

Школа – Инженерная школа информационных технологий и робототехники  
 Направление подготовки – 09.03.02 Информационные системы и технологии  
 Отделение школы – Отделение информационных технологий

### БАКАЛАВРСКАЯ РАБОТА

Тема работы
<b>Программно-математические средства прогнозирования временных рядов на основе гибридной модели и текущей сегментации данных, поступающих в реальном времени</b>

УДК 004.932.1:004.451.7.031.43:519.246

Студент

Группа	ФИО	Подпись	Дата
8И8Б	Михайлова Яна Андреевна		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Кочегурова Елена Алексеевна	к.т.н., доцент		

### КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН ШБИП	Меньшикова Екатерина Валентиновна	к.ф.н.		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ст. преподаватель	Мезенцева Ирина Леонидовна			

### ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Цапко Ирина Валериевна	к.т.н., доцент		

## ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ООП

Код компетенции	Наименование компетенции
<b>Универсальные компетенции</b>	
<b>УК(У)-1</b>	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач
<b>УК(У)-2</b>	Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений
<b>УК(У)-3</b>	Способен осуществлять социальное взаимодействие и реализовывать свою роль в команде
<b>УК(У)-4</b>	Способен осуществлять деловую коммуникацию в устной и письменной формах на государственном языке Российской Федерации и иностранном(-ых) языке(-ах)
<b>УК(У)-5</b>	Способен воспринимать межкультурное разнообразие общества в социально-историческом, этическом и философском контекстах
<b>УК(У)-6</b>	Способен управлять своим временем, выстраивать и реализовывать траекторию саморазвития на основе принципов образования в течение всей жизни
<b>УК(У)-7</b>	Способен поддерживать должный уровень физической подготовленности для обеспечения полноценной социальной и профессиональной деятельности
<b>УК(У)-8</b>	Способен создавать и поддерживать в повседневной жизни и в профессиональной деятельности безопасные условия жизнедеятельности для сохранения природной среды, обеспечения устойчивого развития общества, в том числе при угрозе и возникновении чрезвычайных ситуаций и военных конфликтов
<b>УК(У)-9</b>	Способен проявлять предприимчивость в практической деятельности, в т.ч. в рамках разработки коммерчески перспективного продукта на основе научно-технической идеи
<b>УК(У)-10</b>	Способен принимать обоснованные экономические решения в различных областях жизнедеятельности
<b>УК(У)-11</b>	Способен формировать нетерпимое отношение к коррупционному поведению
<b>Общепрофессиональные компетенции</b>	
<b>ОПК(У)-1</b>	Способен применять естественнонаучные и общинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности
<b>ОПК(У)-2</b>	Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности
<b>ОПК(У)-3</b>	Способен решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности
<b>ОПК(У)-4</b>	Способен участвовать в разработке технической документации, связанной с профессиональной деятельностью с использованием стандартов, норм и правил

<b>ОПК(У)-5</b>	Способен устанавливать программное и аппаратное обеспечение для информационных и автоматизированных систем
<b>ОПК(У)-6</b>	Способен разрабатывать алгоритмы и программы, пригодные для практического применения в области информационных систем и технологий
<b>ОПК(У)-7</b>	Способен осуществлять выбор платформ и инструментальных программно-аппаратных средств для реализации информационных систем
<b>ОПК(У)-8</b>	Способен применять математические модели, методы и средства проектирования информационных и автоматизированных систем
<b>Профессиональные компетенции</b>	
<b>ПК(У)-1</b>	Способен разрабатывать требования и проектировать программное обеспечение
<b>ПК(У)-2</b>	Способен выполнять работы, связанные со сбором, обработкой и подготовкой картографической информации
<b>ПК(У)-3</b>	Способен выполнять анализ и интерпретацию данных ДЗЗ
<b>ПК(У)-4</b>	Способен выполнять работы и управление работами по созданию (модификации) и сопровождению ИС, автоматизирующих задачи организационного управления и бизнес-процессы
<b>ПК(У)-5</b>	Способен разрабатывать графический дизайн интерфейса
<b>ПК(У)-6</b>	Способен разрабатывать базы данных ИС
<b>ПК(У)-7</b>	Способность обеспечивать безопасность информации в автоматизированных системах

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 – 09.03.02 Информационные системы и технологии  
 Отделение школы – Отделение информационных технологий

УТВЕРЖДАЮ:  
 Руководитель ООП

\_\_\_\_\_ Цапко И.В.  
 (Подпись)     (Дата)     (Ф.И.О.)

### ЗАДАНИЕ

**на выполнение выпускной квалификационной работы**

В форме:

Бакалаврской работы
---------------------

Студенту:

Группа	ФИО
8И8Б	Михайлова Яна Андреевна

Тема работы:

Программно-математические средства прогнозирования временных рядов на основе гибридной модели и текущей сегментации данных, поступающих в реальном времени	
Утверждена приказом директора (дата, номер)	03.02.2022 № 34-65/с

Срок сдачи студентом выполненной работы:

--	--

### ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

<b>Исходные данные к работе</b>	<p>Объект исследования: программно-математические средства прогнозирования временных рядов</p> <p>Цель работы: Разработка гибридного метода прогнозирования на основе штрафного сплайна и текущей сегментации данных</p> <p>Режим работы: непрерывный</p>
---------------------------------	---

<b>Перечень подлежащих исследованию, проектированию и разработке вопросов</b>	<ol style="list-style-type: none"> <li>1. Обзор методов прогнозирования временных рядов;</li> <li>2. обзор методов оптимизации параметров прогнозной модели;</li> <li>3. разработка гибридной модели прогнозирования;</li> <li>4. программная реализация метода прогнозирования;</li> <li>5. анализ результатов разработки;</li> <li>6. финансовый менеджмент, ресурсоэффективность и ресурсосбережение;</li> <li>7. социальная ответственность.</li> </ol>
<b>Перечень графического материала</b>	<ol style="list-style-type: none"> <li>1. Схема классификации методов и моделей прогнозирования;</li> <li>2. схемы работы методов прогнозирования;</li> <li>3. графики временных рядов;</li> <li>4. матрица SWOT-анализа;</li> <li>5. диаграмма Ганта.</li> </ol>

**Консультанты по разделам выпускной квалификационной работы**

*(с указанием разделов)*

<b>Раздел</b>	<b>Консультант</b>
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Меньшикова Екатерина Валентиновна
Социальная ответственность	Мезенцева Ирина Леонидовна

**Названия разделов, которые должны быть написаны на русском и иностранном языках:**

Заключение

<b>Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику</b>	24.01.2022 г.
---	---------------

**Задание выдал руководитель:**

<b>Должность</b>	<b>ФИО</b>	<b>Ученая степень, звание</b>	<b>Подпись</b>	<b>Дата</b>
Доцент ОИТ ИШИТР	Кочегурова Елена Алексеевна	к.т.н., доцент		

**Задание принял к исполнению студент:**

<b>Группа</b>	<b>ФИО</b>	<b>Подпись</b>	<b>Дата</b>
8И8Б	Михайлова Яна Андреевна		

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 – 09.03.02 Информационные системы и технологии  
 Уровень образования – Бакалавриат  
 Отделение школы – Отделение информационных технологий  
 Период выполнения – весенний семестр 2021/2022 учебного года

Форма представления работы:

Бакалаврская работа

### КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН выполнения выпускной квалификационной работы

Срок сдачи студентом выполненной работы:	10.06.2022
--	------------

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
10.06.2022	Основная часть	75
01.06.2022	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	15
01.06.2022	Социальная ответственность	10

**СОСТАВИЛ:**

**Руководитель ВКР**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Кочегурова Елена Алексеевна	к.т.н., доцент		

**СОГЛАСОВАНО:**

**Руководитель ООП**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Цапко Ирина Валериевна	к.т.н., доцент		

**ЗАДАНИЕ ДЛЯ РАЗДЕЛА  
«ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСООБЪЕКТИВНОСТЬ И РЕСУРСОСБЕ-  
РЕЖЕНИЕ»**

Студенту:

<b>Группа</b>	<b>ФИО</b>
8И8Б	Михайлова Яна Андреевна

<b>Школа</b>	<b>ИШИТР</b>	<b>Отделение школы (НОЦ)</b>	<b>Отделение информационных технологий</b>
<b>Уровень образова- ния</b>	Бакалавриат	<b>Направление/специа- льность</b>	09.03.02 Информационные системы и технологии

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

1. <i>Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих</i>	Бюджет проекта – не более 264397 руб., в т.ч. затраты по оплате труда – не более 148213 руб.
2. <i>Нормы и нормативы расходования ресурсов</i>	Значение показателя интегральной ресурсоэффективности – не менее 4 баллов из 5
3. <i>Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования</i>	Налог во внебюджетные фонды 27,1% Районный коэффициент – 1,3 Накладные расходы – 16%

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

1. <i>Оценка коммерческого потенциала инженерных решений (ИР)</i>	Оценка потенциальных потребителей исследования, SWOT – анализ.
2. <i>Формирование плана и бюджета инженерного проекта (ИП)</i>	Планирование этапов работ, определение трудоемкости и построение календарного графика, формирование бюджета.
3. <i>Оценка ресурсной, финансовой, социальной, бюджетной эффективности ИР и потенциальных рисков</i>	Оценка сравнительной эффективности исследования. Интегральный показатель ресурсоэффективности – 4,7 Интегральный показатель эффективности – 4,99 Сравнительная эффективность проекта – 1,31

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

1. Оценка конкурентоспособности НТИ
2. Матрица SWOT
3. Морфологическая матрица
4. План-график разработки
5. Бюджет НТИ
6. Оценка ресурсной, финансовой и экономической эффективности НТИ

<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 ОСГН ШБИП	Меньшикова Екатерина Валентиновна	к.ф.н.		

**Задание принял к исполнению студент:**

<b>Группа</b>	<b>ФИО</b>	<b>Подпись</b>	<b>Дата</b>
8И8Б	Михайлова Яна Андреевна		

## ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

<b>Группа</b>		<b>ФИО</b>	
8И8Б		Михайлова Яна Андреевна	
<b>Школа</b>	ИШИТР	<b>Отделение (НОЦ)</b>	Отделение информационных технологий
<b>Уровень образования</b>	Бакалавриат	<b>Направление/специальность</b>	09.03.02 Информационные системы и технологии

Тема ВКР:

<b>Программно-математические средства прогнозирования временных рядов на основе гибридной модели и текущей сегментации данных, поступающих в реальном времени</b>	
<b>Исходные данные к разделу «Социальная ответственность»:</b>	
<p><b>Введение</b></p> <ul style="list-style-type: none"> <li>– Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика) и области его применения.</li> <li>– Описание рабочей зоны (рабочего места) при разработке проектного решения/при эксплуатации</li> </ul>	<p><i>Объект исследования:</i> программно-математические средства прогнозирования временных рядов  <i>Область применения:</i> исследовательская деятельность, прогнозирование  <i>Рабочая зона:</i> офис  <i>Размеры помещения:</i> 3*6 м  <i>Количество и наименование оборудования рабочей зоны:</i> ПЭВМ - 1 шт.  <i>Рабочие процессы, связанные с объектом исследования, осуществляющиеся в рабочей зоне:</i> написание программного кода</p>
Перечень вопросов, подлежащих исследованию, проектированию и разработке:	
<p><b>1. Правовые и организационные вопросы обеспечения безопасности при разработке проектного решения:</b></p> <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>	<p>ГОСТ 12.2.032-78 ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования.          ГОСТ 21889-76 Система «человек-машина». Кресло человека-оператора. Общие эргономические требования.          СП 2.2.3670-20 «Санитарно-эпидемиологические требования к условиям труда»          Трудовой кодекс Российской Федерации от 30.12.2001 N 197-ФЗ (ред. от 25.02.2022)</p>
<p><b>2. Производственная безопасность при разработке проектного решения:</b></p> <ul style="list-style-type: none"> <li>– Анализ выявленных вредных и опасных производственных факторов</li> </ul>	<p><b>Вредные производственные факторы:</b></p> <ol style="list-style-type: none"> <li>1. Отсутствие или недостаток необходимого естественного освещения.</li> <li>2. Факторы, связанные с аномальными микроклиматическими параметрами воздушной среды на местонахождении работающего.</li> <li>3. Факторы, обладающие свойствами психофизиологического воздействия на организм человека: физические статические перегрузки, связанные с рабочей позой.</li> </ol> <p><b>Опасные производственные факторы:</b></p> <ol style="list-style-type: none"> <li>1. Факторы, связанные с электрическим током, вызываемым разницей электрических потенциалов, под действие которого попадает работающий.</li> </ol> <p><b>Требуемые средства коллективной и индивидуальной защиты от выявленных факторов:</b></p>

	световые проемы, устройства для отопления, устройства для вентиляции воздуха, изолирующие устройства, устройства защитного заземления и зануления.
<b>3. Экологическая безопасность при разработке проектного решения</b>	<p>Воздействие на селитебную зону – выбросы загрязняющих веществ с электростанций, обеспечивающих работу офисной техники</p> <p>Воздействие на литосферу – отходы от утилизации офисной техники</p> <p>Воздействие на гидросферу – сброс загрязняющих веществ с электростанций, обеспечивающих работу офисной техники</p> <p>Воздействие на атмосферу – выбросы загрязняющих веществ с электростанций, обеспечивающих работу офисной техники</p>
<b>4. Безопасность в чрезвычайных ситуациях при разработке проектного решения</b>	<p><b>Возможные ЧС:</b></p> <ol style="list-style-type: none"> <li>1. техногенные (взрывы, пожары, обрушение помещений);</li> <li>2. природные (наводнения, ураганы, природные пожары);</li> <li>3. биологические (эпидемии, пандемии);</li> <li>4. социальные (террор).</li> </ol> <p><b>Наиболее типичная ЧС:</b> пожар.</p>
<b>Дата выдачи задания для раздела по линейному графику</b>	

**Задание выдал консультант:**

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ст. преподаватель	Мезенцева Ирина Леонидовна			

**Задание принял к исполнению студент:**

Группа	ФИО	Подпись	Дата
8И8Б	Михайлова Яна Андреевна		

## РЕФЕРАТ

Выпускная квалификационная работа 84 с., 22 рис., 18 табл., 29 источников, 2 прил.

Ключевые слова: временной ряд, прогнозирование, гибридная модель, сплайн, сегментация, разладка, оптимизация параметров.

Объектом исследования являются программно-математические средства прогнозирования временных рядов.

Цель работы состоит в разработке гибридного метода прогнозирования на основе штрафного сплайна и текущей сегментации данных.

В процессе работы проводилось изучение временных рядов, обзор методов прогнозирования временных рядов, изучение сегментации данных.

В результате исследования разработана гибридная модель прогнозирования временных рядов, основанная на штрафном сплайне и сегментации данных. Получены прогнозы на основе синтетических и реальных данных с оценкой точности результатов.

Область применения: научно-исследовательская, прикладные задачи обработки данных.

## **ОПРЕДЕЛЕНИЯ, ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ**

ВР – временные ряды

РРВ – режим реального времени

НТИ – научно-техническое исследование

Датасет – набор данных

ИПЦ – индекс потребительских цен

## ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	15
1 ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ В РЕАЛЬНОМ ВРЕМЕНИ.....	16
1.1 Задача прогнозирования временных рядов.....	16
1.2 Методы прогнозирования .....	17
1.2.1 Гибридные модели прогнозирования .....	19
1.2.2 Сплайн-функции в задаче прогнозирования.....	19
1.2.3 Разладка во временных рядах.....	20
2 РАЗРАБОТКА ГИБРИДНОЙ МОДЕЛИ ПРОГНОЗИРОВАНИЯ	22
2.1 Модель на основе штрафного сплайна .....	22
2.2 Настройка параметров гибридной модели .....	24
2.2.1 Сегментация временных рядов.....	26
2.3 Оценка качества моделей прогнозирования .....	27
2.4 Подготовка данных для исследования.....	30
2.4.1 Синтетические данные .....	31
2.4.2 Реальные данные.....	32
3 РАЗРАБОТКА ПРОГРАММЫ .....	35
3.1 Реализация прогнозирования на основе штрафного сплайна.	35
3.2 Реализация сегментации данных.....	37
4 ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ.....	40
4.1 Оценка коммерческого потенциала и перспективности проведения научных исследований с позиции ресурсоэффективности и ресурсосбережения .....	40
4.1.1 Потенциальные потребители результатов исследования ...	40

4.1.2	Анализ конкурентных технических решений.....	40
4.1.3	SWOT-анализ.....	42
4.2	Определение возможных альтернатив проведения научных исследований.....	43
4.3	Планирование научно-исследовательских работ .....	44
4.3.1	Определение трудоемкости выполнения работ.....	44
4.3.2	Разработка графика проведения научного исследования...	45
4.3.3	Бюджет научно-технического исследования (НТИ) .....	49
4.3.3.1	Расчет материальных затрат НТИ .....	49
4.3.3.2	Расчет затрат на специальное оборудование для научных (экспериментальных) работ .....	50
4.3.3.3	Основная заработная плата исполнителей темы.....	50
4.3.3.4	Дополнительная заработная плата исполнителей темы...	52
4.3.3.5	Отчисления во внебюджетные фонды (страховые отчисления) .....	52
4.3.3.6	Накладные расходы.....	53
4.3.3.7	Формирование бюджета затрат научно-исследовательского проекта .....	53
4.4	Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования..	54
5	СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ.....	56
5.1	Правовые и организационные вопросы обеспечения безопасности .....	56
5.1.1	Правовые нормы трудового законодательства.....	56
5.1.2	Эргономические требования к правильному расположению и компоновке рабочей зоны .....	57

5.2	Производственная безопасность .....	58
5.2.1	Отсутствие или недостаток необходимого естественного освещения .....	59
5.2.2	Факторы, связанные с аномальными микроклиматическими параметрами воздушной среды на местонахождении работающего.....	59
5.2.3	Факторы, обладающие свойствами психофизиологического воздействия на организм человека: физические статические перегрузки, связанные с рабочей позой.....	61
5.2.4	Факторы, связанные с электрическим током, вызываемым разницей электрических потенциалов, под действие которого попадает работающий .....	61
5.3	Экологическая безопасность .....	62
5.4	Безопасность в чрезвычайных ситуациях .....	63
	Вывод по разделу.....	64
	ЗАКЛЮЧЕНИЕ .....	65
	CONCLUSION .....	66
	СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ.....	67
	Приложение А Графики прогноза на основе Р-сплайна.....	70
	Приложение Б Графики сегментов и кумулятивных сумм .....	77

## ВВЕДЕНИЕ

В век информатизации и цифровизации огромное значение уделяется обработке данных. Прогнозирование ВР является актуальной задачей во множестве прикладных областей. Анализ и прогнозирование ВР применяется в научных исследованиях, в медицине, промышленности, в сельском хозяйстве, в финансовой деятельности и в других сферах. Существует множество подходов к прогнозированию ВР, и особенно интересны к рассмотрению методы, комбинирующие несколько подходов. Такие гибридные методы позволяют повысить эффективность прогноза за счёт объединения преимуществ задействованных методов. Среди различных методов прогнозирования в особенности перспективны те, которые работают в реальном масштабе времени. В таких методах и моделях время расчёта прогнозного значения меньше времени дискретизации реальных значений. Работа в режиме реального времени актуальна для систем, требующих быстрого реагирования.

Целью данной работы является разработка программно-математического средства для прогнозирования ВР на основе метода штрафного сплайна и сегментации данных.

