содержит от 14 до более 22 тысяч наблюдений о 20 технических параметрах. Среди них 7 параметров представлены нулевыми, либо около нулевыми значениями с редкими изменениями: «Вибрация насоса X, -Y, -Z», «Дисбаланс напряжений», «Сопротивление изоляции», «Температура двигателя», «Дисбаланс тока». 3 параметра имеют постоянное значение, либо небольшие интервалы

постоянных значений: «Выходная частота», «Частота турбинного вращения». 1 параметр («Давление на приеме») имеет дискретную природу, и потому его целесообразнее рассматривать методами для анализа марковских цепей (тем не менее тренд выделить возможно – см. Приложение Б). Данные параметры не будут анализироваться, поскольку не содержат достаточной информации, либо требуется иной подход.

Параметры «Активная мощность», «Коэффициент мощности», «Ток двигателя фаза В, -А, -С», «Загрузка ПЭД» (итого: 6 параметров) имеют динамику ВР идентичную и скорее всего принадлежат к одному источнику – установке погружного электродвигателя (СУ ПЭД), эти параметры представлены графиком тока фазы В (рис.11).

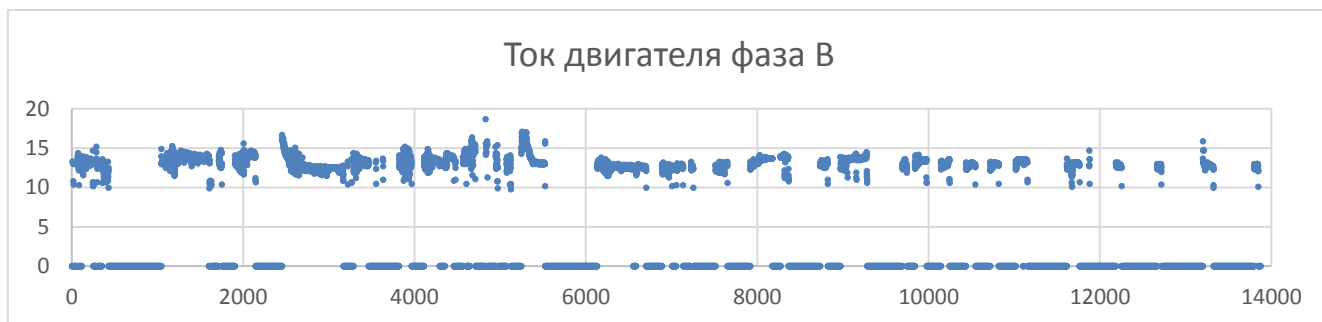


Рис.11 Временной ряд процесса «Ток двигателя фаза В»

Оставшиеся параметры («Дисбаланс напряжений», «Давление буферное», «Давление линейное», «Давление затрубное») имеют индивидуальную природу (Приложение Б).

Процедура проверки рядов на пропуски заключалась в следующем:

- преобразовать даты наблюдений в значения формата «00»;
- просуммировать 4 подряд идущих значения;
- проверить суммы по условию, невыполнение – есть пропуски.

Даты преобразовываются с целью оставить только значения минут: «00», «15», «30» и «45». Реализовано формулой: =ПРАВСИМВ(ЛЕВСИМВ(B1;16);2). Так, например, дата формата «2011-11-24 02:15:00.000» преобразуется в «15».

Затем берется сумма из преобразованных значений, и окном в 4 значения подряд формируется новый ряд значений, где отсутствие пропусков будет равно 90.

Далее значения суммы проверяются по условию: `=ЕСЛИ(G114=90;"";"XXXXXX")`.

Наличие пропуска будет заметно на фоне столбца значений.

После чего необходимо будет выбрать такие фрагменты рядов для рассмотрения, которые не содержат пропусков. Например, в случае анализ периодического ВР пропуски могут исказить и затруднить его анализ, сказаться на будущей модели ряда.

Выбранные образцы сохраняются в файлах текстового формата .csv, при этом следует помнить, что для работы с этими значениями в Python необходимо преобразовать формат числа – заменить запятую точкой.

Таким образом, были сформированы файлы выборок для активной мощности, давления буферного, тока фазы В и выборка из давления буферного для проверки на разладку.

3.3.1 Анализ и моделирование

В ходе работы над ВКР была выработана методика анализа ВР:

- 1) загрузка ряда и выделение из него фрагмента для дальнейшего анализа;
- 2) задание ширины окна L, расчет относительного и общего вклада собственных троек в значения ВР, построение и анализ корреляционной матрицы;
- 3) отображение первых 16 (в зависимости от параметров) компонент по форме их сигналов для визуальной проверки выделенных компонент на возможность их группировки и сама группировка тренда, периодичностей;

- 4) группировка остатков и проверка их распределения на стационарность с помощью АКФ, диаграммы размаха (ящичная), плотности распределения, и отображение результирующего сложения;
- 5) построение модели с выбором компонент по метрике (номера включенных компонент, их количество, критерии MAPE, AIC).

Шаги 2-5 могут повториться в случае получения неудовлетворительной модели.

Критерий MAPE – средняя абсолютная ошибка в процентах, используется в данной работе для сравнения модели с фактическими значениями, рассчитывается по формуле:

$$MAPE = \frac{1}{N} \sum_{t=1}^N \left| \frac{y(t) - \tilde{y}(t)}{y(t)} \right| \times 100\%$$

где:

$y(t)$ – фактическое значение наблюдения t ;

$\tilde{y}(t)$ – значение модели наблюдения t .

Для исключения случая деления на ноль в знаменатель была добавлена поправка 10^{-2} , т.к. наблюдения ВР редко оказываются близкими нулю, то данная поправка не внесет значимым искажений.

Критерий AIC – используется при выборе модели в статистическом анализе, таким образом, чтобы минимальное количество параметров описывало максимально ВР. Хорошим критерием считается минимальный среди рассчитанных моделей.

$$AIC = \frac{k + L}{N} + \ln \frac{\sum_{i=1}^N e^2}{N}$$

где:

k – количество компонент в модели;

e – разность фактического и модельного значений.

В данном исследовании была произведена модификация, без которой наилучшей бы являлась та модель, которая вбирает все компоненты, но тогда критерий лишается смысла. При добавлении ширины окна L критерий адекватно отражает оптимальность выбора компонент модели при определенном значении L . Так при минимальном значении L и количестве компонентов ожидается получение наименьшего, а значит, наилучшего значения, однако при малом значении L как правило обеспечивается слабая разделимость компонент, а при минимальном их выборе получаем модель, часто далекую от реального ряда, о чем будет сигнализировать критерий MAPE. Лучшая оценка метрики – это максимальный MAPE при минимальном AIC.

3.3.2 ВР активной мощности, $L = 72$

Рассмотрим ВР активной мощности как пример ряда с сильной периодичностью.

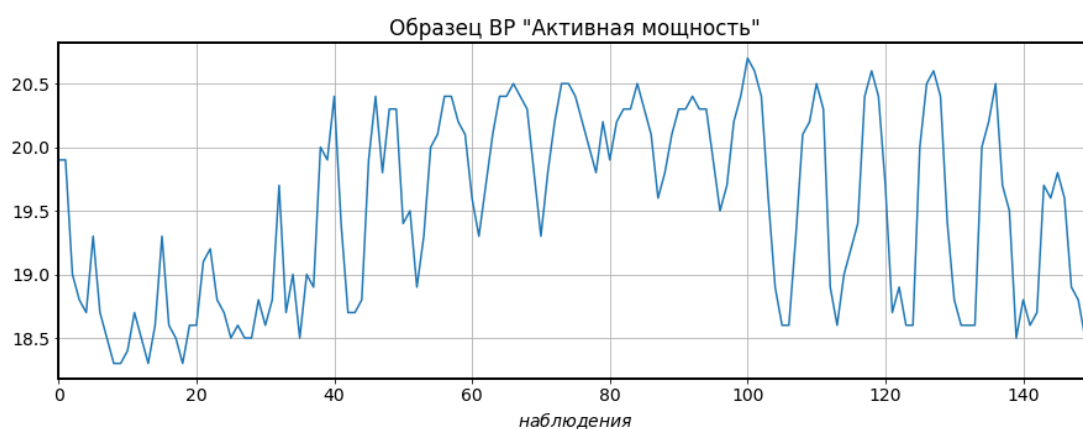


Рис.12 Взятый для анализа ВР активной мощности.

Выделим конец ряда, имеющий стабильную периодику как по амплитуде, так и по фазе, и минимальные искажения.

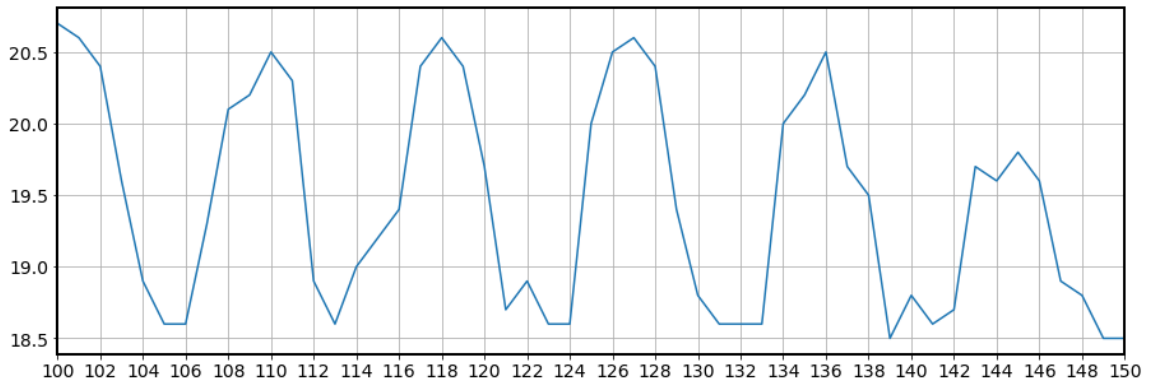


Рис.13 Фрагмент ВР активной мощности для определения периодичности.

Из графика получается, что период равен 9.

Согласно рекомендациям к выбору ширины окна [6, 7, 8] зададим L кратное периоду и максимальное: $L = 72$. Также для дальнейшего анализа будем работать с фрагментом с 6 до 150 метки, длина которого составит 144 наблюдений.

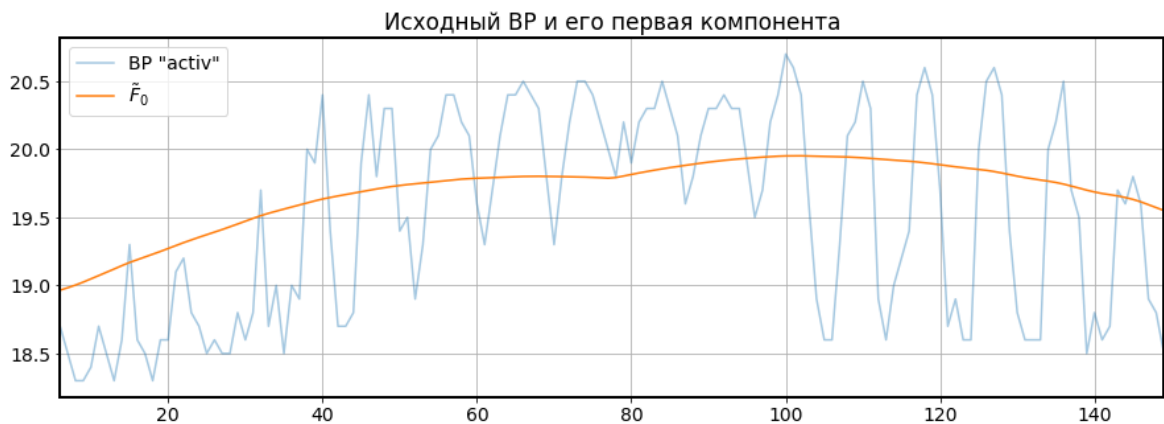


Рис.14 Исходный ВР активной мощности и его первая компонента.

Как видно из диаграммы, при максимальной ширине окна $\tilde{F}_0$ компонента не имеет гармонических составляющих и можно было бы считать ее выпуклой, если бы не 1 перегиб в метке 72 выбранного фрагмента (144) – метка 78 всего ряда. В окрестности этой же точки изменяются колебания исходного ВР – смена дисперсии. Оба признака указывают на наличие разладки.

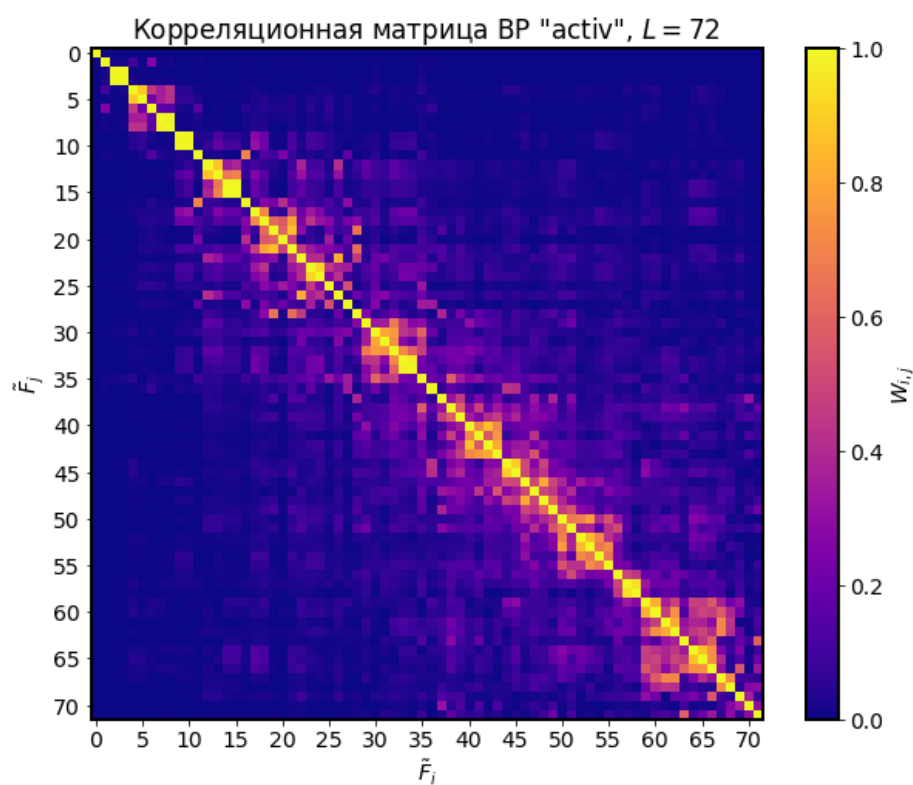


Рис.15 Корреляционная матрица ВР активной мощности с шириной окна $L = 72$

Увеличим первые 11 компонент. Действительно можно выделить эти 11 компонент из общего числа:

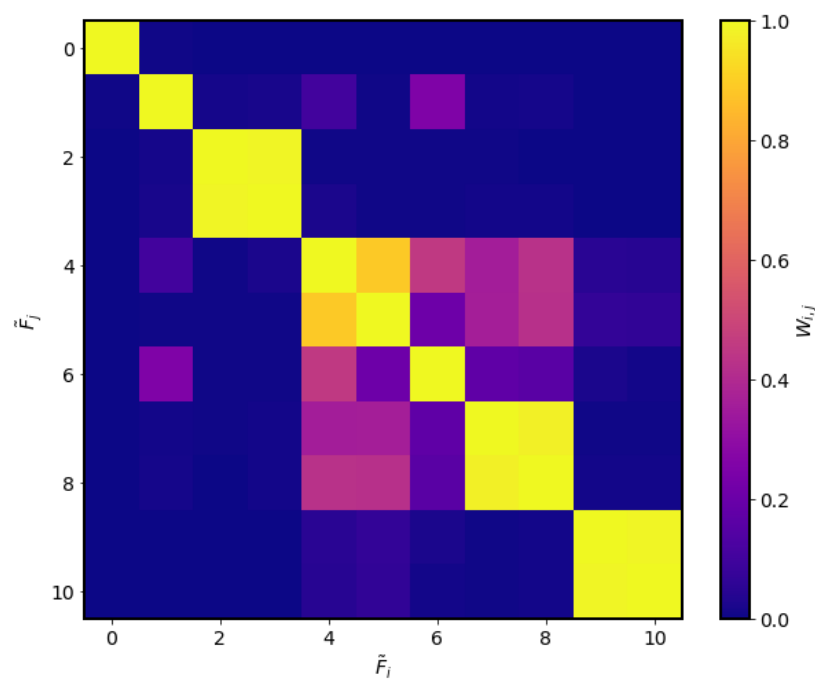


Рис.16 Минор корреляционной матрицы из первых 11 компонент ВР

В целом, получается, что минор 0-8 не имеет ковариации с остальными компонентами, поле которых достаточно рассеянно, что указывает на наличие шума, основной его части.

Рассмотрим первые 16 компонент ВР:

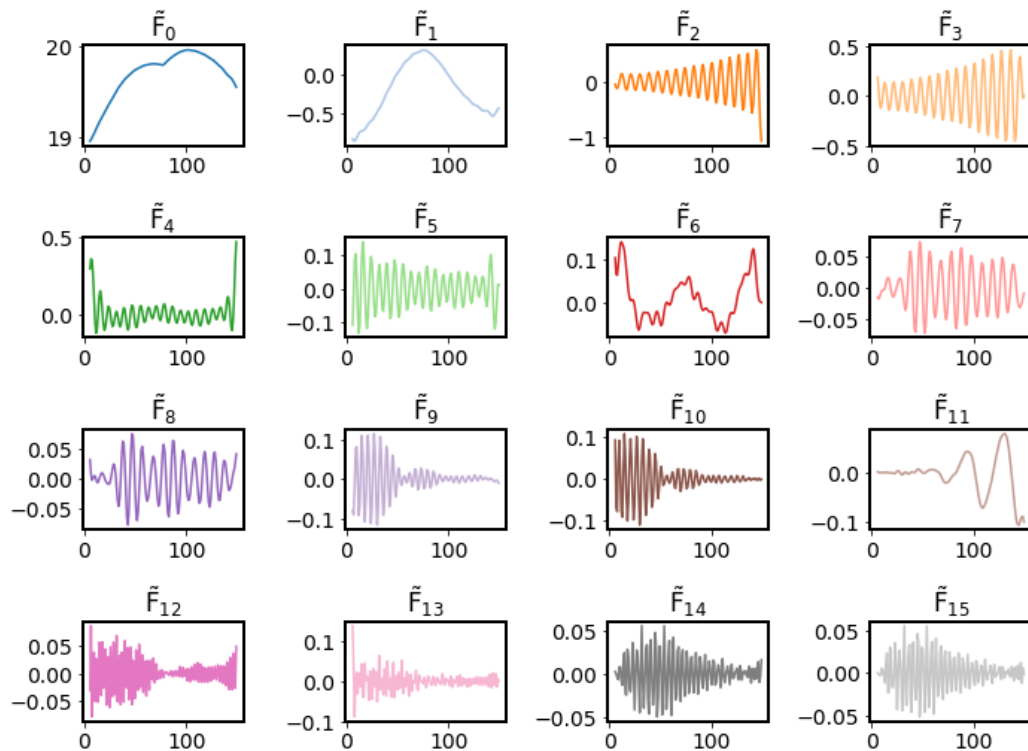


Рис.17 Первые 16 компонент ВР активной мощности при $L = 72$

Как видно из диаграммы, к тренду возможно отнести $\tilde{F}_0, \tilde{F}_1$ компоненты. К периодичной компоненте ВР можно отнести $\tilde{F}_2, \tilde{F}_3, \tilde{F}_5, \tilde{F}_7, \tilde{F}_8$ компоненты. Возможно добавление $\tilde{F}_9, \tilde{F}_{10}$ и $\tilde{F}_{11}$ компоненты – они наряду с трендовыми хорошо показывают момент разладки.

