

Школа – Инженерная школа информационных технологий и робототехники  
 Направление подготовки – 09.04.01 «Информатика и вычислительная техника» Отделение школы (НОЦ) информационных технологий

### МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Тема работы
<b>Разработка алгоритмического и программного обеспечения для исследования временных рядов и использованием метода главных компонент</b>

УДК 004.421.2:004.415.2:519.246

Студент

Группа	ФИО	Подпись	Дата
8ВМ82	Докторов Максим Георгиевич		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Ботыгин И А.	к.т.н. доцент		

### КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН ШБИП	Конотопский В. Ю.	к.э.н		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ООД ШБИП	Горбенко М. В.	к.т.н		

### ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Руководитель ООП	Ботыгин И А.	к.т.н. доцент		

Томск – 2020 г.

## ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ООП

Код	Результаты обучения	Требования ФГОС 3++ ВО, СУОС ТПУ, критерии ассоциации инженерного образования России и международных стандартов, требования профессиональных стандартов России
Р1	Самостоятельно приобретать и применять математические, естественнонаучные, социально-экономические и профессиональные знания в области современных информационно-коммуникационных технологий для решения междисциплинарных инженерных задач.	Требования ФГОС 3++ ВО (ОПК-1, ОПК-4), СУОС ТПУ (УК-1, УК-4, УК-5), критерий 5 АИОР (п. 1.1), требования профессионального стандарта 06.014 (ПК-1).
Р2	Разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач.	Требования ФГОС 3++ ВО (ОПК-2), СУОС ТПУ (УК-1), критерий 5 АИОР (п. 1.1, п. 1.2), соответствующий международным стандартам EUR-ACE и FEANI, требования профессиональных стандартов 06.015 (ПК-2), 06.016 (ПК-3), 06.041 (ПК-11).
Р3	Демонстрировать культуру мышления, способность выстраивать логику рассуждений и высказываний, основанных на интерпретации данных, интегрированных из разных областей науки и техники, выносить суждения на основании неполных данных, анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями.	Требования ФГОС 3++ ВО (ОПК-3), СУОС ТПУ (УК-1, УК-6), критерий 5 АИОР (п. 1.2), соответствующий международным стандартам EUR-ACE и FEANI, требования профессионального стандарта 01.004 (ПК-12, ПК-13, ПК-14).
Р4	Демонстрировать способность к практическому использованию полученных новых знаний, новых научных принципов и новых методов исследований.	Требования ФГОС 3++ ВО (ОПК-3, ОПК-4), СУОС ТПУ (УК-4, УК-6), критерий 5 АИОР (п. 1.6, п. 2.2, 2.6.), соответствующий международным стандартам EUR-ACE и FEANI, требования профессиональных стандартов 06.027 (ПК-7), 06.036 (ПК-8), 06.037 (ПК-9), 06.040 (ПК-10).

<b>P5</b>	Разрабатывать и модернизировать программное и аппаратное обеспечение информационных и автоматизированных систем, адаптировать зарубежные комплексы обработки информации и информационно-коммуникационные системы к нуждам отечественных предприятий.	Требования ФГОС 3++ ВО (ОПК-5, ОПК-6, ОПК-7), СУОС ТПУ (УК-2, УК-3, УК-6), критерий 5 АИОР (п. 2.1, п. 2.3, п. 1.5), соответствующий международным стандартам EUR-ACE и FEANI, требования профессиональных стандартов 06.015 (ПК-2), 06.026 (ПК-6).
<b>P6</b>	Осуществлять эффективное управление разработкой программных средств и проектов, демонстрировать ответственность за результаты работы и готовность следовать корпоративной культуре.	Требования ФГОС 3++ ВО (ОПК-8), СУОС ТПУ (УК-2), требования профессиональных стандартов 06.017 (ПК-4), 06.022 (ПК-5).
<b>P7</b>	Осуществлять авторское сопровождение процессов проектирования, внедрения, эксплуатации и модернизации программно-аппаратного обеспечения информационно-телекоммуникационных систем на всех этапах жизненного цикла.	Требования СУОС ТПУ (УК-2, УК-3, УК-4), критерий 5 АИОР (п. 1.5), соответствующий международным стандартам EUR-ACE и FEANI. Требования профессиональных стандартов 06.026 (ПК-6), 06.036 (ПК-8), 06.037 (ПК-9).
<b>P8</b>	Критически анализировать современные проблемы информатики и вычислительной техники, ставить задачи и разрабатывать программу исследования в индустрии новых информационных технологий, выбирать соответствующие методы решения экспериментальных и теоретических задач, критерии эффективности и ограничения их применимости, прогнозировать тенденции научно-технического развития.	Требования ФГОС 3++ ВО (ОПК-3), СУОС ТПУ (УК-1), требования профессионального стандарта 01.036 (ПК-8).
<b>P9</b>	Способность к профессиональной коммуникации в устной и письменной формах на русском и иностранном языках для решения задач профессиональной деятельности на основе истории и философии нововведений, математических методов и моделей для управления разработкой программных средств и проектов; способность руководить коллективом в сфере профессиональной деятельности, толерантно воспринимая социальные, этнические, конфессиональные и культурные различия; способность публично выступать и отстаивать свою точку зрения.	Требования ФГОС 3++ ВО (ОПК-1), СУОС ТПУ (УК-5, УК-6), требования профессионального стандарта 01.004 (ПК-12, ПК-14).

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

Школа – Инженерная школа информационных технологий и робототехники  
 Направление подготовки (специальность) – 09.04.01 Информатика и вычислительная техника  
 Отделение школы (НОЦ) – Информационных технологий

УТВЕРЖДАЮ:  
 Руководитель ООП  
 \_\_\_\_\_ Ботыгин И. А.  
 (Подпись)    (Дата)    (Ф.И.О.)

**ЗАДАНИЕ**  
**на выполнение выпускной квалификационной работы**

В форме:

магистерской диссертации
--------------------------

(бакалаврской работы, дипломного проекта/работы, магистерской диссертации)

Студенту:

Группа	ФИО
8ВМ82	Докторову Максиму Георгиевичу

Тема работы:

Разработка алгоритмического и программного обеспечения для исследования временных рядов и использованием метода главных компонент
Утверждена приказом директора (дата, номер)

Срок сдачи студентом выполненной работы:	
--	--

**ТЕХНИЧЕСКОЕ ЗАДАНИЕ:**

<p><b>Исходные данные к работе</b></p> <p><i>(наименование объекта исследования или проектирования; производительность или нагрузка; режим работы (непрерывный, периодический, циклический и т. д.); вид сырья или материал изделия; требования к продукту, изделию или процессу; особые требования к особенностям функционирования (эксплуатации) объекта или изделия в плане безопасности эксплуатации, влияния на окружающую среду, энергозатратам; экономический анализ и т. д.).</i></p>	
---	--

<b>Перечень подлежащих исследованию, проектированию и разработке вопросов</b> <i>(аналитический обзор по литературным источникам с целью выяснения достижений мировой науки техники в рассматриваемой области; постановка задачи исследования, проектирования, конструирования; содержание процедуры исследования, проектирования, конструирования; обсуждение результатов выполненной работы; наименование дополнительных разделов, подлежащих разработке; заключение по работе).</i>		
<b>Перечень графического материала</b> <i>(с точным указанием обязательных чертежей)</i>		
<b>Консультанты по разделам выпускной квалификационной работы</b> <i>(с указанием разделов)</i>		
<b>Раздел</b>	<b>Консультант</b>	
Основная часть	Доцент ОИТ ИШИТР, к.т.н., доцент Ботыгин И. А.	
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Доцент ОСГН ШБИП, к.э.н. Конотопский В. Ю.	
Социальная ответственность	Ассистент ООД ШБИП, к.т.н. Горбенко М. В.	
Английский язык	Доцент ОИЯ, к.филол.н. Аксёнова Н. В.	
<b>Названия разделов, которые должны быть написаны на русском и иностранном языках:</b>		
Раздел 1 Review of time series prediction models and methods		

<b>Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику</b>	
---	--

**Задание выдал руководитель / консультант:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Ботыгин И. А.	к.т.н., доцент		

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8ВМ82	Докторов Максим Георгиевич		

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

Школа – Инженерная школа информационных технологий и робототехники  
 Направление подготовки (специальность) – 09.04.01 Информатика и вычислительная техника  
 Уровень образования – Магистратура  
 Отделение школы (НОЦ) – Информационных технологий  
 Период выполнения – весенний семестр 2019/2020 учебного года

Форма представления работы:

Магистерская диссертация
--------------------------

(бакалаврская работа, дипломный проект/работа, магистерская диссертация)

### КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН выполнения выпускной квалификационной работы

Срок сдачи студентом выполненной работы:	
--	--

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
	Основная часть	
	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	
	Социальная ответственность	
	Обязательное приложение на иностранном языке	

**СОСТАВИЛ:**

**Руководитель ВКР**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Ботыгин И. А.	к.т.н. доцент		

**СОГЛАСОВАНО:**

**Руководитель ООП**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Ботыгин И. А.	к.т.н. доцент		

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

Студенту:

<b>Группа</b>	<b>ФИО</b>
8ВМ82	Докторов Максим Георгиевич

<b>Школа</b>	<b>ИШИТР</b>	<b>Отделение</b>	<b>ОИТ</b>
<b>Уровень образования</b>	Магистратура	<b>Направление/специальность</b>	09.04.01 «Информатика и вычислительная техника»

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

1. Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих	Использовать действующие ценники и договорные цены на потребленные материальные и информационные ресурсы, а также указанную в МУ величину тарифа на эл. энергию
2. Нормы и нормативы расходования ресурсов	—
3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования	Действующие ставки единого социального налога и НДС, ставка дисконтирования = 0,1 (см. МУ)

**Перечень вопросов, подлежащих исследованию, проектированию и разработке**

1. Планирование процесса управления НИИ: структура и график проведения, бюджет, риски и организация закупок	Построение плана-графика выполнения ВКР, составление соответствующей сметы затрат, расчет цены результата ВКР.
---	--

**Перечень графического материала (с точным указанием обязательных чертежей):**

1. График проведения и бюджет НИИ
2. Оценка ресурсной, финансовой и экономической эффективности НИИ

<b>Дата выдачи задания для раздела по линейному графику</b>	
---	--

**Задание выдал консультант:**

<b>Должность</b>	<b>ФИО</b>	<b>Ученая степень, звание</b>	<b>Подпись</b>	<b>Дата</b>
Доцент	Конотопский Владимир Юрьевич	Кандидат экономических наук		26.02.2020 г.

**Задание принял к исполнению студент:**

<b>Группа</b>	<b>ФИО</b>	<b>Подпись</b>	<b>Дата</b>
8ВМ82	Докторов Максим Георгиевич		

**ЗАДАНИЕ ДЛЯ РАЗДЕЛА  
«СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»**

Студенту:

<b>Группа</b>	<b>ФИО</b>
8BM82	Докторов Максим Георгиевич

<b>Школа</b>	<b>ИШИТР</b>	<b>Отделение</b>	<b>ОИТ</b>
<b>Уровень образования</b>	Магистратура	<b>Направление/специальность</b>	09.04.01 «Информатика и вычислительная техника»

Тема ВКР:

Разработка алгоритмического и программного обеспечения для исследования временных рядов с использованием РСА	
<b>Исходные данные к разделу «Социальная ответственность»:</b>	
1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	Статистический метод РСА для анализа и прогнозирования различных временных рядов.
Перечень вопросов, подлежащих исследованию, проектированию и разработке:	
<b>1. Правовые и организационные вопросы обеспечения безопасности:</b> <ul style="list-style-type: none"> <li>– специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства;</li> <li>– организационные мероприятия при компоновке рабочей зоны.</li> </ul>	<ul style="list-style-type: none"> <li>– Типовая инструкция по охране труда при работе на персональном компьютере ТОИ Р-45-084-01</li> <li>– Организация рабочего места для работы с ПЭВМ согласно СанПиН 2.2.2/2.4.1340-03</li> <li>– "Трудовой Кодекс Российской Федерации" от 30.12.2001 N 197-ФЗ (ред. от 24.04.2020)</li> </ul>
<b>2. Производственная безопасность:</b> 2.1. Анализ выявленных вредных и опасных факторов 2.2. Обоснование мероприятий по снижению воздействия	2.1 Вредные факторы, сопутствующие разработке и использованию программного продукта: <ul style="list-style-type: none"> <li>• Изменение показателей микроклимата;</li> <li>• Шум;</li> <li>• Недостаточная освещенность рабочей зоны;</li> <li>• Опасность поражения электрическим током;</li> <li>• Короткое замыкание;</li> <li>• Умственное перенапряжение.</li> </ul> 2.2 Меры минимизации: <ul style="list-style-type: none"> <li>• поддержка оптимальных параметров микроклимата;</li> <li>• снижение уровня шума с помощью звукопоглощающих материалов;</li> <li>• поддержка оптимальных условий освещения в области рабочих мест;</li> <li>• минимизация возможностей поражения током с помощью</li> </ul>



	<p>проведения специальных мероприятий;</p> <ul style="list-style-type: none"> <li>• снижение вероятности возникновения короткого замыкания с помощью проведения специальных мероприятий;</li> <li>• организация регламентированных перерывов в процессе работы.</li> </ul>
<b>3. Экологическая безопасность:</b>	3. Полученный в результате выполнения магистерской диссертации, программное обеспечение не наносит вред окружающей среде. Нанести вред окружающей среде могут средства, которые используются при разработке ПО. Для предотвращения такого вреда, рекомендуется организовать утилизацию отходов оргтехники, люминесцентных ламп и макулатуры.
<b>4. Безопасность в чрезвычайных ситуациях:</b>	4.1 Среди возможных ЧС типичной может быть возникновение пожара в помещении 4.2 Действия и меры, направленные на предотвращение возникновения пожаров.

<b>Дата выдачи задания для раздела по линейному графику</b>	
---	--

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Горбенко Михаил Владимирович	к.т.н.		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ВМ82	Докторов Максим Георгиевич		

## РЕФЕРАТ

Выпускная квалификационная работа содержит 93 страницы, 31 рисунок, 15 таблиц, 26 источников, 1 приложение.

Ключевые слова: метод главных компонент, PCA, анализ сингулярного спектра, SSA, сингулярный спектральный анализ, ARIMA, python, анализ данных, временные ряды, метеорологические данные.

Цель работы: разработка алгоритмического и программного обеспечения для исследования временных рядов.

Объектом исследования являются метеорологические данные.

В работе проведено исследование методов PCA, SSA и ARIMA, а также осуществлена их экспериментальная реализация на интерактивной платформе Jupyter Notebook на языке программирования Python. С помощью SSA и ARIMA были проанализированы метеорологические данные, разработан метод прогнозирования на основании изученных моделей и выполнена программная реализация моделей.

Область применения: анализ и проведение прогноза данных, полученных с метеорологических станций.

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

## СПИСОК СОКРАЩЕНИЙ

ARIMA – Auto Regressive Integrated Moving Average.

PCA – метод главных компонент;

SVD – сингулярное разложение;

SSA – анализ сингулярного спектра;

ВР – временной ряд.

## Содержание

СПИСОК СОКРАЩЕНИЙ.....	11
ВВЕДЕНИЕ.....	14
ГЛАВА 1. ОБЗОР МОДЕЛЕЙ И МЕТОДОВ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ .....	16
1.1 Основные понятия.....	16
1.2 Анализ главных компонент.....	19
1.2.1 Описание метода PCA.....	19
1.3 Сингулярный спектральный анализ.....	21
1.3.1 Математика Анализа Сингулярного Спектра.....	22
1.4 Метод главных компонент и Анализ сингулярного спектра .....	25
1.5 ARIMA .....	26
1.5.1 Математика ARIMA.....	26
ГЛАВА 2. ИНСТРУМЕНТАЛЬНЫЕ СРЕДСТВА РАЗРАБОТКИ И СТРУКТУРА ДАННЫХ .....	28
2.1 Среда разработки .....	28
2.2 Метеорологические данные .....	30
ГЛАВА 3. ЭКСПЕРИМЕНТЫ ПО ПОСТРОЕНИЮ ПРОГНОЗНЫХ МОДЕЛЕЙ .....	32
3.1 Построение прогноза с помощью SSA .....	32
3.2 Построение прогноза с помощью ARIMA .....	37
3.3 Применение метода SSA к другим временным рядам.....	42
3.4.1 Прогнозирование валютных курсов .....	42
3.4.2 Прогнозирование нефтяных котировок.....	42
4 ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ.....	43
4.1 Организация и планирование работ.....	43

4.2 Расчет сметы затрат на выполнение проекта .....	48
4.3 Оценка экономической эффективности проекта .....	55
4.4 Вывод по разделу .....	55
5 Социальная ответственность .....	56
5.1 Введение.....	56
5.2 Правовые и организационные вопросы обеспечения безопасности .....	57
5.3 Производственная безопасность .....	59
5.4 Экологическая безопасность.....	70
5.5 Безопасность в чрезвычайных ситуациях .....	71
5.6 Вывод по разделу .....	75
ЗАКЛЮЧЕНИЕ .....	76
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ .....	77
Приложение А .....	81

## ВВЕДЕНИЕ

Временной ряд обеспечивает информацию о физических, биологических, социально-экономических или финансовых системах, которые его произвели. Цель анализа временных рядов основана на модели временных рядов, чтобы определить некоторые основные его свойства на основе прошлых данных. Существует множество известных методов анализа временных рядов.

Метод главных компонент (РСА) является одним из старейших и наиболее известных методов многомерного анализа. Идея состоит в том, чтобы уменьшить размерность набора данных коррелированных переменных при сохранении максимально возможного изменения в наборе данных. Это достигается путем преобразования набора данных в новые некоррелированные переменные, упорядоченные так, чтобы первые несколько содержали большую часть вариаций, присутствующих в исходных переменных. Считается, что РСА был разработан Карлом Пирсоном в 1901 году и Гарольдом Хотеллингом в 1933 году.

Анализ сингулярного спектра (SSA) был впервые представлен в Broomhead and King (1986) [1]. С тех пор анализ временных рядов с использованием SSA стал важной областью статистики с применением в различных областях, таких как экономика, геофизика, инженерия, медицина и многие другие. Поэтому SSA признан чрезвычайно практичным инструментом, который можно использовать для решения проблем без учета какой-либо параметрической модели. SSA также может использоваться для сглаживания, поиска трендов разного разрешения и одновременного извлечения гармоник и компонентов тренда, прогнозирования и т. д. [2].

В настоящей работе исследуются методы анализа временных рядов с последующим построением моделей прогнозирования метеорологических и других временных рядов.

В первой главе проведен обзор моделей и методов прогнозирования временных рядов.

Во второй главе обосновывается выбор средств разработки и структура используемых данных.

В третьей главе описана программная реализация предложенных моделей для решения задач прогнозирования. Приведены скриншоты и способы работы с разработанным программным комплексом.

В четвертом разделе проведены экономические расчеты, определен целевой рынок для реализации, график проведения исследования, оценка экономической эффективности и итоговый бюджет проекта, который составил 159774,64 руб.

В пятом разделе рассмотрены правовые и организационные вопросы обеспечения безопасности. Проведен подробный анализ опасных и вредных факторов при разработке, и меры по их снижению. Рассмотрены вопросы утилизации производственных отходов и порядок действий при возникновении пожара.

# ГЛАВА 1. ОБЗОР МОДЕЛЕЙ И МЕТОДОВ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ

## 1.1 Основные понятия

Временной ряд – это последовательность данных, записанных во времени. Таким образом, при работе с данными временных рядов важен порядок. В частности, значения во временном ряду выражают зависимость от времени. Измерения могут проводиться периодически (час, день, неделя, месяц или год) или с любым другим регулярным интервалом времени.

Обычно данные временных рядов имеют два важных свойства:

- Данные измеряются последовательно и, в лучшем случае, с равными интервалами во времени.
- Каждая единица времени имеет не более одного измерения данных.