В результате работы над ВКР был реализован метод прогнозирования ВР на основе штрафного сплайна. Реализована сегментация данных на основе поиска разладки с помощью алгоритма кумулятивных сумм в апостериорном режиме. Спроектирована сегментация данных на основе поиска разладки с помощью алгоритма кумулятивных сумм, работающая в режиме запаздывания со скользящим окном переменного размера.

Реализованные методы апробированы на искусственных данных и на реальных ВР. Результаты работы говорят об эффективности разработки в задачах прогнозирования.

# 1 ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ В РЕАЛЬНОМ ВРЕМЕНИ

## 1.1 Задача прогнозирования временных рядов

Прогнозирование ВР является актуальной задачей во множестве прикладных областей. ВР представляет собой совокупность значений какого-либо показателя или процесса, собранную в различные моменты времени. Значения в ВР могут зависеть от предыдущих значений и от времени их получения. ВР изображают на графиках, у которых по оси ОХ отложены значения времени, а по оси ОУ – наблюдаемые значения показателя. Две большие группы задач в работе с ВР: их анализ и прогнозирование. Анализ подразумевает извлечение из уже имеющихся данных максимума полезной информации. Задача прогнозирования заключается в предсказании поведения ВР в ещё не наставшие моменты времени.

По наличию тренда ВР делятся на стационарные и нестационарные. Нестационарные ряды имеют направление преимущественного движения, долгосрочную тенденцию увеличения или уменьшения значений. Такая тенденция может быть описана в форме некоторого достаточно простого уравнения (линейного, экспоненциального, степенного логарифмического), наилучшим образом аппроксимирующего фактическую тенденцию динамики. Если в ВР отсутствует тенденция к изменениям, т.е. ВР характеризуется постоянным математическим ожиданием и дисперсией, то такой ряд считают стационарным.

Для ВР характерно наличие сезонной компоненты – периодически повторяющегося явления в данных. Получение информации о недельной, квартальной, годовой или какой-либо другой сезонности помогает лучше узнать данные. Извлечение из ВР сезонной компоненты необходимо для прогнозирования данных. Для ответа на вопрос, характерна ли для данного ВР та или иная сезонность, используют автокорреляционную функцию.

После удаления из ВР трендовой и сезонной компоненты в данных может быть выявлен остаток. Так называют краткосрочные колебания, которые не являются систематическими или предсказуемыми.

Горизонт времени, на который необходимо сделать прогноз, называется временем упреждения, различают долгосрочное, среднесрочное и краткосрочное прогнозирование.

Для краткосрочного прогноза на 1-3 значения вперёд особенно актуальны методы прогнозирования, работающие в режиме реального времени. Данные, полученные при таком прогнозе, позволяют оперативно представлять меняющуюся картинку исследуемого процесса.

## 1.2 Методы прогнозирования

Классификация методов и моделей прогнозирования представлена на рисунке 1.

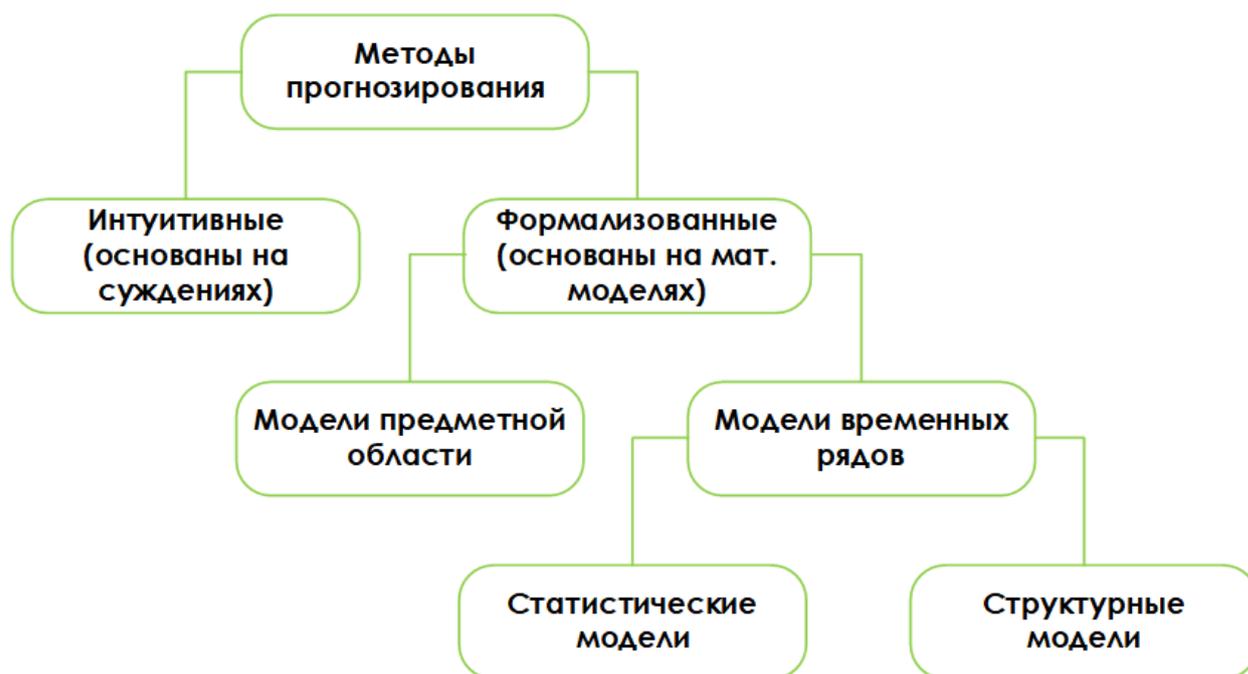


Рисунок 1 – Общая классификация методов и моделей прогнозирования

Все методы прогнозирования можно разделить на две группы: интуитивные и формализованные [1].

Бывает, что система, поведение которой необходимо спрогнозировать, не поддаётся математическому описанию, т.к. слишком сложна. Либо наоборот

система слишком проста и не нуждается в формализации. В таких случаях применяют интуитивные методы прогнозирования. Методы этой группы имеют дело с экспертными оценками и суждениями. Такой подход популярен в маркетинге, политической и экономической сфере.

Формализованные методы прогнозирования исследуют объект для получения его математического описания. Т.е. строится математическая модель, которая позволяет сделать прогноз.

Математические модели в свою очередь могут быть моделями предметной области или моделями ВР.

При построении математической модели предметной области используются законы предметной области изучаемого объекта. Такие модели не могут быть универсальными для любых данных, поскольку они используют характерные лишь для данной предметной области зависимости. К примеру, модель для прогноза погоды содержит уравнения динамики жидкостей и термодинамики. Модели для других предметных областей могут содержать дифференциальные уравнения и даже системы дифференциальных уравнений. Таким моделям необходим индивидуальный подход в разработке, поэтому они более трудоёмки в создании. Провести классификацию моделей предметной области не представляется возможным, т.к. их количество неограниченно.

В противовес специфичным моделям предметных областей существуют более универсальные модели ВР. Модели ВР подходят для данных любого происхождения, любой сферы. В таких моделях ищется зависимость будущих значений от прошлых. На основании этой зависимости строится прогноз. Например, модель, основанная на нейронных сетях, может использоваться и для прогноза числа посетителей магазина, и для прогноза атмосферного давления. Модели ВР можно разделить на статистические и структурные.

В статистических моделях зависимость будущего значения от прошлого задано в виде уравнения. Пример статистических моделей:

- регрессионные модели (линейная регрессия, нелинейная регрессия);
- авторегрессионные модели (ARIMAX, GARCH, ARDLM);
- модель экспоненциального сглаживания и пр.

В структурных моделях зависимость будущего значения от прошлого задается в виде некоторой структуры и правил перехода по ней. К ним относятся:

- нейросетевые модели;
- модели на базе цепей Маркова;
- модели на базе классификационно-регрессионных деревьев и пр.

### **1.2.1 Гибридные модели прогнозирования**

В настоящее время существует множество методов прогнозирования, их комбинирование позволяет получить модель, объединяющую преимущества каждого из задействованных подходов. К гибридным моделям обращаются для получения более качественного прогноза, т.е. для повышения его точности.

Зачастую гибридные модели строятся как комбинация статистической и структурной моделей. Статистические модели используются для поиска линейных закономерностей в рядах, а структурные модели эффективны для ВР со сложными закономерностями. Поэтому комбинация таких методов позволяет достичь необходимого уровня эффективности и универсальности.

Очевидным недостатком гибридных моделей прогнозирования является ресурсоёмкость их разработки. Необходимо комбинировать модели так, чтобы объединить преимущества каждой из них, нейтрализовав слабые стороны.

### **1.2.2 Сплайн-функции в задаче прогнозирования**

Один из подходов к задаче прогнозирования представляет собой использование аппроксимирующих кусочно-полиномиальных функций, заменяющих регрессионные соотношения при моделировании и анализе процессов.

Существуют различные виды сплайнов, применяемых в разнообразных прикладных задачах:

- регрессионные сплайны,
- сглаживающие сплайны,
- В-сплайны (базисные сплайны),
- штрафные Р-сплайны (разновидность В-сплайнов) и т.д.

Сглаживающие сплайны представляют собой метод аппроксимации исходных зашумленных данных и используются для отделения сигнала от шума. Наиболее известным примером сглаживающего сплайна является кубический сглаживающий сплайн.

Регрессионные сплайны отличаются от сглаживающих тем, что данные аппроксимируются определенным набором базисных сплайн-функций с уменьшенным количеством узлов, обычно с использованием метода наименьших квадратов (МНК).

Штрафные сплайны (P-splines) представляют собой сочетание уменьшенного количества узлов аналогично регрессионным сплайнам и штрафами за негладкость сглаживающих сплайнов, которые подавляют нелинейность.

Исследования [2] показывают, что прогнозы, основанные на штрафных сплайнах, более точны по оценкам MSE, MAE, MAPE, чем прогнозы, сделанные на основе кубических сглаживающих сплайнов.

### **1.2.3 Разладка во временных рядах**

Традиционным при исследовании ВР является предположение о том, что статистические свойства наблюдаемых рядов постоянны во времени или медленно изменяются. Но всё же в случайный момент времени могут изменяться статистические свойства процесса, это явление называется разладкой. Сегментация ВР подразумевает под собой разбиение этого ряда на сегменты, которые являются внутренне однородными. Для этого необходимо найти точки изменения,

то есть момент разладки, момент отклонения контролируемых параметров от нормативных пределов [3].

Наибольшей популярностью в отношении поиска момента разладки обладают последовательные алгоритмы, в которых данные обрабатываются в режиме реального времени при поступлении каждого нового измерения, после чего делается вывод об отсутствии разладки или о её появлении. Фундаментальными работами в задаче последовательного обнаружения изменения математического ожидания гауссовской независимой случайной последовательности являются работы Е. С. Пейджа [4], М. Гиршика и Г. Рубина [5].

В результате исследований Ширяева [6] и Лордена [7] было установлено, что алгоритм кумулятивных сумм является оптимальным среди последовательных алгоритмов в смысле минимизации среднего времени запаздывания в обнаружении разладки.

В задаче прогнозирования ВР поиск момента разладки может быть использован для сегментации ВР и перенастройки параметров модели.

## 2 РАЗРАБОТКА ГИБРИДНОЙ МОДЕЛИ ПРОГНОЗИРОВАНИЯ

В данной работе будет разработан гибридный метод прогнозирования ВР, включающий в себя настройку параметров с помощью сегментации и дальнейшее построение прогноза на основе штрафного сплайна с использованием подобранных оптимальных параметров.

### 2.1 Модель на основе штрафного сплайна

Применение сплайна позволяет получить искомое гладкое значение. Сглаживающий сплайн  $S(t)$  основан на оптимизации специального вида функционала [11, 12] (1).

$$J(S) = (1 - \rho)(h\Delta t)^2 \int_{t_0^i}^{t_h^i} [S''(t)]^2 dt + \rho \sum_{j=0}^h [S(t_j^i) - y(t_j^i)]^2, \quad (1)$$

где  $\rho \in [0; 1]$  - весовой коэффициент, устанавливающий баланс между сглаживающими и интерполяционными свойствами сплайна  $S(t)$ ;

$\Delta t$  - интервал дискретизации наблюдаемого процесса;

$h$  - количество измерений внутри  $i$ -го звена сплайна.

Для РРВ используется рекуррентная форма расчета коэффициентов сплайна для каждого  $i$ -го звена на основании  $h$  отсчетов:

$$S = a_0^i + a_1^i m + a_2^i m^2 + a_3^i m^3, \quad -q \leq m \leq h - q, \quad (2)$$

где  $q$  - номер отсчёта  $t_j^i$  внутри  $i$ -го звена ( $j = \overline{0, h}$ ;  $q = \overline{0, h - 1}$ ), в котором сопряжены непрерывные производные сплайна;

$m = t_j^i - t_0^i$  номер отсчёта  $t_j^i$  внутри  $i$ -го звена, в котором сплайн вычисляется как  $m = \overline{-q, h - q}$ ;

$$h = \frac{t_j^i - t_0^i}{\Delta t} - \text{количество отсчётов в } i\text{-м звене сплайна.}$$

Для вычисления прогноза в РРВ используется текущий режим функционирования сплайна, т.е. звено сплайна вычисляется при поступлении нового значения с использованием  $h-1$  предыдущих значений. В (3) приведены формулы для расчёта коэффициентов сплайна (при  $q = 0$  и дефекте  $d = 2$ ).

$$\begin{aligned}
 a_0^i &= a_0^{i-1} + a_1^i h + a_2^i h^2 + a_3^i h^3 \\
 a_1^i &= a_1^{i-1} + 2a_2^i h + 3a_3^i h^2 \\
 a_2^i &= \frac{\rho(F_1^i C - F_2^i A)}{BC - A^2} \\
 a_3^i &= \frac{\rho(F_2^i B - F_1^i A)}{BC - A^2} \\
 A &= 6(1 - \rho)h^4 + \rho H_5 \\
 B &= 4(1 - \rho)h^3 + \rho H_4 \\
 C &= 12(1 - \rho)h^5 + \rho H_6 \\
 F_1^i &= \sum_{k=0}^h y(t_k^i) k^2 - a_0^i H_2 - a_1^i H_3 \\
 F_2^i &= \sum_{k=0}^h y(t_k^i) k^3 - a_0^i H_3 - a_1^i H_4 \\
 H_n &= \sum_{k=0}^h k^n
 \end{aligned} \tag{3}$$

Модель фрагмента ВР в виде звена штрафного Р-сплайна позволяет получать значения прогноза для  $\tau > h$  в реальном времени. На рисунке 2 изображен горизонт прогноза для двух значений  $\tau = h + 1$  и  $\tau = h + 2$ .

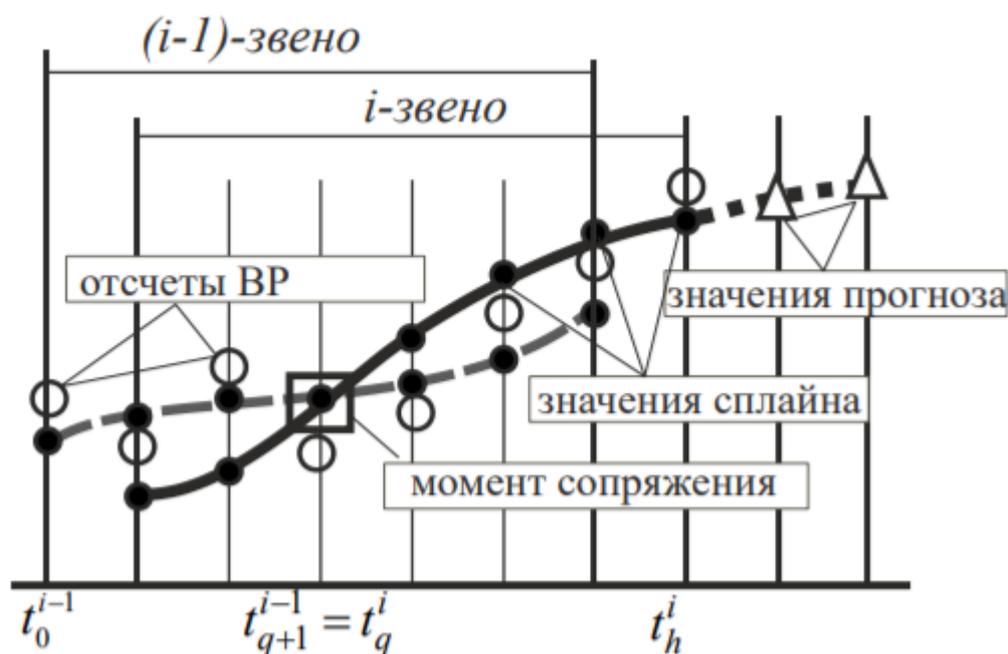


Рисунок 2 – Режим прогнозирования на основе штрафного P-сплайна

Параметр  $h$ , определяющий сколько предыдущих значений учитываются при построении сплайна, может быть константой. Но для повышения точности прогноза этот параметр должен подбираться исходя из статистических свойств текущего однородного сегмента ряда.

## 2.2 Настройка параметров гибридной модели

Гибридные методы как правило подразумевают разбиение ряда на однородные фрагменты и последующее получение прогнозного значения на основе характеристик этих фрагментов и текущего значения ВР. Одним из подходов к разбиению ВР являются методы поиска сходства. К таким относятся различные вариации кластеризации, в том числе продольная кластеризация вдоль временной оси. Другим подходом является обнаружение изменения динамики путём определения моментов разладки и основанной на них сегментации ВР.

Для настройки параметров моделей могут быть использованы биоинспирированные алгоритмы, генетический алгоритм, иммунный алгоритм, метод сегментации на основе кластеризации и обнаружении разладки и другие подходы.

Эволюционные алгоритмы позволяют производить настройку одновременно нескольких параметров, но при этом являются времяёмкими. Такие

алгоритмы тратят много времени на пересчёт нового значения и зачастую требуют большой предыстории значений. Получение эволюционного алгоритма, эффективно работающего в режиме реального времени на данный момент является нерешённой научно-исследовательской задачей.

Глубина прогноза, или размер предыстории ВР, который используется для построения нового значений прогноза, может быть постоянным, возможен вариант, когда он увеличивается с течением работы алгоритма. В первом случае говорят о методе скользящего окна. Под окном здесь понимается временной интервал, значения которого используются для построения прогноза. С поступлением нового значения окно смещается на одну единицу наблюдения вперёд, при этом размер временного интервала остаётся неизменным. Другой подход называется расширяющимся или раскрывающимся окном. Расширяющееся окно содержит в себе все предыдущие значения ряда, т.е. с течением работы алгоритма глубина прогноза увеличивается, для прогноза очередного значения учитываются все предыдущие значения.

Использование скользящего окна постоянного размера может быть недостаточно эффективным, т.к. текущий фрагмент ряда может не отражать в полной мере динамику изменений значений. Например, в предысторию входит резкий скачок значений, после которого значения имеют тенденцию уменьшаться. В таком случае кажется целесообразным сократить величину предыстории для исключения из неё этого скачка. Или, наоборот, тенденция ряда долгое время не меняется, тогда глубину прогноза можно увеличить.

В случае использования расширяющегося окна, учитываются все прошлые значения ряда, т.е. с каждым новым значением необходимо обрабатывать больше и больше данных, что негативно сказывается на скорости расчёта прогноза. В расчёте присутствуют старые значения, которые мало влияют на текущую динамику ряда.

В связи с описанными недостатками и преимуществами рассмотренных вариантов определения глубины истории, было принято решение менять размер окна в зависимости от изменения динамики ряда. Т.е. в разрабатываемом методе

прогнозирования должен производиться подбор параметра  $h$ , отвечающего за величину предыстории.