Если группировать в остатки с 11-й по 71-й (с 10 по 72 в привычной нумерации) компоненты, то получим:

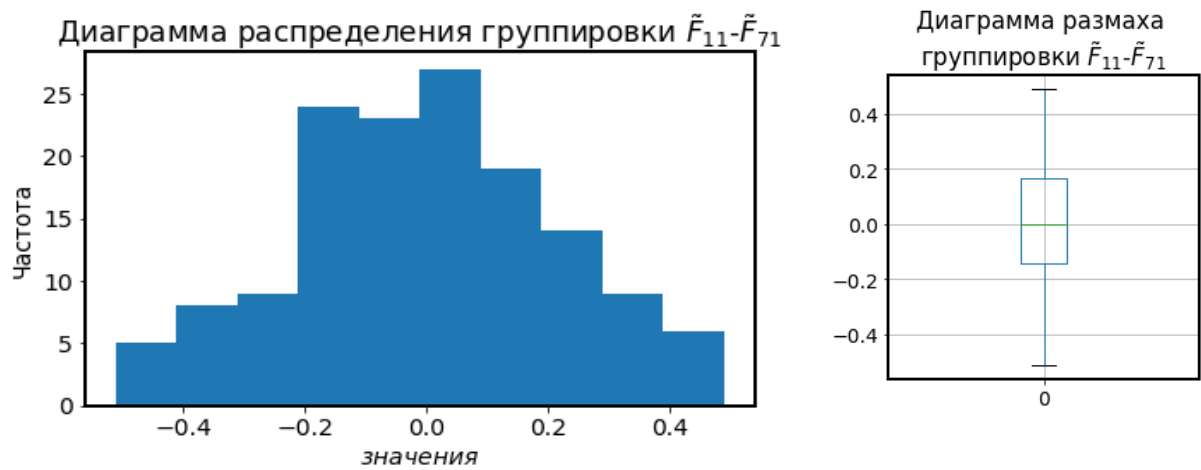


Рис.18 Диаграммы остатков ВР активной мощности при $L = 72$

Отобразим для сравнения коррелограммы АКФ ВР, остатков ряда и, для сравнения, тестового белого шума:

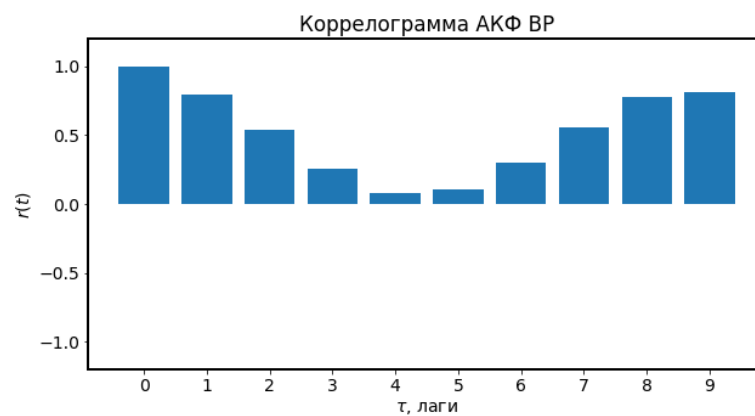


Рис.19 АКФ ряда активной мощности, $L = 72$

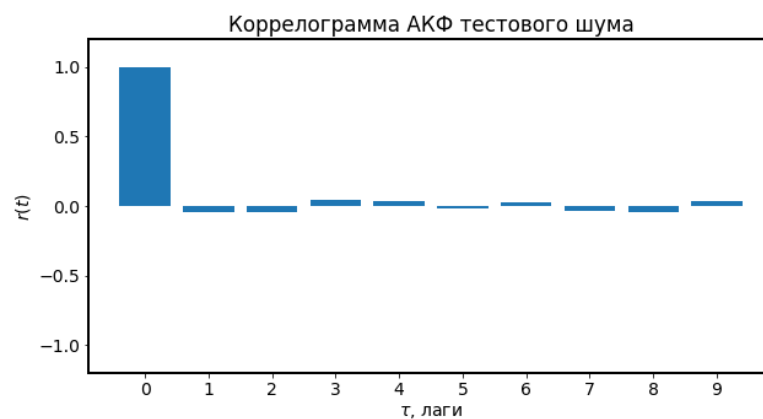


Рис.20 АКФ остатков ряда активной мощности, $L = 72$

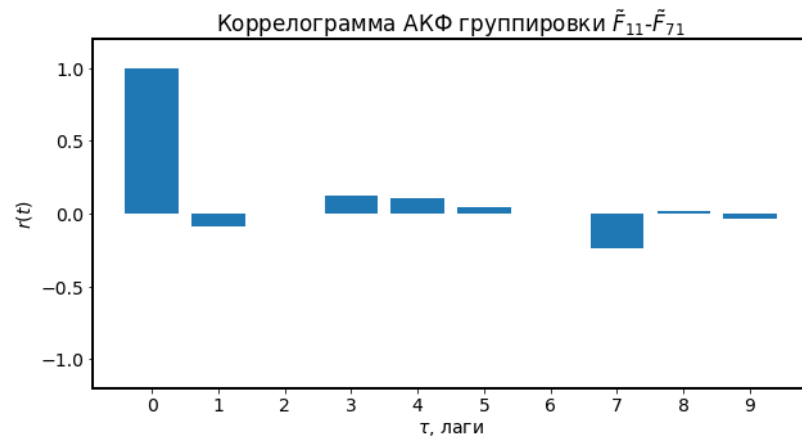


Рис.21 АКФ белого шума.

Здесь следует обратить внимание на то, что нулевой лаг – это по сути значение коэффициента автокорреляции текущего наблюдения самого на себе. Логично, что он будет равен единице. Все последующие коэффициенты отражают взаимосвязь каждого наблюдения с 9 последующими. По АКФ для остатков видно, что автоковариация практически убрана, но есть влияние 7 лага, т.е. какой-то гармонике с периодом в 7 наблюдений. Тем не менее по остальным диаграммам видно, что остатки получились приемлемы для стационарного ряда, и их можно классифицировать как шум, уверенно исключая из дальнейшего рассмотрения.

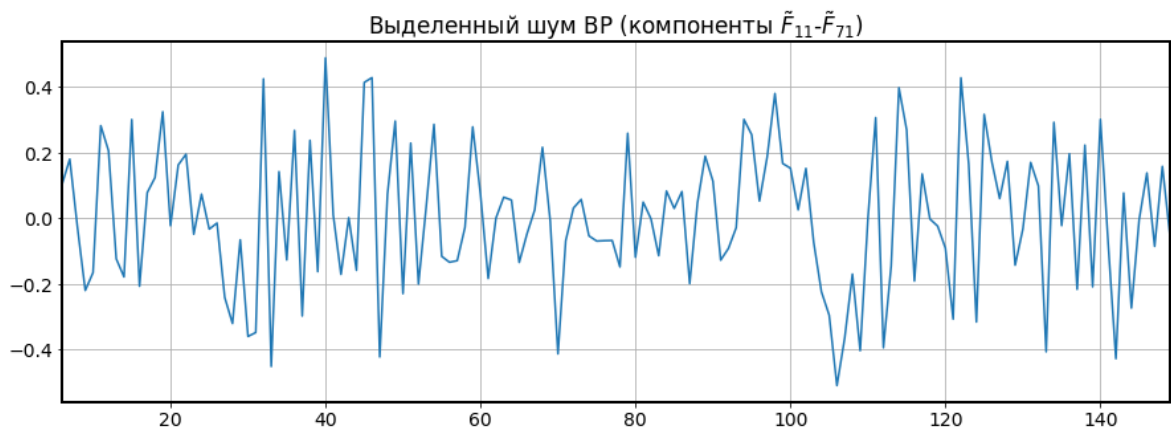


Рис.22 Восстановленный ряд остатков ВР, $L = 72$

Далее приступаем к построению модели. Здесь используем метрику для определения оптимального количества компонент, которые составят модель, и отображаем модель в сравнении с исходным ВР.

0, 1, 2, 3, 4, 5, 6, 7,	0	0, 1, 2, 3, 9, 10	0, 2, 3, 9, 10
-------------------------	---	-------------------	----------------

8, 9, 10, 11, 12, 13, 14 k = 15 MAPE = 0.8% AIC = 12.191	k = 1 MAPE = 2.898% AIC = 0.632	k = 6 MAPE = 1.188% AIC = -1.97	k = 5 MAPE = 1.996% AIC = -1.038
0, 1, 2, 3, 9 k = 5 MAPE = 1.204% AIC = -1.94	0, 1, 2, 3, 10 k = 5 MAPE = 1.196% AIC = -1.952	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 k = 11 MAPE = 0.904% AIC = -2.486	0, 1, 2, 3 k = 4 MAPE = 1.231% AIC = -1.892

Табл.1 Метрика модели ВР при выборе различных наборов компонент.

Исходя из таблицы, определяем наилучший набор компонент.



Рис.23 Исходный и восстановленный ВР (модель), $L = 72$

Таким образом получили модель для фрагмента ВР активной мощности.

3.3.3 ВР активной мощности, $L = 9$

Рассмотрим случай, когда $L = 9$ – минимальная и кратная периодики ряда ширина окна.

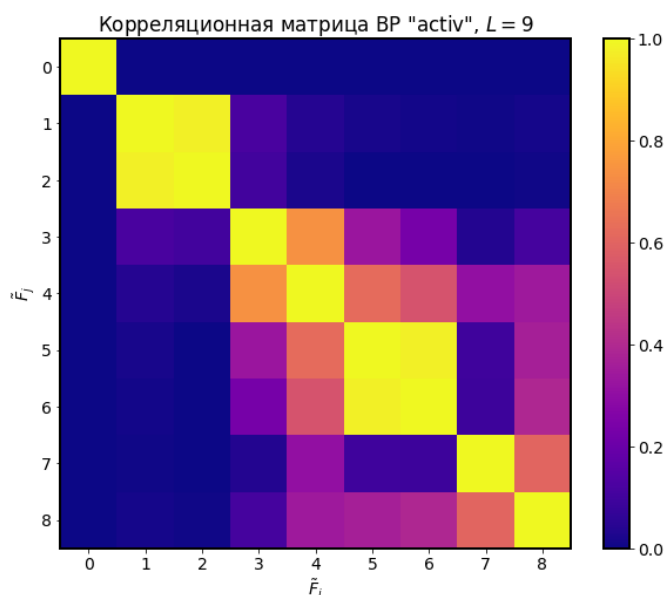


Рис.24 Корреляционная матрица ВР активной мощности, $L = 9$

Видно, что компоненты разделяются на 3 минора: $\tilde{F}_0; \tilde{F}_1, \tilde{F}_2$ и остальные 6 компонент, которые содержат больше смешанных компонент, чем первые 9 в случае с $L = 72$.

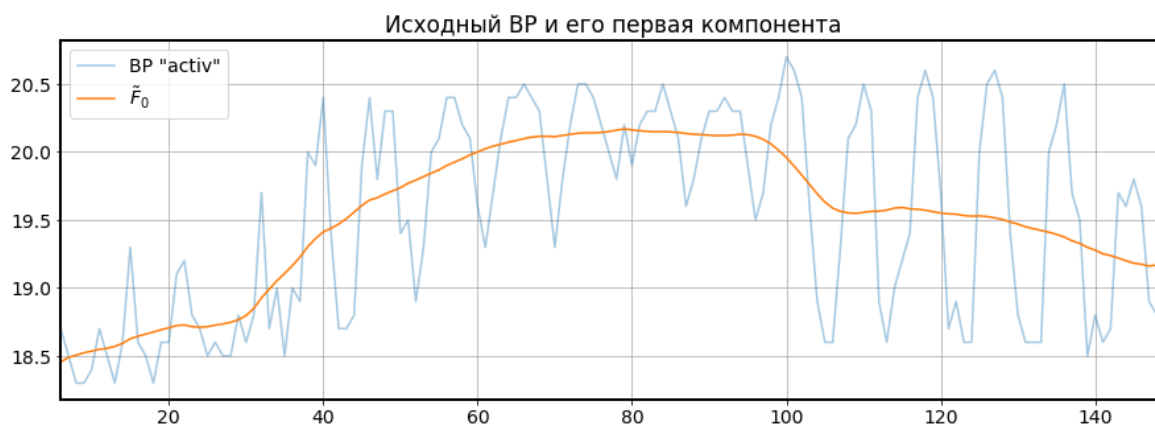


Рис.25 ВР активной мощности и первая компонента, $L = 9$

Первая ($\tilde{F}_0$) компонента при $L = 9$ более убедительно описывает тренд ВР чем при $L = 72$, хотя кривая изменяется менее плавно. Максимальное значение достигается в точке 79 всего ряда. После чего функция компоненты убывает, и меняется дисперсия.

Отообразим входящие компоненты:

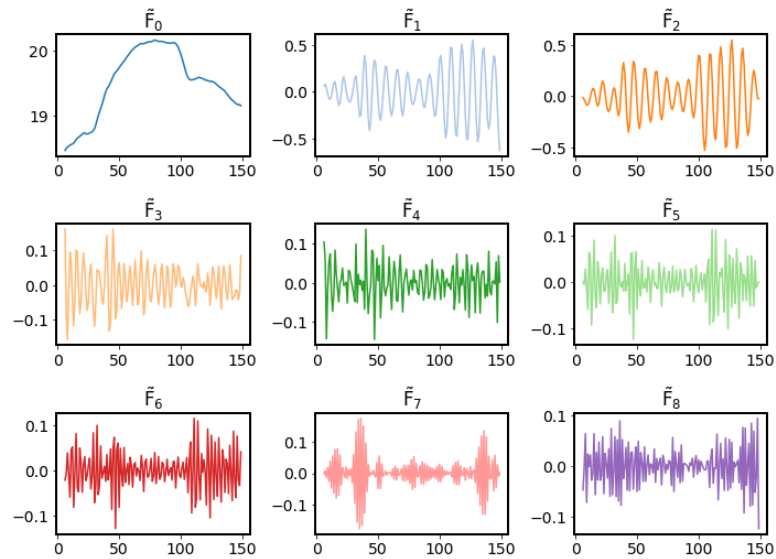


Рис.26 Диаграмма всех 9 компонент ВР

Из диаграммы можно предположить, что $\tilde{F}_0$ – тренд, $\tilde{F}_1, \tilde{F}_2$ – периодики, а остальное – шум. Амплитудные модуляции могут быть обусловлены согласно [7, 8] влиянием периодических компонент.

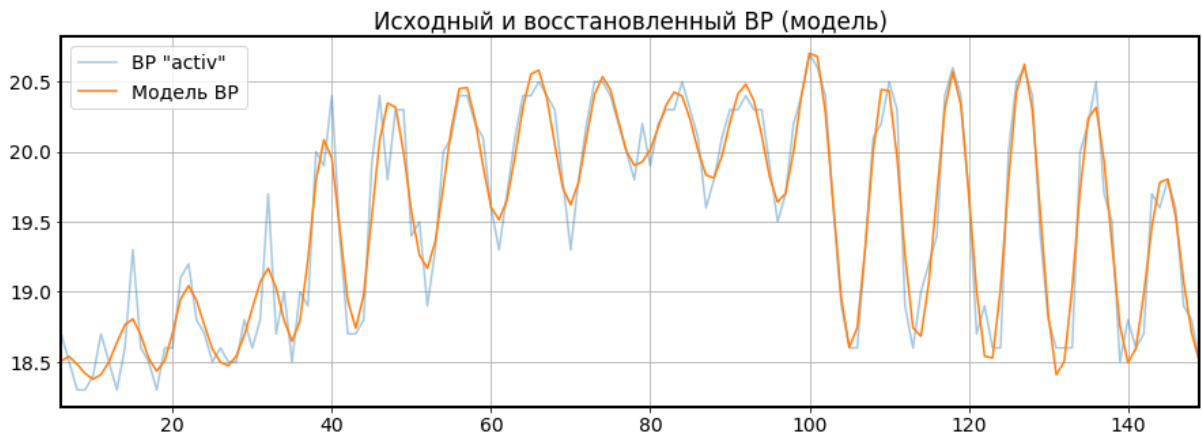


Рис.27 Исходный и восстановленный ВР (модель), $L = 9$

0, 1, 2
 $k = 3$
 $\text{MAPE} = 0.764\%$
 $\text{AIC} = -3.223$

Для более простых и быстрых результатов лучше использовать меньшую ширину окна L , но выделение шума оказалось точнее при большем L .

3.3.4 ВР активной мощности, $L = 18$

Рассмотрим ряд при $L = 18$ – видим, что нулевая ($\tilde{F}_0$) компонента достаточно хорошо и плавно описывает ВР.

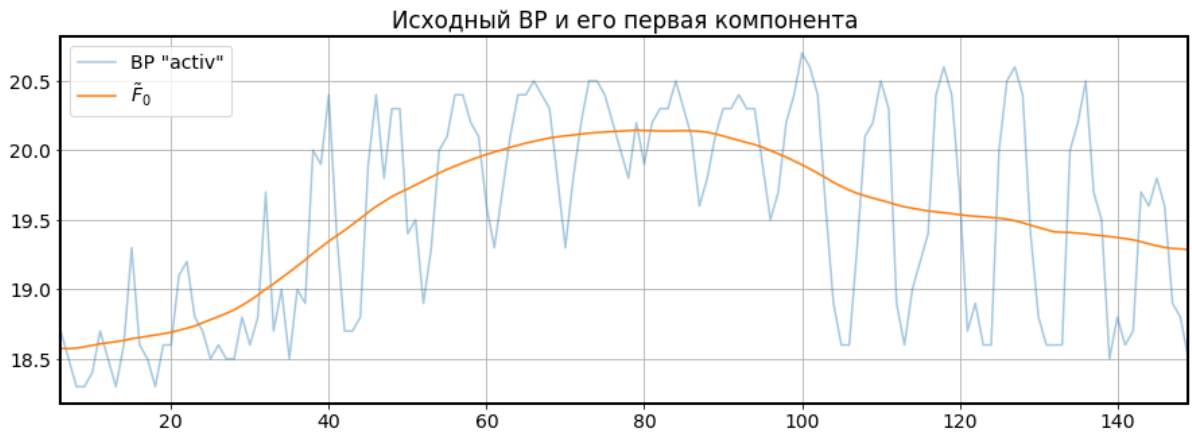


Рис.28 ВР активной мощности и первая компонента, $L = 18$

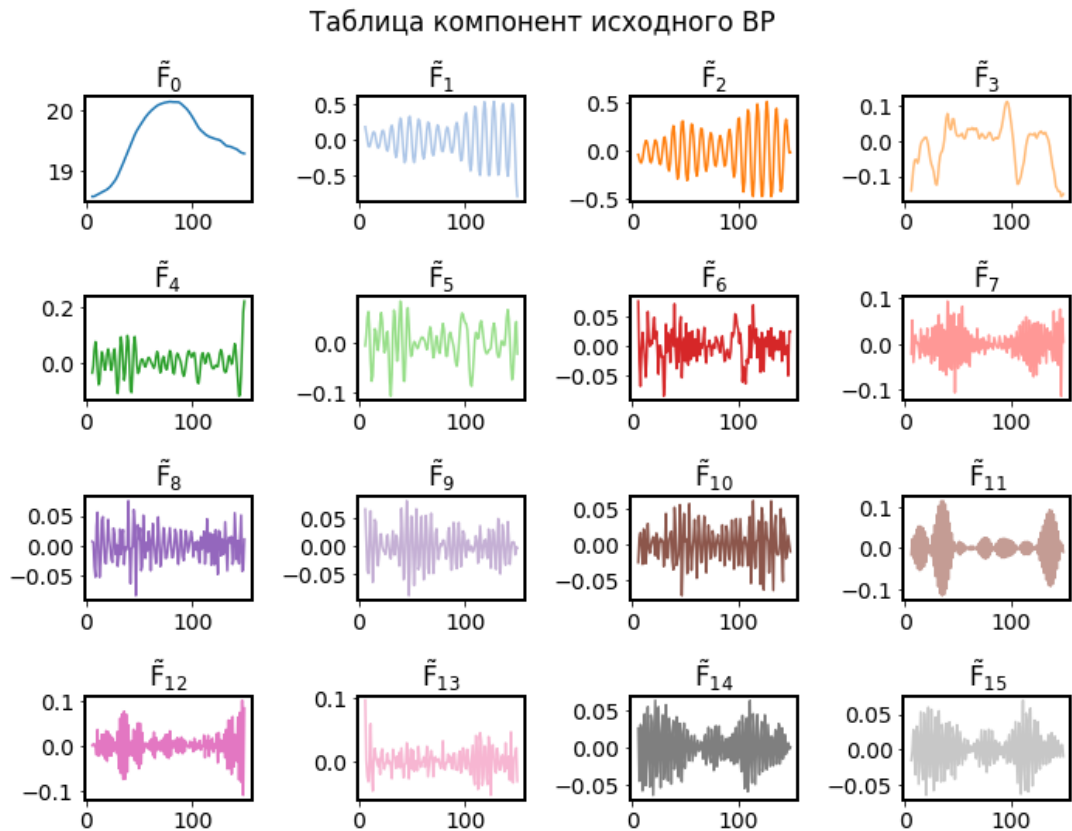


Рис.29 Первые 16 компонент ВР, $L = 18$

Среди компонент $\tilde{F}_1$ и $\tilde{F}_2$ хорошо представляют периодичную составляющую, при этом отражая разладку сменой дисперсии. При этом в сравнении с $L = 9$ они вместе с компонентой тренда стали плавнее. Т.е. разделимость улучшилась, что подтверждается корреляционной матрицей.

