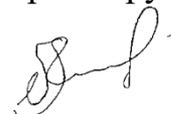


Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Новосибирский государственный университет экономики и управления «НИНХ»

На правах рукописи



НИКУЛИН ВЛАДИМИР СЕРГЕЕВИЧ

**Методика оценки надежности вычислительных систем по малым выборкам
эксплуатационных данных**

Специальность 2.3.8 – Информатика и информационные процессы

Диссертация на соискание учёной степени
кандидата технических наук

Научный руководитель:
канд. физ.-мат. наук, доцент
Пестунов Андрей Игоревич

Новосибирск – 2023

Оглавление

Список принятых сокращений.....	4
Словарь терминов	5
ВВЕДЕНИЕ.....	7
ГЛАВА 1. Проблематика эффективности оценки надежности вычислительных систем	15
1.1 Анализ существующих решений для оценки надежности и их применимость к вычислительным системам.....	15
1.2 Организационно-технические мероприятия контроля и поддержания функциональной надежности вычислительной системы.....	19
1.3 Определение структурных элементов вычислительной системы и контролируемых параметров	32
1.4 Анализ методов статистической оценки надежности сложных технических систем в целом	33
1.5 Выводы по главе.....	38
ГЛАВА 2. Разработка методики оценки надежности вычислительных систем по малым выборкам эксплуатационных данных	40
2.1 Технология сбора и подготовки эксплуатационных данных вычислительных систем	40
2.2 Модели машинного обучения для распознавания отказов в общем потоке информации системы мониторинга	50
2.3 Адаптация непараметрического метода Розенблатта – Парзена для оценки показателей надежности.....	57
2.4 Оценка точности адаптированного метода Розенблатта – Парзена	73
2.5 Выводы по главе.....	76

ГЛАВА 3. Разработка программного обеспечения для оценки надежности вычислительных систем. Экспериментальная проверка методики	77
3.1 Постановка задачи	78
3.2 Разработка программного обеспечения «Модуль СМ» для оценки надежности.....	79
3.3 Интерфейс управления программным обеспечением «Модуль СМ»	81
3.4 Экспериментальная проверка методики оценки эксплуатационной надежности вычислительных систем	87
3.5 Выводы по главе.....	94
ЗАКЛЮЧЕНИЕ	96
Список литературы	97
ПРИЛОЖЕНИЕ А. Свидетельство о государственной регистрации программы для ЭВМ.....	108
ПРИЛОЖЕНИЕ Б. Исследование свойств адаптированного метода Розенблатта – Парзена	109

Список принятых сокращений

АРМ	– автоматизированное рабочее место
АЭ	– аппаратный элемент
БД	– база данных
ВБР	– вероятность безотказной работы
ВС	– вычислительная система
ЗИП	– запасные части, инструменты и принадлежности
ОС	– операционная система
ПО	– программное обеспечение
ПЭ	– программный элемент
СТС	– сложная техническая система
СЭ	– составной элемент
ТОиР	– техническое обслуживание и ремонт
СМ	– система мониторинга
ССОИ	– система сбора и обработки информации
СУДЗ	– система управления диспетчеризации задач
ЦСОД	– центр сбора и обработки данных

Словарь терминов

Группа эксплуатации: технический персонал объекта, на который возложена обязанность по поддержанию ВС в исправном и работоспособном состоянии путем проведения предусмотренных работ по техническому обслуживанию, текущему ремонту (профилактике) и организации других видов ремонта (подача рекламаций и заявок на ремонт оборудования, отправка и прием оборудования), ведению эксплуатационной документации и сбору сведений о надежности.

Запас: совокупность запасных частей одного типа, характеризуемая начальным уровнем, суммарными затратами, стратегией пополнения и показателем достаточности.

Интервал цензурирования: интервал наработки, внутри которого произошел либо произойдет отказ изделия, причем момент наступления отказа неизвестен.

Неполная наработка: наработка от начала испытаний или эксплуатационных наблюдений до прекращения испытаний или эксплуатационных наблюдений, причем в пределах данной наработки отказа не произошло.

Отказ: (по ГОСТ 27.002-89): событие, заключающееся в нарушении работоспособного состояния объекта.

Полная наработка: наработка изделия до отказа.

Профилактика: комплекс операций или операция по поддержанию работоспособности изделия за счет предупредительной замены и (или) регулировки составных частей (узлов, механизмов, приборов, блоков и т.п.), параметры которых вышли за пределы профилактического допуска.

Ремонт: (по ГОСТ 18322-78): комплекс операций по восстановлению исправности или работоспособности изделия и восстановлению ресурсов изделий или их составных частей.

Текущий ремонт: (по ГОСТ 18322-78): ремонт, выполняемый для обеспечения или восстановления работоспособности изделия и состоящий в замене и (или) восстановлении отдельных частей.

Техническое обслуживание: (по ГОСТ 18322-78): комплекс операций или операция по поддержанию работоспособности или исправности изделия при использовании по назначению, ожидании, хранении и транспортировании.

Цензурирование данных: явление, при котором для части испытываемых или контролируемых изделий отсутствуют сведения о моментах наступления отказа за период наблюдения.

Цензурированные интервалом данные: цензурированные данные, для которых интервал неопределенности ограничен и справа и слева.

Цензурированные справа данные: цензурированные данные, для которых интервал неопределенности ограничен слева, а справа открыт.

ВВЕДЕНИЕ

Актуальность и степень разработанности проблемы. Вычислительные системы (далее – ВС) играют важную роль при решении целого ряда задач, включая математическое моделирование, обработку текстового и аудиовизуального контента, обнаружение закономерностей в данных, распознавание образов и речевой информации. Оценка надежности функционирования ВС должна иметь высокую точность, поскольку даже их единичный отказ может иметь критические последствия. Ключевыми показателями надежности ВС являются вероятность безотказной работы в течение заданного интервала времени и коэффициент готовности. Оценить их можно теоретически на этапе проектирования методами биномиальных или ускоренных лабораторных испытаний. Теоретические расчеты надежности ВС зачастую недостоверны, поскольку на этапе проектирования информация о надежности составных элементов, планируемых режимах работы и обслуживания ВС отсутствует.

Для подтверждения достижения вероятности безотказной работы порядка 0,97 методом биномиальных испытаний потребуется поставить в опытную эксплуатацию не менее 128 элементов (при этом допускается не более одного отказа) [24]. Наличие большего числа отказов для подтверждения тех же значений потребует значительного увеличения времени наблюдений либо числа оцениваемых элементов. Подобные масштабы испытаний большинство разработчиков себе позволить не могут. Таким образом, для относительно небольших ВС (менее 100 вычислительных узлов) метод биномиальных испытаний не подходит. Исключением являются крупные системы с большим количеством составных однотипных элементов, например, ВС МГУ «Ломоносов-2», состоящий из 1686 вычислительных узлов. Качественно оценить надежность относительно небольших ВС оказывается возможным только после длительного периода их реальной эксплуатации на основе информации о фактических отказах (далее – эксплуатационные данные).

Другой альтернативой являются методы ускоренных лабораторных испытаний, которые реализуются путем повышения нагрузки на элементы. Как отмечают А.М. Половко и С.В. Гуров, авторы этих методов делают много допущений, например, предположение о том, что законы распределения отказов элементов при увеличении нагрузки остаются неизменны. Подобные допущения сложно доказать и проверить, поэтому модели, используемые при ускоренных лабораторных испытаниях, могут быть неадекватны реальному объекту. Кроме того, лабораторные условия могут существенно отличаться от условий реальной эксплуатации.

К настоящему времени разработан ряд общих методик для более широкого класса устройств, называемых сложными техническими системами. А.М. Половко и С.В. Гуров предлагают использовать инженерную методику анализа надежности техники по эксплуатационным данным, основанную на аналитическом решении интегрального уравнения. Однако, как отмечают сами авторы, ее реализация не всегда возможна на практике, так как предъявляются особые требования к параметру потока отказов, сведения о котором могут отсутствовать.

О.О. Шмидт разработала обобщенную модель процесса восстановления в теории надежности информационных технологий. Модель позволяет получать такие характеристики надежности, как функция восстановления и функция затрат на модельных данных с применением параметрических методов оценивания. Если же априорные данные о распределении отсутствуют, то данная модель не применима.

А.В. Антоновым и М.С. Никулиным, предложена методика, основанная на применении непараметрических методов для оценки показателей надежности насосных агрегатов АЭС. В своих работах авторы рассмотрели варианты использования непараметрических методов, таких как метод Розенблатта – Парзена, метод проекционных оценок и метод корневого оценивания. Метод Розенблатта – Парзена был развит для аппроксимации цензурированной информации.

Применение этих и некоторых другие общих для сложных технических систем методов при оценке надежности именно ВС может быть затруднено по ряду причин. В частности, ВС характеризуются следующими особенностями:

- состоят из аппаратных и управляющих программных элементов, отказы которых могут иметь как общую, так и индивидуальную причину;
- регулярно подвергаются масштабированию и модернизации, что приводит к нестационарному (динамическому) количеству оцениваемых элементов и формирует потребность получения оценок в режиме реального времени;
- генерируют статистическую информацию низкого качества из-за отсутствия стандартизированного способа выявления отказов;
- в качестве основной группы элементов выступают вычислительные серверы, на которых за наблюдаемый период времени фиксируется малое число отказов (менее 30); на ряде элементов отказы фиксируются интервалами (цензурирование интервалом), а часть оборудования может не иметь отказов (цензурирование справа).

В результате анализа общих методик в качестве базового подхода предлагается рассмотреть непараметрические методы статистической оценки. Преимущество этих методов заключается в том, что они развиты для учета цензурированной информации, учитывают исторические данные функционирования и пригодны для малых выборок.

Несмотря на развитие данных методов, в проанализированных работах отсутствует представленное в явном виде аналитическое выражение для функции распределения времени до отказа, объединяющее как полные наработки, так и все виды цензурированных данных. Определение данного выражения необходимо для случаев, когда по элементам системы могут быть зафиксированы как полные наработки, так и цензурированные интервалом. Также представленная авторами оценка с учетом цензурированных справа данных демонстрирует хорошие результаты лишь при малой их доле, в ином случае такая оценка приводит к нарушению вида функции распределения. Способ расчета параметра локальности

метода Розенблатта – Парзена приведен в общем виде, и требуется разработка алгоритма его расчета для ВС. Данные обстоятельства определяют потребность проведения дополнительных исследований и адаптации метода Розенблатта – Парзена в целях его практического использования для оценки реальных показателей надежности элементов ВС по эксплуатационным данным.

