



Школа Инженерная школа информационных технологий и робототехники
Направление подготовки 09.04.01 Информатика и вычислительная техника
ООП/ОПОП Разработка интернет-приложений
Отделение школы (НОЦ) Отделение информационных технологий

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА МАГИСТРАНТА

Тема работы
Оценка параметров структурных моделей временных рядов на основе принципов байесовской статистики

УДК 004.942:519.22

Обучающийся

Группа	ФИО	Подпись	Дата
8ВМ11	Чжу Чжунвэнь		

Руководитель ВКР

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

КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН ШБИП	Былкова Т.В.	к.э.н., доцент		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ООД ШБИП	Федорчук Ю.М.	д.т.н., профессор		

ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП, должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Кочегурова Е.А.	к.т.н., доцент		

ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ООП
по направлению 09.04.01 Информатика и вычислительная техника

Код компетенции	Наименование компетенции
Универсальные компетенции	
УК(У)-1	Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий
УК(У)-2	Способен управлять проектом на всех этапах его жизненного цикла
УК(У)-3	Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели
УК(У)-4	Способен применять современные коммуникативные технологии, в том числе на иностранном (-ых) языке (-ах), для академического и профессионального взаимодействия
УК(У)-5	Способен анализировать и учитывать разнообразие культур в процессе межкультурного взаимодействия
УК(У)-6	Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки
Общепрофессиональные компетенции	
ОПК(У)-1	Способен самостоятельно приобретать, развивать и применять математические, естественно-научные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте
ОПК(У)-2	Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач
ОПК(У)-3	Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями
ОПК(У)-4	Способен применять на практике новые научные принципы и методы исследований
ОПК(У)-5	Способен разрабатывать и модернизировать программное и аппаратное обеспечение информационных и автоматизированных систем
ОПК(У)-6	Способен разрабатывать компоненты программно-аппаратных комплексов обработки информации и автоматизированного проектирования
ОПК(У)-7	Способен адаптировать зарубежные комплексы обработки информации и автоматизированного проектирования к нуждам отечественных предприятий
ОПК(У)-8	Способен осуществлять эффективное управление разработкой программных средств и проектов
Профессиональные компетенции	
ПК(У)-1	Способен разрабатывать и администрировать системы управления базами данных
ПК(У)-2	Способен проектировать сложные пользовательские интерфейсы
ПК(У)-3	Способен управлять процессами и проектами по созданию (модификации) информационных ресурсов
ПК(У)-4	Способен осуществлять руководство разработкой комплексных проектов на всех стадиях и этапах выполнения работ
ПК(У)-5	Способен проектировать и организовывать учебный процесс по образовательным программам с использованием современных образовательных технологий

Школа Инженерная школа информационных технологий и робототехники
 Направление подготовки 09.04.01 Информатика и вычислительная техника
 ООП/ОПОП Разработка интернет-приложений
 Отделение школы Отделение информационных технологий

УТВЕРЖДАЮ:
 Руководитель ООП/ОПОП
 _____ Кочегурова Е.А.
 « ____ » _____ 2023 г.

ЗАДАНИЕ на выполнение выпускной квалификационной работы

В форме:

Магистерской диссертации

Обучающийся:

Группа	ФИО
8ВМ11	Чжу Чжунвэнь

Тема работы:

Оценка параметров структурных моделей временных рядов на основе принципов байесовской статистики	
Утверждена приказом директора (дата, номер)	№ 96-61/с от 06.04.2023

Срок сдачи обучающимся выполненной работы:	01.06.2023
--	------------

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

<p>Исходные данные к работе <i>(наименование объекта исследования или проектирования; производительность или нагрузка; режим работы (непрерывный, периодический, циклический и т. д.); вид сырья или материал изделия; требования к продукту, изделию или процессу; особые требования к функционированию (эксплуатации) объекта или изделия в плане безопасности эксплуатации, влияния на окружающую среду, энергозатратам; экономический анализ и т. д.)</i></p>	<p>Объектом исследования являются временные ряды метеорологических наблюдений.</p>
<p>Перечень разделов пояснительной записки подлежащих исследованию, проектированию и разработке <i>(аналитический обзор литературных источников с целью выяснения достижений мировой науки техники в рассматриваемой области; постановка задачи исследования, проектирования, конструирования; содержание процедуры исследования, проектирования, конструирования; обсуждение результатов выполненной работы; наименование дополнительных разделов,</i></p>	<ol style="list-style-type: none"> 1. Обзор моделей и методов прогнозирования временных рядов. 2. Модели пространства состояний. 3. Общая характеристика и структурная схема анализа временных рядов. 4. Программные эксперименты по построению прогнозных моделей.

<i>подлежащих разработке; заключение по работе)</i>	5. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение. 6. Социальная ответственность.
Перечень графического материала <i>(с точным указанием обязательных чертежей)</i>	1. UML-диаграммы. 2. Скриншоты окон веб-приложения. 3. Таблица трудоемкости выполнения работ. 4. Диаграмма Ганта. 5. Презентация.

Консультанты по разделам выпускной квалификационной работы <i>(с указанием разделов)</i>	
Раздел	Консультант
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Доцент ОСГН ШБИП, к.э.н., Былкова Т.В.
Социальная ответственность	Профессор ООД ШБИП, д.т.н., Федорчук Ю.М.
Английский язык	Доцент ОИЯ ШБИП, к.ф.н., Степура С.Н.
Названия разделов, которые должны быть написаны на иностранном языке:	
Time series forecasting and overview of some methods for its solution	

Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику	01.03.2023г.
---	--------------

Задание выдал руководитель / консультант (при наличии):

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

Задание принял к исполнению обучающийся:

Группа	ФИО	Подпись	Дата
8ВМ11	Чжу Чжунвэнь		

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 09.04.01 Информатика и вычислительная техника
 Уровень образования Магистратура
 ООП/ОПОП Разработка интернет-приложений
 Отделение школы (НОЦ) Отделение информационных технологий
 Период выполнения Осенний / весенний семестры 2022 /2023 учебного года

Форма представления работы:

Магистерская диссертация

КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН
 выполнения выпускной квалификационной работы

Срок сдачи студентом выполненной работы:	01.06.2023
--	------------

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
22.02.2023	Модели и методы прогнозирования временных рядов	20
07.03.2023	Характеристики и структурная схема анализа временных рядов.	20
28.05.2022	Программные эксперименты по построению прогнозных моделей	30
27.05.2023	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	10
28.05.2023	Социальная ответственность	10
28.05.2023	Приложение на английском языке	10

СОСТАВИЛ:

Руководитель ВКР

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

СОГЛАСОВАНО:

Руководитель ООП

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

РЕФЕРАТ

Выпускная квалификационная работа содержит 115 страницы, 28 рисунков, 22 таблицы, 28 источников, 1 приложение.

Ключевые слова: температурный временной ряд, анализ данных, модель пространства состояний, bsts-модель, язык R, прогнозирование.

Цель работы: разработка программного обеспечения для анализа и прогнозирования метеорологических рядов наблюдений.

Для достижения этой цели были поставлены и решены следующие задачи.

1. Исследовать модель прогнозирования временных рядов bsts.
2. Разработать метод прогнозирования на основании модели и выполнить программную реализацию алгоритмов.
3. Оценить эффективность модели прогнозирования.

Объектом исследования являются метеорологические данные.

Предметом исследования является модель прогнозирования bsts.

Программные эксперименты реализованы с использованием интерактивной платформы RStudio на языке статистического моделирования R.

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

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

Оглавление

Введение	9
1. Задача прогнозирования временных рядов и обзор некоторых методов ее решения	10
1.1. Регрессионные модели	10
1.2. Авторегрессионные модели.....	12
1.3. Модели экспоненциального сглаживания.....	14
1.4. Нейросетевые модели	16
1.5. Модели на базе цепей Маркова.....	17
1.6. Обзор современных тенденций в сфере прогнозирования	19
1.7. Байесовский подход к прогнозированию временных рядов.....	21
1.8. Выводы по Главе 1	23
2. Модели пространства состояний	25
2.1. Структурные модели временных рядов	25
2.2. Фильтр и сглаживание Калмана.....	27
2.3. Выводы по Главе 2.....	28
3. Общая характеристика и структурная схема анализа временных рядов ..	29
3.1. Проектирование структуры системы прогнозирования	29
3.2. Схема анализа данных и прогнозирования	31
3.2.1. Предварительная обработка данных.....	32
3.2.2. Предварительный анализ временных рядов	32
3.2.3. Тест на стационарность	33
3.2.4. Декомпозиция временных рядов	36
3.3. Подгонка и оценка модели	38
3.4. Выводы по Главе 3.....	39
4. Программные эксперименты по построению прогнозных моделей	40
4.1. Предварительная обработка данных с использованием пакета bssm	40
4.2. Программная реализация модели bsts	45
4.3. Разработка веб-интерфейса для программного обеспечения.....	50

4.4. Выводы по Главе 4.....	56
5. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	59
5.1. Оценка коммерческого и инновационного потенциала НТИ	59
5.1.1. Потенциальные потребители результатов исследования.....	59
5.1.2. Анализ конкурентных технических решений.....	59
5.1.3. FAST-анализ	61
5.1.4. SWOT-анализ.....	64
5.1.5. Оценка готовности проекта к коммерциализации.....	65
5.1.6. Методы коммерциализации результатов научно-технического исследования.....	67
5.2. Инициация проекта.....	67
5.3. Планирование управления научно-техническим проектом.....	69
5.3.1. План проекта.....	69
5.3.2. Бюджет научного исследования.....	71
5.3.3. Определение ресурсной, финансовой эффективности исследования.....	73
6. Социальная ответственность.....	79
6.1. Производственная безопасность.....	79
6.1.1. Вредные производственные факторы	79
6.1.2. Опасные производственные факторы	87
6.2. Экологическая безопасность.....	91
6.3. Безопасность в чрезвычайных ситуациях	92
Заключение.....	95
Список студенческих публикаций.....	96
Список литературы	98
Приложение А.....	102

Введение

Под прогнозированием понимается процесс оценки будущих значений или состояний системы на основе доступных данных и знаний о прошлых значениях и тенденциях. Прогнозирование является важным инструментом в различных областях, таких как экономика, финансы, метеорология, климатология, наука о данных и др. Оно позволяет принимать более обоснованные решения на основе вероятностных оценок будущего развития системы, что может быть полезно для планирования, управления ресурсами и принятия стратегических решений.

В настоящее время временные ряды являются одними из наиболее распространенных объектов анализа в различных научных и прикладных областях. В связи с этим возникает необходимость разработки эффективных методов оценки параметров таких рядов, которые позволят получать более точные прогнозы будущих значений наблюдаемых параметров и выявлять закономерности в их изменениях.

1. Задача прогнозирования временных рядов и обзор некоторых методов ее решения

1.1. Регрессионные модели

Регрессионная модель – это статистический подход, используемый для анализа связи между зависимой переменной и одной или несколькими независимыми переменными. Он направлен на поиск наиболее подходящей математической функции, описывающей взаимосвязь между этими переменными.

Общая формула представления регрессионной модели может быть записана как:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon$$

Где:

y – зависимая переменная.

β_0 – константа (значение функции регрессии при нулевом значении всех факторов).

$\beta_1, \beta_2, \dots, \beta_n$ – коэффициенты или параметры, связанные с независимыми переменными x_1, x_2, \dots, x_n .

ε – состав ошибки, который фиксирует необъяснимую изменчивость зависимой переменной.

Модели регрессии анализируют функциональную связь между зависимой переменной и независимыми переменными. Они дают возможность количественно прогнозировать значение зависимой переменной на основе значений независимых переменных, оценивать коэффициенты или параметры, определяющие взаимосвязь между переменными.

Модели регрессии включают член ошибки для учета необъяснимой изменчивости в зависимой переменной. Они основаны на определенных допущениях, таких как линейность, независимость от ошибок, гомоскедастичность и нормальность остатков.

Ниже приведены некоторые распространенные методы регрессионного анализа:

- Простая линейная регрессия: включает одну независимую переменную и линейную связь с зависимой переменной.
- Множественная линейная регрессия: включает несколько независимых переменных для моделирования связи с зависимой переменной.
- Полиномиальная регрессия: позволяет использовать нелинейные отношения, включая полиномиальные члены независимых переменных.
- Логистическая регрессия: используется для бинарных или категориальных зависимых переменных, оценивая вероятность принадлежности к определенной категории.

Модели регрессии помогают определить и количественно оценить отношения между переменными. Они позволяют прогнозировать или оценивать зависимую переменную на основе значений независимых переменных. Модели регрессии помогают определить относительную важность независимых переменных для объяснения изменений зависимой переменной. Они также предоставляют интерпретируемые коэффициенты, позволяющие понять влияние независимых переменных на зависимую переменную, облегчают проверку гипотез и оценку доверительных интервалов для коэффициентов.

Регрессионные модели имеют определенные недостатки. Они предполагают линейность, которая может выполняться не во всех случаях. Может произойти переобучение или недообучение, влияющее на обобщение модели. Выбросы или важные точки данных могут повлиять на расчетные коэффициенты. Мультиколлинеарность (высокая корреляция между независимыми переменными) и нарушение допущений могут привести к необъективным или неэффективным оценкам.

В различных областях регрессионные модели находят применение. Они используются в экономике для анализа факторов, влияющих на экономические показатели. В маркетинге они помогают прогнозировать продажи и понимать поведение потребителей. Финансы используют регрессионные модели для анализа рисков и финансового прогнозирования. Здравоохранение использует их для прогнозирования результатов лечения пациентов и анализа эффектов лечения.

Социальные науки используют регрессионные модели для изучения влияния социальных, демографических и психологических факторов на различные результаты.

1.2. Авторегрессионные модели

Модель авторегрессии, называемая моделью AR, представляет собой статистический метод обработки временных рядов с использованием той же переменной, что и предыдущие периоды x , то есть от x_1 до x_{t-1} , для прогнозирования производительности x_t в текущем периоде. Предположим, что они имеют линейную зависимость. Поскольку это разработано на основе линейной регрессии в регрессионном анализе, но вместо использования x для прогнозирования y , x используется для прогнозирования самого себя; поэтому это называется авторегрессией. Авторегрессионные модели широко используются в прогнозировании экономики, информатики и природных явлений.

Модели AR используют прошлые значения ряда для прогнозирования его будущих значений. Эта модель предполагает, что текущее значение ряда зависит от его предыдущих значений. Наиболее популярной моделью AR является модель первого порядка (AR(1)), которая предполагает, что текущее значение ряда зависит только от его предыдущего значения (т.е. $p=1$).

Уравнение авторегрессионной модели, обозначенное AR(p), приведено ниже:

$$X_t = C + \Phi_1 X_{t-1} + E_t$$

Где

X_{t-1} – значение X за предыдущий год/месяц/неделю.

ϕ_1 – коэффициент, на который мы умножаем X_{m-1} . Значение ϕ_1 всегда будет равно 1 или -1.

E_t – разница между значением периода t и правильным значением.

p – порядок.

Таким образом, AR(1) является авторегрессионной моделью первого порядка. Второй и третий порядок будут соответственно AR(2) и AR(3).

Примером использования модели AR является анализ финансовых рынков. Например, используя модель AR, можно прогнозировать будущие цены на акции на основе их прошлых значений. Если взять за основу прошлые цены и применить модель AR, можно определить вероятные цены акций в будущем.

МА-модели используют прошлые значения ошибок – разницу между фактическими значениями ряда и его прогнозируемыми значениями для прогнозирования будущих значений. Наиболее популярной моделью МА является модель первого порядка (МА(1)), которая предполагает, что текущая ошибка зависит только от ее предыдущего значения.

Примером использования модели МА может быть прогнозирование количества пользователей веб-сайта. Например, если предыдущие прогнозы ошибочны на определенный процент, можно использовать эту информацию в модели прогнозирования, чтобы повысить точность прогноза количества пользователей.

ARIMA – это модель, которая сочетает в себе модели авторегрессии и скользящего среднего. ARIMA позволяет моделировать нестационарные данные, что не требуется для моделей AR и МА. ARIMA включает три параметра: параметр авторегрессии (p), параметр скользящего среднего (q) и параметр интегрирования (d).

Модель ARIMA(p,d,q) для нестационарного временного ряда X_t имеет вид:

$$\Delta^d X_t = c + \sum_{i=1}^p a_i \Delta^d X_{t-i} + \sum_{j=1}^q b_j \varepsilon_{t-j} + \varepsilon_t$$

Где

ε_t – стационарный временной ряд;

c, a_i, b_i – параметры модели;

Δ^d – оператор разности временных рядов порядка d.

Примером использования ARIMA может быть прогноз ежемесячного дохода от продукта на основе ежемесячных данных за прошлый год. Если прошлые данные являются трендовыми, сезонными или циклическими, можно использовать ARIMA, чтобы учесть эти факторы в прогнозе доходов.

Преимущество авторегрессионного метода заключается в том, что он не требует большого количества данных и может использовать свои собственные ряды переменных для прогнозирования. Но этот метод имеет определенные ограничения:

Должна быть автокорреляция, ключом является коэффициент автокорреляции. Если коэффициент автокорреляции (R) меньше 0,5, его не следует использовать, иначе результат прогноза будет крайне неточным.

Авторегрессия может применяться только для прогнозирования экономических явлений, связанных с ее собственным предыдущим периодом, то есть экономических явлений, на которые сильно влияют ее собственные исторические факторы, такие как объем добычи, добыча различных природных ресурсов и т.д. Для экономических явлений, на которые большое влияние оказывают социальные факторы, вместо авторегрессии следует использовать векторные авторегрессионные модели, которые могут включать другие переменные.

1.3. Модели экспоненциального сглаживания

Экспоненциальное сглаживание – это метод прогнозирования, используемый при анализе временных рядов для прогнозирования будущих значений на основе исторических данных. Это особенно полезно для данных без сезонности или с минимальной сезонностью и относительно стабильной тенденцией.

Основная идея экспоненциального сглаживания состоит в том, чтобы присвоить веса прошлым наблюдениям таким образом, чтобы придать большее значение последним точкам данных и меньшее значение более старым точкам данных. Этот процесс взвешивания основан на экспоненциально убывающей функции. Самым последним наблюдениям присваивается наибольший вес, в то время как веса уменьшаются экспоненциально по мере того, как осуществляется удаление назад во времени.

Существуют различные варианты методов экспоненциального сглаживания, включая простое экспоненциальное сглаживание, двойное экспоненциальное

сглаживание и тройное экспоненциальное сглаживание (также известное как метод Хольта-Винтерса). Каждый вариант включает в себя разные уровни сложности для фиксации различных закономерностей в данных.

Формула простого экспоненциального сглаживания выглядит следующим образом:

$$F(t+1) = \alpha * Y(t) + (1 - \alpha) * F(t)$$

Где:

$F(t+1)$ – прогноз на следующий период,

$Y(t)$ – фактическое значение в текущем периоде,

$F(t)$ – прогноз на текущий период,

α – параметр сглаживания ($0 \leq \alpha \leq 1$), который определяет вес, придаваемый самому последнему наблюдению.

Экспоненциальное сглаживание относительно легко адаптировать к различным типам данных временных рядов, выбрав соответствующий вариант. Методы экспоненциального сглаживания эффективны в вычислительном отношении, что делает их подходящими для больших наборов данных. По мере поступления новых данных, модели экспоненциального сглаживания можно легко обновлять, чтобы включить в них самую свежую информацию.

Методы экспоненциального сглаживания наиболее эффективны для данных без сезонности или с минимальной сезонностью и устойчивым трендом. Они могут плохо работать со сложными или очень изменчивыми данными. Эта модель в первую очередь ориентирована на прогнозирование будущих значений и не дает понимания основных факторов, влияющих на данные. Начальное значение или начальная точка прогноза могут оказать существенное влияние на результаты, поэтому крайне важно выбрать подходящее начальное значение.

Экспоненциальное сглаживание может применяться в различных областях, в том числе:

- Прогнозирование спроса: обычно используется в управлении запасами и планировании производства для прогнозирования потребительского спроса.

- Финансовое прогнозирование: методы экспоненциального сглаживания используются для прогнозирования цен на акции, рыночных тенденций и финансовых показателей.
- Управление операциями: помогает прогнозировать продажи, спрос на услуги и распределение ресурсов.
- Анализ временных рядов: экспоненциальное сглаживание является фундаментальным методом изучения и анализа данных, зависящих от времени.
- Управление цепочкой поставок: помогает прогнозировать количество заказов, оптимизировать уровни запасов и повышать эффективность цепочки поставок.

Важно отметить, что выбор конкретного метода экспоненциального сглаживания и соответствующих значений параметров зависит от характеристик данных и конкретных целей прогнозирования.

1.4. Нейросетевые модели

Искусственная нейронная сеть абстрагирует нейронную сеть человеческого мозга с точки зрения обработки информации, устанавливает простую модель и формирует различные сети в соответствии с различными методами соединения.

Нейрон – это единица обработки информации в нейронной сети. На Рис. 1-1 показана модель нейрона, лежащего в основе искусственных сетей. Модель содержит 3 основных элемента [1]: набор синапсов, сумматор, функцию активации.

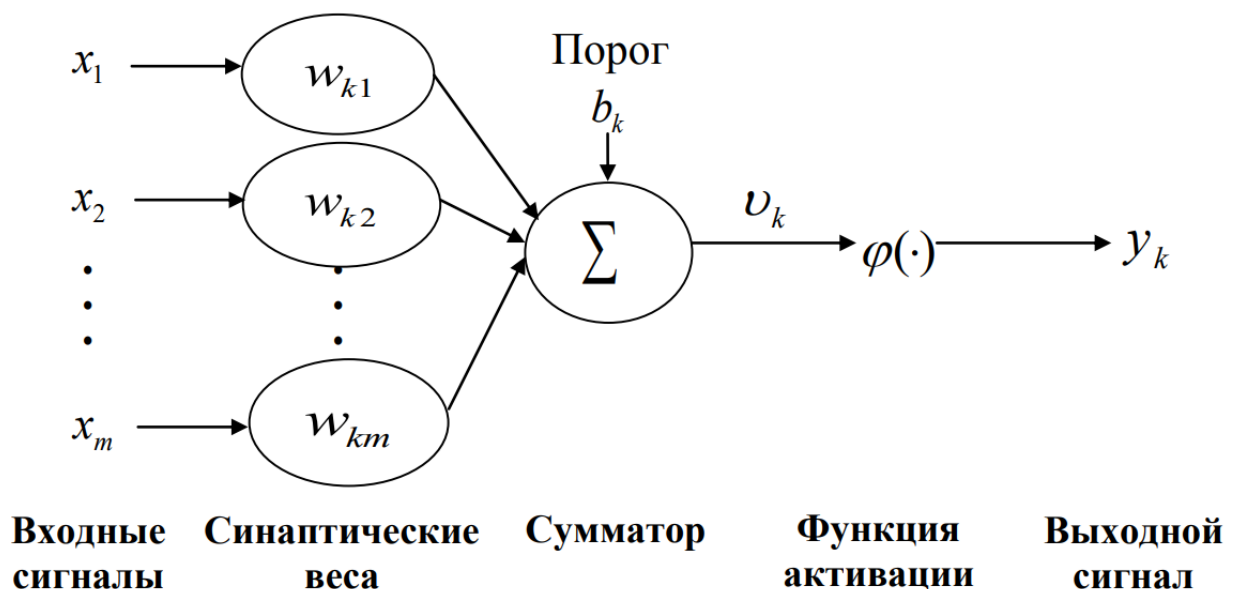


Рис. 1-1. Нелинейная модель нейрона

Каждый узел представляет определенную выходную функцию, называемую функцией активации. Каждое соединение между двумя узлами представляет взвешенное значение сигнала, проходящего через соединение, называемое весом. Выход сети зависит от способа подключения сети, значения веса и функции активации.

В математическом представлении функционирование нейрона k осуществляется следующим образом:

$$u_k = \sum_{j=1}^m W_{kj} X_j$$
$$y_k = \Phi(u_k + b_k)$$

Где

X_1, X_2, \dots, X_m – входные сигналы;

W_{k1}, W_{k2}, W_{km} – синаптические веса нейрона k ;

u_k – линейная комбинация входных воздействий;

b_k – порог;

$\Phi(\cdot)$ – активационная функция;

y_k – выходные сигналы.

1.5. Модели на базе цепей Маркова

Модели, основанные на цепях Маркова, обычно используются для прогнозирования данных временных рядов. Модели цепи Маркова предполагают, что будущее состояние системы зависит только от ее текущего состояния, а не от ее прошлых состояний, что делает их подходящими для моделирования данных временных рядов с определенными закономерностями или зависимостями.

Математическое представление модели цепи Маркова для прогнозирования временных рядов включает вероятности перехода. Матрица вероятности перехода $P(i, j)$ представляет собой вероятность перехода из состояния s_i в состояние s_j . Если данные временного ряда следуют цепи Маркова первого порядка, вероятности перехода могут быть представлены как:

$$P(i, j) = P(X_{t+1} = s_j \mid X_t = s_i)$$

Вероятности перехода должны удовлетворять следующим свойствам:

$$P(i, j) \geq 0 \text{ для всех } i, j.$$

$$\sum P(i, j) = 1 \text{ для каждого } i.$$

Если вероятности перехода были оценены, их можно использовать для прогнозирования будущих значений во временном ряду на основе текущего состояния. Конкретный метод прогнозирования будет зависеть от желаемого подхода, такого как создание вероятностных прогнозов с использованием вероятностей перехода или определение наиболее вероятного следующего состояния.

