

Министерство образования Республики Беларусь
УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ
«ПОЛОЦКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(Полоцкий государственный университет)

УДК 001.891.573

№ госрегистрации 20191726

Инв № _____

Проректор по научной работе
к.т.н., доцент

_____ Ю.П. Голубев
« ____ » _____ 2020г.

ОТЧЕТ
О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ
**«Разработка методов анализа больших данных для решения задачи
прогнозирования потребления газа на магистральном газопроводе ОАО
"Газпром трансгаз Беларусь"»**
(заключительный)
задание 1.08.17 государственной программы научных исследований
«Информатика, космос и безопасность» подпрограмма «Информатика и
космические исследования»
на 2019-2020гг.

Начальник ОСНИ

_____ Т.В. Гончарова
« ____ » _____ 2020г.

Научный руководитель
к.т.н., доцент

_____ Д.О. Глухов
« ____ » _____ 2020г.

Новополоцк 2020

СПИСОК ИСПОЛНИТЕЛЕЙ

Руководитель,

к.т.н., доцент

_____ «__» ____ 20__ г. Д.О. Глухов
(Главы 1-3,
Заключение)

Ответственный

исполнитель

старший

преподаватель

_____ «__» ____ 20__ г. Т.М. Глухова
(Введение, Глава 3)

Исполнители

Научный сотрудник

_____ «__» ____ 20__ г. А.П. Андриевский
(Глава 3)

Научный сотрудник

_____ «__» ____ 20__ г. А.Н. Янушенок
(Глава 3)

Научный сотрудник,

к.т.н., доцент

_____ «__» ____ 20__ г. Р.П. Богуш
(Глава 2)

Нормоконтролер

_____ «__» ____ 20__ г. Л.В. Ищенко

РЕФЕРАТ

Отчет 89 страниц, 44 рисунков, 5 таблиц, 83 источника.

ЗАПАС ГАЗА; МАГИСТРАЛЬНЫЙ ГАЗОПРОВОД; РАСЧЕТ;
МОДЕЛИРОВАНИЕ; БЕЛТРАНСГАЗ

Наименование задания: Разработка методов анализа больших данных для решения задачи прогнозирования потребления газа на магистральном газопроводе ОАО "Газпром трансгаз Беларусь".

Объект исследования или разработки: математическая модель газотранспортной системы.

Цель работы – провести исследования для моделирования и прогнозирования потребления газа в задаче транспортировки газа в стационарном режиме.

При проведении исследований использовались следующие методы и методологии: методы обработки больших данных, нейросети, моделирование на ЭВМ, численные методы: методы повышения скорости сходимости, методы анализа качества.

Результаты НИР вошли в состав программного комплекса расчета запаса газа ОАО «Газпром трансгаз Беларусь» (разработчик Полоцкий государственный университет). А также используются на кафедре трубопроводного транспорта, вентиляции и гидравлики УО «Полоцкий государственный университет» при чтении лекций, проведении лабораторных работ, при выполнении дипломных проектов и при проведении научно-исследовательской работы студентов, магистрантов и аспирантов.

СОДЕРЖАНИЕ

СОДЕРЖАНИЕ	4
ВВЕДЕНИЕ.....	6
ГЛАВА 1 АНАЛИЗ ОСОБЕННОСТЕЙ СИСТЕМЫ ПОТРЕБЛЕНИЯ ГАЗА РЕСПУБЛИКИ БЕЛАРУСЬ.....	7
1.1 Характеристики системы потребления газа	7
1.2 Анализ исходных данных.....	9
Выводы по главе 1	18
ГЛАВА 2 АНАЛИЗ МЕТОДОВ И АЛГОРИТМОВ ПРОГНОЗИРОВАНИЯ ПРОЦЕССА ПОТРЕБЛЕНИЯ ГАЗА.....	19
2.1 Представление данных	21
2.2 Предварительная подготовка данных.....	22
2.3 Прогнозирование потребления газа методом экспоненциального сглаживания.....	23
2.4 Прогнозирование на основе многослойного персептрона	25
2.5 Оценка качества прогноза потребления газа	26
2.6 Обзор перспективных нейросетевых методов прогнозирования .	29
Выводы по главе 2	38
ГЛАВА 3 РАЗРАБОТКА И СРАВНИТЕЛЬНОЕ ИССЛЕДОВАНИЕ МЕТОДОВ ПРОГНОЗИРОВАНИЯ ПОТРЕБЛЕНИЯ ПРИРОДНОГО ГАЗА	39
3.1 Влияние нестационарных процессов на исходные данные.....	39
3.2 Разработка генетического алгоритма аппроксимации многомерных зависимостей	40
3.3 Модель ARIMA.....	51
3.4 Модель экспоненциального сглаживания.....	60
3.5 Прогнозирование с использованием метода группового учета аргументов	63

3.6 Разработка подсистемы визуализации данных прогноза	66
Выводы по главе 3	76
ЗАКЛЮЧЕНИЕ	78
СПИСОК ИСПОЛЬЗУЕМЫХ ЛИТЕРАТУРНЫХ ИСТОЧНИКОВ	80

ВВЕДЕНИЕ

Проводимые исследования являются развитием научно-исследовательской работы "Разработка программного комплекса расчета запаса газа в газотранспортной системе ОАО «Белтрансгаз»", выполненной в рамках хозяйственного договора №10205 от 01.07.2010г., ГПНИ 2016-2018 «Методы и алгоритмы расчета стационарной модели транспортировки газа большой размерности для решения задачи учета запаса газа на магистральном газопроводе ОАО "Газпром трансгаз Беларусь"».

Задача настоящих исследований: Разработка методов анализа больших данных для решения задачи прогнозирования потребления газа на магистральном газопроводе ОАО "Газпром трансгаз Беларусь".

ГЛАВА 1 АНАЛИЗ ОСОБЕННОСТЕЙ СИСТЕМЫ ПОТРЕБЛЕНИЯ ГАЗА РЕСПУБЛИКИ БЕЛАРУСЬ

1.1 Характеристики системы потребления газа

Система потребления газа Республики Беларусь обладает рядом особенностей, определяемых климатом, сезонностью работы предприятий, суточными колебаниями потребления газа.

Исходя из закрытости данных о потреблении и транзите газа, представляющих коммерческую тайну ОАО «Газпром трансгаз Беларусь», мы, разрабатывая алгоритмы прогнозирования и анализа данных, будем оперировать измененными данными без указания конкретных географических названий и населенных пунктов.

В настоящее время учет потребления газа в Республике Беларусь ведется по 234 потребителям, сгруппированным в 8 управлений (УМГ).

Телеметрия в рамках задачи расчета запаса газа предоставляется 1 раз в час по 2500 узлам (температура, давление, состояние задвижек, мгновенные значения расхода). В результате расчета запаса газа мы получаем расчетные значения режима транспортировки газа и значения запаса газа по каждому линейному участку, по группам участков и каждому УМГ в целом, текущую расчетную схему, расчет товарно-транспортной работы. Общий объем данных, поступающих в систему учета каждый час составляет около 30Мб данных. За год объем данных составляет 262.8Гб. За 9 лет эксплуатации системы он составил 2.36Тб данных. Данное обстоятельство требует изучение вопроса применения современных методов обработки больших данных для решения таких задач как прогнозирование последствий управленческих решений, построение прогноза потребления газа.

Пример неоднородности потребления газа в Республике Беларусь различными потребителями представлен на рисунке 1.

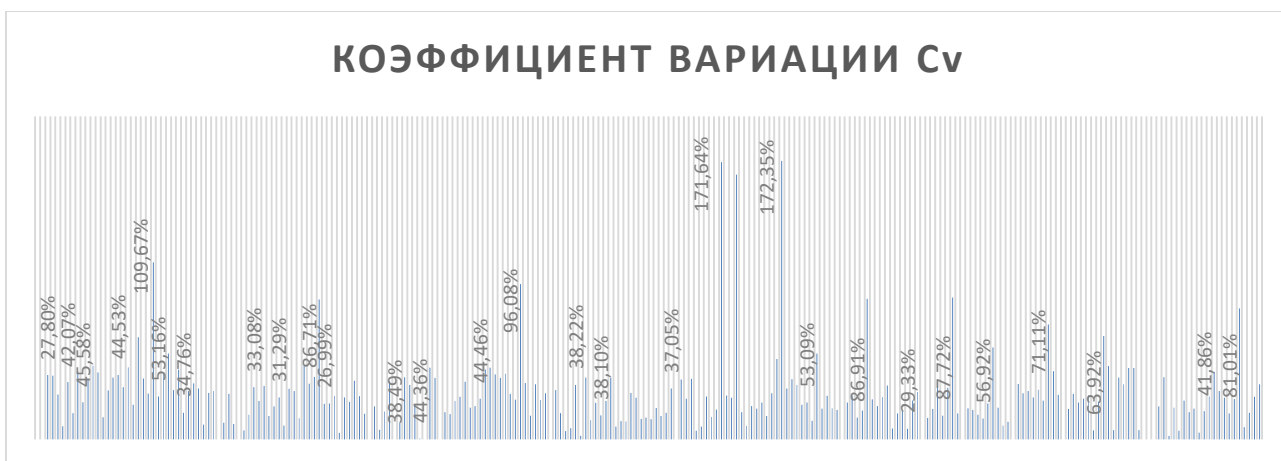
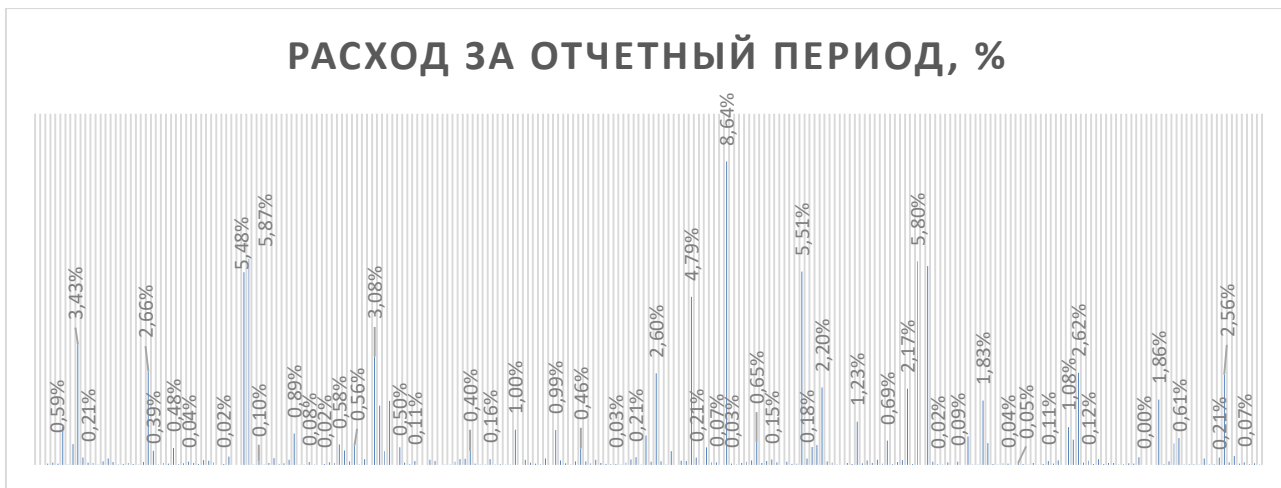


Рисунок 1 – Иллюстрация неоднородности потребления газа в Республике Беларусь

Среднее значение коэффициента вариации потребления газа в течение года составляет 47.3%.

Телеметрические данные характеризуются следующими особенностями:

1. Данные имеют пропуски, связанные с выходом телеметрии из строя, сбоями передачи данных, выводом оборудования в ремонт, регламентных работ по обслуживанию и др.
2. Данные имеют ошибки, связанные с ошибками датчиков, ошибки распределенной системы сбора телеметрии при замене датчиков,

изменения единиц измерения, рассинхронизации имен тегов телеметрии и т.п.

3. Данные имеют не выровненные метки времени и даты, определяемые моментом замера мгновенного значения датчиком телеметрии.

Пример тренда данных представлен на рисунке 2.

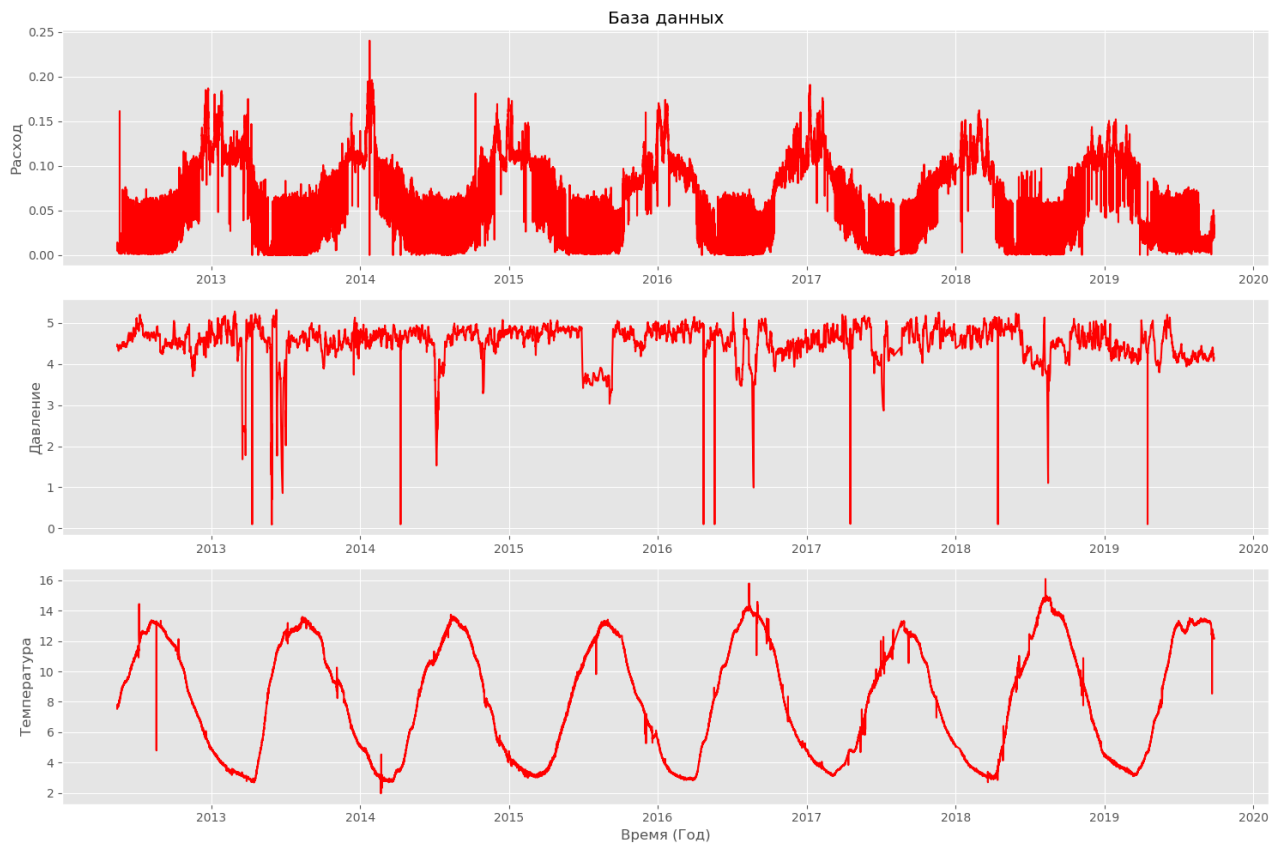


Рисунок 2 – Тренд данных за 7 лет о потреблении газа Потребителем

1.2 Анализ исходных данных

Нами изучены тренды 220 потребителей природного газа Республики Беларусь за последние 8 лет (2012-2020гг.). Причем данные 2012г. появляются с мая и данные по 2020г. неполны.

Все тренды обладают вышеперечисленными особенностями в силу ряда причин:

- В данных наблюдается присутствие ошибок телеметрии, выбросов и нулей, связанных с временным выходом из строя датчиков расхода, или давления и температуры, при вычислении расхода расчетным путем;
- Есть пропуски данных на периоды плановых и экстренных ремонтов газопровода;
- Отсутствует точное выравнивание меток времени;
- Данные отражают не факт потребления газа, а режим работы газотранспортной сети по доставке контрактного газа потребителю, соответственно отражает и нестационарные процессы (закачка газа в систему, разбор газа из системы в режиме баллона, когда движения газа через ГРС или компрессорную станцию не происходит, а потребление газа продолжается);
- Расходы газа, фиксируемые датчиками, лишены информации о плотности газа;
- Часть расходов получено расчетным путем по стационарной неизотермической модели при граничных условиях, заданных через давление и температуру;
- В данных за 8 лет происходили переименования тегов телеметрии, изменение схемы, появление новых и отключение некоторых потребителей;
- Потребление газа различными потребителями отличается друг от друга в силу особенностей работы местных предприятий, сезонности и работы смен.

Все перечисленные особенности требуют специальной подготовки данных перед подбором модели прогнозирования, а, следовательно, анализа исходных данных для выработки решения о виде предварительной обработки.

Рассмотрим статистические характеристики данных для нескольких крупных потребителей.

Названия потребителей в силу того, что эта информация имеет коммерческий характер и с ОАО «Газпром трансгаз Беларусь» у нас подписано соответствующее соглашение о конфиденциальности, мы называть не будем.

Данные за последние 8 лет структурированы в виде файлов расчетных схем газопровода, сохраняемых в архив после расчета по стационарной неизотермической модели каждый час.

Для получения информации нами написан алгоритм миграции данных об основных потребителях газа в реляционную базу данных с нормализацией темпоральных данных до 6-ой нормальной формы (это обеспечивает максимальную компактность хранения и накопление изменений не только значений телеметрии, но и изменения названий тегов телеметрии) и в нереляционную базу данных MongoDB для обеспечения возможности выполнения операций с большими данными в рамках архитектуры Map-Reduce. MongoDB позволяет создавать программы назначения ключей и редуцирования на языке программирования JavaScript прямо в облаке (рисунок 3).

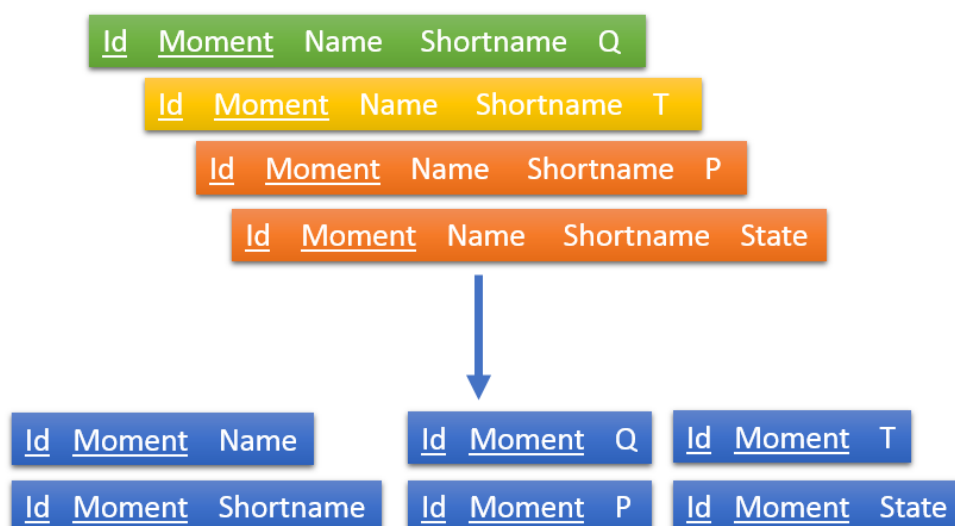


Рисунок 3 – Исходная и нормализованная схема базы данных телеметрии

Для формирования прогноза о расходах газа по месяцам нам необходимы тренды, агрегированные по месяцам. Для выполнения этой операции мы воспользуемся возможности Map-Reduce операции, в рамках которой ключами будут выступать метки, сформированные по номеру месяца, начиная с мая 2012г. и идентификатора потребителя. Схема алгоритма приведена на рисунке 4. После формирования выходных файлов выполняется декодирование ключей и получение соответствующих трендов, соответствующих потребителей.

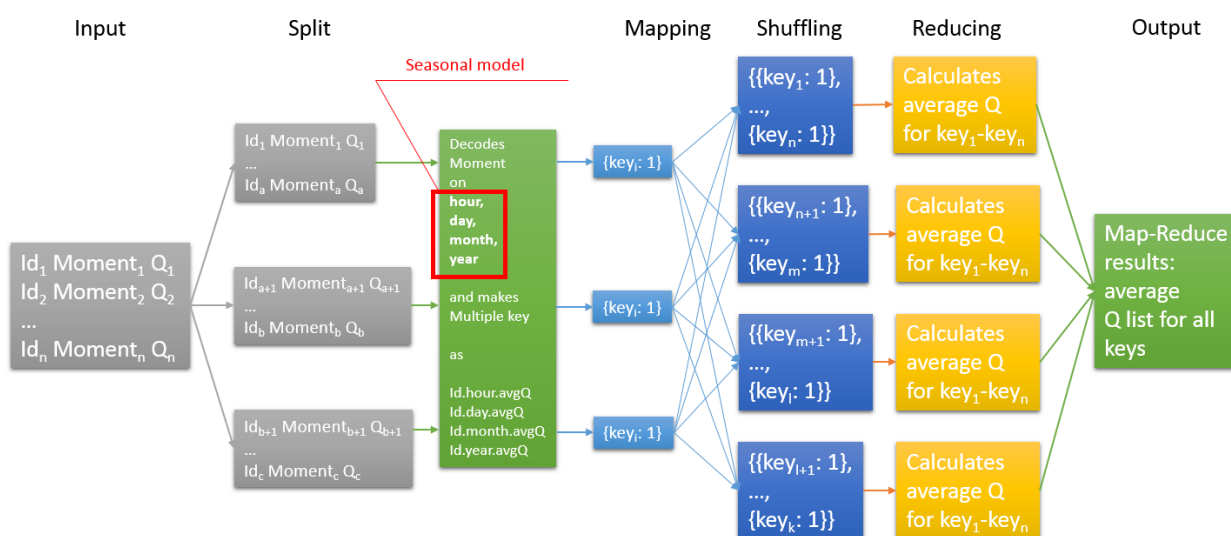


Рисунок 4 – Схема Map-Reduce алгоритма агрегирования трендов по заданным временным интервалам

Анализируя тренды, агрегированные по месяцам, нам необходимо определить такую минимальную степень дифференцирования временного ряда и его сезонной составляющей, при которой остатки ряда будут представлять собой белый шум и результирующий ряд будет стационарным.

Для этого построим коррелограммы автокорреляционной функции (ACF) и коррелограммы частичной автокорреляционной функции (PACF) для каждого потребителя для различных параметров дифференцирования несезонной и сезонной составляющих ряда. Сезонность рядов будет

составлять 12 месяцев. Рассмотрим на примере тренда за 89 месяцев (рисунок 5).

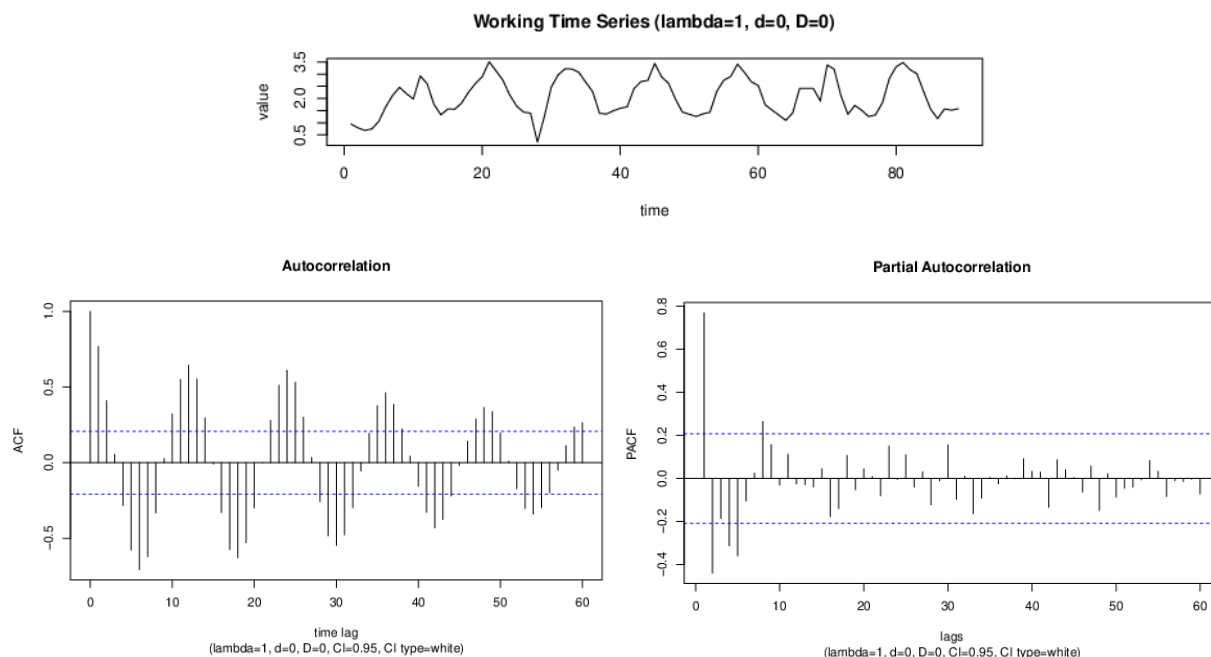


Рисунок 5 – ACF и PACF коррелограммы исходного ряда

Как мы видим, исходный ряд не является стационарным, в частности по причине явно выраженной сезонности. Попробуем проанализировать остатки после вычитания сезонной и трендовой составляющей в соответствии с методом [30]. Для проверки корректно определения сезонной составляющей проанализируем периодичность остатков ряда построением кумулятивной периодограммы (с 95% доверительными интервалами Колмогорова-Смирнова, рисунок 6).