Точность оценки надежности существенно зависит от качества формирования эксплуатационных данных. Несмотря на значимое развитие современных систем мониторинга ВС, сведения об отказах элементов по-прежнему фиксируются обслуживающим персоналом в журналах, рекламационных актах и формулярах. При обобщении таких сведений значимая часть важных событий может быть утрачена. Данная проблема носит общий характер для всех технических систем и была исследована В.С. Викторовой, А.С. Степанянц, М.Б. Успенским, А.Г. Тарасовым, R. Wolski, I. Foster и др.

В результате исследований авторы предлагают исключить «человеческий» фактор и автоматизировать процесс сбора эксплуатационных данных. В частности, А.Г. Тарасов и М.Б. Успенский отмечают, что современный уровень развития систем мониторинга ВС позволяет фиксировать данные эксплуатации в режиме реального времени для большинства аппаратных элементов. При этом актуальной остается задача разработки единой архитектуры системы сбора информации с аппаратных элементов и управляющего программного обеспечения.

При решении задачи формирования массива эксплуатационных данных, возникает проблема обнаружения релевантных событий в общем потоке информации от системы мониторинга. В большинстве случаев информация от системы мониторинга избыточна, может содержать справочный характер и не представлять пользы для оценки надежности. Для этого необходимо разработать модели машинного обучения для распознавания отказов среди такой информации.

Целью диссертационной работы является повышение эффективности оценки надежности вычислительных систем непараметрическими методами по малым выборкам эксплуатационных данных.

Для достижения цели определены следующие **задачи**:

1. Разработать технологию сбора и подготовки эксплуатационных данных для формирования набора отказов по каждому элементу вычислительной системы.
2. Разработать модели машинного обучения для распознавания отказов составных элементов вычислительных систем в процессе их функционирования.
3. Получить аналитическое представление функции плотности распределения отказов, учитывающее полные наработки на отказ, а также наработки, цензурированные интервалом, и наработки, цензурированные справа.
4. Разработать программный комплекс для оценки эксплуатационной надежности вычислительных систем, включающий реализацию технологии сбора и подготовки данных.

Объектом исследования являются вычислительные системы с малым количеством элементов как программно-аппаратные комплексы инфокоммуникационных систем.

Предметом исследования является надежность вычислительных систем, рассчитываемая по эксплуатационным данным.

Методы исследований. Методы теории надежности и математической статистики; технологии структурного и объектно-ориентированного программирования; методы машинного обучения.

Научная новизна диссертационной работы заключается в следующем.

1. Предложена новая методика оценки надежности вычислительных систем по малым выборкам эксплуатационных данных на основе непараметрического метода Розенблатта – Парзена, позволяющая повысить эффективность базового метода за счет учета цензурированных данных и

компенсации смещения оценок плотности распределения отказов составных элементов.

2. Разработана технология сбора и подготовки эксплуатационных данных по каждому элементу вычислительной системы, включающая методы опроса состояния управляющих элементов, архитектуру их взаимодействия и метод выявления релевантных событий в потоке информации.

3. Разработаны модели машинного обучения, впервые позволяющие распознавать отказы программных и аппаратных элементов одновременно, что невозможно осуществить традиционными способами из-за их ориентированности исключительно на аппаратную часть.

Теоретическая значимость. Получено аналитическое представление функции плотности распределения отказов, учитывающее полные наработки на отказ, а также наработки, цензурированные интервалом, и наработки, цензурированные справа.

Практическая значимость работы. Разработан программный комплекс для оценки эксплуатационной надежности вычислительных систем, включающий реализацию технологии сбора и подготовки эксплуатационных данных. Автоматизирован процесс подготовки данных из эксплуатации, на основе которых проводится оценка показателей надежности. Результаты диссертационной работы используются в образовательном процессе кафедры информационных технологий ФГБОУ ВО «Новосибирский государственный университет экономики и управления «НИНХ». Получено свидетельство о государственной регистрации программы для ЭВМ № 2020661451 от 24.09.2020.

Положения, выносимые на защиту.

1. Методика оценки надежности вычислительных систем по малым выборкам эксплуатационных данных с учетом наличия цензурированных данных.

2. Технология сбора и подготовки эксплуатационных данных для формирования набора отказов по каждому элементу вычислительной системы.

3. Модели машинного обучения для распознавания отказов составных элементов вычислительных систем в процессе их функционирования.

4. Программный комплекс для оценки эксплуатационной надежности вычислительных систем, включающий реализацию технологии сбора и подготовки эксплуатационных данных.

Личный вклад автора. Все результаты, выносимые на защиту, получены автором лично.

Соответствие паспорту специальности. Диссертация соответствует паспорту специальности 2.3.8 «Информатика и информационные процессы»:

Пункт 9. Разработка архитектур программно-аппаратных систем поддержки цифровых технологий сбора, хранения и передачи информации в инфокоммуникационных системах, в том числе, с использованием «облачных» интернет-технологий, и оценка их эффективности.

Пункт 17. Разработка методов обеспечения надежной обработки информации и обеспечения помехоустойчивости информационных коммуникаций для целей передачи, хранения и защиты информации; разработка основ теории надежности и безопасности использования информационных технологий.

Апробация результатов работы. Основные результаты исследования представлены на X Всероссийской научно-практической конференции с международным участием «Наука молодых» (Арзамас, 2017); Национальном суперкомпьютерном форуме «НСКФ» (Переславль-Залесский, 2017); XXXVII Всероссийской научно-технической конференции «Проблемы эффективности и безопасности функционирования сложных технических и информационных систем» (Серпухов, 2018); XIV Всероссийской научной конференции молодых ученых «Наука. Технологии. Инновации» (Новосибирск, 2019); VII международной научно-технической конференции студентов, аспирантов и молодых ученых «Математическое моделирование и информационные технологии» (Иваново, 2020); XXI Всероссийской конференции молодых ученых по математическому моделированию и информационным технологиям

(Новосибирск, 2020); XV-ой Международной научно-практической конференции «Современные инструментальные системы, информационные технологии и инновации» (Курск, 2020); XX Международной конференции «Сибирская научная школа-семинар “Компьютерная безопасность и криптография” — SIBECRYPT'21» имени Г. П. Агибалова (Новосибирск, 2021); XVI Всероссийской научной конференции молодых ученых «Наука. Технологии. Инновации» (Новосибирск, 2022); XXIII Всероссийской конференции молодых ученых по математическому моделированию и информационным технологиям (Новосибирск, 2022).

Публикации. Результаты исследований отражены автором в 12 научных работах [45; 54-62; 78; 102], среди них 3 публикации в журналах, рекомендованных высшей аттестационной комиссией для опубликования научных результатов диссертации на соискание ученой степени доктора или кандидата наук [45; 55; 58], 1 публикация в журнале индексируемом Scopus/WoS [78], 1 свидетельство о государственной регистрации программы для ЭВМ [102].

Объем и структура работы. Работа содержит 3 основных главы, введение и заключение. Основное содержание работы изложено на 117 страницах, включая 18 рисунков, 20 таблиц. Список литературы включает 102 наименования.

ГЛАВА 1. Проблематика эффективности оценки надежности вычислительных систем

1.1 Анализ существующих решений для оценки надежности и их применимость к вычислительным системам

Задаче диссертационного исследования отвечают методы, подразумевающие целесообразность их применения заказчиком и организациями заказчика, а также учитывающие особенности функционирования ВС. При выявлении таких методов целесообразно остановиться на следующих аспектах обеспечения эксплуатационной надежности – нормативное регулирование, тенденции, международный опыт стандартизации, методологический и прикладной аспект.

Надежность, как свойство техники, является предметом нормативного регулирования. Анализ стандартов Технического комитета 56 «Надежность» Международной электротехнической комиссии (далее – МЭК/ТК56) и отечественных стандартов серии «Надежность в технике» [1; 3] выявил следующие аспекты:

1. Стандартизация надежности направлена в первую очередь на обеспечение деятельности производителя. Основная часть стандартов напрямую связана с организационным, методическим, технологическим и информационным обеспечением надежности на стадиях разработки, производства и испытаний. Это руководства по сбору данных о надежности, анализу надежности, проектированию и отработке надежных изделий, закрепленный на нормативном уровне инструментарий контроля, проверки (подтверждения) показателей надежности с помощью методов математической статистики, расчета и оценки надежности с помощью специально разработанных математических методов.

2. Существует тенденция расширения области регламентации в сфере надежности на все виды техники. Все больше разрабатывается стандартов по

надежности, носящих комплексный характер и взаимосвязанных со стандартизацией по безопасности, живучести, технической диагностике, применению статистических и других методов и одновременно рассматриваемых как составная часть общей проблематики стандартизации управления качеством [5]. Управление надежностью осмыслено в рамках управления качеством, что предполагает деятельность производителя, направленную на повышение в выпускаемой продукции. Потребитель в системе качества, включающей надежность, выступает в роли индикатора и создает обратную связь с разработчиком и предприятием-изготовителем.

3. Для СТС, со значительным потенциалом риска, деятельность по управлению надежностью направлена на обеспечение безопасности. Эта деятельность осуществляется потребителем в процессе эксплуатации, то есть для потенциально опасных СТС основным субъектом управления выступает не проектирующая организация и производитель, а эксплуатирующая организация. ВС не относятся к изделиям, представляющим опасность для людей и окружающей среды в случае отказов, поэтому применение предложенных в стандартах методов оценки рисков не является обязательным [2].

4. Одним из перспективных направлений развития стандартизации является применение информационных технологий, для создания стандартов-программ. В них, вместо регламентации на бумажном носителе ограниченного числа решений некоторого метода, реализован сам метод. Разработанные и планируемые к разработке методы данного направления не соответствуют задачам обеспечения эксплуатации вычислительных систем.

5. В МЭК/ТК56 руководящие документы по управлению надежностью и посвященные техническому обслуживанию и ремонту (далее – ТОиР) находятся на стадии разработки и планирования. Отечественные стандарты также не позволяют в полной мере организовать эффективное планирование эксплуатации ВС.

Основным выводом из анализа нормативных документов является то, что существующие стандарты в полной мере не регламентируют процесс обеспечения контроля и поддержания надежности ВС, их применение требует гибкого и выборочного подхода. При этом отдельные методы (в первую очередь это касается специальных методов расчета, оценки и подтверждения надежности) могут быть использованы.

Особенности применения наиболее подходящих методов описаны в работе [3]. При этом ни один их известных методов не является универсальным. Основными ограничениями методов, которые предполагают количественную, а не качественную оценку показателей надежности, являются размерность оцениваемой системы и наличие марковского свойства [29].