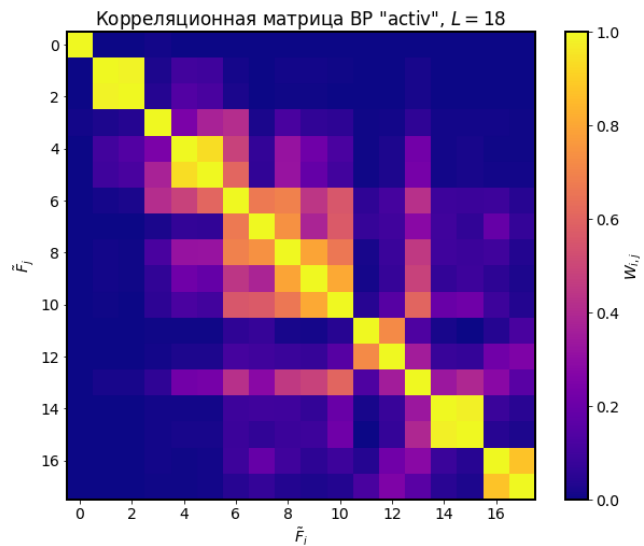


Рис.30 Корреляционная матрица ВР, $L = 18$

Таким образом, $\tilde{F}_0$ представляет тренд ВР, а $\tilde{F}_1$ и $\tilde{F}_2$ – периодичность ВР, остальные 15 компонент выделим как шум. Отобразим их диаграммы.

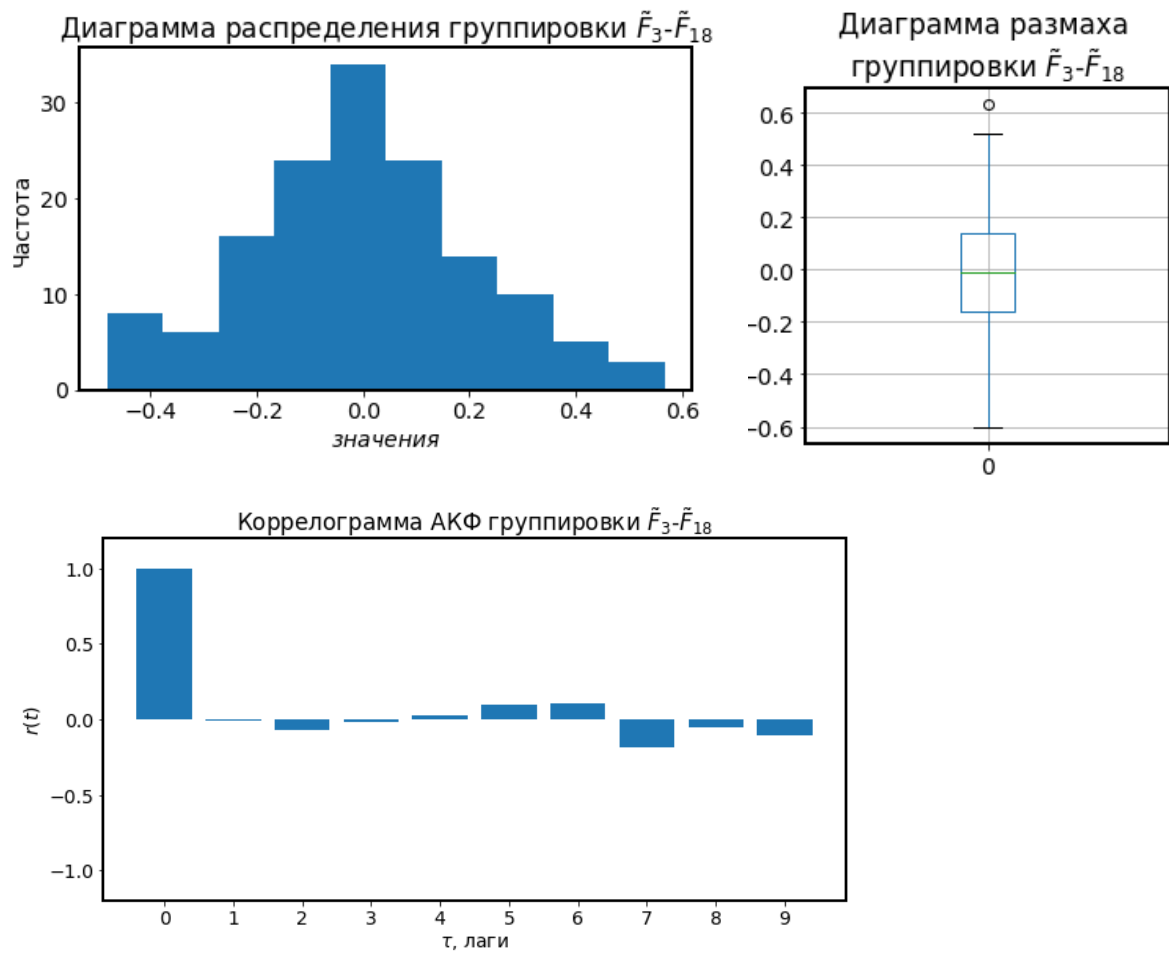


Рис.31 Диаграммы остатков ВР, $L = 18$

С одной стороны, диаграммы размаха и распределения показывают большее несоответствие остатков нормальному распределению, с другой стороны коэффициенты лагов АКФ стали менее выражены, хотя влияние 7 лага все еще видно.

Также частью исследования было рассмотрение ВР при различных значениях L . Было установлено, что разделимость лучше при $L = 9$, чем при $L = 8$ или $L = 10$, что говорит в пользу выбора ширины окна, кратного наблюдаемым периодичностям. Разделимость при $L = 72$ не значительно лучше, чем при $L = 63$. Диаграммы представлены в приложении Б.

Отообразим особенности проведенного анализа в таблице, на основании которых будут проанализированные другие ВР набора данных.

Признак	Ширина окна ТМ ВР		
	$L=9$	$L=18$	$L=72$
Описывание тренда	Хорошее описывание тренда	Лучшее описывание тренда	Будет удобнее всего подбирать функцию, но нулевая компонента описывает тренд плохо
Разделимость компонент	Периодичности содержат гармонические огибающие	Кривая тренда плавная, периодичности выделены как при $L=9$, остатки выделены как при $L=72$	Периодичная компонента и шум выделены наилучшим образом
Выявление разладки	Возможно определить примерную область разладки	Возможно определить примерную область разладки	По максимуму первой компоненты определяется точка разладки
Метрика модели	0, 1, 2 k = 3 MAPE = 0.764% AIC* = -3.223	0, 1, 2 k = 3 MAPE = 0.94% AIC* = -2.761	0, 1, 2, 3, 9, 10, 4, 5, 6, 7, 8 k = 11 MAPE = 0.904% AIC* = -2.486

Табл.2 Итоговая таблица отражающая особенности выбора L

3.3.5 ВР буферного давления

Рассмотрим фрагмент буферного давления длиной 250 наблюдений.



Рис.32 ВР буферного давления

Выделим фрагмент интервалом 20-40:

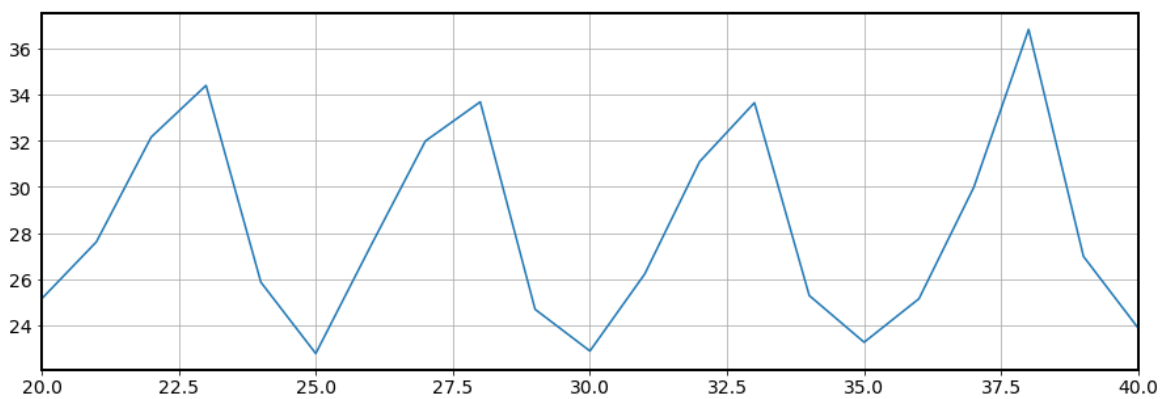


Рис.33 Фрагмент ВР буферного давления для определения периодичности

Из фрагмента видно, что период колебаний повторяется каждые 5 наблюдений.

Выберем ширину окна $L = 125$ и применим к классу SSA.

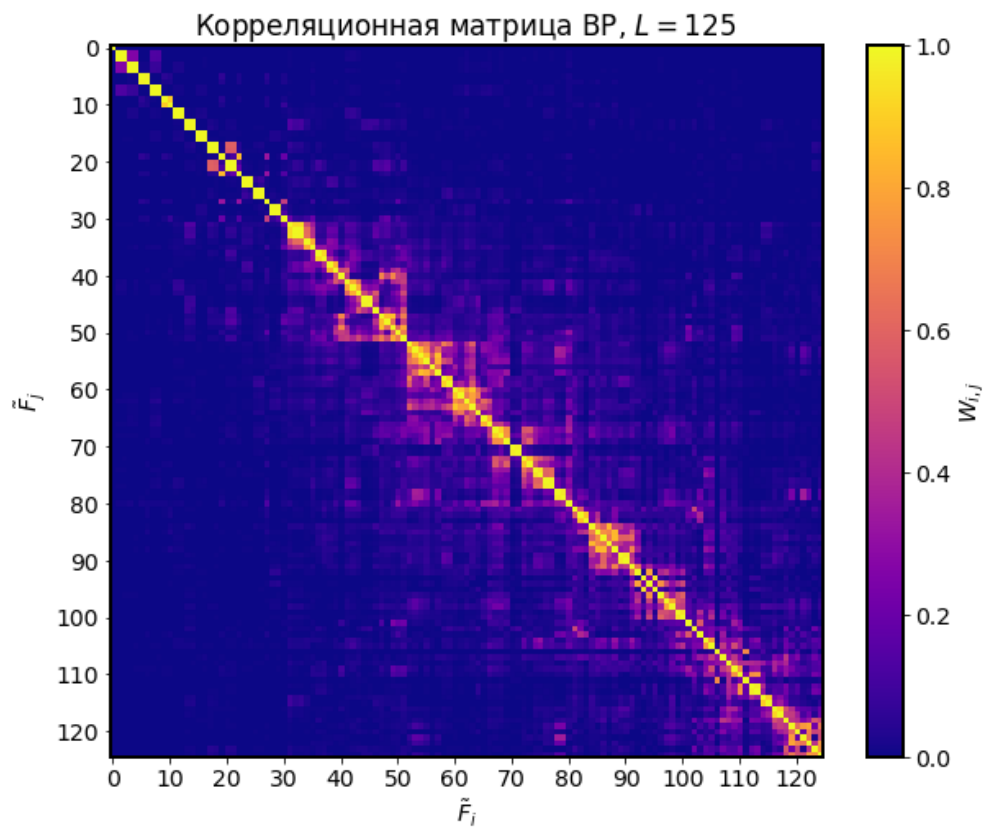


Рис.34 Корреляционная матрица ВР буферного давления, $L = 125$

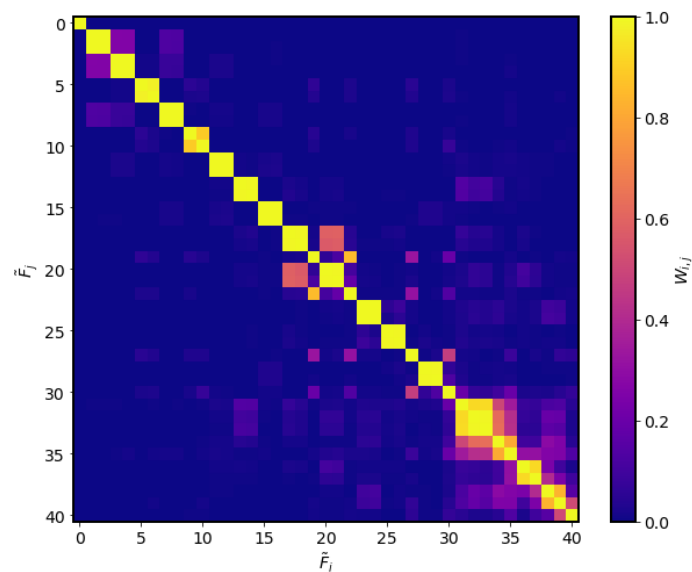


Рис.35 Первые 41 компоненты корреляционной матрицы, $L = 125$

Из построенной корреляционной матрицы видно, что большее число смешанных моментов расположено в миноре 30-124 – основной шум ряда находится здесь. Возможно, что все остальные компоненты придется включить в модель.

Ниже представлены первые 32 компонента ВР, где $\tilde{F}_{31}$ — напоминает шум, т.к. не имеет плавной огибающей, $\tilde{F}_5, \tilde{F}_6, \tilde{F}_9, \tilde{F}_{10}, \tilde{F}_{19}, \tilde{F}_{22}, \tilde{F}_{27}, \tilde{F}_{30}$ — периодичные компоненты, но только низкой частоты. Все остальные компоненты кроме, разумеется, $\tilde{F}_0$, это периодичности высокой частоты, как раз с периодом в 5 наблюдений.

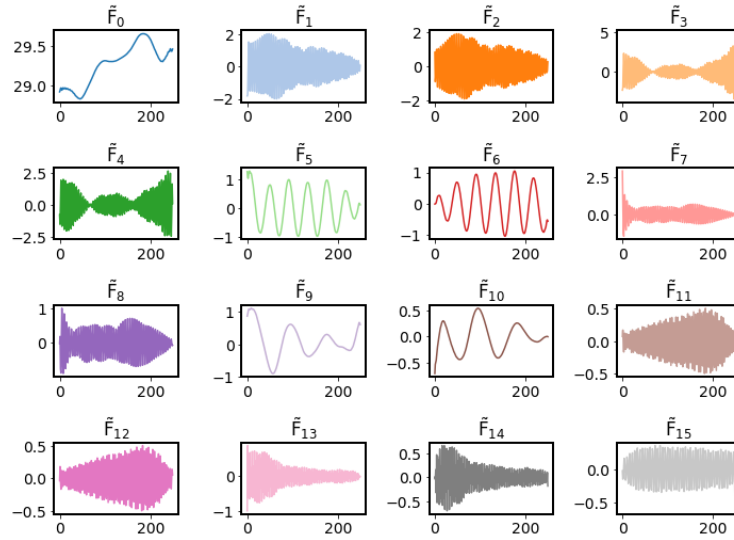


Рис.36 Первые 16 компонент ВР буферного давления, L=125

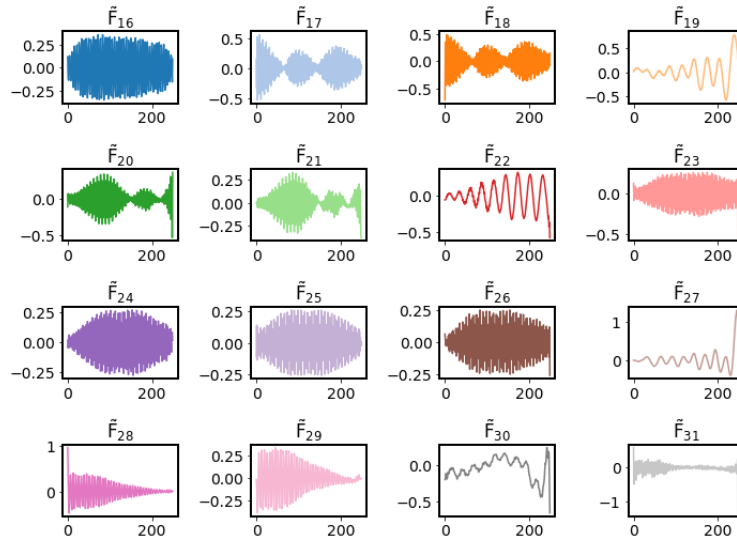


Рис.37 Следующие 16 компонент ВР буферного давления, L=125

Отообразим для наглядности первые 4 НЧ-периодичности.

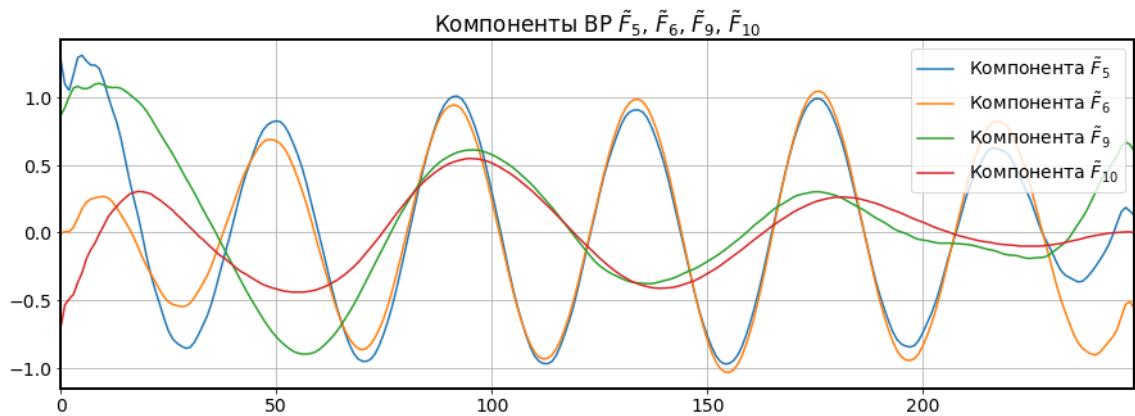


Рис.38 Первые 4 НЧ-компоненты ВР буферного давления, $L = 125$

Отобразим модель, в которую включим их и компоненту тренда на фоне исходного ВР.