Кроме того, при прогнозировании временных рядов обычно ставится две цели:

- Во-первых, нужно определить шаблоны (модели), которые объясняют поведение временных рядов.
- Во-вторых, использовать эти модели для прогнозирования новых значений.

Большинство временных рядов обычно имеют, по крайней мере, один из трех типов паттернов: тренд, сезонность и/или циклы.

Тренд описывает общее поведение временного ряда, который изображен на рисунке 1. Если временной ряд демонстрирует увеличение с течением времени (1), он имеет тенденцию к росту. Если вместо этого он описывает уменьшение (2), он имеет тенденцию к снижению.



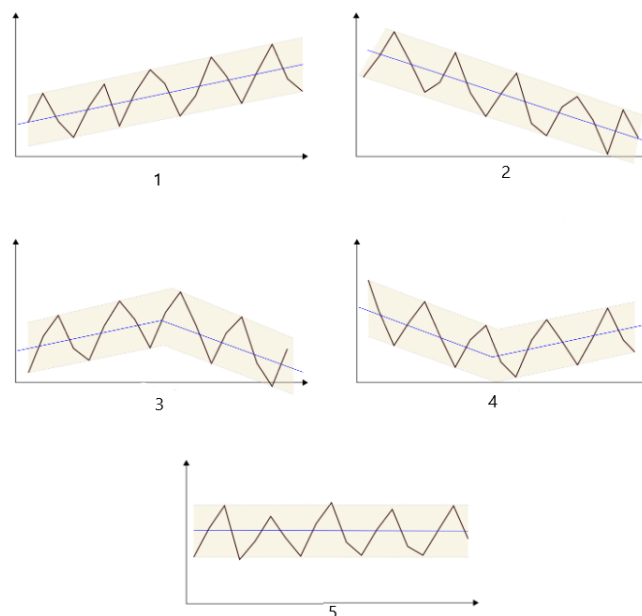


Рисунок 1 – Тренды

Тренд также может изменить направление. Может быть восходяще-нисходящий тренд и наоборот (3, 4). Наконец, стационарный или горизонтальный тренд (5) определяет временной ряд, не имеющий ни положительных, ни отрицательных долгосрочных паттернов.

Сезонность – это любое колебание (изменение) во временном ряду.

Этими изменениями могут быть время года (например, зима или лето), время дня или недели. Сезонность всегда имеет фиксированные частоты. То есть, сезонность всегда начинается и заканчивается в один и тот же период недели, года или в другие периоды (рис. 2).

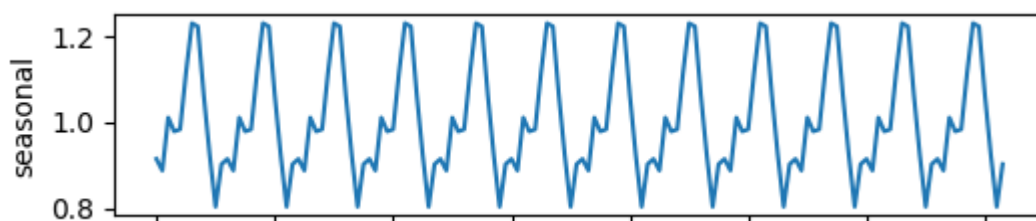


Рисунок 2 – Сезонность

Наконец, цикл во временном ряду является своего рода изменениями, которые не связаны с сезонными факторами. Это повышения и снижения с нефиксированными величинами, которые могут длиться более календарного

года. Циклы не повторяются. Обычно они являются результатом воздействия внешних факторов, что делает их сложными для прогнозирования.

Важно отметить, что не все временные ряды предсказуемы. Более конкретно, некоторые из них непредсказуемы. Такие временные ряды трудно, если невозможно, прогнозировать, так как будущая динамика с одинаковой вероятностью будут резко отличаться.

Прогнозирование временных рядов – это использование статистических методов для прогнозирования дальнейших на основе прошлых данных. Это похоже на другие статистические теории обучения, такие как контролируемое или неконтролируемое обучение. Однако прогнозирование временных рядов имеет множество нюансов, которые отличают его от обычного машинного обучения.

Многие компании изучают прогнозирование временных рядов как способ выбора лучшего решения. Точно также, уверенное понимание будущих событий может помочь широкому кругу отраслей – от традиционного сельского хозяйства до логистики и многого другого.

## 1.2 Анализ главных компонент

Идея PCA состоит в том, чтобы преобразовать набор данных коррелированных переменных в некоррелированные компоненты (называемые главными компонентами) посредством линейных комбинаций исходных переменных. Компоненты упорядочены по количеству отклонений, которые они учитывают. Другими словами, первый компонент – это линейная комбинация, которая учитывает наибольшее возможное отклонение. Вторым компонентом, будучи ортогональным первому, представляет собой линейную комбинацию, которая учитывает наибольшую возможную дисперсию оставшегося отклонения (отклонение, не зафиксированное компонентом один) и так далее. Облако точек исходных переменных не является сферическим, большая часть вариаций отражается в первых нескольких основных компонентах, что делает PCA подходящим методом для:

- Уменьшения размерности данных.
- Анализа данных и для создания прогнозных моделей.
- Поиска важных атрибутов/переменных (выбор объектов в многомерных данных).
- Поиска закономерностей в данных.
- Фильтрации шума.

Основным трудом по PCA является работа его создателя Карла Пирсона [3]. Дополнительный материал по PCA можно найти в [4, 5].

### 1.2.1 Описание метода PCA

Учитывая матрицу  $X$ , которая соответствует  $n$  наблюдениям с  $d$  характеристиками, и  $k$  входов, основная цель PCA состоит в том, чтобы разложить матрицу  $X$  на две меньшие матрицы,  $Z$  и  $W$ , так что  $X = Z \cdot W$ , где  $Z$  имеет размеры  $n \cdot k$  и  $W$  имеет размеры  $k \cdot d$ . Каждый ряд  $Z$  является фактором загрузки. Каждая строка  $W$  называется главным компонентом и изображена на рисунке 3.

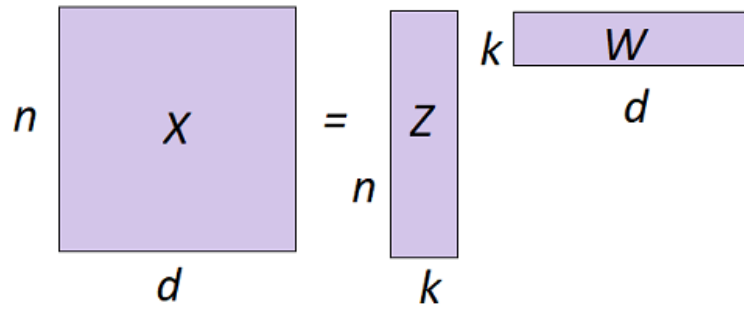


Рисунок 3 – PCA разбивает матрицу  $X$  на две меньшие матрицы  $Z$  и  $W$ .

В PCA минимизируем квадратичную ошибку следующей целевой функцией:

$$f(W, Z) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^d (w_j^T z_i - x_{ij})^2 = \frac{1}{2} \|ZW - X\|_F^2 \quad (1)$$

Есть три общих подхода к решению PCA:

### 1. Singular Value Decomposition (SVD)

Этот подход сначала использует сингулярное разложение (SVD), чтобы найти ортогональную  $W$ . Затем использует ортогональный  $W$  для вычисления  $Z$  следующим образом:

$$X = U \Sigma W^T, \quad (2)$$

$$\nabla_Z f(Z) = Z \cdot W \cdot W^T - X \cdot W^T \rightarrow Z = X \cdot W^T (W \cdot W^T)^{-1} = X \cdot W^T, \quad (3)$$

### 2. Чередование минимизации

Это итеративный подход, который чередуется между:

Исправление  $Z$  и поиск оптимальных значений для  $W$ :

$$\nabla_W f(W) = Z^T \cdot Z \cdot W - Z^T \cdot X, W = (Z^T \cdot Z)^{-1} \cdot (Z^T \cdot X), \quad (4)$$

Исправление  $W$  и поиск оптимальных значений для  $Z$ :

$$\nabla_Z f(Z) = Z \cdot W \cdot W^T - X \cdot W^T, Z = X \cdot W^T \cdot (W \cdot W^T)^{-1}, \quad (5)$$

### 3. Стохастический градиентный спуск

Это итеративный подход, когда матрица  $X$  очень большая. На каждой итерации он выбирает случайный пример  $i$  и показывает  $j$  и обновляет  $W$  и  $Z$ :

$$\text{Set } \omega^j \text{ to } \omega^j - \alpha^t \nabla_{\omega_j}(\omega^j, z_i, x_{ij}), \quad (6)$$

$$\text{Set } z_i \text{ to } z_i - \alpha^t \nabla_{z_j}(\omega^j, z_i, x_{ij}), \quad (7)$$

### 1.3 Сингулярный спектральный анализ

За последние несколько десятилетий в научных исследованиях и научных работах проявился интерес к области прогнозирования и попытка количественно оценить и обосновать широкий спектр используемых методов.

Сингулярный спектральный анализ (SSA), начиная со второй половины XX века, стал быстро развивающимся методом анализа временных рядов. Поскольку это можно назвать анализом главных компонент для временных рядов, SSA определенно станет стандартным методом анализа временных рядов и обработки сигналов в будущем. Более того, проблемы, решаемые SSA, значительно шире, чем проблемы для PCA. В частности, проблемы оценки частоты, прогнозирования и вменения недостающих значений могут быть решены в рамках SSA. Идея SSA исходила от различных научных сообществ, таких как анализ временных рядов (разложение Кархунена-Лоэва), обработка сигналов (аппроксимация низкого ранга и оценка частоты) и анализ многомерных данных (метод главных компонент). Также, в зависимости от области применения, рассматриваются разные точки зрения на одни и те же алгоритмы, выбор параметров и методологию в целом.

Задача SSA – сделать разложение исходного ряда на сумму небольшого числа независимых и интерпретируемых компонентов, таких как медленно меняющийся тренд, колебательные составляющие и бесструктурный шум.

Подобно PCA, SSA можно использовать для нахождения трендов, сглаживания, извлечения сезонности, одновременного извлечения циклов с маленькими и большими периодами, извлечение периодичности с различными амплитудами, одновременное выделение сложных трендов и периодичностей, нахождение структуры в коротких временных рядах и обнаружение точки изменения.

Одними из фундаментальных работ по анализу сингулярного спектра являются монографии [2], в которых описываются и систематизируются алгоритмы, связанные с SSA и их реализация в R-пакете Rssa [6] с

многочисленными примерами. Монографии содержит описание различных аспектов самого SSA.

### 1.3.1 Математика Анализа Сингулярного Спектра

Анализ состоит из четырех шагов:

Первый этап: разложение. Этап разложения состоит из двух этапов: вложения и сингулярного разложения (SVD).

#### Шаг 1. Вложение

Процедура вложения отображает исходный временной ряд в последовательность многомерных векторов с размерностью  $L$  (длина окна) образуя  $K = N - L + 1$  векторов вложения  $X_i = (x_i, \dots, x_{i+L-2})^T, i = 1, \dots, K$ .

Матрица траекторий временного ряда  $F$  имеет вид:

$$X = [X_1: \dots : X_K] = \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_K \\ x_2 & x_3 & x_4 & \dots & x_{K+1} \\ x_3 & x_4 & x_5 & \dots & x_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & x_{L+2} & \dots & x_N \end{pmatrix}, \quad (8)$$

Матрица траекторий является матрицей Ганкеля, потому что  $X$  имеет одинаковые элементы на диагоналях.

#### Шаг 2. Сингулярное разложение (SVD)

Из матрицы  $X$  определяем ковариационную матрицу  $XX^T$ . Сингулярное разложение из  $XX^T$  предоставляет набор из  $L$  собственных значений в порядке убывания величины ( $\lambda_1 \geq \lambda_2 \geq \dots \lambda_L \geq 0$ ) и соответствующих собственных векторов  $U_1, U_2, \dots, U_L$ . Тогда SVD траекторной матрицы  $X$  можно записать в виде:

$$X = X_1 + \dots + X_L, \quad (9)$$

где  $X_i = \sqrt{\lambda_i} \cdot U_i \cdot V_i^T$ . Матрицы  $X_i$  элементарны, то есть имеют ранг 1.

$V_i = \frac{X^T U_i}{\sqrt{\lambda_i}}, i, \dots, L$ , (эквивалентен  $i$ -й векторному коэффициенту вектора  $XX^T$ ). Набор  $(\sqrt{\lambda_i} \cdot U_i \cdot U_i^T)$  называют  $i$ -й собственной тройкой сингулярного разложения.

Второй этап: восстановление

Этот этап состоит из этапа: группирования и усреднения по диагонали.

### Шаг 3. Группировка

Процедура группировки разделяет набор индексов  $\{1, \dots, L\}$  на  $r$  (выбирается пользователем) непересекающиеся подмножества. Пусть  $I = \{i_1, i_2, \dots, i_r\}$  группа из  $r$  выбранных собственных троек и  $X_I = X_{i_1} + X_{i_2} \dots + X_{i_r}$ , где  $X_I$  относится к сигналу  $F$ , в то время как остальные  $(L-r)$  собственные триггеры обозначают временные ряды, относящиеся к члену ошибки и называется группировкой собственных троек.

### Шаг 4. Диагональное усреднение

На последнем шаге базового алгоритма каждая матрица сгруппированного разложения переводится в новый ряд длины  $N$ .

Пусть  $Y$  – некоторая  $L \times K$  матрица с элементами  $y_{ij}$ , где  $1 \leq i \leq L, 1 \leq j \leq K$ . Положим  $L^* = \min(L, K)$ ,  $K^* = \max(L, K)$  и  $N = L + K - 1$ . Пусть  $y_{ij}' = y_{ij}$ , если  $L \leq K$ , и  $y_{ij}' = y_{ij} -$  иначе. Диагональное усреднение переводит матрицу  $Y$  в ряд  $g_0, \dots, g_{N-1}$  по формуле

$$g_k = \begin{cases} \frac{1}{k+1} \sum_{m=1}^{k+1} y_{m, k-m+2}^* , & 0 \leq k < L^* - 1 \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y_{m, k-m+2}^* , & L - 1^* \leq k < K^* , \\ \frac{1}{N-k} \sum_{m=k-K^*+2}^{N-K^*+1} y_{m, k-m+2}^* , & K^* \leq k < N \end{cases} \quad (10)$$

Выражение соответствует усреднению элементов матрицы вдоль «диагоналей»  $i + j = k + 2$ : выбор  $k = 0$  дает  $g_0 = y_{11}$ , для  $k = 1$  получаем  $g_1 = (y_{12} + y_{21}) / 2$  и т.д. Заметим, что если матрица  $Y$  является траекторной матрицей некоторого ряда  $(h_0, \dots, h_{N-1})$  (другими словами, если матрица  $Y$  является ганкелевой), то  $g_i = h_i$  для всех  $i$ .

Применяя диагональное усреднение к результирующим матрицам  $X_{I_k}$  получаем ряды  $\tilde{F}^{(k)} = (\tilde{f}_0^{(k)}, \dots, \tilde{f}_{N-1}^{(k)})$  и, следовательно, исходный ряд  $(f_0, \dots, f_{N-1})$  раскладывается в сумму  $m$  рядов:

$$f_n = \sum_{k=1}^m \tilde{f}_n^{(k)} , \quad (11)$$

Из приведенного выше описания видно, что основным параметром базового алгоритма метода SSA является длина окна  $L$ . Выбор длины окна зависит от решаемой задачи и предварительной информации об анализируемом ВР. В общем случае, нет универсальных правил и безусловных рекомендаций для выбора длины окна [2].



## 1.4 Метод главных компонент и Анализ сингулярного спектра

Если сравнивать SSA с PCA, то результат PCA является, в основном, неструктурированным приближением (с точки зрения временной структуры), в то время как SSA принимает во внимание временную структуру.

К примеру, если у нас  $m$  временных рядов длиной  $N$ ,  $m < N$ , то PCA предоставляет только  $m$  компонент. Для  $m = 1$  бессмысленно применять PCA; при  $m = 2$  двух компонент может быть недостаточно даже для того, чтобы попытаться разложить на тренд, колебания и шум.

Другой пример связан с разложением на сигнал и шум, когда сигнал описывается несколькими компонентами SSA (он выполняется, если временной ряд хорошо аппроксимируется конечной суммой произведений полиномов, экспонент и синусоид).

Например, пусть временные ряды состоят из системы зашумленных синусоид с небольшим отношением сигнал/шум. PCA не помогает извлекать сигнал для любой длины временного ряда  $N$ . А SSA применяет SVD (сингулярное разложение – тот же математический инструмент, что и в PCA) к матрице траекторий. Для достаточно больших  $L$  и  $N$  SSA может приблизительно извлечь сигнал.

Тот же эффект имеет место быть для случая, когда временные ряды состоят из временных компонент, таких как тренд и колебания. PCA не работает для извлечения одного из компонент. SSA может это сделать благодаря SVD. Таким образом, для временного ряда PCA обычно не работает и его можно заменить SSA.

## 1.5 ARIMA

Один из наиболее распространенных методов, используемых в прогнозировании временных рядов, известен как модель ARIMA. Согласно этой модели временной ряд представляет собой некоторые случайные процессы. А задача моделирования заключается в восстановлении характеристик случайного процесса по одной его части. Такое восстановление в общем случае невозможно, однако вполне осуществимо при некоторых ограничениях на класс исследуемых процессов, чаще всего встречающихся в практической деятельности.

Джордж Бокс и Гвилим Дженкинс представили модель авторегрессии с интегрированным скользящим средним или ARIMA в 1970 году в своей книге «Анализ временных рядов: прогнозирование и контроль» [7].

Любые «несезонные» временные ряды, которые демонстрируют закономерности и не являются случайным белым шумом, можно моделировать с помощью моделей ARIMA.

Модель ARIMA характеризуется 3 терминами:  $p$ ,  $d$ ,  $q$ :

Если временной ряд имеет сезонные закономерности, то нужно добавить сезонную модель, и он станет SARIMA.

### 1.5.1 Математика ARIMA

Модель авторегрессии порядка  $p$  показывает зависимость значения нынешнего периода от прошлых значений  $p$  периодов

$$AR(p): y_t = c + \varphi_1 y_{(t-1)} + \varphi_2 y_{(t-2)} + \dots + \varphi_p y_{(t-p)} + \varepsilon_t, \quad (12)$$

Модель скользящего среднего порядка  $q$  показывает зависимость значения нынешнего периода от ошибок предсказания предыдущих  $q$  периодов.

$$MA(q): y_t = c + \varepsilon_t + \theta_1 \varepsilon_{(t-1)} + \theta_2 \varepsilon_{(t-2)} + \dots + \theta_q \varepsilon_{(t-q)}, \quad (13)$$

Модель авторегрессии с интегрированием и скользящим средним порядков  $(p, d, q)$  является суммой  $AR(p)$  и  $MA(q)$  моделей и может быть представлена в виде:

$$ARIMA(p, d, q): (1 - \varphi_1 L - \dots - \varphi_p L^p)((1 - L)^d y_t - \mu) = (1 + \theta_1 L + \dots + \theta_q L^q) \varepsilon_t, \quad (14)$$

где  $d$  – количество дифференцирований исходного временного ряда до достижения его стационарности, а  $L$  – величина лага.

Некоторым ограничением ARIMA является то, что она обрабатывает данные временных рядов в линейной форме и ряд должен быть стационарным, а точность модели прогнозирования значительно зависит от шума набора данных временных рядов.

Модель SARIMA( $p, d, q$ )( $P, D, Q$ ) $s$  является обобщением ARIMA-модели для временных рядов, в которых имеется ярко выраженная сезонная компонента. В такой модели помимо порядка авторегрессионной части, порядка скользящей средней и порядка разностей вводятся сезонные параметры ( $P, D, Q, s$ ), позволяющие учитывать циклические колебания процесса при построении прогноза.