В настоящее время на рынке программного обеспечения (далее – ПО) представлена широкая номенклатура, позволяющая проводить автоматизированные расчеты показателей надежности СТС, а также осуществлять моделирование сбоев и отказов. Существующее ПО в основном реализуют разработанные и апробированные модели и методики расчета надежности систем, представленных в виде связанных структур. Ряд ПО реализует методы марковского анализа, которые используют допущения о наличии экспоненциального распределения времени безотказной работы и восстановлений. Часть ПО описано ниже.

ПО АРБИТР (Россия), Relex (США), Risk Spectrum (Швеция) имеют схожую структуру и позволяют производить логико-вероятностный анализ надежности и безопасности СТС. В качестве средства построения графических моделей надежности используют деревья событий и деревья отказов. ПО Relex и Risk Spectrum используются в основном для оптимизации техногенных рисков и определения оптимальных параметров системы ТОиР потенциально опасных объектов. В частности, ПО Risk Spectrum широко применяется в вероятностном анализе безопасности объектов атомной энергетики на стадиях проектирования и

эксплуатации и позволяет моделировать развитие аварийного процесса. В состав ПО Relex входят расчетно-аналитические модули для решения задач прогнозирования безотказности, анализа видов, последствий и критичности отказов, расчета ремонтпригодности и др. Расчеты производятся в соответствии со стандартами Европы и США. Кроме того, ПО позволяет производить оценку влияния человеческого фактора на надежность и рассчитывать экономические показатели, характеризующие этапы создания, эксплуатации и утилизации системы [68; 75].

ПО АРБИТР – отечественная разработка, которая позволяет рассчитывать вероятность безотказной работы (далее – ВБР), среднюю наработку до отказа, коэффициент готовности, значимость и др. показатели, а также вклад элементов в надежность системы в целом [67].

ПО FavoWeb – представляет собой удаленно доступный интерфейс для расчета и прогнозирования показателей надежности с использованием информации из постоянно обновляемой распределенной базы данных (далее – БД) об отказах и сбоях различного оборудования зарубежного производства. ПК позволяет проектировать системы с учетом тех или иных конфигураций на основе зарубежных регламентирующих документов и статистически проверенных данных об элементной базе из справочников [72].

ПО АСОНИКА-К-РЭС – программное средство решения задач анализа и обеспечения надежности в рамках автоматизированного проектирования радиоэлектронной аппаратуры. Функционал этого ПК аналогичен предлагаемому в ПК «FavoWeb», только расчеты проводятся на основе данных, приведенных в отечественных справочниках «Надежность электрорадиоизделий», «Надежность электрорадиоизделий зарубежных аналогов». Комплекс позволяет проводить расчет надежности с использованием различных видов резервирования составных частей (горячее, скользящее горячее), а также обеспечивает способы контроля работоспособности элементов (непрерывный/периодический) [69].

Применение рассмотренного ПО для оценки надежности ВС не представляется возможным из-за ряда причин. Например, приведенное ПО имеет закрытый исходных код и не предусматривает возможность интеграции сторонних модулей. Отсутствует информация о возможности сбора данных с ПЭ ВС. Также, ВС имеют сложно-резервированную систему и трудно формализуемые условия отнесения к неисправным и работоспособным состояниям. Таким образом требуется разработка методов и методик, учитывающих особенности построения и эксплуатации, а также неопределенность в исходных данных о надежности СЭ ВС.

1.2 Организационно-технические мероприятия контроля и поддержания функциональной надежности вычислительной системы

Организационная составляющая обеспечения функциональной надежности ВС включает в себя проведение испытаний на надежность, организацию сбора, обобщения и обработки эксплуатационных сведений, а также планирование эксплуатации и модернизации.

В результате анализа нормативных и регламентирующих документов по организации эксплуатации ВС установлено, что существует расхождение между требованиями, существующей нормативно-правовой базой и сложившейся практикой эксплуатации. Данное расхождение обусловлено обстоятельствами, описанными ниже.

1. Закреплением функций по ТОиР за сторонними организациями при недостаточной проработке области ответственности и порядка взаимодействия с представителями заказчика.

2. Сокращением численности эксплуатирующего персонала, их возможностей и материально-технической оснащенности при сохранении за ними

в полном объеме задач по планированию эксплуатации, учету движения, хранению, сбору информации о надежности.

3. Многочисленностью планирующих и учетных документов при отсутствии их безусловной необходимости и логической связи с качеством эксплуатации, следование которым отнимает значительные ресурсы.

4. Недостаточной наладкой и настройкой оборудования со стороны предприятий, осуществляющих работы по оснащению, а также приемкой подобных систем в эксплуатацию.

Приведенные обстоятельства отражают потребность в решении актуальных задач, а именно определения информации в планирующих, учетных и эксплуатационных документах, требующей корректировки или отмены, для оптимизации эффективности эксплуатации и организация перехода от бумажного ведения документов к электронному формату в планировании, учете и эксплуатации ВС.

Организация сбора, обобщения, обработки и представления сведений о надежности составных элементов вычислительных систем. Совокупность свойств системы, определяющих ее возможность удовлетворять определенные потребности в соответствии с назначением, называют качеством. Выбрав проект построения ВС, устраивающий по основным характеристикам назначения (вычислительная мощность, наличие графических ускорителей и др.) и осуществив оснащение по данному проекту, заказчик сталкивается с проблемой оценки и контроля предполагаемых и фактических свойств полученной системы в эксплуатации. Относительно ВС заказчика в ходе эксплуатации будут интересоваться, прежде всего, вероятность отказа при проведении расчетов на ВС и возможность восстановления системы при возникновении отказов.

Первое качество характеризует способность ВС стабильно работать в различных условиях. Второе качество определено техническими характеристиками, заложенными на этапах проектирования и производства.

ВС являются сложными изделиями, которые имеют разнообразные условия эксплуатации, причины возникновения отказов и методы ремонта. Оценка надежности такой сложной системы осуществляется путем проведения расчетов на основе известных показателей надежности ее составных элементов. Однако точные значения показателей надежности элементов обычно можно получить только путем анализа эксплуатационных данных или статистических экспериментов.

В ряде источников [5; 21] отмечается, что разработчик технической системы и заказчик редко располагают точной информацией о надежности элементов. В то же время, большинство авторов указывает, что основным источником информации о надежности элементов являются сведения, полученные на этапе эксплуатации. Следовательно, для обеспечения контроля качества на этом этапе требуется сбор статистической информации о надежности и отказах оборудования.

Проведение опытной и подконтрольной эксплуатации. Надежность закладывается на этапе разработки изделия и его производства. Испытания призваны оценить и проверить заявленные характеристики, включая надежность. Требованиям к надежности, как правило, посвящается отдельный раздел нормативно-технических документов. При этом в сложившейся практике проведения испытаний, экспериментальная проверка требований к надежности ВС не проводится. Это вызвано ограниченностью сроков проведения испытаний и небольшим количеством предъявляемых на испытания средств. Выполнение требований к показателям надежности закрывается проверкой корректности расчетов надежности.

Одним из источников данных о надежности ВС являются сведения из опытной эксплуатации. Опытная эксплуатация направлена на улучшение процесса использования продукта, учитывая реальные условия эксплуатации, осуществления контроля характеристик продукта в данных условиях и накопления опыта применения. Однако для снижения трудоемкости при проведении опытной эксплуатации составляется максимально простой план испытаний на надежность,

а в отсутствии плана испытаний на надежность, ограничиваются анализом зафиксированных событий. Типовая ситуация характеризуется ограниченным сроком проведения опытной эксплуатации, дополнительной настройкой оцениваемых систем и минимальным количеством участвующих в опытной эксплуатации СЭ.

Для оценки показателей безотказности в описанной ситуации может быть использован метод биномиальных испытаний. Достоинством метода является его применимость в отсутствии априорной информации о виде закона распределения наработки до отказа и возможность привязки момента окончания опытной эксплуатации к конкретному сроку [27].

Приведем пример, пусть при работе n элементов ВС, имеющих постоянную вероятность отказов Q , наблюдалось d отказов. Тогда при малом числе n и d частота отказов подчиняется биномиальному закону,

$$F(d|n, Q) = \sum_{k=0}^d C_n^k Q^k (1-Q)^{n-k}, \quad (1.1)$$

где $F(d|n, Q)$ – условная функция распределения случайной величины d , Q – вероятность отказов, C_n^k – число сочетаний из n элементов по k .

Верхняя граница Q_B доверительного интервала, является корнем уравнения:

$$\sum_{k=0}^d C_n^k Q_B^k (1-Q_B)^{n-k} = \gamma_1, \quad (1.2)$$

а нижняя граница Q_H – корнем уравнения:

$$\sum_{k=d}^n C_n^k Q_H^k (1-Q_H)^{n-k} = \gamma_1, \quad (1.3)$$

где δ – доверительная вероятность, $\gamma_1 = 1 - \delta$ – уровень значимости для одностороннего доверительного интервала.

Для практического вычисления нижней доверительной границы вероятности безотказной работы используем формулы, приведенные в работе [4]:

$$Q_n = 1 - \exp \left[-\frac{x_{1-q}^2(2N-d)}{2K_2(N, N-d)} \right], d > \frac{N}{2}; \quad Q_n = \exp \left[-\frac{x_q^2(2d+2)}{2K_2(N, d+1)} \right], d \leq \frac{N}{2} \quad (1.4)$$

где N – число испытываемых (наблюдаемых) изделий, $x_q^2(l)$ – квантиль хи-квадрат распределения с числом степеней свободы l , соответствующая вероятности q [4, табл. 75], $2K_2(N, l)$ – коэффициент, учитываемый при оценке нижней доверительной границы вероятности безотказной работы при биномиальных испытаниях [4, табл. 68], d – число отказов.

Учитывая, что ВБР $P = 1 - Q$, минимальное количество СЭ, необходимое для оценки вероятности безотказной работы, может быть оценено по формуле:

$$n = \frac{d}{2} - 0,25 \cdot x[100P\%; 2d] + \frac{x[100P\%; 2d]}{(1-P)} \quad (1.5)$$

где $x[100P\%; d]$ – процентная точка χ^2 -распределения с d степенями свободы.

В табл. 1.1, составленной на основе формулы (1.5), представлено необходимое количество СЭ системы, обеспечивающее расчет нижней границы оценки ВБР за интервал времени Δt , с доверительной вероятностью $\delta = 0.9$.