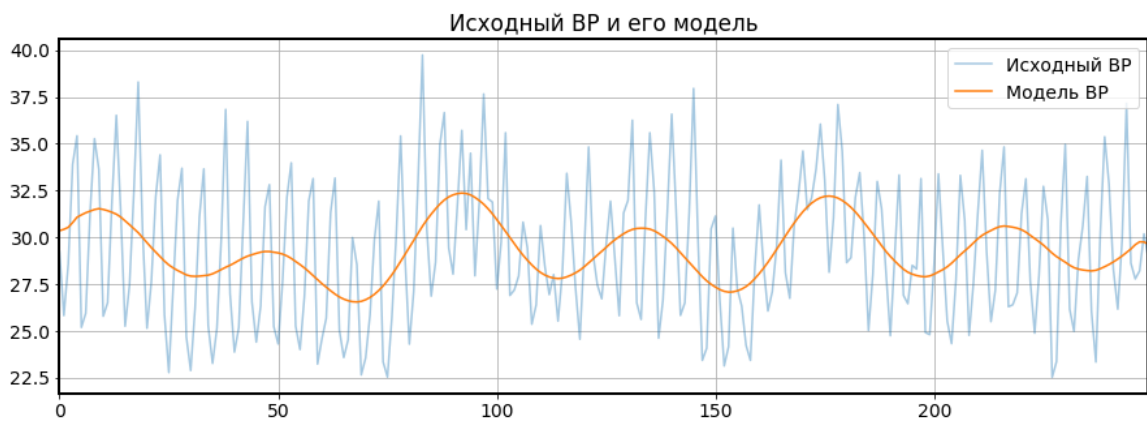


Рисунок 39 Исходный ВР модель из 5 компонент, $L = 125$

Исходя из диаграммы видно, что модель очень хорошо описывает тренд ряда, однако полностью отсутствуют колебания.

0, 5, 6, 9, 10
 $k = 5$
 $MAPE = 10.326\%$
 $AIC = 3.051$

Из метрики видно, что включенные 5 компонент обеспечивают почти 90% точность модели. Возможно, что такая модель уже может быть применена для прогноза, либо для подбора аппроксимации функции для представления в аналитической форме (уравнение, либо система уравнений). Но дополним модель ВЧ-колебаниями.

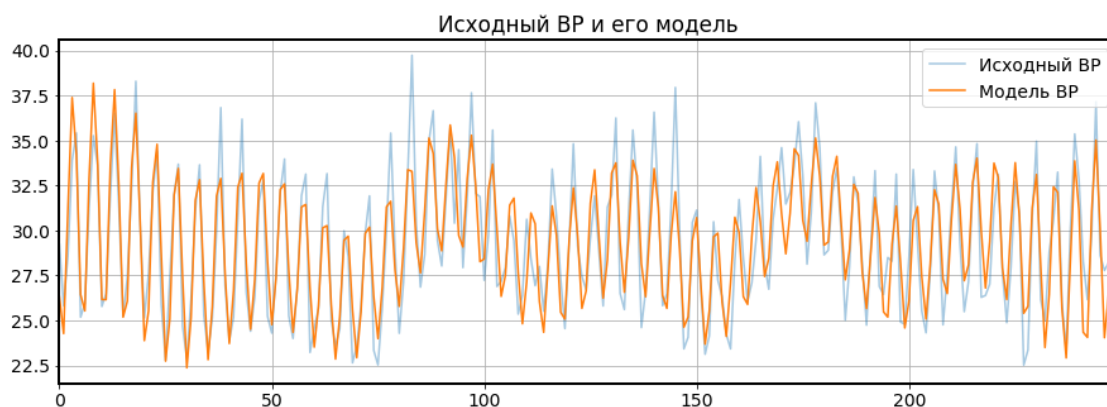


Рис.40 Исходный ВР модель из 9 компонент, $L = 125$

0, 1, 2, 3, 4, 5, 6, 9, 10
 $k = 9$
 $MAPE = 5.212\%$
 $AIC = 1.828$

Были введены высокочастотные компоненты с наиболее значимым сингулярным числом, качество метрики стало лучше.

Если же включить компоненты согласно корреляционной матрицы: с 0 по 30, то получим:

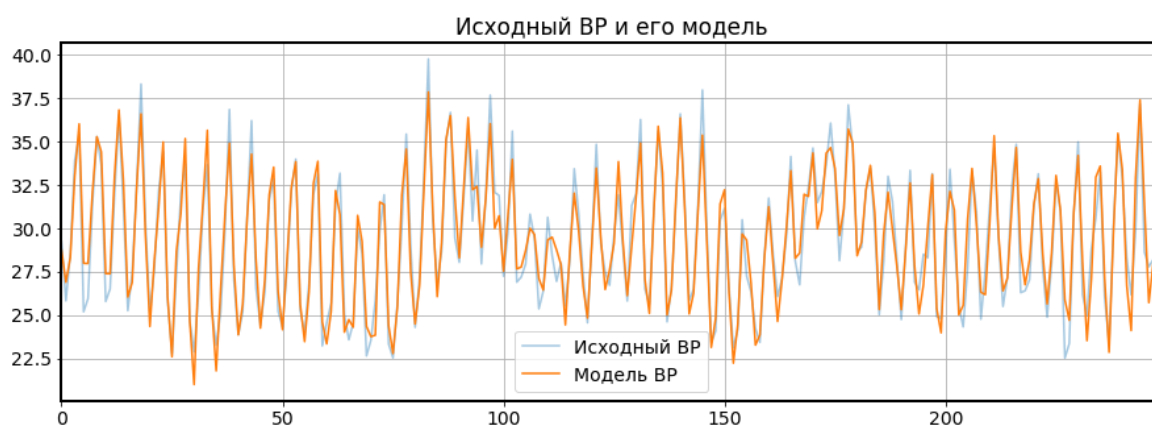


Рис.41 Исходный ВР модель из 31 компоненты, $L = 125$

0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12,
 13, 14, 15, 16, 17, 18, 19, 20, 21, 22,
 23, 24, 25, 26, 27, 28, 29, 30
 $k = 31$
 $MAPE = 3.156\%$
 $AIC = 0.931$

Оставив только первые 10 компонент, получим:

0, 1, 2, 3, 4, 5, 6, 7, 8, 9
 $k = 10$
 $MAPE = 4.739\%$
 $AIC = 1.675$

Теперь точность модели достаточна для порога в 95%.

Подытожив анализ для фрагмента ВР буферного давления получаем, что с помощью корреляционной матрицы выделяем диапазон компонент ряда, которые можно рассмотреть для включения в модель ряда. Затем с помощью формы выделенных компонент и метрики модели подбираем оптимальный набор компонент.

3.3.6 ВР тока двигателя

Рассмотрим фрагмент ВР тока двигателя фазы В (более 500 наблюдений) в качестве примера выделения SSA тренда:

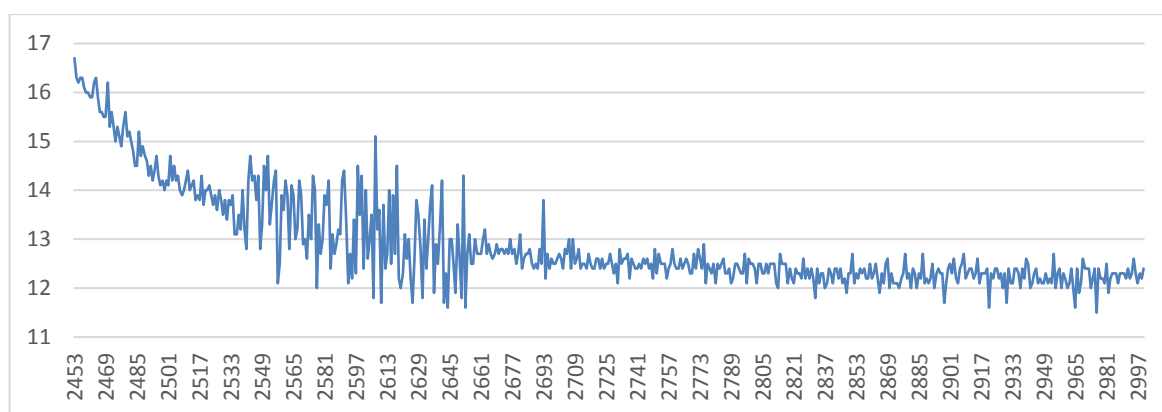


Рис.42 ВР тока двигателя

Исходя из полученного опыта, определим небольшое значение $L = 10$. Получим корреляционную матрицу:

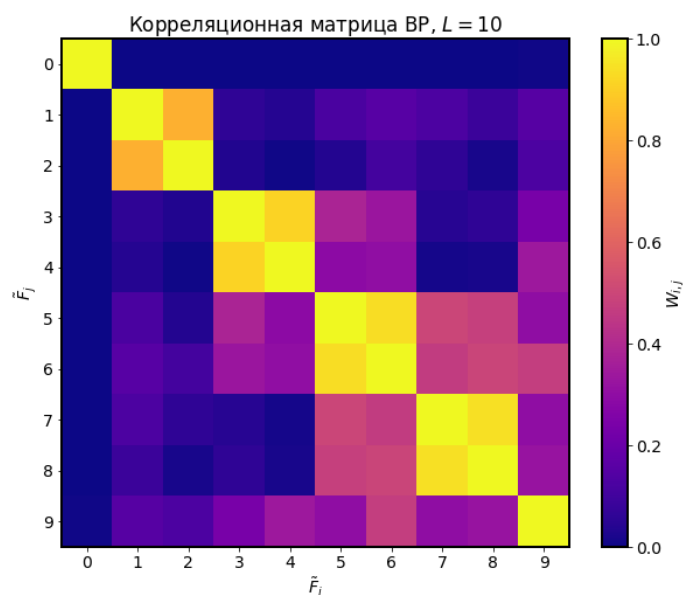


Рис.43 Корреляционная матрица ВР тока двигателя, $L = 10$

Матрица демонстрирует снова хорошую разделимость для нулевой компоненты, также хорошая разделимость имеется у $\tilde{F}_1$ и $\tilde{F}_2$ – это выражено в том, что их чистые моменты (квадраты на главной диагонали) имеют почти нулевую корреляцию с остальными. Однако, оба этих компонента имеют сильную корреляцию σ_{12} (равносильно σ_{21}), что говорит о том, что этим компонента можно сгруппировать вместе. Пара у $\tilde{F}_3$ и $\tilde{F}_4$ имеет примерно ту же характеристику, но разделимость все же хуже. Компоненты $\tilde{F}_5$ - $\tilde{F}_9$ имеют плохую разделимость и скорее всего содержат основной шум.

Дополним характеристику формой компонент:

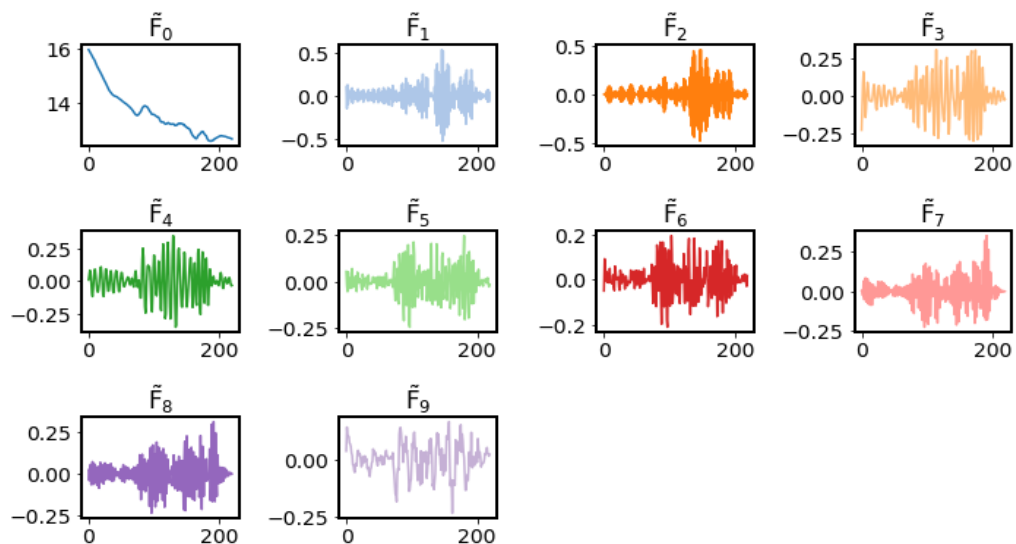


Рис.44 Диаграмма компонент ВР тока двигателя, $L = 10$

Видно, что формы компонент действительно соответствуют представлению корреляционной матрицы.

Отобразим нулевую компоненту и исходный ВР:

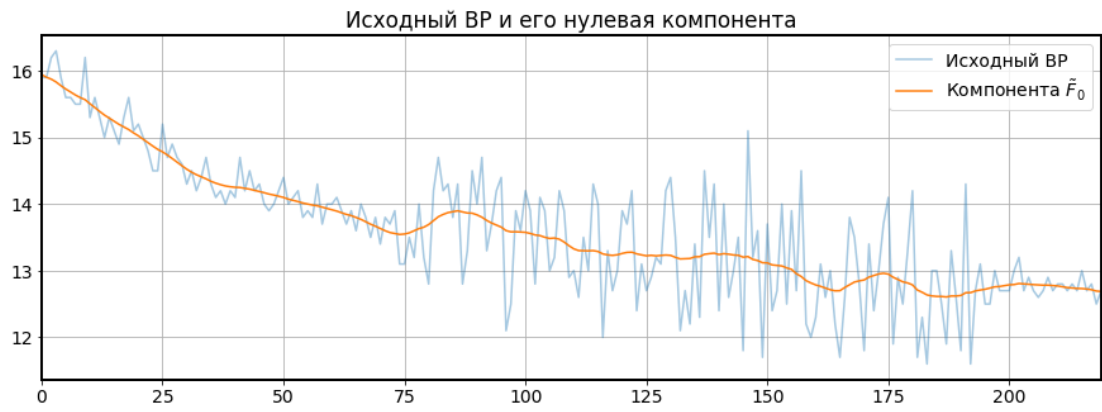


Рис.45 Исходный ВР и его $\tilde{F}_0$ компонента, $L = 10$

Нулевая компонента хорошо описывает тренд ВР, как на прямом участке (промежуток 0-30), так и в случае появления сильных колебаний.

Ниже представлены графики для различных значений L , видно, что с увеличением параметра ряд сглаживается. При $L = 90$ кривая становится монотонно убывающей, тем не менее хорошо описывает тренд.

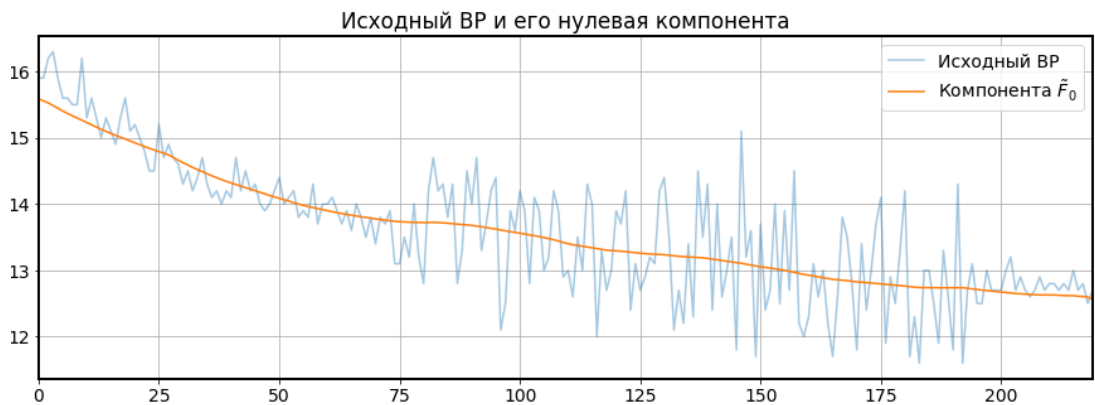


Рис.46 Исходный ВР и его $\tilde{F}_0$ компонента, $L = 28$

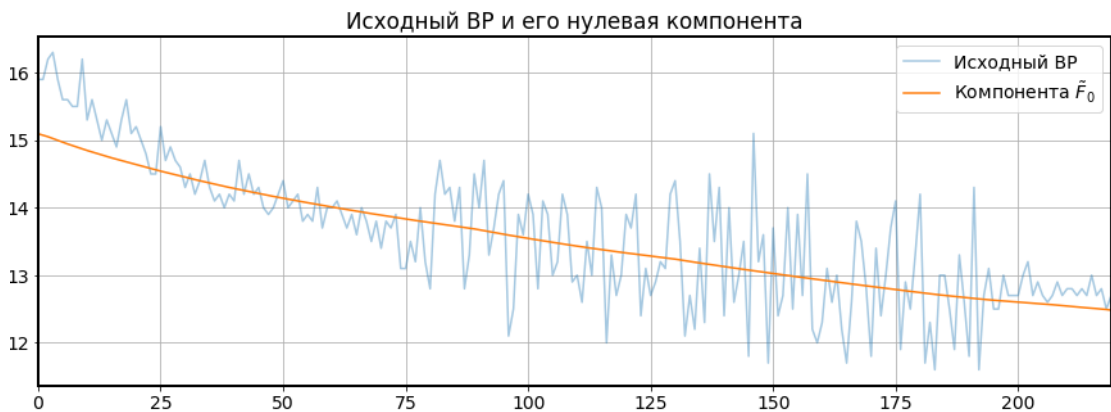


Рис.47 Исходный ВР и его $\tilde{F}_0$ компонента, $L = 90$

Исходя их анализа, делаем вывод, что SSA достаточно хорошо и просто выделяет тренд ВР.

3.3.7 Определение разладки

Для данной задачи используем 2 фрагмента ВР буферного давления, ряд их включающий представлен ниже.

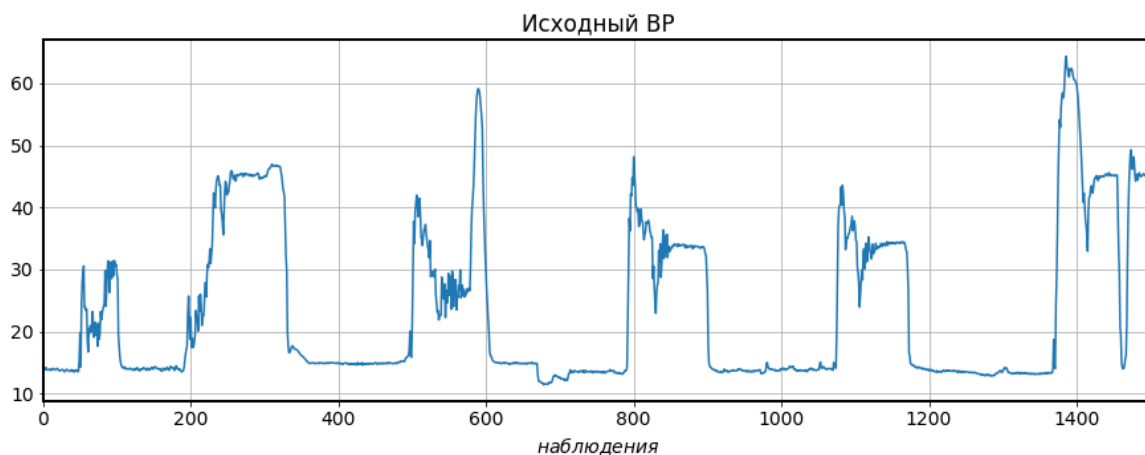


Рис.48 ВР буферного давления, используемый для определения разладки

Первый фрагмент (рис.49 а) является передним фронтом (резкий подъем) второго импульса всего ряда. Фрагмент имеет всплеск в начале, что должно затруднить определение разладки (ложноположительное реагирование).