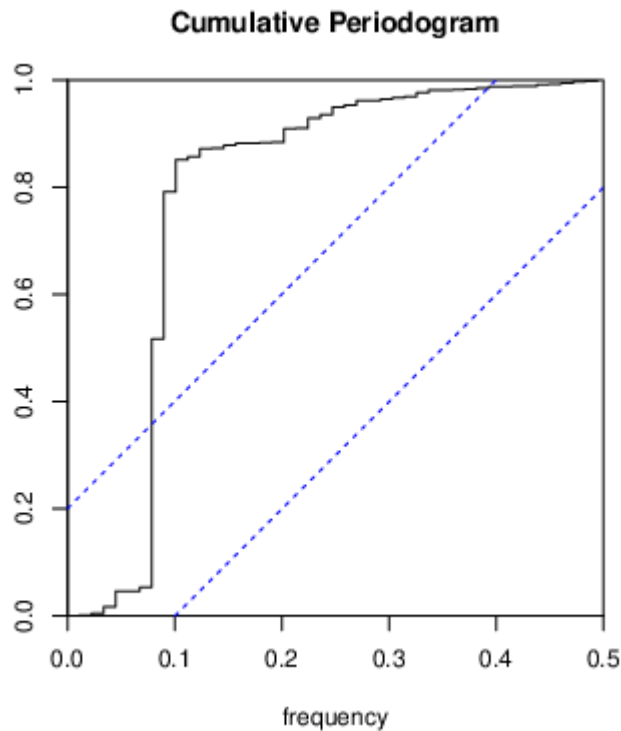


Рисунок 6 – Кумулятивная периодограмма остатков временного ряда

Как мы видим наблюдается существенное отклонение периодограммы от линейности на средних частотах.

Для решения задачи подбора параметров дифференцирования построим для каждого потребителя матрицу уменьшения дисперсии. В случае нашего примера она представлена на рисунке 7.

Variance Reduction Matrix					
$V(Y[t],d=0,D=0)$	0.610425	Range	3.3095	Trim Var.	0.438436
$V(Y[t],d=1,D=0)$	0.268133	Range	2.6778	Trim Var.	0.152753
$V(Y[t],d=2,D=0)$	0.389261	Range	3.91102	Trim Var.	0.16786
$V(Y[t],d=3,D=0)$	1.0213	Range	7.06077	Trim Var.	0.437294
$V(Y[t],d=0,D=1)$	0.267371	Range	3.10766	Trim Var.	0.120245
$V(Y[t],d=1,D=1)$	0.311259	Range	3.60051	Trim Var.	0.137168
$V(Y[t],d=2,D=1)$	0.733927	Range	5.31665	Trim Var.	0.324422
$V(Y[t],d=3,D=1)$	2.12705	Range	8.70267	Trim Var.	0.986151
$V(Y[t],d=0,D=2)$	0.74133	Range	5.27679	Trim Var.	0.266356
$V(Y[t],d=1,D=2)$	0.928353	Range	6.08304	Trim Var.	0.438404
$V(Y[t],d=2,D=2)$	2.18378	Range	8.70267	Trim Var.	1.00499
$V(Y[t],d=3,D=2)$	6.29766	Range	15.074	Trim Var.	2.92183

Рисунок 7 – Матрица уменьшения дисперсии

Минимальные значения усеченной дисперсии достигнуты для параметров $d=0$ $D=1$, $d=1$ $D=1$.

Проанализируем остатки с данными параметрами дифференцирования (рисунок 8).

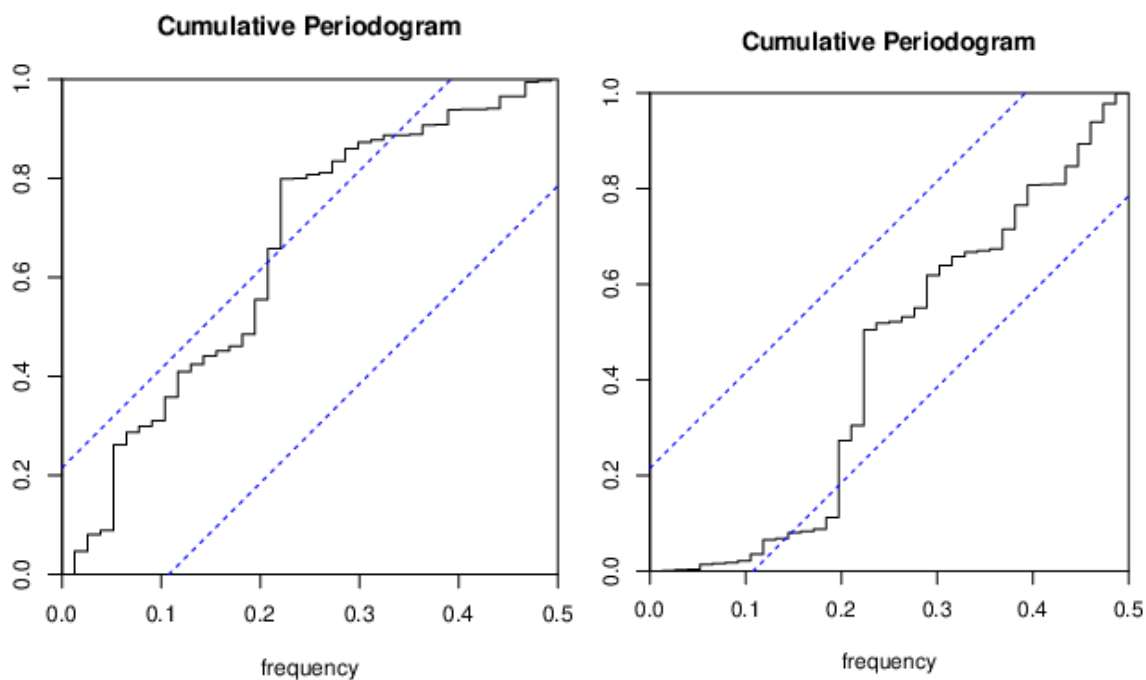


Рисунок 8 – Кумулятивная периодограмма остатков временного ряда для $(0, 1)$ и $(1, 1)$ значениях параметров (d, D)

Наилучшее приближение к линейной зависимости дает значение $d=1$ и $D=1$.

Построим ACF и PACF коррелограммы продифференцированного ряда (см. рисунок 9).

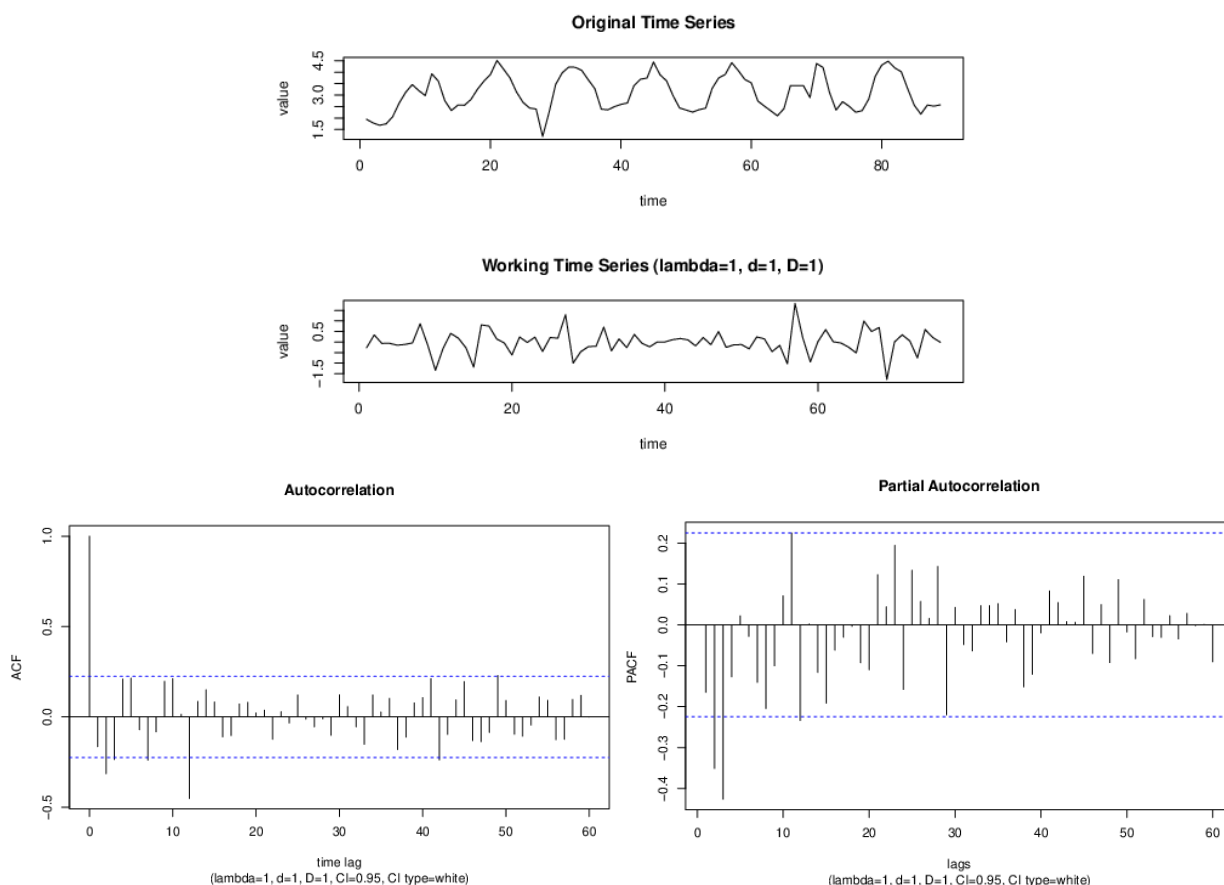


Рисунок 9 – ACF и PACF коррелограммы продифференцированного (1, 1) ряда

Полученный результат можно использовать для построения модели прогноза. Однако, необходимо отметить, что рассмотренный нами пример является примером не крупного потребителя, с явным проявлением нестабильности предприятий региона с учетом менявшейся экономической ситуации. Для крупных потребителей, у которых колебания потребления газа одним предприятием незначительно влияет на общий расход, результат может дать лучшие показатели.

Рассмотрим еще один пример (рисунки 10-12).

Variance Reduction Matrix					
$V(Y[t],d=0,D=0)$	0.15391	Range	1.76092	Trim Var.	0.102738
$V(Y[t],d=1,D=0)$	0.0694401	Range	1.82581	Trim Var.	0.0328994
$V(Y[t],d=2,D=0)$	0.135064	Range	2.83358	Trim Var.	0.0563869
$V(Y[t],d=3,D=0)$	0.422134	Range	5.55366	Trim Var.	0.157061
$V(Y[t],d=0,D=1)$	0.0959963	Range	1.92507	Trim Var.	0.0436091
$V(Y[t],d=1,D=1)$	0.0890773	Range	1.99424	Trim Var.	0.0327191
$V(Y[t],d=2,D=1)$	0.246484	Range	3.27552	Trim Var.	0.11354
$V(Y[t],d=3,D=1)$	0.81734	Range	5.62869	Trim Var.	0.342481
$V(Y[t],d=0,D=2)$	0.366218	Range	3.75067	Trim Var.	0.151088
$V(Y[t],d=1,D=2)$	0.316875	Range	3.27552	Trim Var.	0.137128
$V(Y[t],d=2,D=2)$	0.877825	Range	5.4925	Trim Var.	0.419856
$V(Y[t],d=3,D=2)$	2.97772	Range	10.0201	Trim Var.	1.29041

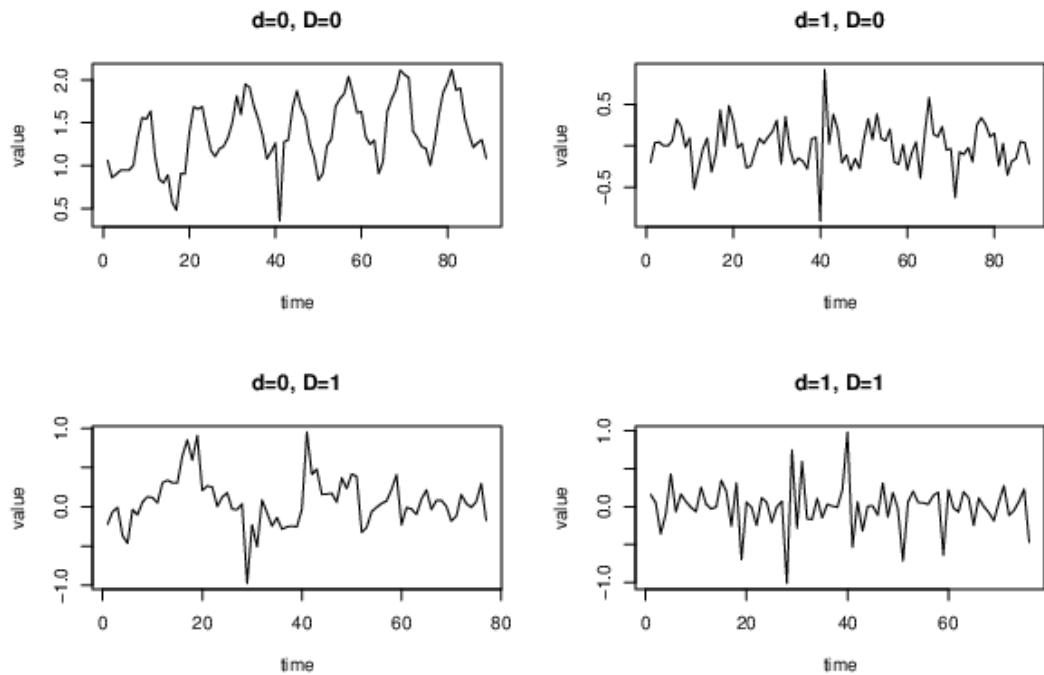


Рисунок 10 – Выбор параметров дифференцирования для крупного потребителя

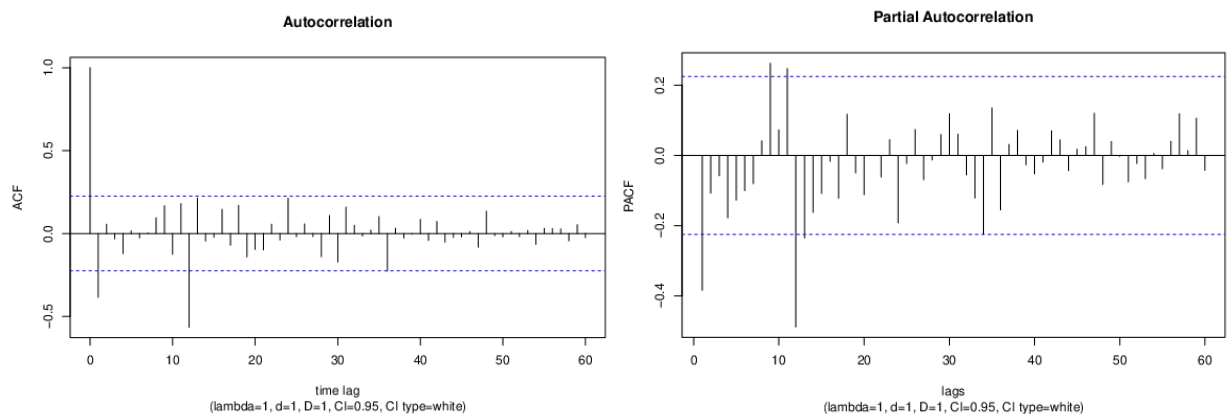


Рисунок 11 – ACF и PACF коррелограммы продифференцированного (1, 1) ряда

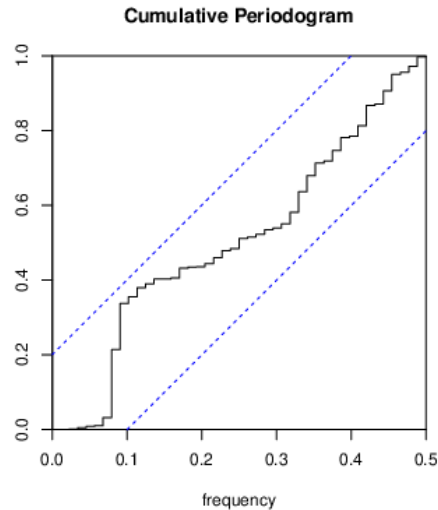


Рисунок 12 – Кумулятивная периодограмма остатков временного ряда

Выводы по главе 1

В главе 1 проанализированы особенности газотранспортной системы Республики Беларусь и исходных данных телеметрии, существенные для решения задачи прогнозирования потребления газа.

Выполнена миграция данных временных рядов о потреблении природного газа основными потребителями Республики Беларусь в реляционную нормализованную до 6NF базу данных. Построен Map-Reduce алгоритм агрегирования данных по заданному временному интервалу и сформированы данные для прогноза среднемесячного потребления газа.

ГЛАВА 2 АНАЛИЗ МЕТОДОВ И АЛГОРИТМОВ ПРОГНОЗИРОВАНИЯ ПРОЦЕССА ПОТРЕБЛЕНИЯ ГАЗА

Газораспределительные сети в настоящее время являются сложными техническими системами и включают большое число средств телеметрии и телемеханики для контроля параметров, необходимых для надежного функционирования и обеспечения безопасности процесса транспортировки газа и уменьшение затрат, связанные с техническим обслуживанием газотранспортной системы в целом. При этом важным инструментом является прогнозирование газопотребления, так как обоснованные прогнозы необходимы для оптимизации режимов транспорта газа с целью оперативного прогнозирования объемов подачи при решении задач диспетчерского управления [51, 52]. Диспетчерское управление предполагает применение планирования на день, месяц, год и включает оперативную корректировку планового режима работы газотранспортной системы при изменении заданий на транспортировку и распределение газа [53, 54].

Следует также учитывать, что в связи с непрерывным повышением уровня информатизации объектов и газотранспортной системы объемы информации, поступающие в распоряжение эксплуатирующих организаций, значительно возрастают, что позволяет принимать более обоснованные решения, но требует применения новых методов обработки, в том числе более совершенных методов прогнозирования, способных получать достоверные прогнозы на основе анализа больших объемов данных, в структуре которых имеется значительная неопределённость, т.к. процесс газопотребления зависит от большого числа случайных факторов.

Основными из них, которые, как правило, учитываются при анализе данных разделяют три типа [55]: хронологические, метеорологические и организационные.

Хронологические факторы определяют зависимость потребления газа от времени суток, от типа дня и от времени года. Метеорологические факторы влияют на потребление газа из-за изменения температуры окружающей среды. Кроме этого, на объемы потребления влияют скорость и направление ветра, влажность, а также интенсивность осадков. Однако, наибольшее влияние на спрос проявляется в периоды экстремальных изменений температуры, т.е. наибольшее их влияние проявляется в зимний период. Однако и в летние месяцы при жаркой погоде потребление газа может возрасти из-за его использования для выработки электроэнергии, необходимой для работы кондиционеров и вентиляции. Организационные факторы представляют зависимость газопотребления из-за модификации газораспределительных систем, принудительного ограничения потребителей при необходимости, выполнения ремонтных работ из-за отказов, что приводит к спаду производственной мощности объектов транспорта и распределения газа [56].

Анализ процессов потребления газа осуществляется на основе статистики объёмов поставок, при этом в качестве потребителей как правило выступают и отдельные объекты, предприятия, и агрегированные, города. Почасовое потребление газа зависит от погодных, хронологических, производственных факторов и структуры потребления, которая характеризуется энергоемкостью промышленных производств, развитием экономики региона, особенностями бытового уклада людей. Суточное потребление снижается, как правило, в ночные часы и увеличивается днем, т.к. для этого периода характерно повышение деловой активности и отопительной нагрузки. Если рассматривать недельный цикл, то, как правило, некоторое присутствие снижения спроса в субботу, воскресенье и праздники. В годовом периоде спрос на газ связан с годовым распределением температур.

Анализ типов влияющих факторов показывает, что наиболее значимыми для объема газопотребления являются температура окружающего воздуха и хронологические факторы. Как правило, для каждой группы факторов

синтезируют модель. Сезонная неравномерность газопотребления учитывается за счет включения в модель прогноза зависимости его от температуры воздуха. Однако применение простых методов построенных на статистическом анализе данных, которые эффективны для краткосрочных прогнозов, не является достаточным для получения достоверных результатов прогноза, т.к. они не учитывают специфику газопотребления. Таким образом несмотря на актуальность данной задачи она в настоящее время не решена полностью, т.к. характеризуется высокой сложностью и необходимостью применения разных методы, соответствующих поставленным целям прогнозирования и уровню заблаговременности, и позволяющим получать достоверные результаты прогноза при обработке больших объемов данных.

2.1 Представление данных

Информация об изменении состояний всей системы может быть представлена в виде последовательности значений контролируемых параметров каждого удаленного объекта в моменты времени t_1, t_2, \dots, t_j . Таким образом, последовательность значений контролируемых параметров в течение некоторого времени представляется с помощью временных рядов [56], где каждой метке времени соответствует определенное значение параметра, которое может быть измерением, отсчетом или уровнем.

Такой временной ряд характеризуется сезонностью, т.е. наличием повторяющихся с фиксированным периодом времени изменений контролируемых величин, а также отдельными пропусками переменной длительности, и в ряде случаев последовательности данных значительно зашумлены. Источниками шумов могут быть погрешности или отказы измерительного оборудования, а также утечки, потери газа при ремонтах на магистрали.

Таким образом для анализа данных применяются модели на основе временных рядов. Присутствие хаотической составляющей в исходных данных приводит к тому, что значения контролируемых параметров могут быть известны лишь с точностью до границ интервалов их изменения. Интервалы присутствуют, если учитывается погрешность измерения датчиками, или только при наличии минимальных и максимальных значения параметра за определенный интервал.

С учетом влияния основных факторов оценивают следующие основные компоненты ряда, влияющие на его уровень: тренд, сезонность и хаотическую составляющую.

2.2 Предварительная подготовка данных

Нами рассмотрены несколько вариантов построения программного обеспечения для прогноза потребления газа:

1. На основе классических методов анализа временных рядов с экспоненциальным сглаживанием;
2. С применением искусственных нейронных сетей в виде многослойных персептронов.

Учитывая выявленные особенности данных перед выполнением процесса обучения искусственной сети или перед обработкой данных методами анализа временных рядов выполняется:

1. Фильтрация грубых промахов (методом медианной фильтрации);
2. Заполнение пропусков данных методом интерполяции;
3. Удаления дублирующихся данных.

2.3 Прогнозирование потребления газа методом экспоненциального сглаживания

Экспоненциальное сглаживание является одним из приемов выравнивания ряда. Экспоненциальное сглаживание можно представить, как фильтр, на вход которого последовательно поступают члены исходного ряда, а на выходе формируются текущие значения экспоненциальной средней.

Пусть $X = \{x_1, \dots, x_T\}$ - временной ряд.

Экспоненциальное сглаживание ряда осуществляется по рекуррентной формуле:

$$S_t = \alpha x_t + (1-\alpha)S_{t-1}, \quad \alpha \in (0,1) \quad (1)$$

Чем меньше α , тем в большей степени фильтруются, подавляются колебания исходного ряда и шума.

Если последовательно использовать рекуррентное это соотношение, то экспоненциальную среднюю S_t можно выразить через значения временного ряда X .

$$S_t = \alpha x_t + (1-\alpha)(\alpha x_{t-1} + (1-\alpha)S_{t-2}) = \dots = \alpha \sum_{i=0}^{t-1} (1-\alpha)^i x_{t-i} + (1-\alpha)^t S_0 \quad (2)$$

Если к моменту начала сглаживания существуют более ранние данные, то в качестве начального значения S_0 можно использовать арифметическую среднюю всех имеющихся данных или какой-то их части.

