

Министерство науки и высшего образования РФ  
Федеральное государственное автономное  
образовательное учреждение  
высшего образования  
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»  
Институт космических и информационных технологий  
Базовая кафедра «Интеллектуальные системы управления»

УТВЕРЖДАЮ

И.о. зав. кафедрой

\_\_\_\_\_ Д.И. Ликсонова

« \_\_\_\_ » \_\_\_\_\_ 2023 г.

**БАКАЛАВРСКАЯ РАБОТА**

27.03.03 «Системный анализ и управление»

Идентификация дискретно-непрерывных процессов в условиях неполной информации

Руководитель \_\_\_\_\_ и.о. зав. каф., канд. тех. наук Д.И. Ликсонова  
подпись, дата

Студент \_\_\_\_\_ Н.А. Усова  
подпись, дата

Нормоконтролер \_\_\_\_\_ Н.Б. Позолотина  
подпись, дата

Красноярск 2023

## РЕФЕРАТ

Бакалаврская работа на тему идентификация дискретно-непрерывных процессов в условиях неполной информации содержит 54 страницы текстового документа, 25 рисунков, 5 таблиц, 40 формул, 15 использованных источников.

ИДЕНТИФИКАЦИЯ, МНОГОМЕРНЫЙ ПРОЦЕСС, ДИСКРЕТНО-НЕПРЕРЫВНЫЙ ПРОЦЕСС, АПРИОРНАЯ ИНФОРМАЦИЯ, СОСТАВНОЙ ВЕКТОР.

Идентификация дискретно-непрерывных процессов является актуальной задачей в современной науке и технике. Она заключается в определении параметров системы на основе измерений выходных сигналов, полученных при воздействии на систему входных сигналов. В условиях неполной информации задача идентификации становится более сложной и требует применения специальных алгоритмов и методов обработки данных.

Рассматриваются основные понятия и определения, связанные с идентификацией дискретно-непрерывных процессов, а также методы и алгоритмы, используемые для решения данной задачи. Обозреваются как классические методы, так и современные методы, используемые в решении практических задач.

Проводятся вычислительные эксперименты, направленные на проверку эффективности применения методов и алгоритмов для решения задачи идентификации дискретно-непрерывных процессов в условиях неполной информации. Рассматриваются как синтетически полученные, так и реальные данные, полученные при измерении выходных сигналов систем.

Основным результатом работы является разработка методов и алгоритмов идентификации дискретно-непрерывных процессов в условиях неполной информации, которые могут быть использованы для решения практических задач в различных областях науки и техники.

## СОДЕРЖАНИЕ

Введение.....	4
1 Идентификация многомерных дискретно-непрерывных процессов.....	5
1.1 Постановка задачи моделирования.....	5
1.2 Параметрические методы моделирования.....	9
1.3 Непараметрические методы моделирования.....	12
Выводы по главе 1.....	15
2 Алгоритмы обработки данных.....	17
2.1 Центрирование и нормирование.....	17
2.2 Непараметрическая оценка функции регрессии Надарая-Ватсона.....	19
2.3 Метод наименьших квадратов.....	20
2.4 Метод стохастической аппроксимации.....	22
2.5 Линейная аппроксимация метода наименьших квадратов.....	25
Выводы по главе 2.....	28
3 Вычислительные эксперименты.....	29
3.1 Эксперименты на имитационной модели дискретно-непрерывного процесса.....	29
3.2 Технологические сведения о процессе гидроочистки и гидродепарафинизации.....	38
3.3 Применение непараметрического алгоритма для моделирования процесса гидроочистки и гидродепарафинизации дизельного топлива.....	42
Выводы по главе 3.....	51
Заключение.....	52
Список использованных источников.....	53

## ВВЕДЕНИЕ

Современный мир не представляет себе жизнь без использования компьютерных технологий и моделирования, в особенности математического. В настоящее время математическое моделирование является неотъемлемой частью многих научных и инженерных исследований. Одной из важных задач математического моделирования является моделирование дискретно-непрерывных процессов в условиях неполной информации. Эта задача актуальна для различных областей науки и техники, например управление производством, финансовая аналитика, транспортные системы и другие. В данной выпускной квалификационной работе будет рассмотрен подход к математическому моделированию дискретно-непрерывных процессов в условиях неполной информации, а также будут предложены методы решения задачи их идентификации.

Цель работы заключается в повышении точности моделирования процессов дискретно-непрерывного типа по наблюдениям с помехами.

Задачи, возникающие в ходе реализации поставленной цели, следующие:

- поиск и исследование значимой литературы в выбранной области исследований;
- исследование параметрических и непараметрических методов идентификации;
- реализация задачи идентификации для многомерных объектов на основе оценки Надарая-Ватсона;
- применение непараметрического алгоритма для моделирования процесса гидродепарафинизации дизельного топлива.

Для достижения поставленной цели применяются методы математического моделирования, такие как имитационное моделирование с использованием компьютерных технологий, а также методы математической статистики.

## **1 Идентификация многомерных дискретно-непрерывных процессов**

Дискретно-непрерывный процесс – процесс, протекающий во времени непрерывно, но измерения его переменных осуществляется через дискретные интервалы времени. Дискретные переменные могут быть описаны числами, которые изменяются только в определенные моменты времени, в то время как непрерывные переменные изменяются непрерывно. Примером дискретно-непрерывного процесса может быть система управления, где входной сигнал является дискретным, а выходной – непрерывным. Задача идентификации дискретно-непрерывных процессов заключается в определении зависимости между входными и выходными сигналами системы на основе их измерений.

### **1.1 Постановка задачи моделирования**

Моделирование еще долгое время будет неотъемлемой частью исследований во многих областях. Например, в наше время моделирование используется для прогнозирования погоды, экономических показателей, поведения потребителей, результатов спортивных соревнований и т.д. Моделирование также позволяет проводить эксперименты в виртуальной среде. Такой подход может быть более эффективным и безопасным, чем проведение экспериментов в реальном мире. Кроме того, моделирование как метод помогает улучшать процессы и оптимизировать ресурсы в ряде отраслей промышленности и производства.

Если говорить о научных исследованиях, моделирование может быть эффективнее исследования реального объекта в случае, когда изучаемый объект является опасным или недоступным для наблюдения в реальном времени. Например, при изучении поведения опасных природных явлений, таких как землетрясения, цунами или вулканические извержения. Моделирование также может быть полезным для исследования процессов, которые происходят при очень малых или очень больших масштабах, которые трудно или невозможно

наблюдать в реальности. К примеру, для изучения процессов, происходящих внутри атомов или молекул, или для изучения космических объектов, находящихся на огромных расстояниях от Земли.

Таким образом, моделирование – это процесс создания формальной структуры, которая отражает взаимодействие элементов реальной системы, и позволяет проводить анализ ее поведения в разных условиях и с разными ее параметрами. Модель может быть использована и для оптимизации реальной системы, прогнозирования ее будущего состояния, тестирования гипотез и принятия решений.

Классификация моделирования может быть различной в зависимости от контекста и области применения. Однако, можно выделить следующие основные типы моделей:

- физические модели – создаются на основе реальных объектов и материалов, частично имитирующих поведение системы в физическом пространстве;

- математические модели – используют математические формулы и уравнения для описания системы и ее поведения;

- аналитические модели – используются для анализа и оптимизации системы на основе математических методов и алгоритмов;

- компьютерные модели – создаются с помощью специальных программных средств;

- статистические модели – используются для анализа статистических данных и прогнозирования будущих событий на основе методов статистики;

- имитационные модели – это математические модели, которые используются для имитации поведения системы в различных условиях. Они включают в себя создание компьютерных моделей, которые имитируют реальную систему и ее динамику во времени.

В настоящее время развитие компьютерных технологий происходит с геометрической прогрессией. С учетом этого, моделирование без использования преимуществ технологического прогресса неэффективно. Самарский А. А. и

Михайлов А. П. выделяют три этапа математического моделирования: модель – алгоритм – программа [1].

Модель – это упрощенное представление реальной системы, которое используется для анализа и прогнозирования ее поведения. Математическая модель является основой для создания компьютерной модели.

Алгоритм – это последовательность действий, которые необходимо выполнить для решения определенной задачи. Алгоритм является частью компьютерной программы и определяет, каким образом модель будет имитировать поведение реальной системы.

Программа – это набор инструкций, написанных на выбранном исследователем языке программирования, которые позволят компьютеру выполнить поставленные перед ним задачи. Компьютерная программа создается на основе математической модели и алгоритма и используется для имитации поведения реальной системы.

Таким образом, этапы модель – алгоритм – программа включают в себя создание математической модели системы, разработку алгоритма для ее имитации и написание компьютерной программы на основе этой модели и алгоритма. Каждый этап является важным для создания эффективной компьютерной модели, которая может быть использована для решения различных задач.

Для проводимого исследования построение математической модели является важным этапом. Математические модели, как правило, отражают те свойства объекта или процесса, которые представляют интерес при исследовании, для достижения нужного уровня абстрагирования. Таким образом, представление объекта в математическом виде делает возможным численную оценку качества его функционирования и оценку принимаемых им состояний по интересующим исследователя критериям.

Существует классическая схема для идентификации объекта. Рассмотрим подробнее рисунок 1.1:

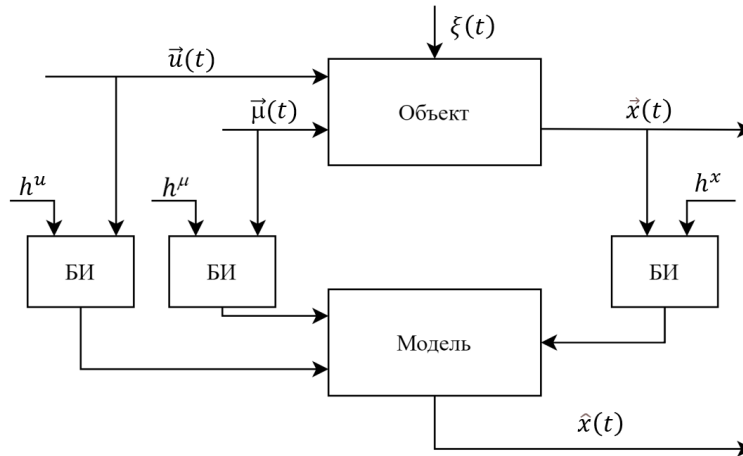


Рисунок 1.1 – Классическая схема при идентификации объекта

Рассмотрим подробнее рисунок 1.1:  $\vec{u}(t)$  – вектор входных управляемых переменных процесса,  $\vec{\mu}(t)$  – вектор входных неуправляемых, но контролируемых переменных процесса,  $\vec{x}(t)$  – вектор выходных переменных, БИ – блоки измерения,  $h^u$ ,  $h^\mu$ ,  $h^x$  – случайные помехи при измерении соответствующих входных и выходных переменных,  $\xi(t)$  – случайные помехи, действующие на процесс,  $\hat{x}(t)$  – выход модели. При моделировании дискретно-непрерывного процесса объект функционирует в реальном времени  $t$ , но измерения его входных и выходных переменных производятся через указанный интервал времени  $\Delta t$ .

В общем случае объект имеет следующий вид:

$$\vec{x}(t) = f(\vec{u}(t), \xi(t)) \quad (1.1)$$

Итак, выявление зависимостей между переменными объекта производится по наблюдениям  $x_i$ ,  $u_i$ ,  $i = \overline{1, S}$ , где  $S$  – объем выборки,  $f$  – набор уравнений, где каждому элементу вектора  $\vec{x}(t)$  ставится уравнение зависимости от набора входных переменных соответственно.

Задача математического моделирования заключается в создании математической модели, которая описывает процесс или систему с помощью



математических уравнений и выражений. Такая модель может использоваться для анализа, прогнозирования и оптимизации систем и процессов в разных областях, включая науку, технологии, экономику, и т.д. При моделировании важным аспектом является выбор подходящей модели, которая наилучшим образом соответствует реальному процессу или системе, а также разработка алгоритмов и методов для оценки параметров модели на основе имеющихся данных.

## **1.2 Параметрические методы моделирования**

Составной вектор является одним из ключевых понятий в моделировании дискретно-непрерывных процессов. Составной вектор – это вектор, состоящий из нескольких компонентов или подвекторов. Например, если у нас есть два вектора  $a$  и  $b$ , то составной вектор  $c$  может быть определен как  $c = [a, b]$ , где запятая обозначает объединение двух векторов в один. Составные векторы используются в различных областях математики и физики для описания систем, состоящих из нескольких элементов или подсистем. Например, в механике составной вектор может описывать положение тела в трехмерном пространстве, где каждая компонента соответствует координате по одной из осей.

В моделировании есть два направления методов: параметрические и непараметрические. Основное отличие между ними заключается в том, что параметрические методы требуют заранее определенной формы модели и параметров. В то же время непараметрические методы позволяют создавать модели без заранее заданной априорной информации о моделируемом объекте или процессе.

Параметрические методы моделирования – это подход к созданию моделей, при котором модель описывается набором параметров, которые могут быть изменены для получения различных вариантов модели [3]. Этот подход используется в различных областях, таких как инженерное проектирование, архитектура, промышленный дизайн и т.д.

В параметрическом моделировании модель создается путем нахождения параметров, которые определяют ее форму, размеры, свойства и другие характеристики. Эти параметры могут быть изменены, чтобы создать новые варианты модели с различными характеристиками.

Одним из широко используемых методов параметрического моделирования можно назвать метод наименьших квадратов. Метод наименьших квадратов (МНК) – это метод статистики, который применяется для оценивания параметров математической модели, которая описывает зависимость между двумя переменными. На сегодняшний день МНК представляет собой один из важнейших разделов математической статистики потому, что широко используется в различных областях науки и техники, особенно в инженерной практике [4].