Пример цепи Маркова для процесса, имеющего три состояния, представлен на Рис. 1-2.

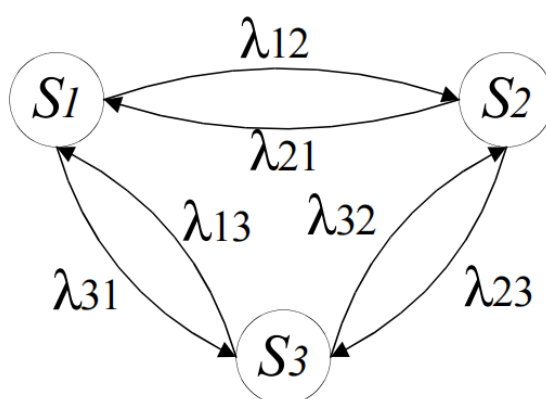


Рис. 1-2. Цепь Маркова с тремя состояниями

Здесь S_1, S_2, S_3 – состояния процесса $Z(t)$; λ_{12} – вероятность перехода из состояния S_1 в состояние S_2 , λ_{23} – вероятность перехода из состояния S_2 в состояние S_3 и т.д. При построении цепи Маркова определяется множество состояний и вероятностей переходов. Если существует текущее состояние процесса S_i , то в качестве будущего состояния процесса выбирается такое состояние S_j , вероятность перехода в которое (величина λ_{ij}) максимальна.

Модели цепей Маркова относительно просты и легки в реализации, что делает их доступными для широкого круга пользователей. Они могут фиксировать зависимости и закономерности в данных временных рядов, включая тенденции, сезонность и другие временные отношения.

Модели цепей Маркова обеспечивают интерпретируемые вероятности перехода, позволяя получить представление о динамике моделируемой системы.

Прогнозирование с помощью моделей цепей Маркова может быть эффективным с точки зрения вычислений, особенно для наборов данных среднего размера.

Модели цепи Маркова также имеют некоторые недостатки: модели предполагают, что будущее состояние зависит только от текущего состояния, игнорируя любую информацию из предыдущих состояний. Это допущение может не выполняться для всех данных временных рядов, что ограничивает способность модели фиксировать долгосрочные зависимости.

Модели обычно используются для краткосрочного прогнозирования, поскольку их точность имеет тенденцию к снижению по мере увеличения горизонта прогнозирования.

Начальное состояние или начальная точка могут оказать существенное влияние на прогнозируемые последовательности, подчеркивая важность выбора подходящего начального состояния.

1.6. Обзор современных тенденций в сфере прогнозирования

Ещё раз подчеркнём, что в настоящее время временные ряды являются одним из наиболее распространенных объектов анализа [1-11]. Ниже приведён краткий анализ публикаций на эту тему.

В статье [1] было проведено сравнение одношагового прогнозирования с использованием смоделированного нелинейного авторегрессионного временного ряда скользящего среднего (NARMA) между двумя группами нейронных сетей. Результаты показывают, что введение обратной связи по ошибкам в нейронные сети помогает в прогнозировании временных рядов NARMA и нейронные сети показывают лучшие результаты по сравнению с моделью ARMA.

В статье [2] были объединены метаобучение и трансфертное обучение для решения проблемы отсутствия данных при прогнозировании временных рядов. Показано, что метатрансферное обучение обеспечивает очень хорошие прогнозы.

В работе [3] предлагается метод краткосрочного прогнозирования туристического спроса, основанный на кластеризации многомерных временных рядов и LSSVM.

В статье [4] представлена двухэтапная структура для прогнозирования одномерных сезонных временных рядов. На первом этапе в явном виде изучается структура долгосрочного временного ряда во временном окне за горизонтом прогноза. Второй этап может повысить точность прогноза на горизонте прогноза. На обоих этапах авторегрессионная модель интегрируется с нейронными сетями для захвата как линейных, так и нелинейных характеристик во временных рядах.

В публикации [5] демонстрируется, как неправильное использование стратегии точечного скользящего прогнозирования может привести к нереалистичным результатам и как многоэтапная стратегия прогнозирования может помочь решить эту проблему.

В работе [6] внимание уделяется существующим моделям прогнозирования трафика данных в транспортных системах, основанным на глубоком обучении, а также изучению и анализу областей применения структурных характеристик различных моделей прогнозирования.

В статье [7] предлагается новый эвристический метод, основанный на хаотических характеристиках экономических временных рядов, который прогнозирует эти временные ряды с помощью искусственных нейронных сетей.

В статье [8] проводится сравнение производительности различных одномерных моделей для задачи прогнозирования. Показано, что метод FFNN хорошо адаптирован для прогнозирования метеорологических одномерных временных рядов с сезонностью и отсутствием тренда с учетом показателей точности, а DTWBI более подходит для учета формы и динамики прогнозных значений.

В публикации [9] представлена модель с глубоким обучением для прогнозирования несбалансированных данных временных рядов.

В работе [10] были объединены модель регрессионного типа и модель нейронной сети для формирования лучшей модели прогнозирования осадков.

В статье [11] предложена улучшенная сетевая модель, основанная на DeepAR, для получения вероятностных прогнозов.

1.7. Байесовский подход к прогнозированию временных рядов

Широко известный современный подход к анализу временных рядов основывается на теореме Байеса. Формула Байеса все более широко применяется в теории и практике, активно используется в самых различных областях знаний для оценки степени риска или шансов на успех, где требуется наилучшее из нескольких вариантов решение [12-16]. Например, в статье [16] на основе метода Байеса выполнены расчеты по определению влияния засорения сопел форсунки и износа топливного насоса на эффективность работы топливных форсунок дизельного двигателя. При выверенной статистике дефектов и их признаков, метод Байеса обеспечивает высокую достоверность во время диагностирования и прогнозирования технического состояния различного оборудования.

Одним из перспективных подходов в анализе временных рядов является их представление в виде байесовского структурного временного ряда (BSTS- модель) [17-24].

Байесовский структурный временной ряд (BSTS) – это статистический метод, используемый для выбора признаков, прогнозирования временных рядов, прогнозирования текущей погоды, определения причинно-следственных связей и других приложений. Модель предназначена для работы с данными временных рядов [17].

В исследовании [18] рассматривается возможности применения модели байесовской векторной авторегрессии (BVAR) для оценки влияния внешних шоков на динамику макроэкономических показателей России.

В работе [19] предпринята попытка выявить основные допущения, предпосылки и приемы методик, использующих теорему Байеса для оценки взаимосвязи признаков социальных акторов и явлений. Обобщены преимущества предположенного подхода по отношению к более традиционным количественным методам.

В книге [20] описана методология по использованию языка программирования и системы статистических вычислений R для анализа временных рядов. Выведены решения нескольких стандартных задач, включая прогнозирование, выявление

структурных изменений и аномалий в данных, а также кластеризацию временных рядов. Представлены соответствующие подходы и программные обеспечения.

В статье [21] описывается система краткосрочного прогнозирования, основанная на ансамблевом прогнозировании, которое усредняет различные комбинации предикторов. Система усредняет потенциальные вклады очень большого набора моделей и дает легко усваиваемые отчеты о том, какие коэффициенты могут быть важными. Результаты показывают, что лежащие в основе статистические методы могут быть применены к более общему краткосрочному прогнозированию с большим количеством одновременных предикторов.

В статье [22] описано, как использовать данные поисковых систем для краткосрочного прогнозирования современных экономических показателей.

В литературе [23] предлагается сделать вывод о причинно-следственном воздействии на основе модели пространства состояний диффузионной регрессии, которая предсказывает контрфактическую реакцию рынка при искусственном контроле (реакцию, если бы вмешательство не произошло). Выявлены сильные стороны и ограничения моделей в пространстве состояний в обеспечении причинно-следственной атрибуции в тех условиях, когда рандомизированный эксперимент недоступен.

В работе [24] рассматриваются особенности применения двух моделей для оценки макроэкономической динамики в США: байесовской векторной авторегрессии (BVAR) и байесовской векторной авторегрессии с марковскими переключениями (MSBVAR), сравнены результаты двух моделей и получен вывод о наиболее подходящей модели.

В структурных моделях временной ряд представлен в виде суммы ненаблюдаемых компонентов, которые можно интерпретировать, например, как тренд, сезонность, цикл, ошибка. Байесовский подход к статистическим проблемам в основе своей является вероятностным. Совместное распределение вероятностей используется для описания взаимосвязей между всеми неизвестными и данными.

Затем вывод основывается на условном распределении вероятностей неизвестных с учетом наблюдаемых данных – апостериорное распределение.

Используя внутреннюю согласованность системы вероятностей, апостериорное распределение извлекает соответствующую информацию из данных и обеспечивает полное и последовательное представление всех переменных состояния (в том числе и ненаблюдаемых компонентов) для обнаружения причинно-следственных связей между прогнозом и наблюдаемыми данными. Использование апостериорного распределения для решения конкретных проблем вывода и принятия решений является в этом случае достаточно простым.

Таким образом, по сравнению с другими моделями BSTS-модель позволяет учитывать неопределенность в данных и сложные взаимосвязи между наблюдаемыми переменными. Подгонка структурных моделей временных рядов выполняется с использованием фильтра Калмана и метода Монте-Карло по схеме марковских цепей (MCMC).

Для оценивания и одновременной регуляризации коэффициентов регрессии применяется т.н. метод «spike-and-slab» (регрессия с пиками и плитами) – тип байесовской линейной регрессии, в которой конкретное иерархическое априорное распределение для коэффициентов регрессии выбирается таким образом, что сохраняется только подмножество возможных регрессоров. Некоторым коэффициентам регрессии присваивается высокая априорная вероятность того, что они равны нулю. В дальнейшем, при MCMC-сэмплировании коэффициентов из полученных апостериорных распределений многие коэффициенты оказываются в точности равными нулю.

Такой механизм регуляризации позволяет эффективно выполнить селекцию наиболее важных предикторов и параллельно избавиться от мультиколлинеарности, благодаря чему в байесовские структурные модели можно включать большое количество предикторов без риска переобучения.

1.8. Выводы по Главе 1

Представлены модели и методы прогнозирования временных рядов, такие как регрессионные модели, нейросетевые модели, модели на основе теоремы Байеса и др. Выполнены обзор современных тенденций в сфере прогнозирования и обзор байесовских подходов. Отмечено, что на сегодняшний день одним из перспективных подходов в анализе временных рядов является их представление в виде байесовского структурного временного ряда (BSTS-модель).

2. Модели пространства состояний

2.1. Структурные модели временных рядов

Выполнить классификацию моделей предметной области, в виду их многообразия, достаточно сложно. Одним их критериев классификации моделей является тип зависимости будущего значения наблюдаемых параметров от прошлых. В этом случае модели временных рядов декомпозируются на два класса: статистические и структурные.

В статистических моделях на первом этапе делается предположение о характере связей между анализируемыми переменными (статистическое предположение), а затем исследуется соответствие реальных данных и данных, полученных с помощью набора статистических предположений. Таким образом, статистические модели представляют, часто в значительно идеализированной форме, процесс генерации данных.

Структурные модели временных рядов – это одно из семейств моделей пространства состояний. В структурных моделях формируется фрейм-структура, которая и определяет функциональную зависимость между имеющимися, прогнозируемыми и внешними факторами. Временной ряд в этом случае представляется в виде суммы ненаблюдаемых компонентов, которые можно интерпретировать, например, как тренд, сезонный компонент, случайный компонент и другие. Эти компоненты временного ряда не наблюдаемые, они являются теоретическими величинами. Под трендом понимают изменение, определяющее общее направление развития, основную тенденцию временного ряда. То есть, тренд – это систематическая составляющая долговременного действия. Сезонные колебания – это колебания, связанные со сменой времен года и имеющие выраженную годовую периодичность. Если из временного ряда удалить тренд и периодические составляющие, то останется нерегулярный компонент (случайный). Случайный компонент является действием большого числа относительно слабых второстепенных факторов. Влияние каждого из второстепенных факторов незначительно, но ощущается их суммарное воздействие.

Пусть y_t обозначает наблюдение в момент t в реальном временном ряду. Структурная модель временного ряда задаётся двумя уравнениями, связывающими y_t с вектором переменных состояния исследуемой системы α_t .

Уравнением наблюдения (1):

$$y_t = Z_t^T \alpha_t + \varepsilon_t; \quad \varepsilon_t \sim N(0, H_t) \quad (1)$$

Где:

y_t – наблюдаемые данные в момент времени t .

Z_t, H_t – структурные параметры (матрицы модели).

α_t – вектор скрытых переменных (состояния исследуемой системы).

Уравнением перехода (2):

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t; \quad \eta_t \sim N(0, Q_t) \quad (2)$$

Где:

T_t, R_t и Q_t – структурные параметры, часто заполнены нулями и единицами.

η_t – может иметь меньшую размерность, чем α_t .

Уравнение (1) называется уравнением наблюдения, поскольку оно связывает наблюдаемые данные y_t с ненаблюдаемым в явном виде (латентном) состоянием α_t . Уравнение (2) называется уравнением перехода, поскольку оно определяет, как латентное состояние изменяется во времени, то есть, задает процесс перехода из одного латентного состояния в другое. Матрицы модели Z_t, T_t , и R_t , обычно содержат смесь известных значений (часто 0 и 1) и неизвестных параметров. Матрица перехода T_t является квадратной, но R_t может быть прямоугольной, если часть переходов состояния детерминирована. Наличие R_t в уравнении (2) позволяет работать с полноранговой матрицей дисперсии Q_t , поскольку любые линейные зависимости в векторе состояния могут быть перенесены из Q_t в R_t .

Состояние системы в каждый момент времени определяется только её состоянием в предыдущий момент, т.е. динамика системы имеет Марковский характер. Схематичное представление модели пространства состояний для конечного временного ряда с T наблюдениями приведено на Рис. 2-1.

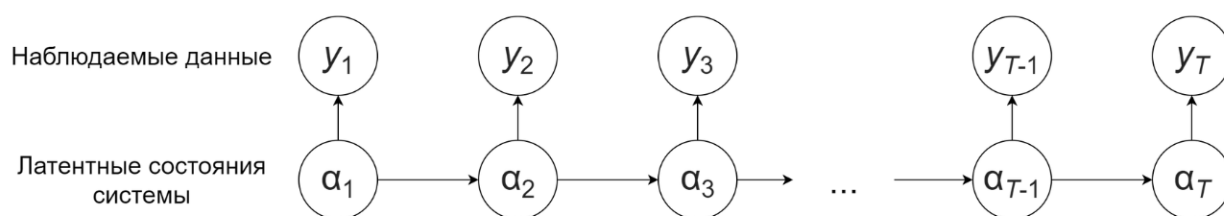


Рис. 2-1. Модель пространства состояний для временного ряда с T наблюдениями

Если модель может быть описана уравнениями (1) и (2), то считается, что она находится в форме пространства состояний, то есть, является моделью пространства состояний. Стоит отметить, что достаточно большой класс моделей может быть выражен в форме пространства состояний. Модели пространства состояний привлекательны потому, что они модульные. Независимые компоненты состояния могут быть объединены путем конкатенации их векторов наблюдений Z_t и расположением других матриц модели как элементов в блочной диагональной матрице. Это обеспечивает значительную гибкость в выборе компонентов для моделирования тренда, сезонности, эффектов регрессии и потенциально других компонентов состояния, которые могут быть необходимы.

2.2. Фильтр и сглаживание Калмана

Основными инструментами для работы с моделями пространства состояний являются фильтр Калмана, сглаживание Калмана и Байесовское расширение данных. Но фильтрация и сглаживание являются традиционными вычислительными операциями, связанными с моделями пространства состояний.

Обозначим всё множество наблюдаемых данных в каждый момент времени t как $y_{1:t} = \{y_1, y_2, \dots, y_t\}$. Фильтр Калмана рекурсивно вычисляет прогнозируемое распределение $p(\alpha_{t+1}|y_{1:t})$, комбинируя $p(\alpha_t|y_{1:t-1})$ с y_t с использованием стандартного набора формул, который логически эквивалентен линейной регрессии. Сглаживание Калмана (сглаживатель Калмана) обновляет выход фильтра Калмана для получения $p(\alpha_t|y_{1:n})$, где n – длина временного ряда при каждом значении t . Поскольку все компоненты модели являются гауссовскими, то и $p(\alpha_{t+1}|y_{1:t-1})$, и $p(\alpha_{t+1}|y_{1:n})$ являются многомерными нормальными распределениями, параметризованными их средним μ_t и дисперсией P_t . Фильтрация Калмана

накапливает информацию о временном ряде по мере его продвижения вперед по списку элементов (μ_t, P_t) . Сглаживатель Калмана перемещается назад по времени, распределяя информацию о более поздних наблюдениях по последовательно более ранним парам (μ_t, P_t) . Стоит отметить, что схема «вперед-назад» является достаточно распространённой для алгоритмов в задачах оптимизации и логического вывода. Фильтр Калмана и сглаживание Калмана являются частными случаями этой схемы.

В Байесовских вычислениях часто желательно моделировать состояние на основе его апостериорного распределения с учетом данных. Пусть $\mathbf{y} = y_{1:n}$ и $\boldsymbol{\alpha} = \alpha_{1:n}$ обозначают полные наборы наблюдаемых и латентных данных. Байесовские методы расширения данных производят моделирование из $p(\boldsymbol{\alpha}|\mathbf{y})$. Нельзя просто взять каждое α_t из $p(\alpha_t|\mathbf{y})$, потому что должна быть соблюдена последовательная корреляция между α_t и α_{t+1} . Вместо этого применяют стохастические версии сглаживателя Калмана, которые могут быть использованы для выборки непосредственно из $p(\boldsymbol{\alpha}|\mathbf{y})$.

2.3. Выводы по Главе 2

Обосновано преимущество использования фильтра и сглаживателя Калмана, которое заключается в том, что не надо предоставлять размеченные данные для обучения (алгоритмы «без учителя») и, кроме того, появляется возможность обработки наблюдений с помехами (восстановить «истинное состояние» отслеживаемого объекта).

3. Общая характеристика и структурная схема анализа временных рядов

В общем случае система анализа и прогнозирования направлена на поиск закономерностей в исходных временных рядах, путём предоставления модели, позволяющей прогнозировать последующие значения характеристик наблюдаемых параметров. При этом необходимо обеспечить и предварительную обработку данных, и анализ исследуемых данных, и управление параметрами модели, и доступ к результату моделирования, и вывода его в соответствующие документы (Рис. 3-1).

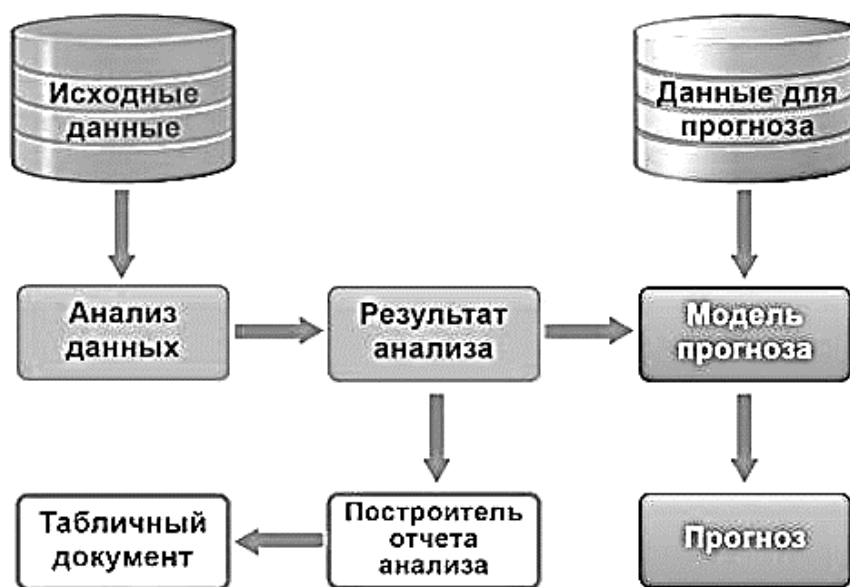


Рис. 3-1. Механизм анализа данных и прогнозирования

3.1. Проектирование структуры системы прогнозирования

Для визуализации, спецификации, конструирования и документирования структуры и поведения системы был использован унифицированный язык моделирования UML [25]. Всего выделяют 13 видов диаграмм. Каждая из диаграмм UML предназначена для представления определенных аспектов системы и обладает своими характеристиками и нотацией.

Диаграмма вариантов использования (Use Case Diagram) – это один из видов UML-диаграмм, который используется для моделирования функциональности

системы из перспективы её акторов. Она позволяет визуализировать и описать взаимодействие между акторами и системой в терминах сценариев использования.

На Рис. 3-2 показана разработанная диаграмма вариантов использования для системы построения прогнозной модели.

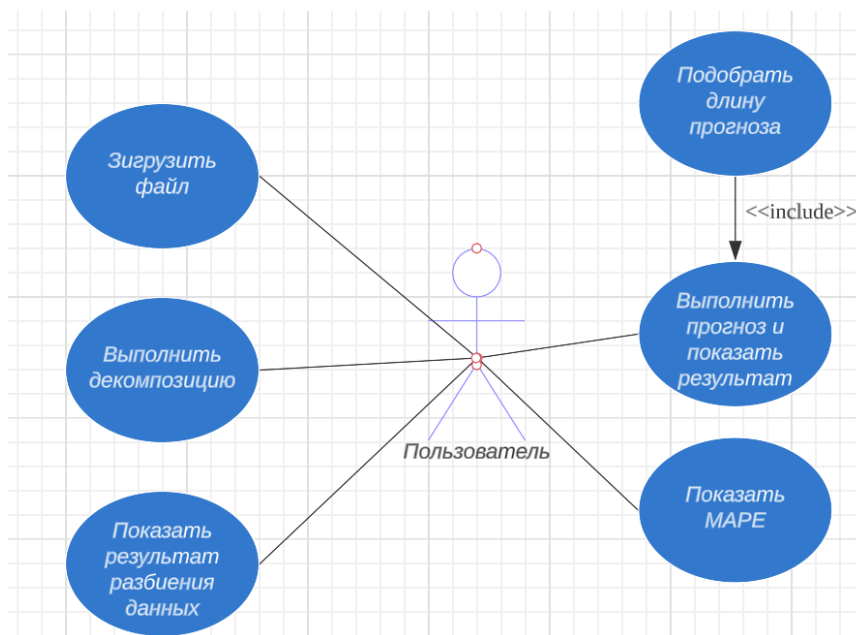


Рис. 3-2. Диаграмма вариантов использования для системы построения прогнозной модели

Диаграмма деятельности (Activity Diagram) – это один из видов UML-диаграмм, который используется для моделирования процессов, действий и потоков управления в системе. Она помогает представить последовательность шагов и условий, определить параллельные процессы, принимать решения и моделировать циклы выполнения. На Рис. 3-3 представлена разработанная диаграмма вариантов использования для системы построения прогнозной модели. Диаграмма последовательно отображает координированные виды деятельности, которые предусматривает данная система.

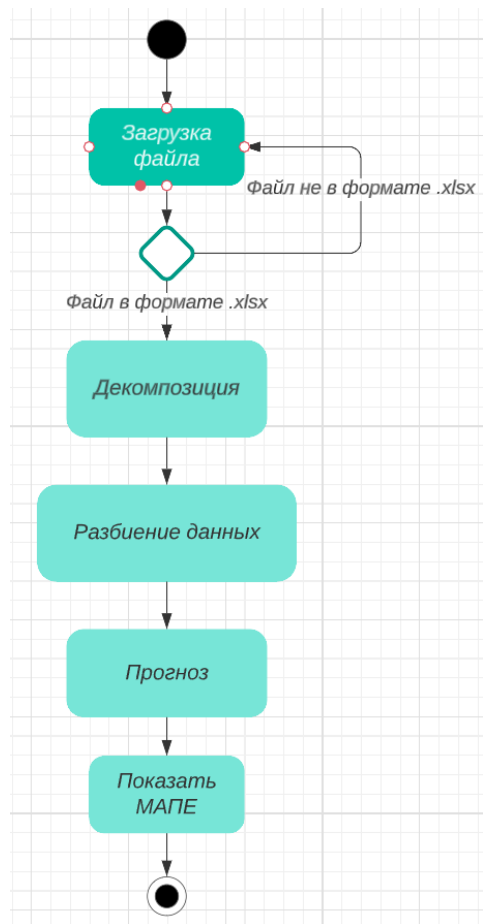


Рис. 3-3. UML-диаграмма деятельности

3.2. Схема анализа данных и прогнозирования

Анализ данных и прогнозирование, в общем случае, является направленной процедурой, которую можно представить в виде алгоритма, нацеленного на получение ответа по поставленной задаче (рис. 3.4).