Второй фрагмент (рис.49 д) является передним фронтом четвертого импульса всего ряда. Фрагмент не имеет всплеска в начале.

Оба фрагмента представляют собой образ «состояние 1 – переход – состояние 2».

Теперь добавим к рассмотрению еще по одному фрагменту, оба – укорочены к моменту начала подъема (рис.49 б, е), приблизительно 25-30 % от всей высоты. Оба эти фрагмента имитируют ситуацию, когда последнее наблюдение во фрагменте – состояние системы «сейчас» – момент, когда пошел подъем, но он еще не достиг максимума.

Проведя разложение фрагментов на их компоненты, была обнаружена интересная особенность, повторяемая во всех 4 случаях: компонента $\tilde{F}_1$ имеет

локальный минимум в месте начала подъема. Который можно считать однозначным, хотя и косвенным признаком начала разладки. Во всех случаях использовалась максимальная ширина окна L , т.к. необходима максимальная разделимость компонент с целью исключения в медленно меняющихся компонентах гармонической составляющей. Диаграммы фрагментов и отмеченные минимумы представлены на рис.49, где в подписи обозначены длина фрагмента N , выбранная ширина окна L и значение локального минимума $lmin$.

Таким образом, метод SSA позволяет обнаруживать разладку в самом начале ее возникновения.

В случае с периодичным примером – перегибом кривой тренда и изменением дисперсии периодик. А в случае с импульсом – по локальному минимуму компоненты $\tilde{F}_1$, определяющему приблизительную область начала подъема.

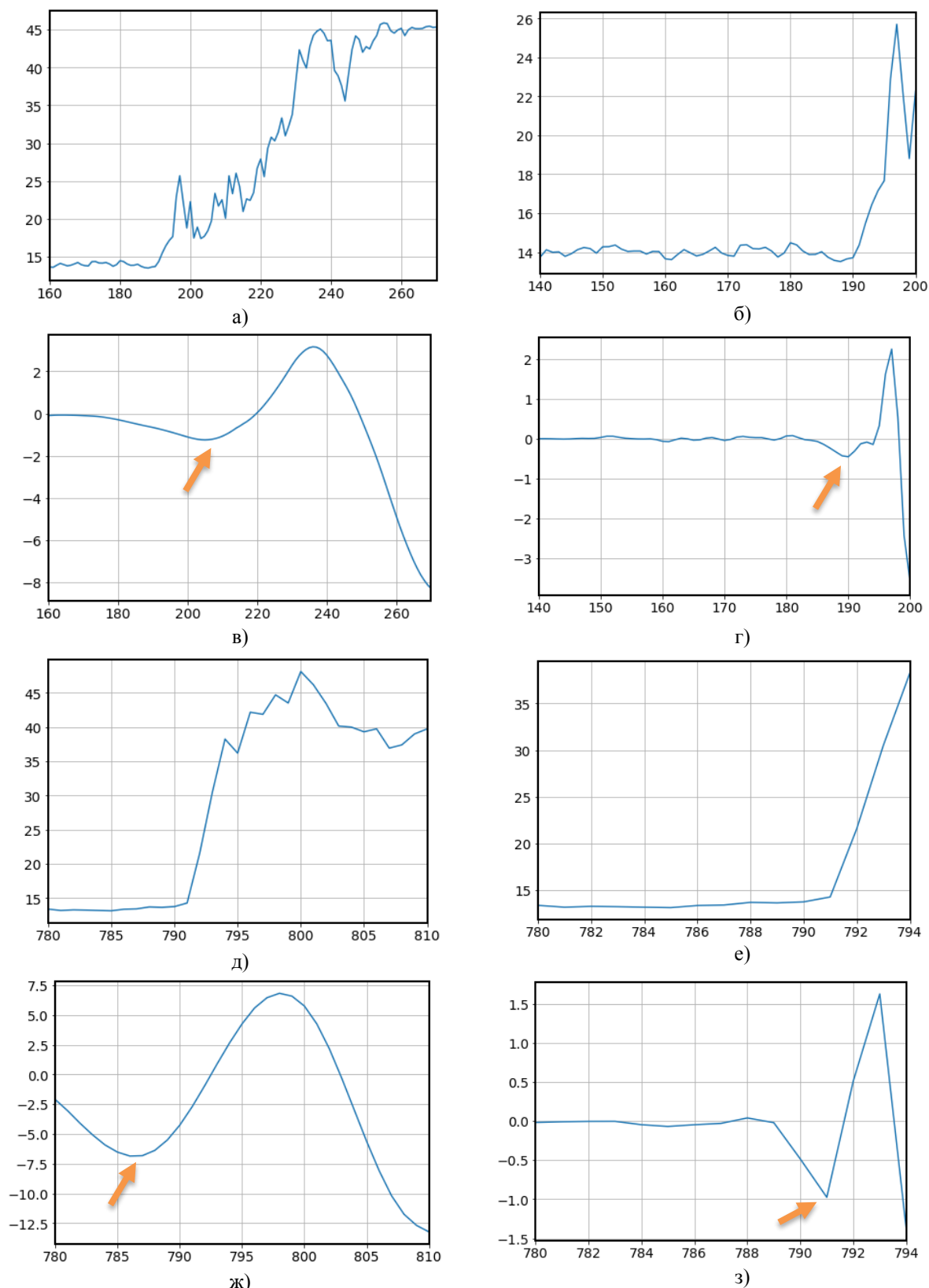


Рис. 49 Фрагменты для определения разладки: а) первый, $N=111$, $L=55$; б) первый укороченный, $N=61$, $L=30$; в) $\tilde{F}_1$ 1-го фрагмента, $l_{\min}=205$; г) $\tilde{F}_1$ 1-го укороченного фрагмента, $l_{\min}=195$; д) второй, $N=31$, $L=15$; е) второй укороченный, $N=15$, $L=7$; ж) $\tilde{F}_1$ 2-го фрагмента, $l_{\min}=789$; з) $\tilde{F}_1$ 2-го укороченного фрагмента, $l_{\min}=791$

3.3.8 Построение прогноза

Для прогнозирования была выбрана модель ВР, полученная в п.3.3.3.

Первые 108 наблюдений ряда являются тестовым набором, на основании которого строится прогноз на 4 периода – 36 значений контрольного набора.

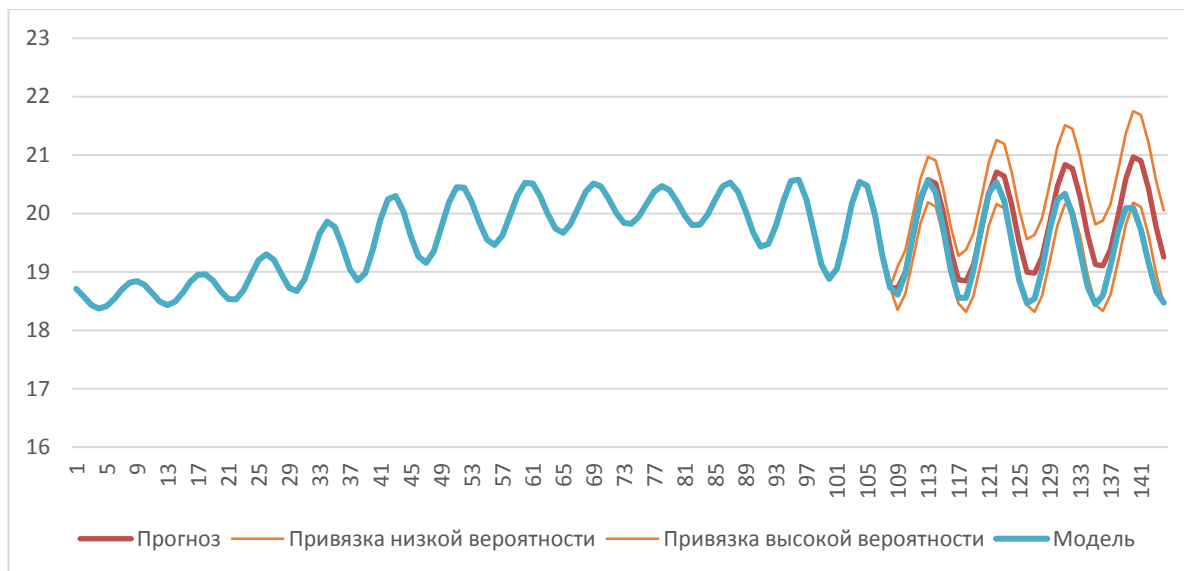


Рис.50 Диаграмма прогноза

Как видно из диаграммы, значения контрольного набора идут в пределах нижней границы доверительного интервала, точность прогноза вписывается в 95%.

Модель, полученную SSA возможно применять для прогнозирования.

4 РЕЗУЛЬТАТЫ ПРОВЕДЕННОГО ИССЛЕДОВАНИЯ

Исследование было проведено с целью по имеющемуся набору данных дать обоснованный ответ на вопрос применимости метода SSA. Сингулярный спектральный анализ демонстрирует наглядно структуру, выделяет составляющие ВР – тренд, периодичности и шум, при этом не имея изначально модели ряда. Метод позволяет получить модель, на основании которой можно делать прогноз, либо для аппроксимации функции с дальнейшим применением в апостериорном анализе. SSA позволяет определять разладку на ограниченном временном окне, и в случае с резким подъемом уровня – определять на начальных моментах.

Для ВР с дискретными состояниями метод применим только в качестве отражения тренда.

5 ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ

Цель раздела – комплексное описание и анализ финансово-экономических аспектов выполненной работы. В разделе «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение» произведена полная оценка всех денежных затрат на реализацию разработки с учетом материальных затрат, заработной платы участников работы, затрат на единый социальный налог и налог на добавленную стоимость, расходов на электроэнергию и прочих накладных расходов при выполнении проекта. Процесс создания работы разбит на этапы, определены сроки проведения отдельных работ, определены исполнители разработки, и их нагрузка по времени.

5.1 Потенциальные потребители результатов исследования

Анализ временных рядов имеет огромную значимость для различных отраслей и потребителей. Так в частности анализ временных данных технологических и производственных процессов на предприятии нефтегазовой отрасли позволяет совершенствовать систему планово-предупредительных работ (своевременная замена комплектующих и узлов оборудования, их ремонт) и точность соблюдения операций, что способствует повышению качества продукции. Потому такие компании как Газпром, Сибур имеют развитые аналитические отделы и подразделения ИТ для решения в том числе и задач, связанных с анализом временных данных, совершенствованием его средств и методов.

Таким образом целевым рынком результатов исследования можно назвать компании нефтегазовой отрасли (в т.ч. нефтесервисные и нефтехимические), а потребителями соответственно компании Роснефть, Газпром, Русснефть, Татнефть, Сургутнефтегаз, Сибур, Schlumberger, Weatherford. Следует все же учитывать исследовательский характер проекта и широкое применение метода SSA.

5.2 SWOT-анализ

SWOT – Strengths (сильные стороны), Weaknesses (слабые стороны), Opportunities (возможности) и Threats (угрозы) – представляет собой комплексный анализ научно-исследовательского проекта. SWOT-анализ применяют для исследования внешней и внутренней среды проекта с целью выбора оптимальной бизнес-стратегии.

	<u>Сильные стороны проекта:</u> С1. Метод можно применять к ВР нестационарного вида => большая применимость. С2. Доступность и популярность ЯП => много инф. Сообществ и справочного материала. С3. Низкие затраты на расходные материалы. С4. Низкая стоимость поддержки проекта.	<u>Слабые стороны проекта:</u> Сл1. Проект исследовательский => индивидуальный подход к каждой задаче. Сл2. Сложная предметная область.
<u>Возможности:</u> В1. Гибкость функционала проекта. В2. Перспектива реализации в виде модуля автоматизированной аналитической системы. В3. Популяризация => увеличение спроса на решение	С1С2В1В2В3В4 позволяет получить коммерческий продукт в виде встраиваемого модуля, либо отдельного ПО. С1С2С3С4В1В4 позволяет использовать исследовательскую перспективу проекта	Сл1Сл2В2В3 реализация проекта в виде модуля и популяризация с поддержкой пользователей нейтрализуют слабые стороны проекта
<u>Угрозы:</u> У1. Отсутствие явной коммерческой прибыли. У2. Снижение спроса У3. Вытеснение автоматизированным решением.	У1С1 при развитии функционала и удобства пользования коммерческий эффект станет явным У2С2 при развитии поддержки пользователей станет проще освоение метода У3С3 – неполное устранение, т.к. почти однозначно метод будет реализован как модуль сторонней автоматизированной системы	Сл1У3 – как взаимосвязанные факторы из-за автоматизации процессов, что может привести к дополнению факторами У1 и У2

Табл.3 SWOT-анализ

На основании Табл.1 выявлены две стратегии: С1С2В1В2В3В4 – развитие коммерческого продукта, что позволит преодолеть все возможные угрозы и снизить значимость слабых сторон; С1С2С3С4В1В4 – дальнейшее развитие исследовательского проекта для получения знаний и проверки

гипотез для лучшего понимания производственных и технологических процессов с целью их совершенствования.

5.3 Планирование научно-исследовательских работ

5.3.1 Структура в рамках научного исследования

Организация и планирование работ непосредственно связаны со всеми элементами финансового менеджмента.

Для организации и планирования, а также определения трудоемкости выполнения ВКР необходимо ее разбить на этапы, количество и содержание которых определяется спецификой темы. Этапом считается крупная часть работы, имеющая самостоятельное значение и является объемом планирования и финансирования.

Помимо определения перечня работ необходимых для реализации разработки, следует планировать занятость каждого из ее участников и рациональную продолжительность отдельных работ для эффективного использования ресурсов исполнителей.

Исполнителями проекта являются научный руководитель (НР) и студент-инженер (И). Описание и перечень этапов, исполнителей и их нагрузка занесены в Табл.2

№, i этапа	Этап работ	Должность и нагрузка
1	Постановка целей и задач, получение исходных данных	НР-100%
2	Разработка ТЗ и календарного плана	НР-100% И-10%
3	Изучение предметной области и подбор материалов	НР-10% И-100%
4	Изучение функционала и работа с базовым примером реализации метода	НР-10% И-100%
5	Применение метода к реальным данным, проверка и анализ результатов	НР-10% И-100%
6	Составление ПЗ, оформление материалов ВКР	И-100%

Табл.4 Этапы работ и их выполнение

5.3.2 Определение трудоемкости выполнения работ

Основная часть стоимости разработки проекта складывается из трудовых затрат, определяемых как трудоемкость работ каждого из участников проекта.

Трудоемкость выполнения проекта оценивается экспертным путем в человеко-днях и носит вероятностный характер, т.к. зависит от множества факторов, которые все в совокупности сложно учесть. Для определения ожидаемого значения трудоемкости используется среднее $t_{ожі}$:

$$t_{ожі} = \frac{3t_{\min i} + 2t_{\max i}}{5},$$

где:

$t_{\min}$ – минимальная трудоемкость работ, чел/дн.;

$t_{\max}$ – максимальная трудоемкость работ, чел/дн..

$$T_{pi} = \frac{t_{ожі}}{Ч_i},$$

T_{pi} – продолжительность одной работы, раб. дн.;

$t_{ожі}$ – ожидаемая трудоемкость выполнения одной работы, чел.-дн.

$Ч_i$ – численность исполнителей, выполняющих одновременно одну и ту же работу на данном этапе, чел..

Результаты расчетов будут отражены в Табл.3 в следующей главе.

5.3.3 Разработка графика проведения научного исследования

ВКР студента – сравнительно небольшой по объему научный проект. Для небольших проектов удобнее и нагляднее применять диаграмму Ганта для построения графика проведения научных работ.

Диаграмма Ганта – горизонтальный ленточный график, на котором работы по теме представляются протяженными во времени отрезками, характеризующимися датами начала и окончания выполнения данных работ.

Для удобства построения графика, длительность каждого из этапов работ из рабочих дней следует перевести в календарные дни. Для этого необходимо воспользоваться следующей формулой:

$$T_{ki} = T_{pi} \cdot k_{\text{кал}} ,$$

где:

T_{ki} – продолжительность выполнения i -й работы в календарных днях;

T_{pi} – продолжительность выполнения i -й работы в рабочих днях;

$k_{\text{кал}}$ – коэффициент календарности.

Коэффициент календарности определяется по следующей формуле:

$$k_{\text{кал}} = \frac{T_{\text{кал}}}{T_{\text{кал}} - T_{\text{вых}} - T_{\text{пр}}} = \frac{365}{365 - 52 - 14} = 1,22 ,$$

где:

$T_{\text{кал}}$ – количество календарных дней в году (365 дней);

$T_{\text{вых}}$ – количество выходных дней в году (52 дней);

$T_{\text{пр}}$ – количество праздничных дней в году (14 дней).

Рассчитанные значения в календарных днях по каждой работе T_{ki} округляются до целого числа. Эти значения будут отражены в Табл.3

Этап работ	Трудоемкость работ			Исполнители	Длительность работ в рабочих днях, T_{pi}		Длительность работ в календарных днях, T_{ki}	
	$t_{\min}$	$t_{\max}$	$t_{\text{ож}}$		НР	И	НР	И
Постановка целей и задач, получение исходных данных	1	3	1,8	НР	1,8	0	3	0
Разработка ТЗ и календарного плана	2	4	2,8	НР, И	2,8	0,28	4	1
Изучение предметной области и подбор материалов	9	20	13,4	НР, И	1,34	13,4	2	16
Изучение функционала и работа с базовым примером реализации метода	13	24	17,4	НР, И	1,74	17,4	3	21
Применение метода к реальным данным, проверка и анализ результатов	12	20	15,2	НР, И	1,52	15,2	2	19
Составление ПЗ, оформление материалов ВКР	3	7	4,6	И	0	4,6	0	6
Итого:	40	78	55,2		9,2	50,88	14	63

Табл.5 Трудозатраты на выполнение проекта

№ работ	Вид работ	Исполнители	$t_{\max}$, кал. дн.	Продолжительность выполнения работ											
				февраль			март			апрель			май		
				1	2	3	1	2	3	1	2	3	1	2	3
1	Постановка целей и задач, получение исходных данных	НР	3												
2	Разработка ТЗ и календарного плана	НР И	4	 											
3	Изучение предметной области и подбор материалов	НР И	20		 										
4	Изучение функционала и работа с базовым примером реализации метода	НР И	24												
5	Применение метода к реальным данным, проверка и анализ результатов	НР И	20									 			
6	Составление ПЗ, оформление материалов ВКР	И	7												

Табл.6 Линейный график работ:  – научный руководитель,  – исполнитель (студент-инженер),  – нерабочее время (сессия, праздники)

5.4 Бюджет научно технического исследования (НТИ)

При планировании бюджета НТИ должно быть обеспечено полное и достоверное отражение всех видов расходов, связанных с его выполнением. В процессе формирования бюджета НТИ используется следующая группировка затрат по статьям:

- материальные затраты НТИ;
- затраты на специальное оборудование для научных (экспериментальных) работ*;
- основная заработная плата исполнителей темы;
- дополнительная заработная плата исполнителей темы;
- отчисления во внебюджетные фонды (страховые отчисления);
- затраты научные и производственные командировки*;
- контрагентные расходы;
- накладные расходы.