Ключевым понятием в этом алгоритме является критерий метода наименьших квадратов, который заключается в том, что при оценке параметров модели выбирается такой набор параметров, сумма квадратов отклонений наблюдаемых значений которого имеет минимальное отклонение от предсказываемых значений. Другими словами, метод наименьших квадратов применяется с целью нахождения таких параметров модели, которые наилучшим образом соответствуют наблюдаемым данным. Т.е., например, если имеется набор данных, представляющих зависимость между двумя переменными, мы можем использовать МНК для построения математической модели, которая сможет описать эту зависимость. Метод наименьших квадратов позволит оценить параметры этой модели и использовать ее для прогнозирования переменных значений в будущем.

Метод наименьших квадратов прост в использовании и позволяет быстро оценить параметры модели. Кроме того, МНК является стандартным подходом к оценке параметров моделей и широко используется в научных и инженерных исследованиях.

Другим, не менее распространенным методом является метод стохастической аппроксимации. Это метод численной оптимизации, который

используется для решения задач оптимизации. Когда невозможно вычислить точный градиент целевой функции, используется последовательность случайных точек, которые приближают оптимальное решение.

Метод стохастической аппроксимации основан на итерационном процессе, в котором каждый шаг состоит из двух этапов:

- генерация случайной точки;
- обновление текущего приближения оптимального решения на основе полученной случайной точки.

Генерация случайной точки может быть выполнена различными способами. Например, случайным выбором из заданного диапазона значений или с использованием распределения вероятностей. Обновление же текущего приближения оптимального решения происходит путем изменения его координат в направлении случайной точки. Это изменение происходит с помощью формулы.

В методе стохастической аппроксимации шаг обучения выбирается таким образом, чтобы обеспечить сходимость итерационного процесса к оптимальному решению. Как правило, шаг обучения уменьшается по мере приближения к оптимальному решению.

Метод стохастической аппроксимации широко используется в машинном обучении для решения задач оптимизации, таких как обучение нейронных сетей, классификация и регрессия.

Существует метод, являющийся синтезом двух предыдущих. Линейная аппроксимация метода наименьших квадратов. Это метод численной оптимизации, который используется для решения задач линейной регрессии, когда данные содержат шум или неопределенность. Т.е. метод наименьших квадратов заключается в минимизации суммы квадратов отклонений между предсказанными значениями и реальными значениями. В линейной регрессии это выражается формулой:

$$y = a * x + b + \xi, \tag{1.2}$$

где  $y$  – зависимая переменная;

$x$  – независимая переменная;

$a$  и  $b$  – коэффициенты, которые нужно определить;

$\xi$  – ошибка или шум.

Линейная аппроксимация метода наименьших квадратов заключается в замене зависимой переменной  $y$  на линейную комбинацию независимой переменной  $x$  и коэффициентов  $a$  и  $b$ :

$$y \approx a * x + b \tag{1.3}$$

Таким образом, задача сводится к определению коэффициентов  $a$  и  $b$ , которые призваны для минимизации суммы квадратов отклонений между реальными и предсказанными значениями. Суть МНК заключается именно в минимизации суммы квадратов отклонений.

Линейная аппроксимация МНК широко используется в статистике, экономике, физике и других областях, где необходим анализ данных и построения моделей.

### **1.3 Непараметрические методы моделирования**

Непараметрическое моделирование – это метод анализа данных, который не предполагает заранее заданных параметров модели. Вместо этого, модель строится на основе априорных данных, используя методы, которые позволяют описать закономерности в данных без предварительной спецификации модели.

Непараметрические методы моделирования используются в случаях, когда неизвестна функциональная форма модели или когда сложно определить параметры модели. С их помощью становится доступным изучение связей

между входными и выходными переменными без предположений о форме зависимости. Непараметрические методы также позволяют учитывать нелинейные взаимодействия и неоднородность данных.

Непараметрические методы включают в себя такие методы, как метод  $k$ -ближайших соседей, ядерная регрессия, локально взвешенная регрессия и другие. Эти методы могут быть использованы для оценки зависимости между переменными, построения прогнозов и классификации данных, для анализа данных с выбросами или отсутствующими значениями.

Особенности непараметрических методов позволяют более гибко моделировать зависимости между входными и выходными переменными процесса и не требуют предположений о распределении данных. Однако результаты могут быть менее точными и требовать большего объема данных для получения надежных результатов. Кроме того, выбор оптимального метода сложен, поскольку часто требует экспериментального подхода.

Таким образом, непараметрические методы моделирования позволяют создавать модели в условиях неполной информации.

При написании данного исследования активно использовался один из традиционных методов моделирования в условиях малой априорной информации. Им является формула Надарая-Ватсона. Данный метод непараметрической регрессии позволяет оценить зависимость между переменными, используя взвешенную сумму значений целевой переменной в окрестности каждой точки выборки.

Основная идея метода заключается в том, что ближайшие к целевой точке значения должны иметь наибольший вес при оценке, чем более далекие значения. Вес каждого значения зависит от расстояния до целевой точки и параметра размытости  $c_s$ . Кроме того, формула Надарая-Ватсона может быть обобщена на случай многомерных данных, где каждая переменная является независимой переменной. Данный метод еще используется для оценки плотности распределения данных, где функция плотности является аналогом зависимости между переменными.

В математическом виде непараметрическая оценка функции регрессии Надарая-Ватсона выглядит следующим образом:

$$x_j^i(u) = \frac{\sum_{i=1}^S x_j^i \prod_{k=1}^m \Phi\left(\frac{u_k - u_k^i}{c_s}\right)}{\sum_{i=1}^S \prod_{k=1}^m \Phi\left(\frac{u_k - u_k^i}{c_s}\right)} \quad (1.4)$$

где  $u_k^i, x_j^i, i = \overline{1, S}, k = \overline{1, m}$  – выборка входных и выходных переменных;

$\Phi(\cdot)$  – ядро;

$u_k, k = \overline{1, m}$  – текущая переменная;

$c_s$  – параметр размытости (параметр метода, который определяет, какие точки будут учитываться при оценке значения функции в конкретной точке  $x$ ).

Параметр размытости (ширина окна)  $c_s$  в формуле Надарая-Ватсона определяет, какие точки будут учитываться при оценке значения функции в конкретной точке  $x$ . Его свойства влияют на точность и устойчивость оценки. В таблице 1.1 перечислены эти свойства.

Таблица 1.1 – Свойства параметра размытости  $c_s$

Название свойства	Обозначение свойства
Сглаживание	При увеличении ширины окна $c_s$ оценка становится более сглаженной, т.е. более устойчивой к случайным колебаниям данных. Однако, слишком большое значение $c_s$ может привести к потере важной информации и сильному сглаживанию реальной зависимости между переменными;
Смещение	Слишком маленькое значение $c_s$ может привести к недооценке реальной зависимости, тогда как слишком большое значение $c_s$ может привести к переоценке зависимости;

Продолжение таблицы 1.1

<b>Название свойства</b>	<b>Обозначение свойства</b>
Оптимальное значение	Идеальное значение параметра размытости $c_s$ зависит от особенностей данных и целей исследования. Оптимальное значение можно выбрать путем кросс-валидации или других методов оптимизации;
Ядро	Выбор ядра также влияет на свойства параметра размытости. Некоторые ядра могут требовать большего значения $c_s$ , чтобы достичь оптимальной оценки, тогда как другие могут требовать меньшего значения.

Таким образом, параметр размытости  $c_s$  существенен в формуле Надарая-Ватсона. Оптимальное значение для параметра размытости  $c_s$  выбирается с учетом особенностей исследуемых данных и целей проводимого исследования.

Непараметрическая оценка функции регрессии Надарая-Ватсона позволяет оценить нелинейную зависимость между переменными без предположений о форме этой зависимости. Однако, недостатком этого метода является то, что он требует большого количества вычислений. С увеличением объема данных, время, затрачиваемое на вычисление, закономерно и пропорционально увеличивается. Таким образом, оптимальное значение параметра размытости  $c_s$  так же избирается с помощью экспериментального подхода.

### **Выводы по главе 1**

В первой главе отмечается широкое применение моделирования процессов или объектов для исследования различных задач. Была рассмотрена классификация моделирования систем. Выбор типа моделирования зависит от

количества и качества априорной информации и требований, поставленных перед аналитиком. Если выборка содержит дефекты, их необходимо устранить перед началом исследования.

Был проведен обзор параметрических методов моделирования, таких как метод наименьших квадратов, метод стохастической аппроксимации и линейная аппроксимация метода наименьших квадратов. Среди непараметрических методов выделяется формула непараметрической оценки функции регрессии Надарая-Ватсона, которая имеет особое значение в исследовании.

Выбор данных для рассмотрения был обусловлен актуальностью соответствующих областей науки и высоким уровнем применимости этих направлений на практике, где часто возникают проблемы нехватки изначальной информации и различных дефектов выборки, требующих устранения для повышения точности модели.



## 2 Алгоритмы обработки данных

В данной главе рассматриваются теоретические сведения о применяемых в ходе исследования алгоритмах обработки данных, а также методах параметрического и непараметрического моделирования.

### 2.1 Центрирование и нормирование

Померанцев А.Л. дает определение центрирования как: «Центрирование – это разница исходной матрицы  $X$  и матрицы  $M$ , т.е.

$$\tilde{X} = X - M \quad (2.1)$$

Как правило, усреднение проводится по столбцам. Т.е. для каждого вектора  $x_j$  вычисляется среднее значение по следующей формуле:

$$p_j = \frac{(x_j^1 + \dots + x_j^i)}{S} \quad (2.2)$$

Тогда  $P = (p_1^1, \dots, p_j^i)$ » [5].

Центрирование данных считается общепринятой процедурой перед применением статистических и оптимизационных методов.

В то же время, нормирование данных не является обязательным этапом. В отличие от центрирования, нормирование не изменяет структуру данных, а лишь регулирует весовые коэффициенты различных компонентов при их обработке. Иначе говоря, это произведение исходной матрицы  $X$  и матрицы  $W$ , т.е.

$$\tilde{X} = XW \quad (2.3)$$

Матрица  $W$  – это диагональная матрица размерности  $J \times J$ . Обычно диагональные элементы  $w_{jj}$  равны обратным значениям стандартного отклонения

$$d_j = \sqrt{\frac{\sum_{i=1}^S (x_j^i - p_j)^2}{S}} \quad (2.4)$$

вычисленным для каждого столбца  $x_j$ . Нормирование по строкам (называемое также нормализацией) – это произведение матрицы  $X$  и диагональной матрицы  $W$ , т.е.

$$\tilde{X} = WX \quad (2.5)$$

При этом размерность  $W$  равна  $N \times N$ , а ее элементы  $w_{ii}$  – это обратные значения стандартных отклонений строк  $x_i^t$

Если скомбинировать методы центрирования и нормирования по столбцам и рассчитывать их для каждого элемента столбца, то получится следующая формула:

$$\tilde{x}_j^i = \frac{(x_j^i - p_j)}{d_j} \quad (2.6)$$

Померанцев А. Л. описывает в своей методической работе назначение описанных методов обработки данных так: «Нормирование данных часто применяют для того, чтобы уравнивать вклад в модель от различных переменных, учесть неравномерные погрешности, или для того, чтобы обрабатывать совместно разные блоки данных. Нормирование также можно рассматривать как метод, позволяющий стабилизировать вычислительные алгоритмы. В то же

время, к этому преобразованию нужно относиться с большой осторожностью, т.к. оно может сильно исказить результаты качественного анализа. Любое преобразование данных – центрирование, нормирование, – всегда делается сначала на обучающем наборе. По этому набору находятся значения  $p_j$  и  $d_j$ , которые затем применяются и к обучающему, и к проверочному набору» [5].

## 2.2 Непараметрическая оценка функции регрессии Надарая-Ватсона

Данный метод уже кратко рассматривался в главе 1. Рассмотрим его более подробно. Для многомерной системы непараметрическая оценка функции регрессии Надарая-Ватсона имеет следующий вид:

$$x_j^i(u) = \frac{\sum_{i=1}^S x_j^i \prod_{k=1}^m \Phi\left(\frac{u_k - u_k^i}{c_s}\right)}{\sum_{i=1}^S \prod_{k=1}^m \Phi\left(\frac{u_k - u_k^i}{c_s}\right)}, j = \overline{1, n}, \quad (2.7)$$

где  $u_k^i, x_j^i, i = \overline{1, S}, k = \overline{1, m}$  – выборка измерений переменных;

$u_k, k = \overline{1, m}$  – текущая переменная;

$c_s$  – параметр размытости (параметр метода, который определяет, какие точки будут учитываться при оценке значения функции в конкретной точке  $x$ ).

Колоколообразные функции  $\Phi\left(\frac{u_k - u_k^i}{c_s}\right)$  в данной работе принимаются в виде треугольного ядра, изображенного на рисунке 2.1:

$$\Phi\left(\frac{u_k - u_k^i}{c_s}\right) = \begin{cases} 1 - \frac{|u_k - u_k^i|}{c_s}, & \frac{|u_k - u_k^i|}{c_s} < 1 \\ 0, & \frac{|u_k - u_k^i|}{c_s} \geq 1 \end{cases} \quad (2.8)$$

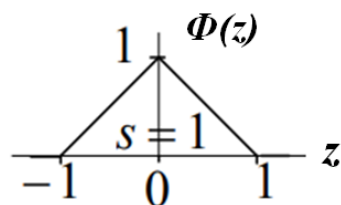


Рисунок 2.1 – Треугольное ядро

Непараметрическая оценка функции регрессии будет использоваться в практической части для исследований с тестовыми данными и с реальными.

### 2.3 Метод наименьших квадратов

Метод наименьших квадратов (МНК) – это статистический метод, который используется для аппроксимации зависимости между двумя переменными. Он используется для оценки параметров линейной регрессии, которая является моделью, предсказывающей значение одной переменной на основе значения другой переменной.

Метод наименьших квадратов заключается в минимизации суммы квадратов отклонений между фактическими значениями зависимой переменной и предсказанными значениями, полученными с помощью линейной регрессии. В результате применения МНК определяются коэффициенты уравнения линейной регрессии, которые могут быть использованы для прогнозирования значений зависимой переменной.