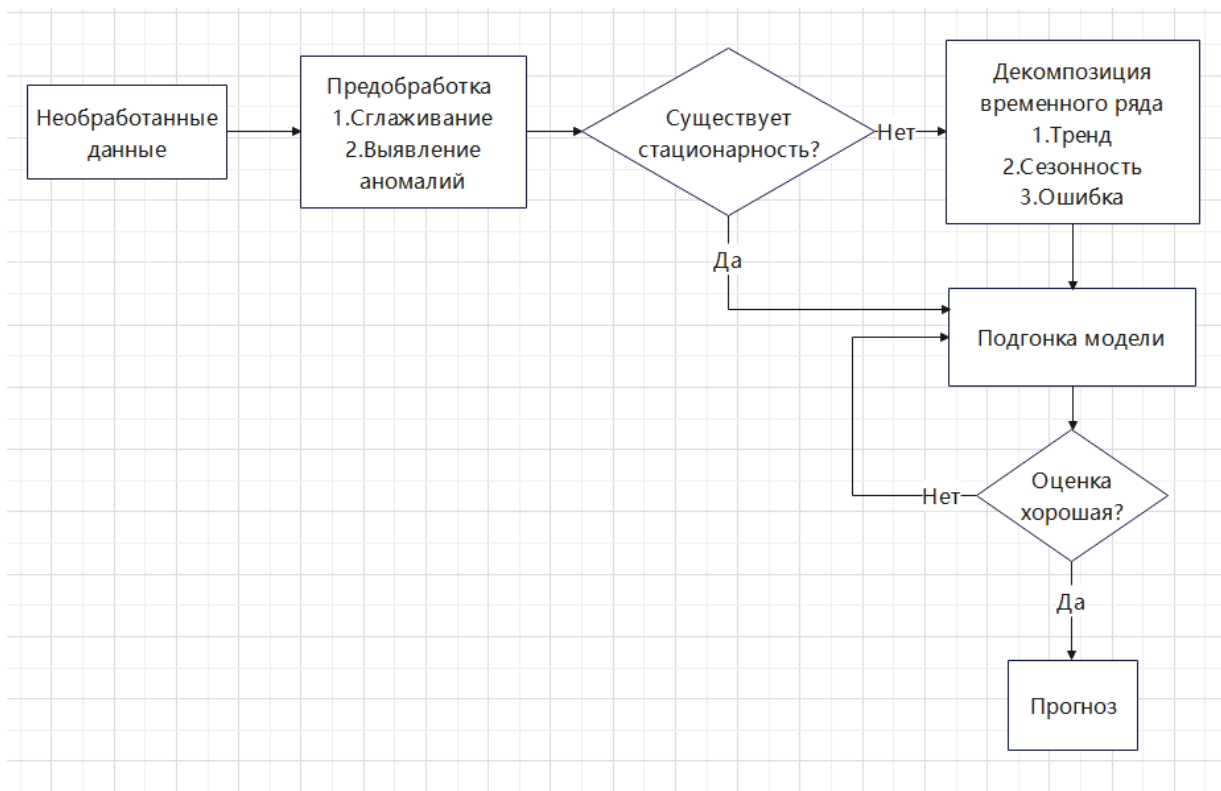


Рис. 3-4. Обобщенная функциональная структура системы построения моделей

3.2.1. Предварительная обработка данных

После получения данных первым шагом является их предобработка, которая заключается в сглаживании данных и выявлении аномальных наблюдений. Перед анализом временных рядов такие ошибки, как скачки и очевидные систематические ошибки, следует убрать путем удаления или использованием математических методов. Алгоритмы машинного поиска аномалий используют данные о работе процесса. В зависимости от предметной области в наблюдаемых временных рядах могут быть аномалии разного вида – точечные, групповые, аномалии контекста. При поиске аномалий во временных рядах применяют методы, основанные на информации о близости параметров, реконструкции фрагментов данных и др.

3.2.2. Предварительный анализ временных рядов

Второй шаг – предварительный анализ временных рядов. Целью этого шага является выявление характеристик изучаемого временного ряда и подбор

подходящей модели, поскольку разные модели способствуют анализу временных рядов с разными характеристиками, и неправильный подбор модели может затруднять анализ данных. К процедурам предварительного анализа относятся тест на стационарность, декомпозиция временных рядов и построение автокорреляционной функции.

3.2.3. Тест на стационарность

При анализе случайных процессов задача проверки стационарности изучаемого процесса занимает важное место. Суждение о стационарности позволяет исследователю уточнить модель процесса или выявить причины нарушения стабильности хода эксперимента.

Тест на стационарность осуществляется путем проверки гипотез о постоянстве среднего, определяемого непараметрическим критерием сдвига, о постоянстве дисперсии процесса или его остатков, определяемых критерием рассеяния.

В случае опровержения гипотезы о постоянстве среднего проводится процедура удаления тренда. Если же процесс имеет ненулевое постоянное среднее, то он сглаживается процедурой удаления среднего. При любом исходе процесс центрируется относительно нуля.

Если при проверке постоянства дисперсии процесса или его остатков после удаления непериодического тренда выявлено нарушение постоянства среднего квадрата, тогда процесс относится к классу нестационарных процессов, причем первичным является предположение о том, что его нестационарность обусловлена только изменением дисперсии с течением времени.

В случае принятия гипотезы о постоянстве дисперсии можно проверить процесс на стационарность по критерию Пирсона или любому другому критерию, позволяющему выявить нестационарность.

Проверка постоянства среднего значения

Существует несколько алгоритмов проверки постоянства среднего значения процесса. Одним из наиболее эффективных непараметрических критериев является

критерий инверсий, применимый как к последовательности отсчетов процесса, так и к последовательностям его характеристик.

Данный критерий строится на понятии инверсии, которая заключается в том, что в последовательности чисел за большим следует меньшее. Если отсчеты отражают стационарный процесс, то число инверсий I независимо от распределения наблюдаемого процесса обладает известным распределением. Так, величина γ , определяемая формулой

$$\gamma = 1 - \frac{4I}{N(N-1)},$$

имеет среднее значение, равное нулю, и дисперсию, равную:

$$D_\gamma = \frac{2(2N-5)}{9N(N-1)}$$

Величина γ , уже начиная с $N > 10$, имеет распределение, близкое к нормальному. Поэтому для проверки нулевой гипотезы о том, что дрейфа среднего значения нет, можно использовать неравенство

$$\left| \frac{\gamma}{\sqrt{D_\gamma}} \right| < Z_\varphi \left(\frac{1+P_D}{2} \right).$$

Проверка постоянства дисперсии

К алгоритмам проверки постоянства дисперсии процесса относятся критерий Кохрена, критерий рассеяния и ряд других алгоритмов.

Критерий Кохрена применяется для выяснения вопроса о том, принадлежат ли несколько оценок дисперсий σ_i^2 одному и тому же нормальному распределению. Если по нескольким участкам реализации вычислены оценки дисперсий σ_i^2 и вычисленная по формуле

$$G = \frac{\sigma_i^2 \max}{\sum_{i=1}^K \sigma_i^2}$$

статистика G превышает критическое значение $Z_G(v, K, P_{\text{дов}})$,

где $v = N - 1$,

N – объём каждой выборки,

K – число участков (выборок),

то предположение о стационарности следует отвергнуть.

Следует отметить, что данный критерий справедлив при предположении о нормальности и очень чувствителен к нормальности исходного распределения. Его применение может приводить к серьезным ошибкам, если предположение о нормальности не справедливо.

Проверка стационарности

Для проверки стационарности распределений может применяться следующая логическая схема: из реализации извлекаются две или более выборки, состоящие из отсчетов x_i, y_i, \dots , снятых на нескольких участках записи. По полученным отсчетам рассчитывается статистика критерия, которая сравнивается с критическим значением при выбранной доверительной вероятности.

Для проверки согласия распределений разных выборок применимы критерии согласия эмпирических распределений с теоретическими, поэтому проверку стационарности распределений основывают на методах проверки согласия распределений.

Критерий X^2 . Пусть имеется n_x отсчетов x_i с первого участка записи, n_y отсчетов y_i со второго участка и т.д. Общее количество отсчетов равно N . Выдвигается нулевая гипотеза о том, что эти выборки принадлежат одной и той же генеральной совокупности, а отклонения в распределениях для каждого участка объясняются лишь случайным характером процесса.

Использование X^2 при анализе временных рядов возможно с применением следующего алгоритма: выборка объема N разбивается на L интервалов и параллельно на две группы с равным числом отсчетов $N/2$. Границы разрядов выбираются по каждой группе отдельно и по всей реализации в целом. В идеале предполагается, что если во всей реализации есть s_i элементов из i -го разряда, то на каждом из участков будет по s_i отсчетов этого же интервала. После сортировки каждой группы и всей реализации по интервалам количества отсчетов, попавшие в разряды, можно описать в форме таблицы (Табл. 3-1):

Табл. 3-1. Разбиение выборки

n_{x1}	n_{x2}	n_{xL}	N_x
n_{y1}	n_{y2}	n_{yL}	N_y
n_1	n_2	n_L	N

По таблице оцениваются вероятности принадлежности отсчетов каждому из участков X и Y, и вычисляется статистика

$$X^2 = 2 \sum_{j=1}^L \frac{1}{n_j} \left[(n_{x_j} - \frac{n_j}{2})^2 + (n_{y_j} - \frac{n_j}{2})^2 \right],$$

которая сравнивается с критическим значением $Z_{\text{критич}}(v, P_{\text{дов}})$,

Где $v = L - 1$. Нулевая гипотеза отвергается при $X^2 > Z_{\text{критич}}(v, P_{\text{дов}})$.

3.2.4. Декомпозиция временных рядов

Использование временных рядов для прогнозирования предполагает инерционность процесса, т.е. факторы, влияющие на исследуемое явление в прошлом и настоящем, будут продолжать действовать схожим образом и не в очень далеком будущем. Поэтому основной целью анализа временных рядов выступает разложение уровней ряда на составные компоненты с целью учета их при прогнозировании, т.е. декомпозиция временных рядов.

Фактическую величину уровня динамического ряда можно представить, как функцию трех компонентов:

$$Y_t = f(T, S, E)$$

Где Y_t – фактический уровень ряда в период времени t ;

T – тренд ряда;

S – периодические колебания (циклические, сезонные);

E – случайная составляющая.

Проведение декомпозиции позволяет понять структуру ряда, построить его модель и экстраполировать уровни ряда на краткосрочную перспективу. Рассматриваемые компоненты динамического ряда не обязательно имеются в каждом временном ряду. Могут быть ряды динамики, в которых отсутствует тенденция, или периодические колебания, или и то, и другое.

Проверка наличия тренда

Для проверки наличия тенденции наиболее широко применяется метод сравнения средних. Данный метод применяется для выявления монотонно возрастающей или монотонно убывающей тенденции. Временной ряд разбивается на две, примерно одинаковые, части y_1, y_2, \dots, y_{n_1} и $y_{n_1+1}, y_{n_1+2}, \dots, y_{n_1+n_2}$, вычисляются средние ($y_{1\text{ср}}$ и $y_{2\text{ср}}$) и выборочные дисперсии (s_1^2 и s_2^2) для обеих частей соответственно.

Далее рассчитывается значение критерия Стьюдента по формуле:

$$t_B = \frac{|y_{1\text{ср}} - y_{2\text{ср}}|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

если предполагается, что значения дисперсий на этих участках не равны между собой.

И по формуле:

$$t_B = \frac{|y_{1\text{ср}} - y_{2\text{ср}}|}{s^2} \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}$$

где s^2 – общая выборочная дисперсия ряда, если предполагается, что дисперсии одинаковы.

Нулевая гипотеза о равенстве средних (об отсутствии тенденции) отвергается, если выполняется условие $t_B > t_{\text{кр}}(1-\alpha, m)$, где $t_{\text{кр}}(1-\alpha, m)$ – табличное значение t -критерия Стьюдента при уровне значимости α и числе степеней свободы $m = n_1 + n_2 - 2$.

Выделение периодических составляющих временного ряда

Для выделения периодических составляющих временного ряда часто применяется спектральный анализ. С помощью преобразования Фурье любой ряд динамики можно представить в виде суммы конечного числа гармоник. Но часто нужны не все гармоники, а лишь основные, порождающие основную часть дисперсии процесса.

Рассмотрим реализацию $x(t)$ длиной T , принадлежащую стационарному случайному процессу и имеющую нулевое среднее значение. При дискретном

временном параметре такая реализация представлена N (т.к. $T=N \cdot \Delta t$) значениями временного ряда $\{x(t)\}$. Дискретное преобразование Фурье дает значения спектральной плотности на частотах:

$$f_k = \frac{k}{T} = \frac{k}{\Delta t \cdot N}, \quad k = 0, N-1$$

При этом коэффициенты Фурье определяются в виде:

$$X(f_k) = \Delta t \cdot \sum_{i=0}^{N-1} x(i) \cdot e^{-\frac{2\pi i k}{N}}$$

Оценку спектральной плотности:

$$\tilde{S}(f_k) = \frac{1}{\Delta t \cdot N} |X_i(f_k)|^2$$

можно рассматривать как характеристику степени тесноты связи между $X(i)$ и гармоническим компонентом с периодом $2\pi/\omega$.

3.3. Подгонка и оценка модели

Для оценки модели данные разделены на обучающую и проверочную выборки. Модель создается на основе обучающих данных и сравнением предсказания с данными из проверочной выборки.

В качестве метрики оценки применяется средняя абсолютная удельная ошибка предсказаний (MAPE). Её значение рассчитывается по формуле

$$\text{MAPE} = \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

где y_i и \hat{y}_i – истинное и предсказанное значения моделируемой переменной, n – количество наблюдений.

Еще одним важным свойством хорошей модели временного ряда является отсутствие автокорреляции в ее остатках. Автокорреляция – обычная корреляция, применённая к самому ряду, сдвинутому на какое-то количество отсчётов. Она характеризует насколько зависит наблюдаемое значение от предыдущих. Ниже представлен примеры распределений автокорреляционной функции остатков моделей.

В идеале центры апостериорных распределений (начиная со сдвига 1 и далее) должны приходиться на 0.

После оценки модели уже можно прогнозировать следующие возможные значения.

3.4. Выводы по Главе 3

Представлено проектирование с использованием языка UML обобщенной структурной схемы анализа и прогнозирования временных рядов. Приведено описание основных этапов (предварительная обработка, оценка на стационарность, декомпозиция, подгонка модели и др.) обобщённого алгоритма, нацеленного на получение прогнозных оценок.

4. Программные эксперименты по построению прогнозных моделей

4.1. Предварительная обработка данных с использованием пакета `bssm`

Объектами исследования в данной части являются временные ряды, составленные из температур воздуха в городе Томске во временном промежутке с 2016.05.31 по 2018.04.11. Интервал измерения – одна минута. Данные были получены с сайта системы климатического мониторинга [26]. На Рис. 4-1 показано распределение температурного ряда.

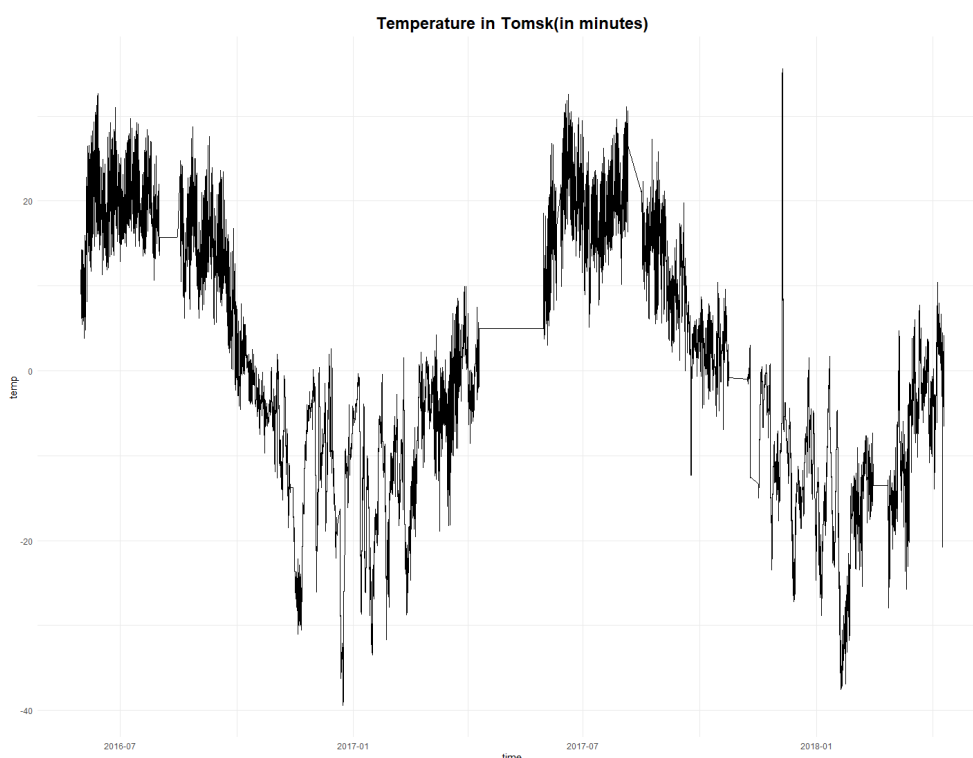


Рис. 4-1. Температуры воздуха в Томске в период с 2016.05.31 по 2018.04.11

В начале импортируем данные и представим их в виде временных рядов.

```
temp <- read_excel("E:/Desktop/data.xlsx",  
                  sheet = "temp", col_types = c("date", "numeric"))
```

В наборе данных есть пропуски, заполним все пропущенные времени и соответствующие температуры, выполняя линейную интерполяцию.

```
full_time_seq <- seq(min(temp$time),
```



```

max(temp$time), by = "1 hour")

full_time_df <- data.frame(time = full_time_seq)
merged_df <- merge(full_time_df, temp3, by = "time", all = TRUE)
merged_df$temp[is.na(merged_df$value)] <- NA
merged_df$temp <- na.interp(merged_df$temp)

```

Для уменьшения чувствительности данных к шумам и затратам времени на обработку изменим шаг дискретизации временного ряда. Преобразуем минутный шаг дискретизации в среднесуточный, вычисляя среднюю температуру воздуха за каждые сутки. Распределение температур после преобразования представлено на Рис. 4-2.

```

temp <- temp %>%
  mutate(date = as.Date(time)) %>%
  group_by(date) %>%
  summarise(temp = round(mean(temp), 2))

```

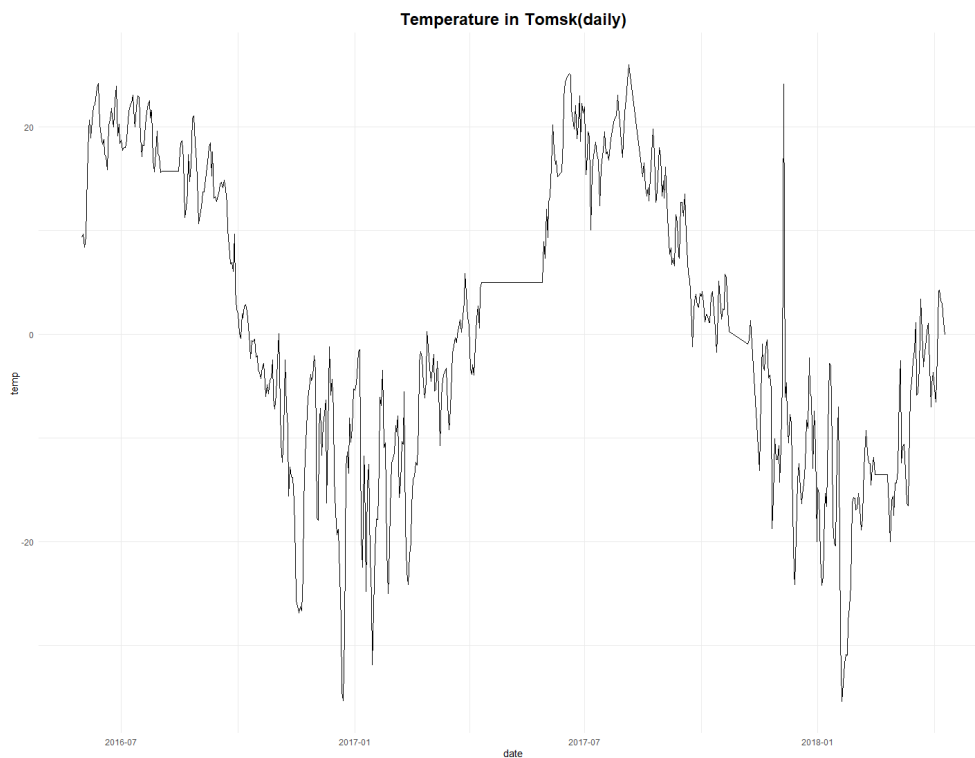


Рис. 4-2. Распределение температур после преобразования

На Рис. 4-2 ещё наблюдаются заметные скачки, которые считаются ошибками, их нужно убрать.

```
clean <- temp %>%  
  time_decompose(temp) %>%  
  anomalize(remainder) %>%  
  time_recompose() %>%  
  plot_anomalies()
```

На Рис. 4-3 представлено выявление точечных аномалий. Значения, обведенные красным кружком, считаются ошибками.

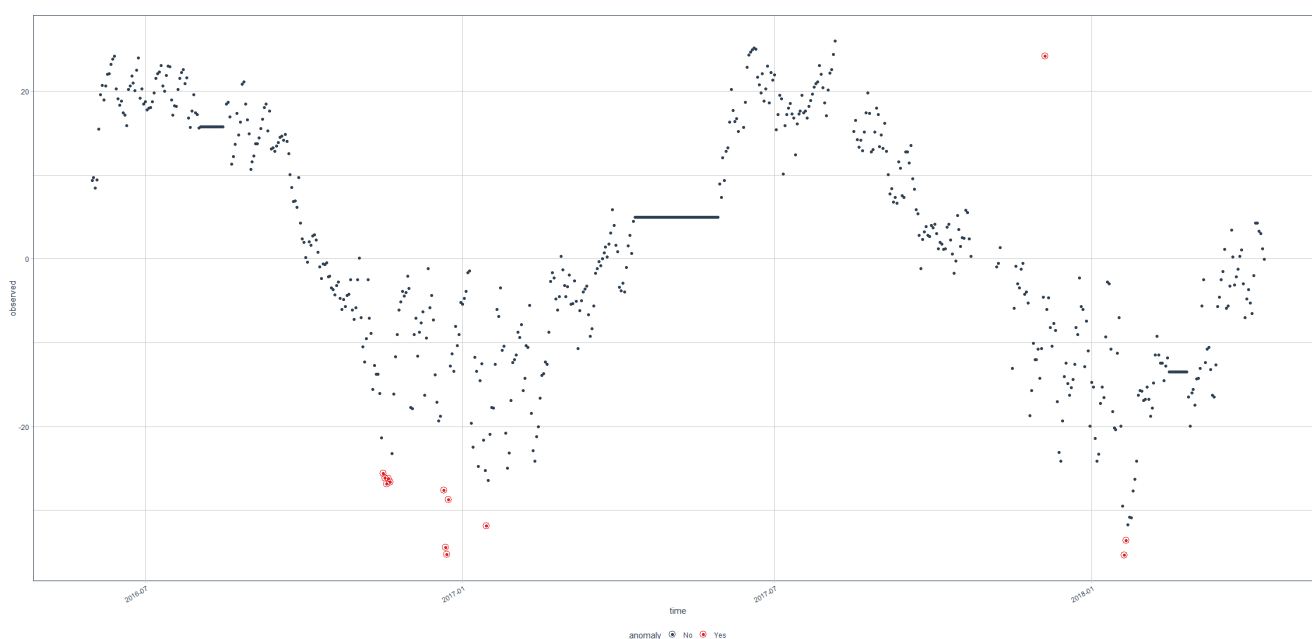


Рис. 4-3. Выявление точечных аномалий

На Рис. 4-3 еще наблюдаются 3 горизонтальных линии, что также считается ошибкой. Их нужно убрать и заполнить места прогнозированными данными, полученными с помощью пакета `bssm` [27].

```
temp_bssm <- read_excel("E:/Desktop/cleaned1.xlsx",  
  col_types = c("date", "numeric"))  
xreg_ <- temp_bssm$time %>% as.matrix()  
sm_model <- bsm_lg(temp_bssm$temp,  
  xreg = xreg_,  
  beta = normal_prior(rep(0, ncol(xreg_)), 0, 1),
```

```

sd_y = gamma_prior(1, 2, 0.01),
sd_level = 0,
sd_slope = gamma_prior(1, 2, 0.01),
sd_seasonal = gamma_prior(1, 2, 0.01),
period = 30)

fit <- run_mcmc(sm_model, iter = 1000, burnin = 100)
obs <- data.frame(Time = temp_bssm$time,
                  Temp = temp_bssm$temp)%>%
  filter(!is.na(temp_bssm$temp))

pred <- fitted(fit, sm_model)
pred$time <- temp_bssm$time

pred %>%
  ggplot(aes(x = Time, y = Mean)) +
  geom_ribbon(aes(ymin = `2.5%`, ymax = `97.5%`),
            alpha = 0.5, fill = "steelblue")+
  geom_line() +
  geom_point(data = obs,
            aes(x = Time, y = Temp), colour = "tomato") +
  theme_bw()

```

Данный код создает объект модели линейного байесовского состояния-прогнозирования с компонентами уровня, сезонности и регрессии, выполняет оценку и прогнозирование модели BSSM для временного ряда, а затем строит график для сравнения прогнозов с наблюдаемыми значениями (Рис. 4-4).