**в данном проекте затраты на статью не требуются.*

5.4.1 Расчет материальных затрат

Данная статья расходов материальных затрат учитывает стоимость материалов, покупных изделий, а также специально приобретенное оборудование, инструменты и другие материальных ценностей, используемые непосредственно в процессе выполнения ВКР. Кроме того, статья включает транспортно-заготовительные расходы, связанные с транспортировкой от поставщика к потребителю, хранением и прочими процессами, обеспечивающими движение (доставку) материальных ресурсов от поставщиков к потребителю.

Расчет материальных затрат осуществляется по следующей формуле:

$$З_m = (1 + k_T) \sum_{i=1}^m Ц_i \cdot N_{расхi}$$

где:

m – количество видов материальных ресурсов, потребляемых при выполнении научного исследования;

$N_{расхi}$ – количество материальных ресурсов i -го вида, планируемых к использованию при выполнении научного исследования (шт., кг, м, м² и т.д.);

C_i – цена приобретения единицы i -го вида потребляемых материальных ресурсов (руб./шт., руб./кг, руб./м, руб./м² и т.д.);

k_T – коэффициент, учитывающий транспортно-заготовительные расходы.

Допустим k_T будет составлять 15% от стоимости материалов.

Одним из преимуществ проекта является то, что ПО, требуемое для реализации бесплатное, поэтому учитывать его в материальных затратах не нужно.

Наименование	Единица измерения	Количество	Цена за ед., руб.	Затраты на материалы, (З _м), руб.
Бумага «Снегурочка» (А4, 80г/кв.м, 500 листов)	1	1	255	255
Заправка картриджа принтера	1	1	400	400
Комплект канцелярских товаров	1	1	210	210
Итого:				865

Табл.7 Расчет затрат на материалы

С учетом $k_T = 0,15$ расходы на материалы составят: $З_m = 996$ руб.

5.4.2 Основная заработная плата исполнителей темы

Данная статья расходов включает заработную плату научного руководителя и студента-инженера, а также премии, входящие в фонд заработной платы. Расчет основной заработной платы выполняется на основе трудоемкости выполнения каждого этапа и величины месячного оклада исполнителя. Величины месячных окладов для сотрудников ТПУ взяты регламентирующих документов с официального сайта ТПУ.

$$З_{зп} = З_{осн} + З_{доп} ,$$

где:

$З_{осн}$ – основная заработная плата;

$З_{доп}$ – дополнительная заработная плата (12-20 % от $З_{осн}$).

Основная заработная плата работника рассчитывается по следующей формуле:

$$З_{осн} = З_{дн} \cdot T_p ,$$

где:

$З_{дн}$ – среднедневная заработная плата работника, руб.

Среднедневная заработная плата рассчитывается по формуле:

$$З_{дн} = \frac{З_{м} \cdot M}{F_{дн}}$$

где:

$З_{м}$ – месячный должностной оклад работника, руб.;

M – количество месяцев работы без отпуска в течение года: при отпуске в 28 раб. дня $M = 11,2$ месяца, 5-дневная неделя; при отпуске в 56 раб. дней $M = 10,4$ месяца, 6-дневная неделя;

$F_{д}$ – действительный годовой фонд рабочего времени научно-технического персонала, раб. дн..

Показатели рабочего времени	Руководитель	Инженер
Календарное число дней	365	365
Количество нерабочих дней:		
- выходные дни	44	48
- праздничные дни	14	14
Потери рабочего времени:		
- отпуск	56	28
- невыходы по болезни	1	1
Действительный годовой фон рабочего времени	250	274

Табл.8 Баланс рабочего времени

Месячный должностной оклад работника:

$$З_{м} = З_{тс} \cdot (1 + k_{пр} + k_{д}) \cdot k_p ,$$

где:

$Z_{\text{ТС}}$ – заработная плата по тарифной ставке, руб.;

$k_{\text{пр}}$ – премиальный коэффициент, равный 0,3 (т.е. 30% от $Z_{\text{ТС}}$);

$k_{\text{д}}$ – коэффициент доплат и надбавок составляет примерно 0,2 – 0,5;

$k_{\text{р}}$ – районный коэффициент, равный 1,3 (для Томска).

Работники	$Z_{\text{ТС}}$, руб.	$k_{\text{пр}}$	$k_{\text{д}}$	$k_{\text{р}}$	$Z_{\text{м}}$, руб.	$Z_{\text{дн}}$, руб.	$T_{\text{р}}$, раб.дн.	$Z_{\text{осн}}$, руб.
Руководитель	33 664	0,3	0,4	1,3	74 397	3 095	9,2	28 474
Инженер	26 300	0,3	0,2	1,3	34 190	1 398	50,9	71 158
Итого:								99 632

Табл.9 Расчет основной заработной платы

5.4.3 Дополнительная заработная плата исполнителей темы

Расчет дополнительной заработной платы ведется по следующей формуле:

$$Z_{\text{доп}} = k_{\text{доп}} \cdot Z_{\text{осн}},$$

где:

$k_{\text{доп}}$ – коэффициент дополнительной заработной платы (на стадии проектирования принимается равным 0,12 – 0,15).

Допустим, $k_{\text{доп}} = 0,15$, тогда $Z_{\text{доп}}$ руководителя составит 4 271 руб., и $Z_{\text{доп}}$ инженера составит 10 674 руб..

5.4.4 Отчисления во внебюджетные фонды (страховые отчисления)

В данную статью входят отчисления в пенсионный фонд, фонд обязательного медицинского страхования, единый социальный налог и пр.

Величина отчислений во внебюджетные фонды определяется исходя из следующей формулы:

$$Z_{\text{внеб}} = k_{\text{внеб}} \cdot (Z_{\text{осн}} + Z_{\text{доп}}),$$

где:

$k_{\text{внеб}}$ – коэффициент отчислений на уплату во внебюджетные фонды (пенсионный фонд, фонд обязательного медицинского страхования и пр.).

На основании пункта 1 ст.58 закона №212-ФЗ для учреждений осуществляющих образовательную и научную деятельность в 2014 году

водится пониженная ставка – 27,1% Таким образом размер отчислений во внебюджетные фонды составит 31 051 руб..

5.4.5 Накладные расходы

Накладные расходы учитывают прочие затраты организации, не попавшие в предыдущие статьи расходов: печать и ксерокопирование материалов исследования, оплата услуг связи, электроэнергии, почтовые и телеграфные расходы, размножение материалов и т.д. Их величина определяется по следующей формуле:

$$З_{\text{накл}} = (\text{сумма статей подпунктов 1 – 4}) \cdot k_{\text{нр}},$$

где:

$k_{\text{нр}}$ – коэффициент, учитывающий накладные расходы.

Допустим коэффициент накладных расходов равен 16%, тогда накладные расходы составят: 23 970 руб..

5.4.6 Формирование бюджета затрат научно-исследовательского проекта

В данном подпункте идет суммирование всех описанных статей затрат для определения бюджета проекта.

Наименование статьи	Сумма, руб.	Примечание
1. Материальные затраты НТИ	996	3.4.1
2. Затраты по основной заработной плате	99 632	3.4.2
3. Затраты по дополнительной заработной плате	14 945	3.4.3
4. Отчисления во внебюджетные фонды	34 373	3.4.4
5. Накладные расходы	23 970	3.4.5
Итого:	170 594	

Табл.10 Расчет бюджета проекта

5.5 Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования

Эффективность исследования во многом определяет его целесообразность. Показатели эффективности дают более детальную картину при наличии конкретных показателей. В данном проекте рассматривается

именно эффективность алгоритма с точки зрения представления неявных знаний о процессе, краткосрочного прогнозирования для оперативных решений. Проект имеет положительный эффект для ресурсной (ресурсосберегающей) и экономической составляющих, который возможно конкретно оценить при наличии дополнительной информации со стороны предприятия и внедрения решения, потому ниже будет определена общая сравнительная оценка эффекта, которая рассчитана аналогично интегральному показателю ресурсоэффективности:

$$I_i = \sum a_i b_i$$

где:

a_i – весовой коэффициент i -го критерия;

b_i – балльная оценка критерия.

Критерии	Весовой коэффициент	Перспективная система	Исходная система
Проще и точнее моделирование технологического процесса (ТП)	0,25	9	7
Сокращение продолжительности простоя в связи с заменой технологического оборудования	0,25	8	6
Ведение ТП в оптимальном режиме	0,2	7	6
Перспектива создания аналитического ПО для ТП	0,3	7	4
Итого:	1	7,75	5,65

Табл.11 Расчет показателя эффективности

Вывод

В ходе выполнения раздела были обозначены целевой рынок и потребители для данного проекта, выработаны стратегические решения для развития с учетом угроз и слабых сторон. Были определены этапы, график работ и подсчитана трудоемкость. Максимальное время выполнения составило 73 календарных дня (с учетом нерабочего времени – 115 дней). Определены и рассчитаны статьи затрат и расходов, бюджет проекта составил 170 594 руб. Выделены критерии и рассчитаны показатели эффективности.

6 СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ

6.1 Специальные правовые требования для рабочей зоны оператора ПК

Существующая на сегодняшний день редакция Трудового кодекса не приравнивает работу за компьютером к вредным условиям труда. Специфика рабочего места оператора ПК определена в ТОО Р-45-084-01 и СанПиН 2.2.2/2.4.1340-03. Так типовой инструкцией ТОО Р-45-084-01 установлено, что при работе с компьютером, пользователь обязан делать перерыв в размере 15 минут каждый час непрерывной работы. СанПиН 2.2.2/2.4.1340-03 устанавливает следующие требования для работников: к работе с ПК допускаются работники, не имеющие медицинских противопоказаний; женщины в период беременности, кормления грудью к выполнению всех видов работ, с использованием компьютера, не допускаются.

6.2 Организационные мероприятия при компоновке рабочей зоны.

Предъявляемые требования к расположению и компоновке рабочего места: «Высота рабочей поверхности стола для взрослых пользователей должна регулироваться в пределах 680÷800 мм, при отсутствии такой возможности высота рабочей поверхности стола должна составлять 725 мм» (по СанПиН 2.2.2/2.4.1340-03).

Рабочий стол должен иметь пространство для ног высотой не менее 600 мм, шириной – не менее 500 мм, глубиной на уровне колен – не менее 450 мм и на уровне вытянутых ног – не менее 650 мм (по СанПиН 2.2.2/2.4.1340-03).

Конструкция рабочего стула должна обеспечивать (по СанПиН 2.2.2/2.4.1340-03):

- ширину и глубину поверхности сиденья не менее 400 мм;
- поверхность сиденья с закругленным передним краем; регулировку высоты поверхности сиденья в пределах 400÷550 мм и углам наклона вперед до 15 град., и назад до 5 град.;
- высоту опорной поверхности спинки 300 ± 20 мм, ширину – не менее

- 380 мм и радиус кривизны горизонтальной плоскости – 400 мм;
- угол наклона спинки в вертикальной плоскости в пределах ± 30 градусов;
- регулировку расстояния спинки от переднего края сиденья в пределах 260÷400 мм;
- стационарные или съемные подлокотники длиной не менее 250мм и шириной – 50÷70 мм;
- регулировку подлокотников по высоте над сиденьем в пределах (230±30) мм и внутреннего расстояния между подлокотниками в пределах (350÷500) мм.

Рабочее место пользователя ПК следует оборудовать подставкой для ног, имеющей ширину не менее 300 мм, глубину не менее 400 мм, регулировку по высоте в пределах до 150 мм и по углу наклона опорной поверхности подставки до 20°. Поверхность подставки должна быть рифленой и иметь по переднему краю бортик высотой 10 мм (по СанПиН 2.2.2/2.4.1340-03).

6.3 Профессиональная социальная безопасность

Основная деятельность при разработке проекта и при пользовании им связана с работой с ПК, где можно выделить такие особенности: нагрузка на глаза, большой объем обрабатываемой информации, малоподвижность. Также ПК может стать причиной поражения электрическим током, например, при отсутствии защитного заземления. Также рабочее место оператора ПК должно иметь возможность искусственного освещения и поддержки циркуляции воздуха. В противном случае это так же может стать вредными факторами.

Согласно ГОСТ 12.0.003-2015 «Опасные и вредные производственные факторы. Классификация» были выделены вредные и опасные факторы, характерные для оператора ПК (Табл.10).

Факторы (ГОСТ 12.0.003-2015)	Этапы работ		Нормативные документы
	Исследование	Эксплуатация	
1. Отклонение показателей микроклимата	+	+	СанПиН 2.2.4.548-96
2. Электромагнитное излучение экрана монитора ПК	+	+	СанПиН 2.2.2/2.4.1340-03 ТСО'03
3. Недостаточная освещенность рабочей зоны	+	+	СП 52.13330.2016 (актуализированная редакция СНиП 23-05-95) СанПиН 2.2.1/2.1.1.1278-03
4. Перенапряжение зрительного анализатора	+	+	ТОИ Р-45-084-01
5. Опасность поражения электрическим током	+	+	ГОСТ 12.1.030-81 ССБТ ТОИ Р-45-084-01

Табл.120 Вредные и опасные факторы исследования

6.4 Отклонение показателей микроклимата в помещении

Показатели микроклимата должны обеспечивать сохранение теплового баланса человека с окружающей средой и поддержание оптимального или допустимого теплового состояния организма. Показателями, характеризующими микроклимат в производственных помещениях, являются: температура воздуха; температура поверхностей; относительная влажность воздуха; скорость движения воздуха; интенсивность теплового облучения.

В производственных помещениях для работы с ПК происходит постоянное выделение тепла самой вычислительной техникой, вспомогательными приборами и средствами освещения. Поскольку оператор расположен в непосредственной близости с источниками выделения тепла, то данный фактор (в случае плохой вентиляции и кондиционирования), является одним из важнейших вредных факторов производственной среды оператора ПК, а высокая температура воздуха способствует быстрому перегреву организма и быстрой утомляемости.

Влажность оказывает большое влияние на терморегуляцию организма. Так, например, высокие показатели относительной влажности (более 85 %) затрудняют терморегуляцию снижая возможность испарения пота, низкие показатели влажности (менее 20 %) вызывают пересыхание слизистых оболочек человека.

Работа на компьютере относится к категории Ia – работы с интенсивностью энергозатрат до 120 ккал/ч (до 139 Вт), производимые сидя и сопровождающиеся незначительным физическим напряжением.

В производственных помещениях, в которых работа на компьютере является основной, обеспечиваются оптимальные параметры микроклимата (Табл.11)

Период года	Температура воздуха, °C	Температура поверхностей, °C	Относительная влажность воздуха, %	Скорость движения воздуха, м/с
Холодный	22-24	21-25	40-60	не более 0,1
Теплый	23-25	22-26	40-60	не более 0,1

Табл.11 Оптимальные величины показателей микроклимата для категории работ Ia (по СанПиН 2.2.4.548-96)

Допустимые микроклиматические условия (Табл.12) установлены по критериям допустимого теплового и функционального состояния человека на период 8-часовой рабочей смены. Они не вызывают повреждений или нарушений состояния здоровья, но могут приводить к возникновению общих и локальных ощущений теплового дискомфорта, напряжению механизмов терморегуляции, ухудшению самочувствия и понижению работоспособности.

Допустимые величины показателей микроклимата устанавливаются в случаях, когда по технологическим требованиям, техническим и экономически обоснованным причинам не могут быть обеспечены оптимальные величины.

Период года	Температура воздуха, °С (диап. отн. оптим. величин)		Температура поверхностей, °С	Относительная влажность воздуха, %	Скорость движения воздуха, м/с (для диап. от оптим. величин)	
	выше	ниже			ниже не более	выше не более
Холодный	20,0-21,9	24,1-25,0	19,0-26,0	15-75*	0,1	0,1
Теплый	21,0-22,9	25,1-28,0	20,0-29,0	15-75*	0,1	0,2

Табл.12 Допустимые величины показателей микроклимата для категории работ Ia (по СанПиН 2.2.4.548-96)

**При температурах воздуха 25°C и выше максимальные величины относительной влажности воздуха должны соответствовать: 70% - 25°C; 65% - 26°C; 60% - 27°C; 55% - 28°C.*

***При температурах воздуха 26-28 °C скорость движения воздуха в теплый период года должна соответствовать 0,1-0,2 м/с*

К мероприятиям по оздоровлению воздушной среды в производственном помещении относятся правильная организация вентиляции и кондиционирования воздуха, отопление помещений. Вентиляция может осуществляться естественным и механическим путём. В зимнее время в помещении необходимо предусмотреть систему отопления. Она должна обеспечивать достаточное, постоянное и равномерное нагревание воздуха. В помещениях с повышенными требованиями к чистоте воздуха должно использоваться водяное отопление.

6.5 Электромагнитное излучение экрана монитора

Персональный компьютер является основным источником повышенных электромагнитных излучений. Они усиливают риск возникновения сердечно-сосудистых заболеваний, а также негативно влияют на нервную систему.

Для защиты пользователя компьютера от данного вида негативного излучения необходимо обеспечить оптимальное расстояние между монитором и работником, равное минимум 500 мм, установленное СанПиН 2.2.2/2.4.1340-03.

Временные допустимые уровни (ВДУ) излучения электромагнитных полей (ЭМП), создаваемых ПК на рабочих местах операторов:

Наименование параметров		ВДУ
Напряженность электрического поля	в диапазоне частот 5 Гц - 2 кГц	25 В/м
	в диапазоне частот 2 кГц - 400 кГц	2,5 В/м
Плотность магнитного потока	в диапазоне частот 5 Гц - 2 кГц	250 нТл
	в диапазоне частот 2 кГц - 400 кГц	25 нТл
Напряженность электростатического поля		15 кВ/м

Табл.13 Нормы ЭМП (по СанПиН 2.2.2/2.4.1340-03)

Последние достижения науки в области производства мониторов позволяют значительно снизить уровень излучения ЭМП. Использование современных ЖК- и LED-мониторов, соответствующих стандарту TCO-2003, гарантирует минимальные значения напряженности электромагнитных полей вблизи экранов (менее 10 В/м в диапазоне частот от 5 Гц до 2 кГц и менее 1 В/м в диапазоне частот 2÷400 кГц).

6.6 Недостаточная освещенность рабочей зоны

Работа с компьютером подразумевает постоянный зрительный контакт с дисплеем ПК и занимает от 80 % рабочего времени. Недостаточность освещения снижает производительность труда, увеличивает утомляемость и количество допускаемых ошибок, а также может привести к появлению профессиональных болезней зрения.

Источником недостаточной освещенности рабочей зоны является недостаток количества естественного и искусственного света, суммарная

норма которых должна составлять 300-500 люкс в соответствии с СП 52.13330.2016.

Недостаток естественного освещения обусловлен некорректным расположением рабочего места, поскольку свет, в соответствии с нормативными документами, должен освещать рабочую зону слева, что не учтено при его организации. Поэтому для минимизации данного воздействия рабочее место должно быть оснащено дополнительными источниками искусственного света.

Светильник освещения должны обеспечивать нормальные условия освещенности с учетом специфики работы, необходимых требований видимости

со стороны работника. В соответствии с СанПиН 2.2.1/2.1.1.1278-03 на рабочем месте желательно применение комбинированной системы освещения: люминесцентные лампы типа ЛД. Люминесцентные лампы имеют ряд существенных преимуществ, таких как: излучаемый ими свет близок к дневному, естественному свету; обладают повышенной светоотдачей и имеют более длительный срок службы.

6.7 Перенапряжение зрительного анализатора

Перенапряжение зрительного анализатора может быть вызвано продолжительной работой за компьютером, яркостью естественного света, попадающего на рабочую поверхность. Данный фактор может привести к снижению остроты зрения пользователя, развитию у него близорукости.

Защитой от данного вредного фактора могут быть зрительная гимнастика, а также обеспечение безопасного расстояния между экраном монитора и пользователем равным 600-700 мм в соответствии с нормативным документом ТОО Р-45-084-01. Также на рабочем месте должны применяться регулируемые жалюзи и шторы, которые в поле зрения

работника снижают яркость при естественном освещении, что определено инструкцией по охране труда при работе на персональном компьютере.