После появления работ Р. Брауна [12, 13] экспоненциальное сглаживание часто используется для решения задачи краткосрочного прогнозирования временных рядов.

Для построения алгоритма прогноза на основе метода экспоненциального сглаживания нами применен ETS алгоритм в версии AAA. [14, 15]

Данное семейство методов выявляет адаптивные сезонные тренды, что позволяет достичь высокой точности и долгосрочных прогнозов. Ключевым отличием данного семейства методов является построение временных иерархий данных. [15]

Исследования способов агрегации прогнозов для узлов различных уровней временной иерархии (иерархии сезонных интервалов) опубликованы в целом ряде работ последнего времени. [16-20]

Недостаток использования единого иерархического уровня для расчета прогнозов состоит в том, что данные на определенном уровне могут быть слишком зашумленными или слишком агрегированными, чтобы предоставлять полезную информацию для целей прогнозирования. Кроме того, сосредоточение внимания на одном уровне агрегации не позволяет в полной мере использовать информацию, предоставляемую из других узлов иерархии. Исследования, сравнивающие восходящие и нисходящие подходы, не предоставили убедительные аргументы в отношении превосходства одного метода над другим. Подход «снизу-вверх» принимался многими авторами как фактический подход во временном измерении, причем прогнозы составлялись с частотой, с которой сбора данных.

Комбинация точечных прогнозов по различным перекрестным [21, 22], временным [15, 23], а также кросс-временным [24–26] уровням агрегации широко изучалась в литературе.

Два важных вывода из этого потока исследований:

- точность прогноза улучшается, если прогнозы с разных уровней агрегации надлежащим образом комбинируются, и
- согласованные прогнозы по уровням агрегации позволяют принимать согласованные решения.

2.4 Прогнозирование на основе многослойного персептрона

Для решения задачи предсказания потребления газа на основе нейронной сети многослойного персептрона (MLP – Multi-layer Perceptron) нами построен контролируемый алгоритм обучения, который изучает функцию $f(\cdot): R^m \rightarrow R^0$ путем обучения на наборе данных. Опираясь на набор функций $X = x_1, x_2, \dots, x_m$ и цель y , он может обучить аппроксиматор нелинейных функций для классификации или регрессии. Он отличается от логистической регрессии тем, что между входным и выходным слоями может быть один или несколько нелинейных слоев, называемых скрытыми слоями. На рисунке 13 показан MLP с одним скрытым слоем и скалярным выводом.

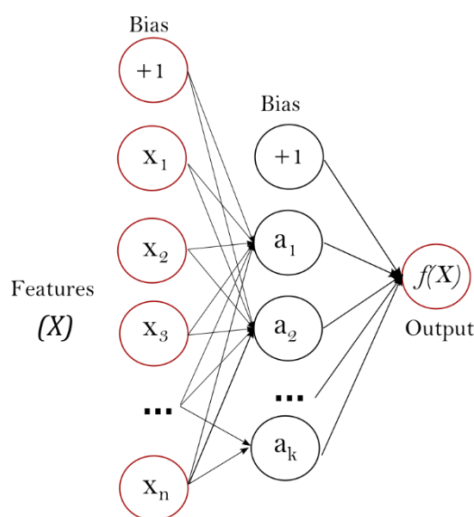


Рисунок 13 – искусственная нейронная сеть многослойного персептрона

Входной слой, состоит из набора нейронов $\{x_i | x_1, x_2, \dots, x_m\}$ представляет входные функции. Каждый нейрон в скрытом слое преобразует значения из предыдущего слоя с помощью взвешенного линейного суммирования $w_1x_1 + w_2x_2 + \dots + w_mx_m$ последующей нелинейной функцией активации $g(\cdot): R \rightarrow R$ - гиперболическая функция. Выходной слой получает значения из последнего скрытого слоя и преобразует их в выходные значения.

Реализация алгоритмов выполнена на языке программирования Python.

Конфигурация проекта: язык программирования Python 3.7, среда разработки PyCharm, интерпретатор Anaconda.

Реализация выполнена с использованием библиотеки `sklearn.neural_network.MLPRegressor`.

2.5 Оценка качества прогноза потребления газа

Для оценки качества прогноза мы использовали оценку среднеквадратического отклонения MSE и коэффициент детерминации R-квадрат.

Коэффициент детерминации – это доля дисперсии зависимой переменной, объясняемая рассматриваемой моделью. Более точно – это единица минус доля необъяснённой дисперсии (дисперсии случайной ошибки модели, или условной по признакам дисперсии зависимой переменной) в дисперсии зависимой переменной.

$$R^2 = 1 - \frac{\hat{\sigma}^2}{\sigma_y^2} = 1 - \frac{RSS/n}{TSS/n} = 1 - \frac{RSS}{TSS}, \quad \text{где} \quad (3)$$

$$RSS = \sum_{t=1}^n e_t^2 = \sum_{t=1}^n (y_t - \hat{y}_t)^2 \quad \text{- сумма квадратов регрессионных остатков,}$$
$$TSS = \sum_{t=1}^n (y_t - \bar{y})^2 = n \sigma_y^2 \quad \text{- общая дисперсия.}$$

В зависимости от уровня коэффициента детерминации, принято разделять модели на три группы:

- 0,8 – 1 — модель хорошего качества;
- 0,5 – 0,8 — модель приемлемого качества;
- 0 – 0,5 — модель плохого качества.

Среднеквадратическая ошибка (MSE) – среднеквадратическая ошибка прогноза применяется в ситуациях, когда нам надо подчеркнуть большие ошибки и выбрать модель, которая дает меньше больших ошибок прогноза.

$$MSE = \frac{RSS}{n}, \text{ где} \quad (4)$$

$$RSS = \sum_{t=1}^n e_t^2 = \sum_{t=1}^n (y_t - \hat{y}_t)^2 \quad \text{– сумма квадратов регрессионных остатков.}$$

Критерий корреляции Пирсона – это метод параметрической статистики, позволяющий определить наличие или отсутствие линейной связи между двумя количественными показателями, а также оценить ее тесноту и статистическую значимость. Другими словами, критерий корреляции Пирсона позволяет определить, есть ли линейная связь между изменениями значений двух переменных. В статистических расчетах и выводах коэффициент корреляции обычно обозначается как r_{xy} или R_{xy} .

Именно этот метод был нами использован, для подсчёта оценки коэффициента детерминации.

Например, результат прогнозирования проиллюстрирован на рисунках 14, 15.

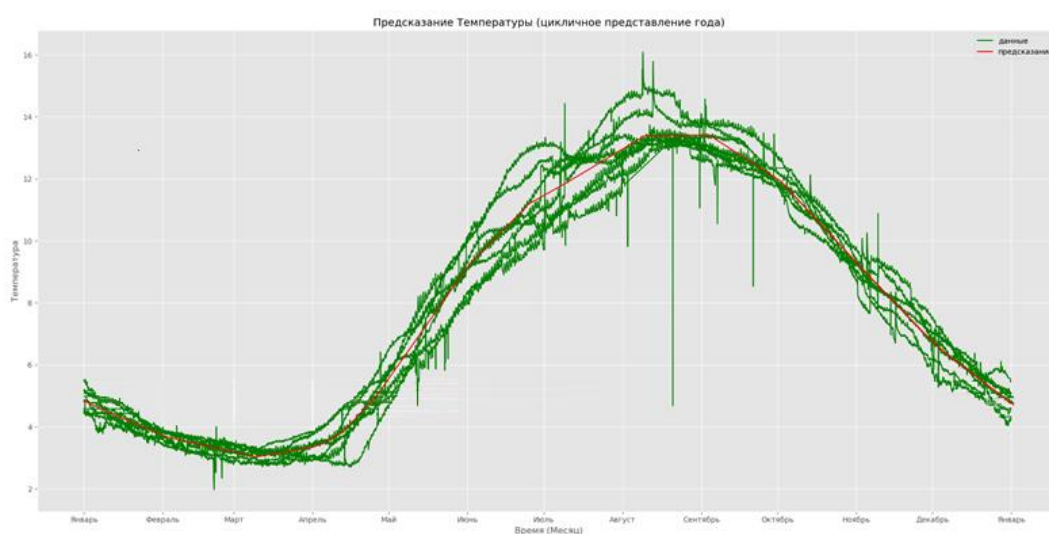


Рисунок 14 – прогноз изменения температуры грунта на 2019г

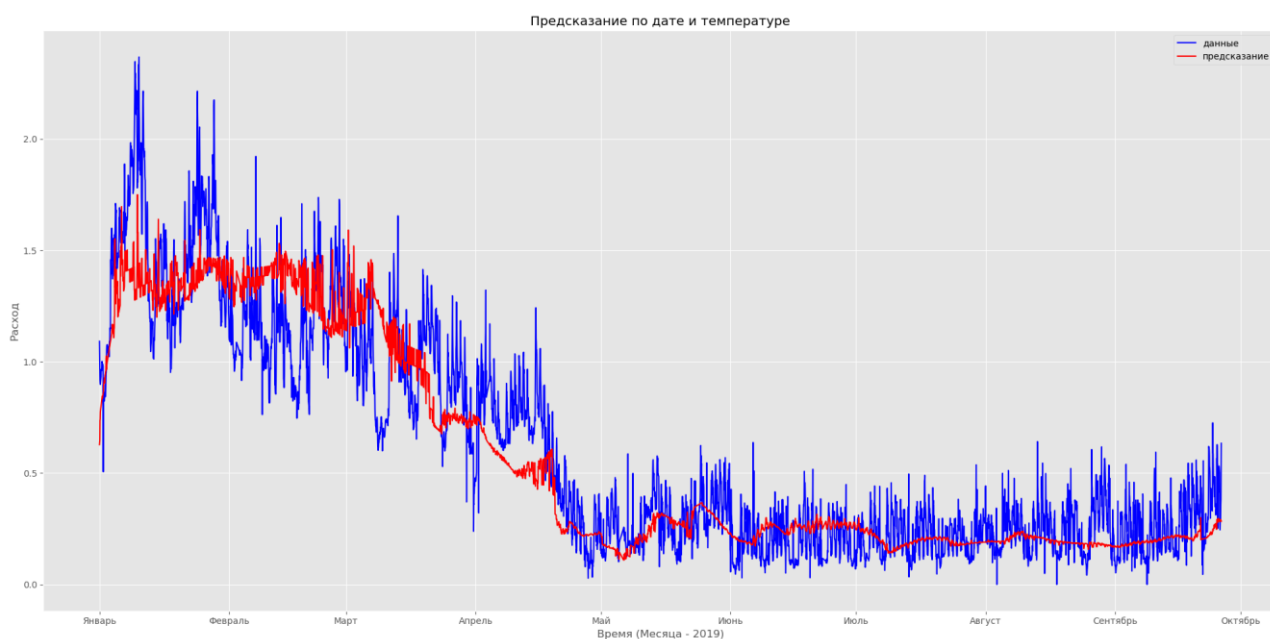


Рисунок 15 – Прогноз потребления газа для единичного потребителя на 2019г

Таблица 1 – Сравнение результатов прогнозирования методами MLP и ETS

Метод	MSE	R ²	Победитель
MLP	0,000349957	0,817151858	MLP
ETS	0,000500124	0,801648983	
MLP	0,000497443	0,895402782	MLP
ETS	0,00070938	0,89313069	
MLP	0,002197816	0,847362195	ETS
ETS	0,001684928	0,896444938	
MLP	0,000715583	0,566817257	ETS
ETS	0,00029087	0,85381516	
MLP	0,041965009	0,830801338	MLP
ETS	0,043377925	0,84355509	
MLP	0.000741670	0,822294676	MLP
ETS	0,000988042	0,838613686	
MLP	0,000344016	0,918363176	MLP
ETS	0,000424404	0,91499993	

2.6 Обзор перспективных нейросетевых методов прогнозирования

Прогнозирование газопотребления выполняется на основе многих методов, включая, метод скользящего среднего, метод экспоненциального сглаживания, авторегрессионные модели, а также на основе искусственных нейронных сетей, обеспечивая различную результативность.

Одной из распространенных и наиболее используемой, в том числе и для прогнозирования потребления газа, является модель ARIMA (Autoregressive moving average), представляющая интегрированную модель авторегрессии скользящего среднего и по-другому называемая моделью Бокса-Дженкинса [60]. Модель применяется к нестационарным временным рядам, которые можно сделать стационарными, путем взятия разности некоторого порядка от исходных значений временного ряда [61]. Для нелинейных моделей характерно представление предыдущих отсчетов в виде определенной структуры и некоторых правил перемещения внутри нее. Благодаря стремительному росту вычислительной мощности компьютерной техники в последние десятилетия совместно с традиционными методами все шире разрабатываются и применяются методы, использующие машинное обучение, искусственные нейросетевые модели [62,63]. Под нейросетевыми моделями понимается совокупность нейронных элементов и связей между ними.

Искусственные нейронные сети строятся на основе нейронов, получающих и преобразующих информацию из нескольких источников по определенному правилу. На рисунке 16 представлена модель нейрона, в которой каждый вход x_i взвешивается определенным значением w_i , затем выполняется суммирование, а суммарный выходной сигнал преобразуется в соответствии с используемой функцией активации. Данная функция как правило, должна быть нелинейной [64], дифференцируемой и применимой для обучения нейронной сети, основные типы которых приведены в таблице 2.

Таким образом, получив набор чисел в качестве входных значений, нейрон формирует некоторое число на выходе.

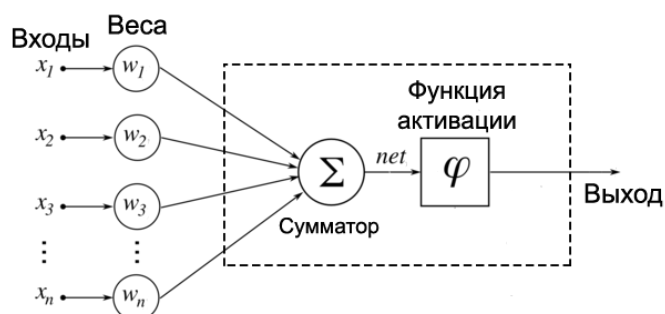


Рисунок 16 – Модель нейрона

С использованием нейронных сетей возможно моделирование нелинейной зависимости последующего значения временного ряда от его предыдущих, а также с учетом внешних факторов. Для прогноза временных рядов различными авторами предлагалось использовать полносвязные нейронные сети прямого распространения [65], мультивейвлетные полиморфные сети [66], машины экстремального обучения [67] и сверточные нейронные сети [68].

Таблица 2 – Основные типы функций активации

Наименование функции	Математическое описание функции
Логистический сигмоид	$\frac{1}{1 + e^{-x}}$
Гиперболический тангенс	$\frac{e^x - e^{-x}}{e^x + e^{-x}}$
Функция Хевисайда	$\begin{cases} 0, & \text{if } x < 0; \\ 1, & \text{if } x \geq 0. \end{cases}$
SoftPlus	$\log(1 + e^{-x})$
ReLU	$\begin{cases} 0, & \text{if } x < 0; \\ x, & \text{if } x \geq 0. \end{cases}$
Parameterized ReLU	$\begin{cases} \alpha x, & \text{if } x < 0; \\ x, & \text{if } x \geq 0. \end{cases}$
ELU	$\begin{cases} \alpha(e^x - 1), & \text{if } x < 0; \\ x, & \text{if } x \geq 0. \end{cases}$

Среди возможных реализаций нейронных сетей для прогнозирования значительное распространение получили модели многослойных персептронов (MLP – multilayer perceptron), которые включают три типа слоев входной (input), скрытый (hidden), выходной (output) (рисунок 17).

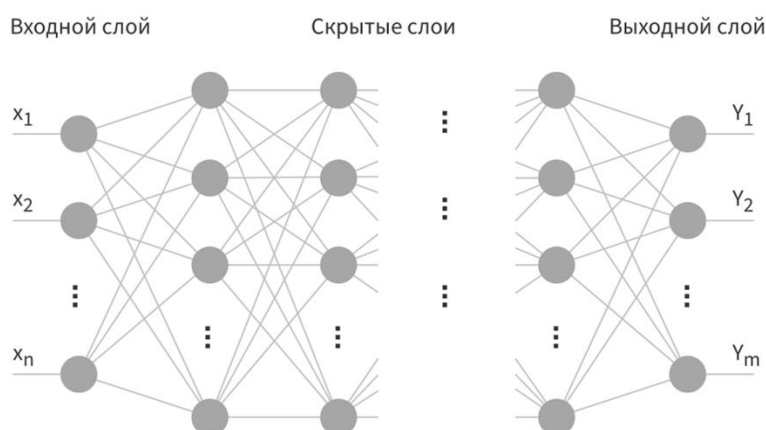


Рисунок 17 – Архитектура многослойной нейронной сети

В данной модели отношения между выходом и входами можно представить с помощью соотношения:

$$y_t = a_0 + \sum_{j=1}^q a_j \cdot g \left(\beta_{0j} + \sum_{i=1}^p \beta_{ij} y_{t-i} \right), \quad (5)$$

где: a_0 - вес отношений между входными нейронами и выходом; β_{0j} характеризуют веса отношений между входными и скрытыми нейронами; a_i, β_{ij} - веса, ассоциируемые с каждым узлом сети; p - количество входных нейронов, определяющих глубину истории временного ряда; q - количество скрытых нейронов; g - нелинейная функция активации, используемая для перехода скрытого слоя.

Эффективность работы данной модели определяется оптимальностью весовых значений, которые устанавливаются в процессе обучения. Для этого,

как правило, используется метод обратного распространения ошибки, с помощью которого весовые коэффициенты рассчитываются путем минимизации среднеквадратичной ошибки с использованием метода градиентного спуска [69].

Для обучения должна быть подготовлена выборка, состоящая из обучающих примеров, формирование которых основывается на исходных данных и специфике решаемой задачи. Для экстраполяции временного ряда исходные данные преобразуются в обучающую выборку с использованием метода скользящих окон [70, 71]. При этом каждый обучающий пример выборки формируется из n значений входного ряда, называемого «окном». Очередной обучающий пример строится за счет сдвига окна на одну единицу времени. Для прогнозирования архитектура нейросетевой системы включает: входной слой, который содержит несколько нейронов и последний слой состоит из единственного нейрона, на выходе которого формируется прогнозируемое значение [72]. Модель MLP на практике применима для прогнозирования технического состояний оборудования [73] авторы использовали сети MLP для диагностики неисправностей подшипников качения. В [84] предлагается метод на основе такого типа сети с предварительной вейвлет-обработкой исходных данных для эффективного мониторинга процессов в информационно-управляющих системах.

MLP характеризуется следующими ограничениями: входные и выходные массивы должны иметь фиксированный, заранее определенный размер, реализуется запоминание небольшого количества предшествующих наблюдений, что не подходит для длинных периодов, каждый пакет входных данных обрабатывается независимо, без сохранения состояния между входом и выходом. Все это приводит к основному недостатку, которым является достаточно быстрое накопление ошибки.

Для установления зависимости между обрабатываемыми пакетами данных были предложены разновидности рекуррентных нейронных сетей

(recurrent neural network – RNN), реализующие обработку временного ряда последовательно с использованием информации, полученной от предыдущих его элементов [85]. Типичная архитектура RNN показана на рисунке 18. Слева показано обозначение свернутой сети, а справа представлена развернутая схема. Главными особенностями RNN являются возможность использовать начальную информацию для текущей задачи. Кроме этого, возможно подавать на вход значения рекурсивно, что и обеспечивает в значительной мере оптимизацию параметров и улучшение качества результата.

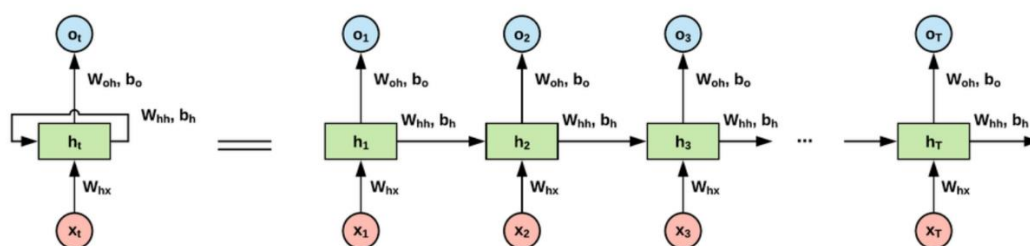


Рисунок 18 – Типичная архитектура RNN

В отличие от обычных нейронных сетей с прямой связью, LSTM (long short-term memory, долговременно-кратковременная память) сети имеют обратные связи [86, 87]. Благодаря этому они способны эффективно обрабатывать последовательности данных. Поэтому LSTM нейронные сети хорошо подходят для классификации, обработки и построения прогнозов на основе временных рядов, где взаимосвязанные явления могут происходить с неопределенным временным лагом, который приводит к затруднениям в традиционных нейронных сетях.

В традиционных рекуррентных нейронных сетях для обратной связи используется комбинация скрытого состояния на предыдущем шаге и текущих входных данных в слое с нелинейной функцией активации, например, такой, как гиперболический тангенс. В LSTM-сети обратная связь реализуется аналогично, но нейронных слоев используется не один, а четыре.

Рисунок 19 поясняет работу ячейки LSTM: на горизонтальной верхней линии показано состояние ячейки c , которое представляет внутреннюю память ячейки; на нижней линии показано скрытое состояние; i, f, o, g -предназначены для решения проблемы затухания градиента и их значения определяются в процессе обучения сети.

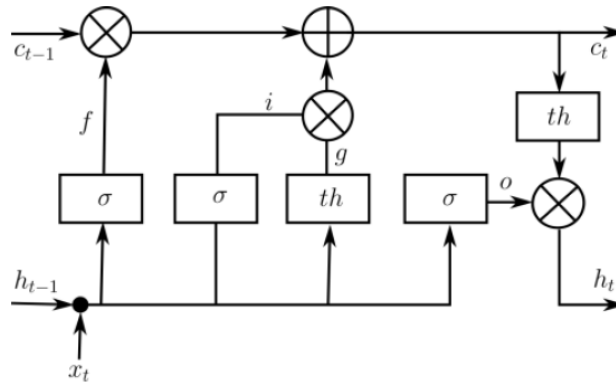


Рисунок 19 – Схема ячейки LSTM

Традиционно, у каждой ячейки сети имеется три дополнительных фильтра: входной – определяет долю информации, поступающей с предыдущей итерации; выходной – задает количество информации, которое поступит на следующий слой; забывания – определяет информацию, которую на текущем шаге нужно игнорировать.

Модель LSTM можно представить в виде [88]:

$$\begin{aligned}
 f_t &= \sigma_g(W_t x_t + U_f h_{t-1} + b_f), \\
 i_t &= \sigma_g(W_t x_t + U_i h_{t-1} + b_i), \\
 o_t &= \sigma_g(W_o x_t + U_o h_{t-1} + b_o), \\
 c_t &= f_t \circ c_{t-1} + i_t \circ \sigma_c(W_c x_t + U_c h_{t-1} + b_c), \\
 h_t &= o_t \circ \sigma_h(c_t),
 \end{aligned}
 \tag{6}$$

где: x_t - входной вектор; h_t - выходной вектор; c_t - вектор состояний; W, U, b - матрицы и вектор параметров; f_t, i_t, o_t - векторы вентилей (забывания,

входных и выходных данных); $\sigma_g, \sigma_c, \sigma_h$ - функции активации; « \circ » - произведение Адамара.

Вначале данные поступают на фильтр забывания, где часть из них отбрасывается. Затем с помощью входного фильтра определяется, какие данные нужно сохранить в состоянии ячейки. Результат, полученный на предыдущих этапах, записывается в состояние ячейки и проходит через выходной фильтр, формируя итоговое значение на выходе ячейки.

В [89] предлагается построение модели на базе сети LSTM с целью прогнозирования по аналогии с моделью MLP с помощью оконного метода. Для этого на вход модели (рисунок 20), поступает вектор, состоящий из чередующихся значений верхних и нижних границ интервала в последовательные моменты времени в течение определенного периода.