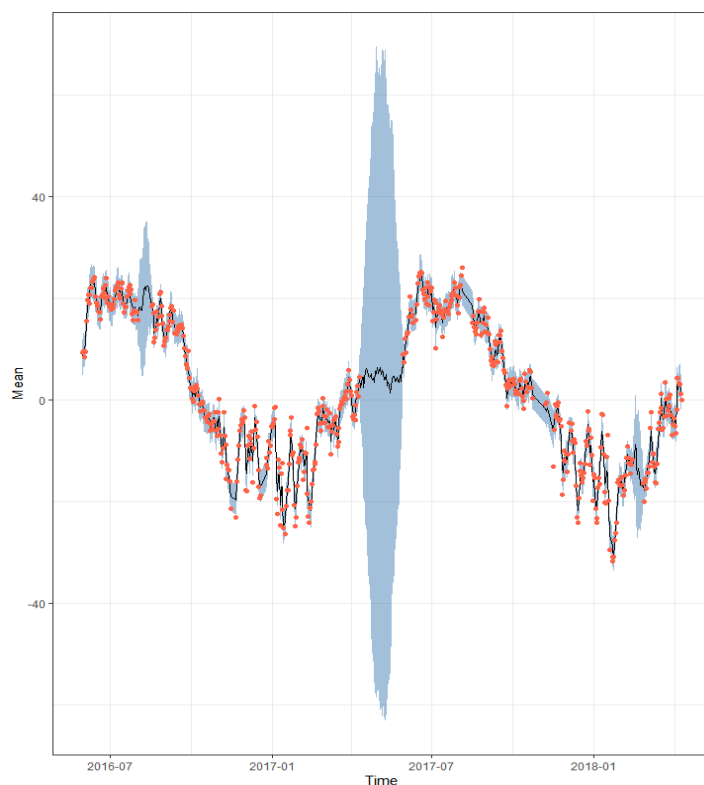


Рис. 4-4. Распределение очищенных данных. Красные точки: наблюдаемые температуры, сплошные линии: прогнозируемые температуры, синие ленты: ограничивают 95%–й доверительный интервал предсказанных значений.

Распределение очищенных данных представлено на Рис. 4-5.

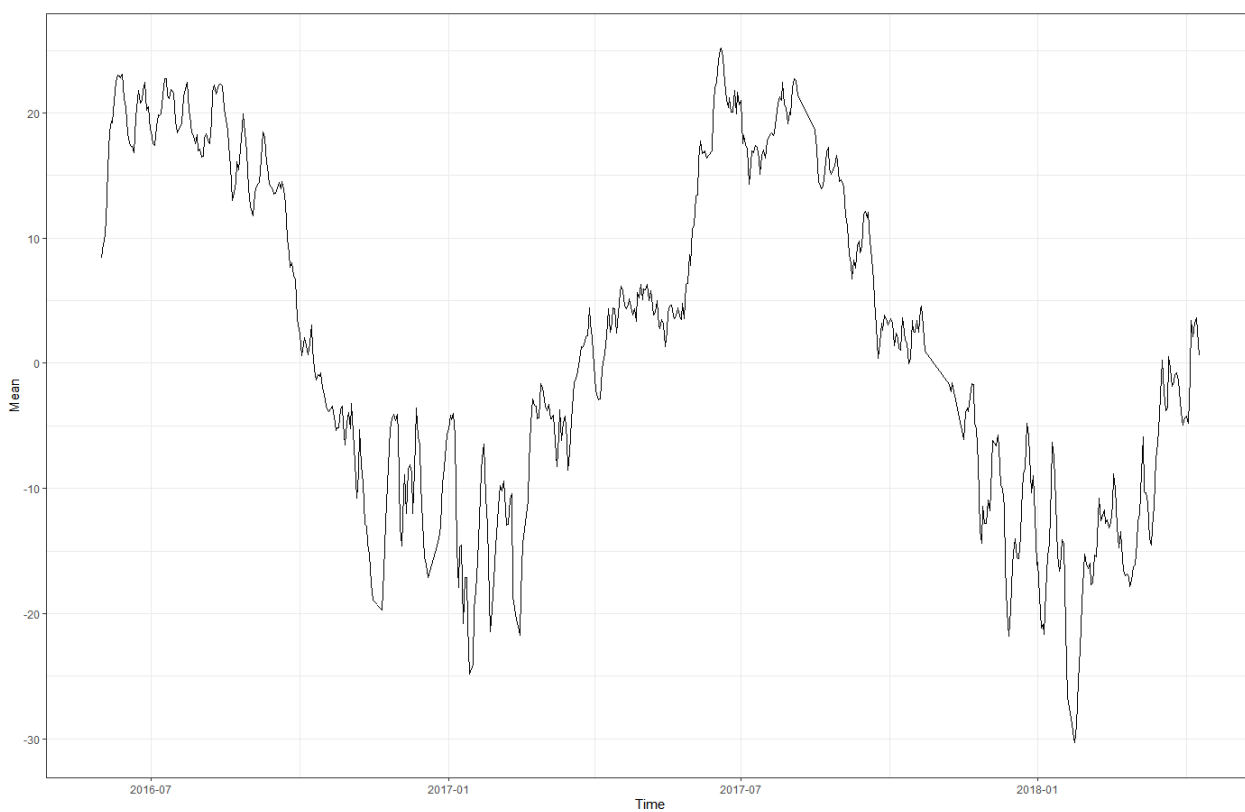


Рис. 4-5. Распределение очищенных данных

4.2. Программная реализация модели `bsts`

Далее декомпозируем временной ряд, чтобы понять его структуру. На Рис. 4-б представлен результат декомпозиции данного ряда. Из рисунка видно, что ряд имеет сезонность с периодом одного года, но не имеет существенного тренда.

```
de <- decompose(temp)
plot(de)
```

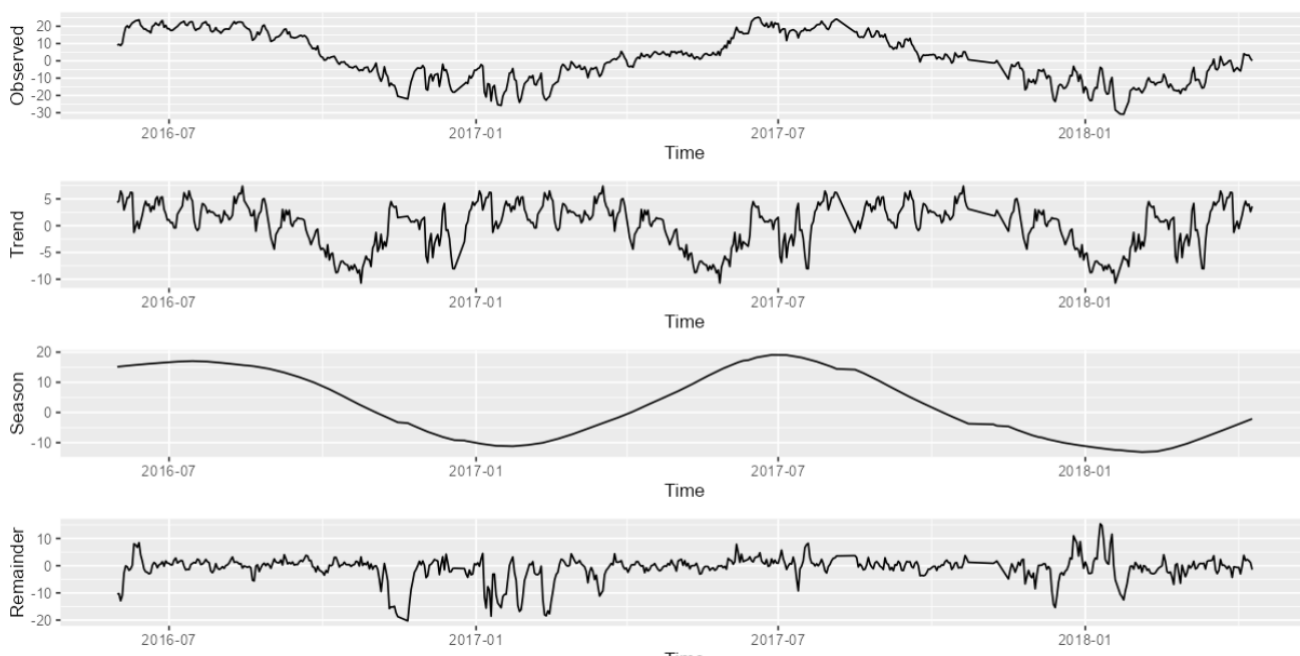


Рис. 4-6. Результат декомпозиции временного ряда

Далее разделяем данные на обучающую и проверочную выборки. Длина прогноза составляет 110 дней. Результат показан на Рис. 4-7.

```
temp <- temp %>%
  index_by(dt = time) %>%
  summarise(y = temp)

cut_point <- as.POSIXct("2018-01-01 00:00:00")
temp_train <- temp %>%
  filter(dt <= cut_point)
```

```

temp_test <- temp %>%
  filter(dt > cut_point)

dplyr::bind_rows(mutate(temp_train, dataset = "train"),
                 mutate(temp_test, dataset = "test")) %>%
  ggplot(aes(dt, y, col = dataset)) +
  geom_line() + geom_point(alpha = 0.4) +
  theme_minimal() +
  scale_color_manual(values = c("blue", "black"))
temp_1 <- temp_train$y
dt_1 <- temp_train$dt

```

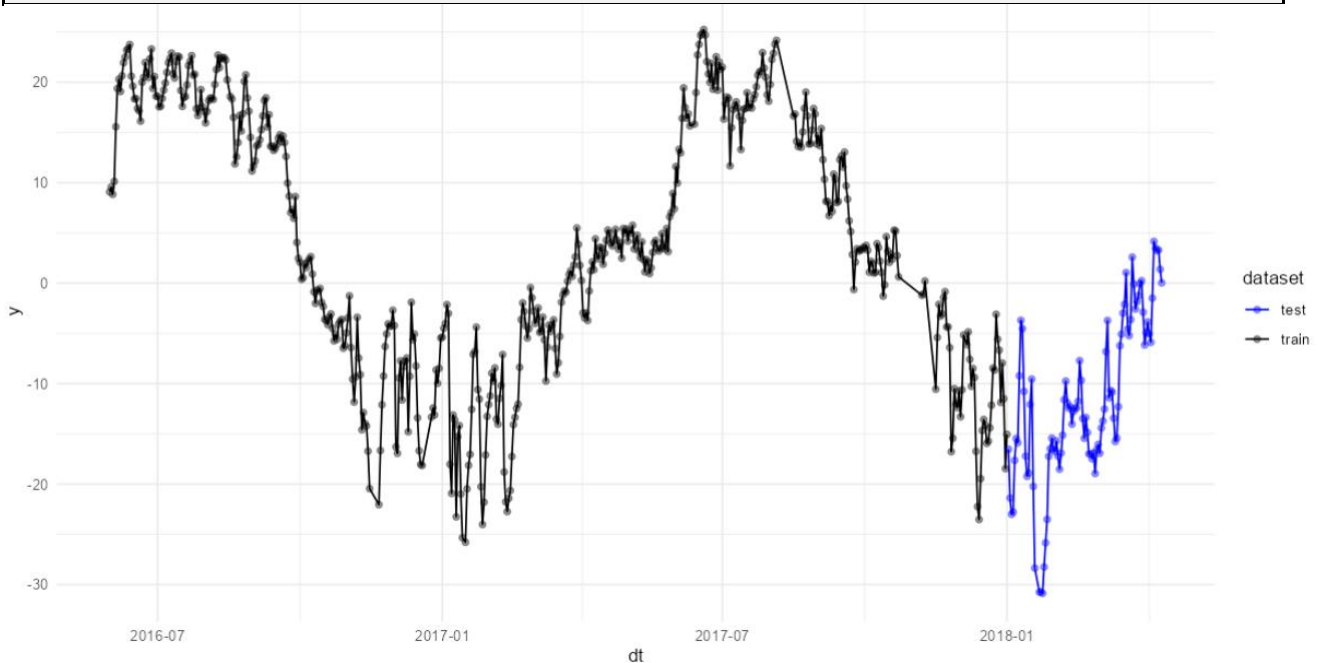


Рис. 4-7. Разделение данных

Далее идет подгонка модели. Добавим в модель компонент годовой сезонности и авторегрессионный процесс. На Рис. 4-8 показан результат подгонки модели.

```

ss <- list()
ss <- AddSeasonal(ss, temp_1, nseasons=365)
ss <- AddSeasonal(ss, temp_1, nseasons=90, season.duration = 4)

```

```

ss <- AddSeasonal(ss, temp_1, nseasons=30, season.duration = 12)
ss <- AddAutoAr(ss, temp_1, lags = 2)
M<- bststs(temp_1, ss,
           timestamps = dt_1,
           niter = 650, ping = 8, seed = 40 )

plot(M)

```

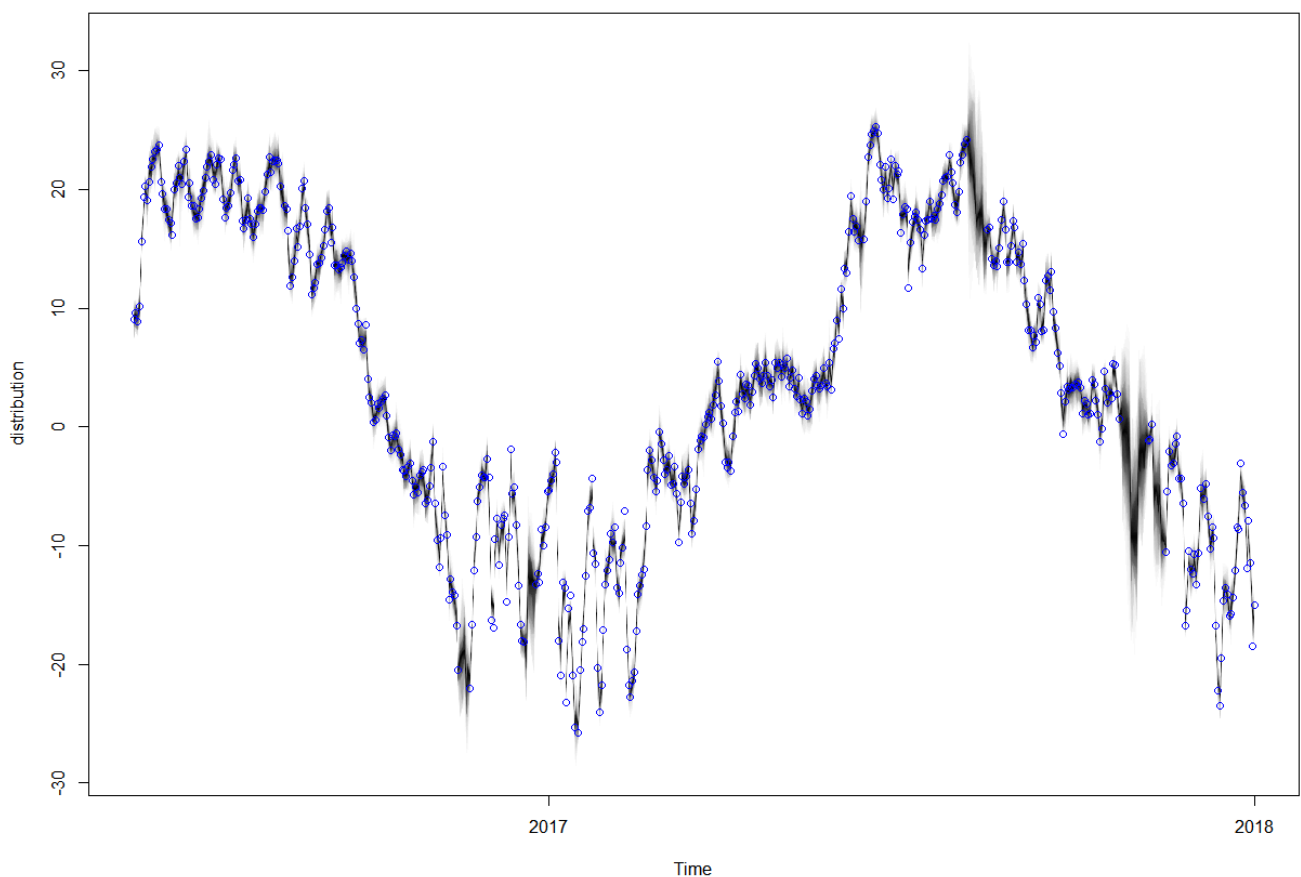


Рис. 4-8. Результат подгонки модели. Синие кружки – обучающие данные. Черные линии – апостериорные распределения компонентов модели.

```

M_pred<- predict(M, horizon = 98)
plot(M_pred, ylim = c(-40,20), plot.original = 10)
with(temp_test, points(dt, y, col = "RED"))

```

Далее идут прогноз и визуальная оценка прогноза. На Рис. 4-9 представлен результат прогноза. Синяя линия – прогноз. Красные кружки – данные из проверочной выборки.

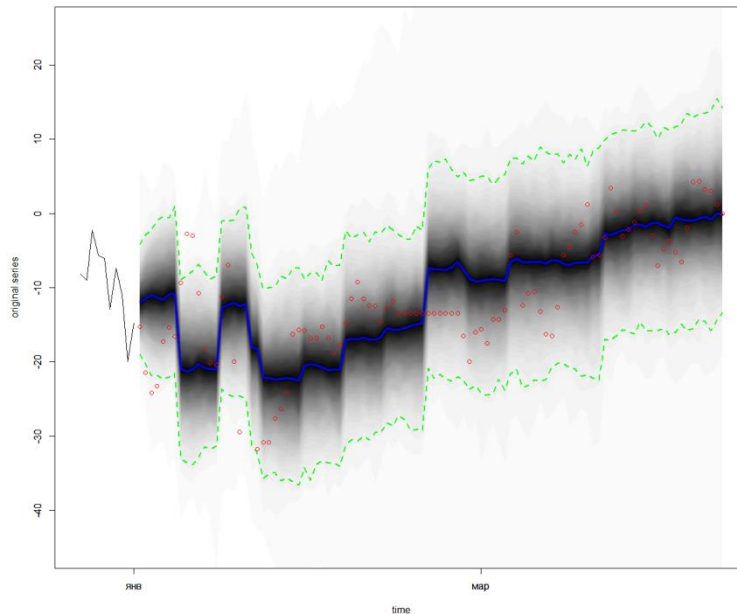


Рис. 4-9. Результат прогноза

Далее построим автокорреляционную функцию модели (Рис. 4-10). Данная форма графика означает, что модель хорошо учитывает все особенности данных.

```
M_resid <- residuals(M4.5)
AcfDist(draws = M_resid)
```

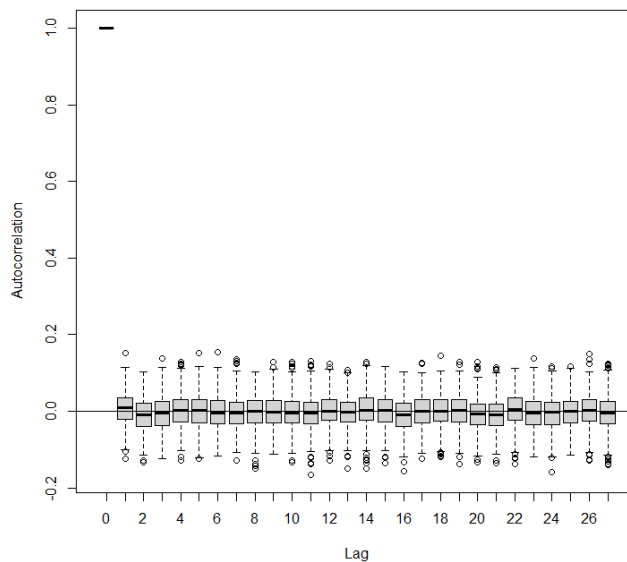


Рис. 4-10. Распределения автокорреляционной функции остатков

Далее рассчитываем среднюю абсолютную удельную ошибку (MAPE). На Рис. 4-11 показано это значение.

```
mape <- function(observed, predicted){  
  mean(abs(observed - predicted)/observed)  
}  
sapply(list("M" = M_pred),  
       mape, observed = temp_test$y) %>%  
round(., 2)
```

MAPE: 0.23

Рис. 4-11. MAPE данной модели

Далее прогнозируем значения будущих 1000 дней. Результат представлен на Рис. 4-12. Визуально можно оценить, что созданная модель хорошо отображает годовую сезонность – температура плавно изменяется с -25°C до $+25^{\circ}\text{C}$. Можно сказать, что модель создана успешно.

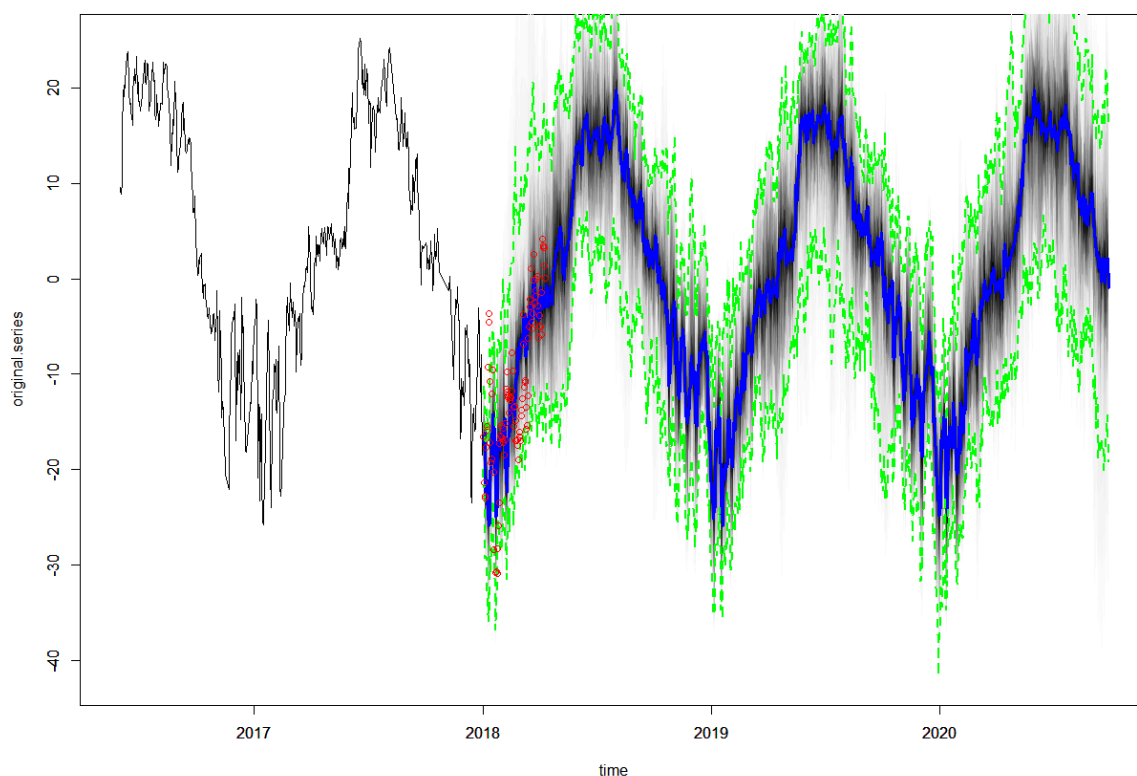


Рис. 4-12. Результат прогнозирования температур будущих 1000 дней

4.3. Разработка веб-интерфейса для программного обеспечения

Для разработки веб-интерфейса используется пакет shiny - пакет R, упрощающий создание интерактивных веб-приложений (приложений) прямо из R [28]. В структуре приложения Shiny сформирован файл app.R, содержащий скрипты ui и server. Ниже представлены коды этих скриптов.

```
ui <- fluidPage(  
  titlePanel("Decompose and forecast of time series data"),  
  sidebarLayout(  
    sidebarPanel(  
      fileInput("file", "upload .xlsx file"),  
      actionButton("decompose_btn", "Decompose")  
    ),  
    mainPanel(  
      plotOutput("combined_plot")  
    ),  
  ),  
  sidebarLayout(  
    sidebarPanel(  
      actionButton("temp_btn", "Into training and test datasets")  
    ),  
    mainPanel(  
      plotOutput("combined_plott")  
    ),  
  ),  
  sidebarLayout(  
    sidebarPanel(  
      actionButton("do_btn", "Forecast"),  
      numericInput("horizon_input", "Length:", value = 1000)
```

```

    ),
    mainPanel(
      plotOutput("combined_plottt")
    )
  ),
  sidebarLayout(
    sidebarPanel(
      actionButton("mape_btn", "show mape")
    ),
    mainPanel(
      textOutput("showmape")
    )
  )
)

```

```

server <- function(input, output) {
  haha5 <- NULL
  temp <- NULL
  temp_1 <- NULL
  malie <- NULL
  mape <- NULL
  temp_test <- NULL

  decomposed_df <- eventReactive(input$decompose_btn, {
    file_path <- input$file$datapath
    data <- read_excel(file_path, col_types = c("date", "numeric"))
    data <- data %>% mutate(time = as.Date(time))
    haha5 <<- as_tsibble(data, key = NULL, index = time)
  })
}