Таблица 1.1 – Схема биномиальных испытаний для оценки ВБР

Число отказов, d	ВБР на периоде, $P(\Delta t)$										
	0.50	0.60	0.70	0.80	0.85	0.90	0.95	0.96	0.97	0.975	0.98
	Минимальное число образцов для оценки										
0	3	6	7	10	14	22	45	56	76	91	114
1	6	9	12	18	25	38	77	96	128	154	193
2	9	12	16	25	34	52	105	131	176	212	264
3	12	15	21	32	43	65	132	167	228	266	332
4	14	18	25	38	52	78	158	198	264	318	398
5	16	21	29	44	60	91	184	230	306	369	462
6	19	24	33	50	68	103	209	261	349	419	>500
7	21	27	37	56	76	116	233	292	390	469	>500
8	23	30	41	62	84	128	258	322	430	>500	>500
9	26	33	45	68	92	140	282	353	471	>500	>700
10	28	36	49	74	100	152	306	382	>500	>500	>700

Используя схему биномиальных испытаний (табл. 1.1), возможна оценка нижней границы для интенсивности отказов. Интенсивность оценивается по следующей формуле:

$$\lambda = \frac{1 - P(d, n(t))}{\Delta t}, \quad (1.6)$$

где $P(d, n(t))$ – выбрана по табл. 1.1 ВБР из множества $\{0.50, \dots, 0.98\}$ такая, чтобы для зафиксированного за период времени Δt числа отказов d в соответствующей строке стояло число $n(t)$ не меньшее, чем количество СЭ, находящихся на момент начала опытной эксплуатации под наблюдением.

Рассмотрим пример, в соответствии с табл. 1.1 и формулой $T = \frac{1}{\lambda}$, для подтверждения средней наработки на отказ T – не менее 5 лет, потребуется поставить в опытную эксплуатацию на срок 1 год не менее 10 образцов. При этом не должно быть зафиксировано ни одного отказа. Наличие одного отказа для 10 образцов для подтверждения тех же значений средней наработки потребует практически двукратного увеличения времени наблюдений.

После проведенных оценок видно, что обеспечение такого масштаба испытаний на надежность даже в рамках опытной эксплуатации может вызвать проблемы. Кроме того, с течением времени происходят деградационные процессы, изменяется интенсивность отказов, и биномиальный анализ надежности, используемый в испытаниях, уже не может адекватно отражать оценку характеристик эффективности систем. Исходя из этого, можно сделать вывод, что результаты опытной эксплуатации и проведение биномиальных испытаний дают лишь приблизительные оценки показателей безотказности. Таким образом, организация сбора информации о надежности вычислительных систем в эксплуатации является необходимой составляющей организационных мероприятий по обеспечению надежности.

Сбор эксплуатационных данных вычислительных систем. До появления цифровых технологий основными источниками эксплуатационных данных СТС являлись рекламационные акты (на период действий гарантийных обязательств),

журналы и формуляры, требующие ручного заполнения. Сбор и обобщение сведений по надежности из всех этих документов трудоемкий, плохо контролируемый и неорганизованный процесс, зачастую осуществляющийся формально [37; 58]. В результате к заказчику и изготовителю поступает недостоверная информация, не позволяющая проводить комплекс работ по обеспечению надежности. Автоматизированная регистрация и обработка информации об отказах и неисправностях является одним из перспективных методов решения данной задачи.

Современные ВС в процессе использования по назначению способны дистанционно проводить контроль технического состояния (автоматическое самотестирование). Объем такого самоконтроля и необходимость проведения ручного контроля остальных параметров зависят от конкретных элементов вычислительной системы [70; 71; 73; 74]. С учетом фактора в целях сбора информации о качестве функционирования предложен метод, изложенный ниже.

Управление объединенных в комплексы ВС, осуществляется системой сбора и обработки информации (далее – ССОИ), функциональные возможности которой обеспечивают, как правило, приведение в действие механизмов (протоколов) проверки работоспособности составных элементов. Ответ на тестовый сигнал позволяет оценить, работоспособно ли устройство, а протоколы общения и наличие адресного пространства позволяют однозначно его идентифицировать. Поступившая по результатам проверки информация обрабатывается в ССОИ и выводится в виде списка генерируемых сообщений на автоматизированное рабочее место оператора (далее – АРМ), или на ряд других АРМ, а также сохраняется в машинной памяти.

Регистрацию и хранение событий по возникшим неполадкам и неисправностям целесообразно реализовывать на основе единой для комплекса БД, которая должна содержать исчерпывающую информацию по каждому СЭ ВС. Это позволит обеспечить не только автоматизированный сбор информации об отказах,

но и электронный учет наличия, движения и эксплуатации СЭ. Соответствующий модуль автоматизированного сбора и анализа поступающих от элементов системы сообщений может быть включен как в программное обеспечение ССОИ, так и быть «перехватывающим», т.е. эти сообщения могут обрабатываться отдельной установленной программой-резидентом.

Алгоритмы функционирования данного модуля должны предусматривать автоматическое формирование и заполнение части полей БД (дата начала эксплуатации, отказа и пр.) при наступлении соответствующих событий в ВС (ввод изделия в эксплуатацию и вывод из нее, отказ, восстановление и пр.). Часть данных (сведения о причинах отказа и пр.) будут заноситься в поля БД вручную, при этом приглашение на заполнение соответствующих полей должно формироваться автоматически. При организации сбора сведений о надежности должно фиксироваться наименование отказавшей функциональной подсистемы, СЭ, а также заводской номер, приведенный в паспорте (формуляре).

Основные фиксируемые причины отказов: климатические факторы, деградационные процессы, нестабильность электропитания, нарушение условий применения, некачественный монтаж, явные конструктивные недостатки, случайные факторы. Основные способы восстановления после отказа в зависимости от вида проводимых при восстановлении изделия работ: замена составной части из ЗИП, ремонт представителями промышленности на объекте, ремонт силами группы эксплуатации объекта, ремонт на предприятии.

В базе данных могут быть предусмотрены поля для возможности занесения дополнительной информации, например:

- время функционирования на момент начала наблюдений (если изделие введено в эксплуатацию ранее ввода в строй системы сбора и обработки информации о надежности);
- конструктивные особенности изделия;
- характеристика внешних условий эксплуатации и их изменения;

- сведения об организации проведения ТОиР;
- способ устранения отказов, число и наименование замененных составных частей;
- продолжительность и трудоемкость поиска и устранения отказа.

БД должна обеспечивать возможность формирования необходимых выборок, сортировки информации по всем введенным в эксплуатацию за заданный период образцам выбранной номенклатуры: общее количество (полный перечень по номерам изделий), в том числе вышедших (вышедших) из строя, сохранивших работоспособность, выведенных из эксплуатации и др. Примеры критериев для сортировки информации представлены ниже:

- наименование составного элемента;
- наименование предприятия-изготовителя;
- дата выпуска (последнего капитального ремонта);
- номера отказавших изделий;
- наименования и номера по каталогу отказавших составных частей;
- наработка изделия до отказа;
- наработка отказавшей составной части.

Таким образом, по каждому введенному в эксплуатацию изделию в базе данных будет содержаться информация, дополняемая при наступлении каждого значимого с точки зрения учета надежности события. Это позволит обеспечить автоматизированный расчет времени от начала функционирования СЭ ВС до момента возникновения отказа.

Стоит отметить, что не для всех элементов вычислительной системы начало наблюдений за ними совпадет с началом работы, а для основной части элементов за текущий период наблюдений отказов не произойдет. Это означает, что большая часть поступающей информации будет цензурированной, причем в основном будет иметь место цензурирование справа.

Планирование эксплуатации. В рамках диссертационного исследования проведен анализ функционирующих ВС с целью определения вопросов планирования мероприятий по эксплуатации, которые требуют математического моделирования и научного обоснования. В результате установлено, что в сводные планы развития и эксплуатации должны быть включены предложения по реконструкции, модернизации, дооснащению и ремонту ВС. В планах развития и эксплуатации должно быть отражено проведение ремонта, замены устаревших и выработавших срок эксплуатации составных элементов ВС, продление сроков их эксплуатации и списание.

В связи с вышеизложенным целесообразна разработка научных методов планирования, обосновывающих выбор оптимального соотношения между различными организационными вариантами обеспечения. Например, определение количества ЗИП в начале эксплуатации, пополнение ЗИП (в процессе эксплуатации, предполагается, что группа эксплуатации способна самостоятельно произвести замену элемента из ЗИП), сроки проведения замены элементов с учетом стоимостей замены.

Эксплуатационно-технические мероприятия. В зависимости от комплектации и степени функциональной законченности изделия возможны различные варианты восстановления его работоспособности. С точки зрения эксплуатации и с учетом свойства ремонтпригодности составные элементы ВС подразделяют на восстанавливаемые сменные модули, являющиеся функционально законченными узлами, и представляющие собой типовые составные элементы, предназначенные для решения установленного объема задач с определенным качеством (допускающие как самостоятельное использование, так и в составе ВС).

Для СЭ ВС установлены различные эксплуатационные требования по режимам и периодичности ТОиР и осуществляется согласно разработанной производителями эксплуатационной документации. Но на уровне ВС задача

поддержания исправности и работоспособности, аналогичная техническому обслуживанию элементов, формулируется уже как задача теории эффективности и исследования операций.

ТОиР СЭ ВС производятся силами имеющейся на объекте группы эксплуатации на месте установки либо в ремонтной мастерской. В последнем случае ремонтируемое изделие должно заменяться аналогичным из состава ЗИП. Средний и капитальный ремонт производится в сервисных центрах, непосредственно на предприятии изготовителя оборудования или с привлечением выездных ремонтных бригад.

В действующих руководящих документах установлено, что текущий ремонт проводится регулировкой либо заменой отказавшего элемента [1]. Это осуществимо при наличии его в ЗИП. Отсутствие неисправного элемента в ЗИП автоматически предполагает необходимость заказа замены в рамках экстренной поставки, отправки в ремонт либо вызова ремонтной бригады. Таким образом, выбор вида ремонта: текущего силами группы эксплуатации или с отправкой (вызовом ремонтной бригады предприятия) определяется не его трудоемкостью, а наличием в ЗИП элемента для замены. Причем последний вариант формально соответствует среднему или капитальному ремонту.

Из опыта эксплуатации продолжительность замены элемента редко может превысить один час, а основные затраты оперативного времени не связаны с непосредственными операциями по ремонту, однако могут занимать практически весь рабочий день. Целесообразно ориентироваться именно на такой срок и устанавливать общее нормативное время восстановления ВС после отказа любого элемента заменой его из ЗИП, равное средней продолжительности рабочего дня или суток в случае обнаружения неисправности в нерабочее время. При наличии более точных сведений и при организации круглосуточного дежурства ремонтного подразделения оперативное время ремонта ВС может быть уточнено.