### 2.2.1 Сегментация временных рядов

Разделение ВР на внутренние однородные сегменты является важной задачей интеллектуального анализа данных. Изменения характеристик ВР обычно расплывчаты и не привязаны к какому-либо конкретному моменту времени. Поэтому нецелесообразно определять четкие границы сегментов.

Сегментация ВР может проводиться на основе поиска моментов разладки в рядах.

Для наблюдений, для которых заранее неизвестна функция распределения, оптимально использовать методы, требующие накопления большого объёма данных. Это апостериорные методы. Для работы в РРВ можно использовать методы со скользящим окном.

Поскольку заранее о функции распределения любого ВР неизвестно, предлагается использовать метод, описанный Воробейчиковым С.Э. [8], основанный на алгоритме кумулятивных сумм с рекуррентными вычислениями [15].

$$F(x) = \begin{cases} F_0(x), & \text{если } n < \theta \\ F_1(x), & \text{если } n \geq \theta \end{cases}, \quad F_1(x) = F_0(x - d), \quad (4)$$

где  $\theta$  – момент разладки

Предполагается, что распределение  $F_i(x)$  неизвестно, а  $M_1 x_n > M_0 x_n$ , где через  $M_i$  обозначено усреднённое распределение  $F_1, F_2$ . И момент разладки  $\theta$  устанавливается на основании наблюдений за  $x_n$ .

Для обнаружения изменения среднего, используется такая статистика, как знаковая функция (5).

$$y_n = \text{sign}(x_n - x_{n-n_0}), \quad n > n_0, \quad (5)$$

где  $n_0$  – глубина памяти.

Среднее значение  $y_n$  больше нуля, если  $\theta \leq n < \theta + n_0$ . При  $n < \theta$  или  $n > \theta + n_0$  среднее  $y_n$  нулевое. Таким образом задача сводится к обнаружению

скачка среднего значения на промежутке конечной длины. Предполагается, что после разладки

Алгоритм имеет параметр, обозначающий чувствительность процедуры (6).

$$\delta = \frac{m}{N} < \Delta, \quad (6)$$

где  $m$  и  $N$  – несократимые натуральные числа.

Формируется последовательность, среднее значение которой меняется с отрицательного на положительное в момент разладки (7).

$$z_n = (y_n - \delta)N \quad (7)$$

К рассчитанной последовательности применяется алгоритм кумулятивных сумм (8)-(10).

$$l = N + m, \quad (8)$$

$$S_{n_0} = l, \quad (9)$$

$$S_n = \max(S_{n-1} + z_n, l). \quad (10)$$

При достижении суммой порогового значения принимается решение о разладке. Пороговое значение должно быть больше значения  $l$ .

### 2.3 Оценка качества моделей прогнозирования

В вопросах оценки точности прогнозной модели можно выделить два основных направления: точностная оценка и временная оценка. Точностная оценка показывает, насколько близко значение прогноза к реальным значениям временного ряда. Временная же оценка представляет, насколько быстро рассчитывается прогноз. В прикладных задачах, в которых в прогнозе наибольшую важность имеет скорость расчётов, точностью пренебрегают в пользу быстрогодействия. Т.е. применяют модели экстраполяции не самой высокой точности. В случаях, когда время дискретизации значений больше времени проведения расчёта, стремятся к более высокой точности моделей. И в таком случае эффективность модели прогнозирования оценивают по точности.

Существует множество показателей, по которым оценивается точность моделей прогнозирования. Расчёт точности прогнозной модели в процессе её разработки позволяет авторам оценивать качество модели для принятия решения о необходимости её доработки. Помимо этого, использование распространённых метрик позволяет сравнивать между собой разработанные модели разных исследователей.

В качестве основного показателя точности прогноза использовалась оценка RMSE, дополненная процентным нормированием – RMSPE:

$$RMSPE = \sqrt{\frac{1}{n-1} \sum_{t=1}^n (Y_t - T_t)^2} \frac{100}{T_{\max} - T_{\min}}, \quad (11)$$

где  $n$  – количество измерений;

$T_t$  – значение ряда в момент времени  $t$ ,

$Y_t$  – спрогнозированное значение в момент времени  $t$ .

В последующих формулах  $T_t$  и  $Y_t$  имеют те же значения.

Для оценки правильности прогноза в данной работе использовались и другие показатели, такие как POCID, Tendency, SLG, Aff, PDA, ARV.

POCID (Prediction on Change In Direction) является одним из показателей правильности прогноза, а именно показателем сохранения тенденций. Данный показатель принимает значения от 0% до 100%, причём чем больше показатель, тем лучше модель предсказывает ВР. POCID вычисляется по формуле:

$$POCID = 100 \frac{\sum_{t=1}^N D_t}{N}, \quad (12)$$

где  $D_t$  вычисляется по формуле:

$$D_t = \begin{cases} 1, & \text{если } (T_t - T_{t-1})(Y_t - Y_{t-1}) > 0 \\ 0, & \text{в остальных случаях} \end{cases} \quad (13)$$

Tendency – показатель несовпадения тенденций. Оценивает сходство и различие направлений изменения прогнозных и реальных значений известных элементов анализируемого ВР. Чем больше несовпадений по направленности,

тем значение *Tendency* выше, и тем хуже модель. Этот показатель принимает значения от 0 до 1 и вычисляется по формуле:

$$Tendency = \frac{h}{N}, \quad (14)$$

где  $h$  – количество отрицательных произведений  $(T_t - T_{t-1})(Y_t - Y_{t-1})$ .

SLG (Sum of Losses and Gains) определяется как среднее значение потерь и прибылей модели. Этот показатель принимает значения от  $-\infty$  до  $+\infty$ , чем больше значение, тем модель лучше. Вычисляется SLG по формуле:

$$SLG = \frac{\sum_{t=1}^N L_t}{N}, \quad (15)$$

где  $L_t$  вычисляется как:

$$L_t = \begin{cases} +|T_t - T_{t-1}|, & \text{если } (T_t - T_{t-1})(Y_t - Y_{t-1}) > 0 \\ -|T_t - T_{t-1}|, & \text{в остальных случаях} \end{cases} \quad (16)$$

Aff (показатель аффинитета, *affinity*) – средняя относительная ошибка прогнозирования, вычисляемая по формуле (17). Показатель аффинитета оценивает сходство и различие прогнозных и реальных значений известных элементов анализируемого ВР. Принимает значения от 0% (наилучшее значение) до 100% (наихудшее значение). Среди нескольких моделей лучшей будет та, у которой оба показателя Aff и *Tendency* меньше таких же показателей других моделей.

$$Aff = \frac{100}{N - k} \sum_{j=k+1}^N \left| \frac{Y^j - T^j}{T^j} \right| \quad (17)$$

PDA (Prediction in Direction and Accuracy) – показатель, позволяющий сравнить модели и найти лучшую относительно тенденций и точности. Это достигается благодаря максимизации ROCID и минимизации относительной погрешности. PDA тем больше, чем лучше модель относительно точности и тенденций. Показатель принимает значения от -1 до 1 и вычисляется по формуле (18).

$$PDA = \frac{\sum_{t=1}^N G_t}{N},$$

$$G_t = \begin{cases} 1 - \frac{re_t}{re_{\max}}, & \text{если } D_t = 1 \text{ и } re_t < re_{\max} \\ 0, & \text{если } D_t = 0 \text{ и } re_t \geq re_{\max} \\ -1 + \frac{re_t}{re_{\max}}, & \text{если } D_t = 0 \text{ и } re_t < re_{\max} \\ -1, & \text{если } D_t = 0 \text{ и } re_t \geq re_{\max} \end{cases}, \quad (18)$$

$$re_t = \frac{e_t}{T_t},$$

$$re_{\max} = 0.002$$

ARV (average relative variance) – средняя относительная дисперсия, позволяет оценить результаты прогнозирования модели применительно к предположению, что значения в будущем будут соответствовать медиане значений ряда на предыдущих шагах. Чем меньше значение ARV, тем точнее прогноз модели относительно прогноза, основанного на предположении, что в дальнейшем значения будут равны медиане значений ВР. Принимает значения от 0 (наилучшее значение) до 1 (наихудшее значение), вычисляется по формуле:

$$ARV = \frac{\sum_{t=1}^N e_t^2}{\sum_{t=1}^N (Y_t - \bar{T})^2}, \quad (19)$$

где  $e_t$  – абсолютная погрешность, т.е. разность между реальным и прогнозным значениями.

Полученные результаты прогнозирования будут оценены по этим показателям.

## 2.4 Подготовка данных для исследования

Одной из задач при анализе данных является предварительная подготовка этих данных. На тематических ресурсах об обработке данных публикуется множество наборов данных для анализа и прогнозирования. ВР важных социальных, экономических и прочих показателей можно найти на сайтах органов власти.

Помимо этого, можно использовать наборы искусственно сформированных данных.

### 2.4.1 Синтетические данные

Для тестирования разрабатываемого метода были сгенерированы искусственные данные. Значения ВР вычисляются по формулам (20) и (21) и подвергаются искусственному добавлению шума.

$$f_1(t) = 10 \sin \frac{2\pi t}{100}, \quad (20)$$

$$f_2(t) = \sin \frac{\pi t}{20} e^{0,02t} + 3. \quad (21)$$

Графики сформированных искусственных ВР представлены на рисунках 3 и 4. Далее  $f_1(t)$  обозначен как ВР 1, а  $f_2(t)$  – ВР 2.

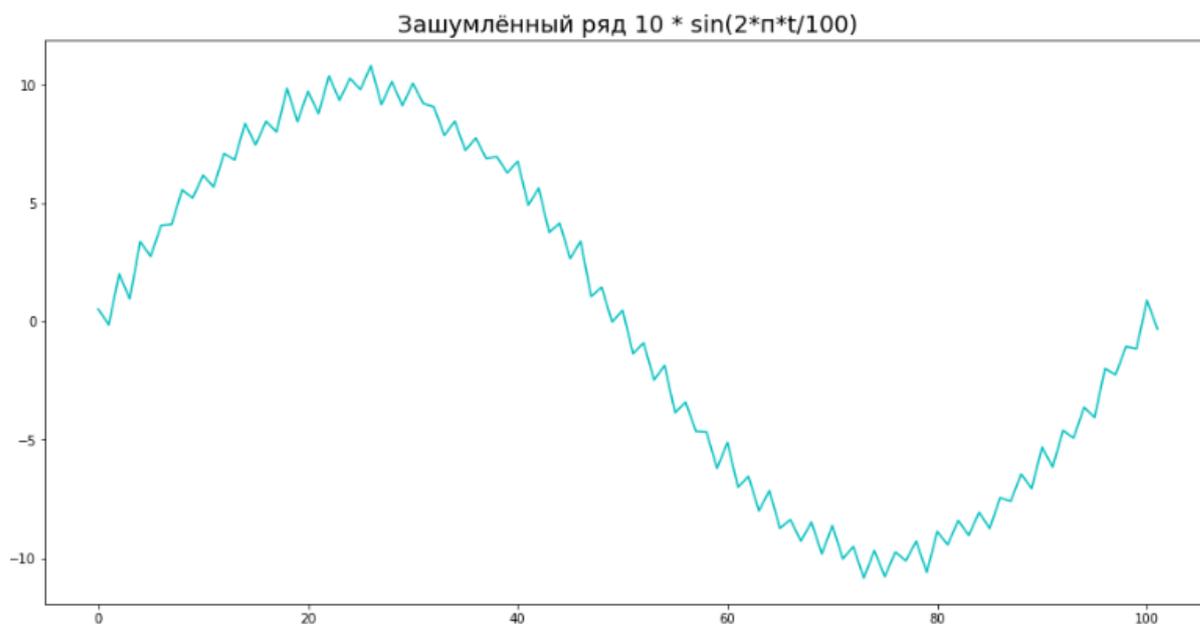


Рисунок 3 – Синтетический ВР 1



Рисунок 4 – Синтетический ВР 2

Описанные ВР содержат по 100 значений.

#### 2.4.2 Реальные данные

Уже несколько лет человечество противоборствует с вирусом, охватившим весь мир, и прогнозирование в этом вопросе является одним из шагов в борьбе с пандемией COVID-19.

Информацию о числе новых заболеваний COVID-19 в России можно получить с сайта Роспотребнадзора, где публикует ежедневные обновления. Однако формат представления данных на этом ресурсе недостаточно удобен для формирования датасета для прогноза.

Следующим источником, к которому обращались в ходе данной работы, был сайт Европейского центра профилактики и контроля заболеваний (ЕЦПКЗ, European Centre for Disease Prevention and Control, ECDC) [9]. ЕЦПКЗ размещал в открытом доступе датасеты по COVID-19 для всех стран (территорий) в мире, но на текущий момент данные не обновлялись более года. Среди поддерживаемых в актуальном состоянии данных есть данные о вакцинации, но Россия не входит в эту выборку.

Данные, использующиеся в этой работе, были взяты с сайта некоммерческого электронного проекта Our World in Data, который публикует в открытом доступе данные о глобальных проблемах человечества [13]. Из Excel-файла был сформирован CSV-файл с новыми случаями COVID-19 в России. ВР содержит значения за период с 31.01.2020 по 25.12.2021. Значения ВР визуализированы на рисунке 5.



Рисунок 5 – ВР ежедневно регистрируемых случаев COVID-19 в России

Всемирная организация здравоохранения относительно источников данных о новых заболеваниях, смертях и вакцинации, связанными с COVID-19 в России, ссылается на уже упомянутый ресурс Our World in Data и на данные от собственного Регионального Европейского Офиса [10].

Так же для прогнозирования был взят ряд индекса потребительских цен (ИПЦ) на все товары и услуги в Томской области в период с мая 2015 года по май 2021 года [14]. Ряд содержит значения для каждого месяца к тому же месяцу прошлого года (ИПЦ в мае 2021 года подсчитан относительно ИПЦ в мае 2020 года). График этого ВР представлен на рисунке 6.

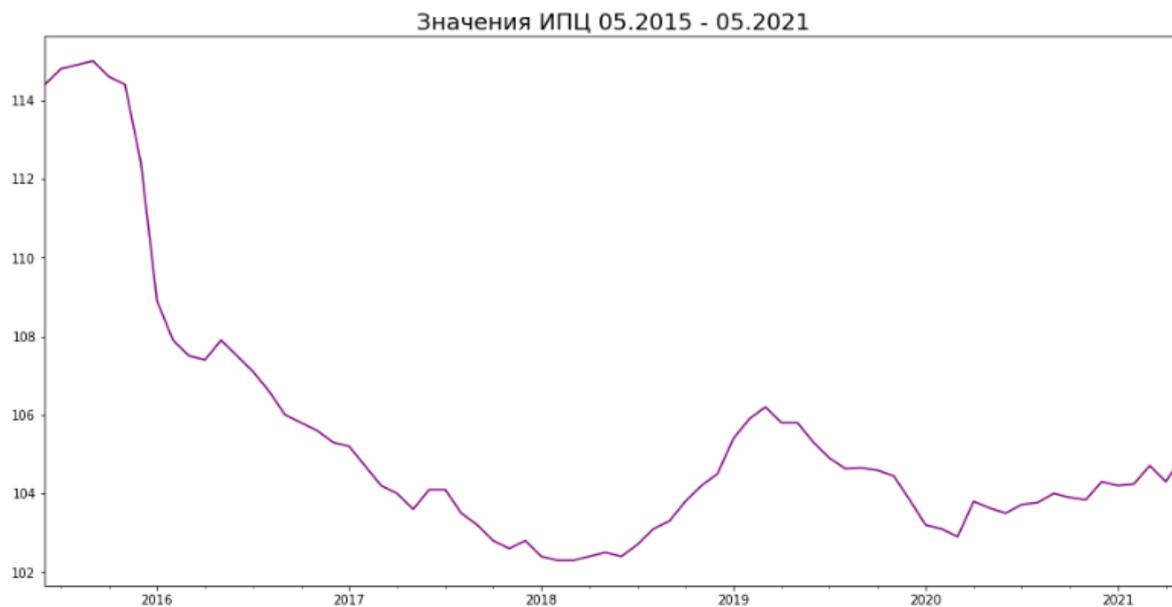


Рисунок 6 – ВР ИПЦ Томской области, 05.2015-05.2021

Подготовленные синтетические и реальные данные используются для прогнозирования разрабатываемым методом.

### 3 РАЗРАБОТКА ПРОГРАММЫ

Было решено разделить процесс разработки на две части. Сперва используемые алгоритмы реализуются независимо друг от друга, а следующим этапом происходит их объединение в гибридный метод.

Алгоритмы разработаны на языке программирования Python в облачной платформе Colaboratory. Python является популярным языком программирования для задач анализа и визуализации данных благодаря большому числу библиотек, предоставляющих удобные средства для обработки данных.

Платформа Colaboratory используется аналитиками и инженерами данных. Colaboratory является облачной средой, предоставляет возможность совместной работы и бесплатна для всех пользователей.

#### 3.1 Реализация прогнозирования на основе штрафного сплайна.

В качестве реальных данных, представляющих интерес для прогнозирования, используется ВР числа новых заболеваний COVID-19 в России. Данные, используемые в этой работе, взяты с сайта некоммерческого электронного проекта Our World in Data, который публикует в открытом доступе данные о глобальных проблемах человечества [13].

Алгоритм принимает на вход .CSV файл, в качестве выходных значений выступает ряд спрогнозированных данных, величина ошибки и график.

Алгоритм запускается для разных значений  $\rho \in [0; 1]$  – весового коэффициента, устанавливающего баланс между сглаживающими и интерполяционными свойствами сплайна:  $\rho = 0,2$ ,  $\rho = 0,5$  и  $\rho = 0,9$ .

Создан прогноз для заболеваемости COVID-19 в России. На рисунке А.1 приложения А представлен график, отражающий реальные данные, 3 варианта прогноза для разных  $\rho$ , и значения соответствующих RMSPE-ошибок. Для большей наглядности различий между кривыми произведено зумирование фрагментов с пиками рядов. Названия графиков содержат даты представленного периода,

для упрощения написаны затронутые месяцы без указания дней. На рисунках А.2-А.4 представлены эти фрагменты, наилучшая оценка  $RMSPE = 1,82\%$  соответствует прогнозу на основе сплайна с  $\rho = 0,5$ . Если сравнивать поведение этой кривой с кривой реальных данных, она довольно сглаженная, не имеет резких скачков. Но визуально красная кривая ( $\rho = 0,9$ ) больше соответствует кривой реальных данных, повторяя скачки.

В таблице 1 представлены значения показателей точности прогноза при  $\rho = 0,5$ . Здесь по метрике AFF и ARV прогноз оценивается как хороший, POCID и Tendency показывают средний результат.

Таблица 1 – Оценка точности прогноза COVID-19 при  $\rho = 0,5$

	<b>POCID</b>	<b>Tendency</b>	<b>SLG</b>	<b>AFF</b>	<b>PDA</b>	<b>ARV</b>
Худшее значение	0	1	$-\infty$	100	-1	1
<b>Результат</b>	<b>48,53</b>	<b>0,49</b>	<b>-41,28</b>	<b>5,51</b>	<b>-0,49</b>	<b>0,01</b>
Лучшее значение	100	0	$+\infty$	0	1	0

Прогноз для ИПЦ Томской области представлен на рисунке А.4 приложения А. Для данного ряда модель показала лучшее значение при  $\rho = 0,9$ , ошибка  $RMSPE$  составила  $5,02\%$ . Таблица 2 содержит оценку точности прогноза для данного ВР при параметре  $\rho = 0,9$ . Все показатели, кроме SLG и PDA, говорят о хорошем прогнозе, а SLG и PDA имеют средний результат.