Агафонов Е. Д. и Шестернева О. В. в своем методическом пособии [3] описывают задачу МНК следующим образом: «Пусть величина  $X$  в течение измерения принимает следующие значения:  $x_1, x_2, \dots, x_5$ . Для простоты будем считать, что все измеренные значения равновероятны. Наилучшей оценкой величины  $X$  будет выборочное среднее:

$$\bar{x} = \frac{1}{S} \sum_{i=1}^S x_j^i \quad (2.9)$$

Отклонением назовем разность между выборочным значением и выборочным средним:  $x_j^i - \bar{x}$ ».

Чтобы охарактеризовать рассеяние наблюдаемых значений вокруг среднего значения, необходимо ввести характеристику, которую принято называть выборочной дисперсией:

$$D = \frac{1}{S} \sum_{i=1}^S (x_j^i - \bar{x})^2 \quad (2.10)$$

Дисперсия имеет важное свойство, которое нельзя не упомянуть:  $D \geq 0$ .

Пусть существует некоторая величина  $\tilde{x} = \bar{x} + \Delta$ , где  $|\Delta| > 0$ . В своем методическом пособии исследователи доказали: «Сумма квадратов отклонений выборочных значений от среднего будет всегда меньше, чем сумма квадратов отклонений выборочных значений от какой бы то ни было другой величины:

$$\begin{aligned} \frac{1}{S} \sum_{i=1}^S (x_j^i - \tilde{x})^2 &= \frac{1}{S} \sum_{i=1}^S (x_j^i - (\bar{x} + \Delta))^2 = \frac{1}{S} \sum_{i=1}^S (x_j^i - 2x_i(\bar{x} + \Delta) + \\ &+ (\bar{x} + \Delta)^2) = \bar{x}^2 - 2\bar{x}(\bar{x} + \Delta) + (\bar{x} + \Delta)^2 = \bar{x}^2 - 2\bar{x}^2 - 2\bar{x}\Delta + \bar{x}^2 + \\ &+ 2\bar{x}\Delta + \Delta^2 = \bar{x}^2 - \bar{x}^2 + \Delta^2 = D + \Delta^2 \end{aligned} \quad (2.11)$$

Таким образом, выборочное среднее наилучшим образом на основании имеющейся выборки приближает к истинному значению величины  $X$ . Критерием здесь служит средний квадрат отклонений выборочных значений от оценки величины  $X$ » [4]. Иначе говоря, среднее значение будет являться решением задачи оптимизации:

$$\frac{1}{S} \sum_{i=1}^S (x_j^i - \tilde{x})^2 = \min_{\tilde{x}}. \quad (2.12)$$

Приведенный критерий именуется критерием метода наименьших квадратов. В теории идентификации критерий МНК позволяет строить параметрические модели регрессионного типа.

Пусть задана функциональная зависимость с точностью до вектора параметров:  $x = x(u, \alpha)$ , а также доступна выборка измерений  $[u_j^i, x_j^i]$ ,  $i = 1, 2, \dots, S$ . Для оценивания параметров этой модели используем критерий МНК:

$$J = \frac{1}{S} \sum_{i=1}^S (x_j^i - x(u_j^i, \alpha))^2 \rightarrow \min_{\alpha}. \quad (2.13)$$

Метод наименьших квадратов является одним из основных инструментов статистического анализа и широко используется в различных областях, таких как экономика, финансы, инженерия и наука о данных.

#### 2.4 Метод стохастической аппроксимации

В методическом пособии Агафонова Е. Д. и Шестерневой О. В. описан также и метод стохастической аппроксимации [3]. Метод стохастической аппроксимации позволяет определить единственный корень уравнения  $x(u) = 0$  в случае, когда значение этой функции измеряется с некоторой случайной ошибкой (помехой):

$$x(u) = \hat{x}(u) + \eta \quad (2.14)$$

Алгоритмы, реализующие метод, получили название алгоритмов Робинса-Монро. Затем подобные алгоритмы были использованы для отыскания экстремума унимодальной функции  $x = x(u)$ , также включающей случайный шум согласно уравнению (2.14). Алгоритмы такого типа называют алгоритмами Кифера-Волфовитца. А. Дворецким была сформулирована и доказана теорема, которая дала наиболее слабые условия сходимости для любых типов алгоритмов стохастической аппроксимации. Пользуясь результатами этой теоремы, можно построить различные алгоритмы идентификации параметров модели.

Получим простейший алгоритм стохастической аппроксимации на примере решения задачи по отысканию оценки истинного значения величины  $X$  на основе ее  $S$  измерений:  $x_i, I = 1, 2, \dots, S$ . Пусть  $x_i = \hat{x}(u) + \eta_i$ , где  $\eta_i$  – некоторая случайная помеха с нулевым средним и конечной дисперсией:  $M\{\eta_i\} = 0, M\{\eta_i^2\} < \infty$ . Из математической статистики известно, что наилучшей оценкой  $X$  будет среднее арифметическое ее  $S$  измерений:

$$\bar{x}_s = \frac{1}{S} \sum_{i=1}^S x_j^i \quad (2.15)$$

С учетом закона больших чисел Колмогорова запишем:

$$\lim_{S \rightarrow \infty} \left\{ \frac{1}{S} \sum_{i=1}^S x_j^i \right\} = x \quad (2.16)$$

Получим рекуррентную формулу для вычисления  $\overline{x_{S+1}}$  по среднему значению выборки  $\bar{x}_S$  после  $S$  измерений и поступившему новому измерению  $x_{S+1}$ . По формуле (2.11) для  $S + 1$  измерений имеем:

$$\overline{x_{S+1}} = \frac{1}{S+1} \sum_{i=1}^{S+1} x_i = \frac{1}{S+1} \sum_{i=1}^S x_j^i + \frac{1}{S+1} x_{S+1} \quad (2.17)$$

Подставляя выражение (2.15) в (2.16), получим:

$$\overline{x_{S+1}} = \frac{S}{S+1} \overline{x_S} + \frac{1}{S+1} x_{S+1} \quad (2.18)$$

Последнее равенство преобразуем к виду:

$$\overline{x_{S+1}} = \overline{x_S} + \frac{1}{S+1} (x_{S+1} - \overline{x_S}) \quad (2.19)$$

Для определения оценки параметра  $x$  по его зашумленным наблюдениям используется алгоритм стохастической аппроксимации, который является рекуррентным выражением. При этом новая оценка будет равна сумме старой оценки и корректирующего члена с некоторым весовым коэффициентом  $\gamma_S = 1/(S+1)$ . С увеличением  $n$  указанные весовые коэффициенты образуют убывающую последовательность

$$\lim_{S \rightarrow \infty} \gamma_S = 0, \quad (2.20)$$

которая обладает еще двумя важнейшими асимптотическими свойствами:

$$\lim_{S \rightarrow \infty} \sum_{i=0}^S \gamma_i = \infty, \quad (2.21)$$

$$\lim_{S \rightarrow \infty} \sum_{i=0}^S \gamma_i^2 < \infty. \quad (2.22)$$



Кроме последовательности вида этим условиям удовлетворяют и многие другие последовательности, например  $\gamma_S = 1/S^q$ , где  $q > 1/2$ . Таким образом, любой рекуррентный алгоритм вида (2.19) или в более общем случае вида  $\alpha_{S+1} = \alpha_S - \gamma_S \mu(\alpha_S)$ , где последовательность весовых коэффициентов при корректирующем члене удовлетворяет условиям (2.20)-(2.22).

Алгоритмы стохастической аппроксимации обладают значительным преимуществом, которое заключается в том, что для их использования не требуется специальных знаний о вероятностном распределении или статистических характеристиках ошибки измерения, за исключением того, что ошибки измерения должны быть независимыми случайными величинами с нулевым средним и ограниченной дисперсией для несмещенности оценок.

На практике, обеспечение этих условий обычно не вызывает затруднений. Кроме того, алгоритмы стохастической аппроксимации являются очень простыми в вычислениях. Однако, они не обладают оптимальными свойствами и медленно сходятся к истинному значению параметра, а также не пригодны для поиска глобального экстремума многоэкстремальных функций. Эта проблема была решена во многих алгоритмах, которые ускоряют процедуру сходимости по сравнению с обычным алгоритмом стохастической аппроксимации.

Один из наиболее эффективных алгоритмов, который обеспечивает сходимость по вероятности и ускоряет скорость сходимости, заключается в том, что коэффициенты остаются постоянными до тех пор, пока корректирующий член сохраняет свой знак, и только после изменения знака этого члена берется следующий элемент последовательности.

## **2.5 Линейная аппроксимация метода наименьших квадратов**

Построим итерационную процедуру расчета параметра модели в соответствии с критерием наименьших квадратов. На каждой итерации будем использовать линейную аппроксимацию выхода модели по параметру

$$x(\alpha^{i+1}) \approx x(\alpha^i) + \frac{dx(\alpha^i)}{d\alpha} \Delta\alpha^{i+1}, \quad (2.23)$$

где  $\alpha^{i+1} = \alpha^i + \Delta\alpha^{i+1}$  – приращение параметра;

$\alpha^i, \alpha^{i+1}$  – значения параметра на соответствующих итерациях.

После подстановки линейной аппроксимации (2.23) в критерий МНК относительно  $\Delta\alpha^{i+1}$  получаем:

$$J(\Delta\alpha^{i+1}) = \sum_{i=1}^s (x_j^i - x(u_j^i, \alpha^i) - \frac{dx(u_j^i, \alpha^i)}{d\alpha} \Delta\alpha^{i+1})^2 \rightarrow \min_{\Delta\alpha^{i+1}}. \quad (2.24)$$

Перейдем к векторной форме МНК. Запишем уравнение для нахождения вектора приращений параметров. Полный дифференциал функции многих переменных выражается как

$$dx = \frac{\partial x}{\partial \alpha_1} d\alpha_1 + \frac{\partial x}{\partial \alpha_2} d\alpha_2 + \dots + \frac{\partial x}{\partial \alpha_j} d\alpha_j \quad (2.25)$$

Оценка приращения функции выхода модели имеет следующий вид:

$$\Delta x \approx \frac{\partial x}{\partial \alpha_1} \Delta\alpha_1 + \frac{\partial x}{\partial \alpha_2} \Delta\alpha_2 + \dots + \frac{\partial x}{\partial \alpha_j} \Delta\alpha_j \quad (2.26)$$

Сформируем вектор приращений параметров, матрицу  $\frac{dx}{d\alpha}$  частных производных по компонентам вектора параметров, а также вектор  $X(a^i)$  выходных значений модели, рассчитанные в выборочных точках:

$$\Delta\alpha^{i+1} = \begin{pmatrix} \Delta\alpha_1^{i+1} \\ \Delta\alpha_2^{i+1} \\ \dots \\ \Delta\alpha_j^{i+1} \end{pmatrix},$$

$$\frac{dX}{d\alpha} = \begin{pmatrix} \frac{\partial x(u_1, \alpha^l)}{\partial \alpha_1} & \frac{\partial x(u_1, \alpha^l)}{\partial \alpha_2} & \dots & \frac{\partial x(u_1, \alpha^i)}{\partial \alpha_j} \\ \frac{\partial x(u_2, \alpha^l)}{\partial \alpha_1} & \frac{\partial x(u_2, \alpha^l)}{\partial \alpha_2} & \dots & \frac{\partial x(u_2, \alpha^i)}{\partial \alpha_j} \\ \dots & \dots & \dots & \dots \\ \frac{\partial x(u_i, \alpha^l)}{\partial \alpha_1} & \frac{\partial x(u_i, \alpha^l)}{\partial \alpha_2} & \dots & \frac{\partial x(u_i, \alpha^i)}{\partial \alpha_j} \end{pmatrix},$$

$$X(\alpha^i) = \begin{pmatrix} x(u_1, \alpha^1) \\ x(u_2, \alpha^1) \\ \dots \\ x(u_i, \alpha^1) \end{pmatrix}.$$

Необходимое условие минимума приводит к системе линейных алгебраических уравнений:

$$\frac{dX}{d\alpha} \Delta\alpha^{i+1} = H - X(\alpha^i) \quad (2.27)$$

решение которой записывается в виде

$$\Delta\alpha^{i+1} = \left[ \left( \frac{dX}{d\alpha} \right)^T \left( \frac{dX}{d\alpha} \right) \right]^{-1} \left( \frac{dX}{d\alpha} \right)^T (H - X(\alpha^1)) \quad (2.28)$$

После этого находим параметры модели на  $(i + 1)$ -ой итерации:

$$\alpha^{i+1} = \alpha^i + \gamma^i \Delta \alpha^{i+1}, i = 0, 1, 2, \dots \quad (2.29)$$

Положительный коэффициент  $\gamma^i > 0$  выбирается по условию  $J(\alpha^{i+1}) \leq J(\alpha^i)$  монотонной сходимости по функции качества. Примером выбора последовательности для  $\gamma$  служит последовательность: 1, 1/2, 1/4, ... .

## **Выводы по главе 2**

Глава 2 включает в себя обзор методов анализа данных, которые применялись в данном исследовании. Были рассмотрены центрирование и нормирование, которые позволяют привести данные к стандартному виду и упростить их анализ. Подробно была изучена непараметрическая оценка функции регрессии Надарая-Ватсона, которая позволяет оценить зависимость между двумя переменными без предположений о распределении данных.

Также, метод наименьших квадратов используется для оценки параметров линейной регрессии и минимизации суммы квадратов отклонений между наблюдаемыми значениями и значениями, предсказанными моделью. Методу стохастической аппроксимации, который используется для оптимизации функций в условиях неопределенности, тоже было уделено внимание. Последней была рассмотрена линейная аппроксимация метода наименьших квадратов. Данный метод позволяет оценить зависимость между переменными с помощью линейной модели.