В случае отсутствия отказавшего элемента в ЗИП время устранения отказа значительно возрастает. Оно будет определяться оперативностью отправки неисправного элемента в ремонтный орган и последующей поставки на объект замены (отремонтированного элемента) или сроками прибытия и работы выездной ремонтной бригады. При оценке временных характеристик следует ориентироваться на действующий порядок и сроки удовлетворения заявок по неисправному оборудованию или на требования и обязательства технического и авторского надзора при наличии соответствующего договора.

ТОиР ВС организационно представлено как независимое для всех входящих в его состав изделий. Каждое из них имеет свою периодичность, глубину, назначенные мероприятия, устанавливаемые индивидуально в собственной эксплуатационной документации. Прямое вмешательство в систему ТОиР оборудования является нежелательным, поскольку его задание и обоснование является прерогативой производителя, но можно оценить затраты на его проведение.

Таким образом, в качестве основных эксплуатационно-технических мероприятий целесообразно рассматривать ремонтные работы по восстановлению неисправного оборудования путем замены и мероприятия по плановой замене оборудования в целях модернизации или окончания срока службы и отсутствия решения на его продление. Осуществление данных мероприятий (помимо затрат на содержание группы эксплуатации) вносит основной вклад в стоимость поддержания ВС в исправном состоянии.

Исходя из изложенного, основной задачей, требующей решения является выбор оптимального соотношения между различными организационными вариантами обеспечения эксплуатации, включающими определение количества ЗИП, закладываемого в начале эксплуатации, условий пополнения ЗИП, условий осуществления гарантийного обслуживания ВС (в том числе с учетом наличия ЗИП).

Модернизация вычислительной системы. В процессе эксплуатации СТС наступает момент, когда сама система либо ее элементы в исправном состоянии требуют замены. Причиной может быть изменение внешней среды, в которой она работает, развитие производства и элементной базы, совершенствование представлений о функционировании системы и изменение требований в соответствии с ними [26; 47]. В последнем случае говорят о необходимости модернизации. Задача определения сроков модернизации ВС решается с целью планирования затрат. Модернизация может быть организована в виде полного обновления всего оборудования ВС, плановой замены отдельных СЭ ВС, допускающих возможность независимой модернизации.

В первом случае возникают дополнительные вынужденные расходы, связанные с заменой части оборудования, выработавшего свой ресурс частично. Дополнительные расходы во втором случае направлены на обследование объекта, установку, подсоединение и совместную настройку существующего и вновь установленного оборудования.

В случае проведения модернизации в виде плановой замены оборудования по окончании назначенного срока службы учет обновленных требований и расширение функциональности достигается применением более совершенных образцов техники. В точках замены отдельно может быть поставлена задача выбора между оборудованием, по которому уже накоплена информация о качестве и надежности по результатам эксплуатации, и новыми более технологичными образцами.

1.3 Определение структурных элементов вычислительной системы и контролируемых параметров

В общем случае архитектура ВС представляет взаимосвязанную систему программно-аппаратных элементов. Структурными единицами ВС в расчете надежности являются заменяемые элементы [6]. Заменяемый элемент — это компонент системы, который в случае сбоя или отказа может быть заменен новым экземпляром для продолжения работы без прерывания. При выборе заменяемых элементов следует учитывать особенности организационно-технических мероприятий по контролю и поддержанию функциональной надежности, а также возможность реализации автоматизированного сбора информации с данных элементов.

В случае с ВС в качестве структурных элементов были определены типовые элементы, представленные в табл. 1.2. Граничные значения описываемых параметров не должны быть строго фиксированными, сохраняя возможность конфигурации рабочей группой с учетом условий эксплуатации ВС.

Таблица 1.2 – Структурные элементы и минимальные необходимые контролируемые параметры

Элемент	Параметры
Сервер	Текущее состояние, доступность по сети мониторинга и управления, доступность по внутренней сети Ethernet, скорость вращения вентиляторов CPU GPU, температура процессоров CPU GPU, температура материнской платы, состояние блоков питания, состояние жестких дисков
Коммутатор	Текущее состояние, доступность по сетям, температура в корпусе, состояние портов
Блоки климат контроля	Текущее состояние, доступность по сетям, температура в корпусе, влажность в корпусе
Кондиционеры	Текущее состояние, доступность по сетям, темп. воздуха на входе/выходе, влажность воздуха на входе/выходе, потребляемая мощность охлаждения
Служебные сервисы	Версия сервиса, уникальный идентификатор каждого запроса, время отклика и статус, код ошибки (если есть), адрес с которого был отправлен запрос, информация о пользователе
Диспетчеризация задач	Состояние расчетных узлов, очередь заданий, загруженность ВС
Базы данных	Информация о последнем успешном резервировании БД, количество подключений к БД, время последнего запуска сервера БД, утилизация ресурсов
Виртуальные машины	Состояние, объем свободной памяти, объем оперативной памяти, доступность по сети

1.4 Анализ методов статистической оценки надежности сложных технических систем в целом

Выбор показателей надежности сложных технических систем (объектов) зависит от типа решаемой задачи (табл. 1.3) и режима эксплуатации оборудования. Например, одной из главных задач организации наблюдений за эксплуатационной надежностью восстанавливаемых объектов является выявление причин отказов с целью уменьшения их влияния на надежность и эффективность применения составных элементов.

Таблица 1.3 – Показатели, используемые при решении задач анализа надежности

Наименование задачи	Используемые показатели
Технологический анализ надежности	T_{cp} – средняя наработка до отказа $\lambda(t)$ – интенсивность отказов $f(t)$ – плотность распределения времени до отказа
Контроль характеристик, заявленных производителем	T_{cp} – средняя наработка до отказа
Выбор наиболее надежных среди аналогов	T_{cp} – средняя наработка до отказа K_r – коэффициент готовности
Прогнозирование сроков службы и обоснование их продления при наличии запаса надежности	t_γ – γ -процентный срок службы
Учет реальной надежности при оценке эффективности	$K_{o,r}$ – коэффициент оперативной готовности
Расчет показателей надежности систем и комплексов	$P(t)$ – ВБР (функция надежности) $\lambda(t)$ – интенсивность отказов T_{cp} – средняя наработка до отказа $K_r(t)$ – нестационарный коэффициент готовности (функция готовности)
Планирование требуемого количества запасных частей	$M(t)$ – математическое ожидание числа отказов

Решению данных задач в наибольшей степени отвечает статистическое точечное оценивание, заключающееся в нахождении плотности или функции распределения времени до отказа для выбранных объектов анализа. Данные показатели предоставляют наибольшую информацию для анализа надежности

элемента. На их основе могут быть получены практически любые другие показатели безотказности [11; 16; 31].

На выбор метода статистической оценки распределения времени до отказа элементов ВС влияют следующие факторы:

- большое количество оцениваемых номенклатур, а также большой разброс в объемах выборок для различных номенклатур с преобладанием малых выборок;
- отсутствие априорной информации о виде функции распределения вероятности безотказной работы для абсолютного большинства оцениваемых номенклатур элементов;
- разнообразие поступающей статистической информации как по качеству, так и по степени общности, поскольку при эксплуатации не придерживаются какого-либо плана испытаний для оценки показателей надежности;
- цензурирование наработок справа, характерное для статистической информации из эксплуатации.

Учет цензурированных данных. Цензурирование – это возникновение неопределенности момента наступления отказа изделия. Причем интервал неопределенности, внутри которого произошел либо произойдет отказ изделия, известен исследователю. Этот интервал может быть неограниченным справа (тогда говорят о цензурировании справа) либо ограниченным справа (тогда говорят о цензурировании слева). Если интервал неопределенности ограничен и справа, и слева, то говорят о цензурировании интервалом. Получаемые в результате выборки называются цензурированными [28].

Полная наработка – это наработка изделия до отказа. Неполная наработка – это наработка от начала испытаний или эксплуатационных наблюдений до прекращения испытаний или эксплуатационных наблюдений. Предполагается что в пределах данной наработки отказа не произошло. Условная наработка при цензурировании слева – это значение интервала наработки, в

пределах которого произошел отказ изделия. Здесь и далее по умолчанию наработка задана в единицах времени.

Различают цензурирование I-го и II-го типа. Основное отличие заключается в том, что при цензурировании I-го типа число наблюдений с реализовавшимся признаком является случайной величиной, а при цензурировании II-го типа оно постоянно и заранее задано, поэтому в теории цензурирование I-го типа сложнее [9; 37].

Выбор метода статистической оценки надежности ВС. Учитывая отсутствие априорной информации о виде функции распределения и значительный массив данных, которые будут поступать, метод анализа должен быть машинно-ориентированным и достаточно универсальным, не привязанным к конкретным видам законов распределений времени до отказа, развитым для обработки цензурированной статистики. В таких случаях принято обращаться к непараметрическим методам [60; 66].

Возможно использование универсальных распределений, таких как семейства распределений Джонсона – Пирсона. Однако поиск параметров этих распределений достаточно трудоемкий, особенно при наличии цензурированных данных. Высшие моменты сильно зависят от случайных колебаний выборки, что делает более простой метод моментов неэффективным для аппроксимации многопараметрическими семействами, особенно при работе с малыми выборками. Использование одно- или двухпараметрических распределений ограничивает универсальность и возможности по аппроксимации [14].

Использование непараметрических методов позволяет аппроксимировать массивы экспериментальной информации и получать оценки эмпирической плотности распределения вероятностей. В частности, к таким методам относятся метод построения гистограмм, метод Розенблатта – Парзена [60; 62], а также проекционные [53], экспоненциальные [50] и корневые оценки [8; 54].

Несмотря на состоятельность гистограммных оценок, их недостатком является то, что они представляют собой разрывную ступенчатую функцию, что не всегда удобно для практического использования. Другие перечисленные методы позволяют получить гладкие оценки, более близкие по свойствам к реальным функциям плотности времени до отказа. Известны результаты [14], доказывающие состоятельность таких оценок, хотя в различных условиях они демонстрируют разную скорость сходимости. В отличие от методов подгонки типовых распределений, аппроксимация на основе рядов Грамма-Шарлье и применения универсальных распределений Джонсона и Пирсона позволяет строить мультимодальные распределения с «настраиваемой гладкостью». Это может оказаться важным при выяснении причин отказов, так как зоны сгущения точек вариационного ряда могут быть как статистическими выбросами, так и свидетельствовать о наличии закономерностей.