Таблица 2 – Оценка точности прогноза ИПЦ Томской области при  $\rho = 0,9$

	<b>POCID</b>	<b>Tendency</b>	<b>SLG</b>	<b>AFF</b>	<b>PDA</b>	<b>ARV</b>
Худшее значение	0	1	$-\infty$	100	-1	1
<b>Результат</b>	<b>65,52</b>	<b>0,28</b>	<b>0,1</b>	<b>0,3</b>	<b>-0,16</b>	<b>0,1</b>
Лучшее значение	100	0	$+\infty$	0	1	0

При прогнозировании синтетических ВР наибольшая точность получена при коэффициенте  $\rho = 0,9$  для обоих рядов. Наименьшая ошибка  $RMSPE$  для ВР 1 составила  $1,86\%$ , а для ВР 2 – 4,  $37\%$ . Стоит отметить, что при прогнозе ВР 2 получен большой разброс в ошибке  $RMSPE$  для разных  $\rho$ , в то время как для трёх других описанных ВР ошибка  $RMSPE$  меняется меньше.

Таблица 3 – Оценка точности прогноза BP 1 при  $\rho = 0,9$ 

	<b>POCID</b>	<b>Tendency</b>	<b>SLG</b>	<b>AFF</b>	<b>PDA</b>	<b>ARV</b>
Худшее значение	0	1	$-\infty$	100	-1	1
<b>Результат</b>	<b>71,59</b>	<b>0,27</b>	<b>0,19</b>	<b>7,15</b>	<b>-0,27</b>	<b>0</b>
Лучшее значение	100	0	$+\infty$	0	1	0

Таблица 4 – Оценка точности прогноза BP 2 при  $\rho = 0,9$ 

	<b>POCID</b>	<b>Tendency</b>	<b>SLG</b>	<b>AFF</b>	<b>PDA</b>	<b>ARV</b>
Худшее значение	0	1	$-\infty$	100	-1	1
<b>Результат</b>	<b>70,45</b>	<b>0,28</b>	<b>0,24</b>	<b>24,8</b>	<b>-0,27</b>	<b>0,02</b>
Лучшее значение	100	0	$+\infty$	0	1	0

Для обоих рядов показатели POCID, Tendency, ARV представляют хорошие значения. SLG показывает среднее значение, PDA ближе к плохому и Aff – хорошее для BP 1 и среднее значение для BP 2.

В ходе работы сделан вывод о том, что для качественной оценки результатов прогнозирования необходимо учитывать несколько разных показателей.

### 3.2 Реализация сегментации данных

Относительно реализации сегментации в PPB задача была разделена на несколько возможных режимов. Это режим без запаздывания результата и режим с запаздыванием. В режиме сегментации с запаздыванием результата вводится граничное значение размера окна, которое выбирается из соображения реального времени и равно 10 значениям. Это означает, что выбранный момент разладки, который определяет длину сплайна, может иметь запаздывание в связи со скользящим окном переменного размера. Как уже было отмечено, максимальный размер окна в этом режиме равен 10, поскольку прошлые исследования показали, что это значение оптимально для параметра  $h$  сплайна. В связи с тем, что для сплайна третьего порядка минимально требуется 4 значения, размер окна сегментации и соответственно параметра сплайна  $h$  должны лежать в интервале от 4 до 10.

Помимо работы самого режима необходимо затронуть выбор входных параметров сегментации на основе алгоритма кумулятивных сумм. Установка значения чувствительности алгоритма зависит от желаемого результата и не связана со значениями ряда, что нельзя сказать о пороговом значении. Пороговое значение может быть указано на основе экспертной оценки, если заранее существуют ожидания о характеристиках ряда. Однако для достижения наибольшей автоматизации алгоритма прогнозирования, это значение также должно оптимизироваться внутри алгоритма. Для этого предлагается в качестве порога устанавливать значение, равное 10% от диапазона изменения первых  $n_0$  значений ряда (см. формулу 5), но больше 1 (см. формулу 8). Например, если первые  $n_0$  значений ряда при  $l = 4$  лежат в интервале от 100 до 120 единиц, то пороговое значение равно 5 единицам ( $1 + 1$ ), а если от 100 до 160, то 6 единицам.

На данный момент реализована сегментация в апостериорном режиме, т.е. сегментация на основе накопленных данных. В этом режиме все значения временного ряда поступают на обработку, для каждого значения рассчитывается кумулятивная сумма, происходит поиск правых и левых границ сегментов. На выходе получены границы сегментов для данного временного ряда и установленных входных параметров глубины памяти, порогового значения и переменных, отвечающих за чувствительность процедуры.

На рисунке Б.1 и Б.2 приложения Б представлен график исследуемого ВР COVID-19 и график кумулятивных сумм. График ВР изображён с границами сегментов, определёнными с помощью процедуры обнаружения разладки. Входные параметры:  $m = 1$ ,  $N = 3$ , пороговое значение 50, а глубина памяти  $n_0 = 8$ .

Рисунки Б.3 и Б.4 демонстрируют сегментированный ВР ИПЦ Томской области и график разладок для этого ВР. Входные параметры:  $m = 1$ ,  $N = 3$ , пороговое значение 5, а глубина памяти  $n_0 = 8$ .

Сегментированы также и синтетические ВР, ранее описанные в этой работе. Их графики и графики соответствующих кумулятивных сумм представлены на рисунках Б.5 – Б.8.

Границы сегментов обозначают моменты разладки, сегменты содержат однородные по характеристикам участки ряда. При реализации гибридного метода прогнозирования с текущей сегментацией данных параметр штрафного сплайна  $h$  должен подбираться исходя из сегментации.

Данная программная реализация может быть встроена в другие программно-математические средства, являясь частью системы обработки информации.

## **4 ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ**

### **4.1 Оценка коммерческого потенциала и перспективности проведения научных исследований с позиции ресурсоэффективности и ресурсосбережения**

#### **4.1.1 Потенциальные потребители результатов исследования**

Для анализа потенциальных потребителей результатов данного исследования необходимо определить целевой рынок среди всех сегментов рынка. Особенностью методов временных рядов в задаче прогнозирования является универсальность, т.е. прогнозировать можно временные ряды любой природы, будь то курс валюты, температура океана или распространение вируса. Анализ и прогнозирование временных рядов широко применяется во многих сферах: научные исследования, промышленность, финансы, медицина, сельское хозяйство и др. Пользователями результатов работы являются организации, прикладные задачи которых связаны с временными рядами.

#### **4.1.2 Анализ конкурентных технических решений**

Рынок продуктов интеллектуального анализа данных полон решений для прогнозирования временных рядов. Среди них библиотеки кода, надстройки Excel, математические пакеты с большим спектром функций и программное обеспечение конкретно для прогнозирования. Для анализа конкурентных решений был выделен программный продукт SAS Visual Forecasting.

Для оценки конкурентоспособности проектируемого решения была составлена оценочная карта (таблица 5), в которой приведены результаты сравнительного анализа решения с конкурентными решениями. Здесь индекс «ф» соответствует собственной разработке, «к1» - SAS Visual Forecasting.

Рассматриваемые решения были оценены по различным критериям по шкале от 1 до 5 баллов, где 5 – наивысшая оценка. Каждому критерию соответствует свой вес. Для анализа применяется формула:

$$K = \sum (B_i * B_i), \quad (22)$$

где K – конкурентоспособность решения;

$B_i$  – вес показателя в долях единицы;

$B_i$  – балл i-го показателя.

Таблица 5 – Оценочная карта критериев эффективности конкурентных технических решений (разработок)

Показатели конкурентоспособности	Вес критерия	Балльная оценка		Конкурентоспособность	
		$B_{\phi}$	$B_{K1}$	$K_{\phi}$	$K_{K1}$
Простота эксплуатации	0,1	5	3	0,5	0,3
Качество визуализации	0,1	4	5	0,4	0,5
Возможность ручной настройки параметров	0,15	3	5	0,45	0,75
Точность прогноза	0,3	4	5	1,2	1,5
Возможность внедрения в другие системы	0,2	5	3	1	0,6
Цена	0,15	5	3	0,75	0,45
Итого	1			<b>4,3</b>	4,1

Составленная оценочная карта позволяет сделать вывод о том, что проектируемое решение конкурентоспособно. Его основным преимуществом являются простота использования и возможность внедрения в другие системы. Слабой стороной является меньшая функциональная мощность по сравнению с конкурентами, но с расширением возможностей продукта усложняется его эксплуатация, от пользователей требуются дополнительные навыки. В качестве перспективы развития может быть поставлена задача расширения функциональности продукта с сохранением простоты эксплуатации.

### 4.1.3 SWOT-анализ

SWOT-анализ является инструментом анализа проектов с точки зрения соотнесения между собой внешней и внутренней среды проекта. В SWOT-анализе выявляют сильные и слабые стороны проектов, возможности и угрозы. Результатом такого анализа является:

- стратегия, при которой сильные стороны проекта привлекают возможности внешней среды;
- стратегия преодоления слабых сторон проекта за счёт возможностей внешней среды;
- стратегия минимизации внешних угроз посредством сильных сторон проекта;
- стратегия минимизации слабых сторон проекта для избежания внешних угроз.

В таблице 6 представлены результаты SWOT-анализа, включающие в себя возможные траектории развития проекта.

Таблица 6 – SWOT-анализ

	<b>Сильные стороны научно-исследовательского проекта:</b> С1. Простота эксплуатации С2. Относительно невысокая стоимость разработки С3. Возможность расширения функций С4. Простота интеграции	<b>Слабые стороны научно-исследовательского проекта:</b> Сл1. Ограниченность функциональных возможностей Сл2. Применимость только для краткосрочного прогнозирования
<b>Возможности:</b> В1. Может быть выпущено на рынок как самостоятельный продукт В2. Может быть внедрено в другие системы	Может быть выбрано такое направление развития, как интеграция решения в другие системы для отечественных заказчиков.	Вследствие высокого спроса разработка может быть актуальной, несмотря на ограниченность функций.

ВЗ. Увеличение спроса на отечественные разработки	Может быть выбрано такое направление развития, как выпуск простого в эксплуатации и не требующего особых навыков продукта для конечного пользователя, для физических лиц.	
<b>Угрозы:</b> У1. Появление решений с более высокой точностью прогноза У2. Ухудшение конкурентной позиции	Для достижения и сохранения высокого уровня конкурентоспособности функциональность продукта может быть расширена.	Во избежание перечисленных внешних угроз может быть исследована возможность добавления функции среднесрочного прогнозирования.

#### **4.2 Определение возможных альтернатив проведения научных исследований**

Возможные варианты проекта анализируются с помощью морфологического подхода. Морфологический подход нацелен на исследование разных вариантов строения проекта. Чтобы метод показал себя эффективным, необходимо раскрыть ключевые характеристики проекта и варианты этих характеристик. При комбинировании вариантов характеристик могут быть обнаружены уникальные версии проекта, представляющие практический и научный интерес. Для раскрытия возможных вариантов исполнения проекта была составлена морфологическая матрица, в которой перечислены возможные варианты характеристик проекта (таблица 7).

Таблица 7 – Морфологическая матрица проекта

	1	2	3
А. Язык программирования	Python	R	C#
Б. Программное решение	Модуль для интеграции	Отдельное десктоп-приложение	Отдельное web-приложение
В. Способ оптимизации параметров модели	Сегментация	Генетический алгоритм	Нейронные сети
Г. Формат исходных данных для анализа	Формат .CSV	Формат .TXT	Поддержка и формата .CSV, и формата .TXT

Интересными представляются следующие решения:

- А2Б1В3Г3;
- А1Б1В1Г1;
- А3Б2В2Г3.

В данной работе разрабатывается решение А1Б1В1Г1.

### 4.3 Планирование научно-исследовательских работ

#### 4.3.1 Определение трудоемкости выполнения работ

Для оценки стоимости выполнения научного исследования определяют трудовые затраты исполнителей проекта. Для этого рассчитывают трудоемкость работы в человеко-днях. Среднее значение трудоемкости рассчитывается по формуле:

$$t_{ож\ i} = \frac{3t_{\min\ i} + 2t_{\max\ i}}{5}, \quad (23)$$

где  $t_{ож\ i}$  – ожидаемая трудоемкость выполнения  $i$ -ой работы, чел.-дн.;

$t_{\min\ i}$  – минимально возможная трудоемкость выполнения заданной  $i$ -ой работы, чел.-дн.;

$t_{\max\ i}$  – максимально возможная трудоемкость выполнения заданной  $i$ -ой работы, чел.-дн.

Продолжительность каждой работы определяется её ожидаемой трудоёмкостью и числом исполнителей, работающих над ней:

$$T_{pi} = \frac{t_{ожi}}{Ч_i}, \quad (24)$$

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

$Ч_i$  – численность исполнителей, задействованных в этой работе одновременно, чел.

#### 4.3.2 Разработка графика проведения научного исследования

Одним из способов представления графика работ является диаграмма Ганта. Для её построения сперва необходимо определить временные показатели проведения научного исследования. Длительность работ в календарных днях рассчитывается по формуле:

$$T_{ки} = T_{pi} * k_{кал}, \quad (25)$$

где  $T_{ки}$  – продолжительность выполнения  $i$ -той работы в календарных днях, округляется до целого числа;

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

$k_{кал}$  – коэффициент календарности.

Коэффициент календарности рассчитывается по формуле:

$$k_{кал} = \frac{T_{кал}}{T_{кал} - T_{вых} - T_{пр}}, \quad (26)$$

где  $T_{кал}$  – количество календарных дней в году;

$T_{вых}$  – количество выходных дней в году;

$T_{пр}$  – количество праздничных дней в году.

Для пятидневной рабочей недели согласно производственному календарю 2022 года 365 календарных дней, из них 104 выходных дня и 14 праздничных. Исходя из этого  $k_{кал} = 1,48$ .

Продолжительность входящих в исследование работ представлена в таблице 8, где среди исполнителей через «НР» обозначен научный руководитель, «С» – студент.

Таблица 8 – Временные показатели проведения научного исследования

Название работы	Трудоёмкость работы			Исполнители	$T_p$ , рабочие дни	$T_k$ , календарные дни
	$t_{min}$ , чел-дни	$t_{max}$ , чел-дни	$t_{ож}$ , чел-дни			
Составление и утверждение темы бакалаврской работы	1	10	4,6	НР	4,6	7
Постановка целей и задач	2	5	3,2	НР	3,2	5
Изучение литературы по предметной области	10	14	11,6	С	11,6	18
Составление календарного плана-графика выполнения бакалаврской работы	1	1	1	С	1	2
Проектирование алгоритма	15	20	17	С	17	26
Поиск и подготовка данных для прогнозирования	1	3	1,8	НР, С	0,9	2
Реализация алгоритма	15	30	21	С	21	32
Тестирование работы алгоритма	10	15	12	С	12	18
Анализ результатов прогнозирования	5	10	7	НР, С	3,5	6
Выполнение других частей работы (финансовый менеджмент, социальная ответственность)	8	14	10,4	С	10,4	16
Составление пояснительной записки	3	5	3,8	С	3,8	6

На основе таблицы 8 построен календарный план-график в виде Диаграммы Ганта. План-график приведён на рисунке 7. На этом рисунке задачи, выполняемые только научным руководителем, обозначены зелёным цветом, только

студентом – голубым. Оранжевым цветом выделены задачи, над которыми работают оба исполнителя.

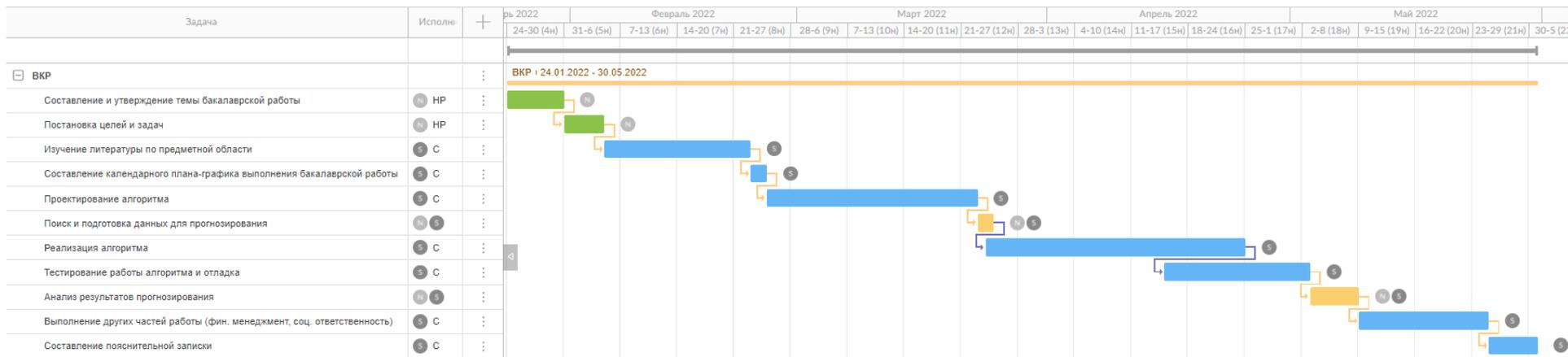


Рисунок 7 – Диаграмма Ганта проекта

Диаграмма составлена в сервисе планирования и управления проектами GanttPRO.

### 4.3.3 Бюджет научно-технического исследования (НТИ)

Для грамотного планирования бюджета НТИ необходимо в полной мере отразить все виды предполагаемых расходов. Бюджет данного НТИ состоит из материальных затрат НТИ, затрат на специальное оборудование, основной и дополнительной заработной платы, страховых отчислений, накладных расходов.

#### 4.3.3.1 Расчет материальных затрат НТИ

В материальные затраты данного НТИ включены траты на канцелярские принадлежности, а именно на бумагу для печати и для черновиков, ручку и канцелярскую папку. Для расчёта материальных затрат необходимо пользоваться формулой:

$$Z_M = (1 + k_T) * \sum_{i=1}^m C_i * N_{расхi}, \quad (27)$$

где  $m$  – количество материальных ресурсов  $i$ -го вида;

$C_i$  – цена приобретения единицы  $i$ -го вида ресурсов;

$N_{расхi}$  – цена приобретения  $i$ -го вида ресурсов;

$k_T$  – коэффициент транспортных расходов.

Транспортные расходы на доставку материалов отсутствуют, поэтому  $k_T$  принят равному 0. Затраты по данной группе расходов приведены в таблице 9.

Таблица 9 – Материальные затраты

Наименование	Единица измерения	Количество	Цена за ед., руб.	Затраты на материалы, руб.
Бумага А4	уп.	1	500	500
Ручка шариковая	шт.	1	50	50
Папка	шт.	1	200	200
Итого:				750

Поскольку стоимость оборудования не превышает 40 тыс. руб., амортизационные отчисления не рассчитываются.

#### 4.3.3.2 Расчет затрат на специальное оборудование для научных (экспериментальных) работ

Для проведения работ по данному исследованию необходим ноутбук, его стоимость включается в затраты на специальное оборудование для научных работ (таблица 10).

Таблица 10 – Затраты на специальное оборудование

Наименование оборудования	Количество единиц оборудования	Цена единицы, руб.	Общая стоимость оборудования, руб.
Ноутбук	1	38000	38000
Компьютерная мышь	1	800	800
Итого:			38800

Поскольку стоимость оборудования не превышает 40 тыс. руб., амортизационные отчисления не рассчитываются.