Данные методы анализа данных используются в различных областях, включая экономику, финансы, маркетинг, медицину, науку о данных и многие другие. Каждый из этих методов может быть применен в зависимости от конкретной задачи и характера данных. Итак, глава 2 представляет собой обзор различных методов анализа данных, каждый из которых может быть применен в зависимости от конкретной задачи и характера данных.

### 3 Вычислительные эксперименты

Основная часть данной главы посвящена вычислительным экспериментам на выборках данных: тестовой и реальной. Также будут представлены общие и технологические сведения о процессе гидроочистки и гидродепарафинизации дизельного топлива, который лежит в основе исследований на реальных данных.

#### 3.1 Эксперименты на имитационной модели дискретно-непрерывного процесса

Важной частью исследования стали эксперименты на модели. Для этого понадобилось создать имитационную модель на основе математических выражений. Была сгенерирована выборка входов и выходов с помощью предварительных вспомогательных расчетов.

Пусть на вход объекта исследования подаются управляющие воздействия  $u_1, u_2, u_3, u_4, u_5$ , а на выходе объекта измеряются переменные  $x_1, x_2, x_3, x_4$  соответственно. Для входных переменных были установлены отношения, представленные в таблице 3.1:  $u_1$  – независимая переменная, принимаемые значения которой равномерно распределяются в интервале [5;11]. Значения генерируются для каждого нового запуска программы с помощью модуля random (язык Python). Переменные  $u_2$ - $u_5$  связаны между собой зависимостями.

Таблица 3.1 – Вид зависимости между входными переменным объекта

Переменная	Вид зависимости
$u_1$	[5, 11]
$u_2$	$0,1 \cdot u_1$
$u_3$	$u_1 + u_2$
$u_4$	$0,6 \cdot u_1$

$u_5$	$0,4 \cdot u_3$
-------	-----------------

Для генерации выходов использовались вспомогательные переменные  $f_1$ - $f_4$ . Они также содержат в себе сгенерированные входы:

$$\begin{cases} f_1 = u_1^2 - 3,4 \cdot u_3 + 5 \cdot u_4 \\ f_2 = 10 \cdot u_2 \\ f_3 = u_3 + u_4 + u_5 \\ f_4 = 6 \cdot u_1 - u_4^2 \end{cases} \quad (3.1)$$

Наконец, выходные переменные генерировались по следующим уравнениям:

$$\begin{cases} x_1 = f_1 + 0,4 \cdot x_2 + 1,2 \cdot x_3 \\ x_2 = f_2 + 0,2 \cdot x_4 \\ x_3 = 0,6 \cdot x_1 - 1,4 \cdot x_2 + f_3 \\ x_4 = f_4 - 1,5 \cdot x_2 \end{cases} \quad (3.2)$$

Таким образом задача идентификации дискретно-непрерывного процесса на тестовой выборке сводится к решению линейной системы уравнений. Для описания на языке программирования Python задаваемой математической модели ее необходимо было описать в следующем виде:

$$\begin{cases} x_1^i = f_1 + \frac{448 \cdot f_4 + 1960 \cdot f_2 + 1170 \cdot f_1 + 1950 \cdot f_3}{455} + \xi_i \\ x_2^i = \frac{0,2 \cdot f_4 - f_2}{1,3} + \xi_i \\ x_3^i = \frac{0,78 \cdot f_1 + 1,4 \cdot f_2 + 1,3 \cdot f_3 - 0,28 \cdot f_4}{0,364} + \xi_i \\ x_4^i = \frac{f_4 - 1,5 \cdot f_2}{1,3} + \xi_i \end{cases} \quad (3.3)$$

Помеха  $\xi_i$ , приложенная к выходным переменной объекта, для каждого такта измерений  $i$  генерируется по следующей формуле:

$$\xi_i = k \cdot x_j^i \cdot b, \quad (3.4)$$

где  $x_j^i$  – значение выходной переменной  $x_j^i$   $j$ -того объекта в  $i$ -ый такт измерений;

$k$  – коэффициент помехи измерения (при первой генерации, для 5-ти % помехи  $k=0.05$ );

$b$  – случайное число из интервала  $[-1; 1]$ .

Точность моделирования оценивалась с помощью следующей формулы:

$$\delta_j = \frac{\sum_{i=1}^s |x_j^i - x_j^i(u)|}{\sum_{i=1}^s |x_i - \hat{x}|}, \quad j = \overline{1, n}, \quad (3.5)$$

где  $x_j^i$  – наблюдения на модели объекта;

$x_j^i(u)$  – прогноз выхода модели объекта;

$\hat{x}$  – среднее значение по каждому элементу выхода  $\bar{x}$ .

Примем, что ошибка нормирована и её значения принадлежат интервалу  $[0, 1]$ . Интерпретировать полученные значения будем следующим образом: чем меньше значение, тем точнее модель описывает рассматриваемый объект.

Модель строилась с использованием непараметрической оценки функции регрессии (2.7). На рисунке 3.1 представлен результат моделирования выходной переменной  $x_4$  объекта при помехе в 5%.

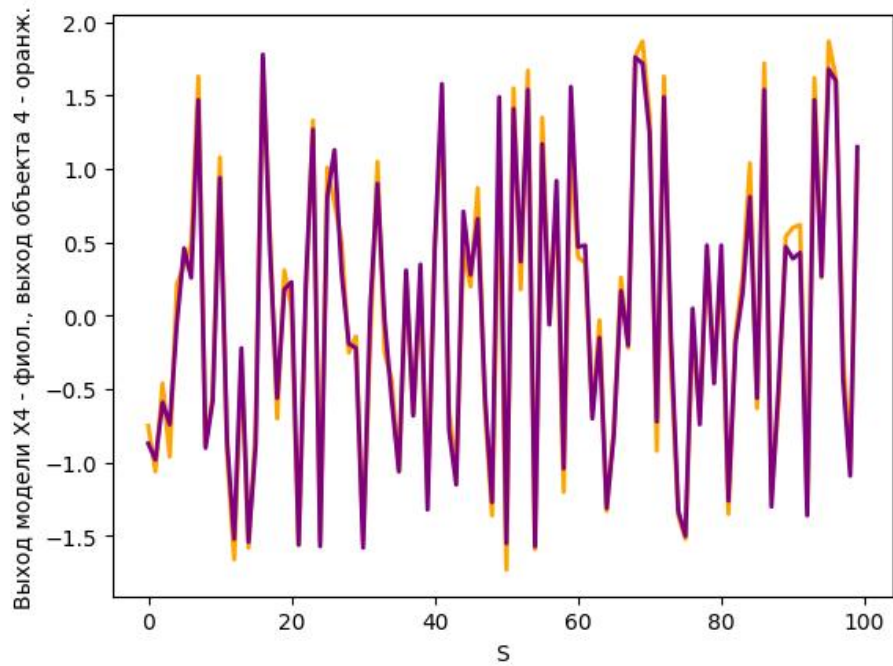


Рисунок 3.1 – Результат моделирования выходной переменной  $x_4$  объекта

Рассмотрим изменения точности моделирования в зависимости от значений, принимаемых параметром размытости  $c_s$ . Значения параметра размытости  $c_s$  постепенно увеличиваются от 0,001 до 5 с шагом 0,01. Точность моделирования  $\delta$  выходов  $x_j^i$  вычисляется для каждой итерации  $i$ .

На рисунках 3.2-3.5 изображены зависимости значений точности моделирования  $\delta(3.5)$  от изменения параметра размытости  $c_s$ .

При первом успешном запуске модели получены минимальные значения ошибок при моделировании выходов  $x_i^{<1>}, x_i^{<2>}, x_i^{<3>}, x_i^{<4>}$  с округлением до тысячных получились  $\delta_1 = 0.089$ ,  $\delta_2 = 0.150$ ,  $\delta_3 = 0.084$ ,  $\delta_4 = 0.141$  соответственно.



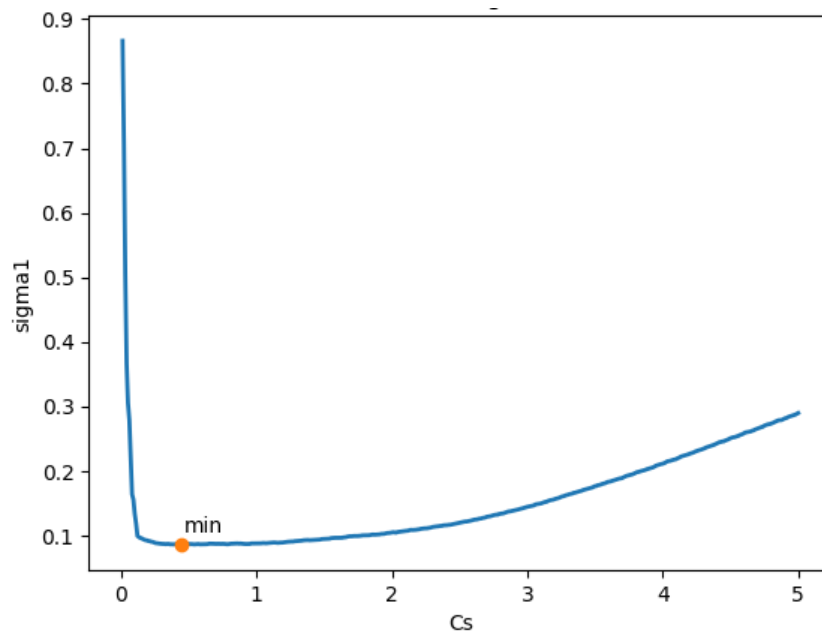


Рисунок 3.2 – Зависимость  $\delta_1$  от значения  $c_s$  для 1-го выхода

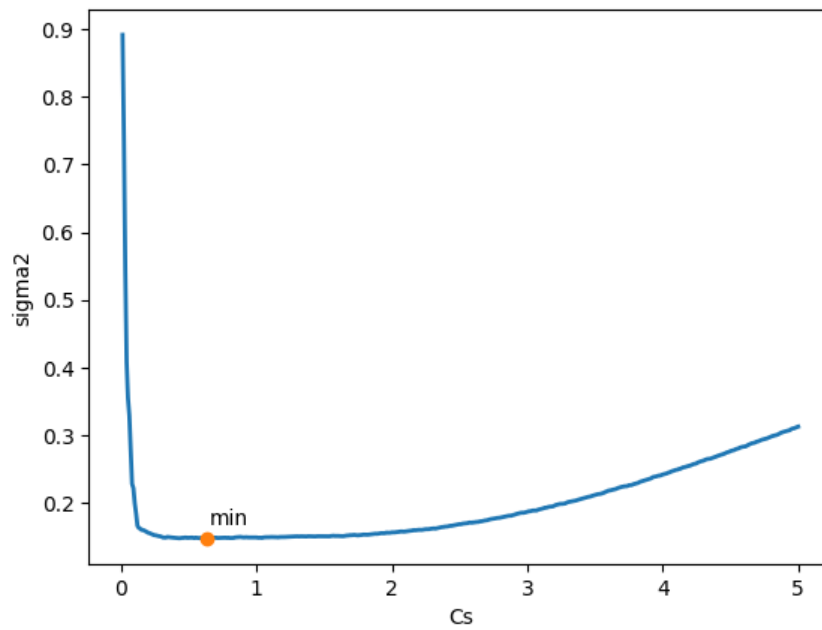


Рисунок 3.3 – Зависимость  $\delta_2$  от значения  $c_s$  для 2-го выхода

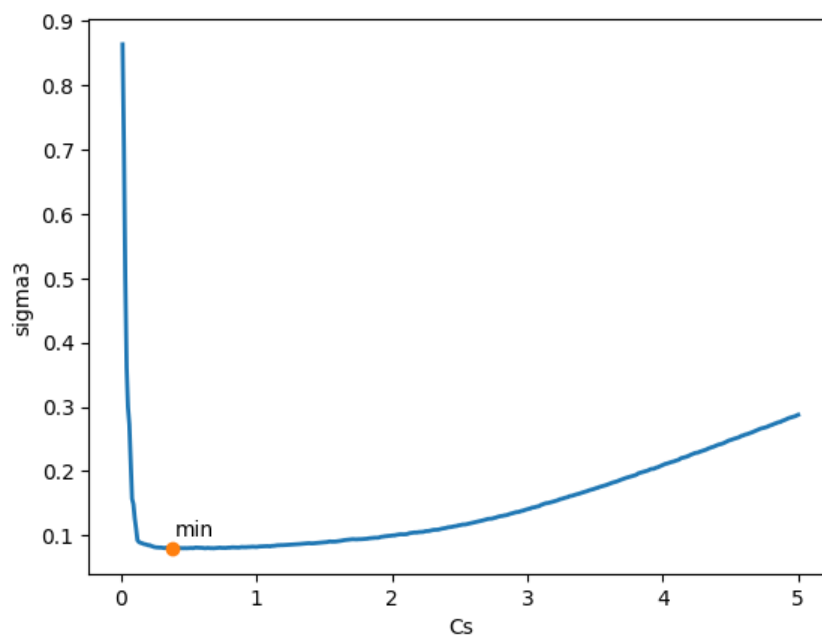


Рисунок 3.4 – Зависимость  $\delta_3$  от значения  $c_s$  для 3-го выхода

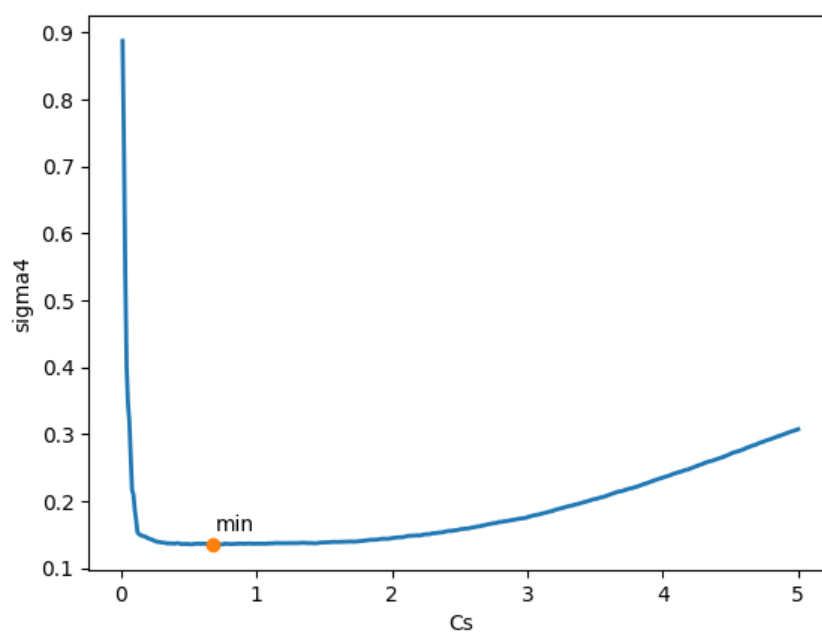


Рисунок 3.5 – Зависимость  $\delta_4$  от значения  $c_s$  для 4-го выхода

В режиме скользящего экзамена поведения этих графиков имеют много общего. Например:

- в промежутке от 0 до 1 наиболее заметно уменьшение значения ошибки;

- после достижения оптимума, значение ошибки медленно начинает возрастать;

- наиболее очевидное увеличение значений ошибки на промежутке [3;5].