## ГЛАВА 2. ИНСТРУМЕНТАЛЬНЫЕ СРЕДСТВА РАЗРАБОТКИ И СТРУКТУРА ДАННЫХ

### 2.1 Среда разработки

В качестве инструмента для исследования была выбрана интерактивная среда для разработки – Anaconda [8] на языке программирования Python и R, включающая в себя веб-приложение Jupyter Notebook [9] и библиотеки для анализа данных.

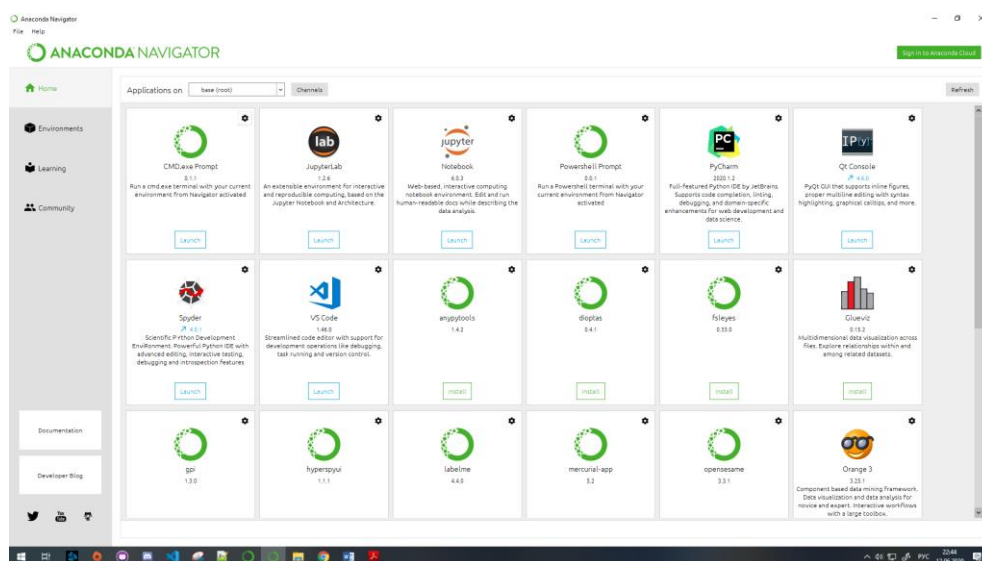


Рисунок 4 – Интерфейс Anaconda

Jupyter Notebook – это бесплатный интерактивный веб-инструмент с открытым исходным кодом, которое позволяет создавать и обмениваться файлами, содержащими код, математические уравнения, карты, графики, визуализацию и текстовую информацию (рис. 5). Инструмент взаимодействует со многими языками программирования, такими как Python, PHP, R, C # и т.д. Блоки кода или форматированного текста содержатся в ячейках. Эти ячейки могут быть запущены индивидуально (и не обязательно в последовательности). Такое сочетание блоков кода и текста делает их более удобными по сравнению с другими средами разработки. Интерфейс прост в освоении, что делает его хорошим инструментом для новичков.

Включает в себя широкий спектр различных библиотек для анализа и исследования временных рядов.

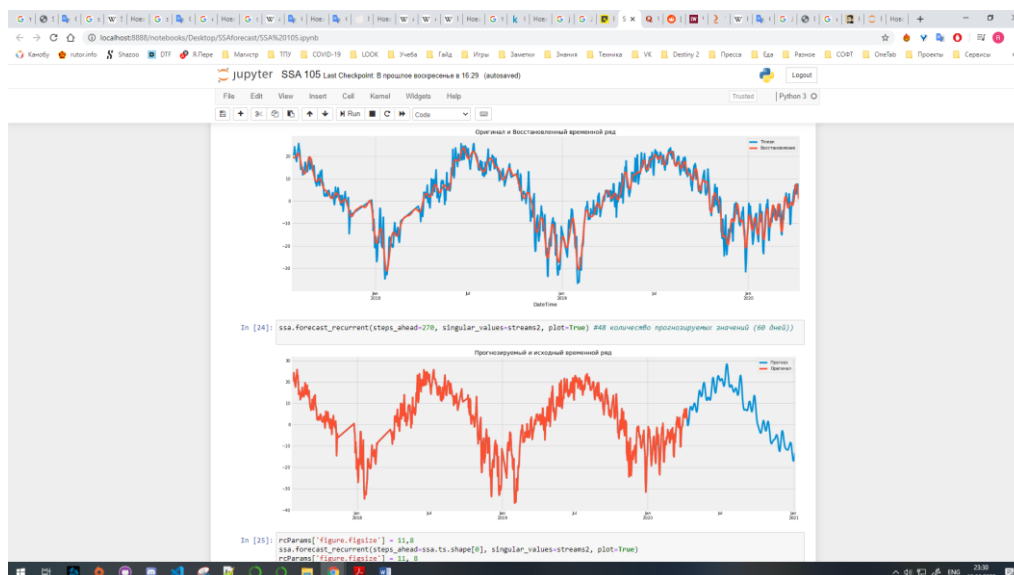


Рисунок 5 – Интерфейс Jupyter Notebook

Еще одна интересная функция Jupyter Notebook – возможность создавать контрольные точки. Создав контрольную точку, можно сохранить текущее состояние проекта, чтобы впоследствии откатиться к этой контрольной точке и отменить изменения, внесенные в записную книжку за это время.

## 2.2 Метеорологические данные

Объектом проведения анализа выбраны данные, представленные Системой климатического мониторинга Института мониторинга климатических и экологических систем СО РАН в г. Томске (рисунок 6) [10].

Тип измерителя	Значение
АКЛ Жидкие осадки	17884
АКЛ Профиль температуры	401801
Направление ветра	35510
АКЛ (осадки)	2920
<b>АКЛ (температура)</b>	<b>931439</b>
АКЛ (р)	699
Аналоговая станция	105716
ДТТ-01 (огнеопасный)	123372
ДТТ-02 (огнеопасный)	0
Датчик электрического поля	1808387
EPS-2-10	221562
Датчик галса фона	15
Координаты станций	0
Ой ветер	0
Ой температура	0
Осцилятор СЕТТИОС	437427
Оптис 2018	0
Поток фот активной радиации	457069
Ручной геоинтерпретатор	0
Влажность почвы	889748
Высота свежего покрыва	0

№ ст.	Широта	Долгота
56.4813	85.0099	160
56.2382	84.8692	158.1
56.6297	84.9947	151
56.464	84.9332	105
56.4757	85.0547	202
54.8577	85.1118	155
56.474	85.0543	167
56.4153	84.0738	80
N	N	N
56.813	82.458	105
56.462	84.9503	126
37.0753	86.2374	122
56.8022	86.0918	123

Серийный номер
15407АМК-03
15408АМК-03

Рисунок 6 – Сайт Системы климатического мониторинга

Данные содержат в себе основные метеорологические параметры (температура, скорость ветра, атм. давление, влажность воздуха и т.д.) в формате .txt со станции №105 (12427АМК-03) за период с 2017-07-20 по 2020-04-09 года.

Исследуемый материал скачивался последовательно по 2 месяца. Сайт с метеорологическими данными не предусматривает загрузку всей доступной им информации одним файлом.

Данные были предварительно обработаны и добавлены заголовки в соответствии с метеопараметрами. Заголовки представлены в таблице 1.

Первоначально данные состоят из нескольких файлов в txt формате, их нужно объединить в один, преобразовать в формат csv.

Структура файла с метеорологическими обработанными данными в формате .csv представлена на рисунке 7.

```

DateTime,AffinityIndex,SerialNumber,H,Tmean,WmeanH,WindDirectionH(^{\circ}),WminH,WmaxH,WmeanV,AtmosphericPressure,Relat
2017-07-22 00:01:00,6,12427AMK-03,15,15.86,0.2,94.53,0.01,0.44,-0.13,750.45,100,15.86,17.99,14.53,1.1954,342.12
2017-07-22 00:02:00,6,12427AMK-03,15,15.89,0.26,104.28,0.15,0.53,-0.19,750.45,100,15.89,18.01,14.55,1.1953,342.14
2017-07-22 00:03:00,6,12427AMK-03,15,15.87,0.23,68.53,0.11,0.36,-0.19,750.45,100,15.87,17.99,14.54,1.1954,342.13
2017-07-22 00:04:00,6,12427AMK-03,15,15.95,0.17,47.04,0.01,0.33,-0.18,750.45,100,15.95,18.05,14.62,1.195,342.18
2017-07-22 00:05:00,6,12427AMK-03,15,16.01,0.06,11.99,0.01,0.22,-0.19,750.45,100,16.01,18.16,14.68,1.1947,342.22
2017-07-22 00:06:00,6,12427AMK-03,15,16.04,0.18,31.14,0.01,0.41,-0.05,750.45,100,16.04,18.19,14.7,1.1946,342.24
2017-07-22 00:07:00,6,12427AMK-03,15,15.91,0.51,15.8,0.2,0.68,-0.11,750.45,100,15.91,18.04,14.58,1.1952,342.16
2017-07-22 00:08:00,6,12427AMK-03,15,15.97,0.56,21.2,0.36,0.76,-0.12,750.45,100,15.97,18.1,14.63,1.1949,342.19
2017-07-22 00:09:00,6,12427AMK-03,15,15.91,0.43,66.64,0.25,0.69,-0.15,750.45,100,15.91,18.05,14.58,1.1952,342.16
2017-07-22 00:10:00,6,12427AMK-03,15,15.98,0.38,73.46,0.18,0.59,-0.18,750.45,100,15.98,18.12,14.64,1.1949,342.2
2017-07-22 00:11:00,6,12427AMK-03,15,15.99,0.18,62.19,0.01,0.37,-0.12,750.45,100,15.99,18.13,14.65,1.1948,342.21
2017-07-22 00:12:00,6,12427AMK-03,15,16.01,0.11,49.23,0.01,0.27,-0.15,750.45,100,16.01,18.15,14.67,1.1947,342.22
2017-07-22 00:13:00,6,12427AMK-03,15,16.06,0.1,16.79,0.01,0.23,-0.09,750.46,100,16.06,18.21,14.72,1.1945,342.25
2017-07-22 00:14:00,6,12427AMK-03,15,16.32,0.19,9.72,0.01,0.51,0.08,750.47,100,16.32,18.52,14.97,1.1933,342.43
2017-07-22 00:15:00,6,12427AMK-03,15,16.14,0.27,27.41,0.02,0.51,-0.1,750.47,100,16.14,18.31,14.8,1.1941,342.31
2017-07-22 00:16:00,6,12427AMK-03,15,16.01,0.36,48.63,0.2,0.5,-0.17,750.47,100,16.01,18.16,14.68,1.1947,342.22
2017-07-22 00:17:00,6,12427AMK-03,15,15.97,0.5,53.93,0.26,0.87,-0.17,750.47,100,15.97,18.11,14.63,1.1949,342.19
2017-07-22 00:18:00,6,12427AMK-03,15,15.65,1.39,22.2,0.78,2.03,-0.19,750.47,100,15.65,17.75,14.33,1.1964,341.99
2017-07-22 00:19:00,6,12427AMK-03,15,15.72,0.89,43.95,0.25,1.58,-0.16,750.47,100,15.72,17.82,14.39,1.1961,342.03
2017-07-22 00:20:00,6,12427AMK-03,15,15.76,0.41,65.91,0.01,0.83,-0.08,750.47,100,15.76,17.87,14.44,1.1959,342.06
2017-07-22 00:21:00,6,12427AMK-03,15,15.81,0.56,58.56,0.22,1.08,-0.11,750.47,100,15.81,17.93,14.49,1.1957,342.09
2017-07-22 00:22:00,6,12427AMK-03,15,15.72,0.63,63.09,0.05,1.49,-0.14,750.47,100,15.72,17.82,14.4,1.1961,342.03
2017-07-22 00:23:00,6,12427AMK-03,15,15.61,0.67,45.13,0.29,1.2,-0.27,750.46,100,15.61,17.7,14.3,1.1966,341.96
2017-07-22 00:24:00,6,12427AMK-03,15,15.68,0.48,58.25,0.14,0.9,-0.16,750.46,100,15.68,17.78,14.36,1.1963,342.01
2017-07-22 00:25:00,6,12427AMK-03,15,15.73,0.58,37.24,0.14,1.1,-0.08,750.46,100,15.73,17.84,14.41,1.196,342.04

```

Рисунок 7 – Структура файла с метеорологическими данными

Таблица 1 – Метеорологические параметры

Наименование	
Время измерения	DateTime
Индекс группы на момент измерения	AffinityIndex
Серийный номер	SerialNumber
Высота установки от уровня земли	H
Средняя температура, градусы Цельсия	Tmean
Средняя скорость горизонтального ветра	WmeanH
Среднее направление гор. ветра, градусы	WindDirectionH( $\hat{\text{A}}^{\circ}$ )
Мин. скорость мгновенного гор. ветра, м/с	WminH
Макс. скорость мгновенного гор. ветра, м/с	WmaxH
Средняя скорость верт. ветра, м/с	WmeanV
Атмосферное давление, мм.рт.ст.	AtmosphericPressure
Относительная влажность воздуха, %	RelativeHumidity
Температура точки росы, градусы Цельсия	TdewPoint( $\hat{\text{A}}^{\circ}\text{C}$ )
Упругость (давление) водяного пара, гПа	WaterVaporElasticity
Абсолютная влажность, г/м <sup>3</sup>	AbsolutusHumidity
Плотность воздуха, кг/м <sup>3</sup>	DensityofAir
Скорость звука в воздухе, в м/с	SpeedofSoundinAir

## ГЛАВА 3. ЭКСПЕРИМЕНТЫ ПО ПОСТРОЕНИЮ ПРОГНОЗНЫХ МОДЕЛЕЙ

### 3.1 Построение прогноза с помощью SSA

В программных экспериментах по реализации SSA использовались среднесуточные данные за весь период наблюдения. Частота может быть изменена: например, H - час, D - день, W - неделя, M - месяц и т.д.

Ниже приведен фрагмент программного кода для получения среднесуточных данных:

```
means105 = data105.groupby(pd.Grouper(freq='D')).mean() # ПЕРИОД
```

Результат выполнения кода представлен на рисунке 2.

DateTime	AffinityIndex	H	Tmean	WmeanH	WindDirectionH(°)	WminH	WmaxH	WmeanV	AtmosphericPressure	RelativeHumidity	TdewPoint(°C)
2017-07-22	6.0	15.0	19.476312	1.202660	176.703170	0.322553	2.263915	-0.122234	750.230475	77.639915	15.043355
2017-07-23	6.0	15.0	20.851142	1.164227	231.695564	0.324827	2.181821	0.055816	749.047753	68.623974	14.043468
2017-07-24	6.0	15.0	20.588375	1.556807	260.066591	0.429341	2.968989	0.220068	746.373318	77.926284	16.303727
2017-07-25	6.0	15.0	21.122393	2.011374	166.887422	0.407171	4.149158	0.133744	742.115037	77.207245	16.613722
2017-07-26	6.0	15.0	21.097521	1.113449	130.933364	0.223541	2.328545	-0.164357	744.684319	80.063233	17.087560

Рисунок 8 – Среднесуточные значения метеопараметров

Для исследования была выбрана среднесуточная температура (Tmean) и общее количество значений составило 993 значений (2,5 года). И может быть изменена на другие параметры, что позволяет нам проводить прогнозирование любых метеорологических величин, которые есть в исследуемом файле.

```
#pd.set_option('display.max_rows', 500)
per_temp105 = means105['Tmean'] # параметр
per_temp105
```

На рисунке 9 приведен линейный график изменения температуры.



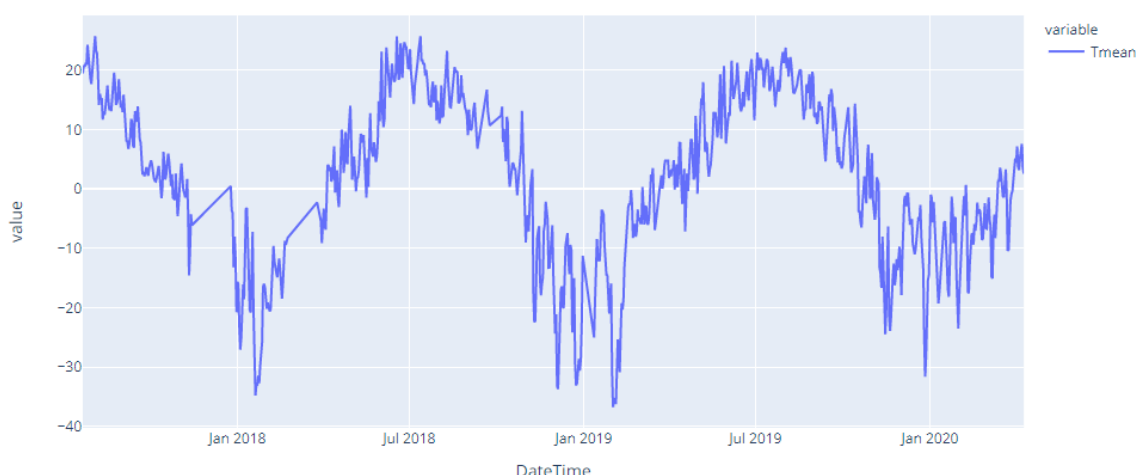


Рисунок 9 – Визуализация среднесуточной температуры

Как и было описано выше, метод SSA включает два отдельных этапа: этап декомпозиции и этап реконструкции. На первом этапе временной ряд декомпозируется на компоненты, а на втором этапе – восстанавливается исходный ряд и используется для прогнозирования новых точек данных.

На рисунке 10 иллюстрируется вклад каждого из компонентов. Видно, что большая часть переменных содержится в первом компоненте. В этом примере выбрано 365 компонентов.

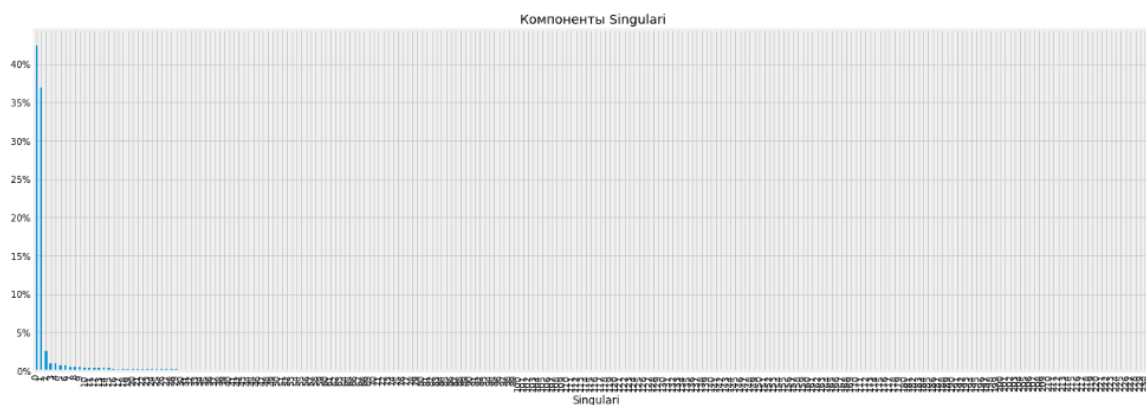


Рисунок 10 – Разложенный температурный ряд

Используя разложенные компоненты ряда восстанавливаем исходный временной ряд.