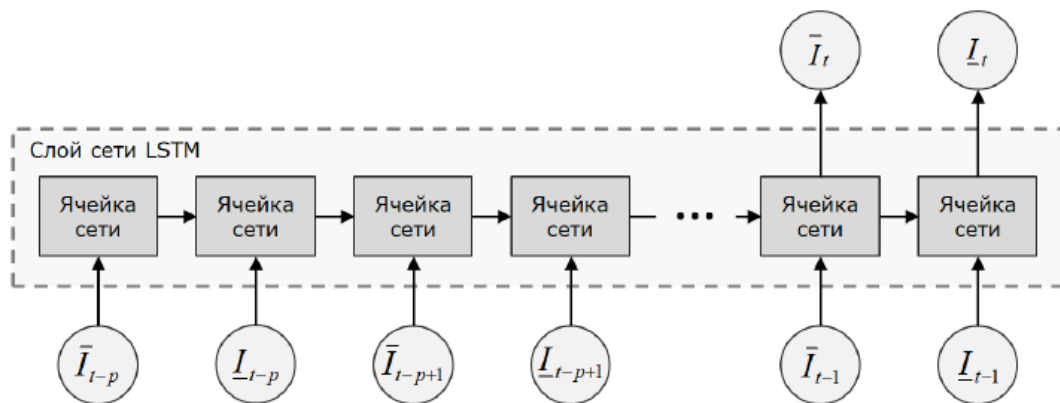


Рисунок 20 – Структура модифицированной модели на базе сети LSTM

Для данной структуры модели с количеством ячеек сети LSTM определяется длиной входного вектора. На вход модели, аналогично модели MLP для интервальных временных рядов, подаются данные с шириной окна, равной периоду времени в одни сутки. Для прогнозирования применительно к длине $2p$ исходного вектора применяется входной слой с числом нейронов, равным $2p$, слой с $q = 2p$ скрытыми ячейками и слой с двумя выходными нейронами, которые необходимы для оперирования значениями \bar{I}_t и \underline{I}_t :

$$X_t = \left\| \bar{I}_{t-1}, \underline{I}_{t-1}, \bar{I}_{t-2}, \underline{I}_{t-2}, \dots, \bar{I}_{t-p}, \underline{I}_{t-p} \right\|^T$$

Выходной вектор формируется после прохождения данных через все ячейки сети и включает информацию о верхней и нижней границах прогнозируемого интервала. Для обучения модели предлагается применять метод обратного распространения с разворачиванием сети во времени ВРТТ (backpropagation through time), который представляет собой модификацию метода градиентного спуска [89, 90]. Выполненные автором эксперименты показали, что предложенная модификация модели на базе LSTM формирует прогнозы приемлемой точности, погрешность около 1,8%, не накапливая ошибки с увеличением глубины прогноза и обеспечивая попадание в пределы интервалов реальных показаний.

В работе [91] представлены результаты нейросетевого прогнозирования расхода газа на основе реальных данных потребления а, предоставленных Харцызским управлением по газоснабжению и газификации в виде значений ежемесячного потребления газа городом за период 1998 – 2009 гг. В качестве типа нормировки использовалась линейная, которая приводит все значения к единому диапазону [0,1]. В качестве модели выбрана MLP-сеть с архитектурой, состоящая из входного слоя, имеющего 4 нейрона, скрытого слоя, состоящего из 7 нейронов и выходного слоя. Функция активации для на скрытом слое - гиперболический тангенс, на выходном слое - линейная. Эксперименты проведены с использованием пакета Matlab 7.0, а именно пакета neural network toolbox.

Данные потребления газа за 12 лет распределены были как: за 11 лет использованы для обучения, а за последний год использовались в качестве проверочной выборки. После обучения нейронной сети выполнено ее тестирование, результаты представлены на рисунке. Эксперименты показали, что прогноз на первые три месяца выполнен существенно точнее, чем на год,

причем ошибка прогнозирования на ближайшие три месяца составляет менее 1 %, т.е. предложенный подход удовлетворителен только для краткосрочного прогнозирования потребления газа.

Слой управляемых рекуррентных блоков (Gated Recurrent Unit, GRU) использует принцип слоя LSTM, однако рекуррентные блоки представляют собой более простые структуры (рисунок 21) по сравнению с LSTM, что приводит к уменьшению объема вычислений и обучается он быстрее.

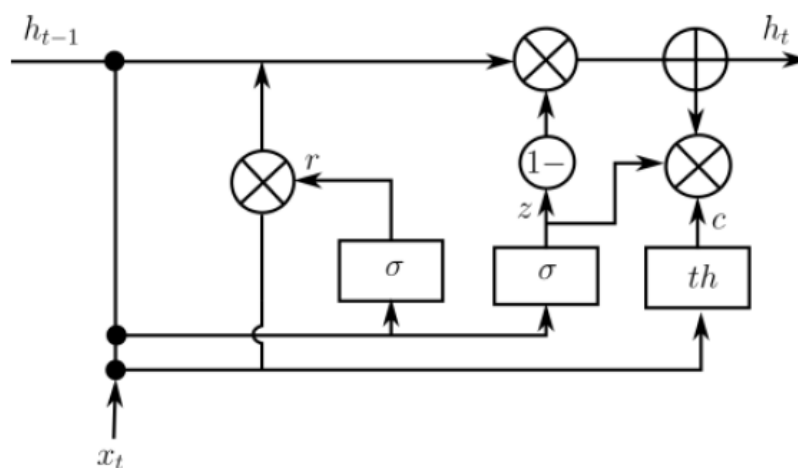


Рисунок 21 – Схема ячейки GRU

Несмотря на то, что GRU эффективнее по вычислительным затратам из-за меньшего числа вентиляей, но данная модель все равно отнесена на второе место после LSTM в плане результативности получения выходных значений [92]. Поэтому модель GRU целесообразно использовать в случае, когда необходимо быстро обучить модель при небольших вычислительных мощностях.

Выводы по главе 2

На основе анализа прогнозирования расхода газа в газотранспортных системах с использованием современных методов прогнозирования можно выделить следующие аспекты:

- данные, собранные датчиками телеметрии, характеризуются сложной структурой: многомерные временные ряды с большим числом показателей; большие объемы данных; пропущенные измерения; нестационарный шум; иерархическая структура.

- первичная подготовка данных во многом определяет результаты прогнозирования;

- точность прогнозов во многом зависит не только от применяемых методов, но и от структуры исходных данных, имеющегося набора для исследования и требуемой глубины прогноза.

С целью построения моделей для прогнозирования потребления природного газа нами выбраны модели ARIMA, ETS и модель, построенная на основе искусственных нейронных сетей (многослойного персептрона).

В качестве моделей для дальнейшего исследования целесообразно рассматривать модели RNN и LSTM сетей.

ГЛАВА 3 РАЗРАБОТКА И СРАВНИТЕЛЬНОЕ ИССЛЕДОВАНИЕ МЕТОДОВ ПРОГНОЗИРОВАНИЯ ПОТРЕБЛЕНИЯ ПРИРОДНОГО ГАЗА

3.1 Влияние нестационарных процессов на исходные данные

Поскольку модели прогнозирования потребления природного газа используют не только данные телеметрии, но и в ряде случаев значения расходов, определенное расчетным путем для узлов, граничные условия в которых заданы парой давление и температура, то мы попытались устранить влияние на оценку расхода нестационарных процессов. В результате корреляционного анализа нами были выявлены проблемы стационарной модели, которые можно сформулировать следующим образом: при включении компрессорной станции начинается подогрев газа и, несмотря на то, что нагретый газ еще не распространился по трубопроводу, стационарная модель описывает такое состояние трубопровода, которое возникнет после перехода в установившейся режим. Таким образом, существенные участки трубопровода в модели оказываются с завышенным значением температуры газа

и, учитывая сильную зависимость плотности газа от температуры, с заниженным значением запаса газа.

Мы рассчитали коэффициент корреляции, который для рассматриваемых участков с ростом температуры показал высокую отрицательную корреляцию. Так, для приведенных на рисунке 22 в качестве примера графиках коэффициент корреляции составил $\text{Correl}(X, Y) = \{ -0,84369, -0,88839, -0,88509, -0,85517 \}$.

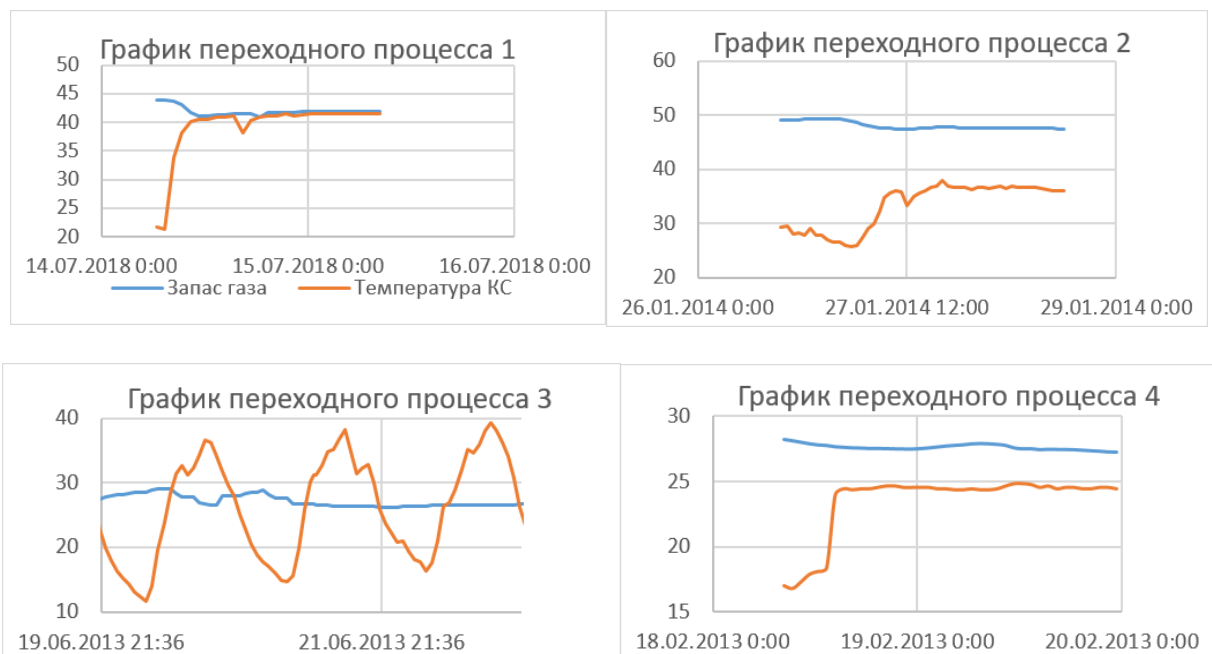


Рисунок 22 – Примеры графиков переходных процессов
в рамках модели магистрального газопровода ОАО «Газпром трансгаз
Беларусь»

3.2 Разработка генетического алгоритма аппроксимации многомерных зависимостей

Под сложностью многомерной зависимости мы будем понимать неопределенность относительно присутствующих в данных функциональных зависимостей между входными переменными и выходным значением функции. Сложные многомерные зависимости встречаются, в частности, в таких задачах, как:

- идентификация параметров аналитических моделей аппроксимации многомерных зависимостей;
- идентификация параметров теоретических вероятностных распределений;
- идентификация параметров моделей трендов при прогнозировании.

Аппроксимация данных в условиях отсутствия априорной информации о реальной модели требует применения универсальных аппроксиматоров, таких как искусственные нейронные сети, нечеткие логические аппроксиматоры. Однако аналитическое представление модели является наиболее компактным и наиболее точным. Методы аппроксимации, основанные на идентификации параметров некоторой теоретической модели, решают задачу, используя неизменяемую модель, принимаемую из каких-либо предположений разумности.

Нами делается попытка применения генетического алгоритма для поиска наиболее точной аналитической модели аппроксиматора. Генетический алгоритм выполняет не только идентификацию параметров теоретической модели, но и поиск самой модели в пространстве возможных модификаций структуры. Целенаправленное движение популяции моделей к моделям, обеспечивающим качественную аппроксимацию, осуществляется благодаря закону эволюционного отбора по критерию минимизации среднеквадратического отклонения (СКО).

Генетический алгоритм (ГА) – это эвристический алгоритм поиска, используемый для решения задач оптимизации и моделирования путём последовательного подбора, комбинирования и вариации искомым параметров с использованием механизмов, напоминающих биологическую эволюцию.

Ранее нами были предприняты попытки решения данной задачи на многоагентных системах, которые для аналитических моделей в пространстве 3-х переменных показали результаты, превосходящие результаты искусственных нейронных сетей, но на ограниченном разнообразии моделей [1, 2, 3]. Основным недостатком таких аппроксиматоров является ограниченный набор преобразований структуры, сформулированный в виде локальных эвристик. Эвристики являются очевидными для человека-эксперта способами изменения структуры. Соответственно неочевидные для эксперта

модификации становятся невозможными. На наш взгляд, применение ГА снимает данное ограничение.

Кроме этого нами были построены аппроксиматоры на нечеткой логике (НЛА). Нечеткие логические системы, предложенные Лофти Заде [8], получили развитие в работах [7, 9]. Если рассматривать неизвестный параметр как непрерывный, то в этом случае можно провести параллель между выводом о значении неизвестного параметра и приближением функции и говорить о свойстве нечеткой системы выступать в роли лингвистического аппроксиматора. Преимуществом таких аппроксиматоров является их быстродействие, а основным недостатком – ограниченные возможности по повышению точности аппроксимации.

Контекстно-свободная грамматика G_{exp} класса LR(1) алгебраических выражений, предложенная нами, представлена ниже в форме Бекуса-Наура:

```

formula:      exp    {ANode* o = (ANode*)$1; createdNodes.push_back(o);}
|
formula ';' exp
              {ANode* o = (ANode*)$3; createdNodes.push_back(o);}
;

exp :  NUM
      {ANodeNUM* o = new ANodeNUM($1); $$ = (void*) o;}
|
  '(' exp ')'      {$$ = $2;}
|
  exp '+' exp
      {ANode* o = (ANode*) new ANodeOp((ANode*)$1, (ANode*)$3, PLUS); $$ = (void*) o;}
|
  exp '-' exp
      {ANode* o = (ANode*) new ANodeOp((ANode*)$1, (ANode*)$3, MINUS); $$ = (void*)
o;}
|
  exp '/' exp
      {ANode* o = (ANode*) new ANodeOp((ANode*)$1, (ANode*)$3, DIV); $$ = (void*) o;}
|
  exp '*' exp
      {ANode* o = (ANode*) new ANodeOp((ANode*)$1, (ANode*)$3, MUL); $$ = (void*) o;}
|
  'p' 'o' 'w' '(' exp ',' exp ')'

```

```

    {ANode* o = (ANode*) new ANodeOp((ANode*)$5, (ANode*)$7, POW); $$ = (void*) o;}
|
's' 'i' 'n' '(' exp ')'
    {ANode* o = (ANode*) new ANodeFunc((ANode*)$5, SIN); $$ = (void*) o;}
|
'c' 'o' 's' '(' exp ')'
    {ANode* o = (ANode*) new ANodeFunc((ANode*)$5, COS); $$ = (void*) o;}
|
'l' 'o' 'g' '(' exp ')'
    {ANode* o = (ANode*) new ANodeFunc((ANode*)$5, LOG); $$ = (void*) o;}
|
'e' 'x' 'p' '(' exp ')'
    {ANode* o = (ANode*) new ANodeFunc((ANode*)$5, EXP); $$ = (void*) o;}
|
't' 'a' 'n' '(' exp ')'
    {ANode* o = (ANode*) new ANodeFunc((ANode*)$5, TAN); $$ = (void*) o;}
|
'c' 't' 'a' 'n' '(' exp ')'
    {ANode* o = (ANode*) new ANodeFunc((ANode*)$6, CTAN); $$ = (void*) o;}
|
NAME
    {ANode* o = (ANode*) new ANodeNAME(string((char*)$1)); $$ = (void*) o; delete $1;}
;

```

Как мы видим, в процессе синтаксического разбора строится дерево правостороннего вывода в виде объектно-ориентированной структуры. Для получения возможности компиляции алгебраического выражения в дерево вложенных друг в друга объектов нами разработана система классов, представленная на диаграмме классов (рисунок 23).

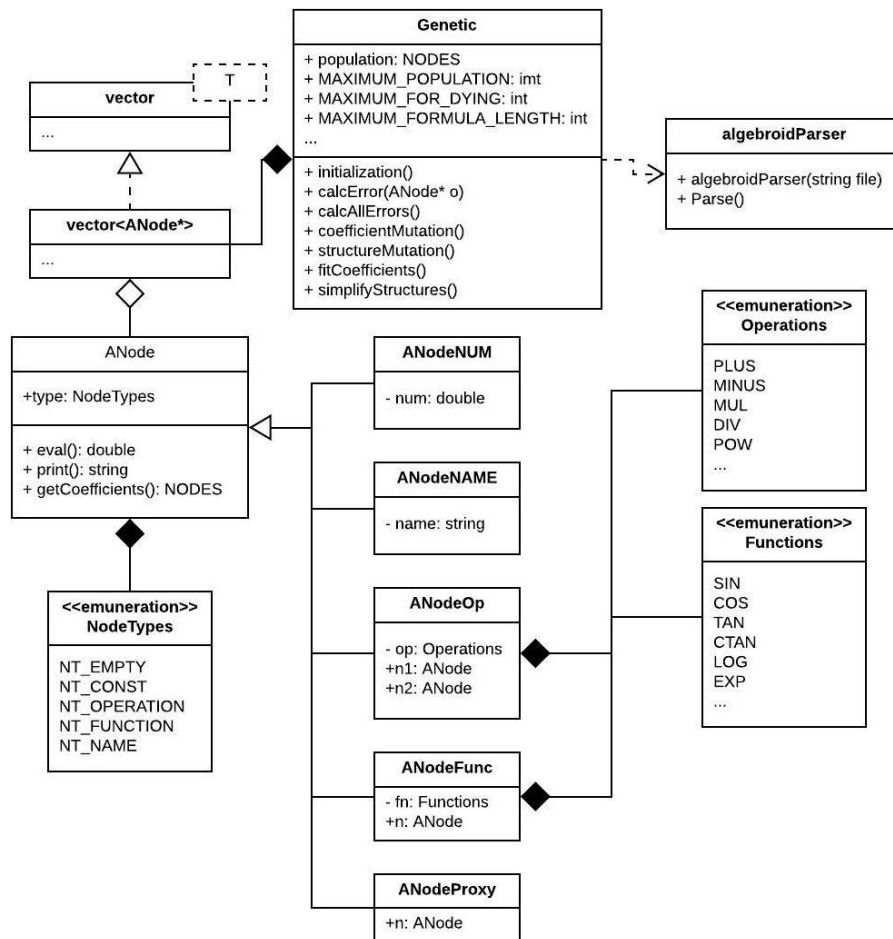


Рисунок 23 – Диаграмма классов программного обеспечения ГА поиска модели аналитического n -мерного аппроксиматора

В результате синтаксического разбора файла инициализации создается начальная популяция аналитических n -мерных моделей. Начальные варианты моделей формулируются экспертом на формальном языке грамматики G_{exp} .

Объекты базового класса ANode и производных классов ANodeNUM, ANodeNAME, ANodeOp, ANodeFunc, ANodeProxy атрибутируют узлы семантического стека и формируют дерево правостороннего разбора алгебраического выражения.

Класс Genetic реализует основные этапы ГА:

- 1) initialization() – инициализация популяции экспертно-заданными для случая n -переменных алгебраическими выражениями;

2) `calcAllErrors()` – вычисление ошибки аппроксимации для каждой особи в ее текущем состоянии;

3) `coefficientMutation()` – процесс случайного клонирования особей с мутацией коэффициентов для обеспечения возможности выхода из локальных экстремумов;

4) `structureMutation()` – процесс клонирования с мутацией структуры, при условии положительного эффекта от мутации;

5) `fitCoefficients()` – процесс поиска экстремума целевой функции в пространстве коэффициентов. В реализации данного процесса нами выбран алгоритм градиентного спуска, но с применением координатной релаксации.

6) `simplifyStructure()` – упрощает структуру выражения в специфических случаях.

Упрощение структуры приводит к удалению и замене узлов семантического дерева (именно для процедуры упрощения предусмотрен класс `ANodeProxy`) и предусмотрено в следующих случаях:

$$x * 0 \mid 0 * x \rightarrow 0;$$

$$x^0 \rightarrow 1; \quad x^1 \rightarrow x;$$

$$x * 1 \mid 1 * x \rightarrow x;$$

Упрощение срабатывает, когда алгоритм обнаруживает сходимость коэффициента к 0 или 1. Блок-схема алгоритма эволюционного поиска представлена на рисунке 24.

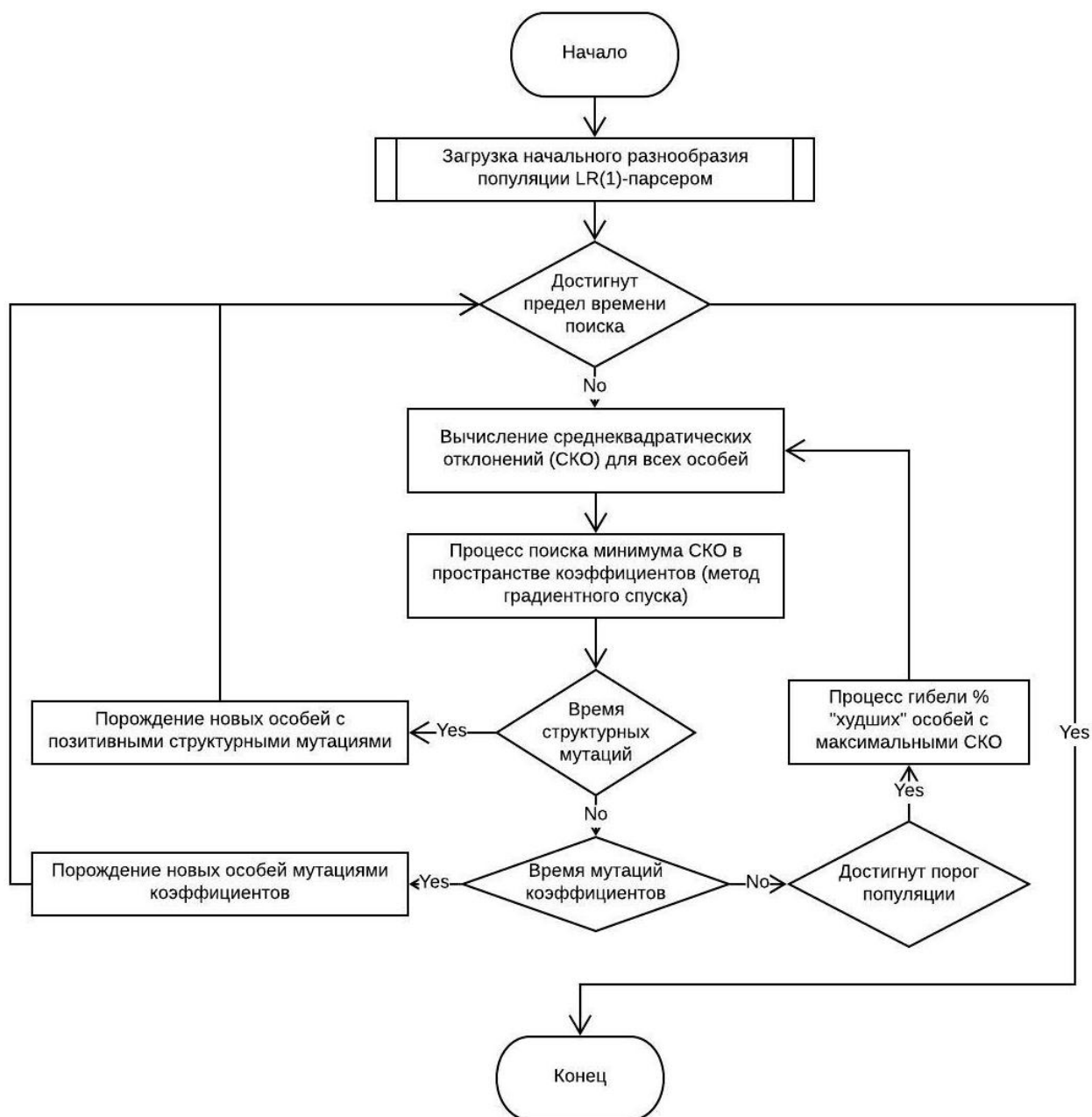


Рисунок 24 – Блок-схема алгоритма эволюционного поиска разработанного генетического алгоритма

В качестве целевой функции нами выбрано среднеквадратичное отклонение между теоретическими и расчетными значениями функции. В m -мерном пространстве коэффициентов аналитических моделей поверхность СКО может иметь локальные экстремумы в связи с ограниченностью области данных и сложностью аппроксимируемых зависимостей. Однако характер поверхности позволяет выдвинуть предположение о возможности применения

алгоритма градиентного спуска для поиска локального или глобального экстремума.