Оценки, основанные на ортогональных функциях, содержат относительно небольшое число слагаемых, а расчеты требуют меньшего объема памяти и времени, чем ядерные оценки. К их числу принадлежит впервые предложенный Н.Н. Ченцовым метод проекционного оценивания [53]. Для них разработаны сравнительно простые алгоритмы восстановления плотности. Для получения проекционной оценки Н.Н. Ченцов использовал разложение в ряд Фурье. Методы на основе применения ортонормированных базисных функций (Лежандра, Лагерра, Дирихле) получили развитие в Самарском государственном аэрокосмическом университете им. академика С.П. Королева [24; 39].

Развитием проекционных оценок являются методы корневого и экспонентного оценивания. Суть этих методов заключается в том, что для первого вместо функции плотности в ортонормированной системе разлагается ее квадрат, а для второго – логарифм.

Проекционные и основанные на ортогональных разложениях оценки обладают следующими общими недостатками: полученная оценка может не

удовлетворять условию нормировки, а также условию неотрицательности (т. е. в общем случае они не являются плотностями). К тому же, применение в методах проекционного и корневого оценивания периодических функций приводит к тому, что в полученной плотности распределения времени до отказа будут иметь место всплески и понижения, не обусловленные характером исходных данных. Они способны вводить в заблуждение исследователя при анализе надежности и выявлении причин отказов. Для исправления указанных недостатков, к примеру, отрицательные значения могут быть заменены на ноль с последующей нормировкой по единичному интегралу, при этом погрешность, оцененная в метрике L_1 , уменьшится. Однако, как отмечал Н.Н. Ченцов в своей статье [53], в этом случае для плотности распределения времени до отказа увеличивается сложность представления оценки и хранения информации.

Таким образом, проекционные и гистограммно-аппроксимационные оценки как, например, в диссертации О.А. Дегтяревой, использующие в качестве базиса функции Лагерра, Лежандра, Дирихле и Эрмита, сложнее в реализации, чем метод Розенблатта – Парзена. Существенно также то, что для них отсутствуют теоретические наработки по учету цензурированной справа информации.

Доказательство состоятельности метода Розенблатта – Парзена приводится, например, в монографии [14]. Непараметрические методы оценки функции плотности распределения времени до отказа, включая данный метод, нашли практическое применение для определения характеристик надежности оборудования АЭС в Обнинском государственном техническом университете атомной энергетики, где при участии и под руководством А.В. Антонова [7; 41] они были развиты для аппроксимации цензурированной информации. Там же исследованы методы, устраняющие граничные эффекты ядерных оценок плотностей для неотрицательных случайных величин (нарушение условий нормировки) [33].

Вместе с тем в проанализированных источниках отсутствует представленное в явном виде аналитическое выражение для функции распределения времени до отказа, объединяющее как полные наработки, так и все виды цензурированных данных. Расчет для включенных в оценку цензурированных справа данных демонстрирует хорошие результаты лишь при малой их доле, так как он приводит к нарушению нормировки. Способ расчета параметра локальности приведен лишь в общем виде, и требуется разработка более подробного алгоритма. Данные обстоятельства обуславливают необходимость проведения дополнительных исследований и совершенствования метода ядерных оценок в целях его практического использования для оценки показателей безотказности СЭ ВС.

1.5 Выводы по главе

В рамках первой главы получены следующие результаты:

1. Обоснована необходимость разработки системы сбора информации о надежности элементов с учетом современного уровня развития техники. В настоящее время у эксплуатирующих ВС организаций отсутствует достоверная информация о параметрах надежности как предлагаемых к принятию на снабжение, так и уже функционирующих систем.

2. Представлено описание методики автоматизированного сбора и подготовки данных из эксплуатации для выявления отказов и неисправностей элементов ВС. Предложена структура БД, для формирования информации из эксплуатации СЭ ВС. Предложены критерии сортировки информации и порядок ведения БД.

3. Применение схемы биномиальных испытаний для подтверждения свойств надежности при проведении опытной эксплуатации нецелесообразно. Показаны условия, обеспечивающие возможность подтверждения некоторых характеристик надежности ВС при проведении опытной и подконтрольной эксплуатации. Для

подтверждения с доверительной вероятностью $\delta = 0,9$ средней наработки на отказ не менее 5 лет потребуется поставить в опытную эксплуатацию на срок 1 год не менее 10 образцов. Таким образом, при проведении опытной и подконтрольной эксплуатации требуются значительное число образцов и продолжительные сроки наблюдений. Данное обстоятельство подтверждает необходимость организации сбора информации о надежности элементов ВС непосредственно в эксплуатации.

4. Продемонстрировано, что для решения задачи оценки показателей надежности СЭ ВС недостаточно существующего научно-методического обеспечения. Разрабатываемая новая методика расчета надежности должна учитывать особенности построения и эксплуатации ВС. Обоснована потребность в адаптации метода Розенблатта – Парзена для оценки надежности СЭ ВС.

ГЛАВА 2. Разработка методики оценки надежности вычислительных систем по малым выборкам эксплуатационных данных

2.1 Технология сбора и подготовки эксплуатационных данных вычислительных систем

Подготовка данных является важным этапом в создании эффективных моделей интеллектуального анализа данных и занимает до 70-80% общего времени проекта [55; 56]. В связи с тем, что реальные данные могут быть неполными, зашумленными и противоречивыми, экспертам требуются систематизированные, подробные методики и автоматизированные инструменты, для уменьшения усилий и временных затрат, необходимых для выполнения этого этапа.

Так, например, в работе [22] описана методика процесса подготовки данных, применяемая к различным областям исследований, но без детализации этапов решения задач. Методики [35; 55] ориентированы только на конкретный этап подготовки данных, очистку или выбор данных. В работе [35] авторы предлагают разработку новой методики анализа данных в области энергетической безопасности из-за ограничений существующих методологий. Авторы [64] разрабатывают приложение на основе методики интеллектуального анализа данных. Однако проанализированные работы носят ограниченный характер и применимы только к своим отраслям.

В первой главе диссертационной работы выявлены особенности и проблемы данных, поступающих в процессе эксплуатации ВС. Исходя из основной цели сформированы требования по наполнению набора данных. Учитывая, архитектуру ВС и установленный порядок фиксации контролируемых значений и отказов, необходимо формировать наборы данных по каждому контролируемому устройству с учетом возможности автоматизации данного процесса [19].

Разработка архитектуры сбора данных. Проведенный анализ решений в области мониторинга аппаратных и программных средств показал, что реализация требуемого функционала по сбору данных может быть достигнута при использовании трехуровневой архитектуры, представленной на рис. 2.1. При таком решении передача информации осуществляется между соседними уровнями, позволяя модифицировать ВС и ее СЭ не нарушая общую работоспособность разработанного программного модуля [51].



Рисунок 2.1 – Архитектура сбора данных

Архитектура логически разделена на три формальных уровня, с возможностью передачи данных только между соседними уровнями. Такой подход позволяет модифицировать и обновлять уровни, не нарушая общей работоспособности. Существуют реализации, использующие меньшее число уровней или выделяющие уровни явно [23], но для решения задач, указанных в предыдущей главе, минимально возможным числом для распределения структурных элементов являются три уровня.

Предполагается, что ВС имеет иерархическую структуру, так как программные элементы не могут существовать без аппаратной части. Примером иерархической структуры является управляющий сервер, где дочерним элементом выступает система управления диспетчеризации задач (далее – СУДЗ), а также другие СЭ объединения вычислительных серверов в единое расчетное поле.

В работе [57] проведено исследование эффективности различных структур сбора данных. В результате использование иерархического сбора данных (пример на рис. 2.2) в сравнении с одноранговой системой приводит к повышению производительности от 10 до 600 раз для различных уровней обработки.

Сборку всех метрик в единое дерево предлагается осуществлять корнем иерархии ЦСОД, что позволяет оптимизировать распределение ресурсов на промежуточном уровне.

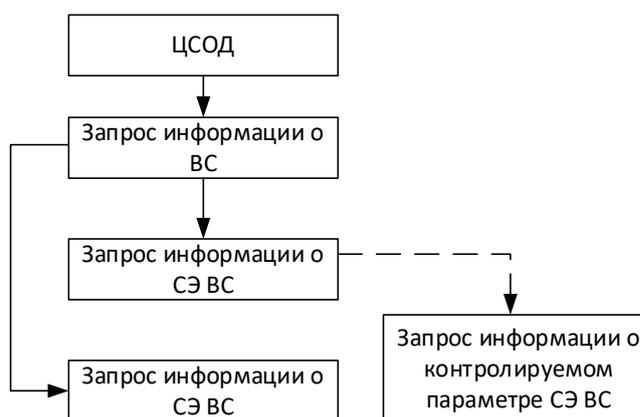


Рисунок 2.2 – Уровни запроса данных

При получении метрик от датчиков используются соединения, сеансы и кадры. Кадр — это набор данных, характеризующих метрики и их значения в определенный момент времени. В общем случае в одном кадре содержатся данные для нескольких метрик.

Метрика — это набор значений формализованного типа (строки, целые числа, числа с плавающей запятой и логические переменные и т. д.), отражающий текущее состояние или изменение контролируемого параметра составного элемента ВС. Кроме того, составной элемент может иметь атрибуты.

Атрибут отличается от метрики тем, что он всегда хранит только одно значение. Атрибуты могут выступать в качестве априорных данных и использоваться в работе. Например, в случае программных компонент ими могут являться параметры терминальной строки, переданные приложению пользователя, а в случае аппаратных элементов информация о их введении в эксплуатацию (дата, организация и т.д.).

Отметим, что в предложенной архитектуре полученные значения метрик остаются неизменными до поступления следующего кадра и являются ориентированными. Кадр является логическим представлением данных, не зависящим от типа соединения.

Соединение — это инструмент, при помощи которого обеспечивается доступ к данным посредством сетевых коммуникаций, а также внутренних интерфейсов устройств, открытия локального файла, запроса к БД. Результатом работы соединения является возможность открытия сеансов. Также его можно рассматривать как реализацию механизма получения метрик в неструктурированном формате от источников, к которым при инициализации устанавливается подключение.

Через заданные промежутки времени источник данных делает мгновенный снимок состояния метрик и передает его в процессе сеанса обмена данными на ЦСОД, ответственному за обработку информации. Источник данных работает на нижнем уровне архитектуры, ЦСОД на промежуточном или уровне приложений.

Сеанс — это объект, идентифицирующий итерацию процесса получения данных. Созданием сеансов занимаются ЦСОД.