#### 4.3.3.3 Основная заработная плата исполнителей темы

Эта статья расходов включает в себя основную заработную плату научных и инженерно-технических работников, участвующих в выполнении работ. Над данным исследованием работы проводят два исполнителя: научный руководитель и студент. Величина расходов на основную заработную плату определяется исходя из трудоёмкости работ и текущей системы окладов. Научный руководитель, задействованный в этом проекте, занимает должность доцента и имеет степень кандидата наук, его оклад, определяемый ТПУ, составляет 37700 рублей. Оклад студента принимается равным окладу инженера низкой квалификации и составляет 19200 рублей.

Формула расчёта основной заработной платы:

$$Z_{\text{осн}} = Z_{\text{дн}} * T_p, \quad (28)$$

где  $Z_{\text{осн}}$  – основная заработная плата, руб.;

$Z_{\text{дн}}$  – среднедневная заработная плата работника, руб.;

$T_p$  – продолжительность работ, выполняемых работником, раб. дн.  
(таблица 8).

Среднедневная заработная плата рассчитывается по формуле:

$$Z_{\text{дн}} = \frac{Z_{\text{м}} * M}{F_{\text{д}}}, \quad (29)$$

где  $Z_{\text{м}}$  – месячный оклад сотрудника;

$M$  – количество месяцев работы без отпуска в течение года;

$F_{\text{д}}$  – действительный годовой фонд рабочего времени, раб.дн.

При пятидневной рабочей неделе и отпуске в 24 рабочих дня  $M = 11,2$ .  
Баланс рабочего времени для пятидневной рабочей недели в 2022 году приведён в таблице 11.

Таблица 11 – Баланс рабочего времени

Показатели рабочего времени	Научный руководитель	Студент
Календарное число дней	365	365
Количество нерабочих дней (выходные дни, праздничные дни)	118	118
Потери рабочего времени (отпуск)	24	24
Действительный годовой фонд рабочего времени	223	223

Месячный оклад  $Z_{\text{м}}$  рассчитывается по формуле:

$$Z_{\text{м}} = Z_{\text{оклад}} * (1 + k_{\text{пр}} + k_{\text{д}}) * k_{\text{р}}, \quad (30)$$

где  $Z_{\text{оклад}}$  – оклад сотрудника;

$k_{\text{пр}}$  – преимальный коэффициент;

$k_{\text{д}}$  – коэффициент доплат (за вредные условия);

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

Для данной работы доплаты и премии не предусмотрены, поэтому  $k_{\text{д}}$  и  $k_{\text{р}}$  в расчётах не учитываются. Расчёт основной заработной платы приведён в таблице 12.

Таблица 12 – Расчёт основной заработной платы

Исполнитель	$Z_{\text{оклад}}$ , руб.	$k_{\text{пр}}$	$k_{\text{д}}$	$k_{\text{р}}$	$Z_{\text{м}}$ , руб.	$Z_{\text{дн}}$ , руб.	$T_{\text{р}}$ , раб. дни	$Z_{\text{осн}}$ , руб.
Научный руководитель	37700	-	-	1,3	49010	2461,49	12	29537,87
Студент	19200	-	-	1,3	24960	1253,60	82	102794,91
Итого:								132332,77

#### 4.3.3.4 Дополнительная заработная плата исполнителей темы

Трудовым кодексом РФ предусмотрена дополнительная заработная плата при совмещении с обучением, при предоставлении отпуска и при выполнении прочих условий. Размер дополнительной заработной платы определяется формулой:

$$Z_{\text{доп}} = k_{\text{доп}} * Z_{\text{осн}}, \quad (31)$$

где  $Z_{\text{доп}}$  – дополнительная заработная плата;

$k_{\text{доп}}$  – коэффициент дополнительной заработной платы.

В рамках научной работы  $k_{\text{доп}}$  был принят 0,12. Размер дополнительной заработной платы для научного руководителя и студента внесены в таблицу 13.

#### 4.3.3.5 Отчисления во внебюджетные фонды (страховые отчисления)

Эта статья расходов состоит из обязательных страховых отчислений в Фонд социального страхования, Пенсионный фонд, Федеральный фонд обязательного медицинского страхования. Сумма отчислений вычисляется по формуле:

$$Z_{\text{внеб}} = k_{\text{внеб}} * (Z_{\text{осн}} + Z_{\text{доп}}), \quad (32)$$

где  $Z_{\text{внеб}}$  – величина отчислений, руб.;

$k_{\text{внеб}}$  – коэффициент отчислений.

Для учреждений, осуществляющих образовательную и научную деятельность,  $k_{\text{внеб}} = 27,1\%$ . Расчёт страховых отчислений представлен в таблице 13.

Таблица 13 – Отчисления во внебюджетные фонды

<b>Исполнитель</b>	<b>З<sub>осн</sub>, руб.</b>	<b>З<sub>доп</sub>, руб.</b>
Научный руководитель	29537,87	3544,54
Студент	102794,91	12335,39
<i>k<sub>внеб</sub></i>	27,1%	
	<b>Итого:</b>	<b>40165,65</b>

Сумма отчислений во внебюджетные фонды составляет 40165,65 рублей.

#### **4.3.3.6 Накладные расходы**

Накладные расходы учитывают затраты, не вошедшие в ранее перечисленные статьи расходов. К накладным расходам относится оплата услуг связи, оплата электроэнергии, расходы на печать и пр. Величина накладных расходов определяется как доля от суммы рассчитанных ранее статей расходов. Коэффициент накладных расходов в данной работе равен 16%.

#### **4.3.3.7 Формирование бюджета затрат научно-исследовательского проекта**

На основе величины затрат на НТИ формируется бюджет всех затрат проекта, который представляет собой минимальную сумму затрат при формировании договора с заказчиком. Определение бюджета затрат на НТИ приведено в таблице 14.

Таблица 14 – Расчёт бюджета затрат НТИ

<b>Наименование статьи</b>	<b>Сумма, руб.</b>	<b>Примечание</b>
1. Расчет материальных затрат НТИ	750	Раздел 4.3.3.1
2. Расчет затрат на специальное оборудование для научных (экспериментальных) работ	38800	Раздел 4.3.3.2
3. Основная заработная плата исполнителей темы	132332,77	Раздел 4.3.3.3
4. Дополнительная заработная плата исполнителей темы	15879,93	Раздел 4.3.3.4

Наименование статьи	Сумма, руб.	Примечание
5. Отчисления во внебюджетные фонды (страховые отчисления)	40165,65	Раздел 4.3.3.5
6. Накладные расходы	36468,54	16% от суммы ст. 1-5
7. Бюджет затрат НИИ	264396,89	Сумма ст. 1-6

Общий бюджет разработки составляет 264397 рублей.

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

Эффективность проекта можно оценить на основе интегрального показателя эффективности научного исследования, расчёт этого показателя учитывает финансовую эффективность и ресурсоэффективность.

Интегральный финансовый показатель разработки определяется по формуле:

$$I_{\text{финр}}^{\text{исп.}i} = \frac{\Phi_{ri}}{\Phi_{\text{max}}}, \quad (33)$$

где  $I_{\text{финр}}^{\text{исп.}i}$  – интегральный финансовый показатель разработки;

$\Phi_{ri}$  – стоимость  $i$ -го варианта исполнения;

$\Phi_{\text{max}}$  – максимальная стоимость исполнения.

Интегральный показатель ресурсоэффективности  $I_{ri}$  определяется формулой:

$$I_{ri} = \sum a_i * b_i, \quad (34)$$

где  $a_i$  – весовой коэффициент  $i$ -го варианта исполнения разработки;

$b_i$  – оценка  $i$ -го варианта в баллах по шкале от 1 до 5, где 5 – высшая оценка.

Расчёт интегрального показателя ресурсоэффективности представлен в таблице 15. Выбранное к реализации исполнение 1 сравнивается с исполнением 2 (см. раздел 4.2).

Таблица 15 – Сравнительная оценка характеристик вариантов исполнения проекта

<b>Критерий</b>	<b>Весовой коэффициент критерия</b>	<b>Исп.1</b>	<b>Исп.2</b>
Простота разработки	0,3	4	3
Удобство эксплуатации	0,2	5	4
Простота внедрения в другие системы	0,2	5	3
Точность результатов прогнозов	0,3	5	5
Итого	1	4,7	3,8

Интегральный показатель эффективности исполнения рассчитывается как отношение интегрального показателя ресурсоэффективности ( $I_{pi}$ ) к интегральному финансовому показателю разработки ( $I_{финр}^{исп.i}$ ). Сравнительная эффективность проектов вычисляется отношением разных интегральных показателей эффективности друг к другу. Рассчитанные показатели представлены в таблице 16.

Таблица 16 – Сравнительная эффективность разработки

<b>Показатели</b>	<b>Исп.1</b>	<b>Исп.2</b>
Интегральный финансовый показатель разработки	0,94	1,00
Интегральный показатель ресурсоэффективности разработки	4,70	3,80
Интегральный показатель эффективности	4,99	3,80
Сравнительная эффективность вариантов исполнения	1,31	0,76

Исходя из рассчитанных показателей наиболее выгодным с точки зрения ресурсоэффективности и финансовой эффективности является исполнение 1, принятое к реализации.

## **5 СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ**

В рамках данной ВКР разрабатывается программно-математическое средство прогнозирования временных рядов на основе гибридной модели и текущей сегментации данных, поступающих в реальном времени. Разрабатываемое решение может использоваться в исследовательской деятельности и в любой сфере, требующей прогнозирования временных рядов. Рабочее место исполнителя работы представляет собой офисное помещение размером 3 м \*6 м и оборудовано 1 ПЭВМ. Работа включает в себя умственную деятельность, изучение тематической литературы, написание программного кода. Трудовой процесс характеризуется физическим напряжением, вызванным долгим нахождением в позе сидя, умственным напряжением.

Раздел посвящён рассмотрению правовых и организационных вопросов обеспечения безопасности, вредных и опасных производственных факторы, которые могут возникать в процессе разработки, а также мер по их предотвращению. Проанализированы аспекты экологической безопасности и возможные чрезвычайные ситуации.

### **5.1 Правовые и организационные вопросы обеспечения безопасности**

#### **5.1.1 Правовые нормы трудового законодательства**

Вопросы охраны труда, трудоустройства, заработной платы, а также права и обязанности работника и работодателя устанавливаются в соответствии с Трудовым кодексом РФ [19].

Продолжительность рабочего времени не должна превышать 40 часов в неделю. Для отдельных категорий работников устанавливается сокращённая продолжительность рабочей недели, а именно: для работников до 16 лет – не более 24 часов в неделю, для работников от 16 до 18 лет, а также для работников I и II группы инвалидности – не более 35 часов в неделю. При работе во вредных

условиях 3 и 4 степени и при работе в опасных условиях труда продолжительность работы не должна превышать 36 часов в неделю. Для конкретного работника продолжительность рабочего времени устанавливается трудовым договором на основании отраслевого (межотраслевого) соглашения и коллективного договора, при этом учитываются результаты специальной оценки условий труда. При работе в ночное время (в период с 22 часов до 6 часов) длительность смены сокращается на один час.

В течение рабочего дня предоставляется перерыв для питания и отдыха, это время не включается в рабочее время. Продолжительность перерыва, время его начала определяются внутренним трудовым распорядком. Сотруднику положен ежегодный оплачиваемый отпуск продолжительностью 28 дней с сохранением должности и среднего заработка. Условия предоставления более длительного основного отпуска и дополнительного отпуска установлены федеральными законами. Помимо перечисленных предусмотрены такие виды отдыха, как межсменный, выходные дни, нерабочие праздничные дни.

### **5.1.2 Эргономические требования к правильному расположению и компоновке рабочей зоны**

Компоновка рабочего места должна соответствовать характеру работы, антропометрическим и психологическим требованиям. Правильное взаимное расположение элементов рабочей зоны, удобная и физиологически верная поза человека повышают эффективность работы и предотвращают развитие профессиональных заболеваний.

Площадь постоянного рабочего места сотрудника, работающего с ПК, должна составлять не менее 4,5 м<sup>2</sup>. Световая среда помещения должна регулироваться оконными проёмами [18]. Рассматриваемая разработка предполагает длительное взаимодействие с ПК, что подразумевает сидячий характер работы. Регулированием высоты рабочей поверхности и пространства для ног исходя из роста сотрудника достигается его оптимальное положение [16]. Если рабочая

поверхность нерегулируемая, варьируется высота сиденья кресла. Кресло сотрудника, работающего с ПК, должно обеспечивать физиологически рациональную рабочую позу для сохранения позвоночника в естественном состоянии. Кресло должно иметь механизм регулировки высоты сиденья и угла наклона спинки, может дополнительно комплектоваться подлокотниками и подставкой для ног [17].

## 5.2 Производственная безопасность

При разработке проектируемого решения могут возникать вредные и опасные производственные факторы, их перечень, а также регламентирующие их действие нормативные документы представлены в таблице 17.

Таблица 17 – Возможные опасные и вредные производственные факторы при разработке программно-математического средства

Факторы (ГОСТ 12.0.003-2015 [20])	Нормативные документы
<b>Вредные производственные факторы:</b>	
Отсутствие или недостаток необходимого естественного освещения.	СП 52.13330.2016 Естественное и искусственное освещение. Актуализированная редакция СНиП 23-05-95*
Факторы, связанные с аномальными микроклиматическими параметрами воздушной среды на местонахождении работающего.	СанПиН 1.2.3685-21 Гигиенические нормативы и требования к обеспечению безопасности и (или) безвредности для человека факторов среды обитания
Факторы, обладающие свойствами психофизиологического воздействия на организм человека: физические статические перегрузки, связанные с рабочей позой.	ГОСТ 12.2.032-78 ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования.
<b>Опасные производственные факторы:</b>	
Факторы, связанные с электрическим током, вызываемым разницей электрических потенциалов, под действие которого попадает работающий.	ГОСТ 12.1.038-82 ССБТ. Электробезопасность. Предельно допустимые уровни напряжений прикосновения и токов. Федеральный закон от 22.07.2013 г. №123 – ФЗ, Технический регламент о требованиях пожарной безопасности.

Далее для каждого из перечисленных факторов будут рассмотрены источники возникновения, воздействие на человека, допустимые нормы и средства индивидуальной защиты (СИЗ).

### **5.2.1 Отсутствие или недостаток необходимого естественного освещения**

При постоянной работе в помещении с ПК может возникать недостаток или отсутствие естественного освещения, что повышает уровень утомляемости и вызывает снижение работоспособности. При дефиците естественного света могут развиваться заболевания зрительного аппарата.

В соответствии с СП 52.13330.2016 [24] зрительная работа, выполняемая исполнителем проектируемого решения за ПК в жилом помещении или общественном здании, может быть охарактеризована как зрительная работа разряда Б подразряда 1. Это работа с наименьшим или эквивалентным размером объекта различения 0,3-0,5 мм с относительной продолжительностью зрительной работы при направлении зрения на рабочую поверхность не менее 70%. Для зрительной работы разряда Б-1 коэффициент естественного освещения (КЕО) должен соответствовать следующим требованиям:

- при верхнем или комбинированном освещении:  $КЕО = 3\%$ ;
- при боковом освещении:  $КЕО = 1\%$ .

Для минимизации воздействия недостатка естественного освещения может применяться расширение оконного проёма, создание дополнительных оконных проёмов.

### **5.2.2 Факторы, связанные с аномальными микроклиматическими параметрами воздушной среды на местонахождении работающего**

Микроклиматические условия рабочей зоны оказывают влияние на утомляемость и работоспособность человека. Аномальные значения температуры воздуха, влажности и подвижности воздуха могут приводить к ухудшению состояния работника и развитию болезней. Так при слишком низких

температурах человек подвержен переохлаждению, что ведёт к снижению иммунитета. Недостаточный уровень влажности вызывает обезвоживание организма.

Причиной неудовлетворительных показателей микроклимата может являться некорректная работа системы вентиляции в помещении.

Нормативные показатели микроклимата регламентируются

Оптимальные и допустимые значения показателей микроклимата регламентируются СанПиН 1.2.3685-21 Гигиенические нормативы и требования к обеспечению безопасности и (или) безвредности для человека факторов среды обитания [21]. Работа, выполняемая исполнителем проектируемого решения, по уровню энергозатрат организма, классифицируется как работа категории Ia, т.к. не требует физического напряжения, постоянного передвижения и ходьбы. В таблице 18 приведены допустимые значения микроклиматических показателей в помещении для данной категории в холодный и тёплый период года.

Таблица 18 – Допустимые величины параметров микроклимата на рабочих местах в помещениях, категория энергозатрат Ia

Период года	Температура воздуха, °С		Температура поверхностей, °С	Относительная влажность воздуха, %	Скорость движения воздуха, м/с	
	диапазон ниже оптимальных величин	диапазон выше оптимальных величин			для диапазона температур воздуха ниже оптимальных величин, не более	для диапазона температур воздуха выше оптимальных величин, не более
Холодный	20,0-21,9	24,1-25,0	19,0-26,0	15-75	0,1	0,1
Тёплый	21,0-22,9	25,1-28,0	20,0-29,0	15-75	0,1	0,2

Для приведения показателей микроклимата к допустимым значениям необходимо наличие в помещении кондиционеров, обогревателей, увлажнителей воздуха. С помощью указанных приборов и/или окон может регулироваться уровень влажности, температуры, скорости воздуха. Рациональная компоновка рабочих мест также влияет на микроклимат.

### **5.2.3 Факторы, обладающие свойствами психофизиологического воздействия на организм человека: физические статические перегрузки, связанные с рабочей позой**

Работа над исследуемым проектом малоподвижна и подразумевает сидячую позу. Низкий уровень физической активности негативно влияет на состояние организма человека и способствует атрофии мышц, повышенной утомляемости, сонливости. Если же при этом сотрудник принимает физиологически неправильную позу, у него могут развиваться заболевания опорно-двигательного аппарата и возникать боли в шее, спине и других частях тела.

Сотруднику необходимо в течение дня делать перерывы на лёгкую гимнастику, направленную на разминку мышц, суставов, улучшение кровообращения. Регулярное выполнение упражнений благотворно влияет на производительность человека.

### **5.2.4 Факторы, связанные с электрическим током, вызываемым разницей электрических потенциалов, под действие которого попадает работающий**

Офисная техника, которой пользуется исполнитель проектируемого решения, работает от электросети, в связи с чем может являться источником опасности. Электрический ток оказывает механическое, электролитическое, биологическое и термическое воздействие, приводит к травмам и гибели.

Вероятность поражения током возрастает при высокой влажности, отсутствии защитных конструкций для проводов.

ГОСТ 12.1.038-82 ССБТ. Электробезопасность. Предельно допустимые уровни напряжений прикосновения и токов. [22] устанавливает, что при переменном токе частотой 50 Гц напряжение не должно превышать 2 В, а сила тока должна быть не более 0,3 мА.

Превентивными мерами в данном вопросе являются регулярные инструктажи техники безопасности на рабочем месте, работа с ПК сухими руками,

периодическая проверка целостности и исправности электроприборов, розеток, кабелей. Необходимо наличие первичных средств пожаротушения (огнетушителей) [25].

### **5.3 Экологическая безопасность**

В разработке используется такая офисная техника, как системные блоки, мониторы, ноутбуки, клавиатуры, компьютерные мыши. Перечисленные приборы состоят из множества деталей, которые созданы из разных веществ, поэтому разные детали относятся к разным классам опасности отходов. Некоторые компоненты офисной техники содержат тяжёлые металлы, которые загрязняют почву, т.е. негативно воздействуют на литосферу. Независимо от класса опасности отходов, должна проводиться их грамотная утилизация. Все отходы должны собираться, сортироваться и быть направлены на переработку организациям, имеющими государственную лицензию на обращение с отходами.