В ряде работ, посвященных данной тематике, предлагается подбирать минимизирующие СКО коэффициенты в результате мутации коэффициентов и применять алгоритм Нелдера-Мида для поиска локального экстремума [5, 6].

С целью минимизации среднеквадратической ошибки отклонения некоторой аналитической функции от дискретного набора данных мы провели сравнительный анализ применимости различных численных методов оптимизации и предоставляем пользователю системы выбор конкретного метода. Наш подход основан на двух положениях:

1. Мутация коэффициентов необходима, для скачкообразного перехода в новую точку поиска экстремума, что позволяет алгоритму запустить несколько стратегий поиска экстремума из различных областей n -мерного пространства поиска, а, следовательно, избежать попадания в локальные экстремумы.
2. Применение наиболее качественных методов поиска экстремума в зависимости от специфики задачи (метод градиентного спуска, симплекс метод, и др.) для идентификации коэффициентов с заданной точностью.

В численных экспериментах наилучшую скорость сходимости показал метод покоординатного спуска, при котором шаг выполняется по координате, дающей максимальный выигрыш. Однако поиск и реализация более эффективных алгоритмов оптимизации остаются предметом дальнейшего исследования.

Важно отметить, что нами предложен новый метод отсева эквивалентных формул. Поскольку одна и та же формула в силу коммутативности операций (+, -, /, *) и дистрибутивности соответствующих

пар операций, а также возможности произвольной расстановки скобок, может быть записана множеством различных вариантов, мы предложили формировать числовые хеши формул, используя специальные коды переменных, коэффициентов и функций и зеркально измененную алгебру (+ вместо *, – вместо /, * вместо +, / вместо –).

Пример работы данной хеш-функции выглядит следующим образом:

Формула	Хеш
$2 * \log(x1) * x2$	16994
$\log(x1) * 2 * x2$	16994
$x2 * 2 * \log(x1)$	16994
$\log(x1) * x2 * 2$	16994
$2 * (\log(x1) + x2)$	2.89382e+007
$2 * \log(x1) + x2 * 2$	2.89722e+007

Пример работы метода покоординатного спуска для семи особей популяции с точностью до 0.001 приведен в таблице 3. Из примера также видно, что нами разработан механизм обработки результатов в виде нечисел и бесконечностей, как больших значений СКО. Такой подход позволяет продолжать поиск без возникновения исключительных ситуаций переполнения и деления на ноль.

Таблица 3 – Пример результата идентификации коэффициентов модели методом покоординатного спуска

Формула	Среднеквадратичное отклонение
До идентификации коэффициентов	
1) $5 * x1 * x1 + 3$	0.00000
2) $2 * x1 + 3 * x2 + 4$	56.92715
3) $2 * \text{pow}(x2, 2) + 2$	154.52993
4) $15 * \text{pow}(x1, 2) + 13$	166.70333
5) $2 * x1 * x1 + 3 * x2 * x2 + 2 * x1 * x2 + 3 * x1 + 3 * x2 + 3$	359.15568
6) $11 * \text{pow}(x1, 3) + 3$	+INF
7) $\text{pow}(x1, x2) * 5$	+INF

После идентификации коэффициентов	
1) $5 * x1 * x1 + 3$	0.00000
2) $27.9921 * x1 + -2.02022 * x2 + 3.12347$	17.37498
3) $1.8 * \text{pow}(x2, 1.5) + 1.579$	57.33454
4) $5.00068 * \text{pow}(x1, 1.99993) + 2.99844$	0.00100
5) $4.99973 * x1 * x1 - 0.48889 * x2 * x2 + 0.33344 * x1 * x2 - 3.33275 * x1 + 4.86652 * x2 + 3.2222$	0.00099
6) $5.00068 * \text{pow}(x1, 1.99993) + 2.99844$	0.00100
7) $\text{pow}(x1, x2) * 3.3e-006$	50.52150

Являясь авторами программного комплекса расчета запаса газа на магистральном газопроводе ОАО «Газпром трансгаз Беларусь», мы попытались устранить влияние таких переходных процессов путем введения инерционности изменения температуры на выходе компрессорной станции в течение времени перехода системы к установившемуся режиму.

Нами был применен ГА для поиска аналитической зависимости между значением времени переходного процесса u и текущего значения расхода газа Q , температуры грунта T_g и перепада температур на выходе компрессорной станции dT . Задачей аппроксиматора был поиск такой аналитической зависимости, которая обладает минимальным СКО для описания $u(Q, T_g, dT)$. Сравнительные результаты работы ГА приведены в таблице 4.

Таблица 4 – Сравнительные результаты работы ГА

Метод аппроксимации	Модель	СКО
ГА	$1.66597 * \text{pow}(1.7573, 1.1225 * \log(3.21177 * \text{pow}(\Delta T * 4.23434, 1.1878))) / \text{pow}(\log(0.0858575 * Q), 0.0742341) - 4.15915 * \Delta T$	1,078879
1) ГА	$52 - 14.48 * \text{pow}(\Delta T, 0.08961) / \text{pow}(Q, -0.1755);$	1,330954
2) ГА	$102.607 * \text{pow}(\Delta T, -0.369501) / \log(0.265488 * Q);$	1,200513
3) НЛА	Модель описана в работе [10]	1,080616

Найденное ГА решение превзошло по точности нечеткий логический аппроксиматор с подобранными оптимальными настройками сглаживания, дав при этом наиболее компактное представление аппроксиматора.

Динамику улучшения качества аппроксимации можно продемонстрировать графиком позитивных мутаций, представленном на рисунке 25.

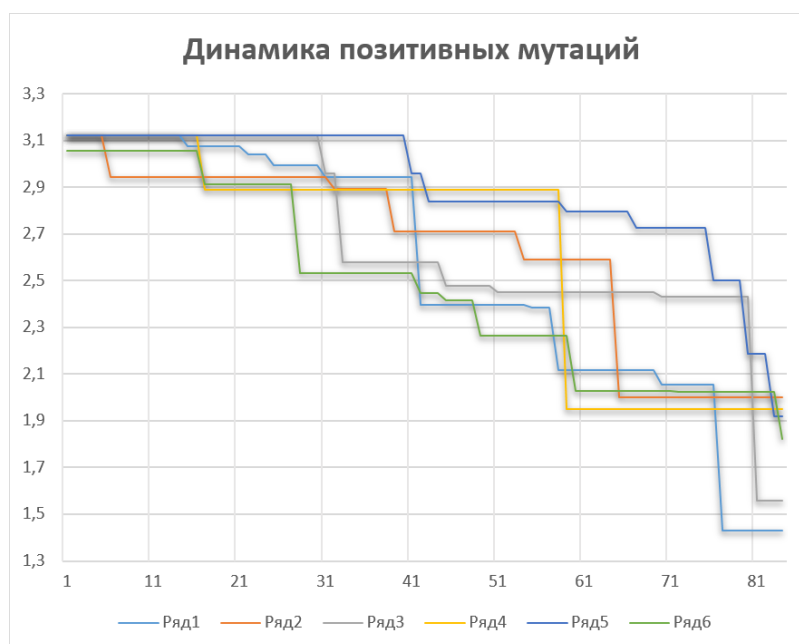


Рисунок 25 – График изменения СКО в результате шагов мутации

Одним из параметров предложенного ГА является предельный размер формулы, что позволяет получать большое разнообразие компактных вариантов аппроксиматоров.

Разработанный ГА имеет возможность гибкой настройки предельных порогов численности популяции, процента и периода гибели неперспективных особей, вероятности периода и вида мутаций структуры и коэффициентов, формулировки инструментов упрощения моделей, алгоритмов поиска экстремума. В частности, для решения задачи учета переходных процессов в стационарных неизотермических моделях

транспортировки газа, нас интересует не только время, но и характер инерции температуры, что требует сбора и обработки дополнительной информации.

3.3 Модель ARIMA

ARIMA – модель авторегрессии скользящего среднего. ARIMA (The Autoregressive Integrated Moving Average Model) – это стандартизированная статистическая модель для прогнозирования и анализа временных рядов. [29]

Уравнение модели, записанное в операторной форме:

$$\Phi_p(L)\Phi_p(L^S)(1-L)^d(1-L^S)^D X_t = \Theta_q(L)\Theta_q(L^S)\varepsilon_t,$$

где операторы сдвига раскрываются аналогично

$$\Phi_p(L) = 1 - \varphi_1 L - \dots - \varphi_p L^p$$

Вместе с разработкой данной модели, авторы Бокс и Дженкинс так же предлагают процесс идентификации, оценки и проверки модели на специфичных наборах данных. Данный процесс получил название метода Бокса-Дженкинса.

Для построения первого приближения модели $ARIMA(d, p, q)(D, P, Q)_L$ мы будем придерживаться следующей последовательности шагов:

1. Обеспечиваем выполнение критерия стационарности временного ряда дифференцированием данных. Выбираем D , и d при которых остатки авторегрессии распределены по закону белого шума (стационарный шум с равномерным спектром), начиная с сезонного дифференцирования.
2. Q – номер последнего сезонного лага ACF, при котором автокорреляционная функция значима;
3. q – номер последнего несезонного лага ACF, при котором автокорреляционная функция значима;
4. P – номер последнего сезонного лага PACF, при котором автокорреляционная функция значима;

5. p – номер последнего несезонного лага PACF, при котором автокорреляционная функция значима;

6. Если есть положительная автокорреляция на лаге 1, мы используем модель AR, если есть отрицательная автокорреляция на лаге 1, мы используем модель MA.

Построим модель ARIMA для двух рассмотренных примеров.

В первом приближении по примеру 1 мы зададимся моделью $ARIMA(1, 0, 2)(1, 0, 1)_{12}$, т.е. $d=1$, $AR(0)$, $MA(2)$, $D=1$, $SAR(0)$, $SMA(1)$.

Для прогноза выберем интервал 12 месяцев (рисунок 26).

Univariate ARIMA Extrapolation Forecast Performance

time	% S.E.	PE	MAPE	sMAPE	Sq.E	MSE	RMSE	ScaledE	MASE
78	0.1099	-0.15	0.15	0.1395	0.1785	0	0	-1.0412	1.0412
79	0.1094	0.0394	0.0947	0.0898	0.0225	0.1005	0.3171	0.3701	0.7056
80	0.1062	0.1077	0.099	0.0978	0.2151	0.1387	0.3724	1.1429	0.8514
81	0.1033	0.1034	0.1001	0.1006	0.2143	0.1576	0.397	1.1408	0.9238
82	0.1056	0.0464	0.0894	0.09	0.0378	0.1336	0.3656	0.479	0.8348
83	0.1104	0.034	0.0802	0.0808	0.0187	0.1145	0.3384	0.3368	0.7518
84	0.1303	-0.023	0.072	0.0725	0.0056	0.0989	0.3145	-0.1848	0.6708
85	0.1688	-0.0218	0.0657	0.0661	0.0031	0.087	0.2949	-0.1376	0.6041
86	0.1791	-0.1505	0.0751	0.0743	0.1066	0.0891	0.2986	-0.8048	0.6264
87	0.1843	0.0419	0.0718	0.0712	0.0115	0.0814	0.2853	0.2647	0.5903
88	0.2035	0.1043	0.0748	0.0747	0.069	0.0803	0.2833	0.6475	0.5955
89	0.1817	0.0053	0.069	0.0689	2e-04	0.0736	0.2713	0.0334	0.5486

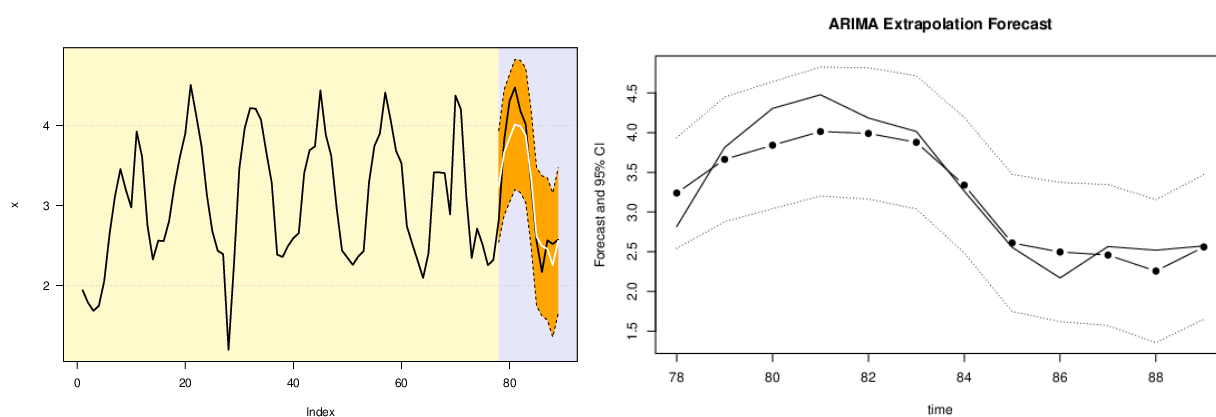


Рисунок 26 – Результат прогноза для примера 1 по модели $ARIMA(1, 0, 2)(1, 0, 1)_{12}$ $AIC = 70.66$

Оценка качества построенной модели может осуществляться по целому ряду критериев (AIC, AICс, BIC, RMSE, MAPE). [28]

В частности, по информационному критерию Акаике AIC [27], вычисляемому в общем случае по формуле $AIC = 2k - 2 \ln(L)$, где k – число параметров статистической модели, L – максимальное значение функции правдоподобия.

Попробовав улучшить модель прогноза меняя параметры модели, мы получили ухудшение:

$$ARIMA(1, 0, 2)(1, 0, 2)_{12} AIC = 72.48$$

$$ARIMA(1, 0, 3)(1, 0, 2)_{12} AIC = 74.86$$

Поэтому будем рассматривать модель $ARIMA(1, 0, 2)(1, 0, 1)_{12}$ как оптимальную и проанализируем распределение остатков по методу обратного выбора Ромена Франсуа [31, 32] (рисунок 27).

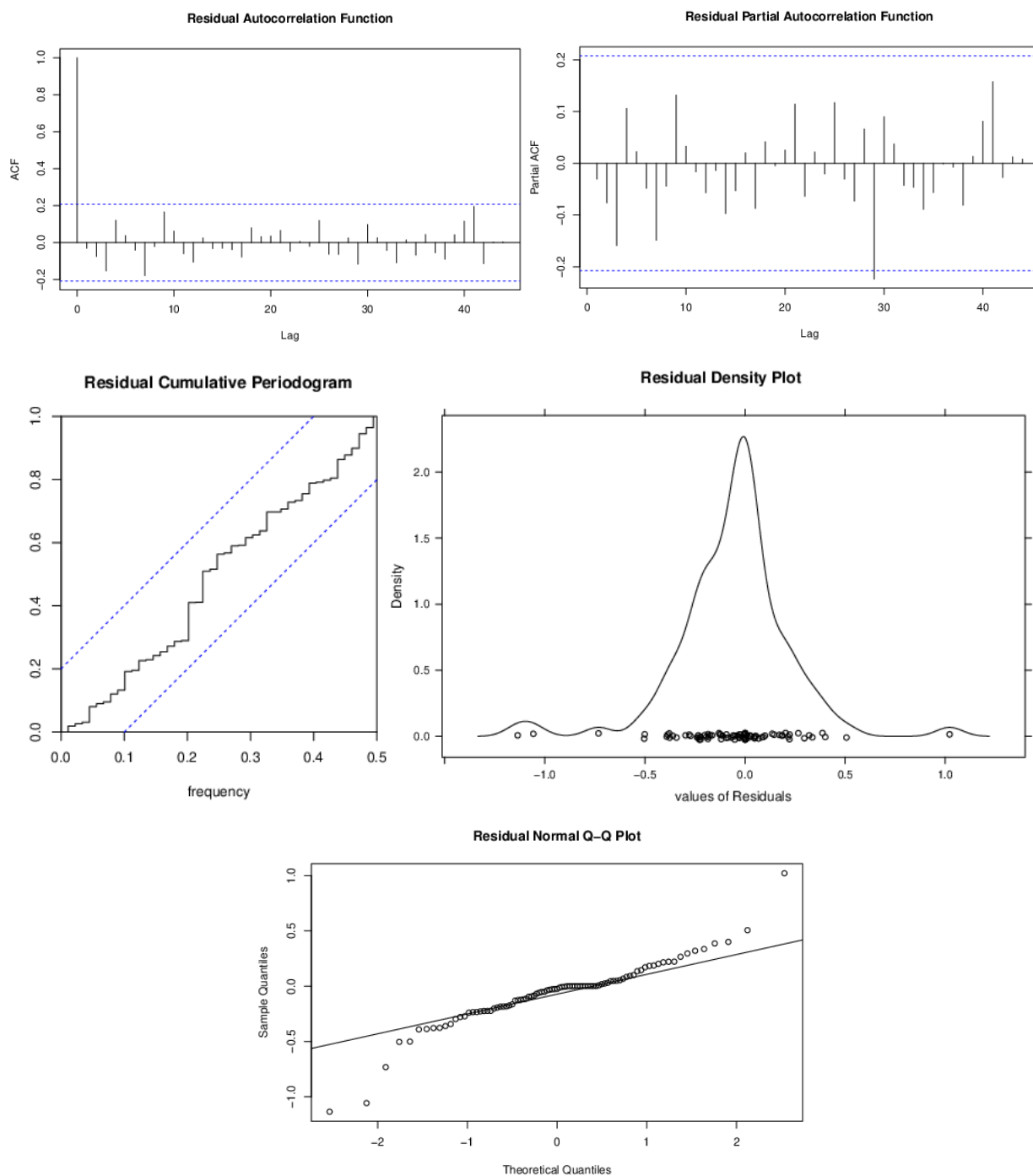


Рисунок 27 – Анализ остатков модели. Распределение остатков лишено цикличности и существенной значимости всего диапазона лагов

По примеру 2 выполним аналогичные операции. В качестве первого приближения будем рассматривать модель $ARIMA(1, 0, 1)(1, 0, 1)_{12}$.

Построим прогноз по данной модели (рисунок 28).

Univariate ARIMA Extrapolation Forecast Performance

time	% S.E.	PE	MAPE	sMAPE	Sq.E	MSE	RMSE	ScaledE	MASE
78	0.1311	0.0369	0.0369	0.0376	0.0035	0	0	0.3648	0.3648
79	0.137	0.1146	0.0757	0.0796	0.0451	0.0243	0.1559	1.3156	0.8402
80	0.1344	0.0683	0.0732	0.0766	0.018	0.0222	0.149	0.8304	0.8369
81	0.1276	0.0171	0.0592	0.0618	0.0013	0.017	0.1303	0.2241	0.6837
82	0.1423	-0.0608	0.0595	0.0612	0.013	0.0162	0.1272	-0.707	0.6884
83	0.157	-0.0049	0.0504	0.0518	1e-04	0.0135	0.1162	-0.0583	0.5834
84	0.1971	-0.0343	0.0481	0.0492	0.0028	0.012	0.1095	-0.3303	0.5472
85	0.2361	-0.028	0.0456	0.0465	0.0015	0.0107	0.1033	-0.2375	0.5085
86	0.2756	-0.0342	0.0443	0.0451	0.0017	0.0097	0.0984	-0.2575	0.4806
87	0.274	-0.0393	0.0438	0.0445	0.0025	0.009	0.0946	-0.3074	0.4633
88	0.297	0.033	0.0429	0.0435	0.0018	0.0083	0.0911	0.2653	0.4453
89	0.323	-0.1015	0.0477	0.0479	0.0121	0.0086	0.0929	-0.6819	0.465

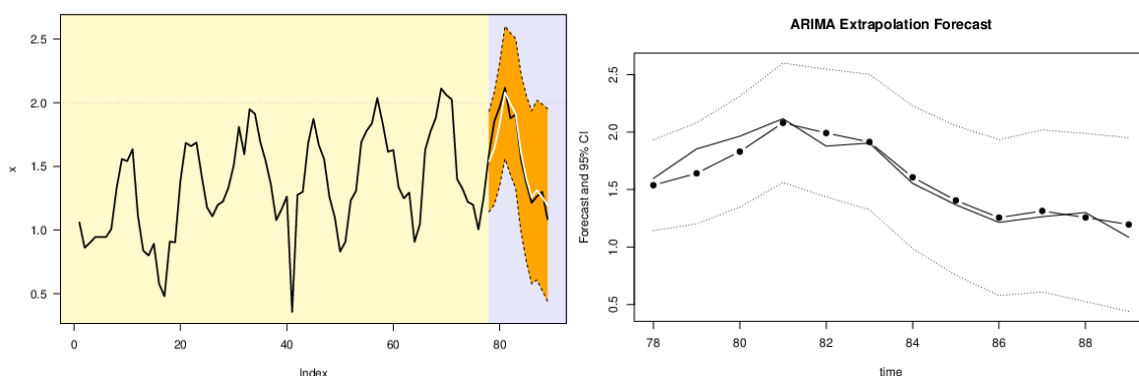


Рисунок 28 – Результат прогноза для примера 2 по модели $ARIMA(1, 0, 1)(1, 0, 1)_{12}$ AIC = -4.8

Попытаемся улучшить модель.

$ARIMA(1, 0, 2)(1, 0, 1)_{12}$ дает AIC = -3.25, что больше предыдущего, однако RMSE незначительно, но уменьшилось, что говорит о том, что в качественном отношении модель дает лучше результат, но возрастает количество параметров модели.

Рассмотрим распределение остатков модели (рисунок 29).

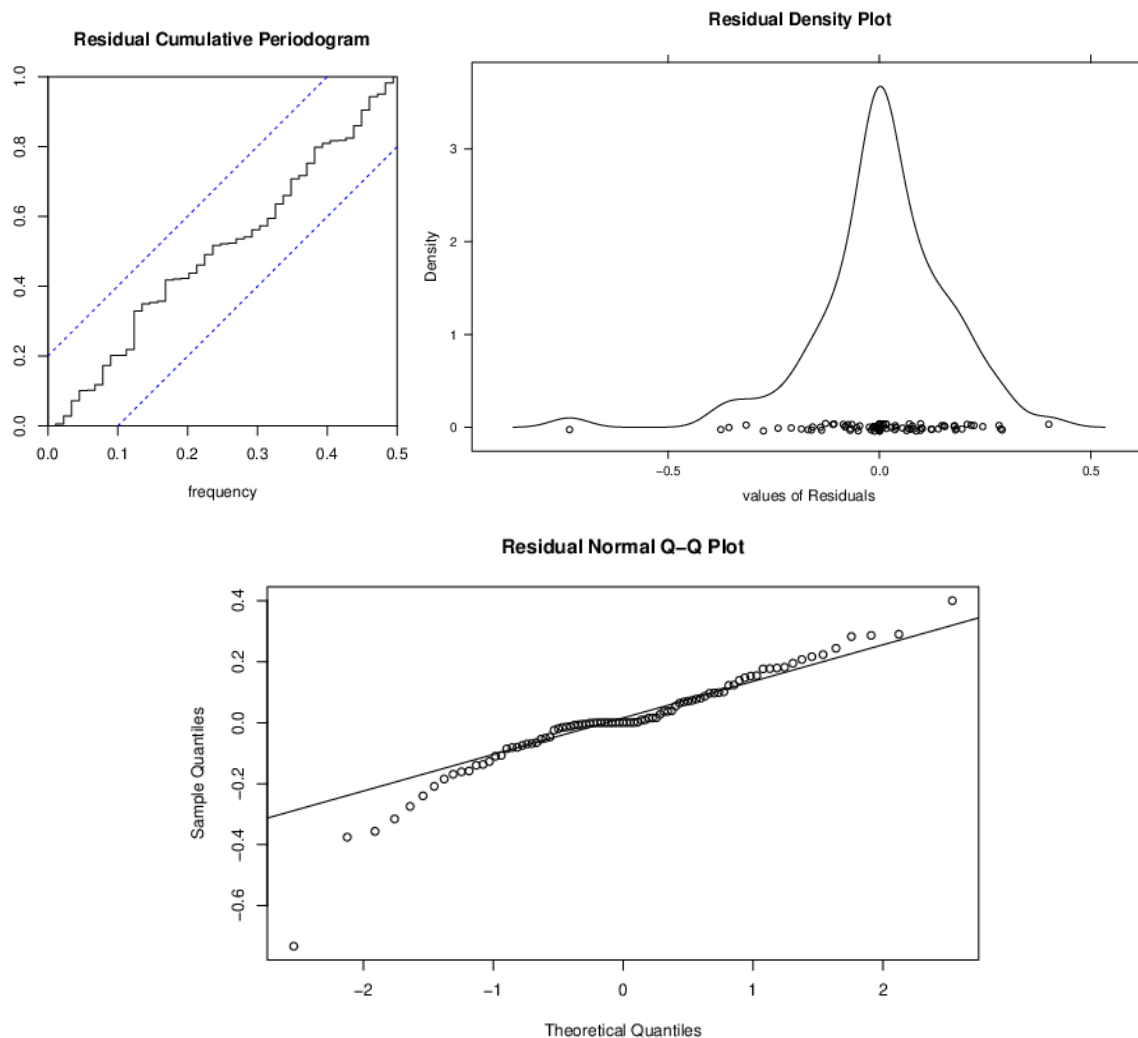


Рисунок 29 – Анализ остатков модели. Распределение остатков лишено цикличности и близко к характеристикам белого шума

Прогнозирование потребления природного газа на месячный период недостаточно, поскольку контрактная основа взаимоотношений поставщика и потребителя оперирует недельными контрактами.