Проведем серию экспериментов с изменением значений некоторых параметров модели: объем выборки  $S$  и помеха  $\xi$ . Результаты будем вносить в таблицу 3.2.

Таблица 3.2 – Результаты серии экспериментов

$N\textcircled{}$	$S$	$C_{S1}$	$C_{S2}$	$C_{S3}$	$C_{S4}$	$\delta_1$	$\delta_2$	$\delta_3$	$\delta_4$	$\xi$
1	100	0.411	0.601	0.361	0.621	0.089	0.150	0.084	0.141	5%
2	200	0.581	0.601	0.381	0.521	0.089	0.152	0.083	0.141	5%
3	500	0.391	0.531	0.401	0.591	0.085	0.145	0.079	0.136	5%
4	700	0.231	0.221	0.341	0.261	0.086	0.149	0.080	0.138	5%
5	1000	0.361	0.431	0.371	0.481	0.085	0.145	0.079	0.136	5%
6	500	0.111	0.081	0.091	0.101	0.004	0.010	0.004	0.004	0%
7	500	0.431	0.411	0.151	0.391	0.052	0.089	0.049	0.083	3%
8	500	0.571	1.111	0.681	1.001	0.119	0.205	0.110	0.192	7%
9	500	0.651	1.171	0.831	1.261	0.170	0.285	0.158	0.268	10%
10	500	0.821	1.301	1.211	1.101	0.248	0.413	0.231	0.388	15%
11	100	0.321	0.371	0.331	0.371	0.053	0.092	0.049	0.086	3%
12	300	0.421	0.921	0.531	0.651	0.124	0.213	0.116	0.199	7%
13	500	0.581	0.721	0.461	0.761	0.164	0.276	0.152	0.259	10%
14	700	0.601	1.061	0.651	1.021	0.251	0.421	0.233	0.395	15%
15	900	0.851	0.981	0.841	1.041	0.334	0.533	0.312	0.504	20%

В ходе исследования стало очевидно влияние помехи  $\xi$  и объема выборки  $S$  на точность моделирования  $\delta$ . Так, при постепенном увеличении выборки ( $N\textcircled{}$  1-5) было замечено, что чем больше объем выборки, тем выше точность моделирования. Минимальное значение ошибки в этой части исследований

достигла значения 0,079. Рассмотрим рисунки 3.6-3.9, иллюстрирующие зависимость точности моделирования  $\delta$  от параметра размытости  $c_s$ .

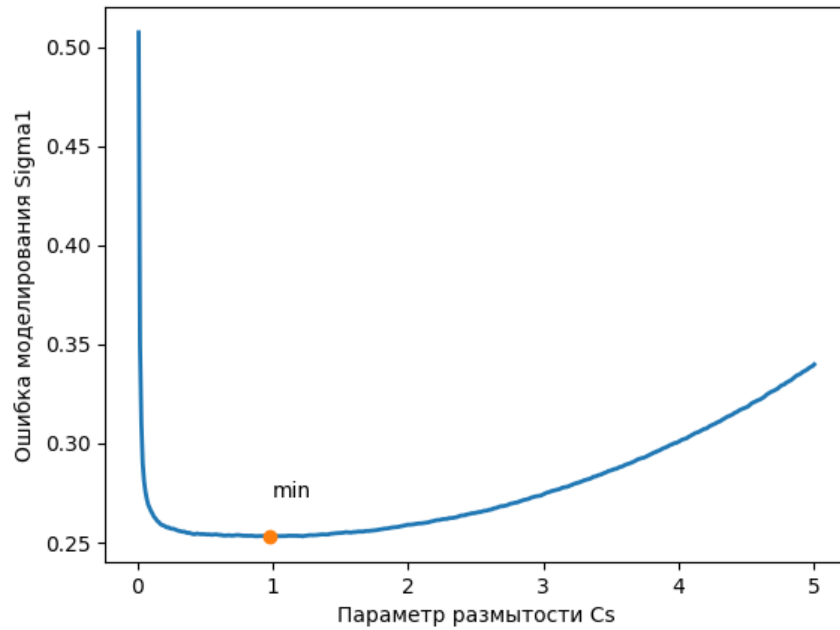


Рисунок 3.6 – Зависимость  $\delta_1$  от значения  $c_s$  для 1-го выхода при  $N = 500$  и  $\xi = 5\%$

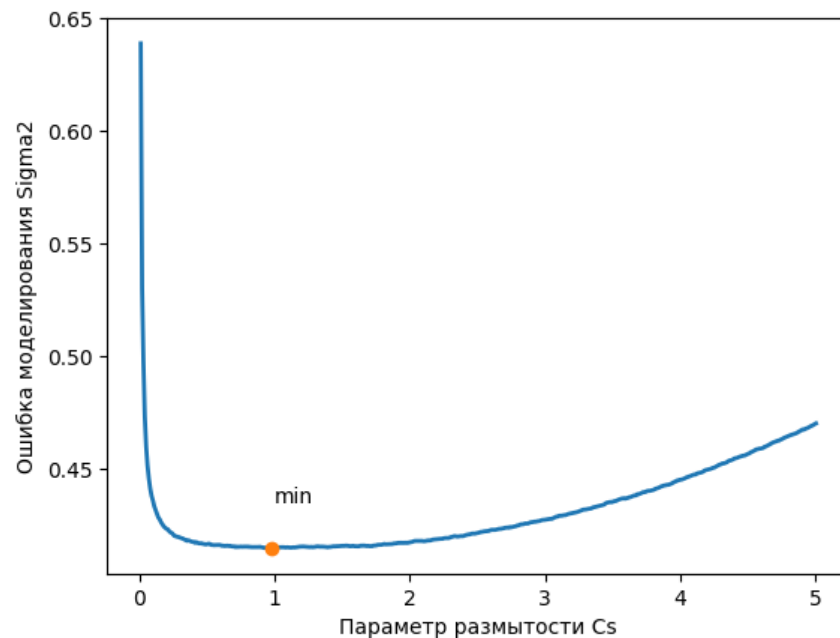


Рисунок 3.7 – Зависимость  $\delta_2$  от значения  $c_s$  для 2-го выхода при  $N = 500$  и  $\xi = 5\%$

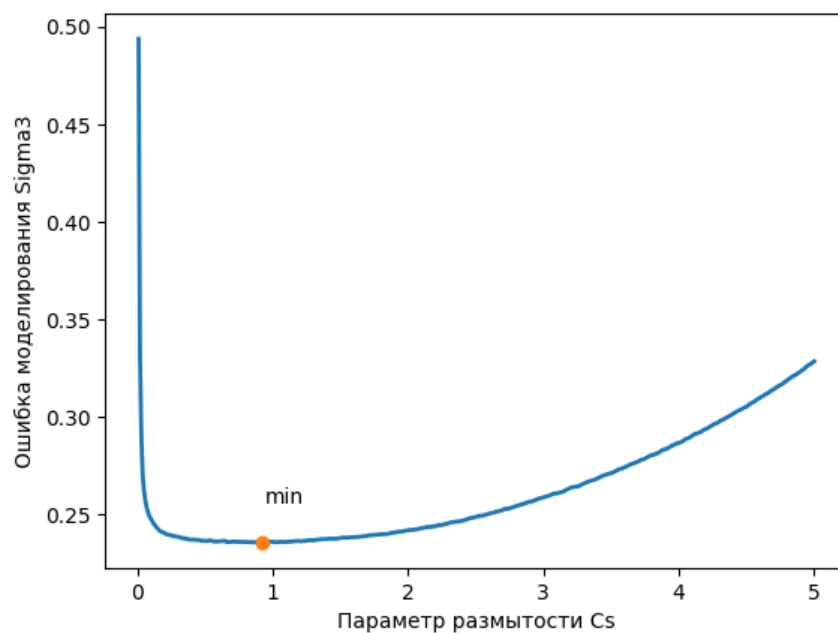


Рисунок 3.8 – Зависимость  $\delta_3$  от значения  $c_s$  для 3-го выхода при  $N = 500$  и  $\xi = 5\%$

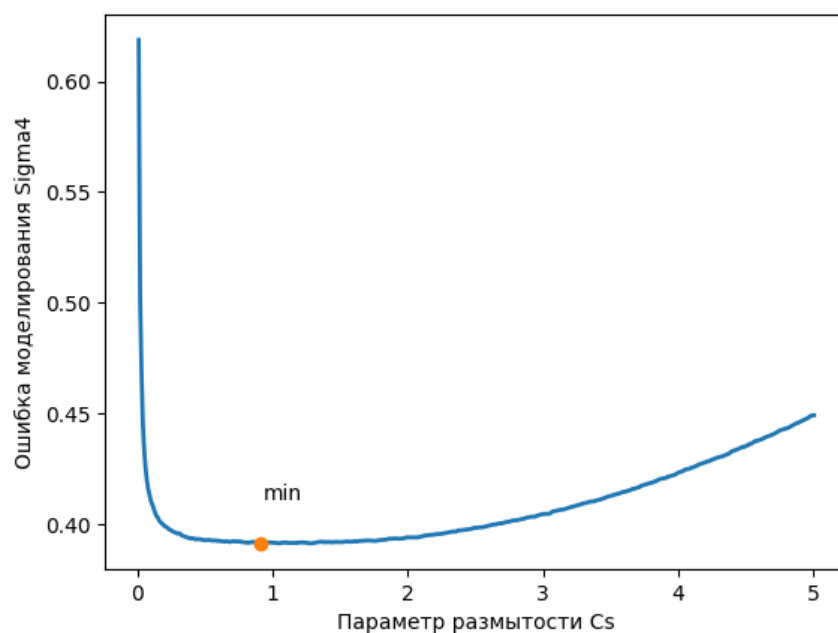


Рисунок 3.9 – Зависимость  $\delta_4$  от значения  $c_s$  для 4-го выхода при  $N = 500$  и  $\xi = 5\%$

Зависимость точности моделирования  $\delta$  от параметра размытости  $c_s$  тоже изменилась с увеличением выборки. На рисунках 3.6-3.9 можно заметить, как после достижения наименьшего значения параметром размытости  $c_s$  ошибка

моделирования  $\delta$  увеличивается. Как правило, после резкого падения следует плавный подъем. В качестве примера, на рисунке 3.10 представлен результат моделирования выходной переменной  $x_2$  объекта при  $N = 500$  и  $\xi = 5\%$ .

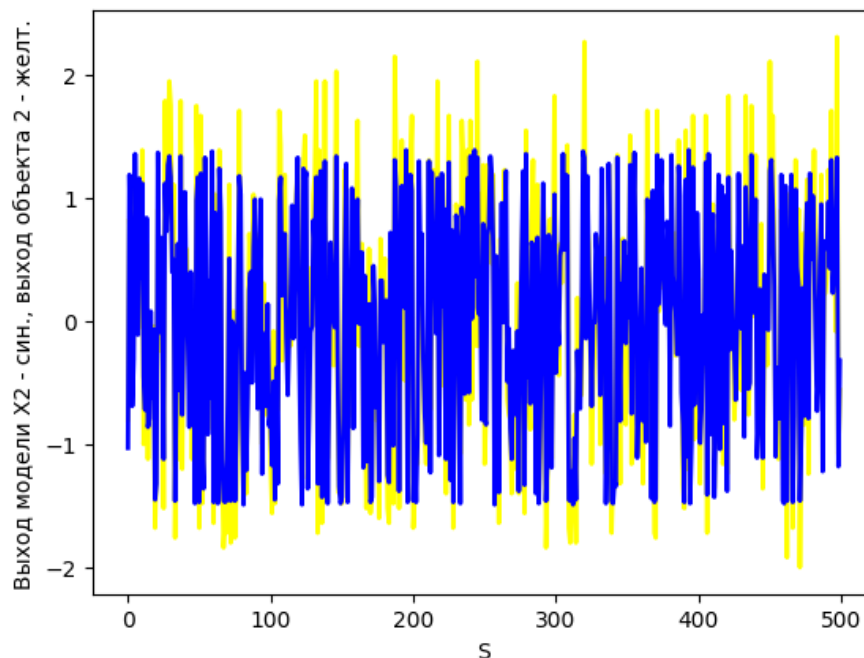


Рисунок 3.10 – Результат моделирования выходной переменной  $x_2$  объекта при  $N = 500$  и  $\xi = 5\%$

Таким образом, высокая точность была достигнута при большом объеме выборке и низком значении помехи.

### **3.2 Технологические сведения о процессе гидроочистки и гидродепарафинизации**

Процесс каталитической депарафинизации (гидродепарафинизации) предназначен для улучшения показателей текучести различного углеводородного сырья при низких температурах. Он применяется для снижения температуры застывания базовых компонентов смазочных масел и средних дистиллятов, температуры помутнения дизельных топлив и

температуры кристаллизации авиационных топлив. Эти свойства определяют поведение указанных продуктов при низких температурах [2].