```

```

        decomposed <- haha5 %>% time_decompose(temp, method = "stl",
frequency = "1 year", trend = "auto")
    })

    forecast_df <- eventReactive(input$temp_btn, {
        temp <- haha5 %>%
            index_by(dt = time) %>%
            summarise(y = temp)
        temp
    })

    forecast_dff <- eventReactive(input$do_btn, {
ss4.5 <- list()
ss4.5 <- AddSeasonal(ss4.5, temp_1, nseasons = 90, season.duration = 4)
ss4.5 <- AddSeasonal(ss4.5, temp_1, nseasons = 30, season.duration = 12)
ss4.5 <- AddAutoAr(ss4.5, temp_1, lags = 2)
ss4.5
    })

    showmape_df <- eventReactive(input$mape_btn, {
M_pred_4.5 <- predict(M4.5, horizon = 98) %>% .$median
malie <- sapply(list("M" = M_pred_4.5), mape, observed = temp_test$y)
        malie
    })

    output$combined_plot <- renderPlot({
        if (!is.null(decomposed_df())) {

```

```

plot_observed <- ggplot(decomposed_df(), aes(x = time, y = observed)) +
  geom_line() +
  labs(x = "Time", y = "Observed")

plot_season <- ggplot(decomposed_df(), aes(x = time, y = season)) +
  geom_line() +
  labs(x = "Time", y = "Trend")

plot_trend <- ggplot(decomposed_df(), aes(x = time, y = trend)) +
  geom_line() +
  labs(x = "Time", y = "Season")

plot_remainder <- ggplot(decomposed_df(), aes(x = time, y = remainder)) +
  geom_line() +
  labs(x = "Time", y = "Remainder")

combined_plot <- grid.arrange(plot_observed, plot_season, plot_trend,
plot_remainder, nrow = 4)
  print(combined_plot)
}
})

output$combined_plott <- renderPlot({
  if (!is.null(forecast_df())) {
    cut_point <- as.POSIXct("2018-01-01 00:00:00")

    temp_train <-<- temp %>%
      filter(dt <= cut_point)

```

```

temp_test <- temp %>%
  filter(dt > cut_point)

plot_result <- dplyr::bind_rows(mutate(temp_train, dataset = "train"),
  mutate(temp_test, dataset = "test")) %>%
  ggplot(aes(dt, y, col = dataset)) +
  geom_line() + geom_point(alpha = 0.4) +
  theme_minimal() +
  scale_color_manual(values = c("blue", "black"))

print(plot_result)

temp_1 <- temp_train$y
dt_1 <- temp_train$dt

mape <- function(observed, predicted) {
  abs(mean(abs(observed - predicted)/observed))
}

})

output$combined_plottt <- renderPlot({
  if (!is.null(forecast_dff())) {
    M4.5 <- bsts(temp_1, ss4.5,
      timestamps = dt_1,
      niter = 50, ping = 150, seed = 40, burnin=5)
  }
})

```

```

M_pred_14.5 <- predict(M4.5, horizon = input$horizon_input)

plot(M_pred_14.5, ylim = c(-42, 25), plot.original = 1000)
with(temp_test, points(dt, y, col = "RED"))
}
})

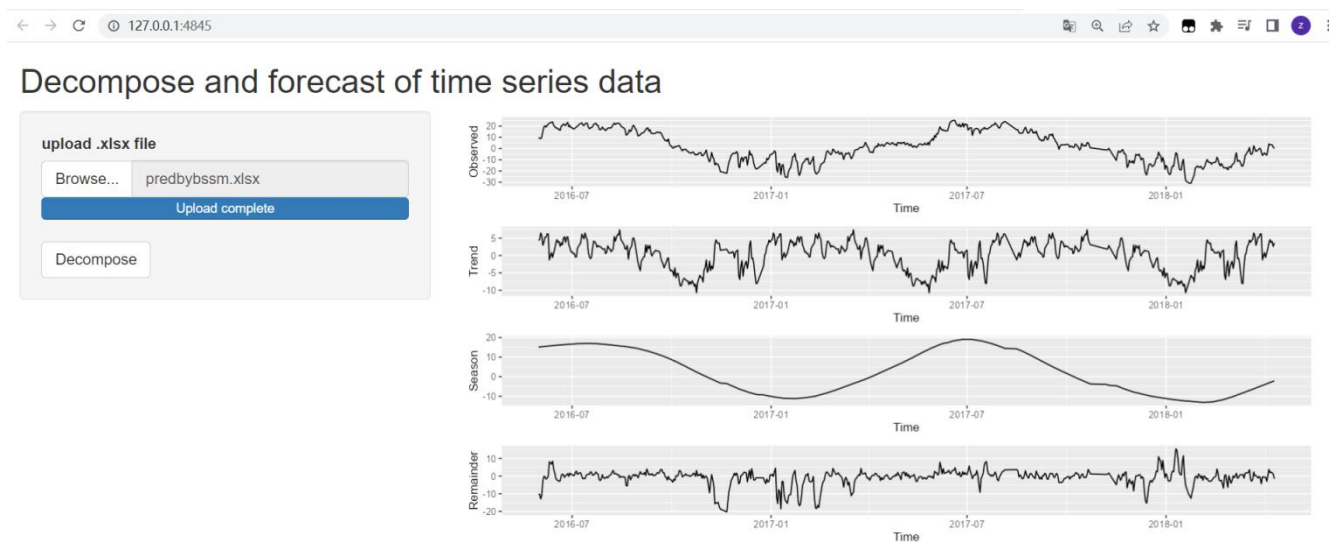
output$showmape <- renderText({
  if (!is.null(showmape_df())) {
    paste("MAPE:", round(showmape_df(), 2))
  }
})
}

shinyApp(ui, server)

```

Данный приложение позволяет пользователю по порядку загрузить файл, совершить декомпозицию временного ряда, разделить данные на обучающие и тестовые датасеты, установить период прогноза, визуализировать результат прогноза и получить значение MAPE.

Готовый веб-интерфейс показан на Рис. 4-13.



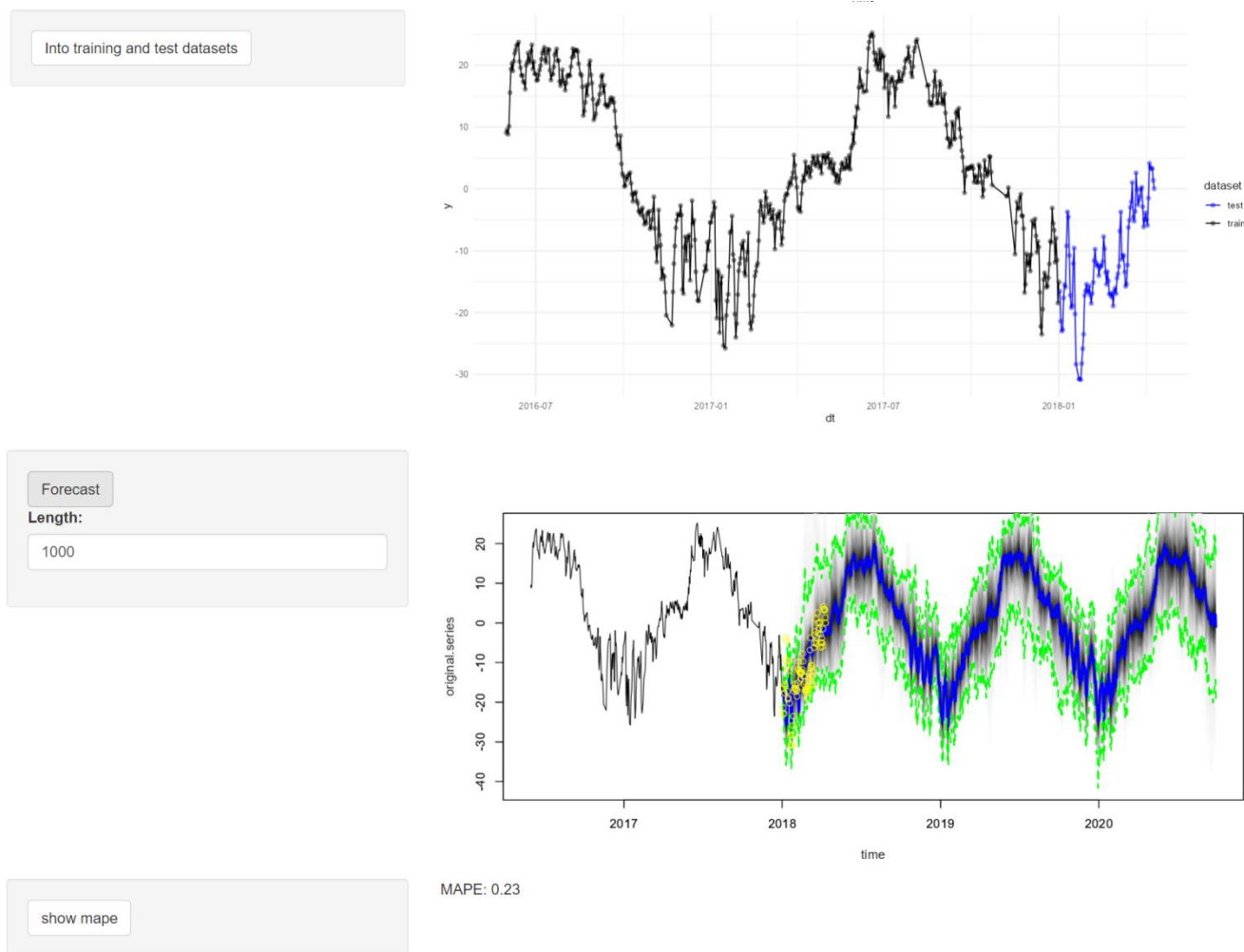


Рис. 4-13. Веб-интерфейс разработанного программного обеспечения

4.4. Выводы по Главе 4

Представлены результаты разработки прогнозной модели и программных экспериментов по исследованию данных, представленных в виде структурных временных рядов. Эксперименты проводились с реальными данными, полученными с ультразвуковой метеостанции ИМКЭС СО РАН.

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ»

Студенту:

Группа	ФИО
8ВМ11	Чжу Чжунвэнь

Школа	Инженерная школа информационных технологий и робототехники	Отделение школы (НОЦ)	Отделение информационных технологий
Уровень образования	Магистратура	Направление/специальность	09.04.01 Информатика и вычислительная техника

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

1. <i>Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих</i>	Среднерыночные цены РФ для определения стоимости материальных ресурсов. Нормативные документы НИ ТПУ, ФЗ «О минимальном размере оплаты труда» для определения оплаты труда исполнителей проекта.
2. <i>Нормы и нормативы расходования ресурсов</i>	Тариф электроэнергии 3,16 руб. кВт/ч., Премияльный коэффициент 30%; Доплаты и надбавки руководителя 20%; Дополнительная заработная плата 12%; Накладные расходы 16%; Районный коэффициент 30;.
3. <i>Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования</i>	Отчисления в социальные внебюджетные фонды 30 %.

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. <i>Оценка коммерческого и инновационного потенциала НТИ</i>	Провести предпроектный анализ
2. <i>Разработка устава научно-технического проекта</i>	Представить Устав научного проекта магистерской работы
3. <i>Планирование процесса управления НТИ: структура и график проведения, бюджет, риски и организация закупок</i>	Разработать план управления НТИ
4. <i>Определение ресурсной, финансовой, экономической эффективности</i>	Рассчитать сравнительную эффективность исследования

Перечень графического материала (с точным указанием обязательных чертежей):

1. Оценочная карта для сравнения конкурентных технических решений. 2. Матрица SWOT-анализа 3. Перечень этапов, работ и распределение исполнителей 4. Временные показатели проведения НИ 5. Бюджет НИ 6. Оценка характеристик вариантов исполнения 7. Сравнительная эффективность разработки.

Дата выдачи задания для раздела по линейному графику	01.03.2023 г.
---	---------------

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
доцент ОСГН, ШБИП	Былкова Татьяна Васильевна	к.э.н.		01.03.2023

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ВМ11	Чжу Чжунвэнь		01.03.2023

5. Финансовый менеджмент, ресурсоэффективность и ресурсосбережение

5.1. Оценка коммерческого и инновационного потенциала НИИ

5.1.1. Потенциальные потребители результатов исследования

Временной ряд представляет собой последовательность данных, описывающих объект в последовательные моменты времени. Существует две основные цели анализа временных рядов: определение природы ряда и прогнозирование, т.е. предсказание будущих значений временного ряда по настоящим и прошлым значениям. Исходя из этого можно сделать вывод, что к потенциальным потребителям можно отнести такие области как: астрономия – графики солнечной активности, экономика – изменения уровня безработицы или процентных ставок, метеорология – прогнозирование погодных условий, а также к планированию пропускных способностей, вычислительные системы.

		Метод прогнозирования			
		Метод д врем. рядов	Экспоненциально е сглаживание	Регрессионны е методы	Нейросетевы е методы
Отрасль	Астрономические обсерватории				
	Консалтинговые компании				
	Метеорологически е компании				

Рис. 5.1. Карта сегментирования рынка услуг по методам прогнозирования



5.1.2. Анализ конкурентных технических решений

На сегодняшний день существует множество конкурентных моделей прогнозирования: регрессионные, авторегрессионные, нейросетевые, экспоненциального сглаживания, цепи Маркова, классификационные модели и другие. Самыми очевидными конкурентными методами прогнозирования в данном исследовании являются: метод экспоненциального сглаживания и регрессивный метод. Приведем оценочную карту для сравнения конкурентных методов прогнозирования. Позиция разработки и конкурентов оценивается по каждому показателю экспертным путем по пятибалльной шкале, где 1 – слабая позиция, а 5 – сильная. Веса показателей в сумме должны составлять 1.

Таблица 5.1 – Оценочная карта для сравнения конкурентных методов прогнозирования

Критерии оценки	Вес критерия	Баллы			Конкуренто-способность		
		Б _ф	Б _{к1}	Б _{к2}	К _ф	К _{к1}	К _{к2}
1	2	3	4	5	6	7	8
Технические критерии оценки ресурсоэффективности							
1. Точность модели	0,3	5	4	4	1,5	1,2	1,2
2. Сложность модели	0,17	4	3	4	0,68	0,51	0,68
3. Гибкость модели	0,13	4	4	3	0,52	0,52	0,39
Экономические критерии оценки эффективности							
1. Уровень проникновения на рынок	0,12	3	4	3	0,36	0,48	0,36
2. Цена	0,15	4	3	4	0,6	0,45	0,6
3. Предполагаемый срок реализации	0,13	4	4	4	0,52	0,52	0,52
Итого	1	24	22	22	4,18	3,68	3,75

Анализ конкурентных технических решений определяется по формуле:

$$K = \sum B_i * B_i$$

Где K – конкурентоспособность научной разработки или конкурента;

B_i – вес показателя (в долях единицы);

B_i – балл i -го показателя.

Коэффициент конкурентоспособности предприятия:

$$K_k = \frac{K_\phi}{K_{\text{ср}}} = \frac{4,18}{(3,68+3,75)/2} = 1,12$$

так как $K_k > 1$, следовательно, предприятие конкурентоспособно.

5.1.3. FAST-анализ

Проведение FAST-анализа предполагает шесть стадий:

1. Выбор объекта FAST-анализа;
2. Описание главной, основных и вспомогательных функций, выполняемых объектом;
3. Определение значимости выполняемых функций объектом;
4. Анализ стоимости функций выполняемых объектом исследования;
5. Построение функционально-стоимостной диаграммы объекта и ее анализ;
6. Оптимизация функций выполняемых объектом.

Стадия 1. Выбор объекта FAST-анализа;

В рамках магистерской диссертации в качестве объекта FAST-анализа выступает объект исследования — данные временных рядов.

Стадия 2. Описание главной, основных и вспомогательных функций, выполняемых объектом;

Главная функция — записывать изменения температур.

Основные функции — хранить данные.

Вспомогательные функции — прогнозировать дальнейшие или проанализировать имеющиеся значения.

Таблица 5.2 – Классификация функций, выполняемых объектом исследования

Выполняемая функция	Ранг функции		
	Главная	Основная	Вспомогательная
1. Записывать изменения температур.	X		

2.Хранять данные		X	
3.Прогнозировать дальнейшие значения.			X
4.Проанализировать имеющие значения.			X
5.Выявление аномалий.			X

Стадия 3. Определение значимости выполняемых функций объектом;

Таблица 5.3 – матрица смежности

	Функция 1	Функция 2	Функция 3	Функция 4	Функция 5
Функция 1	=	<	<	<	<
Функция 2	>	=	>	>	>
Функция 3	>	<	=	<	<
Функция 4	>	<	>	=	<
Функция 5	>	<	>	>	=

Таблица 5.4 – матрица количественных соотношений функций

	Функция 1	Функция 2	Функция 2	Функция 3	Функция 4	ИТОГО
Функция 1	1	0,5	0,5	0,5	0,5	3
Функция 2	1,5	1	1,5	1,5	1,5	7
Функция 3	1,5	0,5	1	0,5	0,5	4
Функция 4	1,5	0,5	1,5	1	0,5	5

Функция 5	1,5	0,5	1,5	1,5	1	6
						25

Стадия 4. Анализ стоимости функций выполняемых объектом исследования;

Таблица 5.5 – определение стоимости функций, выполняемых объектом исследования

Выполняемая функция	Себестоимость, руб.
1.Записывать изменения температур.	500
2.Хранять данные	300
3.Прогнозировать дальнейшие значения.	800
4.Проанализировать имеющиеся значения.	1000
5.Выявление аномалий.	650

Стадия 5. Построение функционально-стоимостной диаграммы объекта и ее анализ;

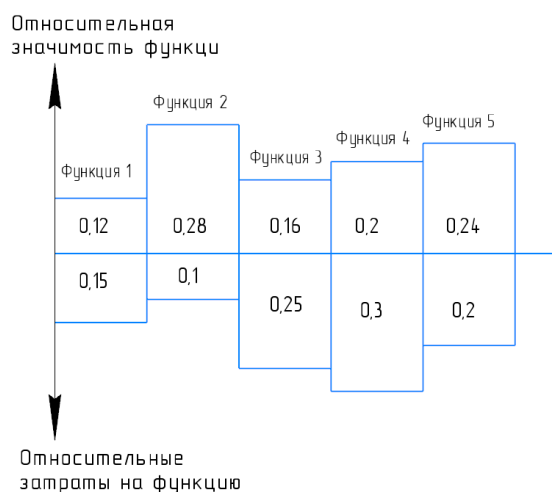


Рис.5.2. Функционально-стоимостная диаграмма

Анализ приведенной диаграммы показывает явное наличие рассогласования по функции 2. За счет того, что функция 2 занимает важное место, но затрат на нее мало, оптимизация функций не нужна.

5.1.4. SWOT-анализ

Слабые и сильные стороны, возможности и угрозы представлены в табличном виде (таблица 5.6).

Таблица 5.6 – Матрица SWOT-анализа

	<p>Сильные стороны научно-исследовательского проекта:</p> <ol style="list-style-type: none"> 1. Высокая точность прогнозирования. 2. Низкие затраты на реализацию. 3. Проект применяется в различных областях. 4. Автоматизация анализа временных рядов. 	<p>Слабые стороны научно-исследовательского проекта:</p> <ol style="list-style-type: none"> 1. Необходимость дополнительного обучения кадров для работы с программным продуктом. 2. Отсутствие репутации на рынке.
<p>Возможности:</p> <ol style="list-style-type: none"> 1. Новые отрасли применения. 2. Новые алгоритмы. 	<p>Новые приложения увеличат финансирование разработки. Заинтересованность бизнеса помогает понять конечные требования к продукту.</p>	<p>Новые области применения технологий помогут найти область, в которой требования могут быть легко выполнены.</p>

<p>3. Интерес ведущих университетов и организаций.</p> <p>4. Повышение стоимости конкурентных разработок.</p>		
<p>Угрозы:</p> <p>1. Малый спрос на разработанный проект.</p> <p>2. Появление альтернативных систем и технологий.</p> <p>3. Ухудшение экономической ситуации.</p>	<p>Повышение производительности системы за счет оптимизации решений различных проблем и новых технологий может стать важным аргументом для перехода на новую систему.</p>	<p>Недостаточный спрос, решить который можно путем поиска новых областей применения и демонстрации преимуществ программных продуктов.</p>

По результатам SWOT-анализа можно сделать вывод, что при разработке системы задача оптимизации и повышения производительности должна быть поставлена на первое место, так как благодаря этому преимуществу системы обеспечивается защита от угроз и предоставляются возможности для расширения такой системы.

5.1.5. Оценка готовности проекта к коммерциализации

Перечень вопросов приведен в таблице 5.7.

Таблица 5.7 – Бланк оценки степени готовности научного проекта к коммерциализации

№ п/п	Наименование	Степень проработанности научного проекта	Уровень имеющихся знаний у разработчика
1	Определен имеющийся научно-технический задел	4	4
2	Определены перспективные направления коммерциализации научно-технического задела	4	3

3	Определены отрасли и технологии для предложения на рынке	4	3
4	Определена товарная форма научно-технического задела для представления на рынок	3	3
5	Определены авторы и осуществлена охрана их прав	3	3
6	Проведена оценка стоимости интеллектуальной собственности	3	3
7	Проведены маркетинговые исследования рынков сбыта	4	4
8	Разработан бизнес-план коммерциализации научной разработки	4	3
9	Определены пути продвижения научной разработки на рынок	4	4
10	Разработана стратегия (форма) реализации научной разработки	4	4
11	Проработаны вопросы международного сотрудничества и выхода на зарубежный рынок	2	2
12	Проработаны вопросы использования услуг инфраструктуры поддержки, получения льгот	3	3
13	Проработаны вопросы финансирования коммерциализации научной разработки	4	3
14	Имеется команда для коммерциализации научной разработки	3	2

15	Проработан механизм реализации научного проекта	4	4
	ИТОГО БАЛЛОВ	53	48

Анализируя данные, представленные в таблице 10, можно заключить, что итоговые значения по обоим критериям попадают в диапазон от 45 до 59. Это указывает на то, что перспективность проекта и уровень знаний разработчика превышают среднее значение, следовательно, разработка имеет перспективы для получения инвестиций. Для улучшения показателей перспективности проекта можно привлечь специализированный персонал, который будет заниматься реализацией и коммерциализацией научной разработки. Например, имеет смысл воспользоваться услугами бизнес-консультантов или экспертов по экспорту и импорту. Они занимаются разработкой и реализацией стратегий для компаний, которые хотят расширить свою деятельность за пределы своей страны. Эти специалисты обладают знаниями в области международного бизнеса, торговых принципов, международного права, маркетинга и логистики, и помогают организациям успешно проникнуть на зарубежные рынки и наладить международные партнерства.

5.1.6. Методы коммерциализации результатов научно-технического исследования

Наиболее перспективным и подходящим методом коммерциализации результатов исследования, полученных в ходе выполнения данной работы, является передача интеллектуальной собственности в уставной капитал предприятия. К преимуществам способа можно отнести невысокий уровень затрат, маловероятность рисков, быструю окупаемость. Среди недостатков основными считаются: высокие затраты на поиск и привлечение покупателя прав, затраты на юридические консультации, риск недополучения значительных доходов от использования разработки в будущем.

5.2. Инициация проекта

Устав научного проекта магистерской работы:

1. Цели и результат проекта.

Были определены заинтересованные стороны (таблица 5.8).

Таблица 5.8 – Заинтересованные стороны проекта

Заинтересованные стороны проекта	Ожидания заинтересованных сторон
Разработчик системы	Получение знаний по специальности, пополнение портфолио, получение материальной выгоды
НИ ТПУ	Увеличение числа научных публикаций, дипломов на научно-практических конференциях.
Компании, желающие повысить эффективность прогнозирования временных рядов	Увеличение надежности, точности результатов анализа данных.

В таблице 5.9 представим цель проекта

Таблица 5.9 – Цели и результат проекта

Цели проекта	Исследование применимости BSTS-модели для оценивания и прогнозирования параметров температурных временных рядов.
Ожидаемые результаты проекта	Разработка и оценка метода прогнозирования на основании BSTS-модели и выполнение ПО.

2. Организационная структура проекта.

В таблице 5.10 отражена организационная структура, роль и функции каждого члена команды.

Таблица 5.10 – Рабочая группа проекта

№ п/п	ФИО, основное место работы, должность	Роль в проекте	Функции	Трудозатраты, час
1	Ботыгин Игорь Александрович, Томский	Научный	Координация участников проекта,	30

	политехнический университет, доцент	руководитель	консультирование по вопросам проекта, предоставление необходимых материалов для реализации проекта	
2	Чжу Чжунвэнь, Томский политехнический университет, магистр	Исполнитель (Инженер)	Выполнение отдельных видов работ по проекту	530
ИТОГО:				560

5.3. Планирование управления научно-техническим проектом

5.3.1. План проекта

В рамках планирования научного проекта необходимо построить календарный график проекта.

Линейный график представляется в виде таблице 5.11.