Временной ряд усредненных недельных расходов газа за 8 лет содержит 334 недели с лагом сезонности 51 неделя.

Нестационарность ряда устранима с параметрами дифференцирования (1, 1) как показано на рисунке 30.

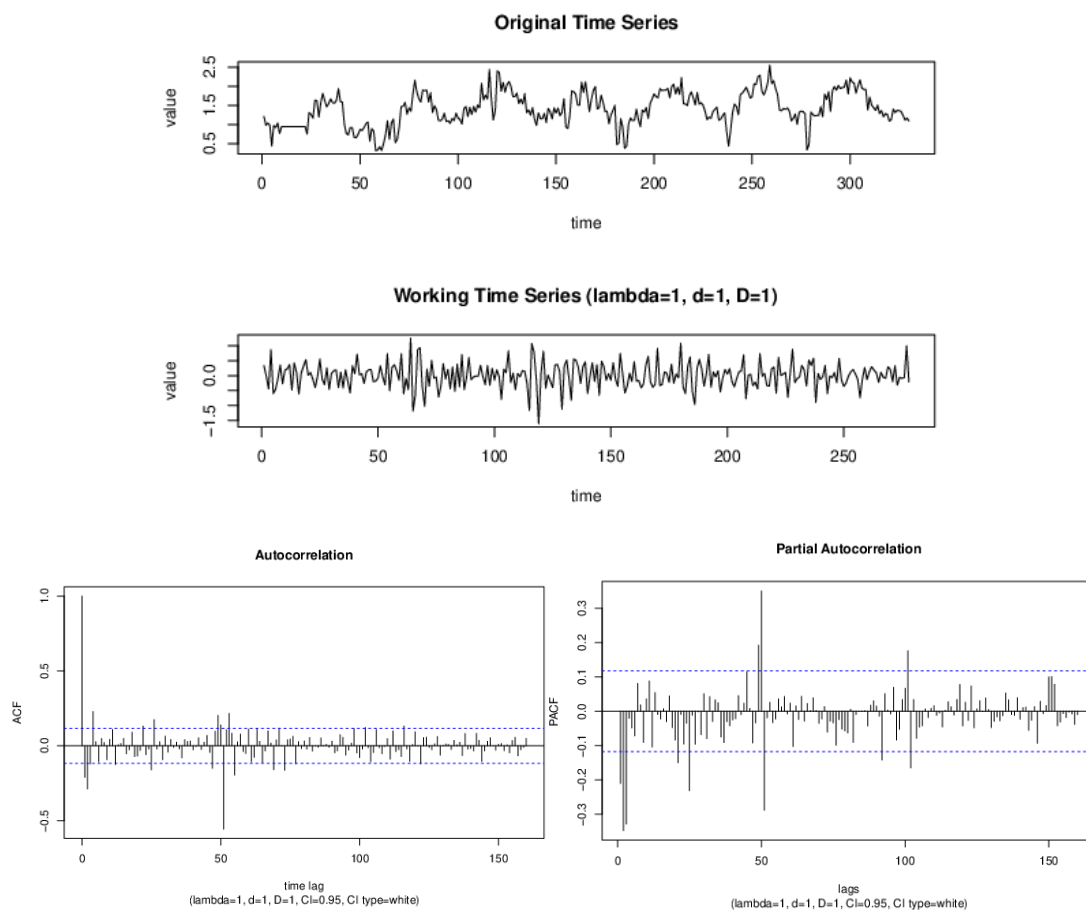


Рисунок 30 – Устранение нестационарности недельного временного ряда с параметрами дифференцирования $d=1, D=1$

Поскольку на лаге 1 наблюдается отрицательная автокорреляция, то используем модель MA. По ACF номер последнего значимого несезонного лага 2, следовательно выбираем модель MA(2), номер последнего значимого сезонного лага 1, следовательно выбираем модель SMA(1).

В первом приближении строим модель ARIMA(2, 0, 1)(1, 0, 1)₅₁ (рисунок 31).

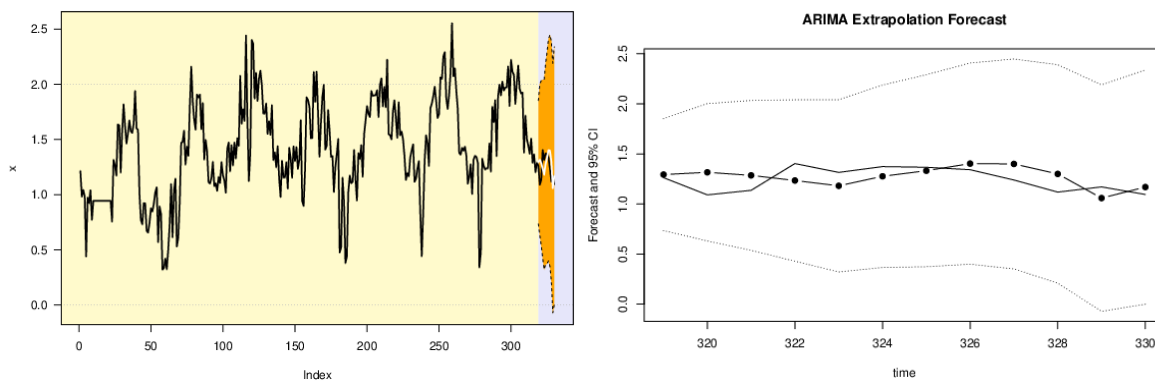


Рисунок 31 – Прогноз на 12 недель по модели $ARIMA(2, 0, 1)(1, 0, 1)_{51}$
 $AIC = 189$, $sMAPE = 0.0964$, $RMSE = 0.1385$

Анализ распределения остатков показывает на серьезную периодичность на малых частотах (больших периодах, см. на рисунке 32).

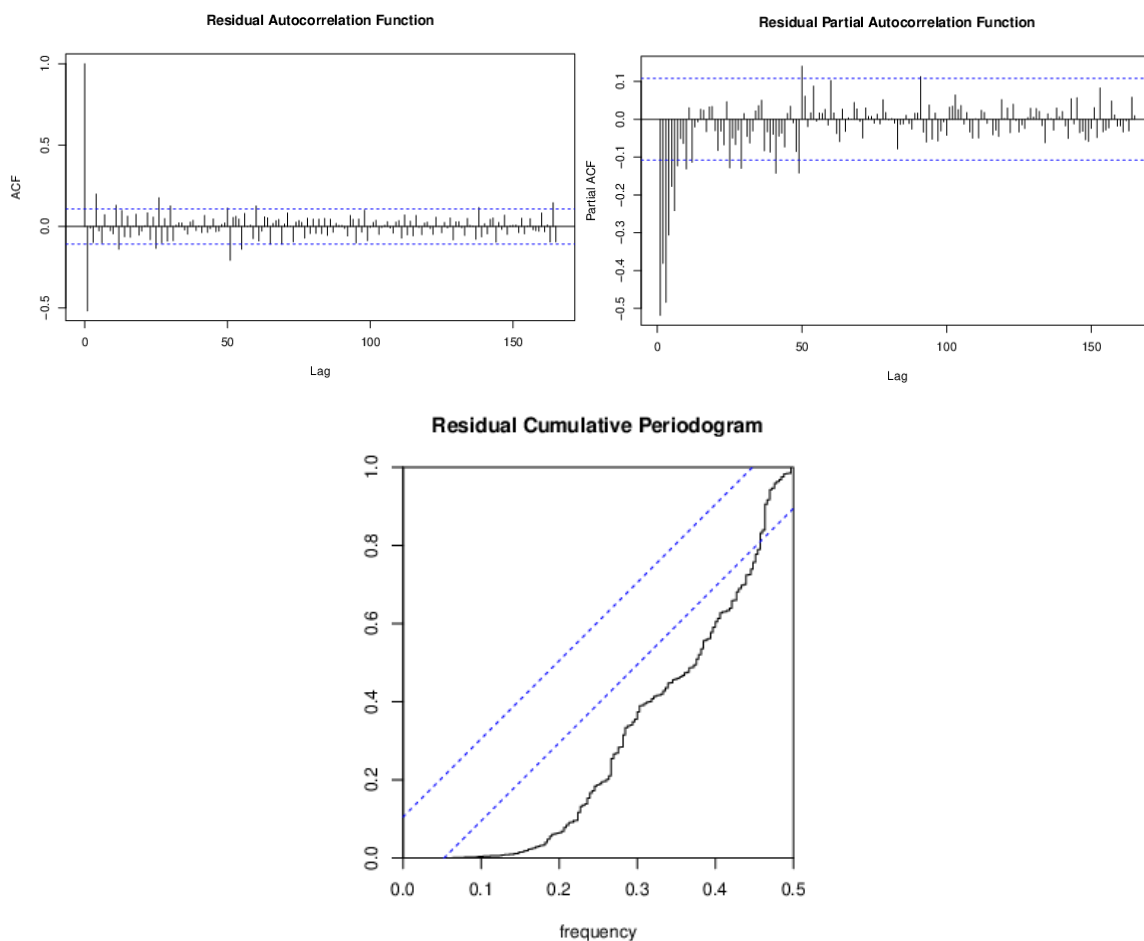


Рисунок 32 – Распределение остатков модели $ARIMA(2, 0, 1)(1, 0, 1)_{51}$

Мы попробовали перебрать параметры и получили удовлетворительный результат снизив дифференцирование по несезонному тренду до 1 и увеличив параметр MA до 2 (рисунок 33).

Univariate ARIMA Extrapolation Forecast Performance

time	% S.E.	PE	MAPE	sMAPE	Sq.E	MSE	RMSE	ScaledE	MASE
319	0.2265	-0.017	0.017	0.0169	5e-04	0	0	-0.2332	0.2332
320	0.2614	-0.1875	0.1023	0.0941	0.0418	0.0211	0.1454	-2.2204	1.2268
321	0.3009	-0.113	0.1059	0.0984	0.0165	0.0196	0.14	-1.3943	1.2826
322	0.3447	0.1355	0.1133	0.1101	0.0362	0.0237	0.1541	2.064	1.4779
323	0.3902	0.1177	0.1141	0.1131	0.024	0.0238	0.1542	1.6802	1.5184
324	0.3862	0.0857	0.1094	0.1092	0.0139	0.0221	0.1488	1.2779	1.4783
325	0.3928	0.0405	0.0996	0.0995	0.0031	0.0194	0.1393	0.6013	1.353
326	0.3932	-0.0298	0.0908	0.0907	0.0016	0.0172	0.1311	-0.4348	1.2382
327	0.4138	-0.1127	0.0933	0.0925	0.0195	0.0174	0.1321	-1.5159	1.2691
328	0.4662	-0.1436	0.0983	0.0967	0.0258	0.0183	0.1352	-1.7442	1.3166
329	0.5985	0.1133	0.0997	0.0988	0.0176	0.0182	0.135	1.4394	1.3278
330	0.5618	-0.0482	0.0954	0.0945	0.0028	0.0169	0.1301	-0.5734	1.2649

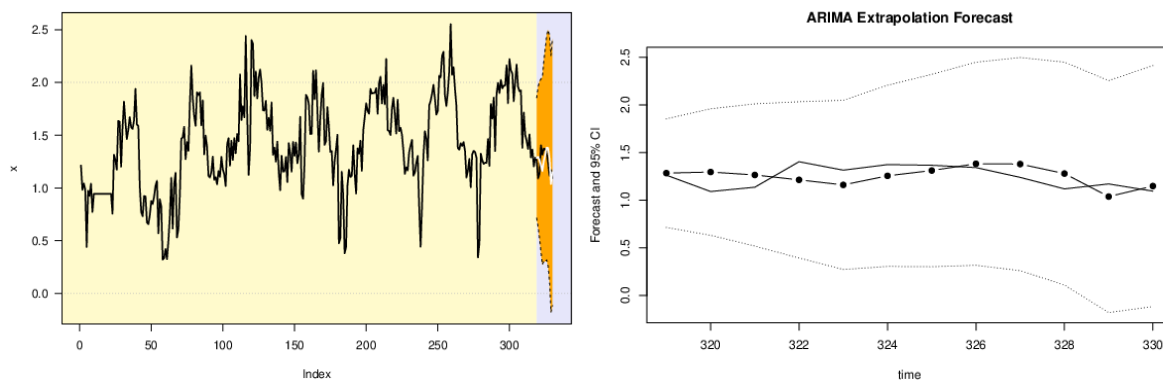


Рисунок 33 – Прогноз на 12 недель по модели $ARIMA(1, 0, 2)(1, 0, 1)_{51}$

$$AIC = 168, sMAPE = 0.0955, RMSE = 0.1364$$

При этом распределение остатков приняло удовлетворительный вид (рисунок 34).

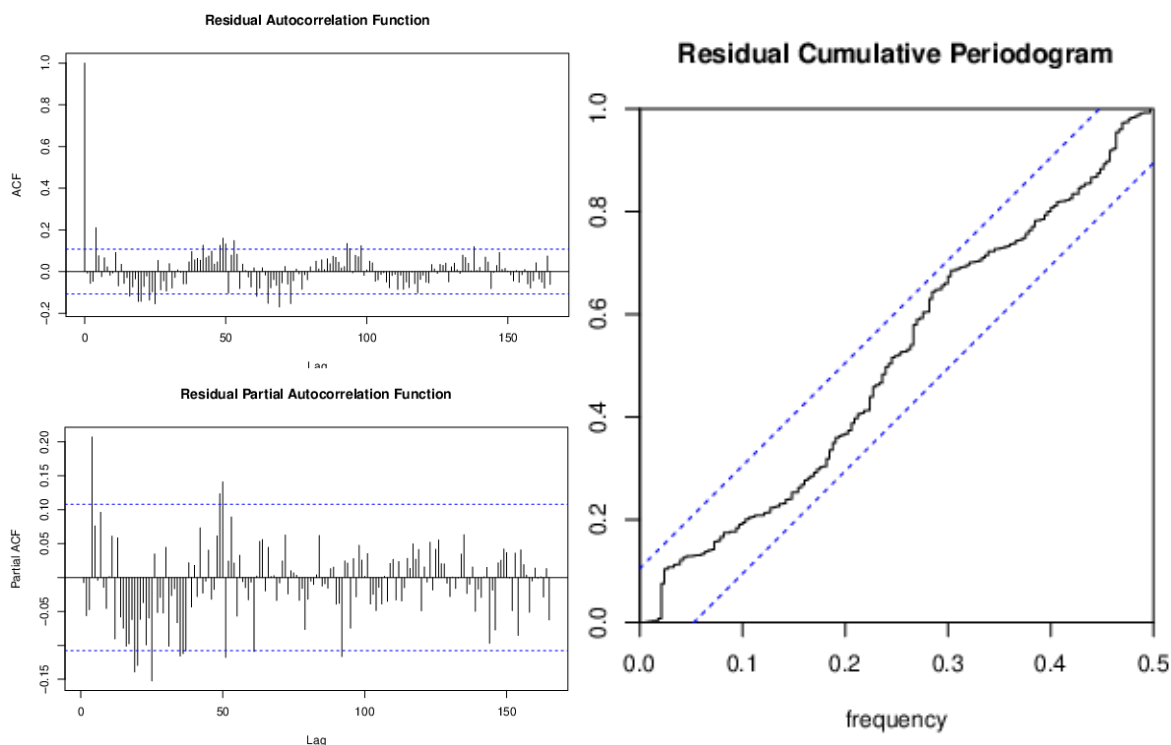


Рисунок 34 – Распределение остатков модели $ARIMA(1, 0, 2)(1, 0, 1)_{51}$

Учитывая, что исходные данные имеют пропуски и грубые промахи в силу ошибок телеметрии, периодов планового и внепланового ремонта газопровода, то помимо фильтрации входных данных временного ряда интересной является идея восстановления пропусков обратным прогнозированием по построенной модели. [33]

В рамках данного исследования такие эксперименты проведены не были, и они могут составлять предмет дальнейшего исследования.

3.4 Модель экспоненциального сглаживания

Методы экспоненциального сглаживания – это семейство моделей, использующих экспоненциально уменьшающиеся веса для предыдущих значений при прогнозировании нового значения временного ряда.

Мы воспользуемся моделью экспоненциального сглаживания с трендом Хольта и сезонностью Винтерса (метод Хольта-Винтерса, рисунок 35, 36).

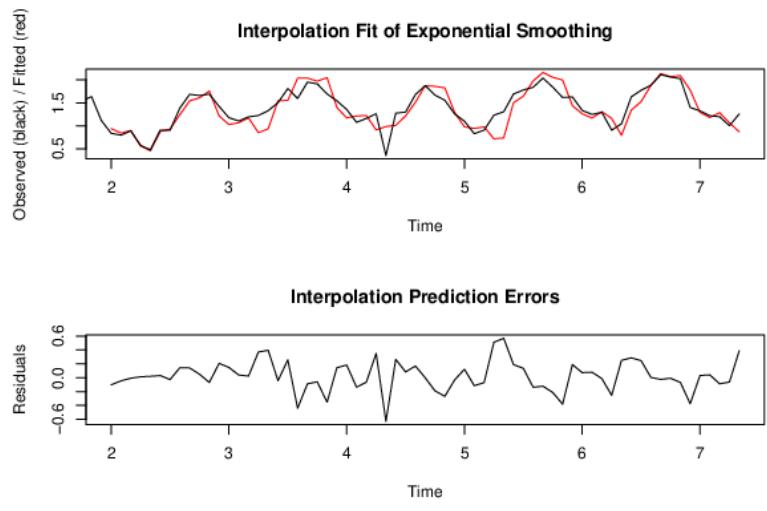


Рисунок 35 – Экспоненциально сглаженный ряд и остатки

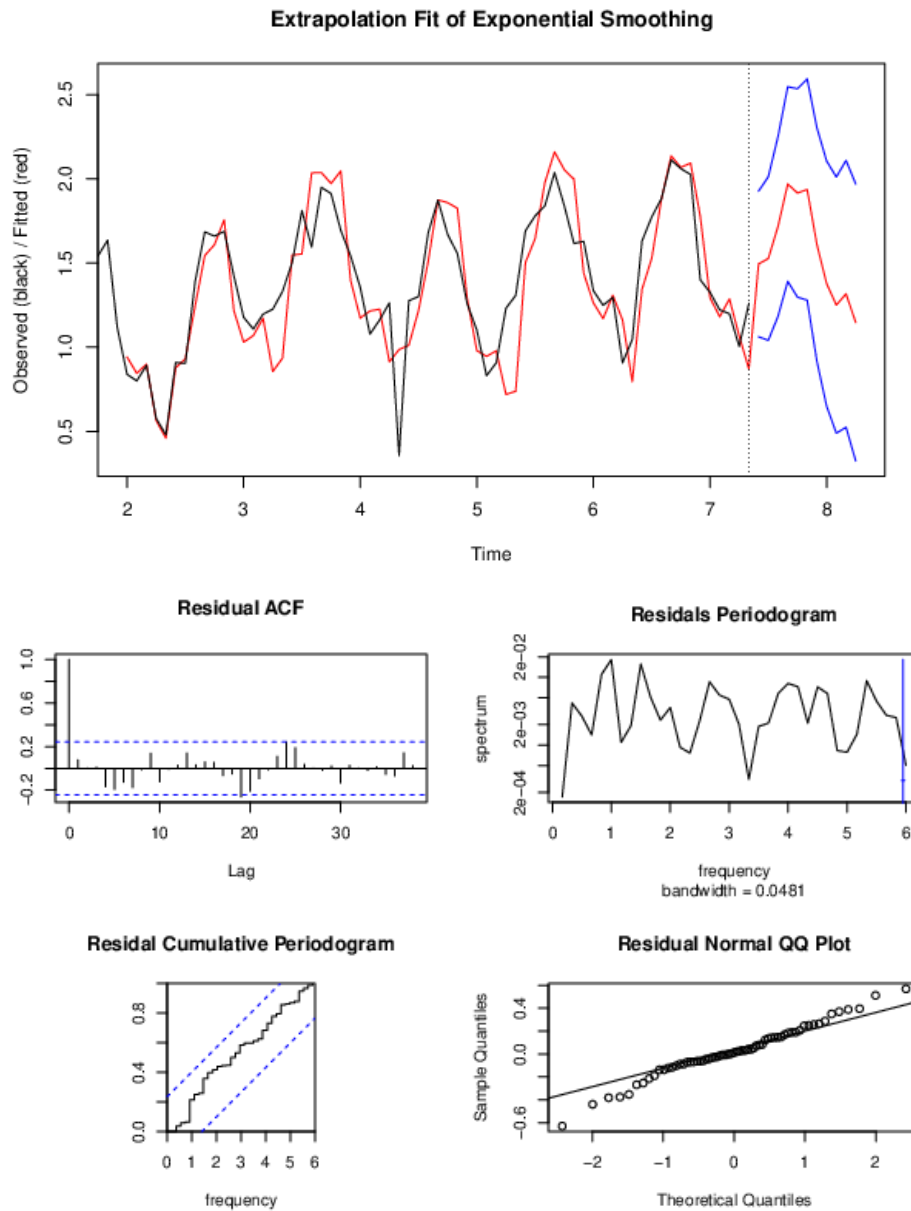


Рисунок 36 – Результат прогнозирования по модели экспоненциального сглаживания с сезонностью Хольта-Винтерса $RMSE=0.256$

В нашем случае модель экспоненциального сглаживания (ETS) показала худшую точность.

3.5 Прогнозирование с использованием метода группового учета аргументов

Метод группового учета аргументов (GMDH) является методом структурно-параметрической идентификации сложных объектов, процессов и систем по данным наблюдений в условиях неполноты информации [19].

А.Г. Ивахненко разработал метод группового учета аргументов (group method of data handling) в 1970 году [20, 21], позволяющий не только вычислять веса связей между нейронами, но и определять количество слоев в сети и нейронов в них в зависимости от потребностей прикладной задачи. Используя подход обучения с учителем, уровни сети строятся и обучаются на основе обучающего множества с использованием регрессионного анализа. Затем происходит этап упрощения сети с применением отдельного множества объектов с известными правильными ответами, которое не использовалось при обучении (проверочное множество, validation set). Для исключения ненужных нейронов из сети используется регуляризация. В работе [22] описано применение метода группового учета аргументов для обучения глубокой нейронной сети, состоящей из восьми слоев. Метод широко использовался на практике [22–24].

На раннем этапе развития теории GMDH утверждалось сходство с нейронными сетями и были выделены многослойные алгоритмы GMDH. А.Г. Ивахненко в одной из вступительных статей утверждает, что, поскольку различия между персептроном и GMDH несущественны, то системы GMDH уместно называть «системами персептронного типа».

Отличие состоит в том, что после процедуры упрощения модели, в модели остаются только значимые узлы (рисунок 37).

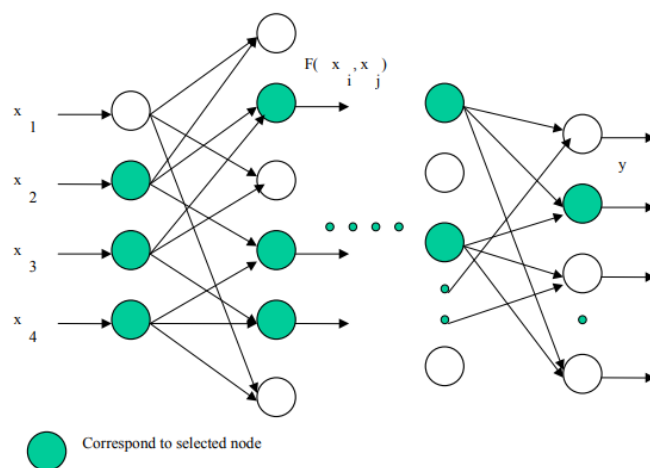


Рисунок 37 – Многоуровневый алгоритм GMDH [1]

Мы построили модель GMDH для примера 2. Она в результате обучения представляла собой 6 слоев по 13 нейронов (рисунок 38).