Триггер — это логический элемент, содержащий выражение или условие, истинность которого необходимо проверить для последнего значения метрики. Если условие выполнено, то состояние триггера изменяется, что приводит к выполнению заранее определенных действий.

Действие — набор системных инструкций, реакция на срабатывание триггера. Действия могут быть разнообразными: посылка уведомления администратору кластера; запись диагностического сообщения в журнал событий; запуск приложений; проверка еще одного триггера. Одному триггеру может быть сопоставлено несколько действий и наоборот: одному действию может быть сопоставлено несколько триггеров. Поэтому было предложено внести в архитектуру механизм подписки на уведомления о событиях.

Возможность уведомлений о событиях позволяет устранить перечисленные выше недостатки. Каждый контролируемый узел системы мониторинга может генерировать сообщение о возникшем событии. Уведомление может поставляться в рамках автоматически создаваемой метрики, например, с типом данных «JSON – объект». Эта метрика может поставляться как по обычному соединению, по которому передаются кадры данных. Если ЦСОД сохраняет историю данных метрик, то метрика события, как и всякая другая, будет сохранена для дальнейшего анализа.

Пользовательские приложения регистрируются на ЦСОД уже в процессе работы. ЦСОД получает информацию о событии и определяет необходимость уведомления подписчиков, если оно подходит под ранее установленный фильтр. При наличии зарегистрированного приложения ему отправляется уведомление. Далее, с целью проверки подписки сообщение транслируется на другие ЦСОД.

Клиентское приложение имеет возможность обработать события, игнорировать его (в этом случае событие передается для дальнейшей обработки на ретрансляторе) или же генерировать новые события. Например, при возникновении события отказа сетевой платы, подписавшееся приложение может сгенерировать более конкретные события: «необходимость миграции задач с узла»; «необходимость уведомления администратора»; «необходимость удаления узла из списка ресурсов вычислительного кластера» — которые, в свою очередь, могут быть обработаны другими подписчиками.

Формализация уровней архитектуры. Разработанная архитектура позволяет динамически формировать структуру данных, элементами которой может выступать любой определяемый аппаратный или программный компонент. При этом базовой единицей хранения данных являются значения метрик [52].

Первый уровень архитектуры отвечает за сбор и представление данных в виде метрик. В случае с аппаратными элементами источниками данных выступают сенсоры логических или физических устройств, например, данные служб сторонних СМ или SNMP-статистика. Более подробно контролируемые показатели представлены в табл. 1.2.

Второй уровень отвечает за сбор данных с группы источников, преобразование их к единообразному формату данных, проверку простейших триггеров и выполнение соответствующих им взаимодействий. Этими задачами занимаются центры сбора и обработки данных.

Третий уровень отвечает за добавление и удаление СЭ ВС в исследуемую систему, контроль показателей надежности, анализ статистических данных и представление их пользователю в удобном интерфейсе.

Описание протокола взаимодействия. Протокол взаимодействия описывает порядок представления данных между уровнями архитектуры. В случае с СЭ ВС с закрытым кодом доступные метрики определяются архитектурой разработчика данных элементов. Если СЭ ВС имеет открытый код или разрабатываются впервые, протоколом взаимодействия определяется Open-API/Rest-API для программных элементов и SNMP для аппаратных элементов.

В случае наличия неконтролируемых аппаратных элементов, например, системы кондиционирования, (не предоставляющие сведения о функционировании) предлагается использовать сторонние модули для опроса их текущего состояния.

Поэтапный процесс подготовки данных для обучения машинных моделей. Согласно [22], предварительный анализ состоит из последовательных этапов, необходимых для определения целевой переменной, выявления отклонений и особенностей исходных данных. Процесс подготовки данных можно разделить на два этапа, что позволит автоматизировать часть задач процесса подготовки данных.

Первый этап. Общая подготовка данных. К первому этапу относятся задачи, не зависящие от исследуемого объекта, и могут быть применены в других областях СТС. На этом этапе осуществляется сбор и получение данных, необходимых для перехода к решению специального набора задач.

После сбора данных необходимо произвести их извлечение и преобразование в удобный для анализа формат. Для этого может потребоваться применение различных методов и инструментов, таких как очистка данных от лишних символов или форматирование данных для унификации.

Очистка данных. Степень преобразования и очистки исходных данных зависит от степени их зашумленности. Удаления пропусков из набора данных может привести к потере зависимостей [18; 46]. Наиболее эффективным является использование методов восстановления пропущенных данных. В соответствии с проведенным сравнительным анализом [43] в нашем случае целесообразно использование гибридного адаптационного метода для заполнения пропущенных значений в массивах данных, содержащих информацию о процессах, происходящих в сложных динамических системах [42]. Еще одним важным аспектом первого этапа является оценка качества данных. Это включает в себя проверку данных на наличие ошибок, пропущенных значений, дубликатов и противоречий. Также может потребоваться проверка соответствия полученных данных требуемым критериям и стандартам.

Второй этап. Специальная подготовка данных. На втором этапе, имеющий более тесную связь с конечным объектом исследования решаются задачи форматирования, формирования и интеграции данных.

Форматирование данных. Изменение типа данных и синтаксической структуры данных атрибутов и значений (при необходимости). На данном этапе определяется целевая переменная. Целевая переменная является фактором с некоторым количеством уровней (классов). При наличии большего количества классов принимаются специальные дополнительные меры для решения таких задач [21]. В нашем случае, при решении задачи формирования целевой переменной необходимо определить состояние устройства «В РАБОТЕ» или «ОТКАЗ» в заданный момент времени на основе показателей контролируемых параметров.

Формирование данных. Выявление, получение и создание новых атрибутов или заполнение пропущенных значений с помощью арифметических операций. Для части параметров требуется формирование производных признаков с дальнейшим преобразованием в векторы.

Интеграция данных. При решении данной задачи происходит интеграция с хранилищем данных и определение схемы интеграции. Уделяется внимание выявлению и исправлению конфликтов интеграции, а также установлению соотношения данных. Необходимым этапом является создание таблиц с атрибутами: ключ, время события, параметры, состояние контролируемого устройства. Окончательная интеграция набора данных заключается в формировании итогового набора, с содержанием всех событий за указанный период эксплуатации [30; 32].

Автоматизация этапов подготовки данных. Учитывая регулярность появления задач по обработке данных и особенности работы СМ, выполнена автоматизация специализированного набора задач по предварительной обработке данных, с целью повышения эффективности оценки показателей надежности и оптимизации временных затрат в долгосрочной перспективе. Это необходимо для минимизации

пропусков, выявленных в ходе предварительного анализа. Основная идея конфигурации СМ заключается в синхронизации ответов от применяемых в эксплуатации датчиков опроса состояния ВС. Данный шаг позволит снизить нагрузку на процесс автоматизации и увеличить точность восстанавливаемых данных.

Схема автоматизации этапов подготовки данных представлена на рис. 2.3. В качестве инструментов для разработки использован язык программирования общего назначения *Python 3.10* с библиотеками *scikit-learn*, *pandas*, *numpy* и язык *SQL* для установления связи между приложением и хранилищем данных.

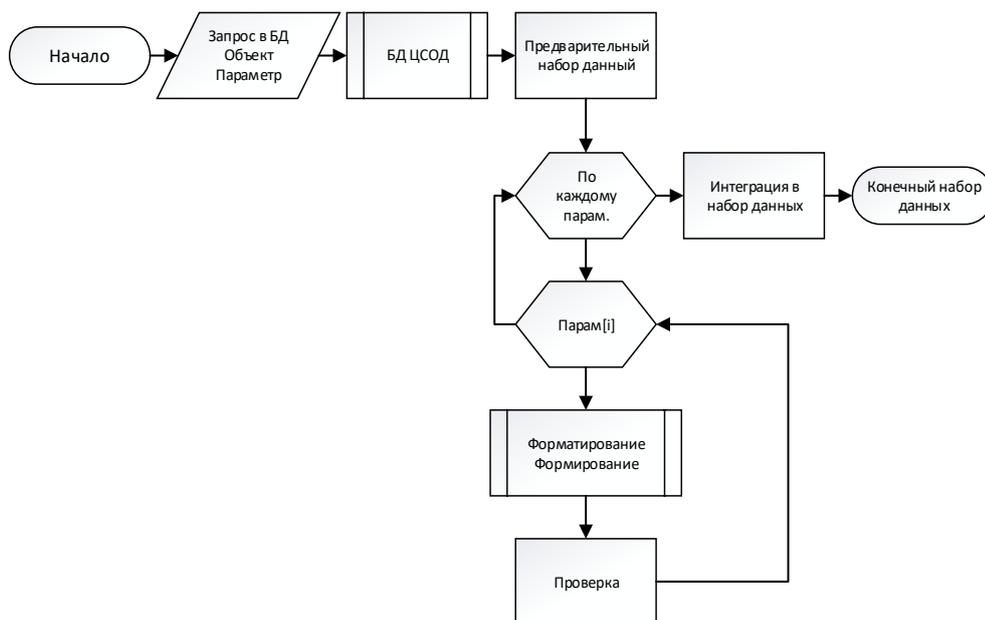


Рисунок 2.3 – Схема автоматизации этапов подготовки данных

Реализация схемы автоматизации подготовки данных состоит из нескольких этапов: выполнение запросов в БД с помощью правил *Python DB-API* и *SQL*-запросов, формирование предварительных выборок с использованием библиотеки *matplotlib.pandas*, отбор параметров и объектов с учетом их релевантности, качества и технических ограничений, очистка данных путем удаления некорректных значений, исключения дублей и однотипных параметров, а также восстановления уникальности, целостности и логических связей с помощью гибридного адаптационного метода. Затем производится создание производных

признаков, их преобразование в векторы для моделей машинного обучения, трансформация для повышения точности алгоритмов машинного обучения, интеграция данных из различных источников, включая агрегацию и вычисление новых значений путем суммирования информации из множества записей, а также форматирование и формирование набора данных [15; 49]. Полученные результаты применены в разработке программного обеспечения для оценки надежности вычислительных систем.

Экспериментальная проверка автоматизации этапов подготовки данных. Проверка метода автоматизации проведена на данных, в сравнении с ручным способом подготовки данных. Из наборов случайным образом удалялось некоторое количество значений параметров, тем самым имитировалось наличие пропусков. Оценка качества восстановления таких значений, проводилась по доле пропущенных значений (точности) и средней квадратичной ошибки (погрешности) [18]. Проверка проводилась на двух наборах данных. «Набор 1» – выборка значений, сформированных в период с 02.10.2019 по 02.12.2019. Число показателей – 18, число событий – 2330. «Набор 2» – выборка значений, сформированных в период с 01.02.2021 по 01.03.2021. Число показателей – 17, число событий – 1250. Результаты эксперимента представлены в табл. 2.1. В табл. 2.2 показано сравнение времени выполнения этапов подготовки данных при выполнении вручную и автоматически.