Таблица 5.11 – Календарный план проекта

№	Название	Длительность, дни	Участники
1	Составление задания	1	Руководитель Студент
2	Сбор исходных материалов	10	Студент
3	Разработка функциональной структуры программной системы	15	Студент
4	Реализация конкретных моделей прогнозирования	8	Студент
5	Проверка результата	4	Студент

	прогнозирования		
6	Оформление ВКР	8	Студент
7	Утверждение ВКР руководителем	1	Руководитель Студент

График-ганта представляется в виде таблице 5.12.

Таблица 5.12 – график Ганнта

№	Название	Т _к , кол, дни	Участники	Продолжительность выполнения работ																
				март		апрель			май			Июнь								
				2	3	1	2	3	1	2	3	1								
1	Составление задания	1	Руководитель Студент	■																
2	Сбор исходных материалов	15	Студент		■															
3	Разработка функциональной структуры программной системы	25	Студент			■														
4	Реализация конкретных моделей прогнозирования	12	Студент					■												
5	Проверка результата прогнозирования	8	Студент							■										
6	Оформление ВКР	15	Студент								■									

7	Утверждение ВКР руководителем	1	Руководитель Студент											
---	-------------------------------------	---	-----------------------------	--	--	--	--	--	--	--	--	--	--	--

■ - студент ■ - научный руководитель

5.3.2. Бюджет научного исследования

Таблица 5.13 – Группировка затрат по статьям

№	Наименование статьи	Сумма, руб.		
		Исп.1	Исп.2	Исп.3
1	Затраты на материалы	1691	1691	1691
2	Затраты на специальное оборудование	50000	60000	70000
3	Затраты по основной заработной плате исполнителей темы	102400	102400	102400
4	Отчисления во внебюджетные фонды	30720,07	30720,07	30720,07
5	Накладные расходы	29570	31170	32770
	Итого	214381	225981	237581

Сырьем и материалами в данном случае являются оплата за электричество и интернет. Тарифный план за свет 3,16 рубля за кВт, по приблизительным усредненным показателям, компьютер потребляет около 800 Вт. Оплата за интернет составляет 350 рублей. Траты на сырье и материалы за год составят 1691 рубль.

Таблица 5.14 – расчет затрат по статье «Спецоборудование для научных работ»

Наименование оборудования	Кол-во, ед.	Общая стоимость, тыс. руб		
		Исп.1	Исп.1	Исп.1

Программное обеспечение R-studio	1	0	10	20
Персональный компьютер	1	50	50	50
Итого		50	60	70

Основной заработной платой является вознаграждение за выполненную работу в соответствии с установленными нормами труда.

Статья включает сумму основной заработной платы и дополнительной заработной платы работников, которые непосредственно заняты выполнением НИИ:

$$Z_{\text{общ}} = Z_{\text{осн}} + Z_{\text{доп}}$$

где $Z_{\text{осн}}$ – основная зарплата, $Z_{\text{доп}}$ - дополнительная зарплата.

Расчет месячной зарплаты для руководителя:

$$Z_{\text{м}} = Z_{\text{тс}} * (1 + k_{\text{пр}} + k_{\text{д}}) * k_{\text{р}} = 40 * (1 + 0,3 + 0,2) * 1,3 = 78 \text{ тыс.руб}$$

где $Z_{\text{тс}}$ – заработная плата по тарифной ставке, 40 тыс. руб. (заработная плата руководителя);

$k_{\text{пр}}$ – премиальный коэффициент, равный 0,3;

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

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

Основная заработная плата руководителя ($Z_{\text{осн}}$):

$$Z_{\text{осн}} = Z_{\text{дн}} * T_{\text{р}}$$

где,

$Z_{\text{осн}}$ – основная заработная плата одного работника;

$T_{\text{р}}$ – продолжительность работ выполняемых научно-техническим работником, рабочие дни;

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

Среднедневная заработная плата руководителя:

$$Z_{\text{дн}} = (Z_{\text{м}} * M) / F_{\text{д}} = (78 * 10,4) / 243 = 3,4 \text{ тыс.руб}$$

где, $Z_{\text{м}}$ – месячный должностной оклад работника, руб.;

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

F_d – действительный годовой фонд рабочего времени научнотехнического персонала, рабочие дни.

Величина отчислений во внебюджетные фонды определяется исходя из следующей формулы:

$$Z_{вн} = k_{вн} * Z_{осн} = 0,30 * (102400) = 30720,07 \text{ руб.}$$

где $k_{вн}$ – коэффициент отчислений на уплату во внебюджетные фонды (пенсионный фонд, фонд обязательного медицинского страхования и пр.), 30 %.

Таблица 5.15 – Расчет основной зарплаты и отчислений в социальные внебюджетные фонды

Исполнитель	$Z_{тс}$,руб	$K_{пр}$	K_d	K_p	$Z_{м}$,руб	$Z_{дн}$,руб	T_p ,раб дни	$Z_{осн}$,руб	$Z_{вн}$, руб.
Руководитель	39300	0,3	0,2	1	70740	3028	2	6055	1816,533
Студент	16242	0,3	0,2	1	29235,6	1251	77	96345	28903,54
								102400	30720,07

5.3.3. Определение ресурсной, финансовой эффективности исследования

Интегральный финансовый показатель проекта рассчитывается как:

$$I_{финр}^{исп.i} = \frac{\Phi_{pi}}{\Phi_{max}}$$

где $I_{финр}^{исп.i}$ – интегральный финансовый показатель разработки; Φ_{pi} – стоимость i-го варианта исполнения; Φ_{max} – максимальная стоимость исполнения.

Таблица 5.16 – Расчёт интегрального финансового показателя

Вариант исполнения	Φ_{pi}	Φ_{max}	$I_{финр}^{исп.i}$
1	214381	237581	0,90
2	225981		0,95
3	237581		1

Интегральный показатель ресурсоэффективности вариантов исполнения объекта исследования можно определить следующим образом:

$$I_{pi} = \sum a_i \cdot b_i.$$

Где I_{pi} – интегральный показатель ресурсоэффективности для i -го варианта исполнения разработки;

a_i – весовой коэффициент i -го варианта исполнения разработки;

b_i – бальная оценка i -го варианта исполнения разработки,

устанавливается экспертным путем по выбранной шкале оценивания;

n – число параметров сравнения.

Расчёт показателя приведён в таблице 20.

Таблица 5.17 - Сравнительная оценка характеристик вариантов исполнения проекта

Критерии Объект	Весовой Коэффициент параметра	Текущий проект	Исп.2	Исп.3
1. Способствует сокращению времени на первичный анализ	0,2	5	3	5
2. Простота использования	0,15	3	3	3
3. Надежность	0,15	4	3	3
4. Воспроизво- димость	0,15	4	4	4
5. Ресурсо- эффективность	0,15	4	4	4
6. Конкурен- тоспособность	0,2	4	3	4
Итого	1	4.05	3.3	3.9

Интегральный показатель эффективности вариантов исполнения разработки ($I_{исп.i}$) определяется на основании интегрального показателя ресурсоэффективности и интегрального финансового показателя по формуле:

$$I_{исп.1} = \frac{I_{p-исп1}}{I_{финр}^{исп.1}}, \quad I_{исп.1} = \frac{I_{p-исп1}}{I_{финр}^{исп.1}} \text{ и т.д.}$$

Сравнение интегрального показателя эффективности вариантов исполнения разработки позволит определить сравнительную эффективность проекта (таблица 21) и выбрать наиболее целесообразный вариант из предложенных. Сравнительную эффективность проекта определим по следующей формуле:

$$\mathcal{E}_{cp} = \frac{I_{исп.1}}{I_{исп.2}}$$

Таблица 5.18 - Сравнительная эффективность разработки.

№ П/П	Показатели	Текущий проект	Исп.2	Исп.3
1	Интегральный финансовый показатель разработки ($I_{финр}^{исп.i}$)	0,90	0,95	1
2	Интегральный показатель ресурсоэффективности разработки ($I_{p-исп}$)	4.05	3.3	3.9
3	Интегральный показатель эффективности ($I_{испi}$)	4,5	3,47	3,9
4	Сравнительная эффективность вариантов исполнения (\mathcal{E}_{cp})	1	1,29	1,15

В результате проведения разработки раздела финансовый менеджмент, ресурсоэффективность и ресурсосбережение было выяснено, что данный проект оправдывает физические и материальные затраты.

Детальный анализ конкурирующих разработок, существующих на рынке, показал, что коэффициент конкурентоспособности (K_K) предприятия равен 1,12. Так как $K_K > 1$, следовательно, предприятие конкурентоспособно.

На разработку данного проекта потребовалось 46 дн., для подробного описания трудозатрат был представлен календарный план проекта.

Оценка сравнительной эффективности показала, что разработка в первом варианте исполнения эффективнее второго и третьего варианта исполнения на 29% и 15%.

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа	ФИО
8BM11	Чжу Чжунвэнь

Тема магистерской диссертации:

«Оценка параметров структурных моделей временных рядов на основе принципов байесовской статистики»

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

Исходные данные к разделу «Социальная ответственность»

<p>1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения</p>	<p>Объект исследования: данные временных рядов. Область применения: финансовые сферы, метеорология. Рабочее место: офис с персональным компьютером. Количество и наименование оборудования рабочей зоны: персональный компьютер.</p>
---	--


Перечень вопросов, подлежащих исследованию, проектированию и разработке:

<p>1. Производственная безопасность:</p> <p>1.1. Анализ выявленных вредных факторов</p> <ul style="list-style-type: none"> • Природа воздействия • Действие на организм человека • Нормы воздействия и нормативные документы (для вредных факторов) • СИЗ коллективные и индивидуальные <p>1.2. Анализ выявленных опасных факторов</p> <ul style="list-style-type: none"> • Термические источники опасности • Электробезопасность • Пожаробезопасности 	<p>1. Вредные факторы:</p> <p>1.1. Недостаточная освещенность;</p> <p>1.2. Нарушения микроклимата, оптимальные и допустимые параметры;</p> <p>1.2. Шум, ПДУ, СКЗ, СИЗ;</p> <p>1.4. Повышенный уровень электромагнитного излучения, ПДУ, СКЗ, СИЗ;</p> <p>2. Опасные факторы:</p> <p>2.1. Электроопасность; класс электроопасности помещения, безопасные номиналы I, U, R_{заземления}, СКЗ, СИЗ;</p> <p>2.2. Пожароопасность, категория пожароопасности помещения, марки огнетушителей, их назначение и ограничение применения; Приведена схема эвакуации.</p>
<p>2. Экологическая безопасность:</p> <ul style="list-style-type: none"> • Выбросы в окружающую среду • Решения по обеспечению экологической безопасности 	<p>Наличие промышленных отходов (перегоревшие люминесцентные лампы) и способы их утилизации.</p>

<p>3. Безопасность в чрезвычайных ситуациях:</p> <p>1. Перечень возможных ЧС при разработке и эксплуатации проектируемого решения</p> <p>2. Разработка превентивных мер по предупреждению ЧС</p> <p>3. Разработка действий в результате возникшей ЧС и мер по ликвидации её последствий</p>	<p>Рассмотреть 2 ситуации ЧС:</p> <p>1) Природная – сильные морозы зимой, (аварии на электро-, тепло-коммуникациях, водоканале, транспорте).</p> <p>2) Техногенная – несанкционированное проникновение посторонних на рабочее место (возможны проявления вандализма, диверсии, промышленного шпионажа).</p> <p>3) Представить мероприятия по обеспечению устойчивой работы производства в том и другом случае.</p>
<p>4. Перечень нормативно-технической документации</p>	<p>– ГОСТы, СанПиНы, СНиПы</p>

Дата выдачи задания для раздела по линейному графику	16.02.2023
--	------------

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Профессор ТПУ	Федорчук Ю.М.	д.т.н.		16.03.2023

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ВМ11	Чжу Чжунвэнь		16.03.2023

6. Социальная ответственность

Введение

Социальная ответственность – ответственность отдельного ученого и научного сообщества перед обществом. Первостепенное значение при этом имеет безопасность применения технологий, которые создаются на основе достижений науки, предотвращение или минимизация возможных негативных последствий их применения, обеспечение безопасного как для испытуемых, как и для окружающей среды проведения исследований.

Разработанный проект в рамках магистерской диссертации представляет собой программно-вычислительную систему, направленную на построение моделей для прогнозирования температурных величин, тем самым повышая предсказуемость этих параметров во времени. Работа выполнялась в лаборатории кибернетического центра ТПУ с использованием компьютера. Раздел также включает в себя оценку условий труда на рабочем месте, анализ вредных и опасных факторов труда, разработку мер защиты от них.

6.1. Производственная безопасность

6.1.1. Вредные производственные факторы

6.1.1.1. Недостаточная освещенность рабочей зоны

Для обеспечения требуемой освещенности необходимо использовать совмещенное освещение, создаваемое сочетанием естественного и искусственного освещения. При данном этапе развития осветительной техники целесообразно использовать люминесцентные лампы, которые по сравнению с лампами накаливания имеют большую светоотдачу на ватт потребляемой мощности и более естественный спектр.

Минимальный уровень средней освещенности на рабочих местах с постоянным пребыванием людей должен быть не менее 200 лк.

В расчётном задании должны быть решены следующие вопросы:

- выбор системы освещения;
- выбор источников света;

- выбор светильников и их размещение;
- выбор нормируемой освещённости;
- расчёт освещения методом светового потока.

В данном расчётном задании для всех помещений рассчитывается общее равномерное освещение.

Таблица 6.1. Параметры помещения

Параметр	Обозначение	Значение, м
Длина	A	7
Ширина	B	6
Высота помещения	H	3,5
Свес	h _с	0,4
Высота Р.П.	h _{рп}	0,8
Высота от светильника до Р.П.	h	H- h _{рп} - h _с
Коэффициент отражения стен	ρ _{ст}	70 %
Коэффициент отражения потолка	ρ _п	70
Коэффициент запаса	K _з	1.5
Коэффициент неравномерности	Z	1.1

Расчёт общего равномерного искусственного освещения горизонтальной рабочей поверхности выполняется методом коэффициента светового потока, учитывающим световой поток, отражённый от потолка и стен.

Световой поток лампы определяется по формуле:

$$\Phi_{\text{рас}} = \frac{E_{\text{н}} * S * K_{\text{з}} * Z}{N * \eta}$$

Где $E_{\text{н}}$ – нормируемая минимальная освещённость по СНиП 23-05-95, лк; S – площадь освещаемого помещения, м²; $K_{\text{з}}$ – коэффициент запаса, учитывающий загрязнение светильника (источника света, свето-технической арматуры, стен и пр., т. е. отражающих поверхностей), наличие в атмосфере цеха дыма, пыли; Z – коэффициент неравномерности освещения, отношение $E_{\text{ср}}/E_{\text{min}}$. Для люминесцентных ламп при расчётах берётся равным 1,1; N – число ламп в помещении; η – коэффициент использования светового потока.

Коэффициент использования светового потока показывает, какая часть светового потока ламп попадает на рабочую поверхность. Он зависит от индекса помещения i , типа светильника, высоты светильников над рабочей поверхностью h и коэффициентов отражения стен ρ_c и потолка ρ_n .

Индекс помещения определяется по формуле:

$$i = S / h(A+B)$$

Проведем расчет индекса помещения:

Площадь помещения :

$$S = A * B = 7 * 6 = 42 \text{ m}^2$$

Индекс:

$$i = \frac{S}{h * (A + B)} = \frac{42}{(3.5 - 0.8 - 0.4) * (8 + 7)} = 1.4$$

Согласно этим данным коэффициент использования светового потока будет равен 48 % или в долях = 0,48.

Коэффициенты отражения оцениваются субъективно (табл. 4.10) [БЖД Практикум 2009-2020].

Согласно указанной методике выбираем тип источника света.

Наиболее подходящим вариантом является 40 ваттная лампа ЛБ, у которой $\Phi=2800$ лм. Для выбранного типа лампы подходит светильник ОД-2-40 с размерами: длина = 1230 мм, ширина = 266 мм.

Количество ламп для помещения:

$$N = \frac{E_n * S * K_z * Z}{\Phi * \eta} = \frac{200 * 42 * 1.5 * 1.1}{2800 * 0.48} = 10.3$$

Принимаем $N=12$ лампы или 6 светильников.

Размещаем светильники в 2 ряда по 3 светильников в ряду с соблюдением условий: L – расстояние между соседними светильниками или рядами (если по длине (А) и ширине (В) помещения расстояния различны, то они обозначаются L_A и L_B), l – расстояние от крайних светильников или рядов до стены.

Оптимальное расстояние l от крайнего ряда светильников до стены рекомендуется принимать равным $L/3$.

Сначала определим световой поток расчетный.

$$\Phi = \frac{E_H * S * K_3 * Z}{N * \eta} = \frac{200 * 42 * 1.5 * 1.1}{12 * 0.48} = 2406 \text{ лм};$$

Теперь определим расстояния между светильниками по длине помещения.

$$7000 = 2 * L_A + 3 * 1230 + 2/3 * L_A; L_A = (7000 - 3690) * 3/8 = 1241 \text{ мм};$$

$$L_A / 3 = 414 \text{ мм};$$

$$6000 = L_B + 2 * 266 + 2/3 * L_B; L_B = (6000 - 532) * 3/5 = 3281 \text{ мм};$$

$$L_B / 3 = 1094 \text{ мм};$$

Рисуем схему размещения светильников на потолке для обеспечения общего равномерного освещения.

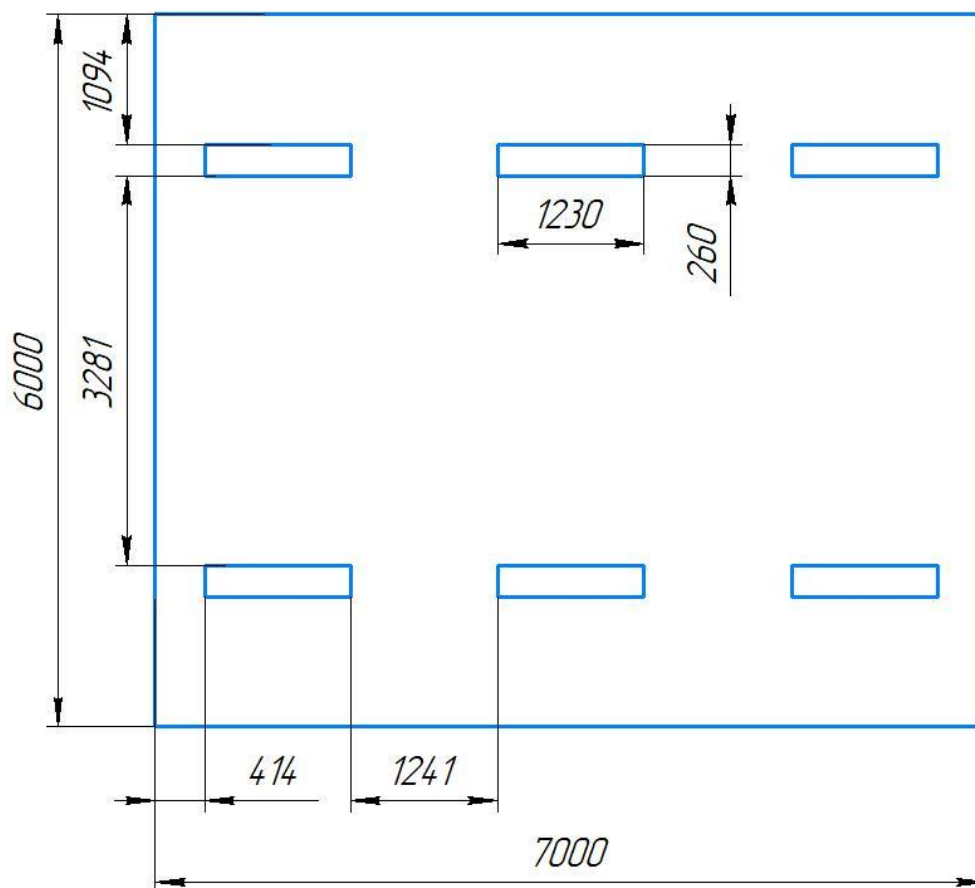


Рис. 6.1. План размещения светильников на потолке.

Проведем проверку выполнения условия соответствия:

$$-10\% \leq \frac{\Phi_{\text{л.станд}} - \Phi_{\text{л.расч}}}{\Phi_{\text{л.станд}}} * 100\% \leq +20\%$$

$$-10\% \leq \frac{2800 - 2406}{2800} * 100\% \leq +20\%$$

$$-10\% \leq +14\% \leq +20\%$$

Результат расчета укладывается в поле допуска.

Определим мощность осветительной установки:

$$P = N * P_i = 12 * 40 \text{ Вт} = 480 \text{ Вт}$$

6.1.1.2. Отклонение показателей микроклимата в помещении

Проанализируем микроклимат в помещении, где находится рабочее место. Микроклимат производственных помещений определяют следующие параметры: температура, относительная влажность, скорость движения воздуха. Эти факторы влияют на организм человека, определяя его самочувствие.

Оптимальные параметры микроклимата на рабочих местах должны соответствовать величинам, приведенным в Табл. 6-2 и Табл. 6-3.

Табл. 6-2. Оптимальные нормы микроклимата

Период года	Температура воздуха, С°	Относительная влажность воздуха, %	Скорость движения воздуха, м/с
Холодный	19-23	40-60	0.1
Теплый	23-25		0.2

Табл. 6-3. Допустимые нормы микроклимата

Период года	Температура воздуха, С°		Относительная влажность воздуха, %	Скорость движения воздуха, м/с
	Нижняя допустимая граница	Верхняя допустимая граница		
Холодный	15	24	20-80	<0.5
Теплый	22	28	20-80	<0.5

Общая площадь рабочего помещения составляет 42м², объем составляет 147м³. По СанПиН 2.2.2/2.4.1340-03 санитарные нормы составляют 6,5 м² и 20 м³ объема на одного человека. Исходя из приведенных выше данных, можно сказать, что количество рабочих мест соответствует размерам помещения по санитарным нормам.

После анализа габаритных размеров рассмотрим микроклимат в этой комнате. В качестве параметров микроклимата рассмотрим температуру, влажность воздуха, скорость ветра.

В помещении осуществляется естественная вентиляция посредством наличия легко открываемого оконного проема (форточки), а также дверного проема. По зоне действия такая вентиляция является общеобменной. Основным недостатком - приточный воздух поступает в помещение без предварительной очистки и нагревания. Согласно нормам СанПиН 2.2.2/2.4.1340-03 объем воздуха необходимый на одного человека в помещении без дополнительной вентиляции должен быть более 40м³[1]. В нашем случае объем воздуха на одного человека составляет 42 м³, из этого следует, что дополнительная вентиляция не требуется. Параметры микроклимата поддерживаются в холодное время года за счет систем водяного отопления с нагревом воды до 100°С, а в теплое время года – за счет кондиционирования, с параметрами согласно [2]. Нормируемые параметры микроклимата, ионного состава воздуха, содержания вредных веществ должны соответствовать требованиям [3].

6.1.1.3. Превышение уровней шума

Одним из наиболее распространенных в производстве вредных факторов является шум. Он создается вентиляционным и рабочим оборудованием, преобразователями напряжения, рабочими лампами дневного света, а также проникает снаружи. Шум вызывает головную боль, усталость, бессонницу или сонливость, ослабляет внимание, память ухудшается, реакция уменьшается.

Основным источником шума в комнате являются компьютерные охлаждающие вентиляторы и. Уровень шума варьируется от 35 до 42 дБА. Согласно СанПиН 2.2.2 / 2.4.1340-03, при выполнении основных работ на ПЭВМ уровень шума на рабочем месте не должен превышать 80 дБА [4].

При значениях выше допустимого уровня необходимо предусмотреть средства индивидуальной защиты(СИЗ) и средства коллективной защиты (СКЗ) от шума.

Средства коллективной защиты:

1. устранение причин шума или существенное его ослабление в источнике образования;
2. изоляция источников шума от окружающей среды (применение глушителей, экранов, звукопоглощающих строительных материалов, например любой пористый материал – шамотный кирпич, микропористая резина, поролон и др.);
3. применение средств, снижающих шум и вибрацию на пути их распространения;

Средства индивидуальной защиты;

1. применение спецодежды и защитных средств органов слуха: наушники, беруши, антифоны.

6.1.1.4. Повышенный уровень электромагнитных излучений

Источником электромагнитных излучений в нашем случае являются дисплеи ПЭВМ. Монитор компьютера включает в себя излучения рентгеновской, ультрафиолетовой и инфракрасной области, а также широкий диапазон электромагнитных волн других частот.

Согласно СанПиН 2.2.2/2.4.1340-03 напряженность электромагнитного поля по электрической составляющей на расстоянии 50 см вокруг ВДТ не должна превышать 25В/м в диапазоне от 5Гц до 2кГц, 2,5В/м в диапазоне от 2 до 400кГц [1]. Плотность магнитного потока не должна превышать в диапазоне от 5 Гц до 2 кГц 250нТл, и 25нТл в диапазоне от 2 до 400кГц. Поверхностный электростатический потенциал не должен превышать 500В [1].

В ходе работы использовалась ПЭВМ типа ASUS GL703VD со следующими характеристиками: напряженность электромагнитного поля 3В/м; поверхностный потенциал составляет 560 В (основы противопожарной защиты предприятий ГОСТ 12.1.004 и ГОСТ 12.1.010 – 76.)[5].