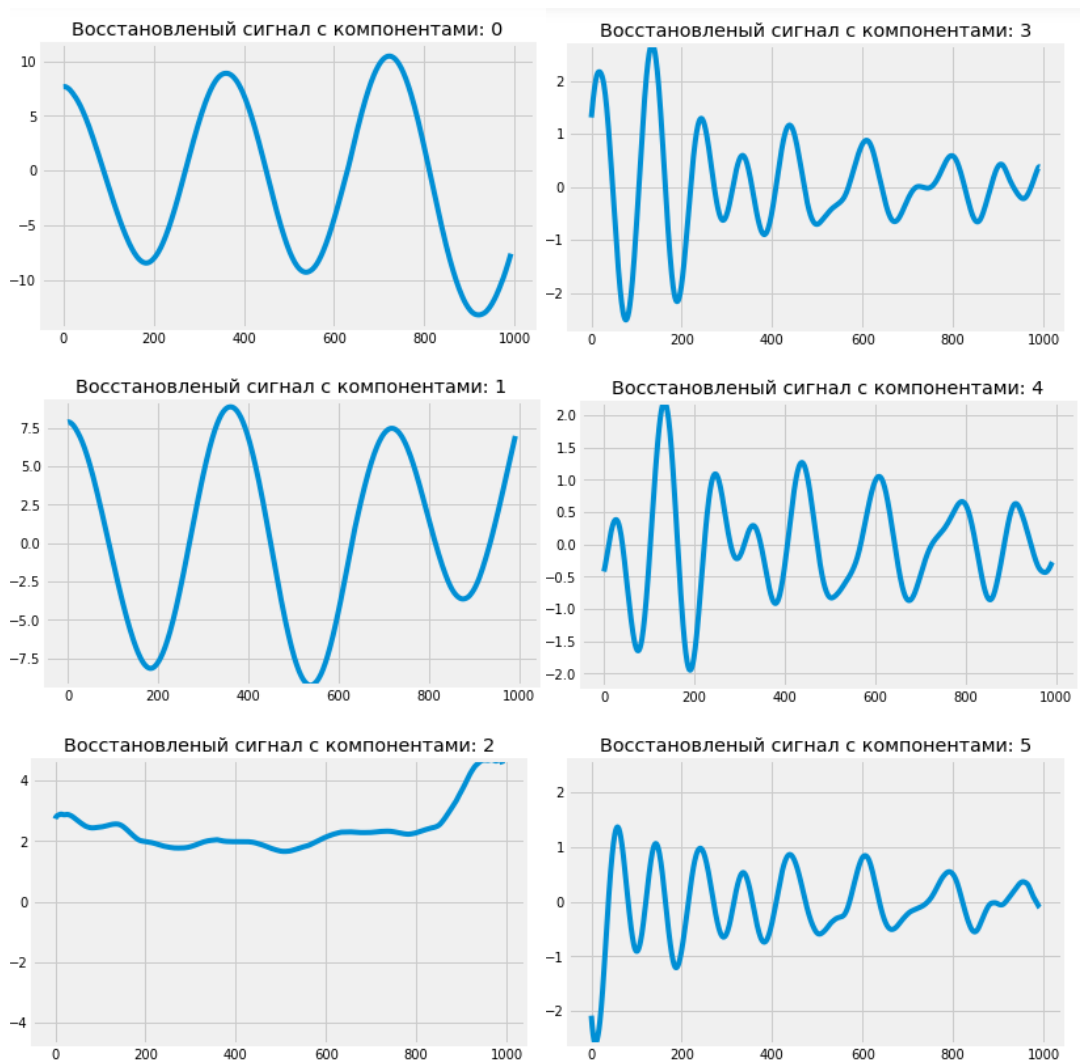


Рисунок 11 – Первые 6 компонент разложенного температурного ряда

На рисунке 12 приведен восстановленный температурный ряда. Чем больше компонентов используем (range), тем точнее будет восстанавливаемый сигнал, но возможны ошибки.

```
ssa.ts.plot(title='Оригинальный временной ряд'); # оригинальный график для сравнения
streams1 = [i for i in range(25)] # количество компонентов.
reconstructed1 = ssa.view_reconstruction(*[ssa.Xs[i] for i in streams1], names=streams1,
return_df=True)
```

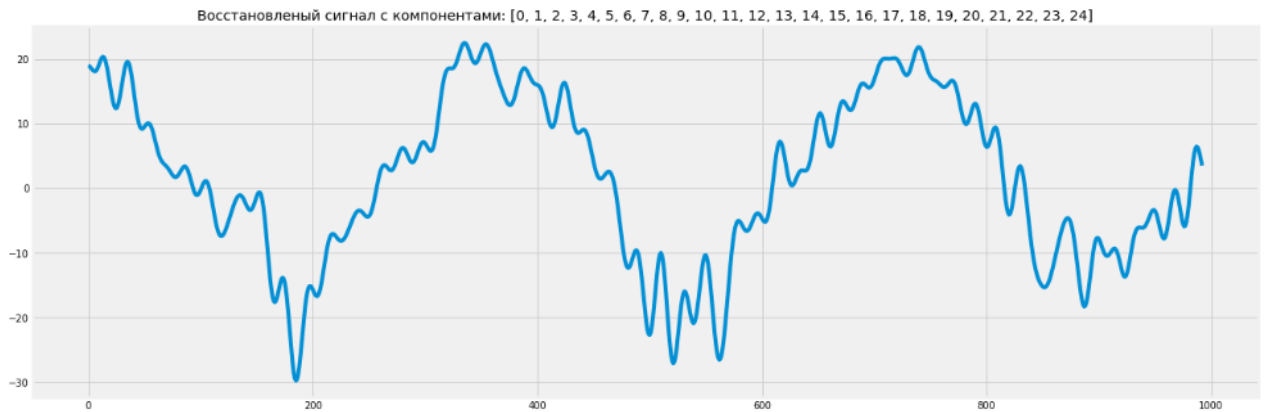


Рисунок 12 – Восстановленный ряд на основе разложенного

На рисунке 13 представлены совмещенные оригинальный и реконструированный временные ряды за весь период. Видно, что, в целом, модель реконструкции достаточно правдоподобно восстанавливает исходный температурный ряд.



Рисунок 13 – Совмещенные временные ряды

Количество прогнозируемых значений можно увеличить или уменьшить, изменив значение `steps`, в данном примере прогнозируется следующие 270 дней.

```
ssa.forecast(steps=270, singular_values=ts, plot=True)
```

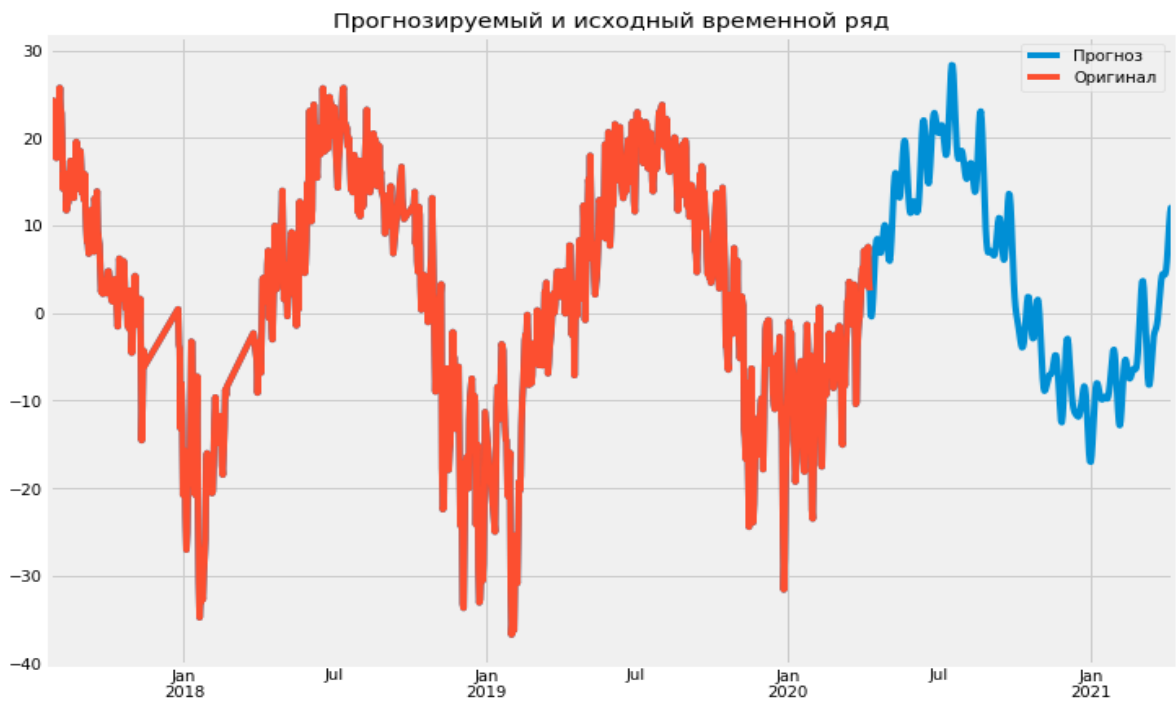


Рисунок 17 – Прогноз на следующие месяцы

Как видим, метод показал достаточно правдоподобный прогноз. Но модель недостоверно прогнозирует зимние сезоны.

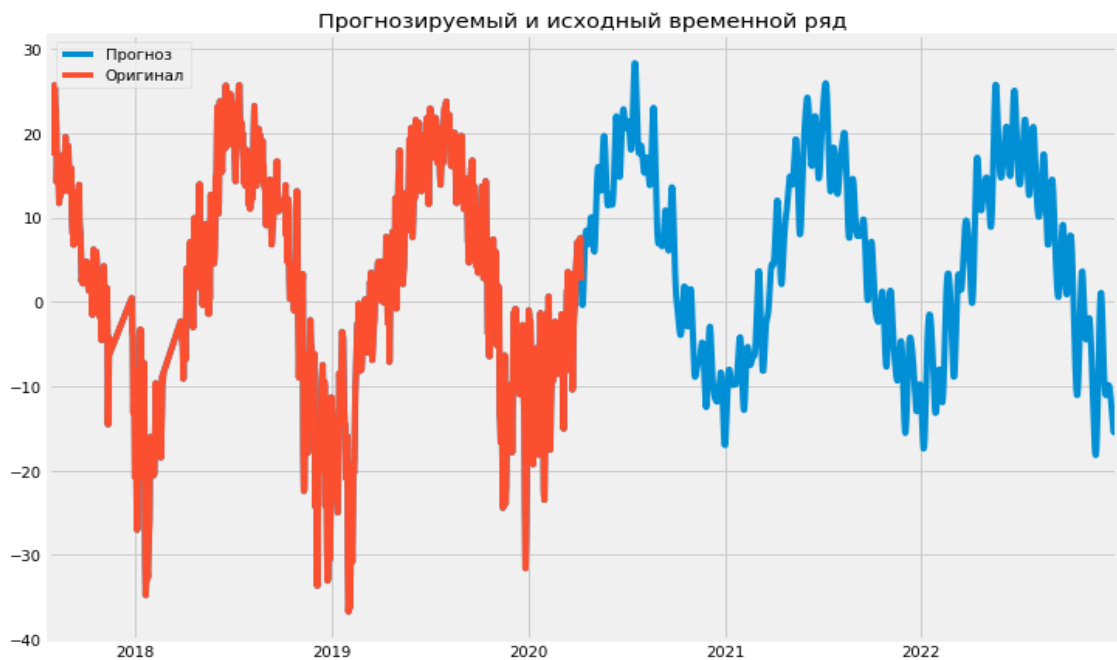


Рисунок 18 – Прогноз на 3 года

На рисунке 18 представлен прогноз и использованием 25 выделенных КОМПОНЕНТ.

### 3.2 Построение прогноза с помощью ARIMA

В программных экспериментах по реализации ARIMA использовались среднемесячные данные за весь период наблюдения. Частота может быть изменена: например, H - час, D - день, W - неделя, M - месяц и т.д.

Ниже приведен фрагмент программного кода для получения среднемесячных данных:

```
y = data105['Tmean'].resample('MS').mean() # MS - месяц, D - день
```

Результат выполнения кода представлен на рисунке 9.

```
DateTime
2017-07-01    20.591039
2017-08-01    16.711198
2017-09-01     8.238210
2017-10-01     1.822292
2017-11-01    -2.027126
2017-12-01    -8.771682
2018-01-01   -20.350205
2018-02-01   -14.975657
2018-03-01    -6.291295
2018-04-01     3.413716
2018-05-01     5.837447
2018-06-01    19.910579
2018-07-01    18.709187
2018-08-01    16.044498
2018-09-01    11.896020
2018-10-01     5.511829
2018-11-01   -10.063837
2018-12-01   -20.772218
2019-01-01   -13.442275
2019-02-01   -15.492891
2019-03-01    -1.165985
2019-04-01     3.083001
2019-05-01    10.045493
2019-06-01    16.250023
2019-07-01    19.421309
2019-08-01    18.038544
2019-09-01    10.635962
2019-10-01     4.499305
2019-11-01   -11.185783
2019-12-01   -10.089139
2020-01-01   -10.087288
2020-02-01    -7.281037
2020-03-01    -2.759580
2020-04-01     4.888799
Freq: MS, Name: Tmean, dtype: float64
```

Рисунок 19 – Значения метеопараметров

Для исследования была выбрана среднесуточная температура (Tmean) и общее количество значений составило 34 значений (2,5 года). И может быть изменена на другие параметры, что позволяет проводить прогнозирование любых метеорологических величин, которые есть в исследуемом файле.

На рисунке 10 приведен линейный график изменения температуры.

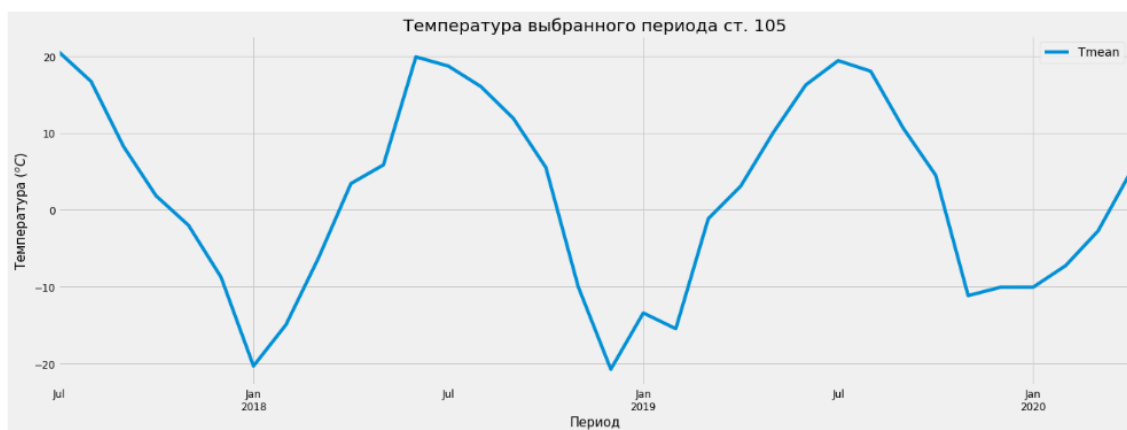


Рисунок 20 – Визуализация среднесуточной температуры

Как видно из рисунка 20, временные ряды имеют сезонность, температура повышается летом. Используя метод декомпозиции, разделим компоненты на тренд и сезонность. На рисунке 21 показан метод декомпозиции временных рядов.

```
decomposition = sm.tsa.seasonal_decompose(y, model='additive')
fig = decomposition.plot()
plt.show()
```

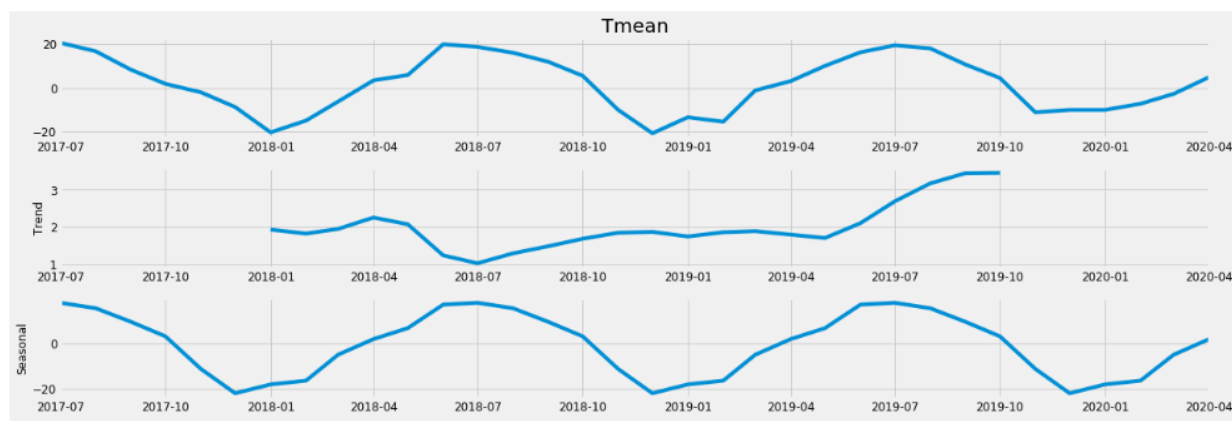


Рисунок 21 – Декомпозиция

Видно, что на рисунке 21 изменения температуры имеют сезонный характер.

На рисунке 22 представлен выбор параметров для выполнения ARIMA.

```

p = d = q = range(0, 2)

pdq = list(itertools.product(p, d, q))

seasonal_pdq = [(x[0], x[1], x[2], 12) for x in list(itertools.product(p, d, q))]
print('Examples of parameter combinations for Seasonal ARIMA...')
print('SARIMAX: {} x {}'.format(pdq[1], seasonal_pdq[1]))
print('SARIMAX: {} x {}'.format(pdq[1], seasonal_pdq[2]))
print('SARIMAX: {} x {}'.format(pdq[2], seasonal_pdq[3]))
print('SARIMAX: {} x {}'.format(pdq[2], seasonal_pdq[4]))

```

Рисунок 22 – Выбор параметров для выполнения модели ARIMA

На рисунках 23 и 24 представлены код и результаты поиска оптимального набора параметров.

```

warnings.filterwarnings("ignore")

for param in pdq:
    for param_seasonal in seasonal_pdq:
        try:
            mod = sm.tsa.statespace.SARIMAX(y, order=param, seasonal_order=param_seasonal,
                                             enforce_stationarity=False, enforce_invertibility=False)

            results = mod.fit()
            print('ARIMA({})x({})12 - AIC:{}'.format(param, param_seasonal, results.aic))
        except:
            continue