Как уже было отмечено, проектируемое решение разрабатывается с использованием электротехники, а значит использует электроэнергию, выработанную на электростанциях. Выбросы электростанций загрязняют селитебную зону района расположения этих электростанций. Негативно воздействуют выбросы электростанций и на атмосферу, гидросферу. Самыми распространёнными являются тепловые электростанции, которые выбрасывают в атмосферу продукты сгорания топлива и примеси различного состава (углекислый газ, оксид углерода, оксиды металлов, свинец, ртуть и др.). Водоёмы загрязняются сбросами тёплой воды с примесями, что ведёт к нарушению экосистемы водоёмов, сокращению биоразнообразия.

Для сокращения последствий работы электростанций в рамках данной разработки необходимо использовать ПК с режимом пониженного потребления электроэнергии для сокращения потребления электроэнергии во время простоя ПК.

## 5.4 Безопасность в чрезвычайных ситуациях

При работе над проектируемым решением могут возникнуть такие чрезвычайные ситуации:

- техногенные (взрывы, пожары, обрушение помещений);
- природные (наводнения, ураганы, природные пожары);
- биологические (эпидемии, пандемии);
- социальные (террор).

Наиболее вероятной чрезвычайной ситуацией является пожар в помещении рабочей зоны, вызванный перегревом и воспламенением электротехники, неисправностью электросети, током короткого замыкания. Пожар может возникнуть из-за несоблюдения правил пожарной безопасности.

В соответствии с Федеральным законом "Технический регламент о требованиях пожарной безопасности" [25] для защиты от воздействия пожара должны соблюдаться следующие нормы:

- применение ограничивающих распространение огня средств;
- использование устройств оповещения и обнаружения пожара;
- применение огнетушителей и других первичных средств пожаротушения;
- наличие и возможность использования эвакуационных путей и пр.

В случае обнаружения пожара необходимо сообщить о происшествии в пожарную охрану. Далее необходимо принять меры по эвакуации людей и материальных ценностей согласно плану эвакуации. До прибытия сотрудников пожарной охраны нужно самостоятельно начать тушение пожара с помощью первичных средств пожаротушения. Прибывшей на место пожарной охране необходимо оказывать содействие в тушении пожара.

В помещении возможен пожар твёрдых горючих веществ и материалов, т.е. пожар класса А. В целях борьбы с пожаром должны использоваться хранящиеся в офисе такие первичные средства пожаротушения, как переносные огнетушители и покрывала для изоляции очага возгорания [25].

## Вывод по разделу

Фактические значения значимых для разработки производственных факторов удовлетворяют нормативным значениям.

В помещении, где ведётся работа над проектируемым решением, нет условий для возникновения повышенной или особой опасности, поэтому по ПУЭ [26] оно отнесено к 1 категории электробезопасности. Персонал, занятый разработкой, относится к I группе по электробезопасности согласно «Правилам по охране труда при эксплуатации электроустановок» [27]. Работа, выполняемая исполнителем проектируемого решения, по уровню энерготрат организма, классифицируется как работа категории Ia, т.к. не требует физического напряжения, постоянного передвижения и ходьбы [21]. По взрывопожарной и пожарной опасности рабочее помещение имеет категорию В, т.е. является пожароопасным помещением с наличием горючих веществ и материалов [28]. Продукт разработки относится к 4 категории объектов, оказывающих значительное негативное воздействие на окружающую среду [29].

## ЗАКЛЮЧЕНИЕ

В ходе работы над ВКР было изучено прогнозирование ВР, рассмотрена классификация подходов к прогнозированию ВР. На языке программирования Python был реализован алгоритм прогнозирования на основе штрафного сплайна в реальном масштабе времени. Получены прогнозы для реальных данных и искусственно сгенерированных. Качество прогноза с точки зрения точности оценивается по ошибке RMSPE, а также по таким показателям тенденций и точности как ROCID, Tendency, SLG, Aff, PDA, ARV. Для реальных ВР лучшие значения RMSPE составляют для ряда ИПЦ – 5,02%, для ряда по заражениям COVID-19 – 1,82%, что является хорошим результатом прогнозирования. Для оптимизации параметров сплайна было решено использовать сегментацию ВР на основе метода, представляющего собой модификацию алгоритма кумулятивных сумм. Алгоритмом фиксируются разладки во временном ряду, то есть моменты смены распределения величин. Была реализована сегментация в апостериорном режиме.

## CONCLUSION

During the work on the graduation work, time series forecasting was studied, the classification of approaches to time series forecasting was considered. A prediction algorithm based on a real-time penalty spline was implemented in the Python programming language. Predictions are obtained for real data and artificially generated ones. The quality of the forecast in terms of accuracy is assessed by RMSPE error, as well as by such indicators of trends and accuracy as POCID, Tendency, SLG, Aff, PDA, ARV. For real time series, the best RMSPE values are 5.02% for the CPI series and 1.82% for the COVID-19 infection series, which is a good forecasting result. To optimize the spline parameters, it was decided to use time series segmentation based on a method that is a modification of the CUSUM algorithm. The algorithm fixes discords in the time series, that is, the moments of change in the distribution of values. Segmentation was implemented in a posteriori mode.

## СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

1. Чучуева И.А. МОДЕЛЬ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ ПО ВЫБОРКЕ МАКСИМАЛЬНОГО ПОДОБИЯ, диссертация... канд. тех. наук / Московский государственный технический университет им. Н.Э. Баумана. Москва, 2012.
2. Низамитдинов А.И. Использование штрафного сплайна (p-spline) и кубического сглаживающего сплайна в прогнозировании временных рядов// Вестник Таджикского государственного университета права, бизнеса и политики. Серия гуманитарных наук - 2012. - № 4. - [С. 124-131]. URL: <https://ezproxy.ha.tpu.ru:2154/item.asp?id=18259124> (дата обращения 28.04.2022)
3. Cluster Analysis for Data Mining and System Identification János Abonyi, Balázs Feil (2007) URL: <https://link.springer.com/book/10.1007/978-3-7643-7988-9> (дата обращения: 10.05.2022)
4. Page E.S. Continuous Inspection Schemes // Biometrika. 1954. № 1. P. 141-154.
5. M. A. Girshick. Herman Rubin. "A Bayes Approach to a Quality Control Model." Ann. Math. Statist. 23 (1) 114 - 125, March, 1952. URL: <https://doi.org/10.1214/aoms/1177729489> (дата обращения: 27.04.2022)
6. Ширяев А.Н. Об оптимальных методах в задачах скорейшего обнаружения // Теория вероятностей и ее применения. 1963. № 8. Вып. 1. С. 26 — 51.
7. Lorden G. Procedures for reacting to a change in distribution // Ann. Math. Stat. 1971. v. 42. N 6. P. 1897 — 1908.
8. С. Э. Воробейчиков, “Об обнаружении изменения среднего в последовательности случайных величин”, Автомат. и телемех., 1998, № 3, 50–56 URL: <http://mi.mathnet.ru/at2457> (дата обращения: 12.05.2022)

9. Download COVID-19 datasets // European Centre for Disease Prevention and Control. URL: <https://www.ecdc.europa.eu/en/covid-19/data> (дата обращения: 18.11.2021).
- 10.Data Information. URL: [covid19.who.int/data](https://covid19.who.int/data) (дата обращения: 15.04.2022).
- 11.Кочегурова, Е.А. Гибридный подход для краткосрочного прогнозирования временных рядов на основе штрафного P-сплайна и эволюционной оптимизации /Е.А. Кочегурова, Е.Ю. Репина, О.Б. Цехан // Компьютерная оптика. – 2020. – Т. 44, № 5. – С. 821- 829. – DOI: 10.18287/2412-6179-CO-667.
- 12.Kochegurova, E., I. Khozhaev, and E. Repina. 2020. Adaptive Time Series Prediction Model Based on a Smoothing P-Spline // Advances in Intelligent Systems and Computing. Vol. 1156 AISC. p. 445-455. doi:10.1007/978-3-030-50097-9\_45 [https://link.springer.com/chapter/10.1007%2F978-3-030-50097-9\\_45](https://link.springer.com/chapter/10.1007%2F978-3-030-50097-9_45)
- 13.Russia: Coronavirus Pandemic Country Profile// Our World in Data. URL: <https://ourworldindata.org/coronavirus/country/russia#how-many-tests-are-performed-each-day> (дата обращения: 18.01.2022).
- 14.Территориальный орган Федеральной службы государственной статистики по Томской области. URL: [tmsk.gks.ru](https://tmsk.gks.ru) (дата обращения: 19.06.2021)
- 15.Заикин Н. Д. Сегментация временных рядов для выделения шаблонов в больших массивах данных : бакалаврская работа / Н. Д. Заикин ; Национальный исследовательский Томский политехнический университет (ТПУ), Инженерная школа информационных технологий и робототехники (ИШИТР), Отделение информационных технологий (ОИТ) ; науч. рук. Е. А. Кочегурова. — Томск, 2021.
- 16.ГОСТ 12.2.032-78 ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования.
- 17.ГОСТ 21889-76 Система «человек-машина». Кресло человека-оператора. Общие эргономические требования.

- 18.СП 2.2.3670-20 «Санитарно-эпидемиологические требования к условиям труда»
- 19.Трудовой кодекс Российской Федерации от 30.12.2001 N 197-ФЗ (ред. от 25.02.2022)
- 20.ГОСТ 12.0.003-2015 ССБТ. Опасные и вредные производственные факторы. Классификация.
- 21.СанПиН 1.2.3685-21 Гигиенические нормативы и требования к обеспечению безопасности и (или) безвредности для человека факторов среды обитания
- 22.ГОСТ 12.1.038-82 ССБТ. Электробезопасность. Предельно допустимые уровни напряжений прикосновения и токов.
- 23.ГОСТ 12.1.019-2017 ССБТ. Электробезопасность. Общие требования и номенклатура видов защиты.
- 24.СП 52.13330.2016 Естественное и искусственное освещение. Актуализированная редакция СНиП 23-05-95\*
- 25.Федеральный закон от 22.07.2013 г. №123 – ФЗ, Технический регламент о требованиях пожарной безопасности.
- 26.ПРАВИЛА УСТРОЙСТВА ЭЛЕКТРОУСТАНОВОК Седьмое издание  
Раздел 1 ОБЩИЕ ПРАВИЛА Глава 1.1 ОБЩАЯ ЧАСТЬ
- 27.ПРИКАЗ от 15 декабря 2020 года N 903н Об утверждении Правил по охране труда при эксплуатации электроустановок.
- 28.СП 12.13130.2009 Определение категорий помещений, зданий и наружных установок по взрывопожарной и пожарной опасности.
- 29.Постановление Правительства РФ от 31 декабря 2020 г. N 2398 «Об утверждении критериев отнесения объектов, оказывающих негативное воздействие на окружающую среду, к объектам I, II, III и IV категорий».