Чтобы обеспечить эффективное управление процессом каталитической гидроочистки дизельного топлива, включая удаление сернистых соединений и гидродепарафинизацию, а также повышение хладотекучести, необходимо тщательно контролировать технологические процессы. Качество окончательного дизельного топлива зависит от качества исходного сырья и параметров, используемых в реакторе. Однако, эту задачу возможно решить с помощью математического моделирования.

Для повышения точности моделирования процесса каталитической гидродепарафинизации предлагается использовать непараметрические алгоритмы и методы идентификации. Входные и выходные переменные процесса гидродепарафинизации взаимосвязаны между собой, и для построения качественной модели необходимо учитывать все зависимости между ними. Однако исследователю эти зависимости не известны.

Для решения задачи идентификации предлагается использовать обучающую выборку, сформированную на основе измерений входных и выходных переменных процесса. За неимением априорной информации о параметрической структуре зависимостей процесса, задача моделирования соответствует идентификации или непараметрической идентификации. Для решения этой задачи предлагается использовать непараметрическое моделирование, основанное на непараметрических оценках Надарая-Ватсона регрессионных зависимостей.

В рассматриваемом процессе были выделены следующие основные входные переменные:

- плотность при 15 °С;
- фракционный состав (температуры: начала кипения, перегонки 50%, 96%, конца кипения);
- давление на входе;
- температура на входе.

Основными выходными переменными стали:

- плотность при 15 °С;
- фракционный состав (температуры: начала кипения, перегонки 50%, 96 %, конца кипения);
- температура помутнения.

На нефтеперерабатывающих заводах России работают установки гидроочистки дизельного топлива, совмещенные с процессом гидродепарафинизации (секция 300/1 комбинированной установки под названием ЛК-6Ус).

На рисунке 3.1 представлен реакторный блок Р-301, в котором совмещены процессы гидроочистки и гидродепарафинизации, а также блоки очистки циркуляционного водородсодержащего газа С-301а (ВСГ), стабилизации дизельного топлива, с извлечением бокового погона К-301 и очистки углеводородных газов; стабилизации бензина отгона.

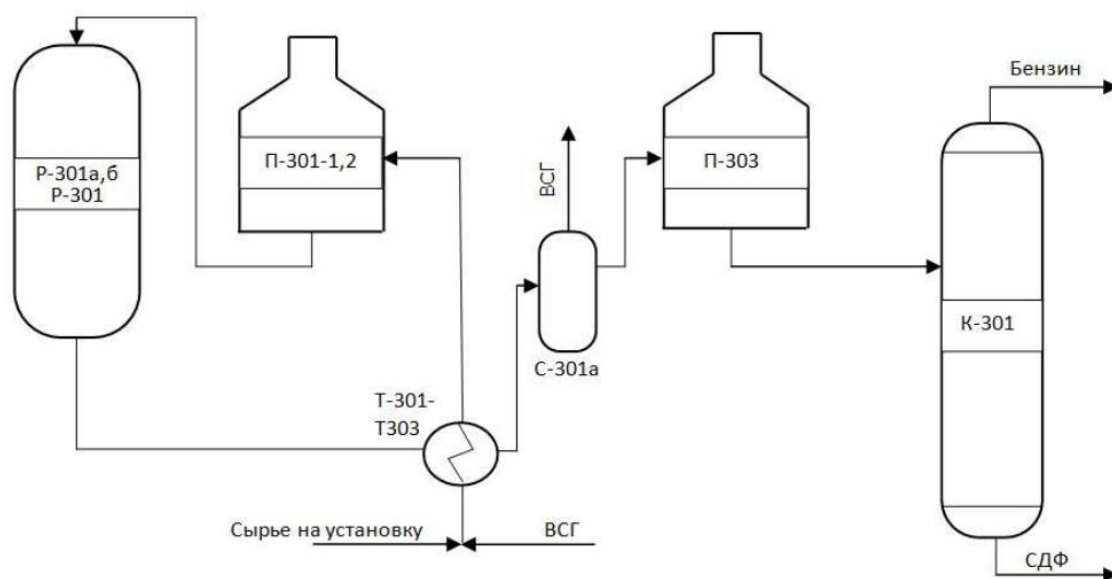


Рисунок 3.11 – Схема гидроочистки дизельного топлива, совмещенная с процессом гидродепарафинизации



Рассмотрим рисунок 3.11, где входными материалами являются «Сырье на установку» и «ВСГ» (водородсодержащий газ), а выходными продуктами – «Бензин» и «СДФ2» (сероочищенная дизельная фракция) [2]. Реактор Р-301, используемый в процессе гидродепарафинизации, имеет комбинированную загрузку. На верхнюю часть реактора загружают катализатор гидродепарафинизации, где далее происходит реакция расщепления и гидрирования углеводородов нормального строения с целью снижения температуры застывания дизельной фракции и образования легких углеводородов. Затем на нижнюю часть реактора загружают катализатор гидроочистки, выполняющий функцию гидрирования. Секция 300/1 гидроочистки предназначена для процесса гидродепарафинизации, повышения хладотекучести дизельного топлива и гидроочистки дизельного топлива от сернистых соединений.

В процессе технологической обработки получают отдельные фракции нефти, которые должны соответствовать требованиям к компонентам товарных продуктов после вторичных процессов комбинированной установки, представленной на рисунке 3.11. Секция 300/1 производит готовый компонент дизельного зимнего топлива (боковой погон колонны К-301 гидроочищенный) и дизельную гидроочищенную фракцию (куб колонны К-301). В установке установлены датчики для измерения технологических параметров процесса. Параметры, которые измеряют на входе и выходе технологической установки являются основными входными и выходными параметрами, описанными выше.

Таким образом, при эксплуатации установки собирается информация о параметрах технологического процесса, которую необходимо далее обработать для мониторинга и управления процессом гидроочистки и гидродепарафинизации в последствии. В условиях ограниченной априорной информации о процессе используются методы непараметрического моделирования. Эти методы также помогают определить текущее состояние входных и выходных потоков, выявить неточности данных, предсказать и смоделировать качество продукции на выходе.

### 3.3 Применение непараметрического алгоритма для моделирования процесса гидроочистки и гидродепарафинизации дизельного топлива

Процессы гидроочистки и гидродепарафинизации в установке являются дискретно-непрерывными, так как они происходят непрерывно во времени, но контроль технологических параметров осуществляется в дискретные моменты времени. Из-за отсутствия достаточной априорной информации невозможно описать различные взаимосвязи между технологическими переменными с помощью параметрических зависимостей. Вместо этого используются непараметрические модели, которые не требуют предварительной параметризации модели.

Для формирования обучающей выборки  $u_k^i, x_j^i, i = \overline{1, S}$  производилось измерение входных и выходных переменных процесса гидроочистки и гидродепарафинизации. Далее, переменные  $u$  подверглись методам обработки данных: центрированию и нормированию – для того, чтобы при дальнейшем исследовании возможно было использовать скалярный параметр размытости  $c_s$ . Итак, на входе измеряются следующие переменные:  $u_1(t)$  – плотность при 15 °С, кг/м<sup>3</sup>, фракционный состав, °С: ( $u_2(t)$  – температура начала кипения,  $u_3(t)$  – температура выкипания 50%,  $u_4(t)$  – температура выкипания 96% и  $u_5(t)$  – температура конца кипения). На выходе измеряются следующие переменные:  $x_1(t)$ – плотность при 15 °С, кг/м<sup>3</sup>, фракционный состав, °С: ( $x_2(t)$  – температура начала кипения,  $x_3(t)$  – температура выкипания 50%,  $x_4(t)$  – температура выкипания 96%,  $x_5(t)$  – температура конца кипения) и  $x_6(t)$  – температура помутнения.

Из исходных данных видно, что входные переменные процесса принимают значения из следующих интервалов:  $u_1(t) \in [-5,83; 2,29]$ ,  $u_2(t) \in [-2,11; 3,25]$ ,  $u_3(t) \in [-3,44; 2,54]$ ,  $u_4(t) \in [-2,84; 2,55]$ ,  $u_5(t) \in [-4,34; 2,25]$ . А выходные переменные:  $x_1(t) \in [838,3; 848]$ ,

$$x_2(t) \in [178; 214], \quad x_3(t) \in [264; 275], \quad x_4(t) \in [318; 345], \quad x_5(t) \in [342; 355], \\ x_6(t) \in [-31; -10].$$

Обозначим обучающую выборку, полученную на реальном объекте в виде  $(u_i, x_i, i = \overline{1, S})$  (для исследуемой выборки  $S = 115$ ). Выборка подвергается методам обработки данных: нормированию и центрированию (2.6).

Затем находятся прогнозные значения выходных переменных  $x(t)$  процесса:

$$\hat{x}_j = \frac{\sum_{i=1}^{115} x_j^i \cdot \prod_{k=1}^5 \Phi\left(\frac{u_k - u_k^i}{c_s}\right)}{\sum_{i=1}^{115} \prod_{k=1}^5 \Phi\left(\frac{u_k - u_k^i}{c_s}\right)}, \quad (3.6)$$

где  $j = \overline{1, n}$ ;

$\langle n \rangle$  – размерность составного вектора входных переменных  $x_j$ ;

$\langle m \rangle$  – размерность составного вектора выходных переменных  $u_k$ .

Колоколообразные функции принимаются в виде треугольного ядра (2.8).

С помощью данного алгоритма предлагается искать прогнозные значения для всех выходных переменных  $x(t)$  объекта по известным нам входным  $u(t)$ .

При построении модели проводим процедуру скользящего экзамена [6]. Настраиваемый параметр будет один – параметр размытости  $c_s$ , для которого реализован поиск оптимального значения, при котором ошибка моделирования  $\delta(3.5)$  минимальна на промежутке  $[0; 5]$ .

При экспериментальном прогнозировании выхода объекта зависимость точности моделирования от параметра размытости сохраняет вид, полученный при моделировании на тестовой выборке.

На рисунках 3.12-3.17 представлены результаты прогнозирования выходных переменных исследуемого процесса. В таблицу 3.3 будем вносить значения, принимаемые параметром размытости  $c_s$ , а также значения ошибки моделирования  $\delta$  для выходов объекта  $x_1-x_6$