При длительном постоянном воздействии электромагнитного поля (ЭМП) радиочастотного диапазона при работе на ПЭВМ у человеческого организма

сердечно-сосудистые, респираторные и нервные расстройства, головные боли, усталость, ухудшение состояния здоровья, гипотония, изменения сердечной мышцы проводимости. Тепловой эффект ЭМП характеризуется увеличением температуры тела, локальным селективным нагревом тканей, органов, клеток за счет перехода ЭМП на теплую энергию.

Предельно допустимые уровни (ПДУ) облучения (по ОСТ 54 30013-83):

- а) до 10 мкВт/см² , время работы (8 часов);
- б) от 10 до 100 мкВт/см² , время работы не более 2 часов;
- в) от 100 до 1000 мкВт/см² , время работы не более 20 мин. при условии пользования защитными очками;
- г) для населения в целом ППИМ не должен превышать 1 мкВт/см².

Защита человека от опасного воздействия электромагнитного излучения осуществляется следующими способами:

СКЗ

1. защита временем;
2. защита расстоянием;
3. снижение интенсивности излучения непосредственно в самом источнике излучения;
4. заземление экрана вокруг источника;
5. защита рабочего места от излучения;

СИЗ

1. Очки и специальная одежда, выполненная из металлизированной ткани (кольчуга). При этом следует отметить, что использование СИЗ возможно при кратковременных работах и является мерой аварийного характера. Ежедневная защита обслуживающего персонала должна обеспечиваться другими средствами.

2. Вместо обычных стекол используют стекла, покрытые тонким слоем золота или диоксида олова (SnO₂).

6.1.2. Опасные производственные факторы

6.1.2.1. Поражение электрическим током

К опасным факторам можно отнести наличие в помещении большого количества аппаратуры, использующей однофазный электрический ток напряжением 220 В и частотой 50Гц. По опасности электропоражения комната относится к помещениям без повышенной опасности, так как отсутствует повышенная влажность, высокая температура, токопроводящая пыль и возможность одновременного сприкосновения токоведущих элементов с заземленными металлическими корпусами оборудования [6].

Лаборатория относится к помещению без повышенной опасности поражения электрическим током. Безопасными номиналами являются: $I < 0,1 \text{ А}$; $U < (6-42) \text{ В}$; $R_{\text{зазем}} < 4 \text{ Ом}$.

Для защиты от поражения электрическим током используют СКЗ и СИЗ.

Средства коллективной защиты:

1. — защитное заземление, зануление;
2. — малое напряжение;
3. — электрическое разделение сетей;
4. — защитное отключение;
5. — изоляция токоведущих частей;
6. — оградительные устройства.

Использование щитов, барьеров, клеток, ширм, а также заземляющих и шунтирующих штанг, специальных знаков и плакатов.

Средства индивидуальной защиты:

1. — диэлектрические перчатки, изолирующие клещи и штанги;
2. — слесарные инструменты с изолированными рукоятками;
3. — указатели величины напряжения, калоши, боты, подставки и коврики.

6.1.2.2. Пожароопасность

По взрывопожарной и пожарной опасности помещения подразделяются на категории А, Б, В1-В4, Г и Д.

Согласно НПБ 105-03 лаборатория относится к категории В – горючие и трудно горючие жидкости, твердые горючие и трудно горючие вещества и материалы, вещества и материалы, способные при взаимодействии с водой, кислородом воздуха или друг с другом только гореть, при условии, что помещения, в которых находится, не относятся к категории наиболее опасных А или Б.

По степени огнестойкости данное помещение относится к 1-й степени огнестойкости по СНиП 2.01.02-85 (выполнено из кирпича, которое относится к трудносгораемым материалам).

Возникновение пожара при работе с электронной аппаратурой может быть по причинам как электрического, так и неэлектрического характера.

Причины возникновения пожара неэлектрического характера:

а) халатное неосторожное обращение с огнем (курение, оставленные без присмотра нагревательные приборы, использование открытого огня);

Причины возникновения пожара электрического характера: короткое замыкание, перегрузки по току, искрение и электрические дуги, статическое электричество и т. п.

Для локализации или ликвидации загорания на начальной стадии используются первичные средства пожаротушения. Первичные средства пожаротушения обычно применяют до прибытия пожарной команды.

Огнетушители водо-пенные (ОХВП-10) используют для тушения очагов пожара без наличия электроэнергии. Углекислотные (ОУ-2) и порошковые огнетушители предназначены для тушения электроустановок, находящихся под напряжением до 1000В. Для тушения токоведущих частей и электроустановок применяется переносной порошковый огнетушитель, например ОП-5.

В общественных зданиях и сооружениях на каждом этаже должно размещаться не менее двух переносных огнетушителей. Огнетушители следует располагать на видных местах вблизи от выходов из помещений на высоте не более 1,35 м. Размещение первичных средств пожаротушения в коридорах, переходах не должно препятствовать безопасной эвакуации людей.

Для предупреждения пожара и взрыва необходимо предусмотреть:

1. специальные изолированные помещения для хранения и разлива легковоспламеняющихся жидкостей (ЛВЖ), оборудованные приточно-вытяжной вентиляцией во взрывобезопасном исполнении - соответствии с ГОСТ 12.4.021-75 и СНиП 2.04.05-86;

2. специальные помещения (для хранения в таре пылеобразной канифоли), изолированные от нагревательных приборов и нагретых частей оборудования;

3. первичные средства пожаротушения на производственных участках (передвижные углекислые огнетушители ГОСТ 9230-77, пенные огнетушители ТУ 22-4720-80, ящики с песком, войлок, кошма или асбестовое полотно);

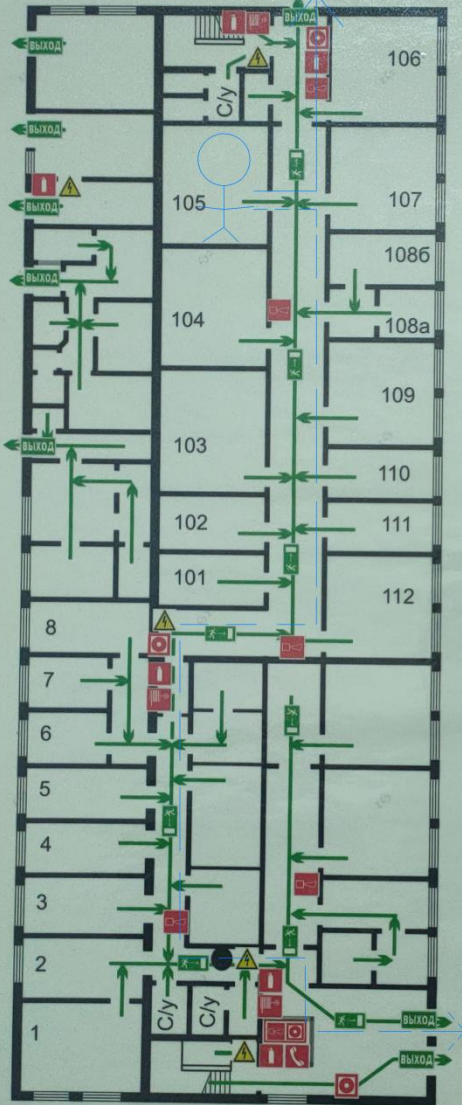
4. автоматические сигнализаторы (типа СВК-3 М 1) для сигнализации о присутствии в воздухе помещений предвзрывных концентраций горючих паров растворителей и их смесей.

Лаборатория полностью соответствует требованиям пожарной безопасности, а именно, наличие охранно-пожарной сигнализации, плана эвакуации, изображенного на рисунке 1, порошковых огнетушителей с поверенным клеймом, табличек с указанием направления к запасному (эвакуационному) выходу.

**ПЛАН ЭВАКУАЦИИ ПРИ ПОЖАРЕ И ДРУГИХ ЧС
из помещений учебного корпуса № 22,
ул. Советская, 84/3 - 1 этаж**



Утверждаю
Начальник УЭО
ФГАОУ ВО НИ ТПУ
П. Н. Ковылин



Действия при пожаре
Сохранять спокойствие!

- 1 Сообщить по телефону - адрес объекта
- место возникновения пожара
01, сотовый 101 - свою фамилию
- 2 Эвакуировать людей - ориентироваться по знакам направления движения
- взять с собой пострадавших
- 3 По возможности принять меры по тушению пожара - использовать средства противопожарной защиты
- при необходимости обесточить помещения

Действия при ЧС
Сохранять спокойствие!

- 1 Сообщить по телефону - адрес объекта
- что случилось
112 - имеются ли пострадавшие
- свою фамилию
- предотвратить развитие аварии
- 2 Локализовать аварию - использовать средства защиты
- обозначить место аварии
- 3 Эвакуировать людей - оказать помощь пострадавшим
 - ориентироваться по знакам
- взять с собой пострадавших

Задействовать систему оповещения о пожаре

УСЛОВНЫЕ ОБОЗНАЧЕНИЯ

- | | | | | | | |
|--|------------------------------------|--|--------------------------|--------------------------|--|--------------------------------------|
| | звуковой оповещатель | | пожарный кран | 01-112- номера помещений | | эл.щит |
| | огнетушитель | | эвакуационный выход | | | кнопка включения пожарной автоматики |
| | направление движения при эвакуации | | Вы находитесь здесь | | | блок АУПС |
| | путь к эвакуационному выходу | | путь к аварийному выходу | | | телефон |

ЗАО НИИ и НТЦ "Ириат-Т" г. Томск, ул. Яковлева, 29, тел. 66-09-01
Лицензия № 6-БД/1008 от 07 июня 2013 г. МЧС России

Изготовлено согласно ГОСТ Р 12.2.143-2009, ГОСТ 12.4.026-2015 206-25-1440-ЖЗ январь 2019 года

Рис. 6.2. План эвакуации. Исполнитель в аудитории 105, сплошные линии – основной выход, штриховые линии – запасный выход.

6.2. Экологическая безопасность

В компьютерах огромное количество компонентов, которые содержат токсичные вещества и представляют угрозу, как для человека, так и для окружающей среды.

К таким веществам относятся:

- свинец (накапливается в организме, поражая почки, нервную систему);
- ртуть (поражает мозг и нервную систему);
- никель и цинк (могут вызывать дерматит);
- щелочи (прожигают слизистые оболочки и кожу);

Поэтому компьютер требует специальных комплексных методов утилизации.

Таким образом утилизацию компьютера можно провести следующим образом:

- отделить металлические детали от неметаллов;
- разделить углеродистые металлы от цветмета;
- пластмассовые изделия (крупногабаритные) измельчить для уменьшения объема;
- копир-порошок упаковать в отдельную упаковку, точно также, как и все проклассифицированные и измельченные компоненты оргтехники, и после накопления на складе транспортных количеств отправить предприятиям и фирмам, специализирующимся по переработке отдельных видов материалов.

Люминесцентные лампы утилизируют следующим образом. Не работающие лампы немедленно после удаления из светильника должны быть упакованы в картонную коробку, бумагу или тонкий мягкий картон, предохраняющий лампы от взаимного соприкосновения и случайного механического повреждения. После накопления ламп объемом в 1 транспортную единицу их сдают на переработку на соответствующее предприятие. Недопустимо выбрасывать отработанные энергосберегающие лампы вместе с обычным мусором, превращая его в ртутьсодержащие отходы, которые загрязняют ртутными парами

6.3. Безопасность в чрезвычайных ситуациях

Природная чрезвычайная ситуация – обстановка на определенной территории или акватории, сложившейся в результате возникновения источника природной чрезвычайной ситуации, который может повлечь или повлечет за собой человеческие жертвы, ущерб здоровью людей и (или) окружающей природной среде, значительные материальные потери и нарушение условий жизнедеятельности людей.

Производство находится в городе Томске с континентально-циклоническим климатом. Природные явления (землетрясения, наводнения, засухи, ураганы и т. д.), в данном городе отсутствуют.

Возможными ЧС на объекте в данном случае, могут быть сильные морозы и диверсия (вандализм, хулиганство, шпионаж).

Для Сибири в зимнее время года характерны морозы. Достижение критически низких температур приводит к авариям систем тепло- и водоснабжения, сантехнических коммуникаций и электроснабжения, приостановке работы. В этом случае при подготовке к зиме следует предусмотреть

а) газобаллонные калориферы (запасные обогреватели),

б) дизель или бензоэлектрогенераторы;

в) запасы питьевой и технической воды на складе (не менее 30 л на 1 человека);

г) теплый транспорт для доставки работников на работу и с работы домой в случае отказа муниципального транспорта. Их количества и мощности должно хватать для того, чтобы работа на производстве не прекратилась.

В лаборатории кибернетического центра наиболее вероятно возникновение чрезвычайных ситуаций (ЧС) техногенного характера.

Для предупреждения вероятности осуществления вышесказанных диверсии предприятие необходимо оборудовать системой видеонаблюдения, круглосуточной охраной, пропускной системой, надежной системой связи, а также исключения распространения информации о системе охраны объекта, расположении помещений и оборудования в помещениях, системах

охраны, сигнализаторах, их местах установки и количестве. Должностные лица раз в полгода проводят тренировки по отработке действий на случай экстренной эвакуации.

Перечень НТД

[1] ГОСТ 54 30013-83. Электромагнитные излучения СВЧ. Предельно допустимые уровни облучения. Требования безопасности.

[2] ГОСТ 12.4.154-85. "ССБТ. "Устройства экранирующие для защиты от электрических полей промышленной частоты".

[3] ГН 2.2.5.1313-03. Предельно допустимые концентрации (ПДК) вредных веществ в воздухе рабочей зоны.

[4] СанПиН 2.2.4/2.1.8.055-96. "Электромагнитные излучения радиочастотного диапазона(ЭМИРЧ)".

[5] СанПиН 2.2.4.548-96. Гигиенические требования к микроклимату производственных помещений.

[6] ГОСТ Р 12.1.019-2009. Электробезопасность. Общие требования и номенклатура видов защиты.

[7] ГОСТ 12.4.123-83. Средства коллективной защиты от инфракрасных излучений. Общие технические требования.

[8] ГОСТ Р 12.1.019-2009. Электробезопасность. Общие требования и номенклатура видов защиты.

[9] ГОСТ 12.1.030-81. Электробезопасность. Защитное заземление. Зануление.

[10] ГОСТ 12.1.004-91. Пожарная безопасность. Общие требования.

[11] ОСТ 12.2.037-78. Техника пожарная. Требования безопасности.

[12] СанПиН 2.1.6.1032-01. Гигиенические требования к качеству атмосферного воздуха.

[13] ГОСТ 30775-2001. Ресурсосбережение. Обращение с отходами. Классификация, идентификация и кодирование отходов.

[14] СНиП 21-01-97. Противопожарные нормы.

[15] ГОСТ 12.4.154. Система стандартов безопасности труда. Устройства экранирующие для защиты от электрических полей промышленной частоты. Общие технические требования, основные параметры и размеры.

Заключение

В результате выполнения выпускной квалификационной работы был произведен анализ метеорологических данных, полученных из системы климатического мониторинга ИМКЭС СО РАН. Данные систематизированы по основным метеостанциям томского региона.

Был осуществлен обзор публикаций по байесовской статистике, рассмотрено применение метода Байеса в различных отраслях, рассмотрена модель *bsts*.

Было разработано программное обеспечение с возможностью построения моделей для анализа и прогнозирования метеорологических величин, спроектирована обобщенная функциональная структура программного комплекса.

Был разработан веб-интерфейс созданного метода прогнозирования временных рядов с использованием пакета *shiny* языка *R*.

На основе пакета *bssm* языка *R* были выполнены эксперименты для прогнозирования пропущенных данных.

На основе пакета *bsts* языка *R* был проведён программный эксперимент для создания и оценивания *bsts*-модели. Полученный результат показывает, что созданная модель хорошо учитывает все особенности температурных временных рядов и успешно прогнозирует будущие изменения данных.

В разделе «Финансовый менеджмент» была обоснована ресурсоэффективность разработки программного обеспечения.

В разделе «Социальная ответственность» был проведен анализ рабочего помещения, вредных и опасных факторов, и методы минимизации их воздействия на человеческое здоровье. Также рассмотрены аспекты экологической, производственной безопасности, безопасности в чрезвычайных ситуациях.

На английский язык переведена глава: Задача прогнозирования временных рядов и обзор некоторых методов ее решения.

Список студенческих публикаций

1. Чжу Чжунвэнь “Практика разработки технологического процесса изготовления специальных технических объектов” // Вестник научных конференций. – 2022. – № 2-3(78). – С. 103-106. – URL: <https://ukonf.com/doc/cn.2022.02.03.pdf>

2. Чжу Чжунвэнь “Прогнозирование с помощью байесовских структурных временных рядов” // Наука и технологии – 2023. Сборник статей Международной научно-практической конференции. – Петрозаводск: Новая наука, 2023. – С. 180-190. – «Диплом победителя I степени» в секции «Информатика» в Международной научно-практической конференции «Наука и технологии – 2023». – URL: <https://www.elibrary.ru/item.asp?id=52663848&pff=1>. – URL: <https://www.sciencen.org/assets/Kontent/Konferencii/Arhiv-konferencij/KOF-782.pdf>.

3. Zhu Zhongwen “Using R language BSTS package for modelling Bayesian structural time series” // Евразийский научный журнал. – 2023. – № 5. – С. 4-10. – URL: <https://journalpro.ru/articles/using-r-language-bsts-package-for-modelling-bayesian-structural-time-series/>

«Диплом победителя I степени» в секции «Информатика» в Международной научно-практической конференции «Наука и технологии – 2023».



Список литературы

1. W. Waheeb, R. Ghazali and H. Shah, "Nonlinear Autoregressive Moving-average (NARMA) Time Series Forecasting Using Neural Networks," 2019 International Conference on Computer and Information Sciences (ICCIS), Sakaka, Saudi Arabia, 2019, pp. 1-5, doi: 10.1109/ICCISci.2019.8716417.
2. M. Maya, W. Yu and X. Li, "Time series forecasting with missing data using neural network and meta-transfer learning," 2021 IEEE Symposium Series on Computational Intelligence (SSCI), Orlando, FL, USA, 2021, pp. 1-6, doi: 10.1109/SSCI50451.2021.9659864.
3. F. Liu and W. Wang, "Forecasting of Short-term Tourism Demand Based on Multivariate Time Series Clustering and LSSVM," 2022 IEEE 6th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Beijing, China, 2022, pp. 174-178, doi: 10.1109/IAEAC54830.2022.9929603.
4. Q. Xu, Q. Wen and L. Sun, "Two-Stage Framework for Seasonal Time Series Forecasting," ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 2021, pp. 3530-3534, doi: 10.1109/ICASSP39728.2021.9414118.
5. S. Masum, Y. Liu and J. Chiverton, "Comparative analysis of the outcomes of differing time series forecasting strategies," 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Guilin, China, 2017, pp. 1964-1968, doi: 10.1109/FSKD.2017.8393069.
6. L. Wang, J. Chen, W. Wang, R. Song, Z. Zhang and G. Yang, "Review of Time Series Traffic Forecasting Methods," 2022 4th International Conference on Control and Robotics (ICCR), Guangzhou, China, 2022, pp. 1-5, doi: 10.1109/ICCR55715.2022.10053870.
7. R. Reyhani and A. M. E. Moghadam, "A heuristic method for forecasting chaotic time series based on economic variables," 2011 Sixth International Conference on Digital Information Management, Melbourne, VIC, Australia, 2011, pp. 300-304, doi: 10.1109/ICDIM.2011.6093338.

8. T. -T. -H. Phan, É. P. Caillault and A. Bigand, "Comparative Study on Univariate Forecasting Methods for Meteorological Time Series," 2018 26th European Signal Processing Conference (EUSIPCO), Rome, Italy, 2018, pp. 2380-2384, doi: 10.23919/EUSIPCO.2018.8553576.
9. C. Hou, J. Wu, B. Cao and J. Fan, "A deep-learning prediction model for imbalanced time series data forecasting," in Big Data Mining and Analytics, vol. 4, no. 4, pp. 266-278, Dec. 2021, doi: 10.26599/BDMA.2021.9020011.
10. N. Ponnampereuma and L. Rajapakse, "Comparison of Time Series Forecast Models for Rainfall and Drought Prediction," 2021 Moratuwa Engineering Research Conference (MERCon), Moratuwa, Sri Lanka, 2021, pp. 626-631, doi: 10.1109/MERCon52712.2021.9525690.
11. Y. Liao and C. Liang, "A Temperature Time Series Forecasting Model Based on DeepAR," 2021 7th International Conference on Computer and Communications (ICCC), Chengdu, China, 2021, pp. 1588-1593, doi: 10.1109/ICCC54389.2021.9674623.
12. Гублер Е.В. Вычислительные методы анализа и распознавания патологических процессов. – Ленинград: Медицина, 1978. – 294 с.
13. Эфендиев Г.Д., Халилзаде Р.С., Гасанов Р.С. Анализ факторов, влияющих на исходы хирургического лечения геморрагических инсультов // Азербайджанский медицинский журнал. – 2003. – № 4. – С. 71-74.
14. Шаров В.Д. Применение байесовского подхода для уточнения вероятности событий в автоматизированной системы прогнозирования и предотвращения авиационных происшествий // Управление большими системами. – 2013. – № 43. – С. 240-253. URL: <https://cyberleninka.ru/article/n/primeneniye-bayesovskogo-podhoda-dlya-utochneniya-veroyatnostey-sobytiy-v-avtomatizirovannoy-sisteme-prognozirovaniya-i/pdf>.
15. Рустамов Я.И., Эфендиев Г.Д. Построение прогностической модели безотказной работы субартезианских скважин // URL: <https://cyberleninka.ru/article/n/postroenie-prognosticheskoy-modeli-bezotkaznoy-raboty-subartezianskih-skvazhin/pdf> (дата обращения: 16.03.2023).

16. Кирдищев Д.В. Применение метода Байеса при выявлении дефектов топливной аппаратуры по виброакустическим характеристикам во время работы дизеля // Вестник ГГТУ им. П.О. Сухого. – 2021. – № 1. – С. 92-99.
17. Байесовский структурный временной ряд – Bayesian structural time series // wiki5.ru. URL: https://wiki5.ru/wiki/Bayesian_structural_time_series (дата обращения: 17.03.2023).
18. Шевелев А. А. Байесовский подход к выявлению внешнего шока на макроэкономических показателях России // Мир экономики и управления. – 2017. – Т. 17. – № 1. – С. 26-40. URL: <https://cyberleninka.ru/article/n/bayesovskiy-podhod-k-otsenke-vozdeystviya-vneshnih-shokov-na-makroekonomicheskie-pokazateli-rossii>
19. Доманов А. О. Основы байесовского подхода к количественному анализу (на примере евроскептицизма) // Политическая наука. – 2021. – № 1. – С. 301-321. URL: <https://cyberleninka.ru/article/n/osnovy-bayesovskogo-podhoda-k-kolichestvennomu-analizu-na-primere-evroskeptitsizma>
20. Мастицкий С. Э. Анализ временных рядов с помощью R // ranalytics.github.io. [2020]. URL: <https://ranalytics.github.io/tsa-with-r> (дата обращения: 17.03.2023).
21. Scott S. L., Varian H. R. Predicting the present with Bayesian structural time series // ResearchGate GmbH. [2008-2023]. URL: https://www.researchgate.net/profile/Steven-Scott-7/publication/264816307_Predicting_the_Present_with_Bayesian_Structural_Time_Series/links/5ff4f3faa6fdccdc8339d17/Predicting-the-Present-with-Bayesian-Structural-Time-Series.pdf (accessed 17.03.2023).
22. Scott S. L., Varian H. R. Bayesian variable selection for nowcasting economic time series // nber.org. URL: <https://www.nber.org/system/files/chapters/c12995/c12995.pdf> (accessed 17.03.2023).
23. Brodersen K. H., Gallusser F., Koehler J., Remy N., Scott S. L. Inferring causal impact using Bayesian structural time-series models // Annals of Applied Statistics. – 2015. – vol. 9. – pp. 247-274. URL: <https://storage.googleapis.com/pub-tools-public-publication-data/pdf/41854.pdf>

24. Гусева М. Е., Силаев А. М. Использование байесовских методов для макроэкономического моделирования фаз бизнес-цикла // Вестник Санкт-Петербургского университета. Экономика. – 2021. – Т. 37. – Вып. 2. – С. 298-317. URL: <https://economicsjournal.spbu.ru/article/view/5065/8152>

25. На каком языке рисуют схемы: что такое UML и почему его понимают во всём мире // АНО ДПО «Образовательные технологии Яндекса», ООО «Яндекс». [2022]. URL: <https://practicum.yandex.ru/blog/uml-diagrammy/> (дата обращения: 17.03.2023).

26. Система климатического мониторинга // URL: <http://mon.imces.ru/> (дата обращения: 15.02.2023).

27. Package ‘bssm’ // URL: <https://cran.r-project.org/web/packages/bssm/bssm.pdf> (дата обращения: 18.04.2023).

28. Package ‘shiny’ // URL: <https://cran.r-project.org/web/packages/shiny/shiny.pdf> (дата обращения: 18.04.2023).

Приложение А

(справочное)

Time series forecasting and overview of some methods for its solution

Студенту:

Группа	ФИО	Подпись	Дата
8BM11	Чжу Чжунвэнь		16.03.2023 г.