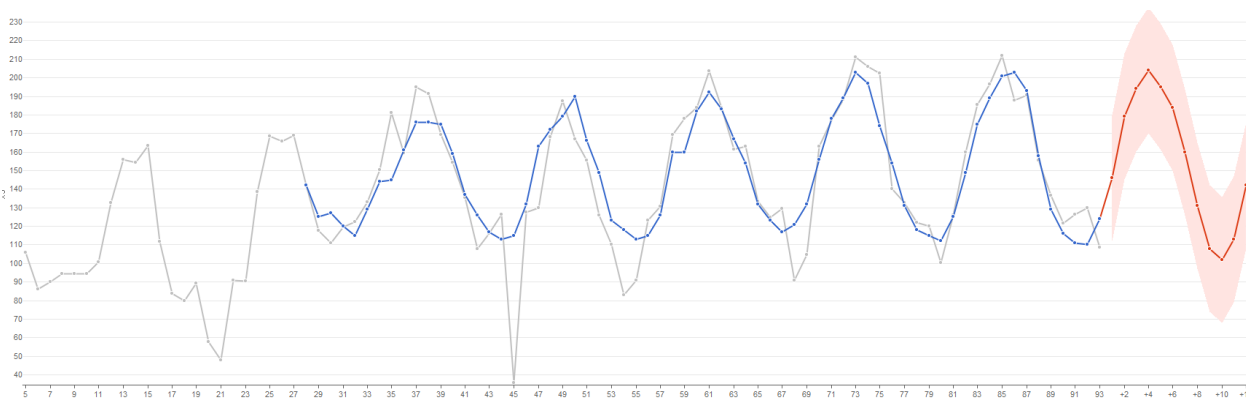


Рисунок 38 – Результат прогноза с использованием GMDH модели

Анализ результатов прогнозирования показал, что точность прогнозирования по построенной модели хуже, чем для моделей класса ARIMA, но лучше, чем модели экспоненциального сглаживания $RMSE=0.169$.

Анализ распределения остатков также показал удовлетворительное распределение (см. рисунок 39)

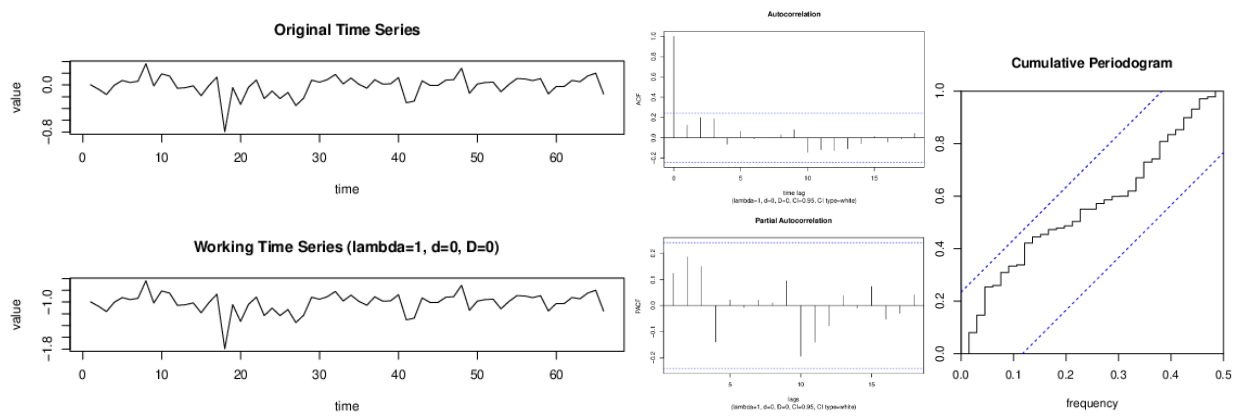


Рисунок 39 – Распределение остатков для модели GMDH

Для прогноза по усредненным недельным данным была сформирована модель из 8 слоев с 13 нейронами в каждом. Результат данной модели представлен на рисунке 40. Результат прогнозирования значительно хуже $RMSE=0.392$.

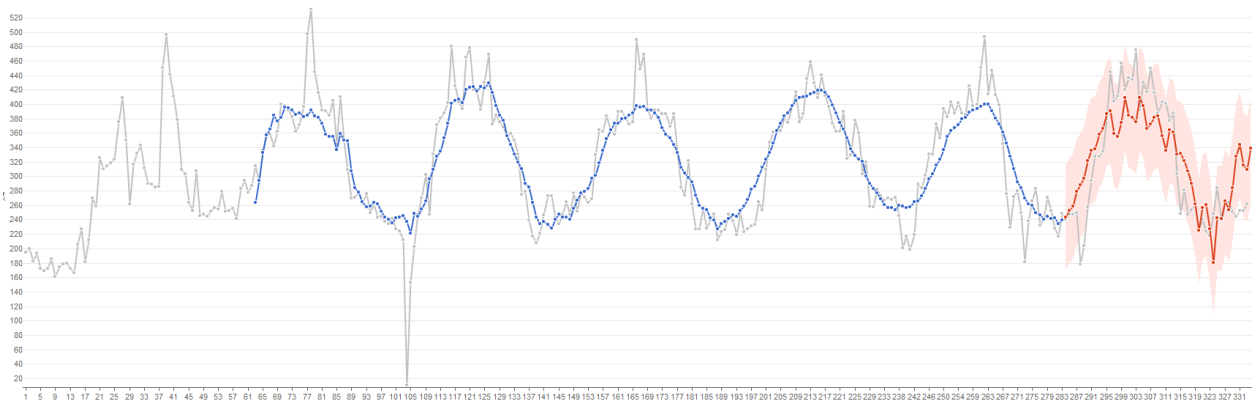


Рисунок 40 – Результат прогноза по усредненным недельным данным с использованием GMDH модели

И распределение остатков имеет существенное отклонение от линейности на периодограмме в области средних частот (рисунок 41).

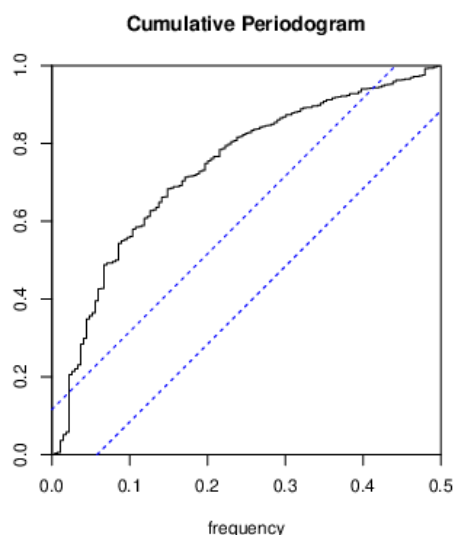


Рисунок 41 – Распределение остатков для модели по усредненным недельным данным GMDH

Таблица 5 – Точность прогнозирования различных моделей по RMSE

Тип данных	ARIMA	ETS	GMDH
Месяц	0.138	0.256	0.169
неделя	0.136	1.270	0.392

3.6 Разработка подсистемы визуализации данных прогноза

В данной работе предлагается новый метод построения аппроксимации географических данных на основе оригинальной реализации нечеткой логики. Применение математического аппарата нечеткой логики обусловлено возможностью гибкой настройки аппроксиматора под условия решаемой задачи при работе с данными большой размерности.

Географические данные обеспечивают привязку информации (Lat, Lon) координатам или преобразованным (x, y) сферической проекции Меркатора (система координат WGS-84/UTM).

Для построения топографической карты поверхности по дискретному множеству точек традиционно применяются различные методы пространственной интерполяции. Методы геостатистического оценивания, такие как крикинг, требуют большого объема вычислений, но позволяют получать оптимальные в определенном смысле интерполяции. [36]

Идея применения нечеткого логического аппроксиматора для построения топографической карты следует из аналогии. Совокупность точечных данных можно рассматривать как систему знаний, в которой каждая точка – это факт, описываемый с точки зрения формальной логики, как правило:

ЕСЛИ координаты X , Y И время t ТО расход Q , температура T , и т.д.

В этом случае мы имеем систему продукций, в рамках которой строится логический вывод о значении характеристик неизвестной точки. Нечеткая логика, предложенная Лофти Заде [37], естественным образом представляет работу нашего мозга при решении сходных задач. Наше сознание оперирует нечеткими понятиями, которые мы привыкли на вербальном уровне выражать словами. Мы передаем информацию и выполняем рассуждения, отталкиваясь от нечеткой модели вербальных категорий на соответствующей шкале измерения. Такое утверждение верифицировано рядом исследований, в частности, исследованием личных психологических конструкторов основоположника клинической психологии Джорджа Келли. [38]

В этом смысле применение нечеткой логической аппроксимации подобно реализации искусственного интеллекта, ориентированного на достраивание рельефа и иных характеристик неизвестных областей по данным окружающих замеров.

В данной работе в качестве универсального адаптивного аппроксиматора предлагается использовать аппарат специфической

реализации нечеткой логики, разработанной в 1998г. в рамках диссертационного исследования Глухова Д.О. и получившей дальнейшее развитие в работах авторов данной публикации [39, 40, 41, 42]. Адаптивность предложенной реализации заключается в возможности настройки параметров функций принадлежности под условия конкретного объекта аппроксимации. Учитывая присутствие существенной неопределенности, мы специально не формулируем задачу, как задачу получения оптимальной в том или ином смысле аппроксимации, а оставляем право на выбор параметров и оценку качества результата стороннему наблюдателю, имеющему возможность верификации результата по данным наблюдений.

Рассмотрим формализацию задачи построения карты. Для представления информации о геометрии будем придерживаться объектно-ориентированной парадигмы в постановке Гради Буча [44].

Обращение к внутренней структуре вводимых сущностей будем изображать, используя оператор точка «.». Тогда обозначение вида $a.b$ будет обозначать обращение к свойству b элемента a .

Для хранения информации о точке мы будем оперировать конструкцией, включающей помимо информации о глубине, информацию о всех вычисляемых характеристиках: $p = \langle x, y, z, r_1, r_2, \dots \rangle$

Пусть $P = \{p_i\}$ – множество узлов аппроксимации, представляющее собой точки трека эхолота и контура водного объекта (внешние границы и границы островов).

Введем оператор определения расстояния между точками как $L(p, p_i)$.

Продолжая аналогию с продукционной системой, знание о фрагменте водоема распространяется на окружающие фрагменты в соответствии с унимодальной функцией принадлежности, имеющей максимум в заданной точке. Нами предлагается функция принадлежности в пространстве расстояний от точки узла аппроксимации p_i до точки p вида:

$$\varphi(p, p_i) = \frac{1}{L(p, p_i)^n} | L(p, p_i) < L_{limit} \quad (7)$$

Значение искомого параметра, например, глубины z , в неизвестной точке p будет определяться дефазификацией методом центра тяжести при рассмотрении полученных единиц влияния (валидных правил) как синглетонов, а именно:

$$p.z = \frac{\sum_i \varphi(p, p_i) p_i \cdot z}{\sum_i \varphi_i} \quad (8)$$

Аппарат нечеткой логики дает нам пространство выбора различных вариантов формы и параметров функции принадлежности, в том числе, форма функции принадлежности может отражать вероятностный характер изменчивости рельефа.

В нечеткой логике Бозе логические операции И и ИЛИ реализуются в виде операций взятия минимума и максимума соответственно. Поэтому вывод о значении неизвестного параметра строится, как говорят, по минимаксному принципу. Если предикат представляет собой условие в пространстве расстояний между неизвестной точкой p и узлом аппроксимации p_i , а пространство неизвестного параметра (например, глубины z), то вывод методом центра тяжести будет выглядеть следующим образом:

$$p.z = \frac{\int \max(\varphi(p, p_i)) z dz}{\int \max(\varphi(p, p_i)) dz} \quad (9)$$

Нами также предложены модификации метода дефазификации, устраняющие диспропорцию влияния множества близкорасположенных точек, путем дискретизации пространства и замены влияния группы правил, попавших в один интервал дискретизации, на влияние одного правила с максимумом функции принадлежности в точке p .

Интересным вариантом устранения влияния неравномерности расположения узлов аппроксимации является дискретизация угла между неизвестной точкой и узлом аппроксимации. Мы рассмотрели 8 и 16 секторное

разбиение круга. В каждом секторе влияние на результат аппроксимации оказывает только самая ближняя точка (точка с максимальным значением функции принадлежности). Остальные узлы отбрасываются.

Введем оператор $angle(p, p_i)$, возвращающий номер сектора круга в который попадает угол между неизвестной точкой p и узлом аппроксимации – точкой p_i (1..8 или 1..16 в зависимости от варианта разбиения).

Определим ближайшую точку для каждого сектора следующим образом:

$$PA_{angle(p, p_i)} = p_i \mid \forall p_j \in P \ p_j \neq p_i, \ angle(p, p_j) = angle(p, p_i), \ \varphi(p, p_i) > \varphi(p, p_j), \ L(p, p_i) < L_{limit}$$

Значение неизвестного параметра, на примере глубины z , для 8-секторного разбиения:

$$p.z = \frac{\sum_{k=1}^8 \varphi(p, PA_k) PA_k.z}{\sum_{k=1}^8 \varphi(p, PA_k)} \quad (10)$$

Если интервал дискретизации по углу стремится к 0, то выражение 2 примет вид:

$$p.z = \frac{\int_{-\pi}^{\pi} \varphi(p, \alpha) z(\alpha) d\alpha}{\int_{-\pi}^{\pi} \varphi(p, \alpha) d\alpha}, \quad (11)$$

где $\varphi(p, \alpha)$ - это кусочно-линейная интерполяция в полярных координатах значений функций принадлежности ближайших в направлении α точек, а $z(\alpha)$ – кусочно-линейная интерполяция в полярных координатах значение глубин.

Сравнительный результат применения дискретизации, как метода устранения влияния неоднородности расположения узлов аппроксимации, приведен на рисунке 42. На карте показан тренд эхолота. На рисунке 42а пространство между трендами эхолота получилось приподнятым в результате влияния узлов контура озера. На рисунке 42б данный эффект существенно

снижен и 42в пространство между трендами корректно аппроксимируется как переход глубин между глубинами трендов.

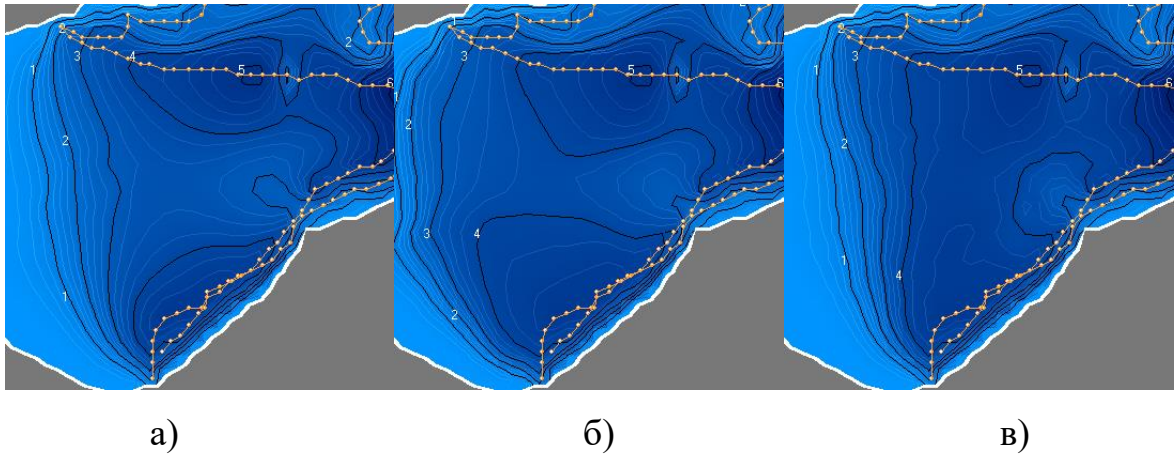


Рисунок 42 – Результат работы алгоритмов аппроксимации а) – согласно формуле 1, б) – согласно формуле 2 и в) – с применением дискретизации угла (формула 4)

Ключевое отличие нечеткой логической аппроксимации от традиционных способов аппроксимации проявляется в возможности учета нескольких предикатов.

Например, мы формулируем условие аппроксимации, которое должно учитывать, помимо информации о глубине, информацию о структуре дна, для того, чтобы устранить влияние скачков глубины, возникающих от находящихся на дне артефактов.

ЕСЛИ координаты X , Y **И** структура дна без артефактов
ТО глубина Z , температура воды T , и т.д.

В этом случае значение предикатов должны быть нормализованы и нами предлагается следующая модификация функции принадлежности:

$$\varphi(p, p_i) = \left(1 - \frac{L(p, p_i)}{L_{limit}}\right)^n \mid L(p, p_i) < L_{limit}$$

Пусть значение неровности дна определено для точки p в рамках одно-предикатной аппроксимации и равно $R(p)$. Нормализуем значение данного параметра и построим оценку вербального термина «дно без артефактов» следующим образом:

$$\varphi_R(p) = \left(1 - \frac{R(p) - R_{min}}{R_{max} - R_{min}}\right)^m$$

Тогда, вычисление значения неизвестного параметра примет вид:

$$p.z = \frac{\int \max(\min(\varphi(p,p_i), \varphi_R(p)))z dz}{\int \max(\min(\varphi(p,p_i), \varphi_R(p)))dz} \quad (12)$$

В итоге мы получаем классическое минимаксное представление логического вывода о значении неизвестного параметра и, соответственно, аппроксимацию поверхности дна со снижением влияния артефактов. Степень снижения данного влияния управляется параметром m . Результат работы такой логики приведен на рисунке 43.



без устранения артефактов структура дна влияние артефакта устранено

Рисунок 43 – Пример аппроксимации поверхности дна двух-предикатным аппроксиматором

Дополним описание нашего аппроксиматора параметром интервала дискретизации, или шага изолинии (Counter line step) CLS.

Учитывая особенность исходных данных, а именно скачкообразное изменение GPS координат для множества точек, большую размерность задачи, введем для нашего аппроксиматора дополнительные параметры, такие как шаг накопления данных (Filter interval) FI.

Исходя из принятых допущений и ограничений предложенный аппроксиматор можно описать четверкой параметров: $App = \langle p, L_{lim}, CLS, FI \rangle$

Введем понятие отрезка прямой линии $L = \langle p_1, p_2 \rangle$, которое нам понадобится для дальнейших рассуждений.

Введем оператор аппроксимации для произвольной точки p_x : $p_x = app(P, p_{x.x}, p_{x.y})$, где P множество узлов аппроксимации.

Для построения карты изолиний по уровням дискретизации введем понятие высот дискретизации, которое представляет собой множество высот $D = \{z_i\}$. Если рассматривать решение задачи построения изолиний на регулярном разбиении пространства на элементарные участки прямоугольной формы, то достаточно решить задачу построения изолиний в пределах прямоугольного фрагмента пространства.

Такой фрагмент (далее прямоугольник разбиения) будет описываться четверкой угловых точек $frag = \langle p_1, p_2, p_3, p_4 \rangle$.

Каждая угловая точка прямоугольника разбиения определяется аппроксимацией

$$p_{i.z} = app(P, p_{i.x}, p_{i.y})$$

Введем оператор, который для каждой грани прямоугольника разбиения определит подмножества уровней дискретизации, попадающих в интервал глубин в пределах грани, и соответствующих точек на гранях:

$$D(p_1, p_2) = \{z_i \in D \mid z_i \leq p_{1.z} \wedge z_i \geq p_{2.z} \vee z_i \leq p_{2.z} \wedge z_i \geq p_{1.z} \}$$

$$P(p_1, p_2) = \{p \mid \forall z_i \in D(p_1, p_2), p.z = z_i,$$

$$p.x = p_{1.x} + (p.z - p_{1.z})(p_{2.x} - p_{1.x}) / (p_{2.z} - p_{1.z}),$$

$$p.y = p_{1.y} + (p.z - p_{1.z})(p_{2.y} - p_{1.y}) / (p_{2.z} - p_{1.z}), \}$$

Тогда для прямоугольника разбиения в целом имеем

$$D(frag) =$$

$$D(\text{frag.p}_1, \text{frag.p}_2) \cup D(\text{frag.p}_2, \text{frag.p}_3) \cup D(\text{frag.p}_3, \text{frag.p}_4) \cup D(\text{frag.p}_4, \text{frag.p}_1)$$

$$P(\text{frag}) =$$

$$P(\text{frag.p}_1, \text{frag.p}_2) \cup P(\text{frag.p}_2, \text{frag.p}_3) \cup P(\text{frag.p}_3, \text{frag.p}_4) \cup P(\text{frag.p}_4, \text{frag.p}_1)$$

Сгруппируем эти точки по уровням дискретизации

$$\forall z_i \in D(\text{frag}), D_i(\text{frag}) = \{p_j \in P(\text{frag}) \mid p_j.z = z_i\}$$

Далее введем искусственные допущения:

- Проходящая через прямоугольник разбиения изолиния пересекает его в одной или максимум в двух точках;
- В пределах прямоугольника, проходящая через него изолиния, представляет собой отрезок прямой.

Такие допущения для большого шага разбиения будут приводить к некорректному отображению карты изолиний, но при приближении размера шага к размеру точки изображения они перестают влиять на качество отображения карты.

С учетом принятых допущений можно построить множество отрезков изолиний для прямоугольника разбиения:

$$L_i(\text{frag}) = \{L_j \mid \exists! L_j, L_j.p_1 \in D_i(\text{frag}) \wedge L_j.p_2 \in D_i(\text{frag}) \wedge L_j.p_1 \neq L_j.p_2\}$$

Объединение связанных отрезков для всех уровней дискретизации даст нам полное векторное представление карты изолиний.

Для усиления визуального эффекта и подчеркивания микрорельефа описываемой аппроксимацией поверхности выполним расчет освещения карты под некоторым углом.

Для этого у каждого фрагмента определим нормали к поверхности аппроксимации для всех его углов, т.е. расширим описание фрагмента до следующей структуры

$$\text{frag} = \langle \text{app}(P, p_{1.x}, p_{1.y}), \text{app}(P, p_{2.x}, p_{2.y}), \text{app}(P, p_{3.x}, p_{3.y}), \text{app}(P, p_{4.x}, p_{4.y}), n_1, n_2, n_3, n_4 \rangle$$

Нормали вычислим по треугольникам малых размеров, построенных в точках p_1, p_2, p_3, p_4 .

Так для p_i , построим малый треугольник $\langle p_{i1}, p_{i2}, p_{i3} \rangle$, с каждой вершиной которого связан вектор v_{i1}, v_{i2}, v_{i3} . Тогда вектор нормали будет

$$n_i = v \times u, \quad v = v_{i2} - v_{i1}, \quad u = v_{i3} - v_{i1}$$

Пусть задан вектор освещения $light$. Тогда косинус угла отклонения нормали поверхности от луча освещения находим как

$$\cos(\alpha) = \frac{|light \cdot n_i|}{|n_i| \cdot |light|}$$

Пример карты изменения потребления природного газа за 8 лет приведен на рисунке 44.