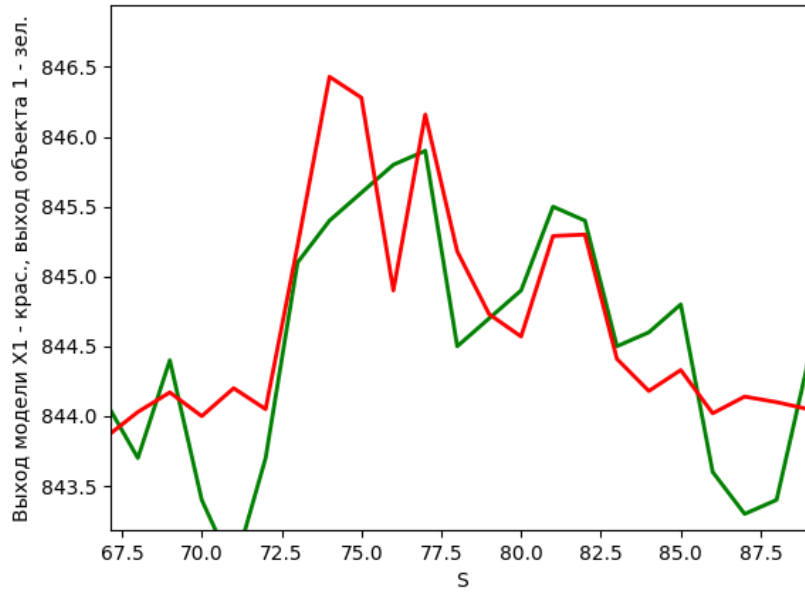


Рисунок 3.12 – Результат прогнозирования выходной переменной  $x_1$  объекта

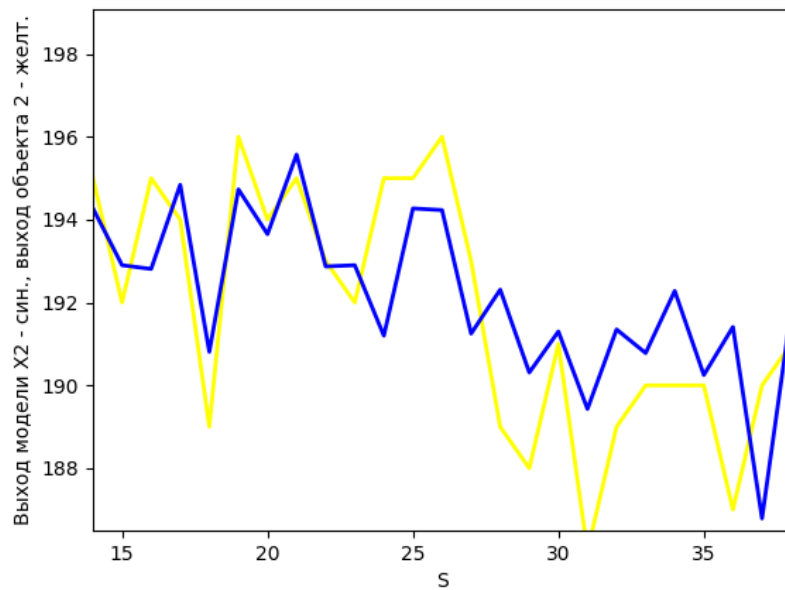


Рисунок 3.13 – Результат прогнозирования выходной переменной  $x_2$  объекта

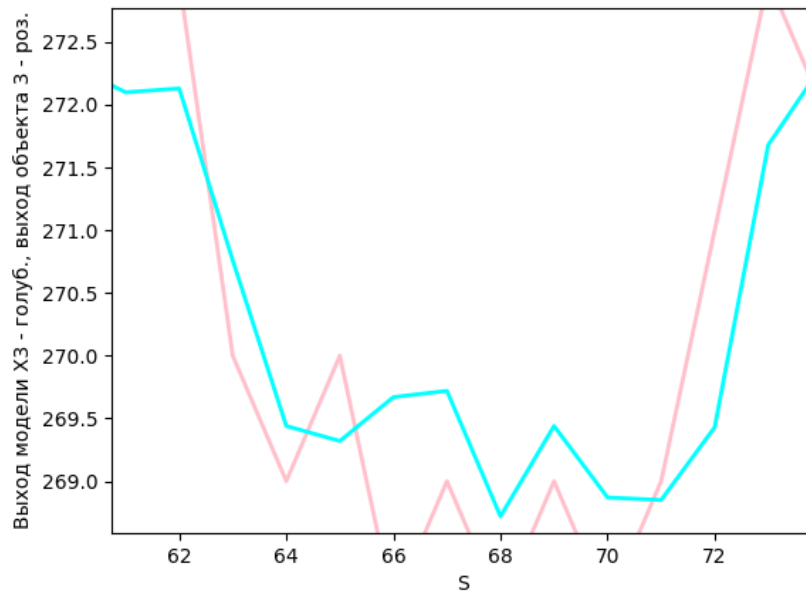


Рисунок 3.14 – Результат прогнозирования выходной переменной  $x_3$  объекта

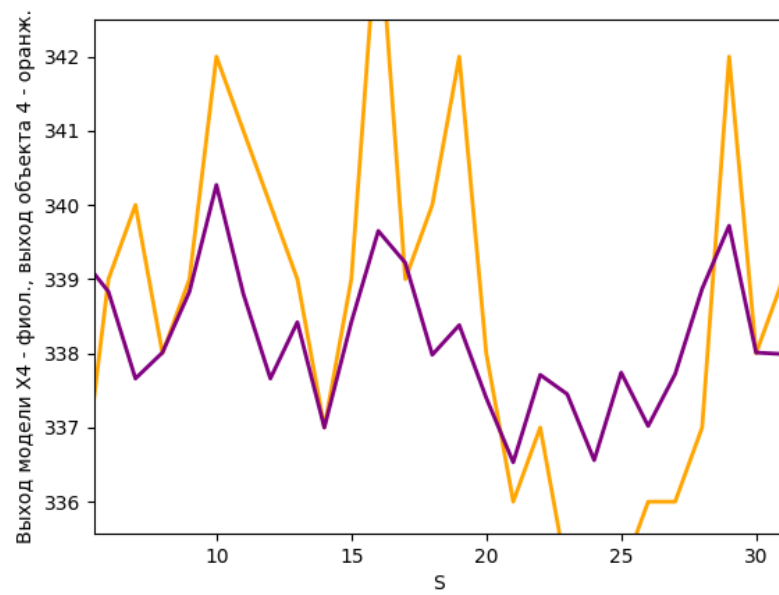


Рисунок 3.15 – Результат прогнозирования выходной переменной  $x_4$  объекта

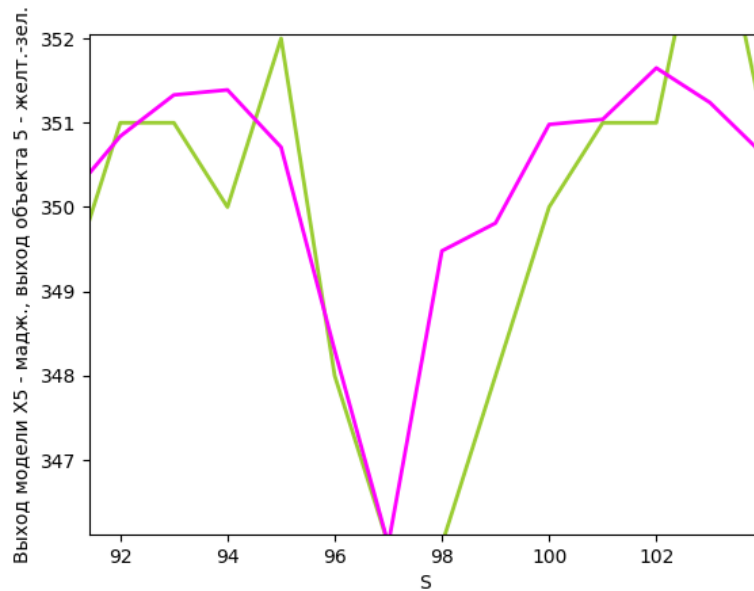


Рисунок 3.16 – Результат прогнозирования выходной переменной  $x_5$  объекта

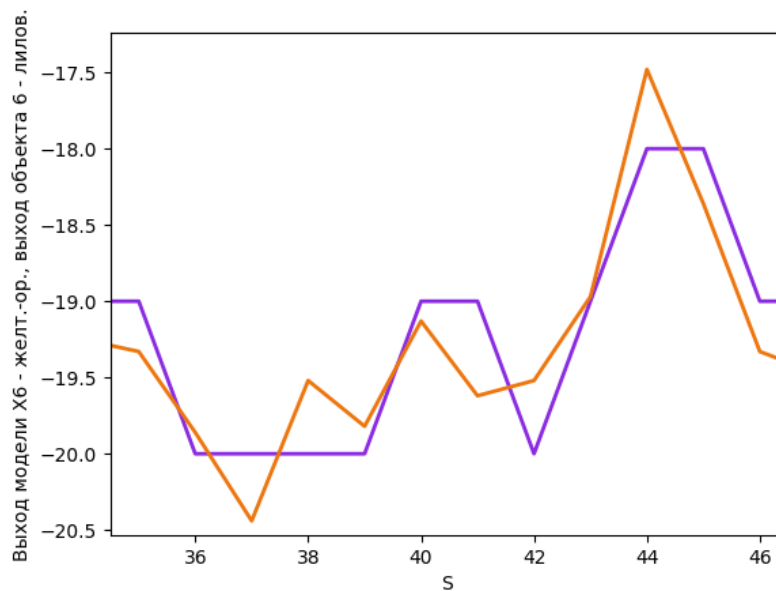


Рисунок 3.17 – Результат прогнозирования выходной переменной  $x_6$  объекта

Таблица 3.3 – Значения параметра размытости  $c_s$  и ошибки  $\delta$  при моделировании выходов объекта  $x_1-x_6$

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$
$c_s$	0.701	0.801	1.701	1.801	0.901	1.601
$\delta$	0.680	0.646	0.631	0.886	0.874	0.761

Рассмотрим так же зависимости точности моделирования  $\delta$  от изменения параметра размытости  $c_s$ , представленные на рисунках 3.18-3.23.