6.8 Опасность поражения электрическим током

Поражение электрическим током является опасным производственным фактором и, поскольку оператор ПК имеет дело с электрооборудованием, то вопросам электробезопасности на его рабочем месте должно уделяться внимание. Нормы электробезопасности на рабочем месте регламентируются ТОО Р-45-084-01, где определены и требования к защите от поражения электрическим током, также – ГОСТ 12.1.030-81 ССБТ.

Электробезопасность – система организационных и технических мероприятий и средств, обеспечивающих защиту людей от вредного и опасного воздействия электрического тока, электрической дуги, электромагнитного поля и статического электричества.

Опасность поражения электрическим током усугубляется тем, что человек не в состоянии без специальных приборов обнаружить напряжение дистанционно.

Помещение, где расположено рабочее место оператора ПК, относится к помещениям без повышенной опасности ввиду отсутствия следующих факторов: сырость, токопроводящая пыль, токопроводящие полы, высокая температура, возможность одновременного прикосновения человека к имеющим соединение с землёй металлоконструкциям зданий, технологическим аппаратам, механизмам и металлическим корпусам электрооборудования.

Для оператора ПК при работе с электрическим оборудованием обязательны следующие меры предосторожности:

- перед началом работы нужно убедиться, что выключатели и розетка закреплены и не имеют оголённых токоведущих частей;

- при обнаружении неисправности оборудования и приборов необходимо, не делая никаких самостоятельных исправлений, сообщить человеку, ответственному за оборудование.

6.9 Экологическая безопасность

Исследование статистического метода и работа за ПК не являются экологически опасными работами, потому что объект, на котором производилась разработка продукта, а также объекты, на которых будет производиться его использование операторами ПК относятся к предприятиям пятого класса, размер санитарной зоны для которых равен 50 м.

Аппаратные средства, необходимые для разработки и эксплуатации программного комплекса, могут наносить вред окружающей среде. Современные ПК производят практически без использования вредных веществ, опасных для человека и окружающей среды.

Люминесцентные лампы, применяемые для искусственного освещения рабочих мест, требуют утилизации, т.к. в них присутствует от 10 до 70 мг ртути, которая относится к чрезвычайно опасным химическим веществам и может стать причиной отравления живых существ, а также загрязнения атмосферы, гидросферы и литосферы. Сроки службы таких ламп составляют около 5-ти лет, после чего их необходимо сдавать на переработку в специальных пунктах приема. Юридические лица обязаны сдавать лампы на переработку и вести паспорт для данного вида отходов. В соответствии с ГОСТ Р 52105-2003 разработаны методы переработки таких отходов в зависимости от их типа, которые представлены в таблице ниже (Табл.14).

Наименование	Обозначение отходов по ГОСТ 30775	Группа	Подгруппа	Позиция	Метод обработки
Бракованные люминесцентные лампы	Q2, Q3, Q6	16	1	3	ХМ, ТМ, ВО
Ступа ртутьсодержащая от переработки люминесцентных ламп	Q9	19	9	99	ТМ

Люминофор, ртутьсодержащий от переработки люминесцентных ламп	Q9	20	3	10	TM, BO
---	----	----	---	----	--------

Табл.14 Классификация и методы переработки ртутьсодержащих отходов по ГОСТ 52105-2003

6.10 Безопасность в чрезвычайных ситуациях

Среди возможных ЧС в работе оператором ПК наибольшую вероятность имеет возникновение пожара, которые могут быть вызваны коротким замыканием, перегрузкой, нагреванием участков электросети, неисправностью оборудования и несоблюдения правил пожарной безопасности.

К пожарно-профилактическим мероприятиям можно отнести: правильный выбор оборудования и способы его монтажа, систематический контроль исправности защитных устройств на электрооборудовании, создание условий, обеспечивающих пожарную безопасность.

С целью оповещения о возникшей чрезвычайной ситуации учебный корпус №10 НИ ТПУ оборудован речевой системой оповещений, а также на каждом этаже вывешен план эвакуации.

В случае с предприятиями-потребителями проектного решения соблюдение правил пожарной безопасности ведется на основе документации, разработанной отделом ОТиТБ предприятия.

6.11 Общий порядок действий при возникновении пожара

Каждый сотрудник организации ознакомлен с инструкцией по пожарной безопасности, проходит инструктаж по технике безопасности и обязан строго соблюдать его.

Запрещается использовать электроприборы в условиях, не соответствующих требованиям инструкций изготовителей, или имеющие неисправности, которые в соответствии с инструкцией по эксплуатации

могут привести к пожару, а также эксплуатировать электропровода и кабели с повреждённой или потерявшей защитные свойства изоляцией.

Перед уходом из служебного помещения работник обязан провести его осмотр, закрыть окна, и убедиться в том, что в помещении отсутствуют источники возможного возгорания, все электроприборы отключены и выключено освещение. С периодичностью не реже одного раза в три года проводятся замеры сопротивления изоляции токоведущих частей силового и осветительного оборудования.

Повышение устойчивости достигается за счёт проведения соответствующих организационно-технических мероприятий, подготовки персонала к работе в ЧС.

Работник при обнаружении пожара или признаков горения должен:

- немедленно прекратить работу, воспользоваться ближайшим ручным извещателем пожара и вызвать службу МЧС по телефону «101», сообщив при этом адрес, место возникновения пожара и свою фамилию;
- принять по возможности меры по эвакуации людей и материальных ценностей;
- отключить от сети закреплённое за ним электрооборудование;
- приступить к тушению пожара имеющимися средствами пожаротушения;
- сообщить непосредственному или вышестоящему начальнику и оповестить окружающих сотрудников.

При общем сигнале опасности покинуть здание согласно «Плану эвакуации людей при пожаре и других ЧС»

Для тушения пожара применять ручные углекислотные огнетушители (типа ОУ-2, ОУ-5), находящиеся в помещениях офиса, и пожарный кран внутреннего противопожарного водопровода. Они предназначены для

тушения начальных возгораний различных веществ и материалов, за исключением веществ, горение которых происходит без доступа воздуха. Огнетушители должны постоянно содержаться в исправном состоянии и быть готовыми к действию. Категорически запрещается тушить возгорания в помещениях офиса при помощи химических пенных огнетушителей (типа ОХП-10).

Вывод

В ходе выполнения раздела были определены специфичные для исследования и его пользования правовые и организационные особенности рабочего места, выделены вредные и опасные факторы воздействия на работника, возможные ЧС и описаны меры предупреждения или борьбы с ними. Также были описаны меры по утилизации вредных отходов, образуемых в ходе эксплуатации оборудования рабочего места в целях соблюдения экологической безопасности. Непосредственно само исследование и его применение не несут экологической опасности.

ЗАКЛЮЧЕНИЕ

В результате выполнения ВКР был проведен обзор информации по методу SSA и его программной реализации, определены инструменты для исследования. Был освоен базовый метод SSA, его реализация на ЯП Python, представлено наглядное описание метода, определены оптимальные параметры для анализа и построения моделей ВР, выработана методика их анализа, используемое ПО было дополнено функциями и методами для проведения анализа. С помощью SSA были проанализированные данные нефтегазовой отрасли, определена и продемонстрирована применимость метода – построены модели, описана структура, определена разладка, сделан прогноз.

В разделе «Финансовый менеджмент» были определены потребители метода и его импликации, выделены стратегии дальнейшего развития проекта.

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

1. Колесов В.В. Элементарное введение в высшую математику: учебное пособие / В.В. Колесов, М.Н. Романов. – Ростов н/Д: Феникс, 2013. – 476с.
2. Письменный, Д.Т. Конспект лекций по высшей математике: полный курс / Д.Т. Письменный. - 4-е изд. - М.: Айрис-пресс, 2006.-608 с.
3. Анализ данных: учебник для академического бакалавриата / под. Ред. В.С. Мхитаряна. – М.: Издательство Юрайт, 2016 – 490 с.
4. Бендат Дж., Пирсол А. Прикладной анализ случайных данных: пер. с англ. М.: Мир, 1989. – 540 с.
5. Бокс Дж. Анализ временных рядов. Прогноз и управление / Дж. Бокс, Г. Дженкинс. – М.: Мир, 1974 – Вып. 1,2.
6. Моделирование динамических процессов по временным рядам, [Электронный ресурс], режим доступа:
https://www.researchgate.net/publication/315665081_Modelirovanie_dinamicheskikh_processov_po_vremennym_radam, свободный, дата обращения [01.06.2019]
7. Голяндина Н.Э. Метод "Гусеница"-SSA: анализ временных рядов: Учеб. пособие – СПб., 2004 – 76 с.
8. Golyandina N. , Nekrutkin V. , and Zhigljavsky A . Analysis of Time Series Structure: SSA and Related Techniques. London: Chapman & Hall/CRC , 2001 . 305 p.
9. Стрижов В. В. Сингулярное разложение: конспект лекций – М., 2010 – 6 с.
10. Introducing SSA for Time Series Decomposition, [Электронный ресурс], режим доступа:
<https://www.kaggle.com/jdarcy/introducing-ssa-for-time-series-decomposition/notebook>, свободный, дата обращения [01.06.2019]
11. NumPy v1.16 Manual, [Электронный ресурс], режим доступа:
<https://docs.scipy.org/doc/numpy/>, свободный, дата обращения [01.06.2019]
12. matplotlib Tutorial, [Электронный ресурс], режим доступа:
13. <https://matplotlib.org/tutorials/>, свободный, дата обращения [01.06.2019]
14. Python Data Analysis Library, [Электронный ресурс], режим доступа:
<https://pandas.pydata.org/>, свободный, дата обращения [01.06.2019]
15. Кудинов А.В. Проблемы автоматизации производства газодобывающих компаний: монография / А.В. Кудинов, Н.Г. Марков. – Томск: Изд-во ТПУ, 2012. – 247 с.
16. Relationship between Singular Spectrum Analysis and Fourier analysis: Theory and application to the monitoring of volcanic activity, [Электронный ресурс], режим доступа:
<https://www.sciencedirect.com/science/article/pii/S0898122110003858>, свободный [01.06.2019]

17. Применение метода SSA для анализа технологической информации, собираемой информационным комплексом ТЭЦ, [Электронный ресурс], режим доступа: <http://earchive.tpu.ru/handle/11683/2453>, свободный [01.06.2019]
18. Применение метода сингулярно-спектрального анализа для идентификации сигналов электрической активности мозга, [Электронный ресурс], режим доступа: <http://www.vestnik.vsu.ru/pdf/analiz/2012/02/2012-02-07.pdf>, свободный [01.06.2019]
19. СанПиН 2.2.2/2.4.1340-03 "Гигиенические требования к персональным электронно-вычислительным машинам и организации работы" (с изменениями на 21 июня 2016 года).
20. ТСО'03 Стандарт добровольной сертификации на эргономичность и безопасность электронного оборудования, [электронный ресурс, 02.05.19г.], режим доступа: <http://poradme.se/images/d/d7/Tco03.pdf>
21. СанПиН 2.2.4.548-96 Гигиенические требования к микроклимату производственных помещений.
22. СП 52.13330.2016 Естественное и искусственное освещение. (Актуализированная редакция СНиП 23-05-95).
23. ГОСТ Р 12.1.019-2009 Система стандартов безопасности труда (ССБТ). Электробезопасность. Общие требования и номенклатура видов защиты
24. ГОСТ Р 52105-2003 Ресурсосбережение. Обращение с отходами. Классификация и методы переработки ртутьсодержащих отходов. Основные положения.
25. Матрюков, Борис Степанович. Безопасность в чрезвычайных ситуациях в природно-техногенной сфере. Прогнозирование последствий: учебное пособие / Б. С. Матрюков. — Москва: Академия, 2011. — 368 с.: ил. — Высшее профессиональное образование. Безопасность жизнедеятельности. — Библиогр.: с. 364-365.
26. СП 9.13130.2009 Техника пожарная. Огнетушители. Требования к эксплуатации, [электронный ресурс, 02.05.19г.], режим доступа: <http://polyset.ru/GOST/all-doc/SP/SP-9-13130-2009/>

ПРИЛОЖЕНИЕ А

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
plt.rcParams['figure.figsize'] = (10,8) # размеры рисунка
plt.rcParams['font.size'] = 14 # размер шрифта
plt.rcParams['image.cmap'] = 'plasma' # палитра plasma (градиент син-фиол-ор-
желт) для раскраски изображений (для матриц BP)
plt.rcParams['axes.linewidth'] = 2 # ширина рамки графика

from cyciler import cyciler # cyciler задает ротацию значений параметров, таких
как цвет, тип линии
cols = plt.get_cmap('tab10').colors
cols20 = plt.get_cmap('tab20').colors # cols принимает цвета палитры tab20 (в
наборе 20 цветов)
plt.rcParams['axes.prop_cycle'] = cyciler(color=cols)

#основа импликации https://www.kaggle.com/jdarcy/introducing-ssa-for-time-series-decomposition/notebook
class SSA(object):

    __supported_types = (pd.Series, np.ndarray, list)

    def __init__(self, tseries, L, save_mem=True):
        """
        Разделяет данный BP SVD-разложением. Предполагает, что наблюдения BP
        записаны через равные промежутки времени.

        Параметры
        -----
        tseries: Исходный BP в виде Pandas-ряда, NumPy-массива либо обычного
        списка.
        L: Ширина окна, некоторое целое число,  $2 \leq L \leq N/2$ , где N длина BP.
        save_mem: Conserve memory by not retaining the elementary matrices.
        Рекомендовано для больших BP с тысячами наблюдений. Исходное
        значение: True.

        Примечание: Даже если BP был записан в виде NumPy-массива или списка,
        он будет преобразован в объект структуры данных Pandas - Series или
        DataFrame.
        """

        # Проверка BP на поддерживаемый тип данных
        if not isinstance(tseries, self.__supported_types):
            raise TypeError("Неподдерживаемый тип данных. Попробуйте Pandas
            Series, NumPy array или list.")

        # Проверка ширины окна L на соответствие условию
        self.N = len(tseries)
        if not 2 <= L <= self.N/2:
            raise ValueError("Ширина окна должна быть в диапазоне [2, N/2].")

        self.L = L
        self.orig_TS = pd.Series(tseries)
```

```

self.K = self.N - self.L + 1

# Преобразование BP в траекторную матрицу
self.X = np.array([self.orig_TS.values[i:L+i] for i in range(0,
self.K)]).T

# SVD-разложение траекторной матрицы
self.U, self.Sigma, VT = np.linalg.svd(self.X)
self.d = np.linalg.matrix_rank(self.X)

self.TS_comps = np.zeros((self.N, self.d))

if not save_mem:
    # Построить и сохранить в памяти все элементарные матрицы
    self.X_elem = np.array([ self.Sigma[i]*np.outer(self.U[:,i],
VT[i,:]) for i in range(self.d) ])

    # Диагональное усреднение элементарных матриц и расположение их
    столбцами в массиве
    for i in range(self.d):
        X_rev = self.X_elem[i, ::-1]
        self.TS_comps[:,i] = [X_rev.diagonal(j).mean() for j in
range(-X_rev.shape[0]+1, X_rev.shape[1])]

    self.V = VT.T
else:
    # Воссоздать элементарные матрицы без хранения
    for i in range(self.d):
        X_elem = self.Sigma[i]*np.outer(self.U[:,i], VT[i,:])
        X_rev = X_elem[::-1]
        self.TS_comps[:,i] = [X_rev.diagonal(j).mean() for j in
range(-X_rev.shape[0]+1, X_rev.shape[1])]

    self.X_elem = "Перезапустите с save_mem=False, чтобы сохранить в
памяти элементарные матрицы."

    # The V array may also be very large under these circumstances,
    so we won't keep it.
    # V-массив может быть очень большим, в связи с чем мы его не
    оставим в памяти
    self.V = "Перезапустите с save_mem=False, чтобы сохранить в
памяти V-матрицу"

# Расчет корреляционной матрицы
self.calc_wcorr()

def components_to_df(self, n=0):
    """
    Возвращает все компоненты BP в одном объекте Pandas DataFrame.
    """
    if n > 0:
        n = min(n, self.d)
    else:
        n = self.d

    # Создать список столбцов, обозначены F0, F1, F2 и т.д.

```

```

        cols = ["F{}".format(i) for i in range(n)]
        return pd.DataFrame(self.TS_comps[:, :n], columns=cols,
index=self.orig_TS.index)

    def reconstruct(self, indices):
        """
        Восстанавливает BP из его элементарных компонент, используя данные
индексов indices.
        Возвращает BP в виде объекта Pandas Series.

        Параметры
        -----
        indices: целое число, либо список из чисел, либо срез(n,m),
представляющий компоненты BP в виде суммы.
        """
        if isinstance(indices, int): indices = [indices]

        ts_vals = self.TS_comps[:,indices].sum(axis=1)
        return pd.Series(ts_vals, index=self.orig_TS.index)

    def calc_wcorr(self):
        """
        Расчет корреляционной матрицы F_i компонентов BP.
        """

        # Расчет весов
        w = np.array(list(np.arange(self.L)+1) + [self.L]*(self.K-self.L-1) +
list(np.arange(self.L)+1)[::-1])

        def w_inner(F_i, F_j):
            return w.dot(F_i*F_j)

        # Расчет нормированных значений ||F_i||_w, и знаменателя коэффициента
корреляции.
        F_wnorms = np.array([w_inner(self.TS_comps[:,i], self.TS_comps[:,i])
for i in range(self.d)])
        F_wnorms = F_wnorms**-.5

        # Расчет корреляционной матрицы.
        self.Wcorr = np.identity(self.d)
        for i in range(self.d):
            for j in range(i+1,self.d):
                self.Wcorr[i,j] = abs(w_inner(self.TS_comps[:,i],
self.TS_comps[:,j]) * F_wnorms[i] * F_wnorms[j])
                self.Wcorr[j,i] = self.Wcorr[i,j]

    def plot_wcorr(self, min=None, max=None):
        """
        Отображает корреляционную матрицу компонентов BP.
        """
        if min is None:
            min = 0
        if max is None:
            max = self.d

```

```

        if self.Wcorr is None:
            self.calc_wcorr()

        ax = plt.imshow(self.Wcorr)
        plt.xlabel(r"$\tilde{F}_i$")
        plt.ylabel(r"$\tilde{F}_j$")
        plt.colorbar(ax.colorbar, fraction=0.045)
        ax.colorbar.set_label("$W_{i,j}$")
        plt.clim(0,1)

        # Задать границы осей x,y графика:
        if max == self.d:
            max_rnge = self.d-1
        else:
            max_rnge = max

        plt.xlim(min-0.5, max_rnge+0.5)
        plt.ylim(max_rnge+0.5, min-0.5)

    def contrib(self, q):
        self.sigma_sumsq = (self.Sigma**2).sum()
        fig, ax = plt.subplots(1, 2, figsize=(14,5))
        ax[0].plot(self.Sigma**2 / self.sigma_sumsq * 100, lw=2.5)
        ax[0].set_xlim(0,q-1)
        ax[0].set_title("Относительный вклад, %")
        ax[0].set_xlabel("$i$")
        ax[0].set_ylabel("")
        ax[1].plot((self.Sigma**2).cumsum() / self.sigma_sumsq * 100, lw=2.5)
        ax[1].set_xlim(0,q-1)
        ax[1].set_title("Общий вклад, %")
        ax[1].set_xlabel("$i$")
        ax[1].set_ylabel("")

TS_set = pd.read_csv("path/filename.csv")
TS1 = TS_set['buf250']

TS1.plot(figsize=(15, 5), grid=True)
plt.xlabel(r"$наблюдения$")
plt.title("Исходный BP");