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
доцент ОИТ ИШИТР	Ботыгин И.А.	к.т.н.		16.03.2023г.

Консультант-лингвист отделения иностранных языков ШБИП

Должность	ФИО	Ученая степень, звание	Подпись	Дата
доцент ОИЯ ШБИП	Степура С.Н.	к.фил.н.		16.03.2023г.

1. Time series forecasting and overview of some methods for its solution

1.1. Regression models

A regression model is a statistical approach used to analyze the relationship between a dependent variable and one or more independent variables. It aims to find the best-fitting mathematical function that describes the relationship between these variables.

The general formula representation of a regression model can be written as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon$$

Where:

y represents the dependent variable.

β_0 is the intercept (constant term).

$\beta_1, \beta_2, \dots, \beta_n$ are the coefficients or parameters associated with the independent variables x_1, x_2, \dots, x_n .

ε represents the error term, which captures the unexplained variability in the dependent variable.

Regression models analyze the functional relationship between the dependent variable and independent variables. They provide a way to quantitatively predict the value of the dependent variable based on the values of the independent variables, estimate the coefficients or parameters that define the relationship between the variables.

Regression models incorporate an error term to account for unexplained variability in the dependent variable. They rely on certain assumptions, such as linearity, independence of errors, homoscedasticity, and normality of residuals.

The following are some common methods of regression analysis:

- Simple Linear Regression: It involves a single independent variable and a linear relationship with the dependent variable.
- Multiple Linear Regression: It incorporates multiple independent variables to model the relationship with the dependent variable.

- Polynomial Regression: It allows for non-linear relationships by including polynomial terms of the independent variables.
- Logistic Regression: It is used for binary or categorical dependent variables, estimating the probability of belonging to a certain category.

Regression models help identify and quantify the relationships between variables. They enable prediction or estimation of the dependent variable based on the values of independent variables. Regression models assist in determining the relative importance of independent variables in explaining the variation in the dependent variable. They also provide interpretable coefficients, allowing for the understanding of the impact of independent variables on the dependent variable, facilitate hypothesis testing and the estimation of confidence intervals for coefficients.

Regression models have certain disadvantages. They assume linearity, which may not hold in all cases. Overfitting or underfitting can occur, impacting model generalization. Outliers or influential data points can affect estimated coefficients. Multicollinearity (high correlation between independent variables) and violation of assumptions can lead to biased or inefficient estimates.

In various fields regression models find applications. They are used in economics to analyze factors influencing economic indicators. In marketing, they help predict sales and understand consumer behavior. Finance employs regression models for risk analysis and financial forecasting. Healthcare utilizes them to predict patient outcomes and analyze treatment effects. Social sciences employ regression models to study the effects of social, demographic, and psychological factors on various outcomes.

1.2. Autoregressive models

Autoregressive model, referred to as AR model, is a statistical method for processing time series, using the same variable such as the previous periods of x , that is, x_1 to x_{t-1} to predict the performance of x_t in the current period, And assume that they have a linear relationship. Because this is developed from linear regression in regression analysis, but instead of using x to predict y , x is used to predict itself; so it is called

autoregression. Autoregressive models are widely used in the prediction of economics, informatics, and natural phenomena.

AR models use the past values of a series to predict its future values. This model assumes that the current value of a series depends on its previous values. The most popular AR model is the first order model (AR(1)), which assumes that the current value of a series depends only on its previous value (i.e. $p=1$).

The autoregressive model equation, denoted AR(p), is given below:

$$X_t = C + \Phi_1 X_{t-1} + E_t$$

Where

X_{t-1} — X value for the previous year/month/week.

ϕ_1 — The coefficient by which we multiply X_{t-1} . The value of ϕ_1 will always be 1 or -1.

E_t — The difference between the period t value and the correct value.

p — Order. Thus, AR(1) is a first-order autoregressive model. The second and third order will be AR(2) and AR(3) respectively.

An example of using the AR model is the analysis of financial markets. For example, using an AR model, you can predict future stock prices based on their past values. If you take past prices as a basis and apply an AR model, you can determine the likely prices of shares in the future.

MA-models use past error values - the difference between the actual values of a series and its predicted values, to predict future values. The most popular MA model is the first order model (MA(1)), which assumes that the current error depends only on its previous value.

An example of using the MA model would be to predict the number of users of a website. For example, if previous forecasts are wrong by a certain percentage, you can use this information in a forecasting model to improve the accuracy of the user count forecast.

ARIMA is a model that combines autoregressive and moving average models. ARIMA allows you to model data that is not stationary, as is not required for AR and MA

models. ARIMA includes three parameters: an autoregressive parameter (p), a moving average parameter (q), and an integration parameter (d).

The ARIMA(p,d,q) model for the non-stationary time series X_t has the form:

$$\Delta^d X_t = c + \sum_{i=1}^p a_i \Delta^d X_{t-i} + \sum_{j=1}^q b_j \varepsilon_{t-j} + \varepsilon_t$$

Where

ε_t — A stationary time series;

c, a_i, b_i — Model parameters;

Δ^d — The time series difference operator of order d.

An example of using ARIMA would be to forecast the monthly revenue of a product based on last year's monthly data. If past data is trending, seasonal, or cyclical, you can use ARIMA to factor these factors into your revenue forecast.

The advantage of the autoregressive method is that it does not require much data, and it can use its own variable series to make predictions. But this method is subject to certain limitations:

Must have autocorrelation, the autocorrelation coefficient is the key. If the autocorrelation coefficient (R) is less than 0.5, it should not be used, otherwise the prediction result is extremely inaccurate.

Autoregression can only be applied to predict economic phenomena related to its own previous period, that is, economic phenomena that are greatly affected by its own historical factors, such as the amount of mining, the output of various natural resources, etc.; for economic phenomena that are greatly affected by social factors, instead of autoregressive, vector autoregressive models that can incorporate other variables should be used instead.

1.3. Exponential smoothing prediction method

Exponential smoothing is a prediction method used in time series analysis to forecast future values based on historical data. It is particularly useful for data with no or minimal seasonality and a relatively stable trend.

The main idea behind exponential smoothing is to assign weights to past observations in a way that gives more importance to recent data points and less importance to older data points. This weighting process is based on an exponentially decreasing function. The most recent observations are given the highest weight, while the weights decrease exponentially as we move further back in time.

There are different variations of exponential smoothing methods, including simple exponential smoothing, double exponential smoothing, and triple exponential smoothing (also known as Holt-Winters method). Each variation incorporates different levels of complexity to capture different patterns in the data.

The formula for simple exponential smoothing is as follows:

$$F(t+1) = \alpha * Y(t) + (1 - \alpha) * F(t)$$

Where:

$F(t+1)$ is the forecast for the next period,

$Y(t)$ is the actual value in the current period,

$F(t)$ is the forecast for the current period,

α is the smoothing parameter ($0 \leq \alpha \leq 1$) that determines the weight given to the most recent observation.

Exponential smoothing is relatively easy to understand and implement. It can be adapted to different types of time series data by choosing the appropriate variation. Exponential smoothing methods are computationally efficient, making them suitable for large datasets. As new data becomes available, exponential smoothing models can be easily updated to incorporate the latest information.

Exponential smoothing methods are most effective for data with no or minimal seasonality and a stable trend. They may not perform well with complex or highly volatile data. This model is primarily focused on forecasting future values and does not provide insights into the underlying factors driving the data. The initial value or starting point of the forecast can have a significant impact on the results, making it crucial to choose an appropriate initial value.

Exponential smoothing can be applied in various domains, including:

- Demand forecasting: It is commonly used in inventory management and production planning to forecast customer demand.
- Financial forecasting: Exponential smoothing methods are employed to predict stock prices, market trends, and financial indicators.
- Operations management: It helps in predicting sales, demand for services, and resource allocation.
- Time series analysis: Exponential smoothing is a fundamental technique for studying and analyzing time-dependent data.
- Supply chain management: It assists in predicting order quantities, optimizing inventory levels, and improving supply chain performance.

It's important to note that the choice of the specific exponential smoothing method and the appropriate parameter values depends on the characteristics of the data and the specific forecasting objectives.

1.4. Artificial Neural Networks

The artificial neural network abstracts the neuron network of the human brain from the perspective of information processing, establishes a simple model, and forms different networks according to different connection methods.

A neuron is a unit of information processing in a neural network. On Fig.1. shows a model of a neuron that underlies artificial networks. The model contains 3 main elements:

1. A set of connections;
2. The adder;
3. The activation function.

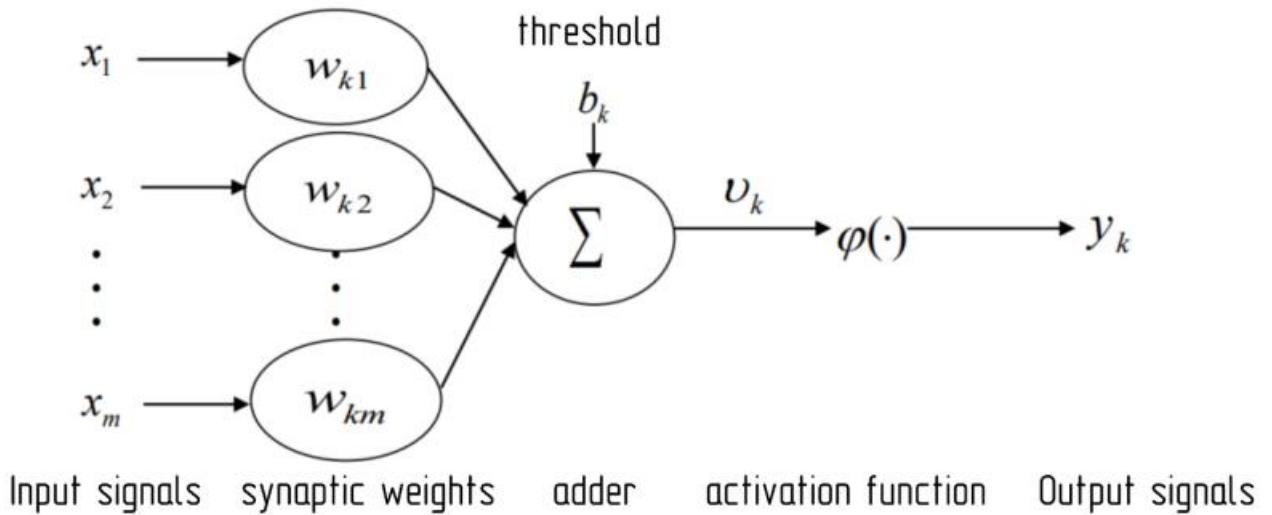


Fig.1. Nonlinear neuron model

Each node represents a specific output function called an activation function. Each connection between two nodes represents a weighted value for the signal passing through the connection, called weight. The output of the network varies according to the way the network is connected, the weight value and the activation function.

In mathematical representation, the functioning of neuron k is carried out as follows:

$$u_k = \sum_{j=1}^m w_{kj} X_j$$

$$y_k = \Phi(u_k + b_k)$$

Where

- x_1, x_2, \dots, x_m – input signals;
- w_{k1}, w_{k2}, w_{km} – synaptic weights of neuron k ;
- u_k – linear combination of input actions;
- b_k – threshold; □
- $\Phi(\cdot)$ – activation function;
- y_k – output signals.

1.5. Models based on Markov chains

Models based on Markov chains are commonly used for forecasting time series data. Markov chain models assume that the future state of a system depends only on its

current state and not on its past states, making them suitable for modeling time series data with certain patterns or dependencies.

The mathematical representation of the Markov chain model for time series forecasting involves the transition probabilities. The transition probability matrix $P(i, j)$ represents the probability of transitioning from state s_i to state s_j . If the time series data follows a first-order Markov chain, the transition probabilities can be represented as:

$$P(i, j) = P(X_{t+1} = s_j \mid X_t = s_i)$$

The transition probabilities should satisfy the following properties:

$$P(i, j) \geq 0 \text{ for all } i, j.$$

$$\sum P(i, j) = 1, \text{ for each } i.$$

If the transition probabilities have been estimated, they could be used to forecast the future values in the time series based on the current state. The specific forecasting method will depend on the desired approach, such as generating probabilistic forecasts using the transition probabilities or determining the most likely next state.

An example of a Markov chain for a process that has three states is shown in Fig.2.

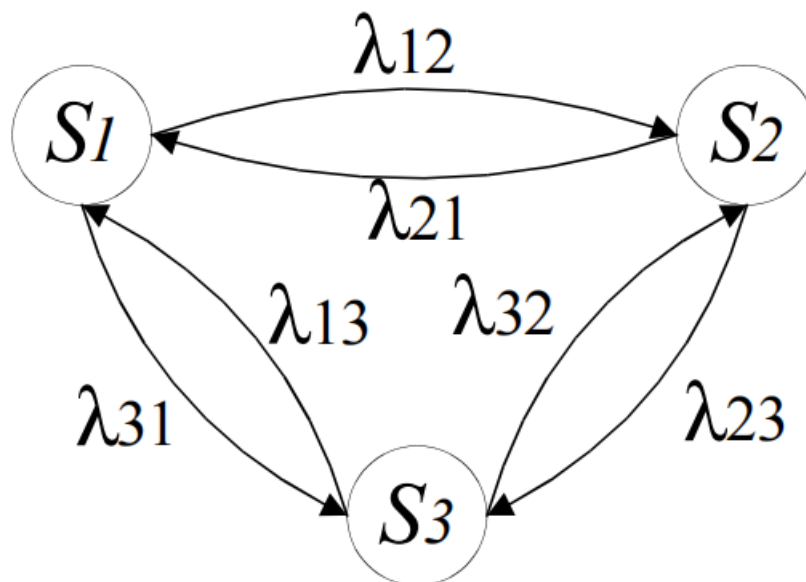


Fig.2. Three-State Markov Chain

Here S_1, \dots, X_3 are the states of the process $Z(t)$; λ_{12} is the probability of transition from state S_1 to state S_2 , λ_{23} is the probability of transition from state S_2 to state S_3 , etc. When constructing a Markov chain, a set of states and transition probabilities are

determined. If there is a current state of the process S_i , then such a state S_i is chosen as the future state of the process, the probability of transition to which (value λ_{ij}) is maximum.

Markov chain models are relatively simple and easy to implement, making them accessible to a wide range of users. They can capture dependencies and patterns in time series data, including trends, seasonality, and other temporal relationships.

Markov chain models provide interpretable transition probabilities, allowing for insights into the dynamics of the system being modeled. Forecasting with Markov chain models can be computationally efficient, particularly for moderate-sized datasets.

Markov chain models also have some disadvantages, the models assume that the future state depends only on the current state, disregarding any information from previous states. This assumption may not hold for all time series data, limiting the model's ability to capture long-term dependencies.

Models are typically used for short-term forecasting since their accuracy tends to degrade over longer forecast horizons.

The initial state or starting point can have a significant impact on the forecasted sequences, highlighting the importance of selecting an appropriate initial state.

1.6. Overview of current trends in time series forecasting

Currently, time series are one of the most common objects of analysis. There are many models for analyzing and forecasting time series:

- * integrated model of autoregression - moving average (ARMA) and its extension for non-stationary time series (ARIMA);

- * SARIMA-model – an extension of the ARIMA model with seasonal conditions;

- * ARFIMA – autoregressive fractionally integrated moving average model;

- * VAR-model – vector autoregression model;

- * exponential smoothing model and its modification TBATS for interaction with several seasonalities;

- * Prophet-model – procedure for fitting additive regression models with trend, annual seasonality, weekly seasonality, holidays;

- * GARCH-model – generalized autoregressive conditionally heteroscedastic model;

- * NNETAR-model – модель авторегрессии нейронной сети;
- * DeepAR-model – probabilistic forecasting model with autoregressive recurrent networks;
- * NARX – nonlinear autoregressive exogenous model;
- * and others.

Article [3] compared one-step forecasting using a simulated non-linear moving average autoregressive time series (NARMA) between two groups of neural networks. The results show that introducing error feedback into neural networks helps in NARMA time series forecasting and neural networks perform better compared to the ARMA model.

Article [4] combined meta-learning and transfer learning to address the lack of data in time series forecasting. The prediction results and comparison show that metatransfer learning has very good predictions.

Paper [5] proposes a method for short-term forecasting of tourist demand based on multivariate time series clustering and LSSVM.

Literature [7] provides an overview of time series forecasting, a classification of time series forecasting, and time series forecasting approaches and strategies. A comparative analysis of strategies using the ARIMA model is demonstrated.

Article [9] proposes a new heuristic method based on the chaotic characteristics of economic time series, which predicts these time series using artificial neural networks.

In the literature [11], it was found that some time series are unbalanced and show different patterns between special and normal periods, a unified model was developed to correct the imbalance to improve the accuracy of forecasts for special periods.

Reference [13] considers an improved network model based on DeepAR for making probabilistic forecasts. The result shows that the proposed model has a good improvement in accuracy over DeepAR and widely used baselines.

1.7. Bayesian inference for time series forecasting

A well-known modern approach to the analysis of time series is based on Bayes' theorem. The Bayes formula is increasingly used in theory and practice, is actively used

in various fields of knowledge to assess the degree of risk or chances of success, where the best of several options is required.

A large number of scientific and educational articles have been devoted to the Bayesian inference to predicting phenomena.

Study [20] examines the possibilities of using the Bayesian vector autoregression (BVAR) model to assess the impact of external shocks on the dynamics of Russia's macroeconomic indicators.

Literature [**Ошибка! Закладка не определена.**] describes a methodology for using the R programming language and statistical computing system for time series analysis. Solutions of several standard problems are derived, including forecasting, detection of structural changes and anomalies in data, as well as clustering of time series. Appropriate approaches and software are presented.

Article [24] develops a model for short-term economic performance forecasting and presents a scalable approach to time series forecasting for fat regressions.

One of the promising approaches in the analysis of time series is their representation in the form of a Bayesian structural time series (BSTS model).

Bayesian structural time series, also known as "state space models" or "dynamic linear models," are a type of time series models that can dynamically fit changes in time series structure. It is a statistical technique used for feature selection, time series forecasting, nowcasting, causation, and other applications.

BSTS uses Monte Carlo Markov Chain to estimate, the MCMC method uses the Monte Carlo integral of the Markov chain. Its basic idea is to construct a Markov chain so that its stationary distribution is the posterior distribution of the parameters to be estimated, and generate samples of the posterior distribution through this Markov chain. And Monte Carlo integration is performed based on the samples (effective samples) when the Markov chain reaches a stationary distribution. The most common MCMC sampling can only approximate the distribution due to the initial sample. The complex MCMC improvement algorithm is like past coupling, but it will consume more computing resources and time. A typical usage is to simulate a random walking pedestrian for path optimization, etc. Each step counts as a state. And the place with the most number of

passes will be more likely to be the destination in the next step. The convergence characteristic of the MCMC algorithm is the global minimum, and the corresponding linear least squares rule is easy to stagnate in the case of local convergence or slow to converge due to nonlinear problems. At the same time, a large number of applications show that the MCMC method is significantly better than the linear model.

For estimation and simultaneous regularization of regression coefficients, spike-and-slab method is a type of Bayesian linear regression in which a particular hierarchical prior for the regression coefficients is chosen in such a way that only a subset of possible regressors is retained. Some regression coefficients are assigned a high prior probability that they are zero. Further, when MCMC-sampling the coefficients from the obtained posterior distributions, many coefficients turn out to be exactly equal to zero.

1.8. Structural models of time series

It is quite difficult to classify domain models, in view of their diversity. One of the criteria for classifying models is the type of dependence of the future value of the observed parameters on the past ones. In this case, time series models are decomposed into two classes: statistical and structural.

In statistical models, at the first stage, an assumption is made about the nature of the relationships between the analyzed variables (statistical assumption), and then the correspondence between real data and data obtained using a set of statistical assumptions is examined. Thus, statistical models represent, often in a greatly idealized form, the process of generating data.

Structural time series models are one of the families of state space models. In structural models, a frame structure is formed, which determines the functional dependence between existing, predictable and external factors. The time series in this case is presented as a sum of unobserved components that can be interpreted, for example, as a trend, a seasonal component, a random component, and others. These components of the time series are not observable, they are theoretical values. A trend is understood as a change that determines the general direction of development, the main trend of the time series. That is, a trend is a systematic component of a long-term action. Seasonal

fluctuations are fluctuations associated with the change of seasons and having a pronounced annual periodicity. If the trend and periodic components are removed from the time series, then an irregular component (random) will remain. The random component is the action of a large number of relatively weak secondary factors. The influence of each of the secondary factors is insignificant, but their total impact is felt.

Let y_t denote the observation at time t in the real time series. The structural model of the time series is given by two equations relating y_t with the state variable vector of the system under study α_t .

Observation equation (1):

$$y_t = Z_t^T \alpha_t + \varepsilon_t; \quad \varepsilon_t \sim N(0, H_t)$$

Where:

y_t – The observed data at time t .

Z_t, H_t – Are structural parameters (model matrices).

α_t – The vector of hidden variables (states of the system under study).

Уравнением перехода (2):

$$\alpha_{t+1} = T_t \alpha_t + R \eta_t; \quad \eta_t \sim N(0, Q_t)$$

Where:

T_t, R_t and Q_t – The observed data at time t .

η_t may have a lower dimension than α_t .

Equation (1) is called the observation equation because it relates the observed data y_t to the explicitly unobservable (latent) state α_t . Equation (2) is called the transition equation because it determines how the latent state changes over time, that is, it specifies the process of transition from one latent state to another. The model matrices Z_t, T_t , and R_t usually contain a mixture of known values (often 0 and 1) and unknown parameters. The transition matrix T_t is square, but R_t can be rectangular if some of the state transitions are deterministic. The presence of R_t in equation (2) makes it possible to work with the full-rank variance matrix Q_t , since any linear dependencies in the state vector can be transferred from Q_t to R_t .

The state of the system at each moment of time is determined only by its state at the previous moment, i.e. the dynamics of the system has a Markov character. A

schematic representation of the state-space model for a finite time series with T observations is shown in Fig. 3.

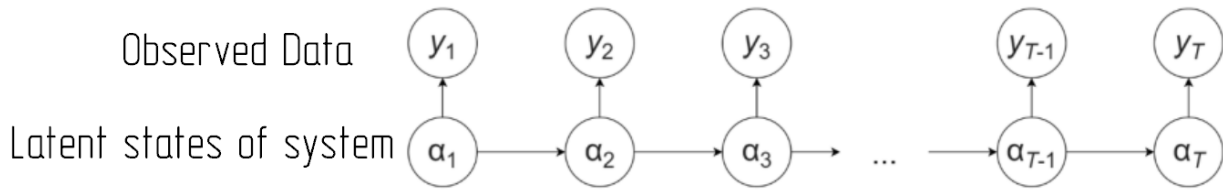


Fig. 3. State space model for a time series with T observations

If the model can be described by equations (1) and (2), then it is considered that it is in the form of a state space, that is, they are state space models. The advantages of time series prediction based on state space model are:

1. The state space model is a structural model that predicts time series based on the state space decomposition model, facilitating analysts to use existing statistical theories to conduct statistical tests on the model.
2. The core of the state space model solving algorithm is Kalman filtering, which is the most ideal recursive process of calculating the state vector based on all available information at time t . When the disturbance term and initial state vector obey normal distribution, Kalman filter can calculate likelihood function through prediction error decomposition, so that all unknown parameters in the model can be estimated, and once new observations are obtained, Kalman filter can be used to continuously modify the estimation of state vector.

1.9. Filter and Kalman smoothing

The main tools for working with state-space models are the Kalman filter, Kalman smoothing, and Bayesian data enhancement. But filtering and smoothing are traditional computational operations associated with state-space models.

Let us denote the entire set of observed data at each time t as $y_{1:t} = \{y_1, y_2, \dots, y_t\}$. The Kalman filter recursively calculates the predicted distribution $p(\alpha_{t+1}|y_{1:t})$ by combining $p(\alpha_t|y_{1:t-1})$ with y_t using a standard set of formulas that is logically equivalent to linear regression. The Kalman smoother updates the output of the Kalman filter to obtain $p(\alpha_t|y_{1:n})$ where n is the length of the time series at each value of t . Since all model

components are Gaussian, both $p(\alpha_{t+1}|y_{1:t-1})$ and $p(\alpha_{t+1}|y_{1:n})$ are multivariate normal distributions parameterized by their mean μ_t and variance P_t . Kalman filtering accumulates information about the time series as it moves forward through the list of elements (μ_t, P_t) . The Kalman smoother moves backward in time, distributing information about later observations over successively earlier pairs (μ_t, P_t) . It should be noted that the forward-backward scheme is quite common for algorithms in optimization and inference problems. The Kalman filter and Kalman smoothing are special cases of this scheme.

In Bayesian computing, it is often desirable to model a state based on its posterior distribution given the data. Let $y = y_{1:n}$ and $\alpha = \alpha_{1:n}$ denote complete sets of observed and latent data. Bayesian data augmentation methods produce simulations from $p(\alpha|y)$. You can not just take every α_t from $p(\alpha_t|y)$ because there must be a serial correlation between α_t and α_{t+1} . Instead, stochastic versions of the Kalman smoother are used which can be used to sample directly from $p(\alpha|y)$.