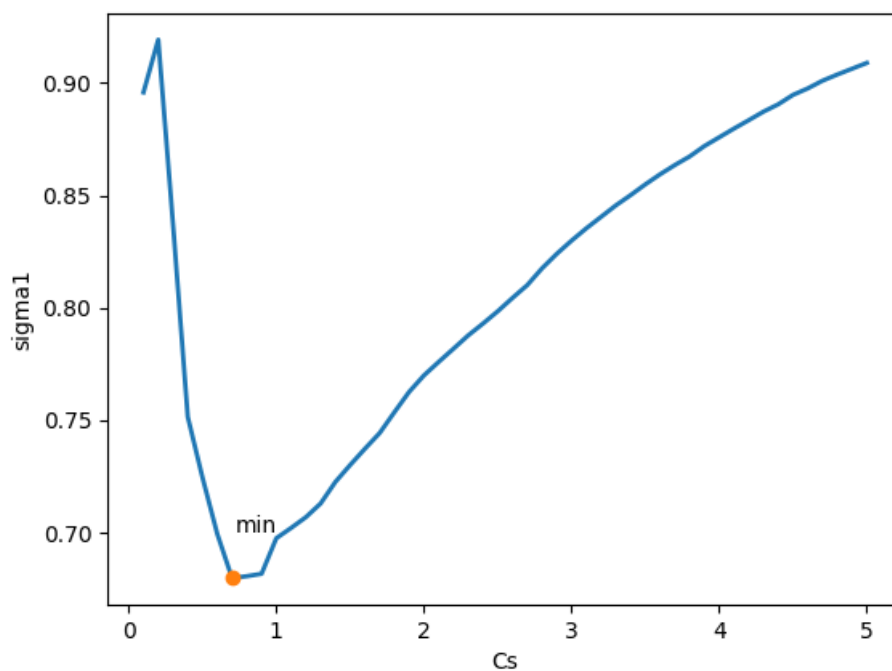


Рисунок 3.18 – Зависимость  $\delta_1$  от значения  $c_s$  для 1-го выхода

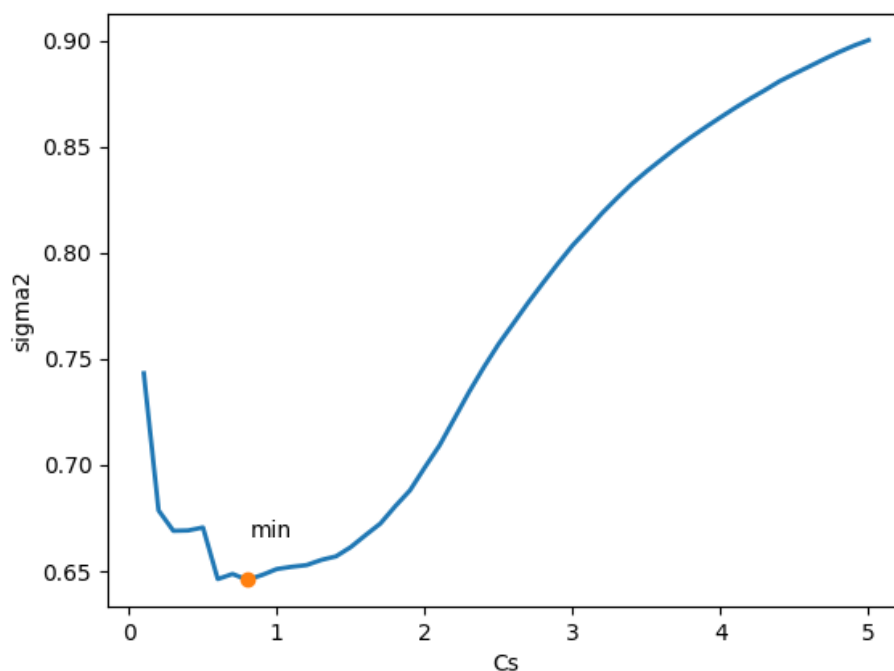


Рисунок 3.19 – Зависимость  $\delta_2$  от значения  $c_s$  для 2-го выхода

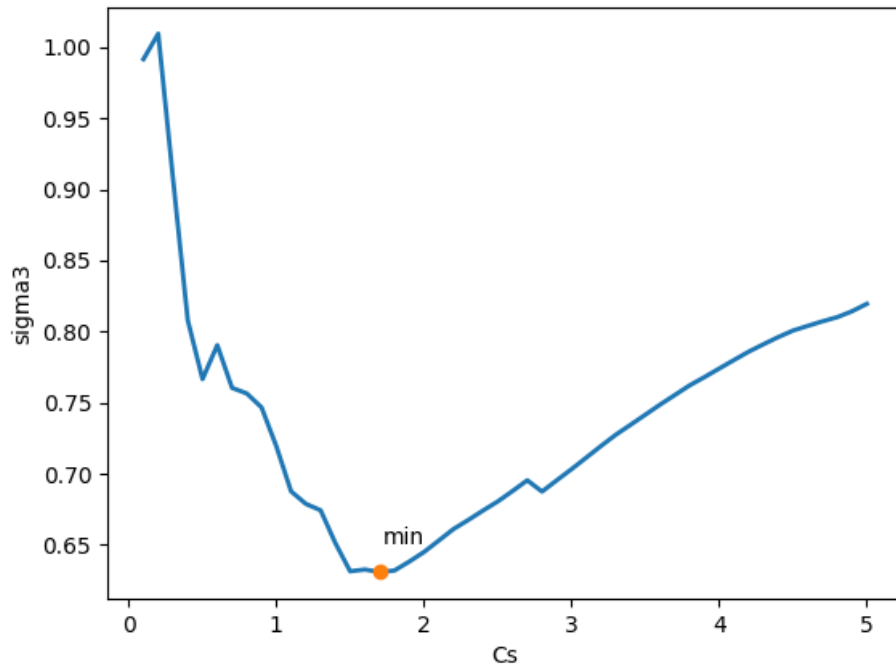


Рисунок 3.20 – Зависимость  $\delta_3$  от значения  $c_s$  для 3-го выхода

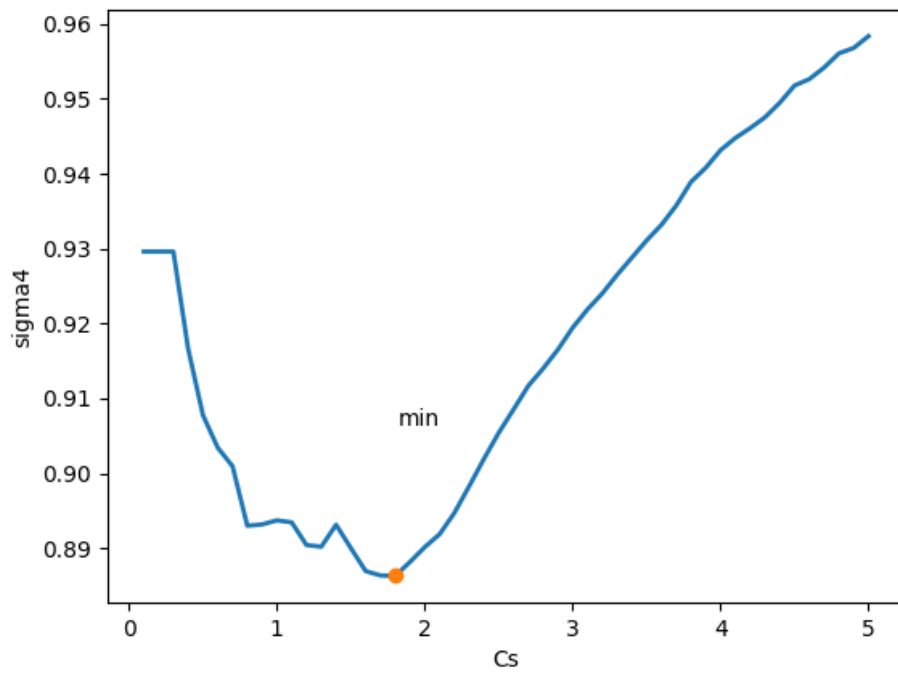


Рисунок 3.21 – Зависимость  $\delta_4$  от значения  $c_s$  для 4-го выхода



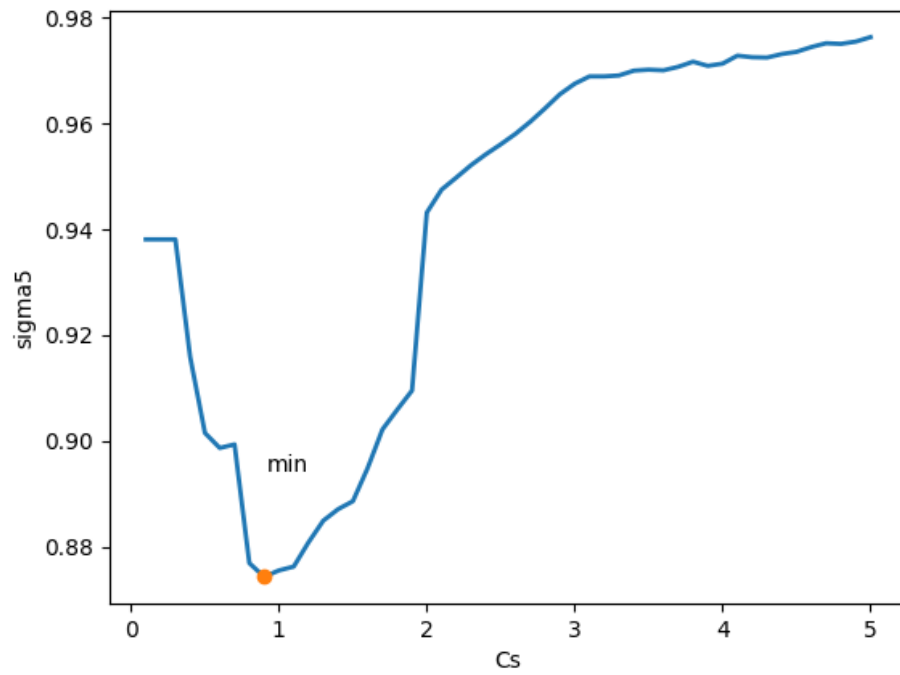


Рисунок 3.22 – Зависимость  $\delta_5$  от значения  $c_s$  для 5-го выхода

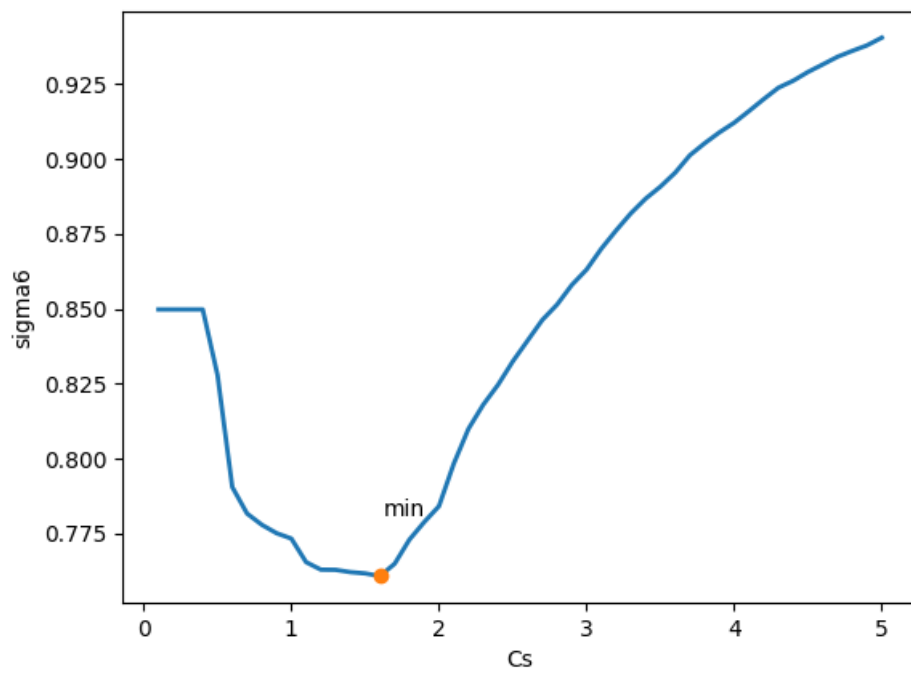


Рисунок 3.23 – Зависимость  $\delta_6$  от значения  $c_s$  для 6-го выхода

Исходя из представленных изображений, можно сделать вывод о том, что поведение зависимости из результатов экспериментов на модели сохранятся.

Далее оценим адекватность модели с помощью коэффициента детерминации  $R^2$  [13]. Расчет ведется по столбцам, т.е. по каждому выходу. Для этого используется следующая формула:

$$R^2 = \frac{Q_R}{Q}, \quad (3.7)$$

где  $Q_R = \sum_{i=1}^S (\hat{x}_i - \bar{x})^2$  – сумма квадратов, объясненная моделью;

$Q = \sum_{i=1}^S (x_i - \bar{x})^2$  – полная сумма квадратов.

Результаты будем вносить в таблицу 3.4.

Свойства коэффициента детерминации:

- при  $R^2 \rightarrow 1$  регрессия хорошо отражает эмпирические данные, наблюдения примыкают к линии регрессии;

- при  $R^2 = 1$  точки  $(x_i, \hat{x}_i)$  лежат на линии регрессии и между переменными существует линейная функциональная зависимость;

- при  $R^2 = 0$  вариация зависимой переменной полностью обусловлена воздействием неучтенных в модели переменных.

Таблица 3.4 – Коэффициент детерминации для выходов объекта  $x_1$ - $x_6$

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$
$R^2$	0.288	0.417	0.413	0.123	0.208	0.255

Результаты построения модели процесса гидроочистки и гидродепарафинизации дизельного топлива показывают, что каналы, на которых производились измерения, имеют неудовлетворительную точность. Это может быть связано с тем, что не были учтены некоторые важные переменные технологического процесса, например температура и давление в реакторе Р-301, концентрация и расход водородосодержащего газа, а также погрешность средств измерения. Таким образом, обратим внимание на тот факт, что неучет

дополнительных параметров процесса при его моделировании является одной из основных причин получения неудовлетворительной модели.

### **Выводы по главе 3**

Эксперименты на модели показали, что объем выборки и помеха влияют на точность моделирования. Так, с увеличением объема выборки повышается и точность моделирования. Однако, без учета влияния помехи получить минимальную погрешность моделирования не представляется возможным. Поэтому, при моделировании процессов дискретно-непрерывного типа требуется минимизировать влияние помехи в обучающей выборке.

Для успешного же моделирования процессов гидроочистки и гидродепарафинизации необходимо использовать методы непараметрического моделирования, так как в данном случае априорная информация недостаточна, а обучающая выборка является малой. Несмотря на это, результаты экспериментов показывают, что прогнозы моделей могут быть близки к реальным значениям в некоторых случаях, но в других – хуже.

Однако, при увеличении обучающей выборки и включении в нее дополнительных технологических показателей, можно получить более точные модели. Это открывает возможности для построения компьютерных систем управления и оптимизации процессов гидроочистки и гидродепарафинизации.

Таким образом, непараметрическое моделирование является перспективным направлением для эффективного использования в системах компьютерного управления процессами гидроочистки и гидродепарафинизации по различным каналам. Для достижения наилучших результатов необходимо разработать хорошо настроенные непараметрические модели.

## ЗАКЛЮЧЕНИЕ

В данной выпускной-квалификационной работе был рассмотрен подход к математическому моделированию дискретно-непрерывных процессов в условиях неполной информации. Целью работы является повышение точности моделирования процессов дискретно-непрерывного типа по наблюдениям с помехами.

В качестве инструментов в исследовании использовались методы математического моделирования, в частности, имитационные методы (с использованием электронных вычислительных средств), а также методы математической статистики.

Был осуществлен краткий обзор параметрических методов моделирования, таких как: метод наименьших квадратов, метод стохастической аппроксимации и линейная аппроксимация метода наименьших квадратов. Из совокупности рассмотренных непараметрических методов выделяется непараметрическая оценка функции регрессии Надарая-Ватсона, так как в исследовании этот метод выполняет ключевую роль.

В ходе вычислительных эксперименты, направленных на проверку эффективности применения методов и алгоритмов, направленных на решение задачи идентификации дискретно-непрерывных процессов в условиях неполной информации, было выявлено, что для повышения точности идентификации дискретно-непрерывных процессов следует увеличивать объем выборки, уменьшать влияние помехи на входы и выходы объекта, а также повышать точность исходной выборки.

Практическая значимость исследования заключается в разработанных методах и алгоритмах идентификации дискретно-непрерывных процессов в условиях неполной информации, которые могут быть использованы для решения практических задач в различных областях науки и техники.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Самарский, А. А. Математическое моделирование: Идеи. Методы. Примеры / А. А. Самарский, А. П. Михайлов. – 2-е изд., испр. – Москва : ФИЗМАТЛИТ, 2005. – С. 7. – ISBN 5-9221-0120-X.
2. Ярещенко, Д. И. Непараметрические алгоритмы моделирования и управления многомерными безынерционными системами с запаздыванием : специальность 05.13.01 «Системный анализ, управление и обработка информации (по отраслям)» : диссертация на соискание ученой степени кандидата технических наук. / Ярещенко Дарья Игоревна ; СибГАУ им. М.Ф. Решетнева. – Красноярск, 2020. – 138 с.
3. Агафонов, Е. Д. Математическое моделирование линейных динамических систем : учебное пособие / Е. Д. Агафонов, О. В. Шестернева ; Сибирский федеральный университет, Институт космических и информационных технологий. – Красноярск, 2011, – Гл. 2. – С. 6–28. – ISBN 978-5-7638-2162-8.
4. Медведев, А. В. Основы теории адаптивных систем : монография / А. В. Медведев ; СибГАУ им. М. Ф. Решетнева. – Красноярск, 2015 – С. 51–117. – ISBN 978-5-86433-657-1.
5. Померанцев, А. Л. Калибровка (градуировка) / А. Л. Померанцев // Wayback Machine: Internet Archive : [сайт]. – Москва : Российское хеометрическое общество, 2008. – URL: <https://web.archive.org> (дата обращения: 21.04.2023).
6. Надарая, Э. А. Непараметрические оценки плотности вероятности и кривой регрессии. / Э. А. Надарая. – Тбилиси : Изд-во Тбил. ун-та, 1983 – 194 с.
7. Бойко, Р. С. О моделировании дискретно-непрерывных процессов / Р. С. Бойко // Молодой ученый : [сайт]. – 2010 – № 4. – С. 93–98. URL: <https://moluch.ru/archive/15/1378> (дата обращения: 22.04.2023).

8. Ликсонова, Д. И. О непараметрических алгоритмах адаптации в многомерных системах с запаздыванием / Д.И. Ликсонова, А.В. Медведев // Сибирский журнал науки и технологий. – Красноярск, 2021. – 10 с.
9. Ярещенко, Д. И. О непараметрической идентификация частично-параметризованного дискретно-непрерывного процесса / Д.И. Ярещенко. // Сибирский журнал науки и технологий. – Красноярск, 2020. – Т. 21, №1. – С. 47–53.
10. Демченко, Я. И. Об исследовании некоторых непараметрических оценок функции регрессии по наблюдениям / Я. И. Демченко, А. И. Орлова // Вестник СибГАУ им. М. Ф. Решетнева. – Красноярск, 2010. – №5 (31). – С. 166–170.
11. Медведев, А. В. Моделирование многомерных N-процессов / А. В. Медведев, Е. А. Чжан // Информационные и математические технологии в науке и управлении. – Красноярск, 2018. – №1 (9). – С. 99–105.
12. Корнеева, А. А. О непараметрическом анализе данных в задаче идентификации / А. А. Корнеева, Н. А. Сергеева, Е. А. Чжан // Вестник Том. Гос. ун-та. Управление, вычислительная техника и информатика. – 2013. – №1 (22). – С. 86–95.
13. Колачева, Н. В. Эконометрика : курс лекций / Н. В. Колачева, С. Ш. Палферова. – Тольятти : Изд-во ТГУ, 2010. – С. 42–43.
14. Гмурман, В. Е. Теория вероятностей и математическая статистика: учебное пособие / В. Е. Гмурман. – Москва, 2003. – 479 с.
15. 1 Тарасенко, Ф. П. Прикладной системный анализ (Наука и искусство решения проблем) : учебник / Ф. П. Тарасенко. – Томск : Изд-во Том. ун-та, 2004. – 224 с.

Министерство науки и высшего образования РФ  
Федеральное государственное автономное  
образовательное учреждение  
высшего образования  
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»  
Институт космических и информационных технологий  
Базовая кафедра «Интеллектуальные системы управления»

УТВЕРЖДАЮ

И.о. зав. кафедрой


 Д.И. Ликсонова

« 21 » июня 2023 г.

**БАКАЛАВРСКАЯ РАБОТА**

27.03.03 «Системный анализ и управление»

Идентификация дискретно-непрерывных процессов в условиях неполной информации

Руководитель  21.06.2023 и.о. зав. каф., канд. тех. наук Д.И. Ликсонова  
подпись, дата

Студент  21.06.23 Н.А. Усова  
подпись, дата

Нормоконтролер  21.06.2023 Н.Б. Позолотина  
подпись, дата

Красноярск 2023