**Приложение А**  
**(обязательное)**  
**Графики прогноза на основе Р-сплайна**

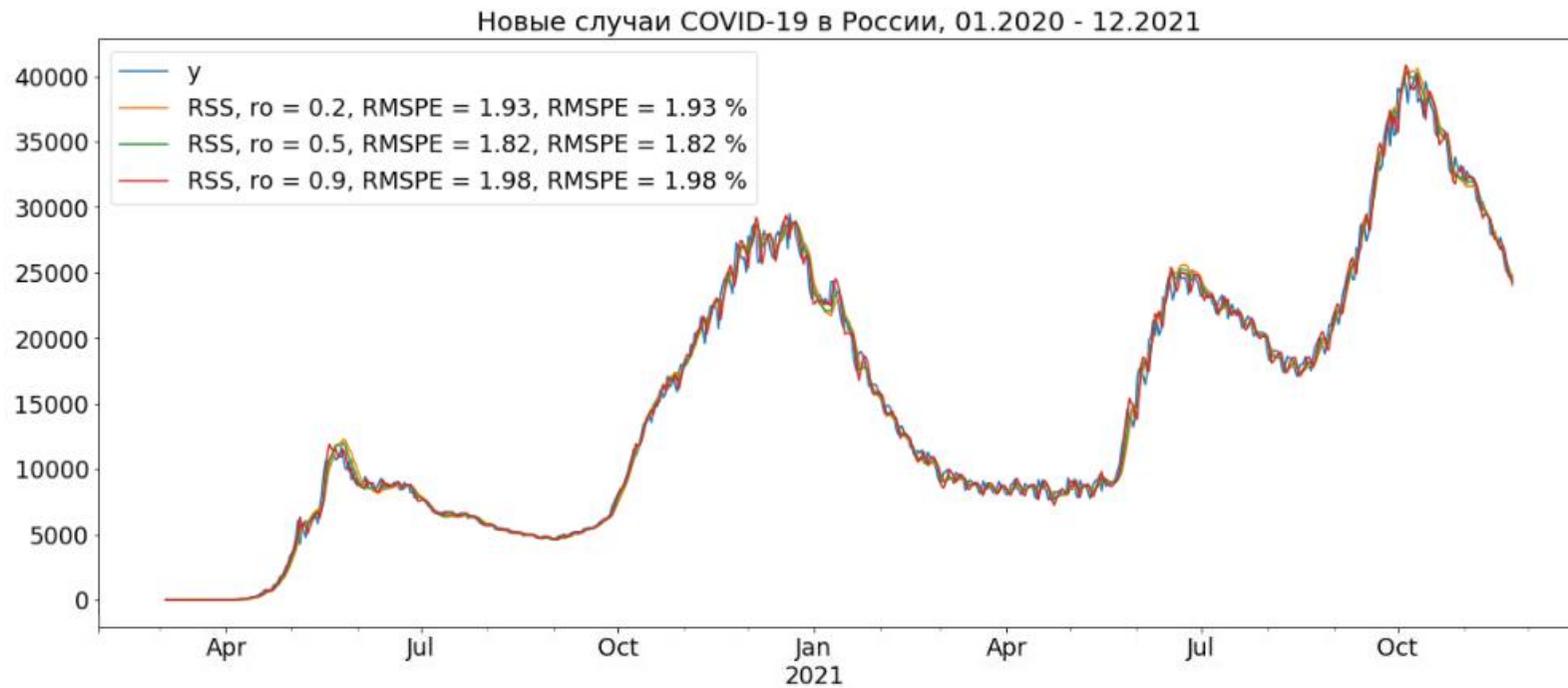


Рисунок А.1 – Новые случаи COVID-19 в России, 02.2020 – 12.2021

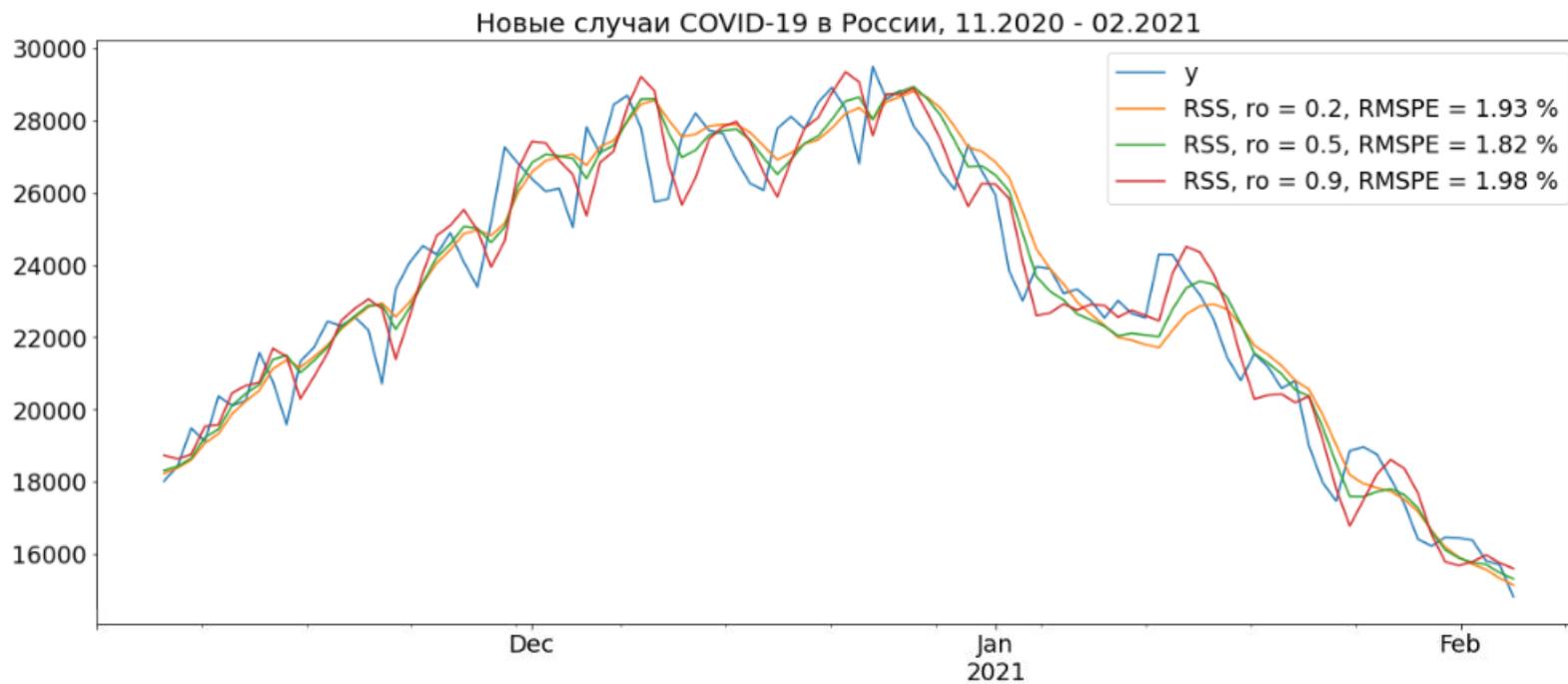


Рисунок А.2 – Новые случаи COVID-19 в России, 11.2020 – 02.2021

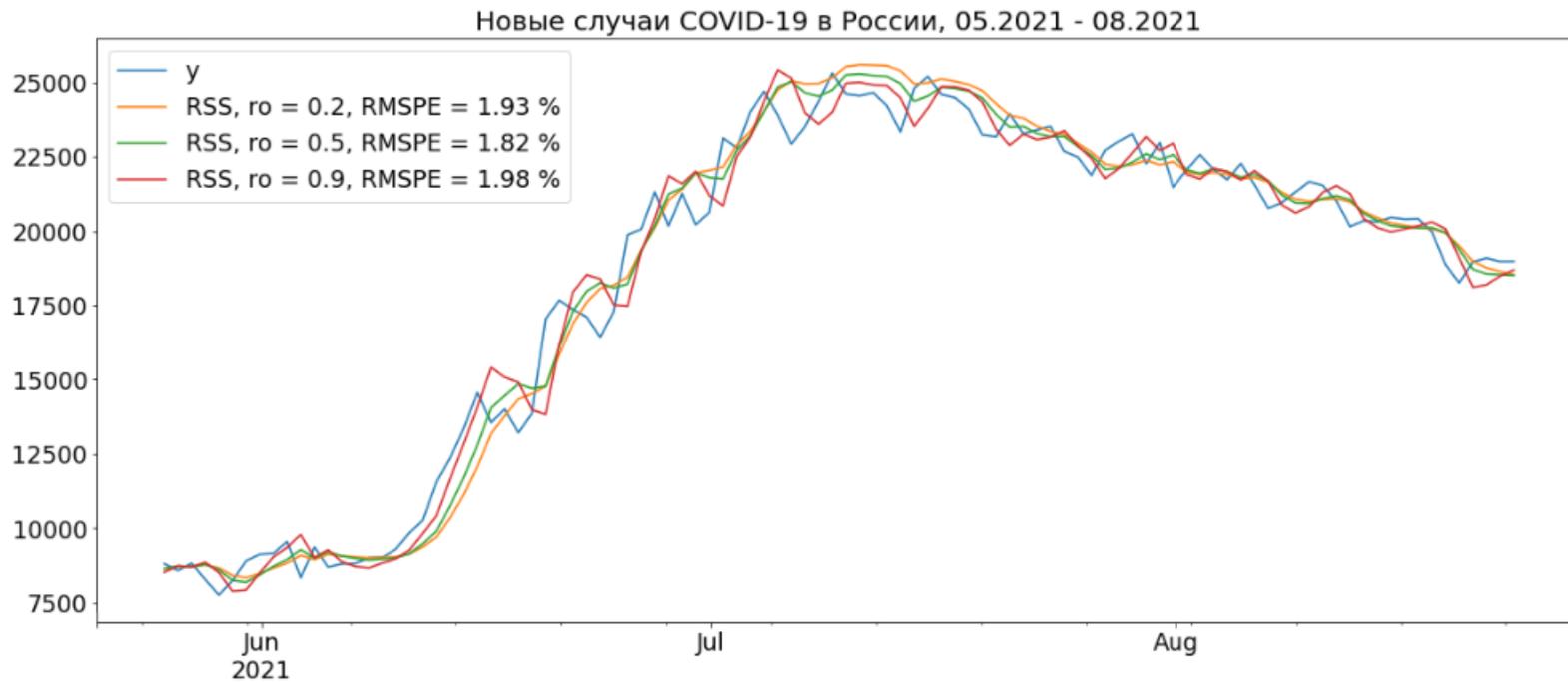


Рисунок А.3 – Новые случаи COVID-19 в России, 05.2021 – 08.2021

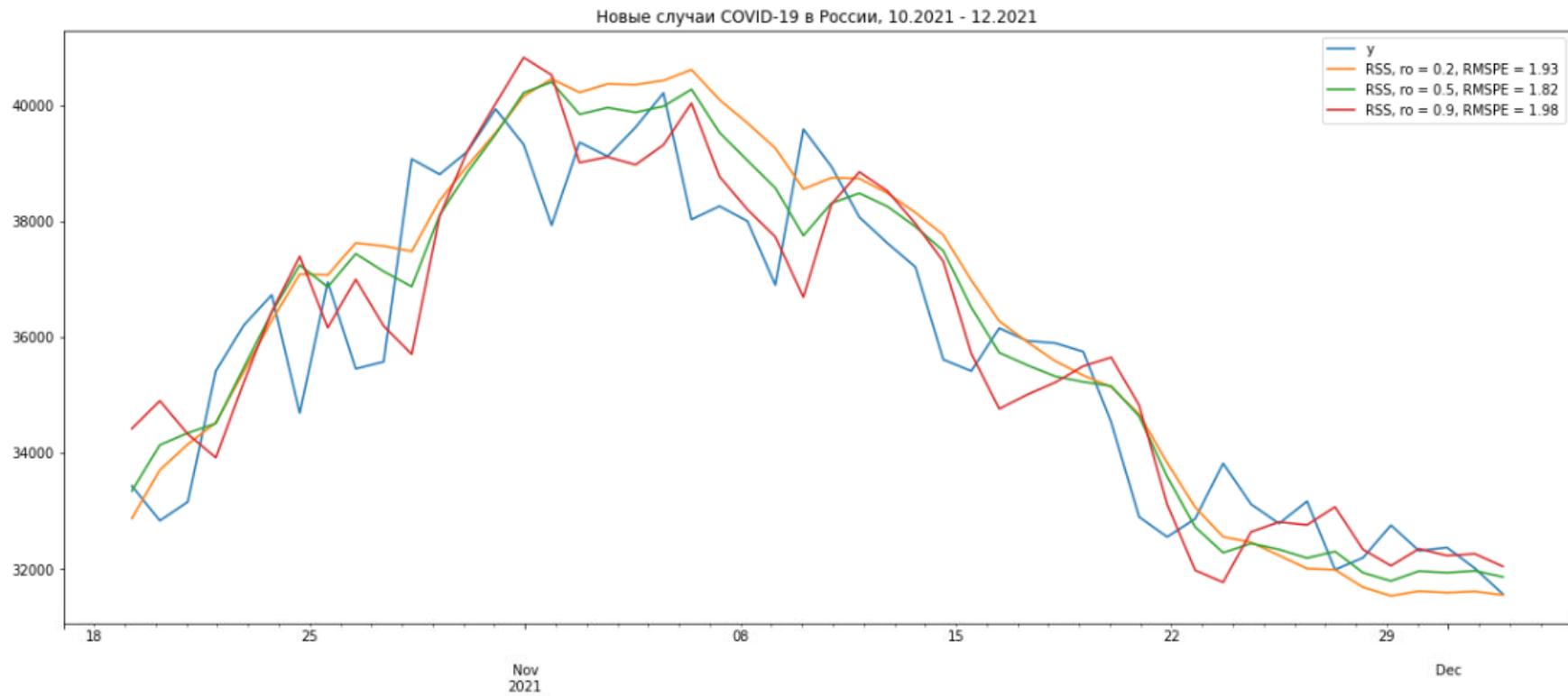


Рисунок А.4 – Новые случаи COVID-19 в России, 10.2021 – 12.2021

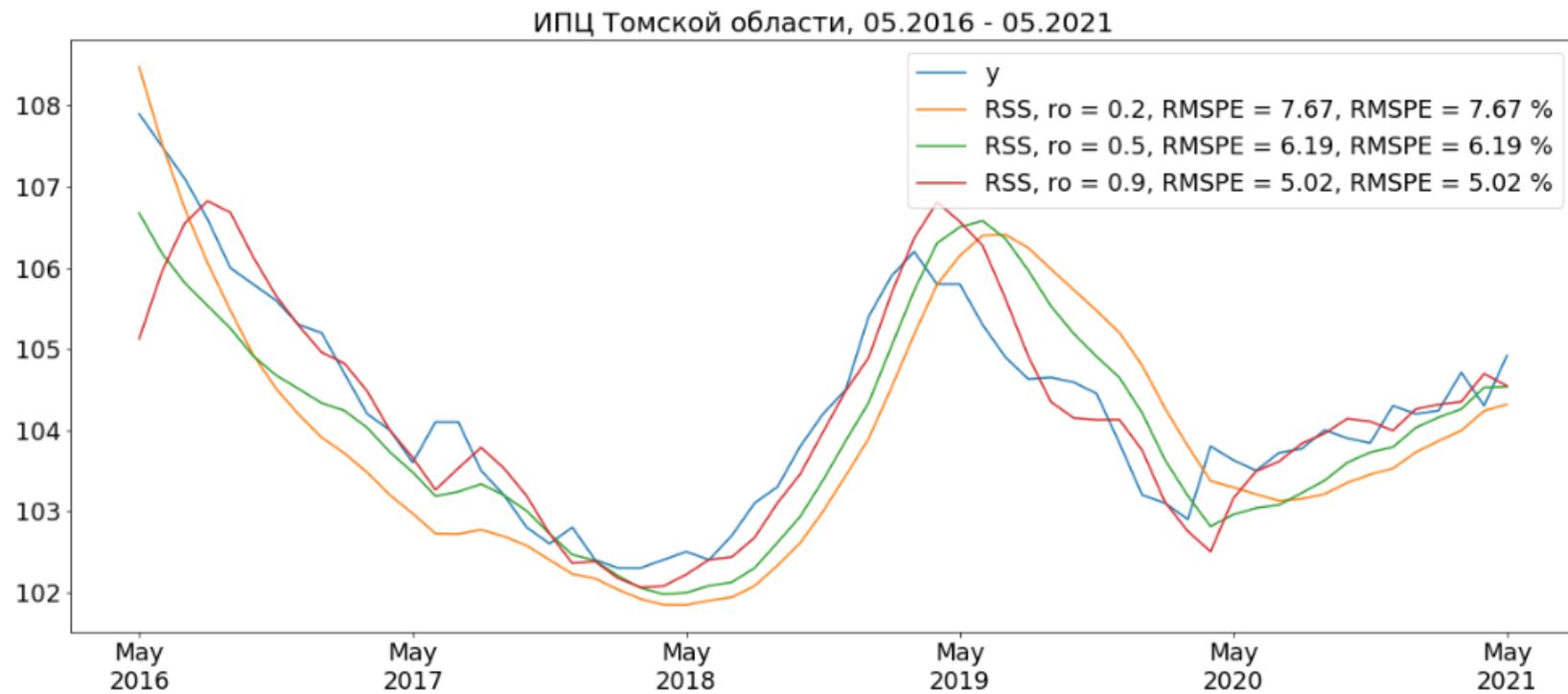


Рисунок А.5 – ИПЦ Томской области

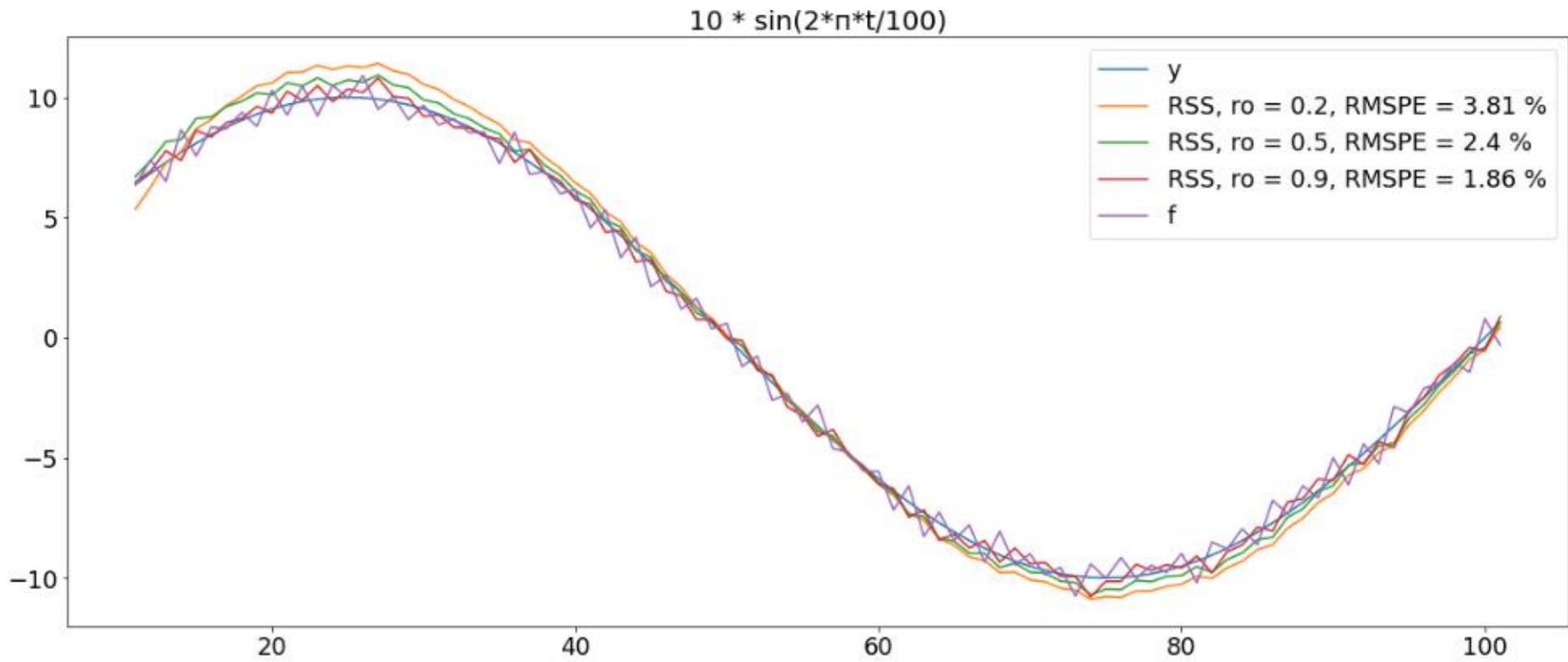


Рисунок А.6 – Синтетический ВР 1

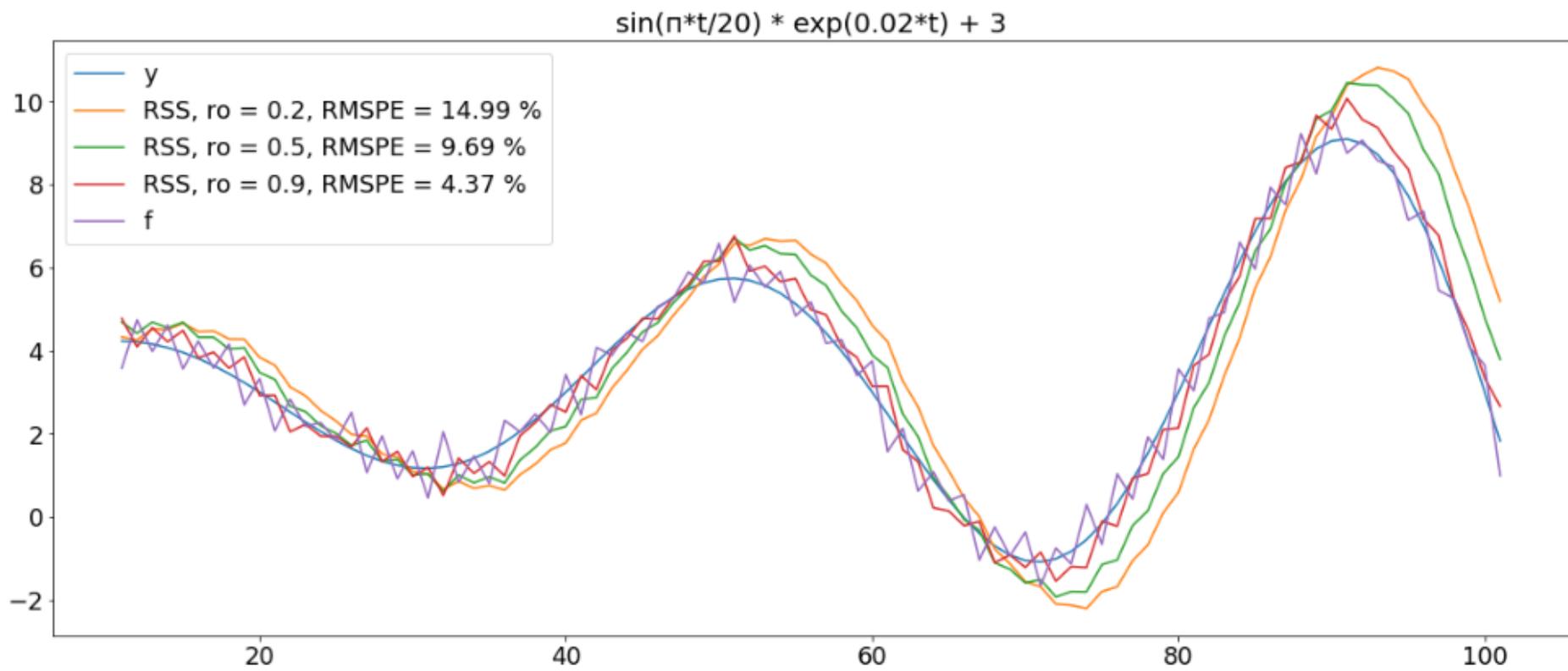


Рисунок А.7 – Синтетический ВР 2

**Приложение Б**  
**(обязательное)**  
**Графики сегментов и кумулятивных сумм**

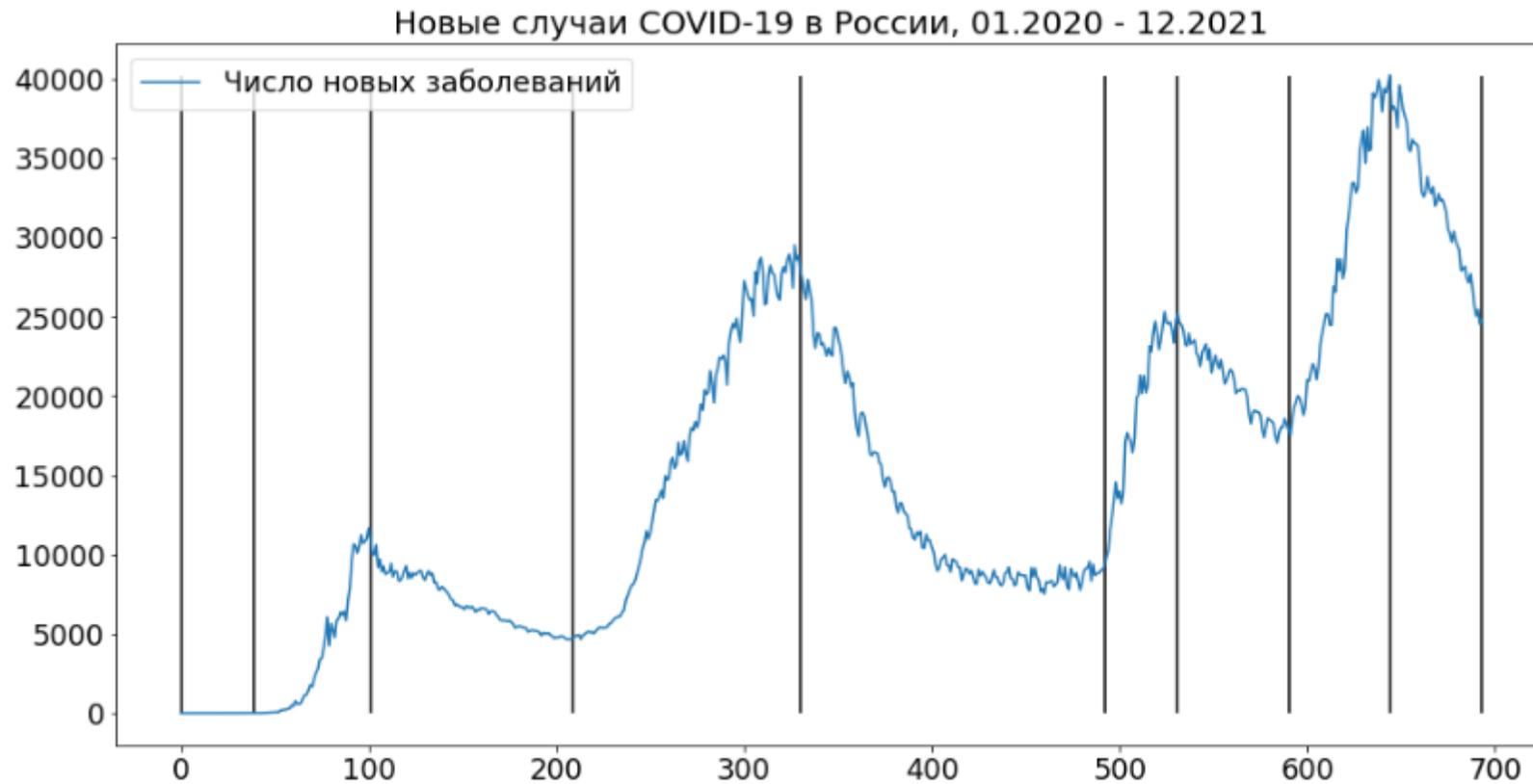


Рисунок Б.1 – Сегментированный ВР Новых случаев COVID-19 в России, 01.2020 – 12.2021



Рисунок Б.2 – График кумулятивных сумм ВР Новых случаев COVID-19 в России, 01.2020 – 12.2021