```

Рисунок 23 – Поиск оптимального набора параметров для модели

```

ARIMA(0, 0, 0)x(0, 0, 0, 12)12 - AIC:260.72959785072857 ARIMA(1, 0, 0)x(0, 0, 0, 12)12 - AIC:227.6161996991362
ARIMA(0, 0, 0)x(0, 0, 1, 12)12 - AIC:160.43168534979046 ARIMA(1, 0, 0)x(0, 0, 1, 12)12 - AIC:142.8013174682833
ARIMA(0, 0, 0)x(0, 1, 0, 12)12 - AIC:130.11080558622996 ARIMA(1, 0, 0)x(0, 1, 0, 12)12 - AIC:132.10858125984075
ARIMA(0, 0, 0)x(0, 1, 1, 12)12 - AIC:55.50284007463304 ARIMA(1, 0, 0)x(0, 1, 1, 12)12 - AIC:55.974310974420135
ARIMA(0, 0, 0)x(1, 0, 0, 12)12 - AIC:134.9719090104639 ARIMA(1, 0, 0)x(1, 0, 0, 12)12 - AIC:131.93316573780925
ARIMA(0, 0, 0)x(1, 0, 1, 12)12 - AIC:130.09558297980624 ARIMA(1, 0, 0)x(1, 0, 1, 12)12 - AIC:131.92533556126034
ARIMA(0, 0, 0)x(1, 1, 0, 12)12 - AIC:60.17182422023405 ARIMA(1, 0, 0)x(1, 1, 0, 12)12 - AIC:55.97435100281341
ARIMA(0, 0, 0)x(1, 1, 1, 12)12 - AIC:57.50289183885912 ARIMA(1, 0, 0)x(1, 1, 1, 12)12 - AIC:57.972125387957064
ARIMA(0, 0, 1)x(0, 0, 0, 12)12 - AIC:230.8798745872176 ARIMA(1, 0, 1)x(0, 0, 0, 12)12 - AIC:218.35893257517205
ARIMA(0, 0, 1)x(0, 0, 1, 12)12 - AIC:144.90940618091037 ARIMA(1, 0, 1)x(0, 0, 1, 12)12 - AIC:138.42558162266414
ARIMA(0, 0, 1)x(0, 1, 0, 12)12 - AIC:126.968606858854 ARIMA(1, 0, 1)x(0, 1, 0, 12)12 - AIC:128.96931850097928
ARIMA(0, 0, 1)x(0, 1, 1, 12)12 - AIC:51.6124419433836 ARIMA(1, 0, 1)x(0, 1, 1, 12)12 - AIC:53.204713752409866
ARIMA(0, 0, 1)x(1, 0, 0, 12)12 - AIC:136.94133191382767 ARIMA(1, 0, 1)x(1, 0, 0, 12)12 - AIC:133.9228883747644
ARIMA(0, 0, 1)x(1, 0, 1, 12)12 - AIC:126.84308089082447 ARIMA(1, 0, 1)x(1, 0, 1, 12)12 - AIC:128.54417860458673
ARIMA(0, 0, 1)x(1, 1, 0, 12)12 - AIC:61.07398150140802 ARIMA(1, 0, 1)x(1, 1, 0, 12)12 - AIC:57.98750639942898
ARIMA(0, 0, 1)x(1, 1, 1, 12)12 - AIC:53.60494407789349 ARIMA(1, 0, 1)x(1, 1, 1, 12)12 - AIC:55.19715813435254
ARIMA(0, 1, 0)x(0, 0, 0, 12)12 - AIC:224.16620404992756 ARIMA(1, 1, 0)x(0, 0, 0, 12)12 - AIC:219.34651405092734
ARIMA(0, 1, 0)x(0, 0, 1, 12)12 - AIC:137.74365450017604 ARIMA(1, 1, 0)x(0, 0, 1, 12)12 - AIC:138.72814450291236
ARIMA(0, 1, 0)x(0, 1, 0, 12)12 - AIC:138.6200416857866 ARIMA(1, 1, 0)x(0, 1, 0, 12)12 - AIC:134.97812276820156
ARIMA(0, 1, 0)x(0, 1, 1, 12)12 - AIC:51.9648002306911 ARIMA(1, 1, 0)x(0, 1, 1, 12)12 - AIC:52.64461864083965
ARIMA(0, 1, 0)x(1, 0, 0, 12)12 - AIC:139.94406960449464 ARIMA(1, 1, 0)x(1, 0, 0, 12)12 - AIC:133.28594846009858
ARIMA(0, 1, 0)x(1, 0, 1, 12)12 - AIC:135.6124239500209 ARIMA(1, 1, 0)x(1, 0, 1, 12)12 - AIC:134.73719644980557
ARIMA(0, 1, 0)x(1, 1, 0, 12)12 - AIC:57.06423133418398 ARIMA(1, 1, 0)x(1, 1, 0, 12)12 - AIC:52.644631432755986
ARIMA(0, 1, 0)x(1, 1, 1, 12)12 - AIC:53.9647100424718 ARIMA(1, 1, 0)x(1, 1, 1, 12)12 - AIC:54.63919819667174
ARIMA(0, 1, 1)x(0, 0, 0, 12)12 - AIC:214.93622998721608 ARIMA(1, 1, 1)x(0, 0, 0, 12)12 - AIC:214.32098829999873
ARIMA(0, 1, 1)x(0, 0, 1, 12)12 - AIC:133.39670797310941 ARIMA(1, 1, 1)x(0, 0, 1, 12)12 - AIC:134.5891939433888
ARIMA(0, 1, 1)x(0, 1, 0, 12)12 - AIC:122.73725620112458 ARIMA(1, 1, 1)x(0, 1, 0, 12)12 - AIC:124.72432507247278
ARIMA(0, 1, 1)x(0, 1, 1, 12)12 - AIC:46.754659253795786 ARIMA(1, 1, 1)x(0, 1, 1, 12)12 - AIC:48.42535472170316
ARIMA(0, 1, 1)x(1, 0, 0, 12)12 - AIC:134.2557092232241 ARIMA(1, 1, 1)x(1, 0, 0, 12)12 - AIC:130.5672635050471
ARIMA(0, 1, 1)x(1, 0, 1, 12)12 - AIC:123.49622774283462 ARIMA(1, 1, 1)x(1, 0, 1, 12)12 - AIC:125.48309250660759
ARIMA(0, 1, 1)x(1, 1, 0, 12)12 - AIC:57.02977165806186 ARIMA(1, 1, 1)x(1, 1, 0, 12)12 - AIC:53.43617910604918
ARIMA(0, 1, 1)x(1, 1, 1, 12)12 - AIC:48.8239915114085964 ARIMA(1, 1, 1)x(1, 1, 1, 12)12 - AIC:50.51539684290921

```

Рисунок 24 – Результаты поиска оптимального набора параметров

Выполненный поиск показывают, что (0, 1, 1)x(0, 1, 1, 12)12 - AIC:46.75 дает самое низкое значение, что является оптимальным вариантом.

Для проведения точности прогноза сравниваем прогнозируемые значения с реальным временным рядом начиная с 2018-09-01 и до конца исследуемых данных. Построенная модель показана на рисунке 26.

```

pred = results.get_prediction(start=pd.to_datetime('2018-09-01'), dynamic=False)
pred_ci = pred.conf_int()

ax = y['2017:'].plot(label='Исходный')
pred.predicted_mean.plot(ax=ax, label='Прогноз на шаг вперед', alpha=.7, figsize=(16, 7))

ax.fill_between(pred_ci.index,
                pred_ci.iloc[:, 0],
                pred_ci.iloc[:, 1], color='k', alpha=.2)

ax.set_xlabel('Date')
ax.set_ylabel('Furniture Sales')
plt.legend()

plt.show()

```

Рисунок 25 – Определение точности прогноза

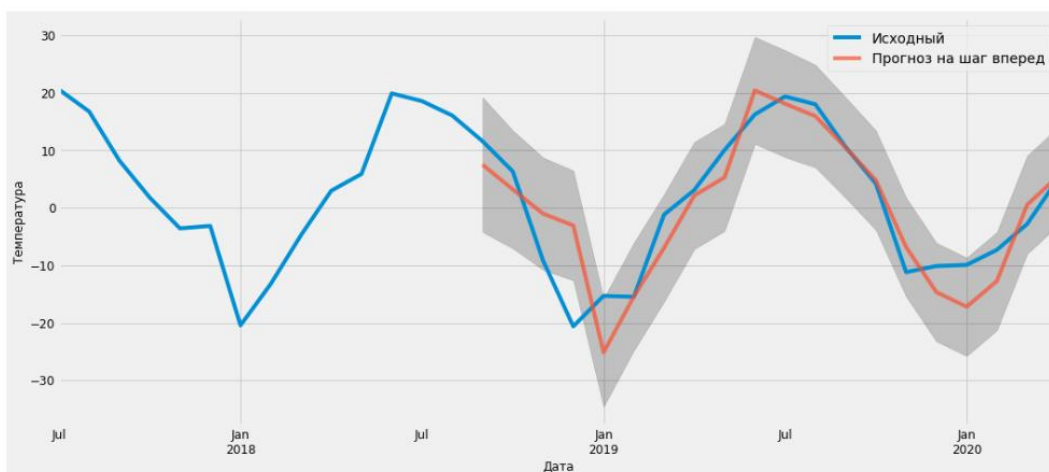


Рисунок 26 – Построение модели

На рисунке 26 наглядно видно, что модель, в целом, хорошо согласуется с реальными значениями, показывая, что восходящая тенденция начинается с середины года (лето) и убывающая к концу года (зима).

На рисунке 18 показано вычисление среднеквадратичной ошибки:

```

y_forecasted = pred.predicted_mean
y_truth = y['2018-09-01:']

mse = ((y_forecasted - y_truth) ** 2).mean()
print('The Mean Squared Error of our forecasts is {}'.format(round(mse, 2)))

```

Рисунок 27 – Среднеквадратическая ошибка

В результате среднеквадратичная ошибка MSE прогнозов составило 35,58. А ее квадратный корень RMSE=5,96.



На рисунке 28 представлен код прогнозирования до 2023 года на основе созданного алгоритма прогнозирования на 2019 год.

```
pred_uc = results.get_forecast(steps=30) # количество данных
pred_ci = pred_uc.conf_int()

ax = y.plot(label='observed', figsize=(14, 7))
pred_uc.predicted_mean.plot(ax=ax, label='Forecast')
ax.fill_between(pred_ci.index,
               pred_ci.iloc[:, 0],
               pred_ci.iloc[:, 1], color='k', alpha=.25)
ax.set_xlabel('Date')
ax.set_ylabel('Furniture Sales')

plt.legend()
plt.show()
```

Рисунок 28 – Прогнозирование до 2023 года

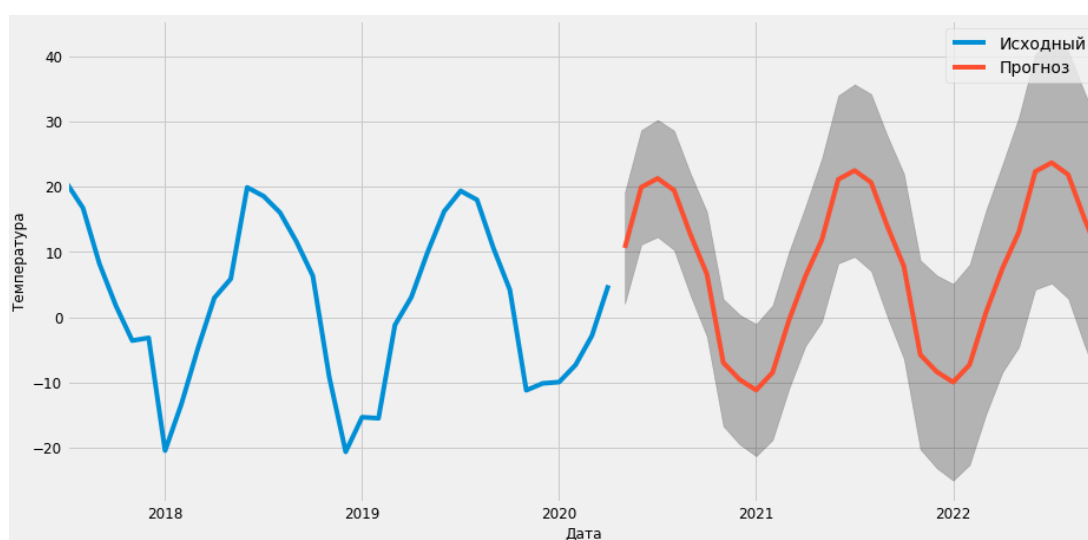


Рисунок 29 – Результат прогноза температуры до 2025 года

На графике видно, что модель четко видит сезонность временного ряда. Но по мере удаления по времени точность прогноза уменьшается. Также вызывают вопросы увеличения температуры в последующие годы, что логически невозможно.

### 3.3 Применение метода SSA к другим временным рядам

Прогнозирование других временных рядов выполнено по той же схеме, что и SSA. Лишь немного изменены периоды исследований значений переменных.

#### 3.4.1 Прогнозирование валютных курсов

Для исследования был выбран курс доллара к рублю.



Рисунок 30 – Результат прогнозирования

Из рисунка 30 видно, что укрепление рубля носит временный характер.

#### 3.4.2 Прогнозирование нефтяных котировок

Для прогнозирования выбраны цены на нефть марки Brent с 2000 года в долларах США за баррель.



Рисунок 31 – Результат прогнозирования

На рисунке 31 видно, что нефть марки Brent продолжит падать.

Выполненные прогнозы носят только математический характер и не отражают реальных событий в политике и экономике.

## 4 ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ

Результатом выполнения магистерской диссертации является программный продукт, служащий инструментом для научно-исследовательских вычислений с последующим анализом полученных данных. Данный программный продукт применяется в области метеорологии и климатологии.

Целевым рынком, на котором будет реализовываться разработка, являются коммерческие организации, научно-исследовательские институты, гидрометцентры предоставляющие услуги по прогнозированию погоды. Потенциальными потребителями могут являться аэропорты, городские, региональные и федеральные службы.

### 4.1 Организация и планирование работ

При организации процесса реализации конкретного проекта необходимо оптимально планировать занятость каждого из его участников и сроки проведения отдельных работ.

На данном этапе составляется полный перечень проводимых работ, и определяются их исполнители и оптимальная продолжительность. Результатом планирования работ является сетевой, либо реализации проекта. Наиболее удобным, простым и наглядным способом для этих целей является использование линейного графика. Для его построения составим перечень работ и соответствие работ своим исполнителям, продолжительность выполнения этих работ и сведем их в таблицу 2.

Таблица 2 – Перечень работ и продолжительность их выполнения

Этапы работы	Исполнители	Загрузка исполнителей
Постановка целей и задач, получение исходных данных	НР	НР – 100%
Составление ТЗ и календарного плана	НР, И	НР – 100% И – 10%
Подбор и изучение материалов по тематике	НР, И	НР – 10% И – 100%
Изучение функционала и работа с базовым примером реализации метода	НР, И	НР – 10% И – 100%
Применение метода к реальным данным, проверка и анализ результатов	НР, И	НР – 10% И – 100%
Составление ПЗ, оформление материалов ВКР	И	И – 100%

НР – научный руководитель.

И – исполнитель.

#### 4.1.1 Продолжительность этапов работ

Расчет продолжительности этапов работ осуществляется двумя методами:

- технико-экономическим;
- опытно-статистическим.

В данном случае используется опытно-статистический метод, который реализуется двумя способами:

- аналоговый;
- экспертный (вероятностный).

Для определения ожидаемого значения продолжительности работ  $t_{ож}$  применяется вероятностный метод – метод двух оценок  $t_{min}$  и  $t_{max}$ .

$$t_{ож} = \frac{3*t_{min} + 2*t_{max}}{5}, \quad (15)$$

где  $t_{min}$  – минимальная продолжительность работы, дн.;

$t_{max}$  – максимальная продолжительность работы, дн.;

Для выполнения перечисленных в таблице 2 работ требуются специалисты:

- инженер – исполнитель ВКР;

- научный руководитель.

Для построения линейного графика необходимо рассчитать длительность этапов в рабочих днях, а затем перевести в календарные дни. Расчет продолжительности выполнения каждого этапа ведется по формуле:

$$T_{РД} = \frac{t_{ОЖ}}{K_{ВН}} * K_{Д}, \quad (16)$$

где  $t_{ОЖ}$  – трудоемкость работы, чел/дн.;

$K_{ВН}$  – коэффициент выполнения работ ( $K_{ВН} = 1$ );

$K_{Д}$  – коэффициент, учитывающий дополнительное время на компенсации и согласование работ ( $K_{Д} = 1.2$ ).

Расчет продолжительности этапа в календарных днях ведется по формуле:

$$T_{КД} = T_{РД} * T_{К}, \quad (17)$$

где  $T_{РД}$  – продолжительность выполнения этапа в рабочих днях;

$T_{КД}$  – продолжительность выполнения этапа в календарных днях;

$T_{К}$  – коэффициент календарности.

Коэффициент календарности рассчитывается по формуле:

$$T_{К} = \frac{T_{КАЛ}}{T_{КАЛ} - T_{ВД} - T_{ПД}}, \quad (18)$$

где  $T_{КАЛ}$  – календарные дни ( $T_{КАЛ} = 365$ );

$T_{ВД}$  – выходные дни ( $T_{ВД} = 52$ );

$T_{ПД}$  – праздничные дни ( $T_{ПД} = 17$ ).

$$T_{К} = \frac{365}{365 - 52 - 17} = 1,233. \quad (19)$$

Таблица 3 – Трудозатраты на выполнение проекта

Этап	Исполнители	Продолжительность работ, дни			Трудоемкость работ по исполнителям чел.-дн.			
					Т <sub>рд</sub>		Т <sub>кд</sub>	
		<i>t<sub>min</sub></i>	<i>t<sub>max</sub></i>	<i>t<sub>ож</sub></i>	НР	И	НР	И
Постановка целей и задач, получение исходных данных	НР	2	5	3,2	3,84	0	4,74	0
Составление ТЗ и календарного плана	НР, И	2	4	2,8	3,36	0,34	4,14	0,41
Подбор и изучение материалов по тематике	НР, И	9	20	13,4	1,61	16,08	1,98	19,82
Изучение функционала и работа с базовым примером реализации метода	НР, И	15	25	19	2,28	22,8	2,81	28,11
Применение метода к реальным данным, проверка и анализ результатов	НР, И	14	25	18,4	2,21	22,08	2,72	27,23
Составление ПЗ, оформление материалов ВКР	И	7	13	9,4	0	11,28	0	13,91
<b>Итого:</b>				<b>66,2</b>	<b>13,3</b>	<b>72,58</b>	<b>16,39</b>	<b>89,48</b>

Таблица 4 – Линейный график работ

Этап	НР	И	Март			Апрель			Май			Июнь	
			10	20	30	40	50	60	70	80	90	100	110
1	4,74	0	■										
2	4,14	0,41	■	■									
3	1,98	19,83		■	■								
4	2,81	28,11			■	■	■	■					
5	2,72	27,22						■	■	■	■		
6	0	13,91										■	■

НР –



## 4.2 Расчет сметы затрат на выполнение проекта

В состав затрат на создание проекта включается величина всех расходов, необходимых для реализации комплекса работ, составляющих содержание данной разработки. Расчет сметной стоимости ее выполнения производится по следующим статьям затрат:

- материалы и покупные изделия;
- заработная плата;
- социальный налог;
- расходы на электроэнергию (без освещения);
- амортизационные отчисления;
- командировочные расходы;
- оплата услуг связи;
- арендная плата за пользование имуществом;
- прочие услуги (сторонних организаций);
- прочие (накладные расходы) расходы.

### 4.2.1 Расчет затрат на материалы

При планировании бюджета работы должно быть обеспечено полное и достоверное описание всех видов планируемых расходов, необходимых для его выполнения.

Разработка данного программного продукта не предполагает больших затрат на приобретение специальных материалов или оборудование. Основными затратами выступает покупка канцелярских принадлежностей. Результаты по данной статье были занесены в таблицу 5.



Таблица 5 – Расчет затрат на материалы

Наименование материалов	Цена за ед., руб.	Количество, шт	Сумма, руб.
Бумага «SvetoCopy» (А4, 80г/кв.м, 500 листов)	250	1	250
Комплект канцелярских принадлежностей	210	1	210
Картридж для принтера	1550	1	1550
<b>Итого:</b>			<b>2010</b>

С учетом ТЗР расходы на материалы составят:  $C_{\text{мат}} = 2010 * 1,05 = 2110,5$  руб.

#### 4.2.2 Расчет заработной платы

Данная статья расходов включает заработную плату научного руководителя и студента, а также премии, входящие в фонд заработной платы. Расчет основной заработной платы выполняется на основе трудоемкости выполнения каждого этапа и величины месячного оклада исполнителя.

Среднедневная заработная плата рассчитывается по формуле:

$$\text{Ср. з/плата} = \frac{\text{Месячный оклад}}{25,25}. \quad (20)$$

Расчеты затрат на основную заработную плату приведены в таблице 5.6. При расчете учитывалось, что в году 303 рабочих дня и, следовательно, в месяце 25,25 рабочих дня. Затраты времени на выполнение работы по каждому исполнителю брались из таблицы 3. Также был принят во внимание коэффициент, учитывающий коэффициент по премиям  $K_{\text{доп.ЗП}} = 1,188$ ,  $K_{\text{ПР}} = 1,1$  и районный коэффициент  $K_{\text{р}} = 1,3$ .  $K_{\text{И}} = 1,1 * 1,188 * 1,3 = 1,699$ .