SSA1 = SSA(TS1, 125)
SSA1.plot_wcorr()
plt.xticks(np.arange(0, SSA1.L, step=10))
plt.yticks(np.arange(0, SSA1.L, step=10))
plt.title("Корреляционная матрица BP, $L=$" + str(SSA1.L) + "$");

plt.suptitle('Таблица компонент исходного BP')
for i in range(min(16,SSA1.L)):
    plt.subplot(4,4,i+1)
    plt.plot(SSA1.reconstruct([i]), color=cols20[i])
    #plt.rcParams['axes.prop_cycle'] = cycler(color=cols)
    plt.title(r"$\tilde{\mathbf{F}}_{\text{" + str(i) + "}}$")
plt.tight_layout(rect=[0, 0.03, 1, 0.95])

SSA1.orig_TS.plot(alpha=0.4)
SSA1.reconstruct([0]).plot(figsize=(15, 5), grid=True) #компонента для тренда

```

```

plt.title('Исходный ВР и его нулевая компонента')
plt.ylabel("")
plt.legend(['Исходный ВР', r'Компонента $\tilde{F}_{\{0\}}$']);
np.savetxt("filename.csv", SSA1.reconstruct([1]), delimiter=",")

SSA1.orig_TS.plot(alpha=0.4)
compon = [0,1,2,3,4,5,6,7,8,9,10]
SSA1.reconstruct(compon).plot(figsize=(15, 5), grid=True)
plt.legend(['Исходный ВР', 'Модель ВР'])
plt.title('Исходный ВР и его модель');

orig = SSA1.orig_TS
model = SSA1.reconstruct(compon)

MAPE = abs((orig - model) / (orig + 0.01)).sum() / len(orig) * 100 #100%-MAPE
- качество модели
AIC = (len(compon) + SSA1.L) / len(orig) + np.log( ((orig - model)**2).sum() /
len(orig) ) #чем ниже критерий, тем лучше

print(*compon, sep=',') # номера задействованных в модели компонент
print( 'k = ' +str(len(compon))) #количество компонент в модели
print( 'MAPE = ' +str(round(MAPE,3)) + '%' )
print( 'AIC = ' +str(round(AIC,3)) )

rest = range(30,SSA1.L)
SSA1.reconstruct(rest).plot(figsize=(15, 5), grid=True)
plt.title('Остатки ВР '+r"(компоненты $\tilde{F}_{\{3\}}$-$\tilde{F}_{\{9\}}$)");

plt.figure(figsize=(7, 4))
plt.hist(SSA1.reconstruct(rest))
plt.title("")
plt.xlabel(r"$значения$")
plt.ylabel("Частота")
plt.title('Диаграмма распределения остатков');

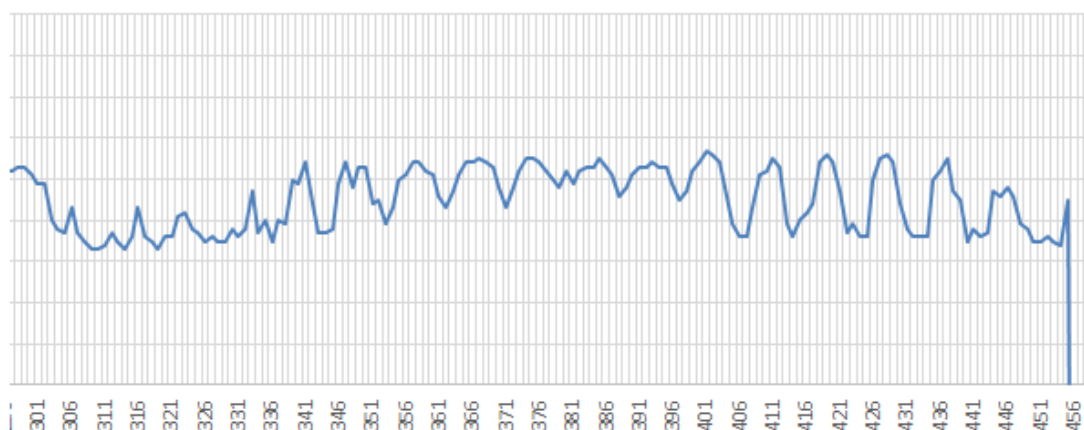
whisker = SSA1.reconstruct(range(30,125)).to_frame()
whisker.boxplot(figsize=(4, 4))
plt.title('Диаграмма размаха остатков');

ACF = [SSA1.reconstruct(range(30,125)).autocorr(i) for i in range(10)]
plt.figure(figsize=(10, 5))
plt.bar(np.arange(len(ACF)), ACF)
plt.xlabel(r"$\tau$"+', лаги')
plt.ylabel(r'$r(t)$')
plt.xticks(np.arange(len(ACF)), rotation=False) #fontsize=5
plt.ylim([-1.2, 1.2])
plt.title('Коррелограмма АКФ остатков');

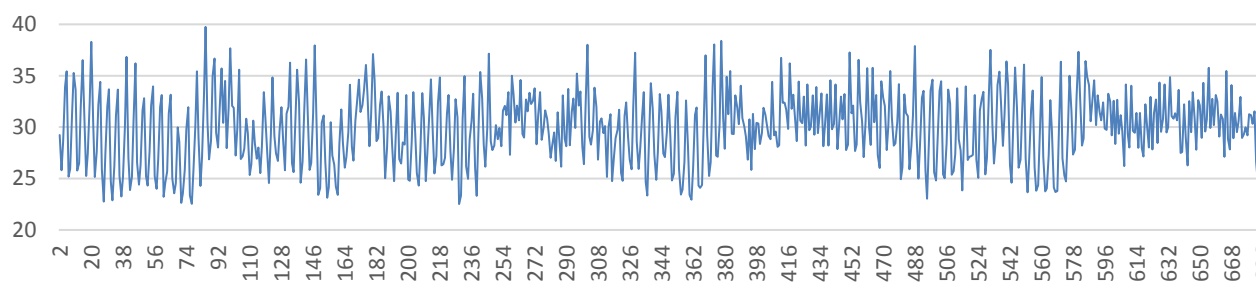
```

ПРИЛОЖЕНИЕ Б

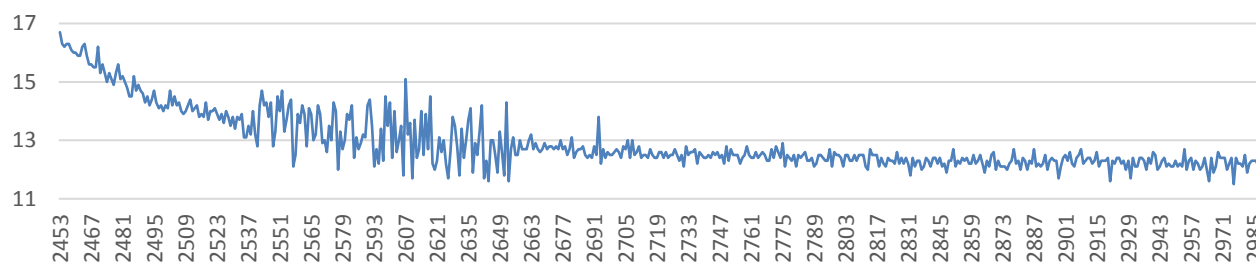
Активная мощность 6100-6700



Давление буферное до 700



Ток двигателя В



Давление буферное 17900-19400

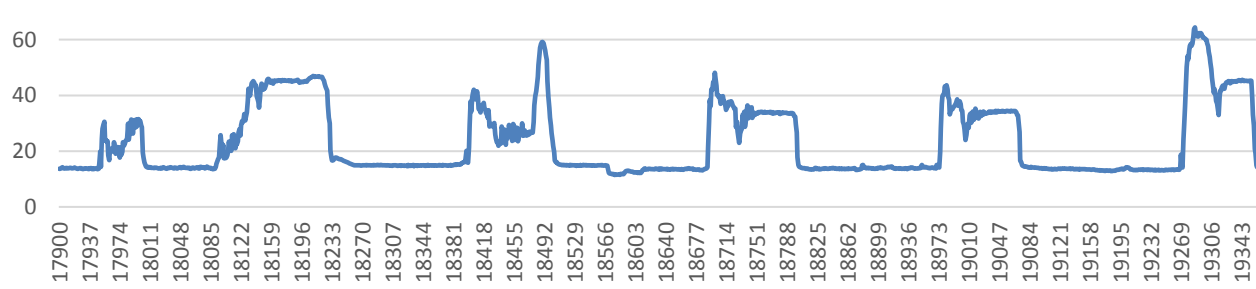


Рис.51 ВР, отобранные в ходе предварительного анализа

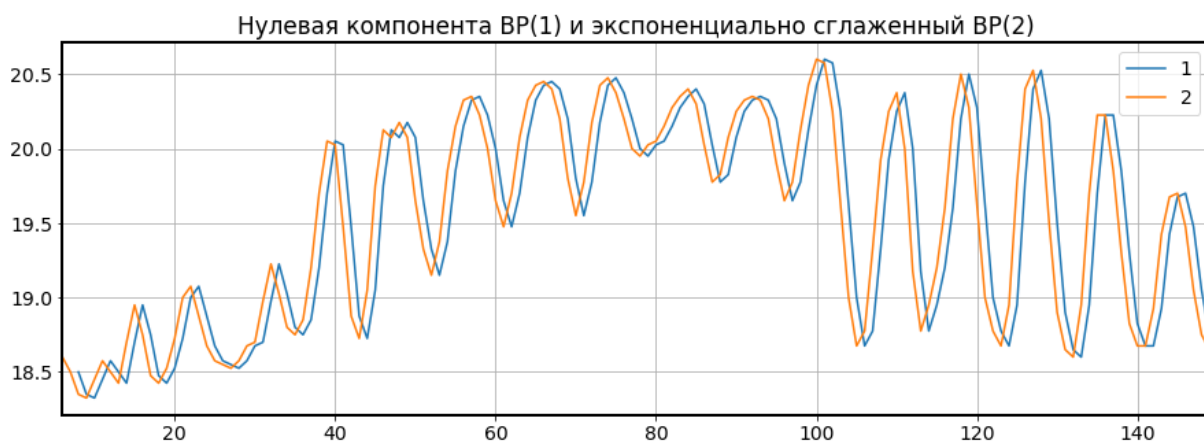


Рис.52 При $L = 2$ нулевая компонента выглядит как сглаженный исходный ВР (1), например экспоненциальное сглаживание может давать тот же результат (2).



Рис.53 ВР давления на приеме цепь, $L = 10$

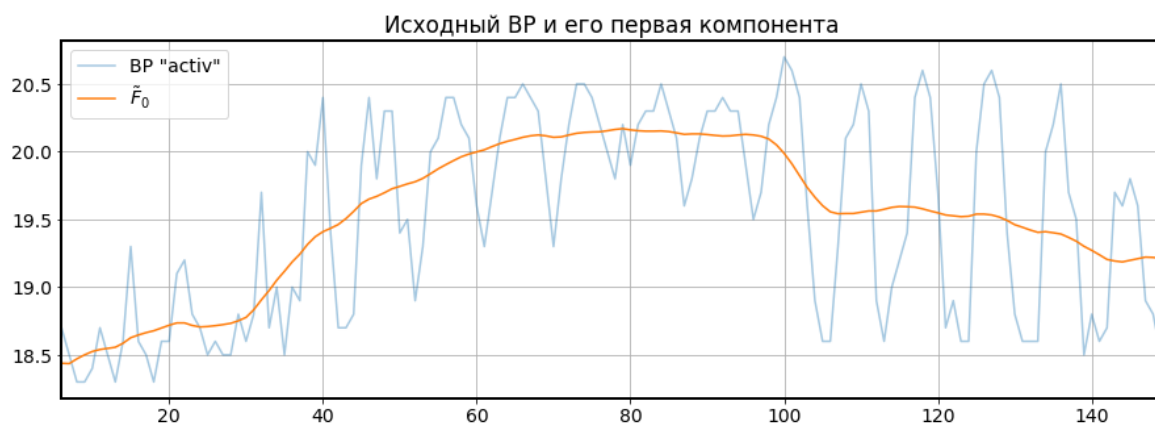


Рис.54 ВР активной мощности и первая компонента, $L = 8$

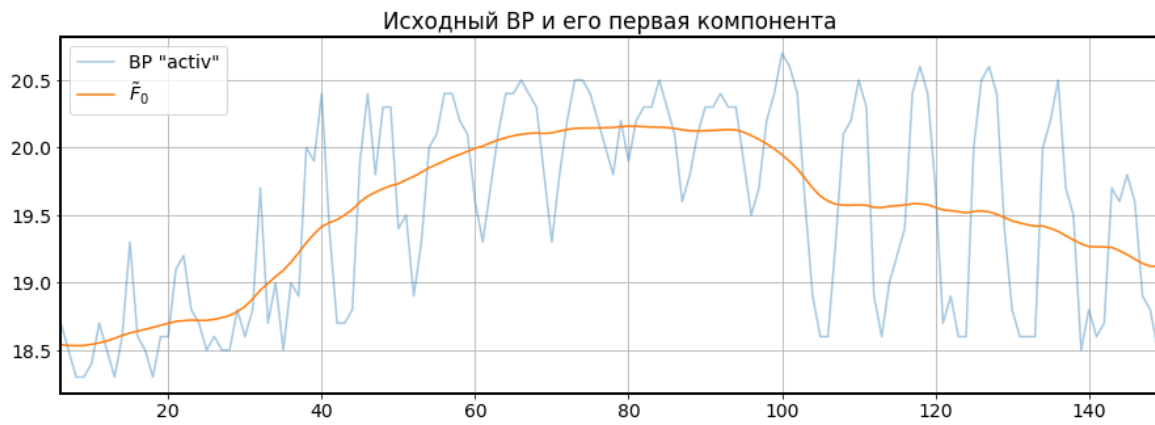


Рис.55 ВР активной мощности и первая компонента, $L = 10$

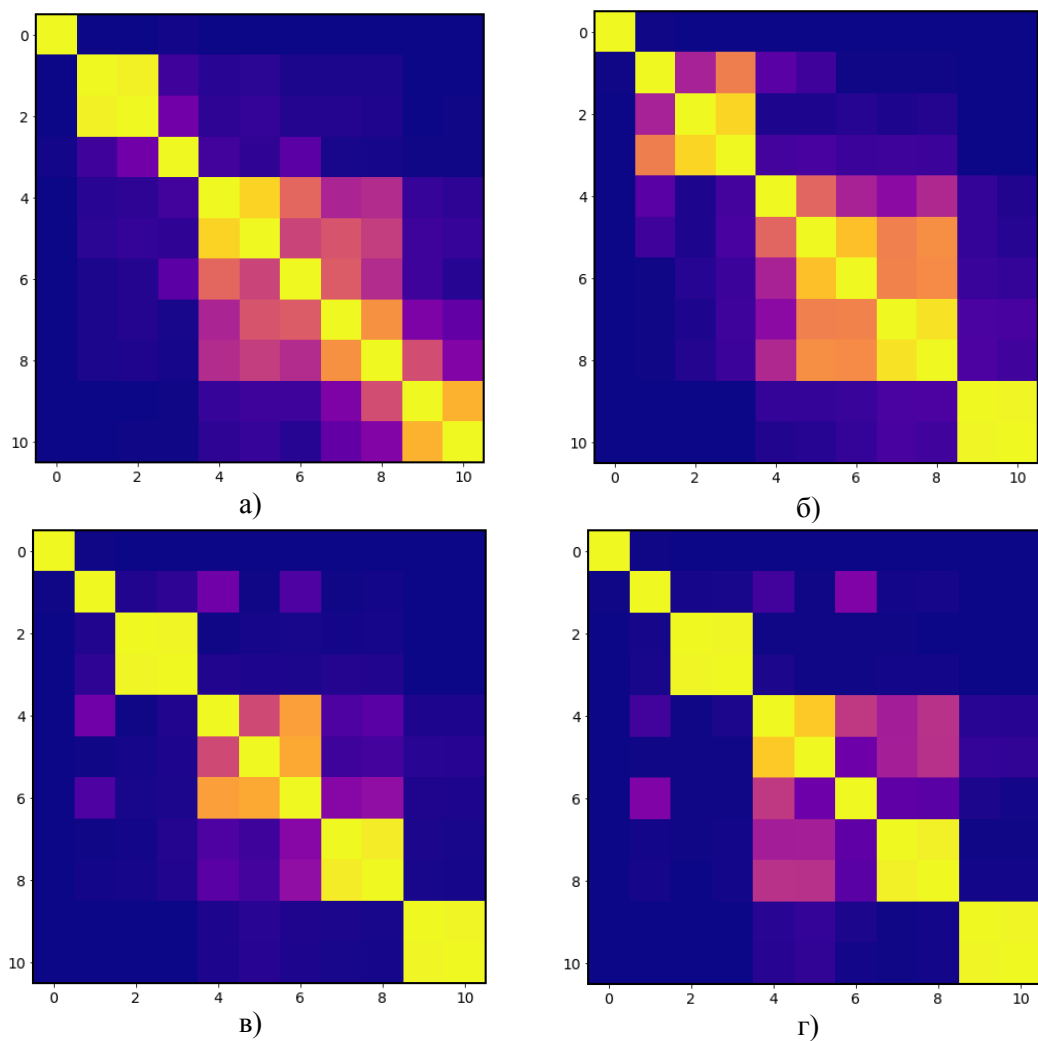


Рис.56 Первые 11 компонент корреляционной матрицы ВР активной мощности: а) $L = 45$; б) $L = 54$; в) $L = 63$; г) $L = 72$;

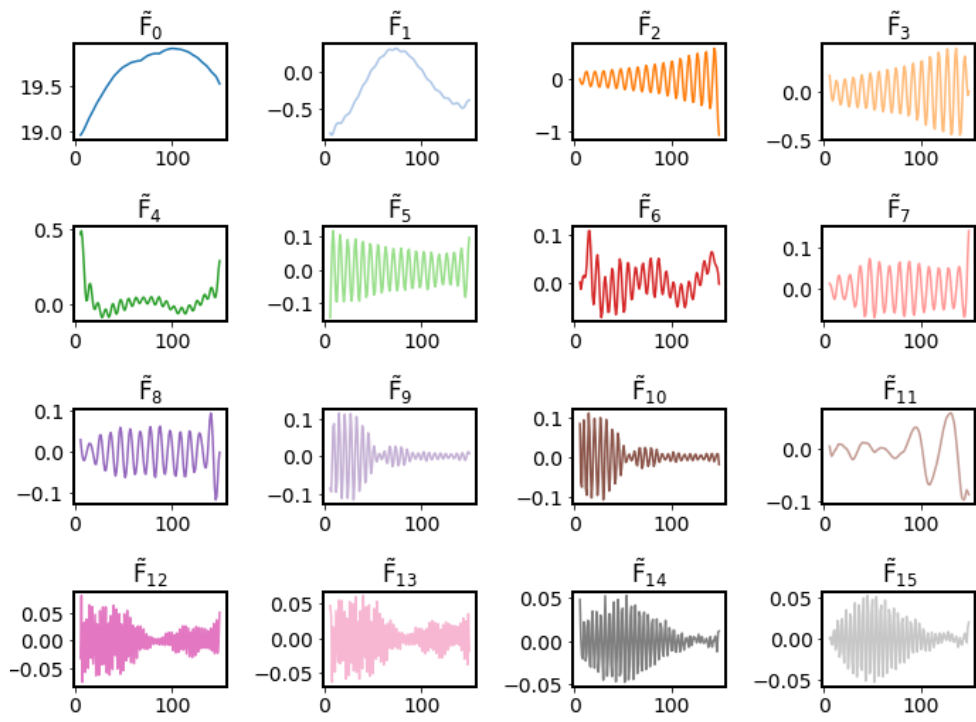


Рис.57 Первые 16 компонент ВР активной мощности, $L = 18$