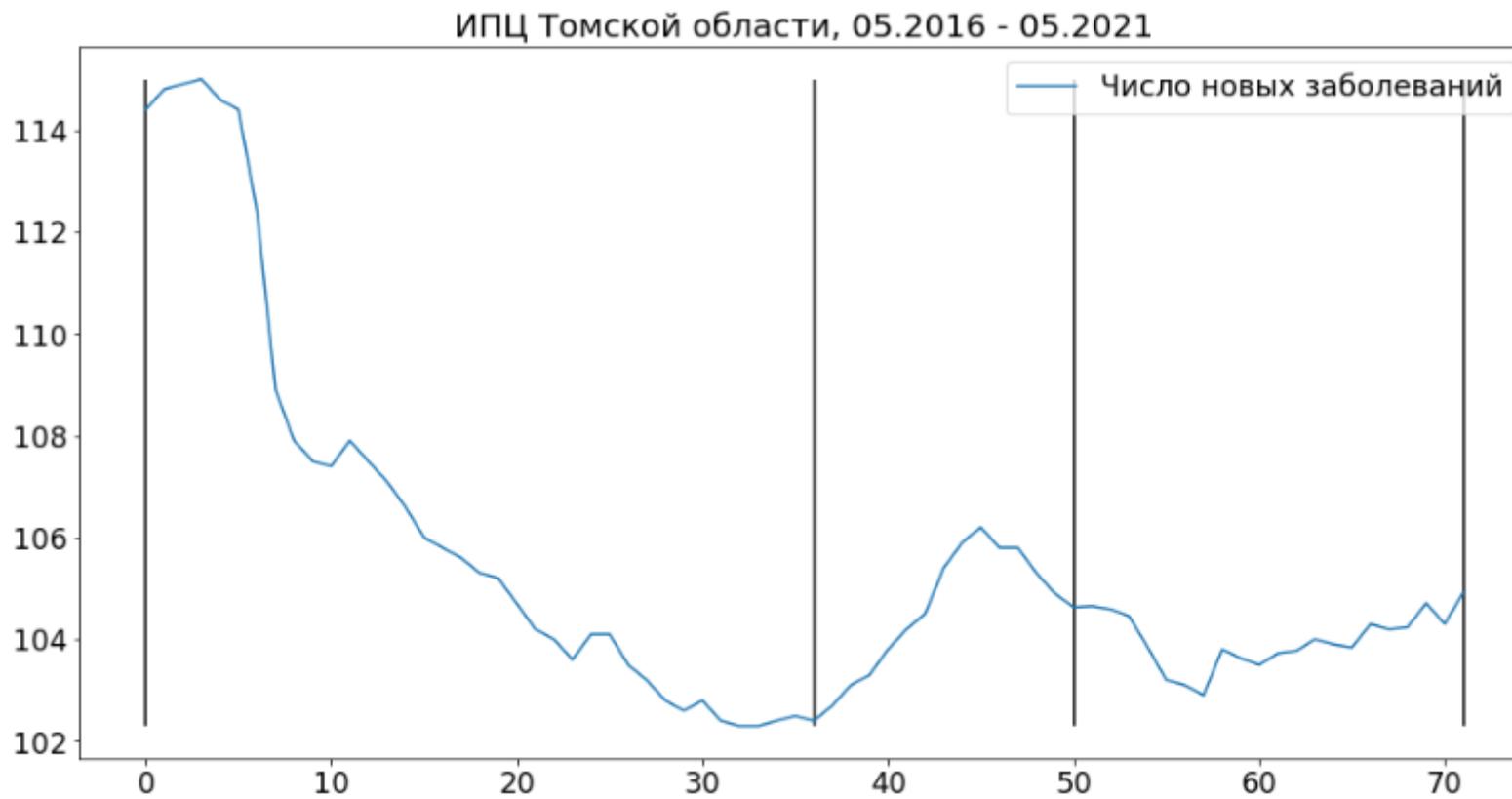


Рисунок Б.3 – Сегментированный ВР ИПЦ Томской области 05.2016 – 05.2021



Рисунок Б.4 – График кумулятивных сумм ВР ИПЦ Томской области 05.2016 – 05.2021

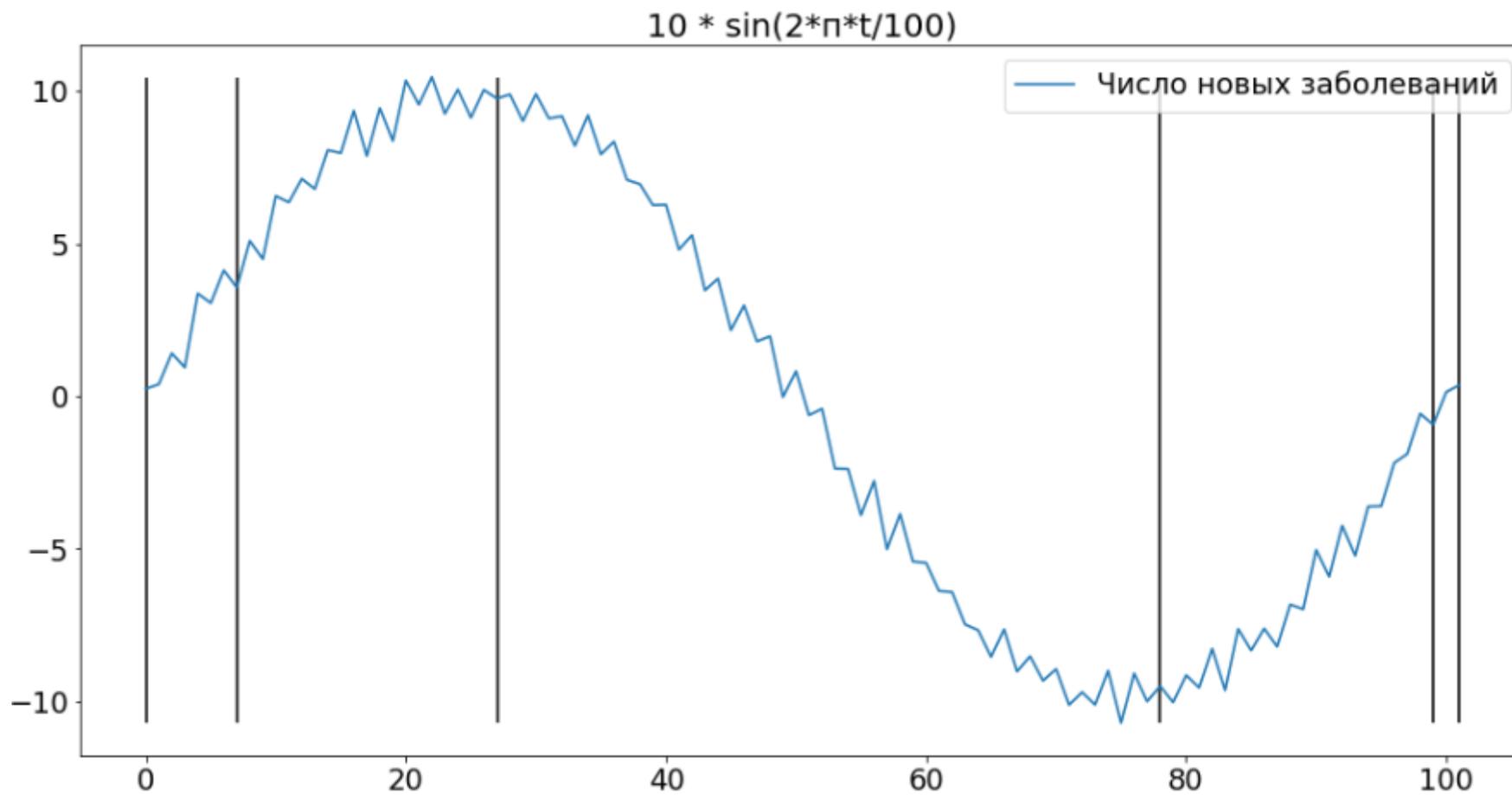


Рисунок Б.5 – Сегментированный синтетический ВР 1

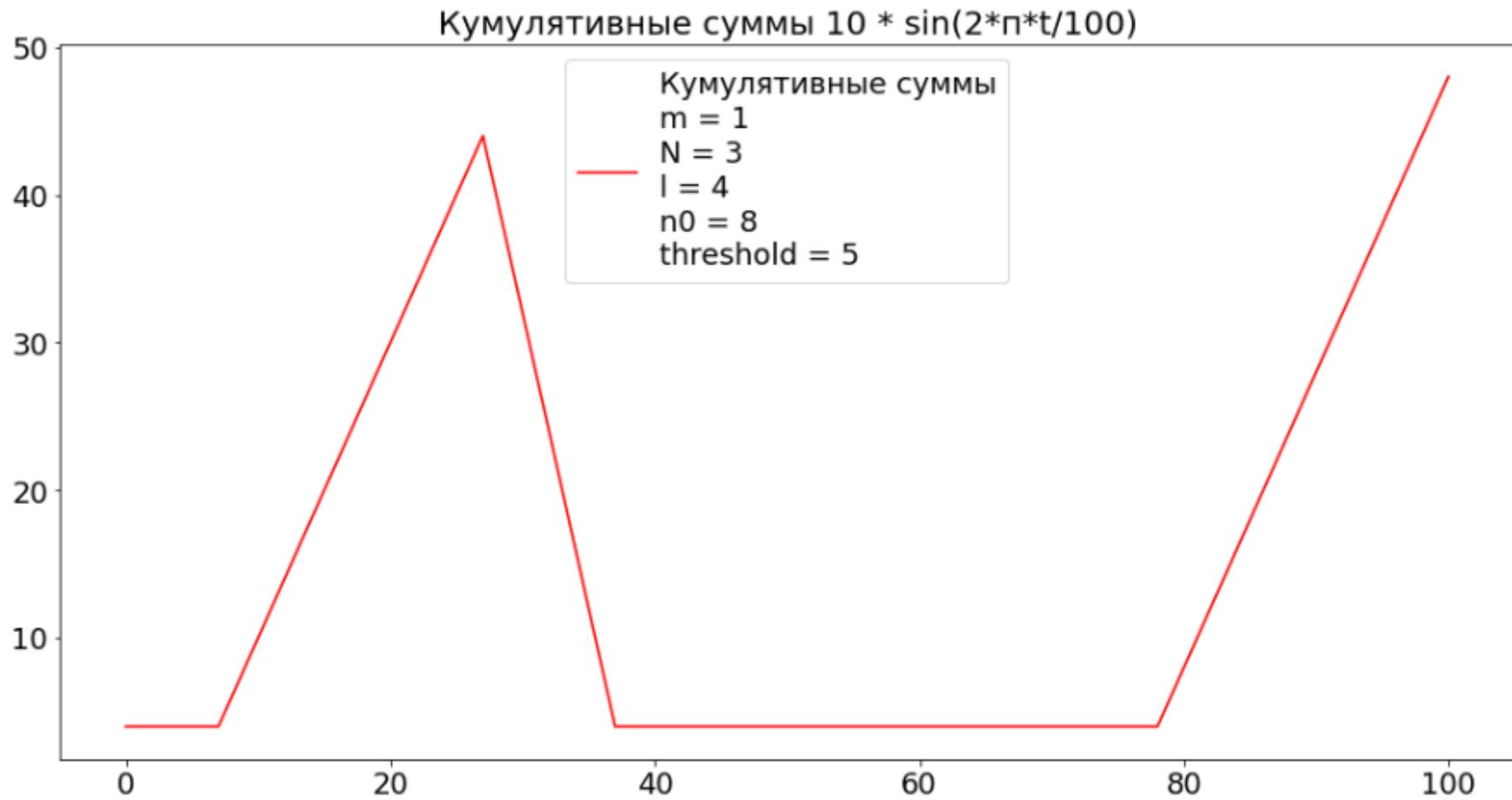


Рисунок Б.6– График кумулятивных сумм синтетического ВР 1

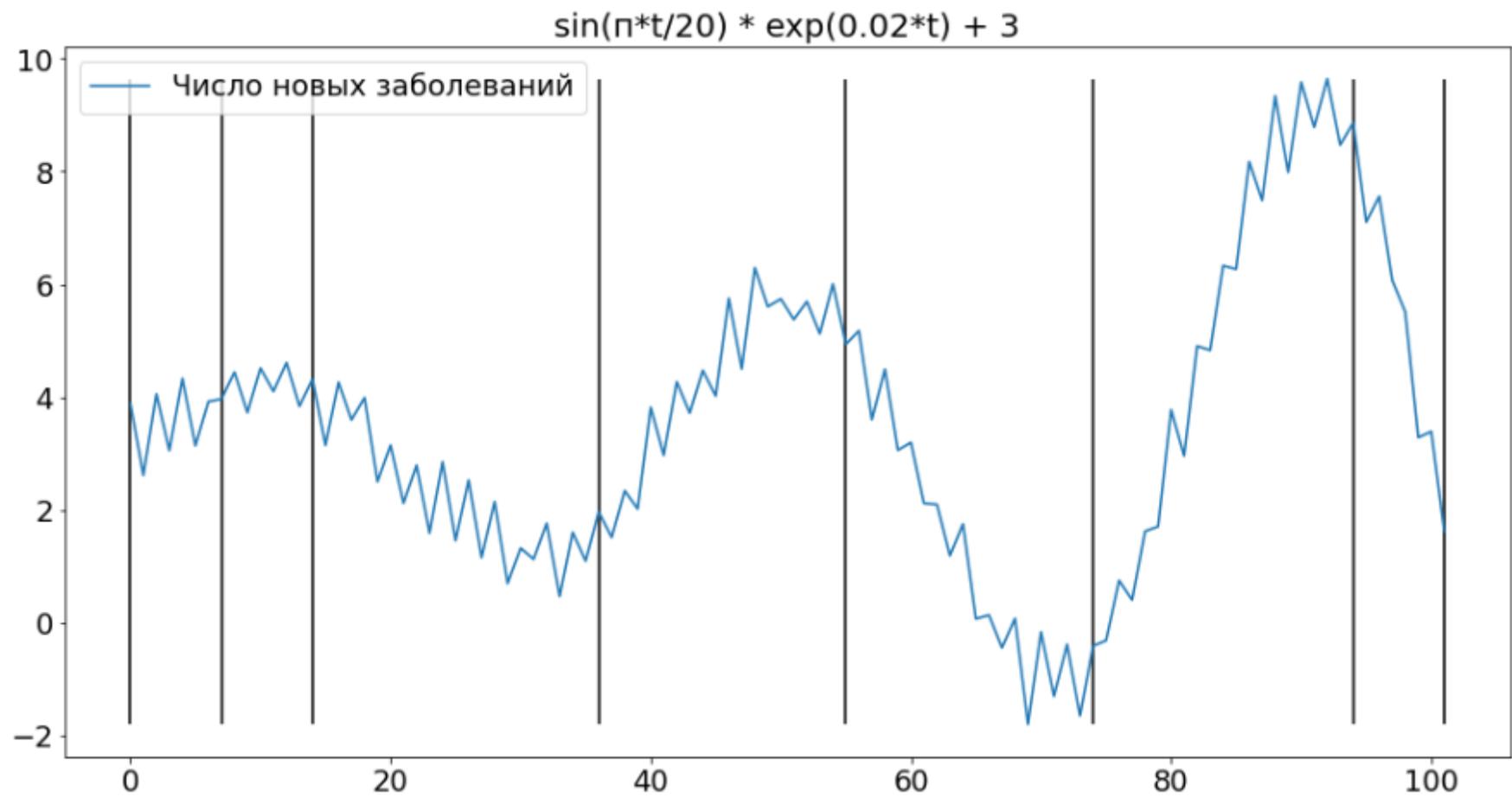


Рисунок Б.7 – Сегментированный синтетический ВР 2

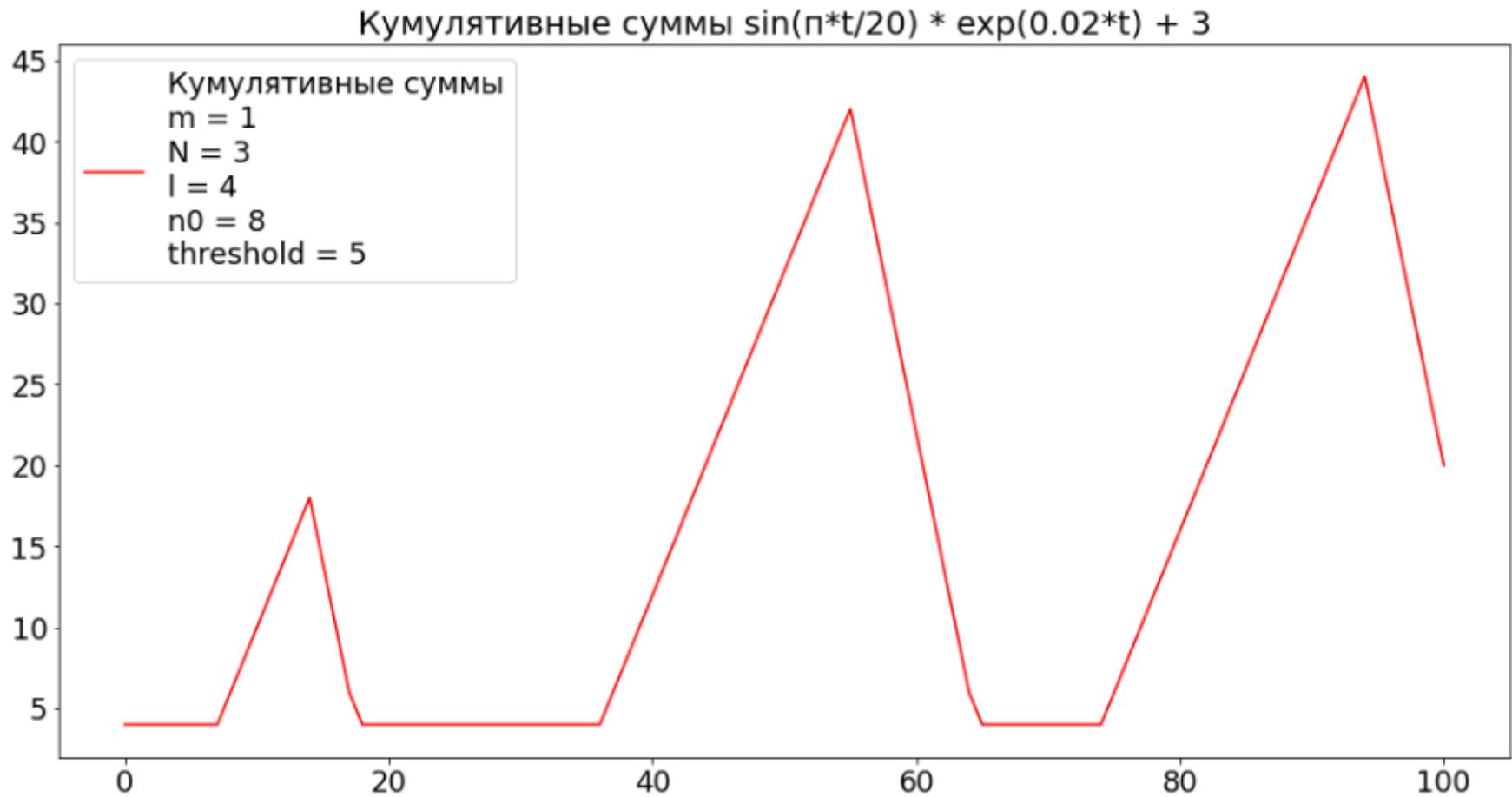


Рисунок Б.8– График кумулятивных сумм синтетического ВР 2