Расчет основной заработной платы исполнителей системы выбирается на основе системы оплаты труда в ТПУ (для руководителя). Для исполнителя (студент) предусматривается расчет оплаты труда исходя из системы оплаты труда предприятия.

Таблица 6 – Затраты на заработную плату

Исполнитель	Оклад, руб./мес.	Среднедневная ставка, руб./раб.день	Затраты времени, раб.дни	Коэффициент	Фонд з/платы, руб.
НР	33664	1333,23	13,3	1,699	<b>30126,6</b>
И	9489	375,80	72,58	1,699	<b>46341,18</b>
<b>Итого:</b>					<b>76467,78</b>

#### 4.2.3 Расчет затрат на социальный налог

Затраты по этой статье составляют отчисления в социальные фонды.

Отчисления по заработной плате определяются по следующими формулами:

$$C_{\text{соц}} = K_{\text{соц}} * C_{\text{ЗП}}, \quad (21)$$

$$C_{\text{соц}} = K_{\text{соц}} * C_{\text{осн}}, \quad (22)$$

где  $K_{\text{соц}}$  – коэффициент, учитывающий размер отчислений из заработной платы. Данный коэффициент составляет 30% от затрат на заработную плату и включает в себя:

- отчисления в пенсионный фонд;
- на социальное страхование;
- на медицинское страхование.

Итак, отчисления из заработной платы составили:

$$C_{\text{соц}} = 0,3 * 76467,78 = 22940,33 \text{ руб.} \quad (23)$$

#### 4.2.4 Расчет затрат на электроэнергию

Данный вид расходов включает в себя затраты на электроэнергию при работе оборудования. Затраты на электроэнергию при работе оборудования для технологических целей рассчитываются по формуле:

$$C_{\text{эл.об}} = P_{\text{об}} * t_{\text{об}} * Ц_{\text{э}}, \quad (24)$$

где  $P_{\text{об}}$  – мощность, потребляемая оборудованием, кВт;

$Ц_{\text{э}}$  – тариф на 1 кВт час;  $Ц_{\text{э}} = 6,59 \text{ руб./квт}\cdot\text{час}$  (с НДС).

$t_{\text{об}}$  – время работы оборудования, час.

Время работы оборудования вычисляется на основе данных для  $T_{\text{рд}}$  таблицы 3 для инженера из расчета, что продолжительность рабочего дня равна 8 часов.

Мощность, потребляемая оборудованием, определяется по формуле:

$$P_{\text{об}} = P_{\text{ном}} * K_{\text{с}}, \quad (25)$$

где  $P_{\text{ном}}$  – номинальная мощность оборудования, кВт;

$K_{\text{с}}$  – коэффициент загрузки, зависящий от средней степени использования номинальной мощности. Для технологического оборудования малой мощности  $K_{\text{с}} = 1$ .

Затраты на электроэнергию для технологических целей приведены в таблице 7.

Таблица 7 – Затраты на электроэнергию для технологических целей

Наименование оборудования	Время работы оборудования $t_{об}$ , час	Потребляемая мощность $P_{об}$ , кВт	Затраты Эоб, руб.
Персональный компьютер	580,64*0,5	0,3	573,93
Струйный принтер HP	10	0,1	6,59
<b>Итого:</b>			<b>580,55</b>

#### 4.2.5 Расчет амортизационных расходов

В статье «Амортизационные отчисления» рассчитывается амортизация используемого оборудования за время выполнения проекта.

$$C_{AM} = \frac{N_A * Ц_{об} * t_{рф} * n}{F_D}, \quad (26)$$

где  $N_A$  – годовая норма амортизации единицы оборудования;

$Ц_{об}$  – балансовая стоимость единицы оборудования с учетом ТЗР. При невозможности получить соответствующие данные из бухгалтерии она может быть заменена действующей ценой, содержащейся в ценниках, прейскурантах и т.п.;

$F_D$  – действительный годовой фонд времени работы соответствующего оборудования, берется из специальных справочников или фактического режима его использования в текущем календарном году. При этом второй вариант позволяет получить более объективную оценку  $C_{AM}$ . Например, для ПК г. можно принять  $F_D = 303 * 8 = 2424$  часа;

$t_{рф}$  – фактическое время работы оборудования в ходе выполнения проекта, учитывается исполнителем проекта;

$n$  – число задействованных однотипных единиц оборудования.

$$\text{Для компьютера: } C_{AM} = \frac{0,5 * 50000 * 290,32 * 1}{2424} = 2994,22 \text{ руб.}$$

$$\text{Для принтера: } C_{AM} = \frac{0,4 * 12000 * 10 * 1}{300} = 160 \text{ руб.}$$

Итого начислено амортизации 3154,22 руб.

#### 4.2.6 Расчет прочих расходов

В статье «Прочие расходы» отражены расходы на разработку проекта, которые не учтены в предыдущих статьях.

Прочие расходы составляют 10 % от единовременных затрат на выполнение технического продукта и проводятся по формуле:

$$C_{\text{проч}} = (C_{\text{мат}} + C_{\text{ЗП}} + C_{\text{СОЦ}} + C_{\text{эл.об}} + C_{\text{ам}}) * 0,1. \quad (27)$$

$$C_{\text{проч}} = (2110,5 + 76467,78 + 22940,33 + 580,55 + 3154,22) * 0,1 = 10525,34 \text{ руб.} \quad (28)$$

#### 4.2.7 Расчет общей себестоимости разработки

Проведя расчет сметы затрат на разработку, можно определить общую стоимость разработки проекта «Алгоритмического и программного обеспечения для исследования временных рядов».

Таблица 8 – Смета затрат на разработку проекта

Статья затрат	Условное обозначение	Сумма, руб.
Материалы и покупные изделия	$C_{\text{мат}}$	2110,5
Основная заработная плата	$C_{\text{ЗП}}$	76467,78
Отчисления в социальные фонды	$C_{\text{СОЦ}}$	22940,33
Расходы на электроэнергию	$C_{\text{ЭЛ}}$	580,55
Амортизационные отчисления	$C_{\text{ам}}$	3154,22
Прочие расходы	$C_{\text{проч}}$	10525,34
<b>Итого:</b>		<b>115778,72</b>

Таким образом, затраты на разработку составили  **$C = 115778,72$  руб.**

#### 4.2.8 Расчет прибыли

Прибыль от реализации проекта в зависимости от конкретной ситуации (масштаб и характер получаемого результата, степень его определенности и коммерциализации, специфика целевого сегмента рынка и т.д.) может

определяться различными способами. В нашем проекте она составляет **17366,81 руб.** (15 %) от расходов на разработку проекта.

#### 4.2.9 Расчет НДС

НДС составляет 20% от суммы затрат на разработку и прибыли. В нашем случае это  $(115778,72 + 17366,81) * 0,2 = 133145,53 * 0,2 = 26629,11$  руб.

#### 4.2.10 Цена разработки НИР

Цена равна сумме полной себестоимости, прибыли и НДС, в нашем случае

$$C_{\text{НИР(КР)}} = 115778,72 + 17366,81 + 26629,11 = 159774,64 \text{ руб.}$$

### 4.3 Оценка экономической эффективности проекта

Экономическая эффективность проекта обусловлена острой необходимостью и потребностью вычислений метеорологических данных за максимально короткие сроки. Достижение необходимой производительности путем использования алгоритмов для анализа больших данных является важнейшей темой в условиях технического прогресса на сегодняшний день. Конечная система позволяет производить расчеты значительно быстрее обычных способов.

Также использование подобной системы в коммерческой и социальной сфере позволяет решать такие задачи, как оперативная обработка данных, анализ данных своевременно, и своевременно давать прогноз происходящим процессам в природе.

### 4.4 Вывод по разделу

На разработку данного проекта потребовалось 92 дней, для подробного описания трудозатрат был представлен календарный план проекта. Итоговый бюджет проекта, исходя из расчетов всех затрат, составил около 159774,64 руб.

Стоит также отметить, что в результате внедрения данной системы, предприятие улучшит процесс оказания услуг, что является основной целью организации, путем исключения из работы человеческого фактора и проведения автоматизации основных аналитических работ. Данные улучшения позволят снизить затраты на некоторые виды работ пользователей, что позволит направить сохраненное время для решения других задач.

## 5 Социальная ответственность

### 5.1 Введение

Разрабатываемый в рамках магистерской диссертации проект является программным продуктом, который предназначен для исследования метеорологических временных рядов и их прогнозирования. Работа над данным проектом проходит с использованием персонального компьютера. Взаимодействие с данной программой осуществляется с помощью персонального компьютера. Данный раздел посвящен анализу различных вредных и опасных факторов производственной среды для пользователей персональных компьютеров при работе с разработанным программным продуктом.

Основная работа при разработке проекта связана с использованием персонального компьютера, на котором выполняется анализ данных для прогнозирования временных рядов. Исследование и обработка данных, осуществляется с помощью ПК. Данный раздел предназначен для классификации различных опасных и вредных факторов рабочей зоны для пользователей персонального компьютера при работе над разработкой программного обеспечения.

В результате внедрения данной системы, предприятие улучшит процесс оказания услуг, что является основной целью организации, путем исключения из работы человеческий фактор и провести автоматизацию основных аналитических работ. Данные улучшения позволят снизить затраты на некоторые виды работ пользователей, что позволит направить сохраненное время для решения других задач.



## 5.2 Правовые и организационные вопросы обеспечения безопасности

### 5.2.1 Правовые нормы трудового законодательства

Сотрудник, постоянно работающий с ПЭВМ, сталкивается с различными негативными воздействиями на организм. Для минимизирования таких воздействий законодательно закреплено установление перерывов.

Согласно классификации видов трудовой деятельности с персональным компьютером, представленной в СанПиН 2.2.2/2.4.1340-03 [11], работу разработчика и пользователей с программным продуктом следует отнести к группе В, предполагающая работу в режиме диалога с компьютером. Согласно данной классификации, суммарное время регламентированных перерывов составляет от 30-120 минут в соответствии с категорией работ(ст. 108 ТК РФ)[12].

Работу разработчика и пользователей системы можно отнести к III категории работ, тогда, при 8-часовой рабочей смене, регламентированные перерывы следует устанавливать через 1,5 - 2,0 часа от начала рабочей смены и через 1,5 - 2,0 часа после обеденного перерыва продолжительностью 20 минут каждый или продолжительностью 15 минут через каждый час работы.

При 12-часовой рабочей смене регламентированные перерывы должны устанавливаться в первые 8 часов работы аналогично перерывам при 8-часовой рабочей смене, а в течение последних 4 часов работы, независимо от категории и вида работ, каждый час продолжительностью 15 минут.

Продолжительность рабочего дня не должна быть меньше указанного времени в договоре, но не больше 40 часов в неделю. Для работников до 16 лет – не более 24 часов в неделю, от 16 до 18 лет и инвалидов I и II группы – не более 35 часов.

Организация обязана предоставлять ежегодный отпуск продолжительностью 28 календарных дней. Дополнительные отпуска предоставляются работникам, занятым на работах с вредными или опасными условиями труда, работникам имеющими особый характер работы, работникам

с ненормированным рабочим днем и работающим в условиях Крайнего Севера и приравненных к нему местностях.

### 5.2.2 Организация рабочего места при работе с ПЭВМ

Правильная организация рабочего места пользователя ПЭВМ позволяет повысить эффективность работ и облегчить работу пользователей, сделать ее максимально удобной и безопасной. Сам разрабатываемый продукт не несет никакого влияния на организацию рабочей зоны.

Требования к помещениям для работы с ПЭВМ устанавливаются согласно СанПиН 2.2.2./2.4.1340-03[11]. К таким требованиям относятся:

При размещении рабочих мест с ПЭВМ расстояние между рабочими столами с видеомониторов, должно быть не менее 2,0 м, а расстояние между боковыми поверхностями видеомониторов - не менее 1,2 м.

- Экран монитора должен находиться от глаз пользователя на расстоянии 600-700 мм, но не ближе 500 мм.
- Конструкция рабочего стула (кресла) должна обеспечивать поддержание рациональной рабочей позы при работе на ПЭВМ, позволять изменять позу с целью снижения статического напряжения мышц шейно-плечевой области и спины для предупреждения развития утомления.

### 5.3 Производственная безопасность

Для того чтобы организовать необходимый уровень производственной безопасности для разработки программного обеспечения, а также при его использовании, нужно провести анализ возможных воздействий на человека вредных и опасных производственных факторов.

Производственный фактор считается *вредным*, если воздействие этого фактора на работника может привести к его заболеванию или к снижению работоспособности. Производственный фактор считается *опасным*, если его воздействие на работника может привести к его травме или к его летальному исходу [13].

Производственные факторы классифицируются по категориям:

- физические;
- химические;
- биологические;
- психофизиологические.

Таблица 9 — Вредные и опасные производственные факторы при выполнении работ за ПЭВМ [14].

Факторы (ГОСТ 12.0.003-2015)	Этапы работ			Нормативные документы
	Разработка	Изготовление	Эксплуатация	
1. Изменение показателей микроклимата.	+	+	+	СанПиН 2.2.4.548-96 [15]
2. Шум.	+	+	+	СанПиН 2.2.4.3359-16 [16]
3. Недостаточная освещенность рабочего места.	+	+	+	СП 52.13330.2016 [17] СанПиН 2.2.1/2.1.1.1278-03 [18]
4. Опасность поражения электрическим током.	+	+	+	ГОСТ 12.1.030-81 ССБТ [19] ТОИ Р-45-084-01 [20]
5. Короткое замыкание.	+	+	+	ГОСТ 12.1.030-81 ССБТ [19]
6. Умственное перенапряжение.	+	+	+	ГОСТ 12.0.003-2015 (ССБТ) [14]

### 5.3.1 Изменение показателей микроклимата

К основным показателям микроклимата в производственных помещениях относятся: температура воздуха, относительная влажность воздуха, температура поверхностей, скорость движения воздуха, интенсивность теплового облучения. Оптимальные параметры микроклимата на рабочих местах должны соответствовать величинам, приведенным в таблице 10, применительно к выполнению работ различных категорий в холодный и теплый периоды.

Категории работ разделяются исходя из энерготрат организма в ккал/ч (Вт). Работы, проводимые в сидячем положении и сопровождающиеся незначительным физическим напряжением, исходя из данных СанПиН 2.2.4.548-96 [15], относятся к категории работ Ia — работы с интенсивностью энерготрат до 120 ккал/ч. В данную категорию можно отнести работу разработчика ПО и его пользователей.

Таблица 10 – Оптимальные показатели микроклимата на рабочих местах [15]

Период года	Категория работ по уровню затрат, Вт	Температура воздуха, °С	Температура поверхности, °С	Относительная влажность воздуха, %	Скорость движения воздуха, м/с
Холодный	Ia (до 139)	22-24	21-25	60-40	0,1
Теплый	Ia (до 139)	23-25	22-26	60-40	0,1

В производственных помещениях при работе с ПК разработчики, и пользователи, сталкиваются с воздействием тепла, выделяемым на них вычислительной техникой, вспомогательными приборами и средствами освещения. Этот фактор (в случае плохой вентиляции и кондиционирования), является одним из основных вредных факторов производственной среды.

Для обеспечения оптимальной температуры воздуха в помещении, где происходит разработка производится проветривание. В теплое время года, в помещении используется кондиционер или вентилятор. В холодное время производится отопление помещения. Все перечисленные меры направлены на установление оптимальных показателей микроклимата на рабочем месте.

### 5.3.2 Шум на рабочих местах

Важным фактором, который влияет на самочувствие человека на работе, его трудоспособности является уровень шума в помещении. В помещениях, где разрабатывается программное обеспечение, и в помещениях, где предполагается использовать данное ПО, к источникам возникновения шума можно отнести: работу всех офисных электроприборов, наружный шум, например, уличный.

Требования к предельно допустимому уровню шуму на рабочих местах установлены в СН 2.2.4/2.1.8.562-96[21]. В таблице 11 представлены допустимые значения уровней звукового давления для рабочих мест в зависимости от трудовой деятельности.

Если уровень шума в помещении, в котором располагается рабочее место, не превышает 50 дБа, соответствует нормативным документам.

Таблица 11 – Предельно допустимые уровни звукового давления, уровни звука и эквивалентные уровни звука для основных наиболее типичных видов трудовой деятельности и рабочих мест [21].

Вид трудовой деятельности, рабочее место	Уровни звукового давления, дБ, в октавных полосах со среднегеометрическими частотами, Гц									Уровни звука и эквивалентные уровни (в дБА)
	31,5	63	125	205	500	1000	2000	4000	8000	
Творческая деятельность, руководящая работа с повышенными требованиями, научная деятельность, конструирование и проектирование, программирование, преподавание и обучение, врачебная деятельность. Рабочие места в помещениях дирекции, проектно-конструкторских бюро, расчетчиков, программистов вычислительных машин, в лабораториях для теоретических работ и обработки данных, приема больных в здравпунктах	86	71	61	54	49	45	42	40	38	50
Работа, выполняемая с часто получаемыми указаниями и акустическими сигналами; работа, требующая постоянного слухового контроля; операторская работа по точному графику с инструкцией; диспетчерская работа. Рабочие места в помещениях диспетчерской службы, кабинетах и помещениях наблюдения и дистанционного управления с речевой связью по телефону; машинописных бюро, на участках точной сборки, на телефонных и телеграфных станциях, в помещениях мастеров, в залах обработки информации на вычислительных машинах	96	83	74	68	63	60	57	55	54	65

### 5.3.3 Недостаточная освещенность рабочего места

Большую часть информации человек получает через зрения и в этом важную роль играет используемое освещение. Недостаток или неправильное распределение света повышает утомляемость, количество допускаемых ошибок, вызывает утомление глаз, провоцирует заболевания зрения, снижет

производительность труда. Требования к освещению регламентируются СП 52.13330.2016[17].

Работа за персональным компьютером, согласно СП 52.13330.2016[17], относится к зрительным работам высокой точности и соответствует третьему разряду зрительных работ. Требования к освещению рабочих мест представлены в таблице 12.

Таблица 12 – Требования к освещению рабочих мест

Допустимая средняя освещенность	от 300 до 500 лк
Оптимальная равномерность освещенности	0,4
Уровень блескости	прямая: не более 200кд/м <sup>2</sup>
	отраженная: не более 40 кд/м <sup>2</sup>
Максимальный коэффициент пульсации	не более 5%
Индекс цветопередачи	от 20 до 90 Ra

Рабочее место разработчика должно быть оборудовано естественным и искусственным освещением. Общее освещение при использовании люминесцентных светильников следует выполнять в виде прерывистых или сплошных линий светильников, расположенных сбоку от рабочих мест, параллельно линии зрения пользователя. При расположении компьютеров по периметру помещения линии светильников должны располагаться локализовано над рабочим столом ближе к его переднему краю, обращенному к оператору. Для рассеивания избыточного естественного света в помещении предусмотрены оконные занавеси.

### 5.3.3.1 Расчет искусственного освещения

Для расчета освещенности помещения необходимо определить расстояния между светильниками и высоты их подвеса над рабочей поверхностью [22].

Расчет искусственного освещения будет проводиться для помещения общей площадью  $S = 16 \text{ м}^2$  (длина помещения  $a = 4 \text{ м}$ , ширина  $b = 4 \text{ м}$ , высота  $h = 3 \text{ м}$ ).

В соответствии с СанПиН 2.2.2/2.4.1340-03[11] «Гигиенические требования к ПЭВМ и организации работы» освещенность на поверхности стола в зоне размещения рабочего документа должна быть 300 – 500 лк. Для расчета требуемый нормативный уровень освещенности  $E_H$  выберем равной 300 лк.

Для освещения данного помещения выбираем тип светильника ЛСП02-2х40 с люминесцентными лампами типа ЛБ40 (в количестве двух штук, мощностью по 40 Вт каждая и световым потоком  $\Phi_L = 3120 \text{ лм}$ ), габаритные размеры которого 158х280х1234 мм.