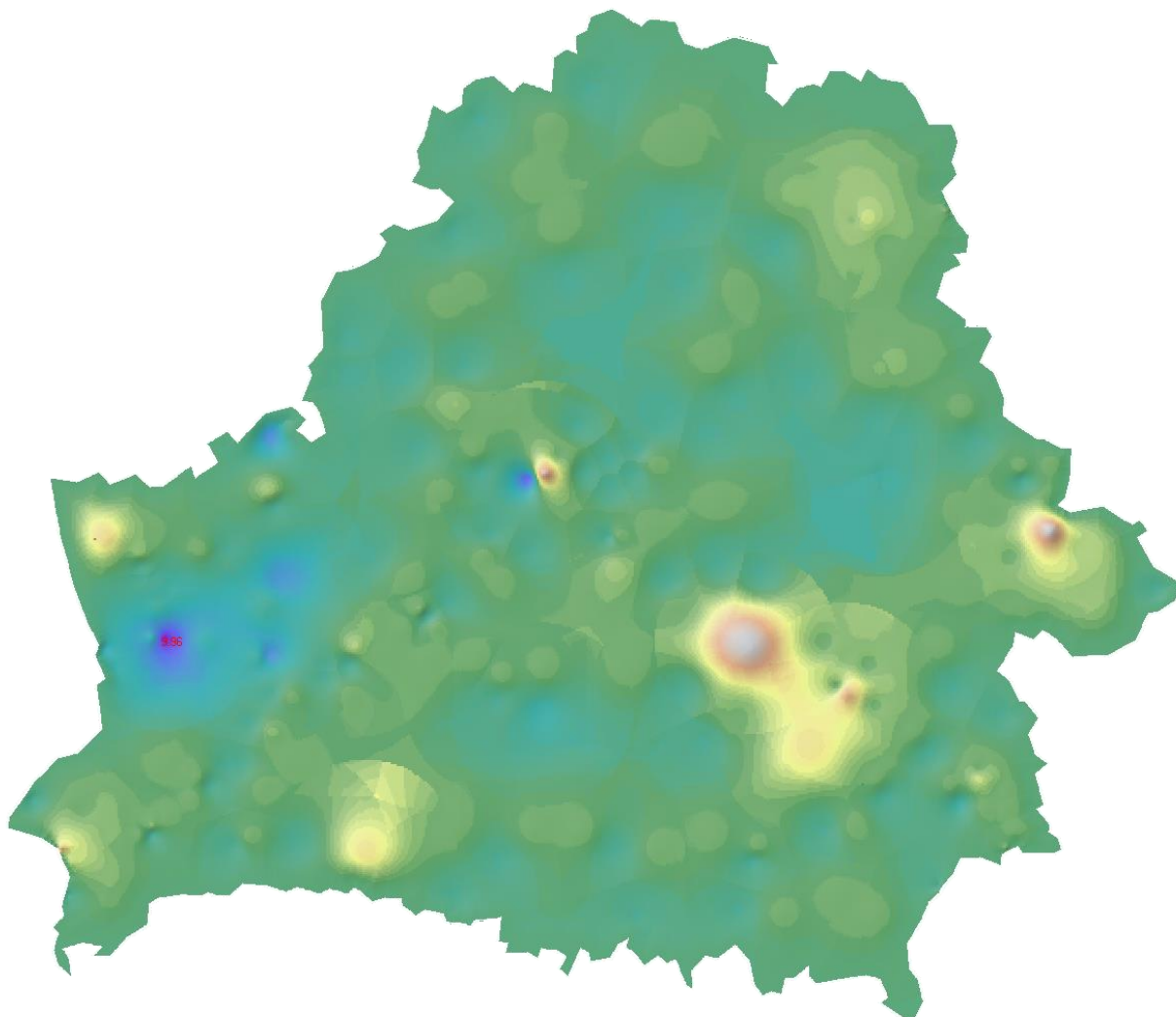


Рисунок 44 – Пример работы подсистемы визуализации карты изменения структуры потребителей природного газа

Выводы по главе 3

Предложен новый метод построения аналитической аппроксимации сложных n -мерных зависимостей, основанный на применении генетического алгоритма.

Разработанный ГА имеет следующие особенности:

кодирование пространства поиска выполняется в виде дерева разбора алгебраического выражения синтаксическим анализатором контекстно-свободной грамматики класса LR(1);

в ходе эволюционного процесса, помимо применения мутаций структуры (при условии их положительного влияния), выполняется этап мутации коэффициентов, позволяющие избегать попадания в локальные экстремумы целевой функции;

на каждом шаге эволюционного процесса предусмотрен этап поиска экстремума в пространстве коэффициентов и этап упрощения аналитической модели.

Разработанный ГА был применен для решения задачи снижения влияния нестационарных процессов в стационарных неизотермических моделях транспортировки газа. Полученные модели были внедрены в рамках программного комплекса расчета запаса газа на магистральном газопроводе ОАО «Газпром трансгаз Беларусь», что обеспечило снижение влияния переходных процессов на дисбаланс газа.

Построены модели ARIMA, ETS и нейросетевая модель GMDH для потребителей газа Республики Беларусь. Определены оптимальные параметры моделей прогнозирования на текущий момент.

Исследование моделей ARIMA для прогнозирования потребления природного газа показало, что каждая модель требует тонкой настройки параметров с учетом особенностей потребителя.

Исследована применимость модели экспоненциального сглаживания с сезонностью Хольта-Винтерса, показано, что данная модель уступает модели ARIMA по точности прогноза.

Показано, что полученные модели прогнозирования по коэффициенту детерминации относятся к моделям хорошего качества и точность прогнозирования потребления газа с использованием нейронных сетей не уступает точности метода экспоненциального сглаживания.

ЗАКЛЮЧЕНИЕ

В ходе выполнения задания разработан метод, алгоритм и программный модуль моделирования климатических изменений температуры грунта, толщины снежного покрова, силы ветра с использованием нечеткого логического аппроксиматора по данным усредненных ежемесячных оценок Республиканского центра по гидрометеорологии.

Построены методы и алгоритмы моделирования потребления газа, основанные на методах анализа временных рядов и искусственных нейронных сетей. Проведен сравнительный анализ алгоритмов прогнозирования потребления природного газа потребителями Республики Беларусь по модели с экспоненциальным сглаживанием, ARIMA и с применением искусственных нейронных сетей. Показано, что полученные модели прогнозирования по коэффициенту детерминации относятся к моделям хорошего качества и точность прогнозирования потребления газа с использованием нейронных сетей не уступает точности метода экспоненциального сглаживания.

Установлен важный научный факт о сильной отрицательной корреляции скачков температур на компрессорных станциях, связанных с пуском компрессоров, и падений оценок запаса газа по стационарной неизотермической модели для рассматриваемых участков.

Проведен анализ данных на предмет влияния различных переходных (нестационарных) процессов на оценку запаса газа по стационарной неизотермической модели. Предложен метод устранения влияния переходных процессов, связанных с ростом температуры газа при запуске компрессорных станций, основанный на введении инерционности по определенному закону роста температуры в модели. Для поиска компактных аналитических зависимостей, описывающих влияние переходного процесса на оценку запаса газа разработан генетический алгоритм поиска оптимальной компактной модели.

Систематизирована и собрана в единую базу данных информация о потреблении природного газа в Республике Беларусь с 2012г. по 2019г.

Разработана улучшенная стационарная модель, реализующая предложенный метод устранения влияния переходных температурных процессов. Разработанный программный модуль моделирования годовых климатических изменений интегрирован в программный комплекс расчета запаса газа на магистральном газопроводе ОАО «Газпром трансгаз Беларусь».

СПИСОК ИСПОЛЬЗУЕМЫХ ЛИТЕРАТУРНЫХ ИСТОЧНИКОВ

1. Глухов, А.О. Многоагентные структуры для решения задачи коммивояжера / А.О. Глухов, В.В. Трофимов. // Проблемы менеджмента : сб. науч. тр. – СПб.: Изд-во СПбГУЭФ, 2000. – Вып. 3. – С. 71–76.
2. Трофимов, В.В. Аппроксимация на многоагентных структурах / В.В. Трофимов, А.О. Глухов // Экономическая кибернетика: системный анализ в экономике и управлении : сб. научн. тр. – СПб. : Изд-во СПбГУЭФ, 2000. – Вып. 1. – С. 40–53.
3. Glukhov, A.O. Modified hybrid genetic algorithm of discreet optimization problems / Glukhov, A.O., Glukhov, D.O., Trofimov, V.V., Trofimova, L.A. // Proceedings of 2017 20th IEEE International Conference on Soft Computing and Measurements, SCM 20176 July 2017, Scopus 7970603. – St. Petersburg; Russian Federation. – Pages 417-419
4. Пожарский, Д.А. Генетический алгоритм для нахождения коэффициентов аппроксимации функции в контактных задачах для цилиндра [Электронный ресурс] / Д.А. Пожарский, Н.Б. Золотов, И.Е. Семенов // Молодой ученый. – 2017. – № 24. – С. 122–125. – Режим доступа: <https://moluch.ru/archive/158/44625/>. – Дата доступа: 09.10.2019.
5. Кильдюшов, М.С. Программа для восстановления аппроксимированных алгебраических функций от нескольких переменных по набору дискретных значений функции [Электронный ресурс] / М.С. Кильдюшов // Интернет-журнал «Науковедение». – 2015. – Т. 7, № 5. – Режим доступа: <http://naukovedenie.ru/PDF/136TVN515.pdf>
6. Кильдюшов М.С. Применение генетических алгоритмов для восстановления аппроксимированных алгебраических функций с определенной точностью / М.С. Кильдюшов // Наука и бизнес: пути развития ; Фонд развития науки и культуры, Тамбов. – 2016. – № 1 (55), – С. 25–31.

7. Driankov, D. An introduction to fuzzy control / D. Driankov, H. Hellendoorn, M. Reinfrank. – Springer-Verlag, 1993. 5. Zadeh, Lotfi A. Fuzzy Sets / Lotfi A. Zadeh // *Information & Control*. – 1965. – Vol. 8. – P. 338 – 353.
8. Zadeh, Lotfi A. Fuzzy Sets / Lotfi A. Zadeh // *Information & Control*. – 1965. – Vol. 8. – P. 338–353.
9. Glukhov, D.O. Dynamic expert system by fuzzy inference rules to automations an examination of complex objects / D.O. Glukhov // *Budownictwo i Inzynieria Srodowiska*. – Zielonogorsk : Politechnika Zielonogorska, 1998. – P. 105–109.
10. Глухов, Д.О. Мягкие вычисления для организации компьютерного представления номограмм на примере вычисления предельного коэффициента ползучести / Д.О. Глухов, Т.М. Глухова, С.П. Кундас *Вестн. Полоц. гос. ун-та. Сер С, Фундам. науки*. – 2010. – № 3. – С. 2–6.
11. Звонков, В.Б. Сравнительное исследование классических методов оптимизации и генетических алгоритмов / В.Б. Звонков, А.М. Попов // *Вестн. Сибир. гос. аэрокосмич. ун-та им. М.Ф. Решетнева*. – 2013. – С. 23–27.
12. Brown R.G. Smoothing forecasting and prediction of discrete time series. - N.Y., 1963.
13. Brown R.G., Meyer R.F. The fundamental theorem of exponential smoothing. *Oper. Res.* - 1961. - Vol.9. -№ 5.
14. Kourentzes, N.; Petropoulos, F. & Trapero, J. R. Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting*, 2014, 30, 291-302
15. Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., & Petropoulos, F. (2017). Forecasting with temporal hierarchies. *European Journal of Operational Research*, 262(1), 60-74. <https://doi.org/10.1016/j.ejor.2017.02.046>
16. Lütkepohl H. Forecasting contemporaneously aggregated vector ARMA processes. *Journal of Business & Economic Statistics*. 1984;2(3):201–214.

17. Gross CW, Sohl JE. Disaggregation methods to expedite product line forecasting. *Journal of Forecasting*. 1990;9(3):233–254.
18. Dangerfield BJ, Morris JS. Top-down or bottom-up: Aggregate versus disaggregate extrapolations. *International Journal of Forecasting*. 1992;8(2):233–241.
19. Fliedner G. An investigation of aggregate variable time series forecast strategies with specific subaggregate time series statistical correlation. *Computers & Operations Research*. 1999;26(10):1133–1149.
20. Zellner A, Tobias J. A note on aggregation, disaggregation and forecasting performance. *Journal of Forecasting*. 2000;19(5):457–465.
21. Hyndman RJ, Lee AJ, Wang E. Fast computation of reconciled forecasts for hierarchical and grouped time series. *Computational Statistics & Data Analysis*. 2016;97:16–32.
22. Abouarghoub W, Nomikos NK, Petropoulos F. On reconciling macro and micro energy transport forecasts for strategic decision making in the tanker industry. *Transportation Research Part E: Logistics and Transportation Review*. 2018;113:225–238.
23. Andrawis RR, Atiya AF, El-Shishiny H. Combination of long term and short term forecasts, with application to tourism demand forecasting. *International Journal of Forecasting*. 2011;27(3):870–886.
24. Spiliotis E, Petropoulos F, Kourentzes N, Assimakopoulos V. Cross-temporal aggregation: Improving the forecast accuracy of hierarchical electricity consumption. *Forecasting and Strategy Unit Working Paper*. 2018;1/18:1.
25. Kourentzes N, Athanasopoulos G. Cross-temporal coherent forecasts for Australian tourism. *Annals of Tourism Research*. 2019;forthcoming.
26. Yagli GM, Yang D, Srinivasan D. Reconciling solar forecasts: Sequential reconciliation. *Solar Energy*. 2019;179:391–397.
27. Akaike, H. A new look at the statistical model identification. — *IEEE Transactions on Automatic Control*. — 1974. — T. 19. — C. 716. — 723 c.

28. Yeni Rahkmawati, I Made Sumertajaya, Muhammad Nur Aidi (2019); Evaluation of Accuracy in Identification of ARIMA Models Based on Model Selection Criteria for Inflation Forecasting with the TSClust Approach; International Journal of Scientific and Research Publications (IJSRP) 9(9). – 2019. - ISSN: 2250-3153, DOI: <http://dx.doi.org/10.29322/IJSRP.9.09.2019.p9355>. – P.439-443
29. Box, George; Jenkins, Gwilym (1970). Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day. – 712p.
30. Durbin, James & Koopman, Siem Jan. (2001). Time Series Analysis by State Space Methods.
31. Wessa P., (2017), ARIMA Backward Selection (v1.0.6) in Free Statistics Software (v1.2.1), Office for Research Development and Education, URL https://www.wessa.net/rwasp_arimabackwardselection.wasp/
32. Wessa, P. (2021), Free Statistics Software, Office for Research Development and Education, version 1.2.1, URL <https://www.wessa.net/>
33. Moahmed, Tawfik & Gayar, Neamat & Atiya, Amir. (2014). Forward and Backward Forecasting Ensembles for the Estimation of Time Series Missing Data. 93-104. 10.1007/978-3-319-11656-3_9.
34. Holt, C.C. (1957) Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages. ONR Memorandum, Vol. 52, Carnegie Institute of Technology, Pittsburgh.
35. P. R. Winters (1960), Forecasting sales by exponentially weighted moving averages, Management Science 6, 324–342.
36. Гандин Л. С. Задача об оптимальной интерполяции. Тр. Глав, геофиз. обсерв. 1959. Т. 99. С. 67-75.
37. Zadeh, Lotfi A. Fuzzy Sets. – Information & Control. – Vol. 8, 1965. – P.338-353.
38. Kelly G.A. The Psychology of Personal Constructs. — N. Y.: Norton, 1955.

39. V. Trofimov, A. Gloukhov, D. Gloukhov Algorithm of ecological monitoring by fuzzy production rules / 2-nd International Conference Ecology and Society's Development Abstracts.- St.P.: МАНЭБ, 1997.- p.166.
40. Dmitry O. Glukhov Dynamic expert system by fuzzy inference rules to automations an examination of complex objects / Budownictwo i Inzynieria Srodowiska. – Zielonogorsk: Politechnika Zielonogorska, 1998. – ISBN 83-85911-60-X. – P.105-109.
41. Driankov, Hellendoorn, Reinfrank, An introduction to fuzzy control. – Springer-verlag, 1993
42. Глухов Д. О., Кастрюк А.П. Глухова Т.М. Применение унимодальных функций принадлежности в нечетких производственных системах для решения задач интеллектуального управления динамическими процессами / Вестник Полоцкого государственного университета. Фундаментальные науки. – №3. Серия С. – Новополоцк: РИО ПГУ, 2009. – 115-119с.
43. Глухов Д. О., Глухова Т.М., Кундас С.П. Мягкие вычисления для организации компьютерного представления номограмм на примере вычисления предельного коэффициента ползучести / Вестник Полоцкого государственного университета. Фундаментальные науки. – №3. Серия С. – Новополоцк: РИО ПГУ, 2010. – 2-6с.
44. Гради Буч и др. Объектно-ориентированный анализ и проектирование с примерами приложений (UML 2). Третье издание = Object-Oriented Analysis and Design with Applications (3rd Edition). — М.: «Вильямс», 2010. — 720 с. — ISBN 978-5-8459-1401-9.
45. Anastasakis L., Mort N. The development of self-organization techniques in modelling: A review of the Group Method of Data Handling (GMDH): Research Report № 813. – Sheffield, United Kingdom: The University of Sheffield, 2001. – 38 p)

46. Ivakhnenko A.G. Heuristic Self-Organization in Problems of Engineering Cybernetics. *Automatica*. 1970. vol. 6, no. 2. pp. 207–219. DOI: 10.1016/0005-1098(70)90092-0.
47. Ivakhnenko A.G. Polynomial Theory of Complex Systems. *IEEE Transactions on Systems, Man and Cybernetics*. 1971. vol. 4. pp. 364–378. DOI: 10.1109/tsmc.1971.4308320.
48. Ikeda S., Ochiai M., Sawaragi Y. Sequential GMDH Algorithm and Its Application to River Flow Prediction. *IEEE Transactions on Systems, Man and Cybernetics*. 1976. vol. 7. pp. 473–479. DOI: 10.1109/tsmc.1976.4309532.
49. Witczak M, Korbicz J, Mrugalski M., et al. A GMDH Neural Network-Based Approach to Robust Fault Diagnosis: Application to the DAMADICS Benchmark Problem. *Control Engineering Practice*. 2006. vol. 14, no. 6. pp. 671–683. DOI: 10.1016/j.conengprac.2005.04.007.
50. Kondo T., Ueno J. Multi-Layered GMDH-type Neural Network Self-Selecting Optimum Neural Network Architecture and Its Application to 3-Dimensional Medical Image Recognition of Blood Vessels. *International Journal of Innovative Computing, Information and Control*. 2008. vol. 4, no. 1. pp. 175–187.
51. Бернер, Л.И. Управление газотранспортной сетью с использованием методов моделирования и прогнозирования/ Л.И. Бернер, А.А. Ковалев, В.В. Киселев // Автоматизация, телемеханизация и связь в нефтяной промышленности. -2013. - №1. -с. 48-54.
52. Сухарев М.Г., Ставровский Е.Р. Оптимизация систем транспорта газа. – М.: Недра. 1975. – 277 с.
53. Подсистема прогнозирования газопотребления крупного промышленного кластера АСУТП магистрального транспорта газа/ Никаноров В.В. и др.//Интеллектуальные информационные технологии в управлении. - 2017. - №3. – с. 20-24.

54. Прогноз потребления газа – основа принятия рациональных решений по структуре и технологическим параметрам при проектировании и реконструкции системы газоснабжения/ Р.А. Кантюков и др. // Электронный научный журнал «Нефтегазовое дело». - 2015. - №1. – с. 201-202.
55. Леонов, Д.Г. Интеграционная архитектура адаптируемого распределённого программного обеспечения для решения задач газоснабжения / Т.М. Папилина, О.А. Степанкина// Proc. of the 6th Int. Conf. Actual Problems of System and Software Engineering, Moscow, Russia, 12-14 November, 2019. - p.259-270.
56. Степанкина, О.А. Системные основы постановки и решения задач прогнозирования для автоматизированного диспетчерского управления процессами нефтегазового производства / О.А. Степанкина, А.С. Абрамов // Автоматизация, телемеханизация и связь в нефтяной промышленности. – 2015. – № 12. – С. 19-23.
57. Тевяшев, А.Д. Сравнительный анализ методов прогнозирования процессов потребления природного газа/ А.Д. Тевяшев, В.Н. Щелкалин //Восточно-Европейский журнал передовых технологий. -2009. - №6/7(42). – с.8-15
58. Информационно-аналитическая система прогнозирования процессов потребления природного газа в газотранспортной системе Украины/ А.Д. Тевяшев и др. // ри, 2011, № 3 . – с.92-98
59. Сухарев М.Г., Самойлов Р.В. Анализ и управление стационарными и нестационарными режимами транспорта газа. – М.: Издательский центр РГУ нефти и газа (НИУ) имени И. М. Губкина, 2016. – 399 с.
60. Бокс, Д. Анализ временных рядов: прогноз и управление : пер. с англ. / Д. Бокс, Г.Дженкинс. – М. : Мир, 1974. – 406 с.
61. Сухарев, М.Г. Методы прогнозирования. Учебное пособие — М.: РГУ нефти и газа, 2009 г., 208 с.

62. Zhang, G. Forecasting with artificial neural networks / G. Zhang, B.E. Patuwo, M.Y. Hu // *International Journal of Forecasting*. - 1998.- Vol 14.- P.35–62.
63. Верзунов, С.Н. Применение глубоких нейронных сетей для краткосрочного прогноза дальности видимости / С.Н. Верзунов // *Проблемы автоматизации и управления*. – 2019. - №1 (36). – с.118-130.
64. Великанова, Л.И. Краткосрочное прогнозирование влажности горных грунтов / Л.И. Великанова // *Проблемы автоматизации и управления*. – 2015. - №4.– с. 158–166.
65. Boyd, M. Designing a neural network for forecasting financial and economic time series / M. Boyd, I. Kaastra // *Neurocomputing*. - 1996.- Vol.10.- p. 215–236.
66. Верзунов С.Н., Лыченко Н.М. Мультивейвлетная полиморфная сеть для прогнозирования геофизических временных рядов / С.Н. Верзунов, Н.М. Лыченко // *Проблемы автоматизации и управления*. – 2017. – № 1 (32). – С. 78–87.
67. Lei, Y. Prediction of length-of-day using extreme learning machine/ Y. Lei, D. Zhao, H. Cai // *Geodesy and Geodynamics*. – 2015. - Vol.6. - №. 2. – p. 151–159.
68. Прокопцев, Н. Г. Использование сверточных нейронных сетей для прогнозирования скоростей транспортного потока на дорожном графе / Н. Г. Прокопцев, А. Е. Алексеенко, Я. А. Холодов // *Компьютерные исследования и моделирование*. – 2018. – Т.10. - №3. -с. 359–367.
69. Головкин, В.А. Нейросетевые технологии обработки данных : учеб. пособие / В. А. Головкин, В. В. Краснопрошин. – Минск : БГУ, 2017. – 263с.
70. Соколова Э.С. Разработка и апробация эмулятора нейросетевого моделирования для целей прогнозирования временных рядов / Д.В. Дмитриев, Д.А. Ляхманов, Э.С. Соколова // *Электронный научный журнал «Современные проблемы науки и образования»*. – 2012. – № 6. - Режим

доступа: <https://www.science-education.ru/ru/issue/view?id=106>, Дата доступа:04.04.2020.

71. Шагалова, П.А. Нейросетевые технологии в решении задач прогнозирования / П.А. Шагалова, Д.А. Ляхманов // Электронный научный журнал «Современные проблемы науки и образования». - 2014. – № 6 . – Режим доступа: <https://www.science-education.ru/ru/article/view?id=16494>, Дата доступа:04.04.2020.
72. Степанкина, О.А. Актуальность применения нейронных сетей для прогнозирования газопотребления/О.А. Степанкина //Автоматизация, телемеханизация и связь в нефтяной промышленности. - 2013. - №11. - С. 49-52.
73. Susto, G. A. Machine learning for predictive maintenance: A multiple classifier approach / G. A. Susto, A. Schirru, S. Pampuri, S. McLoone, A. Beghi // IEEE Transactions on Industrial Informatics. – 2014. – Vol 11(3). – P. 812– 820
74. Susto, G. A., A. Data-driven prognostic method based on Bayesian approaches for direct remaining useful life prediction / A. Mosallam, K. Medjaher, N. Zerhouni // Journal of Intelligent Manufacturing. – 2014. – P. 1–20
75. Hopfield, J.J. Neural networks and physical systems with emergent collective computational abilities / J.J. Hopfield // Proc. of National Academy of Sciences. – 1982.– Vol.79.- №.8.– p.2554-2558.
76. Hochreiter, S. / Long short-term memory // S. Hochreiter, J. Schmidhuber // Neural Computation. - 1997. - Vol.9.- p. 1735–1780.
77. Вьюгин В.В. «Математические основы теории машинного обучения и прогнозирования» М.: 2013. - 387 с.
78. Джулли, А Библиотека Keras – инструмент глубокого обучения / А. Джулли, С. Пал.– М.: ДМК Пресс, 2018. – 294 с.
79. Имильбаев, Р.Р. Методы и алгоритмы прогнозирования значений контролируемых параметров газораспределительной сети по результатам обработки телеметрической информации: автореф. дис. Канд. техн. наук:

05.13.01/ Р.Р. Имильбаев – Самарский гос-ный техн. ун-т. – Самара. – 2018 – 20 с.

80. Mozer M.C. A Focused Backpropagation Algorithm for Temporal Pattern Recognition / M.C. Mozer, Y. Chauvin, D. Rumelhart. - Hillsdale, NJ: Lawrence Erlbaum Associates.- 1995.- p.137-169.
81. Васяева, Т.А. Нейросетевое прогнозирование потребления природного газа Вестник Национального технического университета Харьковский политехнический институт. Серия: Информатика и моделирование / Т.А. Васяева, С.В. Хмелевой, Е.В. Алексеева // 2010. - №31. – с.37-41.
82. Гафаров, Ф.М. Искусственные нейронные сети и приложения: учеб. пособие / Ф.М. Гафаров, А.Ф. Галимянов. – Казань: Изд-во Казан. ун-та, 2018. – 121 с.
83. Демидова, Л. А. Применение рекуррентных нейронных сетей в задаче классификации отказов работы сложных технических систем в рамках проактивного технического обслуживания/ Л. А. Демидова, Д. В. Марчев // Вестник РГРТУ. - 2019. - № 69. – с. 135-148.