Количество светильников с люминесцентными лампами определяется по формуле:

$$N = \frac{E_H \cdot S_n \cdot z \cdot k_3}{\Phi_L \cdot N_{л.св} \cdot \eta}$$

где:  $E_H = 400 \text{ лк}$  – требуемый нормативный уровень освещенности рабочей поверхности;

$S = 16 \text{ м}^2$  – площадь освещаемого помещения;

$k_3 = 1,5$  – коэффициент запаса, учитывающий запыленность светильников в процессе эксплуатации;

$Z = 1,1$  – коэффициент неравномерности освещения для люминесцентных ламп;

$\Phi_L = 3120 \text{ лм}$  – световой поток одной люминесцентной лампы в светильнике;

$N_{л.св} = 2$  – количество ламп в одном светильнике;

$\eta$  – коэффициент используемого излучаемого светильником светового потока при освещении рабочей поверхности, зависящий от коэффициентов



отражения света от потолка ( $\rho_n=0,7$ ) и стен ( $\rho_c=0,5$ ), а также от индекса помещения (i).

Индекс помещения определяется по формуле:

$$i = \frac{a \cdot b}{h_p \cdot (a + b)} = \frac{4 \cdot 4}{0,73 \cdot (4 + 4)} = 2,74$$

где: a = 4 м – ширина помещения;

b = 4 м – длина помещения;

$h_p = 0,73$  м – расчетная высота.

Зная индекс помещения и коэффициенты отражения, из таблицы коэффициентов использования определяем  $\eta = 0,46$ .

$$N = \frac{400 \cdot 16 \cdot 1,1 \cdot 1,5}{3120 \cdot 2 \cdot 0,46} = 3,68$$

В результате количество светильников считаем равным 4, каждый из которых укомплектован двумя люминесцентными лампами.

Проверим фактический уровень освещенности по формуле:

$$E_{\text{факт}} = \frac{\Phi_{\text{л.факт}} \cdot N \cdot n_{\text{л.св}} \cdot \eta_{\text{исп}}}{S_n \cdot z \cdot k_z} = \frac{3120 \cdot 4 \cdot 2 \cdot 0,46}{16 \cdot 1,1 \cdot 1,5} = 434,91 \text{ лк.}$$

Полученное значение не критично отличается от требуемых норм.

#### 5.3.4 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 много внимания. Нормы электробезопасности на рабочем месте регламентируются СанПиН 2.2.2/2.4.1340-03[11], вопросы требований к защите от поражения электрическим током описаны в ГОСТ Р 12.1.019-2009 ССБТ [23]

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 поскольку отсутствуют: сырость, токопроводящая пыль, токопроводящие полы, высокая температура, возможность одновременного прикосновения человека к имеющим соединение с землей 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.

### 5.3.5 Короткое замыкание

Последствия короткого замыкания приводит к серьезным последствиям. Во-первых, достаточно часто оно сопровождается выходом из строя электроустановок и возникновением пожара. Во-вторых, из-за резкого увеличения силы тока в цепи отдельные части кабеля могут быть подвергнуты механическому воздействию, в результате чего появятся механические и термические повреждения. В-третьих, достаточно часто короткое замыкание сопровождается значительным падением напряжения в цепи или на отдельных ее участках. Это, в свою очередь, ведет к ухудшению работы электрооборудования и уменьшению их срока службы. В-четвертых, это явление

оказывает крайне негативное влияние на находящиеся поблизости приборы, провода и другое электрическое оборудование.

Защита от короткого замыкания включает в себя целый комплекс мер, исходным пунктом в которых является профилактика повреждений линий электропередач и оборудования. Кроме того, чтобы предотвратить возникновение пожара, используют специальные приборы – плавкиеставки, которые при замыкании сгорают и размыкают электрическую цепь[19].

### 5.3.6 Умственное напряжение

Главный инструмент пользователя или разработчика при работе с программным продуктом – это умственная деятельность. Такого рода деятельность является наиболее сложной формой труда, протекающей с повышенным эмоциональным напряжением. Как правило, повышенное эмоциональное напряжение ведет к нервно-психическим перегрузкам людей. Поэтому меры, позволяющие снизить воздействие вредного производственного фактора, которые регулируются СанПиН 2.2.2/2.4.1340-03[11] и являются важными в работе оператора ПК.

Работа разработчика, рассматриваемая в данной работе системы, можно отнести к группам А и Б, в то время, как деятельность пользователя, который будет использовать систему в профессиональной деятельности, относится к группе В. Категории туровой деятельности различаются по степени тяжести выполняемых работ. Для снижения воздействия рассматриваемого вредного фактора предусмотрены регламентированные перерывы для каждой группы работ – таблица 13.

Таблица 13 – Суммарное время регламентированных перерывов в зависимости от продолжительности работы, вида категории трудовой деятельности с ПК [11].

Категория работы с ПЭВМ	Уровень нагрузки за рабочую смену при видах работ с ПЭВМ			Суммарное время регламентированных перерывов, мин.	
	группа А, количество знаков	группа Б, количество знаков	группа В, ч	при 8-часовой смене	при 12-часовой смене
I	до 20 000	до 20 000	до 2	50	80
II	од 20 000	до 20 000	до 4	70	110
III	до 20 000	до 20 000	до 6	90	140

Для разработчика программного продукта были предусмотрены перерывы, составляющие 60 мин., что удовлетворяет показателям таблицы 13.

К мерам предупреждения умственного напряжения, можно отнести производственную гимнастику, изменение рабочей позы в процессе работы, а также смену деятельности. В последнее время, все больше предприятий организуют активные перерывы: прогулки на улице, тренажерные залы, посещение бассейна и т.д.

### 5.3.7 Минимизация воздействия вредных и опасных факторов

Обеспечить поддержку оптимальных условий освещений позволяют правила, описанные в СанПиН 2.2.2/2.4.1340-03[11]. Рабочее место должно быть оснащено естественным и искусственным освещением и соответствовать показателям. Для рассеивания избыточного естественного помещения рекомендуется использовать жалюзи, рулонные шторы и др. При искусственном освещении рекомендуется использовать люминесцентные лампы и лампы накаливания, а также светодиодные лампы.

Для поддержания оптимальных параметров микроклимата, рабочие места необходимо оснащать в теплый период системами кондиционирования воздуха, вентиляторами (вентиляторы или кондиционеры), а также проветривать помещение. В зимний период системами отопления. В помещения, где проходит работы на персональных компьютерах необходима влажная уборка, так как пыль способствует ухудшению работы оборудования, перегреву и др.

Для снижения уровня шума предусматриваются следующие меры: различные звукоизоляционные конструкции, например, звукоизоляционные перегородки; уплотнение притворов окон, дверей; укрытия для источников шума.

Когда основная работа подразумевает взаимодействие с компьютером долгое время, необходимо организовать перерывы на каждые 10-15 минут раз в 45-60 минут, это позволит снизить утомляемость и перенапряжение.

Для минимизации возможности поражения электрическим током необходимо проведение мероприятий:

1. При производстве монтажных работ необходимо использовать только исправный инструмент, аттестованный службой КИПиА.
2. С целью защиты от поражения электрическим током, возникающим между корпусом приборов и инструментом при пробое сетевого напряжения на корпус, корпуса приборов и инструментов должны быть заземлены
3. При включенном сетевом напряжении работы на задней панели должны быть запрещены
4. Все работы по устранению неисправностей должен производить квалифицированный персонал
5. Необходимо постоянно следить за исправностью электропроводки и оборудования.

#### 5.4 Экологическая безопасность

Исследование статистического метода и разработка программного обеспечения на ПК не являются экологически опасными работами, потому объекты, на которых будет производиться его создание и использование операторами ПК относятся к предприятиям пятого класса (практически не опасные), размер селитебной зоны для которых равен 50 м[13].

По ГОСТу Р 55102-2012[24] персональный компьютер содержит компоненты, которые должны быть отдельно собраны при выводе его из эксплуатации, представленные в таблице 14.

Таблица 14 – Перечень основных компонентов некоторых типов электротехнического и электронного оборудования

Компоненты	Тип оборудования		
	Системный блок и клавиатура персонального компьютера	Ноутбук	Монитор персонального компьютера
Металлы	+	+	+
Двигатель	+	+	
Пластик	+	+	+
ЖК-экран		+	+
Электропровода	+	+	+
Трансформатор	+	+	
Печатные платы	+	+	+
Пластики, содержащие замедлители горения		+	
ХИТ	+	+	
Внешние электропровода	+	+	+

Перечисленные выше компоненты требуют специальной утилизации, поскольку игнорирование таких работ может оказать значительный вред человечеству и окружающей среде. Для обеспечения правильной утилизации необходимо обращаться в специальные организации, имеющие государственные сертификаты, документации и профессиональное оборудование для осуществления услуг на утилизацию отходов оргтехники.

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 лет, после чего их необходимо сдавать на переработку в специальных пунктах приёма. Юридические лица обязаны сдавать лампы на переработку и вести паспорт для данного вида отходов. В соответствии с ГОСТ Р 52105-2003 [25] разработаны методы переработки таких отходов в зависимости от их типа, которые представлены в таблице 15.

Таблица 15 – Классификация ртутьсодержащих отходов по видам с указанием возможных способов их переработки

Наименование отходов	Обозначение отходов 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	ТМ, ВО
Шламы ртутьсодержащие от переработки люминесцентных ламп	Q9	5	7	1	ТМ, ВО
Примечание — Обозначение методов переработки РСО: ВО – высокотемпературный обжиг, ТМ – термические методы, ХМ – химико-металлургические методы.					

Также работа в помещениях включает в себе использование большого количества бумаги и другой макулатуры, которая также подлежит утилизации.

### 5.5 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.

#### 5.5.1 Минимизация возможности возникновения пожара

В первую очередь для того, чтобы предотвратить возникновение пожара в помещениях с ПЭВМ, должен осуществляться комплекс работ по обеспечению пожарной безопасности. Требования по обеспечению пожарной безопасности устанавливаются согласно СНиП 21-01-97 [26].

В помещении с ПЭВМ должен быть определен порядок обесточивания электрооборудования по окончании рабочего дня, порядок прохождения противопожарного инструктажа, помещения должны оснащаться углекислотными огнетушителями. Дополнительно к огнетушителям на каждые 200 м<sup>2</sup> площади рекомендуется иметь: грубошерстную ткань или войлок размером не менее 1х1 м, асбестовое полотно и пожарный стенд с емкостью для песка не менее 0,1 м<sup>3</sup>.

Система электропитания ЭВМ должна иметь блокировку, обеспечивающую отключение ее в случае остановки системы охлаждения и кондиционирования. Следует предусматривать пожарную автоматику и датчики пожарной сигнализации. Целесообразно разработать план эвакуации имущества и людей на случай пожара. Эффективность эвакуации людей может быть обеспечена достаточным количеством запасных выходов с хорошим освещением и многочисленными сигнальными огнями, и указателями на стенах, полах и т.д.

В случае возникновения пожара в помещении или в здании необходимо предпринять меры по эвакуации людей из него, используя для этого план эвакуации. При необходимости устранения очага возгорания в помещениях с



ПЭВМ использовать огнетушители. Требуется отметить, что в таких помещениях запрещается использовать химические пенные огнетушители.

### 5.5.2 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).

## 5.6 Вывод по разделу

В данном разделе рассматривались такие части как:

- Правовые и организационные вопросы обеспечения безопасности.
- Производственная безопасность.
- Экологическая безопасность.
- Безопасность в чрезвычайных ситуациях.

В производственной безопасности был проведён анализ выявленных опасных и вредных факторов при разработке и эксплуатации проектируемого решения, в результате было установлено, что все пункты, которые включены в данную часть соответствуют нормативным документам. Приведены меры по снижению вреда, оказывающие негативное воздействие на человека в процессе работы.

В экологической безопасности были рассмотрены вопросы утилизации отходов оргтехники, макулатуры. В ходе работы было установлено, что из самых распространённых источников ртутного загрязнения является вышедшие из эксплуатации люминесцентные лампы, которые также подлежат соответствующей утилизации. Непосредственно само проект и его применение не несут экологической опасности.

При рассмотрении безопасности в чрезвычайных ситуациях было выявлено следующее, что наиболее типичной ЧС для помещения, в котором производится выполнение работы, является пожар.

Правовые и организационные вопросы обеспечения безопасности включают в себя требования к организации рабочих мест пользователей, которые соответствуют ГОСТу.

## ЗАКЛЮЧЕНИЕ

В результате выполнения магистерской диссертации был проведен обзор методов PCA, SSA, ARIMA. Были освоены базовые методы SSA и ARIMA с их реализацией в интерактивной вычислительной среде Jupiter Notebook на языке программирования Python. Методом SSA и ARIMA были исследованы метеорологические данные с последующим построением прогнозной модели. Использовались данные, представленные Институтом мониторинга климатических и экологических систем СО РАН в г. Томске за 2,5 года наблюдений. Также с помощью SSA были проанализированные другие временные ряды.

Прогнозирование временного ряда методом ARIMA получилось достаточно успешное, за исключением увеличения температуры в последующие годы. Возможно, что такие аномалии вызваны недостаточным количеством анализируемых данных. Безусловно, построенные прогнозные данные не продемонстрировали эффективных результатов и не учитывают многих факторов при проведении реальных метеорологических прогнозов. Но их можно использовать для изучения других временных рядов, отличных от исследованных.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Broomhead DS, King GP. Extracting qualitative dynamics from experimental data. *Physica D* 20: 217-236 – URL: [https://www.researchgate.net/publication/216638258\\_Broomhead\\_DS\\_King\\_GP\\_Extracting\\_qualitative\\_dynamics\\_from\\_experimental\\_data\\_Physica\\_D\\_20\\_217-236](https://www.researchgate.net/publication/216638258_Broomhead_DS_King_GP_Extracting_qualitative_dynamics_from_experimental_data_Physica_D_20_217-236)
2. Голяндина Н., Некруткин В., Жиглявский А. Анализ структуры временных рядов 2001 года: SSA и смежные методы, [Электронный ресурс], режим доступа: <https://scholar.google.it/citations?user=1druVRYAAAAAJ&hl=it>
3. Pearson K., On lines and planes of closest fit to systems of points in space, *Philosophical Magazine*, (1901) 2, 559—572.
4. Principal Manifolds for Data Visualization and Dimension Reduction., Gorban, A.N., Kégl, B., Wunsch, D.C., Zinovyev, A. 2008.
5. Метод главных компонент (PCA), [Электронный ресурс], режим доступа: [http://help.prognosz.com/ru/mergedProjects/Lib/04\\_other\\_models/uimodelling\\_principalcomponents.htm](http://help.prognosz.com/ru/mergedProjects/Lib/04_other_models/uimodelling_principalcomponents.htm)
6. Package 'Rssa' – CRAN / [Электронный ресурс]. – Режим доступа: <https://cran.r-project.org/web/packages/Rssa/Rssa.pdf>
7. TIME SERIES ANALYSIS Forecasting and Control, [Электронный ресурс], режим доступа: [http://www.ru.ac.bd/stat/wp-content/uploads/sites/25/2019/03/504\\_05\\_Box\\_Time-Series-Analysis-Forecasting-and-Control-2015.pdf](http://www.ru.ac.bd/stat/wp-content/uploads/sites/25/2019/03/504_05_Box_Time-Series-Analysis-Forecasting-and-Control-2015.pdf)
8. Anaconda (Python distribution) [Электронный ресурс], режим доступа: <https://www.anaconda.com/>
9. Project Jupyter [Электронный ресурс], режим доступа: <https://jupyter.org/>
10. Система климатического мониторинга [Электронный ресурс], режим доступа: <http://mon.imces.ru/>
11. СанПиН 2.2.2/2.4.1340-03. Гигиенические требования к персональным электронно-вычислительным машинам и организации работы //

Электронный фонд правовой и нормативно-технической документации. 2003.  
URL: <http://docs.cntd.ru/document/901865498> (дата обращения: 22.05.2020 г.)

12. Трудовой кодекс Российской Федерации от 30.12.2001 N 197-ФЗ (ред. от 24.04.2020). // Электронный фонд правовой и нормативно-технической документации. 2007. URL: [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_34683/](http://www.consultant.ru/document/cons_doc_LAW_34683/) (дата обращения: 28.05.2020 г.)

13. СанПиН 2.2.1/2.1.1.1200-03 Санитарно-защитные зоны и санитарная классификация предприятий, сооружений и иных объектов // Электронный фонд правовой и нормативно-технической документации. 2007. URL: <http://docs.cntd.ru/document/902065388> (дата обращения: 22.05.2020 г.)

14. ГОСТ 12.0.003-2015 Система стандартов безопасности труда (ССБТ). Опасные и вредные производственные факторы. Классификация // Электронный фонд правовой и нормативно-технической документации. 2017. URL: <http://docs.cntd.ru/document/1200136071> (дата обращения: 20.05.2020 г.)

15. СанПиН 2.2.4.548-96. Гигиенические требования к микроклимату производственных помещений // Электронный фонд правовой и нормативно-технической документации. 1996. URL: <http://docs.cntd.ru/document/901704046> (дата обращения: 20.05.2020 г.)

16. СанПиН 2.2.4.3359-16 Санитарно-эпидемиологические требования к физическим факторам на рабочих местах // Электронный фонд правовой и нормативно-технической документации. 2007. URL: <http://docs.cntd.ru/document/420362948> (дата обращения: 28.05.2020 г.)

17. СП 52.13330.2016. Естественное и искусственное освещение // Электронный фонд правовой и нормативно-технической документации. 2011. URL: <http://docs.cntd.ru/document/456054197> (дата обращения: 20.05.2020 г.)

18. СанПиН 2.2.1/2.1.1.1278-03 Гигиенические требования к естественному, искусственному и совмещенному освещению жилых и общественных зданий // Электронный фонд правовой и нормативно-технической

документации. 2007. URL: <http://docs.cntd.ru/document/901859404> (дата обращения: 28.05.2020 г.)

19. ГОСТ 12.1.030-81 Система стандартов безопасности труда (ССБТ) Электробезопасность. Защитное заземление. Зануление. // Электронный фонд правовой и нормативно-технической документации. 2007. URL: <http://docs.cntd.ru/document/5200289> (дата обращения: 28.05.2020 г.)

20. ТОО Р-45-084-01. Типовая инструкция по охране труда при работе на персональном компьютере // Справочная правовая система КонсультантПлюс. URL: [http://www.consultant.ru/document/cons\\_doc\\_LAW\\_79762/8b2ed343ae4bcf0636cse936afa1156fb10b78ae/](http://www.consultant.ru/document/cons_doc_LAW_79762/8b2ed343ae4bcf0636cse936afa1156fb10b78ae/) (дата обращения: 21.05.2020 г.)

21. СН 2.2.4/2.1.8.562-96 Шум на рабочих местах, в помещениях жилых, общественных зданий и на территории жилой застройки. Санитарные нормы // Электронный фонд правовой и нормативно-технической документации. 2007. URL: <http://docs.cntd.ru/document/901703278> (дата обращения: 28.05.2020 г.)

22. Назаренко О. Б. Безопасность жизнедеятельности: РАСЧЁТ ИСКУССТВЕННОГО ОСВЕЩЕНИЯ. Методические указания / О. Б. Назаренко; Томский политехнический университет. – Томск: Изд-во Томского политехнического университета, 2008. – 20 с.

23. ГОСТ Р 12.1.019-2009. Система стандартов безопасности труда (ССБТ). Электробезопасность. Общие требования и номенклатура видов защиты // Электронный фонд правовой и нормативно-технической документации. 2009. URL: <http://docs.cntd.ru/document/1200080203> (дата обращения: 22.05.2020 г.)

24. ГОСТ Р 55102-2012. Ресурсосбережение. Обращение с отходами. Руководство по безопасному сбору, хранению, транспортированию и разборке отработавшего электротехнического и электронного оборудования, за исключением ртути содержащих устройств и приборов // Электронный фонд правовой и нормативно-технической документации. 2012. URL: <http://docs.cntd.ru/document/1200104723> (дата обращения: 25.05.2020 г.)

25. ГОСТ Р 52105-2003 Ресурсосбережение. Обращение с отходами. Классификация и методы переработки ртутьсодержащих отходов. Основные положения. // Электронный фонд правовой и нормативно-технической документации. 2007. URL: <http://docs.cntd.ru/document/1200032452> (дата обращения: 28.05.2020 г.)

26. СНиП 21-01-97. Пожарная безопасность зданий и сооружений // Информационно-правовое обеспечение. 1997 г. URL: <http://base.garant.ru/2305928/> (дата обращения: 25.05.2020 г.)



Приложение А

(обязательное)

Раздел ВКР, выполненный на иностранном языке

**Раздел 1 Review of time series prediction models and methods**

Студент:

<b>Группа</b>	<b>ФИО</b>	<b>Подпись</b>	<b>Дата</b>
8ВМ82	Докторов Максим Георгиевич		

Руководитель ВКР

<b>Должность</b>	<b>ФИО</b>	<b>Ученая степень, звание</b>	<b>Подпись</b>	<b>Дата</b>
Доцент ОИТ ИШИТР	Ботыгин И. А.	к.т.н., доцент		

Консультант – лингвист отделения ОИЯ ШБИП:

<b>Должность</b>	<b>ФИО</b>	<b>Ученая степень, звание</b>	<b>Подпись</b>	<b>Дата</b>
Доцент ОИЯ ШБИП	Аксёнова Н. В.	к.филол.н.		

# 1. Review of time series prediction models and methods

## 1.1 Basic concepts

A time series is a sequence of observations for a certain period. Thus, when working with time series data, order is important. In particular, values in a time series express a time dependence. Measurements can be taken periodically (hour, day, week, month, year) or at any other time interval.

Typically, time series data have two important properties:

- Data are measured sequentially and at the best at equal intervals in time.
- Each time unit has no more than one data dimension.

In addition, when predicting time series, there are usually two targets:

- Firstly, you need to define templates that explain the behavior of time series.
- Secondly, we use these models to predict new values.

Most time series usually have at least one of these three types of patterns: trend, seasonality and/or cycles.

The trend describes the general behavior of a time series, which is shown in Figure 1. If a time series shows an increase over time (1), it tends to increase. If it describes a decrease (2) instead, it tends to decrease.

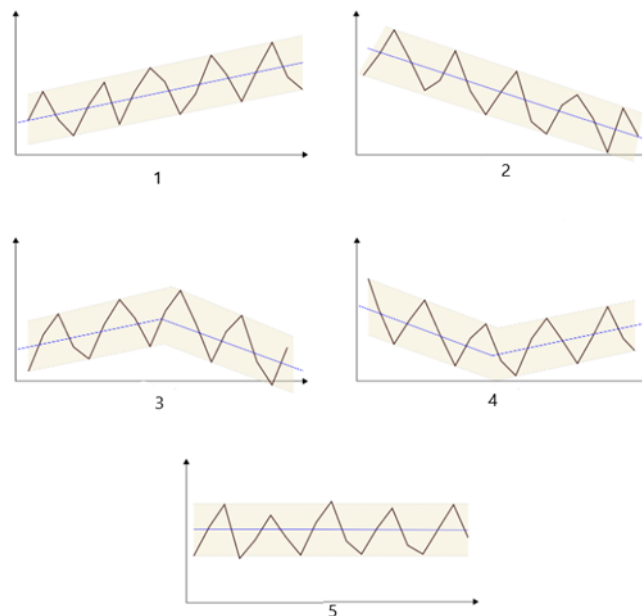


Figure 1 – Trends

Trend can also change direction. It may be an upward or downward trend and vice versa (3.4). Finally, a stationary or horizontal trend (5) defines a time series that has neither positive nor negative long-term patterns.

Seasonality is any fluctuation (change) in a time series.

These changes can be the time of year (e.g. winter or summer) or the time of day or week. Seasonality always has fixed frequencies. That is, the season always starts and ends at the same period of the week, year or other periods.

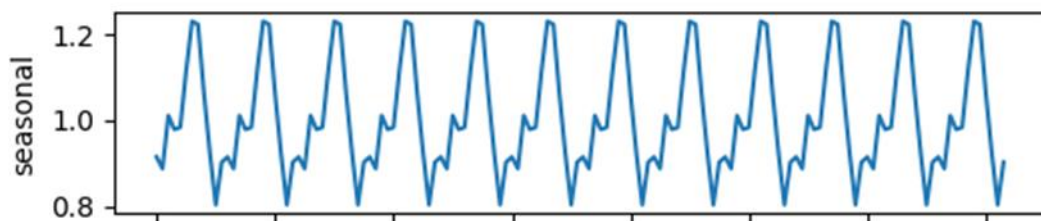


Figure 2 – Seasonality

For example, a data center as an example. If we look at the cooling system as the primary source of power consumption, it is only natural that power costs will increase in summer and power consumption may decrease in winter.

Finally, the cycle in time series is a kind of change that is not related to seasonal factors. These are increases and decreases with non-fixed values, which may last longer than a calendar year. Cycles are not repeated. Usually they are the result of external factors, which makes them difficult to predict.

It is important to note that not all time series are predictable. More specifically, some of them are not predictable. Such time series are difficult, if not impossible, to predict, because future dynamics are equally likely to differ sharply.

Prediction of time series is the use of statistical methods to predict future trends based on past data.

It is similar to other statistical theories of learning, such as controlled or uncontrolled learning. However, time series forecasting has many nuances that distinguish it from conventional machine learning.

Many companies are learning about time series forecasting as a way of choosing the best solution. Take the hotel as an example. If a manager has a clear idea of how many guests to expect next summer, they can use this information to better manage

their staff, increase their budget or expand their facility. Similarly, a confident understanding of future events can help a wide range of industries and problems, from traditional agriculture to logistics and much more.

## 1.2 Principal Component Analysis

The idea behind PCA is to convert a dataset of correlated variables into uncorrelated components (called master components) through linear combinations of source variables. Components are ordered by the number of deviations. In other words, the first component is a linear combination that considers the largest possible deviation. The second component is a linear combination that considers the largest possible variance of the remaining variation (deviation not fixed by component one), being orthogonal to the first component, etc. The point cloud of source variables is not spherical, most of the variations are reflected in the first few major components, which makes PCA a suitable method for this:

Reducing the data size;

- PCA is mainly used as a tool for data analysis and for creating predictive models;
- Find important attributes/variables (object selection in multidimensional data);
- PCA is used for finding patterns in data;
- Noise filtration.

### 1.2.1 Description of PCA method

Given matrix  $X$ , which corresponds to  $n$  observations with  $d$  characteristics, and input  $k$ , the main objective of PCA is to decompose matrix  $X$  into two smaller matrices,  $Z$  and  $W$ , such that  $X = Z \cdot W$ , where  $Z$  has dimensions  $n \cdot k$  and  $W$  has dimensions  $k \cdot d$ . Each row of  $Z$  is a factor loading. Each row of  $W$  is called the principal component.

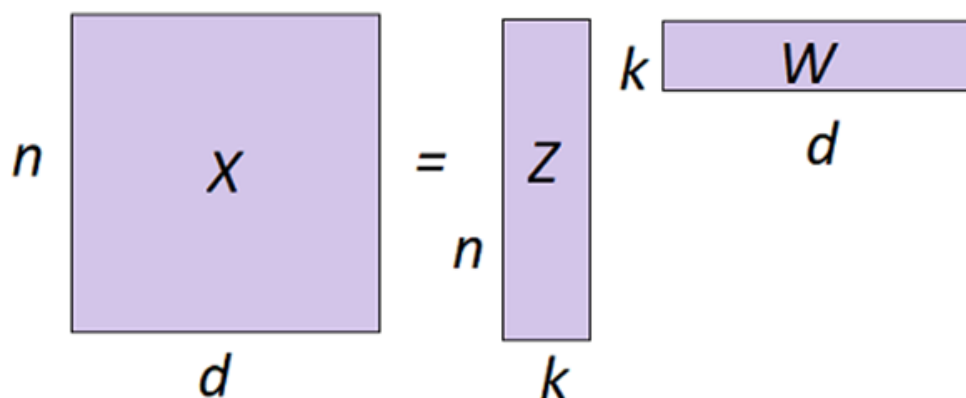


Figure 3 – PCA decomposes matrix  $X$  into two smaller matrices  $Z$  and  $W$

In PCA, it minimizes the square error of the following target function:

$$f(W, Z) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^d (w_j^T z_i - x_{ij})^2 = \frac{1}{2} \|ZW - Z\|_F^2 \quad (1)$$

There are three general approaches to the PCA solution:

### 1. Singular Value Decomposition (SVD)

Firstly, this approach uses singular decomposition (SVD) to find an orthogonal  $W$ . Then, it uses orthogonal  $W$  to calculate  $Z$  as follows:

$$X = U \Sigma W^T, \quad (2)$$

$$\nabla_Z f(Z) = Z \cdot W \cdot W^T - X \cdot W^T \rightarrow Z = X \cdot W^T (W \cdot W^T)^{-1} = X \cdot W^T, \quad (3)$$

### 2. Alternating Minimization

This is an iterative approach, which alternates between:

Fixing  $Z$  and finding optimal values for  $W$ :

$$\nabla_W f(W) = Z^T \cdot Z \cdot W - Z^T \cdot X, W = (Z^T \cdot Z)^{-1} \cdot (Z^T \cdot X), \quad (4)$$

Fixing  $W$  and finding optimal values for  $Z$ :

$$\nabla_Z f(Z) = Z \cdot W \cdot W^T - X \cdot W^T, Z = X \cdot W^T \cdot (W \cdot W^T)^{-1}, \quad (5)$$

### 3. Stochastic Gradient Descent

This is an iterative approach when the  $X$ -matrix is very large. At each iteration, it chooses a random example  $i$  and shows  $j$  and updates  $W$  and  $Z$ :

$$\text{Set } \omega^j \text{ to } \omega^j - \alpha^t \nabla_{\omega^j}(\omega^j, z_i, x_{ij}), \quad (6)$$

$$\text{Set } z_i \text{ to } z_i - \alpha^t \nabla_{z_i}(\omega^j, z_i, x_{ij}), \quad (7)$$

### 1.3 Singular spectrum analysis

Singular Spectrum Analysis (SSA) has been a rapid evolving method for time series analysis since the second half of the 20th century. SSA can be called the analysis of the principal components for time series. In the future, SSA may become one of the standard methods for time series analysis and signal processing. SSA can solve a wider range of tasks than PCA.

The challenge of SSA is to decompose a source series into the sum of a small number of independent and interpreted components, such as slowly changing trends, oscillating components and unstructured noise.

SSA can be used to find trends, smooth out, extract seasonality, extract cycles with small and large periods simultaneously, extract periodicity with different amplitudes, identify complex trends and periodicity simultaneously, find structure in short time series and detect the point of change.

The paper describes the use of SSA for time series forecasting. The algorithm is based on singling out a certain set of its principal components from the time series under study and then building a forecast for the selected period.

#### 1.3.1 Basic SSA

The analysis consists of four steps:

The first stage: decomposition. The decomposition stage consists of two stages: embedding and singular decomposition (SVD).

Step 1: Embedding

The embedding procedure maps the original time series into a sequence of lagged vectors of size  $L$  by forming  $K = N - L + 1$  lagged vectors  $X_i = (x_i, \dots, x_{i+L-1})^T, i = 1, \dots, K$ . The trajectory matrix of the time series F:

$$X = [X_1: \dots : X_K] = \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_K \\ x_2 & x_3 & x_4 & \cdots & x_{K+1} \\ x_3 & x_4 & x_5 & \cdots & x_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & x_{L+2} & \cdots & x_N \end{pmatrix}, \quad (8)$$

The trajectory matrix is a Hankel matrix because X has the same elements on diagonals.

## Step 2. Singular value decomposition (SVD)

From matrix  $X$ , define the covariance matrix  $XX^T$ . The SVD of  $XX^T$  provides a set of  $L$  eigenvalues in the decreasing order of magnitude ( $\lambda_1 \geq \lambda_2 \geq \dots \lambda_L \geq 0$ ) and the corresponding eigenvectors  $U_1, U_2, \dots, U_L$ . Then the SVD of the trajectory matrix  $X$ :

$$X = X_1 + \dots + X_L, \quad (9)$$

where  $X_i = \sqrt{\lambda_i} \cdot U_i \cdot V_i^T$ . The matrices  $X_i$  are elementary, that is, they have rank 1.  $V_i = \frac{X^T U_i}{\sqrt{\lambda_i}}, i, \dots, L,$  (equivalent to the  $i$  factor vector of  $XX^T$ ). The triple  $(\sqrt{\lambda_i} \cdot U_i \cdot V_i^T)$  is referred to as the  $i$  eigentriple.

### Second Stage: Reconstruction

This stage is subdivided into two steps: grouping and diagonal averaging.

### Step 3. Grouping

The grouping procedure partitions the set of indices  $\{1, \dots, L\}$  into  $r$  disjoint subsets. Let  $I = \{i_1, i_2, \dots, i_r\}$  be a group of  $r$  selected eigentriples and  $X_I = X_{i_1} + X_{i_2} + \dots + X_{i_r}$ , where  $X_I$  is related to the signal of  $F$  while the remaining  $(L - r)$  eigentriples denotes the time series related to the error term.

### Step 4. Diagonal Averaging

At the last step of the basic algorithm, each matrix of the grouped decomposition is translated into a new length row  $N$ .

Let  $Y$  be some  $L \cdot K$  matrix with elements  $y_{ij}$ , where  $1 \leq i \leq L, 1 \leq j \leq K$ . Let's put  $L^* = \min(L, K), K^* = \max(L, K)$  and  $N = L + K - 1$ . Let  $y_{ij}^* = y_{ij}$ , if  $L \leq K$ , and  $y_{ij}^* = y_{ij}$ - different. Diagonal averaging translates the  $Y$  matrix into a row  $g_0, \dots, g_{N-1}$  by formula 10.

$$g_k = \begin{cases} \frac{1}{k+1} \sum_{m=1}^{k+1} y_{m, k-m+2}^*, & 0 \leq k < L^* - 1 \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y_{m, k-m+2}^*, & L^* - 1 \leq k < K^* \\ \frac{1}{N-k} \sum_{m=k-K^*+2}^{N-K^*+1} y_{m, k-m+2}^*, & K^* \leq k < N \end{cases}, \quad (10)$$

The expression corresponds to the averaging of matrix elements along the "diagonals"  $i + j = k + 2$ : choice  $k = 0$  gives  $g_0 = y_{11}$ , for  $k = 1 - g_1 =$



$\frac{y_{12}y_{21}}{2}$ , etc. Note that if the matrix  $Y$  is a trajectory matrix of some series  $(h_0, \dots, h_{N-1})$  (if the matrix  $Y$  is a Hannelian), then  $g_i = h_i$  for all  $i$ .

Applying diagonal averaging to the resulting matrices  $X_{I_k}$  get series  $\tilde{F}^{(k)} = (\tilde{f}_0^{(k)}, \dots, \tilde{f}_{N-1}^{(k)})$  and, original series  $(f_0, \dots, f_{N-1})$  is expanded to the sum of  $m$  series:

$$f_n = \sum_{k=1}^m \tilde{f}_n^{(k)}, \quad (11)$$

From the above description we can see that the main parameter of the basic algorithm of the SSA method is the window length  $L$ . The choice of window length depends on the task being solved and preliminary information about the analyzed time series. In general, there are no universal rules and unconditional recommendations for window length selection [9].

## 1.4 PCA and SSA

When comparing SSA to PCA, the result of PCA is basically an unstructured approximation (in terms of time structure), while SSA takes the time structure into account.

For example, if we have  $m$  time series with lengths  $N, m < N$ , then PCA provides only  $m$  components. For  $m = 1$ , it makes no sense to use PCA; for  $m = 2$ , the two components may not be enough even to try to decompose into trend, oscillations and noise.

Another example is related to the decomposition into signal and noise when the signal is described by several SSA components (it is executed if the time series is well approximated by the final sum of polynomial, exponent and sinusoid products).

For example, let the time series of a system consist of noisy sinusoids with a small signal-to-noise ratio (SNR). PCA does not help to extract a signal for any length of time series  $N$ . And SSA applies SVD (singular decomposition; the same mathematical tool as in PCA) to the trajectory matrix. For sufficiently large  $L$  and  $N$ , the SSA can approximately extract the signal.

The same effect can be achieved when time series consist of time components such as trend and oscillations. PCA does not work for extracting one of the components. The SSA can do this with an SVD. Thus, the PCA usually does not work for the time series and can be replaced by an SSA.

## 1.5 ARIMA

One of the most common methods used in time series forecasting is known as the ARIMA model. According to this methodology, time series is the implementation of some random process. The task of modeling is to restore characteristics of a random process on its implementation. In general, such restoring is impossible, but it is quite feasible with some restrictions on class processes being under study, most often being in practice.

George Box and Gwilym Jenkins presented an integrated autoregressive model - the moving average or ARIMA in 1970 in their book "Time series analysis: forecasting and control" [29].

Any "non-seasonal" time series that show patterns and are not random white noise can be simulated with ARIMA models.

ARIMA model is characterized by 3 terms:  $p$ ,  $d$ ,  $q$ :

If the time series has seasonal patterns, you need to add the seasonal model and it will become SARIMA.

### 1.5.1 Basic ARIMA

The autoregressive model of the order  $p$  is described as well as shows the dependence of the current period value on the previous values of  $p$  periods:

$$AR(p): y_t = c + \varphi_1 y_{(t-1)} + \varphi_2 y_{(t-2)} + \dots + \varphi_p y_{(t-p)} + \varepsilon_t, \quad (12)$$

The Moving Average Order Model  $q$  is described as it shows the dependence of the current  $q$  perio value on the previous prediction errors:

$$MA(q): y_t = c + \varepsilon_t + \theta_1 \varepsilon_{(t-1)} + \theta_2 \varepsilon_{(t-2)} + \dots + \theta_q \varepsilon_{(t-q)}, \quad (13)$$

The autoregressive model with integration and moving average orders  $(p, d, q)$  is the sum of  $AR(p)$  and  $MA(q)$  models and can be presented in the form:

$$ARIMA(p, d, q): (1 - \varphi_1 L - \dots - \varphi_p L^p)((1 - L)^d y_t - \mu) = (1 + \theta_1 L + \dots + \theta_q L^q) \varepsilon_t, \quad (14)$$

where  $d$  is the number of differentiations of the initial time series until its stationarity, and  $L$  is the value of the lag.

Some limitation of ARIMA is that it processes the time series data in a linear form and must be stationary, and the accuracy of the forecast model significantly depends on the noise of the time series data set.

To determine the seasonality, one can use the modification of the base model  $ARIMA(p, d, q) - SARIMA(P, D, Q) m$ , where  $m$  - number of measurements per year,  $P, D, Q$  - the orders of seasonal autoregression, seasonal integration and seasonal moving average respectively.

The  $SARIMA(p, d, q)(P, D, Q, s)$  model is a generalization of the ARIMA model for time series with a pronounced seasonal component. In such a model, in addition to the order of the autoregressive part, the order of the moving average and the order of differences, seasonal parameters  $(P, D, Q, s)$  are introduced, which allow considering cyclic process fluctuations when building the forecast.

## **Conclusion**

As a result of the master's thesis, a review of PCA, SSA, ARIMA methods was conducted. The basic methods of SSA and ARIMA with their implementation in an interactive computing environment Jupiter Notebook on the programming language Python were mastered. The SSA and ARIMA methods investigated meteorological data, with subsequent forecast creation. The data presented by the Institute for Monitoring of Climate and Environmental Systems of the Siberian Branch of the Russian Academy of Sciences in Tomsk for 2.5 years of observations were used. Other time series were also analyzed with the help of SSA.

Built forecast data have not shown effective results and do not consider many factors in making real meteorological forecasts. But they can be used in other activities.