Таблица 2.1 – Сравнение качественных показателей подготовки данных

	Набор данных №1		Набор данных №2	
	Точность	Погрешность	Точность	Погрешность
Форматирование данных	90,7 %	12 %	94,7 %	9 %

Таблица 2.2 – Усредненное сравнение временных показателей подготовки данных

Задача	Ручная подготовка (минуты)	Автоматизированная подготовка (минуты)	Снижение временных затрат на:
Форматирование данных	33,53	0,058	99,83%
Формирование признаков	5,16	0,33	99,61%
Интеграция данных	40,5	1,4	96,55%

Результаты сравнения временных показателей подготовки данных, доказывают значительное сокращение времени при автоматизации этапов подготовки данных.

2.2 Модели машинного обучения для распознавания отказов в общем потоке информации системы мониторинга

В большинстве случаев задача распознавания отказов в общем потоке информации, поступающей от системы мониторинга, решается с помощью оценки граничных параметров каждого из контролируемых значений. Однако нередко бывают ситуации, когда контролируемые устройства формируют сигнал работоспособного состояния, при этом такое устройство, по каким-либо причинам, будет недоступно. В этом случае целесообразно применение методов машинного обучения.

Общая логика функционирования обученных моделей заключается в следующем. Показатели контролируемых параметров СЭ ВС, передаются в модель машинного обучения, которая в свою очередь проводит комплексную обработку данных. На основе этой обработки модель формирует прогностическую информацию о текущем состоянии оборудования. Если модель распознает отклонение или предполагает возможный отказ, она генерирует соответствующее предупреждение и рекомендации для операторов или технического персонала. Практическим результатом обработки является сформированный массив значимых событий по каждому СЭ ВС.

Входные данные. Важным фактором при обучении машинных моделей является наличие исторических данных из эксплуатации. В качестве входных данных сформирована и подготовлена выборка за период эксплуатации с 01.01.2017 по 01.01.2021 по СЭ ВС, разработанной ранее технологией сбора и подготовки данных, представленная в табл. 2.3.

Таблица 2.3 – Составные элементы вычислительной системы

Элемент	Тех. характеристика	Значение	Кол-во
Управляющий сервер	Тип процессора	Intel i7 - 8700	1
	Объем оперативной памяти	16 Гб	
	Объем физической памяти	1 Тб	
	Модель платы мониторинга	Supermicro AOC-SIMSO	
	Дополнительный адаптер Gigabit Ethernet	PCI/PCI-X 10/100/1000Mbps	
	Интерфейс ввода-вывода сервисной сети	Консольный порт и порт Fast Ethernet (BMC)	
Вычислительный сервер	Тип процессора	Intel i7 - 8700	20
	Объем оперативной памяти	32 Гб	
	Объем физической памяти	1 Тб	
	Модель платы мониторинга	Supermicro AOC-SIMSO	
	Дополнительный адаптер Gigabit Ethernet	PCI/PCI-X 10/100/1000Mbps	
	Интерфейс ввода-вывода сервисной сети	Консольный порт и порт Fast Ethernet (BMC)	
Система диспетчеризации задач «СУДЗ»	Протокол взаимодействия	Open-API/Rest-API	1
Система мониторинга	Протокол взаимодействия	Open-API/Rest-API	1

Итоговым результатом применения технологии сбора и подготовки эксплуатационных данных является сформированная выборка по каждому СЭ ВС с размеченной целевой переменной. Пример таких данных представлен в табл. 2.4, где отражена частичная выборка по управляющему серверу.

Таблица 2.4 – Составные элементы вычислительной системы

Время фиксации	Парам. 1	Парам. 2	Парам. ...	Целевая переменная
10:24:20 20.11.2019	23,5	24,4	...	1
10:25:20 20.11.2019	23,7	25,2	...	1
...
10:26:24 27.11.2019	23,9	26,3	...	0
10:28:40 27.11.2019	23,5	24,6	...	1

Обучение моделей. Для распознавания отказов СЭ ВС возможно рассмотреть один из следующих подходов к обучению машинных моделей.

Использование общей модели. Общая модель включает в себя обучение на объединенных данных от СЭ ВС. Такой подход в перспективе позволит упростить

архитектуру разрабатываемого программного обеспечения. Так как использование одной модели для всех компонентов облегчает разработку и обслуживание системы. В этом случае нет необходимости создавать и поддерживать отдельные модели для каждого компонента.

Использование частных моделей. В случае с частными моделями распознавания отказов происходит обучение отдельных моделей для каждого типа составного элемента. Частные модели обладают следующими преимуществами:

- независимость от других моделей. Использование отдельных моделей для каждого типа позволяет изолировать процесс мониторинга отказов и не влиять на уже действующие модели. Если одна модель не функционирует должным образом или требует изменений, это не повлияет на остальные модели системы;

- адаптивность к различным компонентам. Разные компоненты системы могут обладать разными характеристиками и поведением при возникновении отказов. Использование отдельных моделей позволяет настраивать параметры для каждого компонента отдельно и получать более точное распознавание отказов.

Функционирование ВС с его особенностями, такими как постоянное масштабирование и модификация, ограничивает реализацию первого варианта и обосновывает целесообразность использования частных моделей распознавания отказов.

Выбор базовой модели. Выбор базовой модели зависит от характеристик входных данных, скорости обучения модели и ее ресурсоемкости в процессе эксплуатации. Так, например, эксплуатационные данные от составных элементов ВС характеризуются дисбалансом классов с преобладанием малых выборок на отказ.

В таком случае, в качестве базовых архитектур при решении задачи распознавания отказов целесообразно использовать легковесные архитектуры моделей машинного обучения. К таким моделям относятся:

Логистическая регрессия. Это статистический метод для предсказания вероятности принадлежности объекта к определенному классу. Она использует линейную комбинацию входных признаков и применяет нелинейную функцию для получения вероятности [20; 34].

Деревья решений. Это метод, основанный на построении дерева с решающими правилами на основе входных данных. Каждый узел дерева представляет тест на одном из признаков, а каждое ребро соответствует возможному результату этого теста. Дерево решений может быть использовано для классификации путем прохождения по дереву от корня до листьев [36].

При этом, существует возможность использовать методы основаны на применении нейронных сетей. Так, авторы работ [44; 45] предлагают использовать решения, основанные на применении сверточных нейронных сетей (англ. convolutional neural network, CNN) при решении задач распознавания отказов в составных элементах летальных аппаратов. Однако, как отмечают сами авторы данные методы требуют большого объема данных для обучения каждой модели и значительных ресурсов для ее обучения [38; 65].

Обучение частных моделей распознавания отказов можно представить в виде блок-схемы (рис. 2.4), а также в виде алгоритма (табл. 2.5).



Рисунок 2.4 – Блок-схема алгоритма обучения машинных моделей

Таблица 2.5 – Алгоритм обучения машинных моделей для распознавания отказов

Шаг 1. Из генеральной совокупности событий формируется обучающая и тестовая выборка в соотношении 80 и 20% от объема соответственно;

Шаг 2. На основе предварительного анализа обучающей выборки производится выбор базовой модели интеллектуального анализа, а также выбор функции для оценки качества обучения модели;

Шаг 3. Определение управляющих параметров базовых моделей;

Шаг 4. Обучение модели с применением метода кросс-валидации и поиском оптимальных параметров по сетке;

Шаг 5. Оценка качества модели. Выбор итоговой модели.

Наиболее распространенный вид кросс-валидации — это кросс-валидация с k -блоками (k -fold cross-validation). В этом случае данные разбиваются на k равных размеров блоков (обычно 5 или 10). После разделения данных модель обучается k раз, каждый раз используя $k-1$ блок в качестве обучающего набора данных и оставшийся блок в качестве тестового набора данных. В конечном итоге производится усреднение показателей производительности модели на всех k итерациях для получения окончательной оценки. Кросс-валидация позволяет максимально использовать доступные данные, так как каждая точка данных используется как часть обучающего и тестового набора данных. Во-вторых, кросс-валидация обеспечивает более объективную оценку производительности модели, поскольку устраняет влияние конкретного разбиения данных на эту оценку. Пример разделения данных на рис. 2.5.

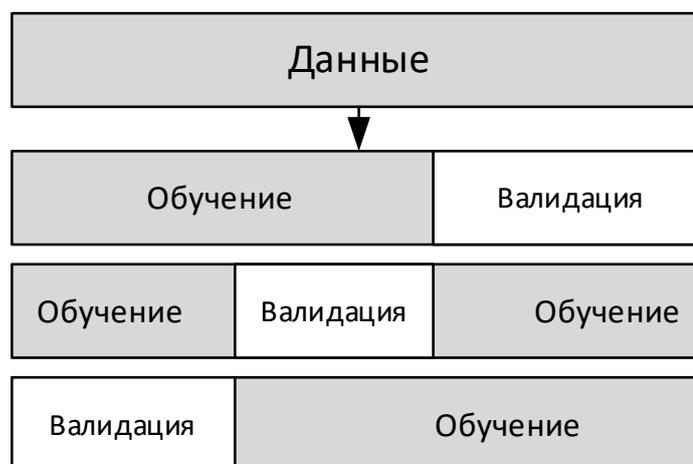


Рисунок 2.5 – Разделение данных на 5 блоков в методе кросс-валидации

Эксперимент по исследованию точности машинных моделей в задаче распознавания отказов. Проведен эксперимент по оценке точности обученных моделей.

Входные данные. Рассмотрим на примере СЭ ВС «Управляющий сервер». В период с 01.01.2020 по 01.01.2021 по составным элементам ВС сформирован набор данных из 8560 событий, из которых 311 событий описывают состояние как «ОТКАЗ».

Шаг 1. Из генеральной совокупности событий формируется обучающая и тестовая выборка в соотношении 80 и 20% от объема, с учетом дисбаланса классов.

Шаг 2. В качестве базовых моделей определены три архитектуры «CatBoostClassifier» [61] и «Xgboost» [63], основанные на алгоритме «дерево решений», и одна модель логистической регрессии. Выводы о качестве основывались на вычислении средней абсолютной ошибки (MAE) по выражению 2.1 [12]:

$$MAE = \frac{1}{N} \sum_{l=1}^N |y_l - \hat{y}_l|, \quad (2.1)$$

где N - число примеров обучающей выборки, y_l - целевое значение l -го примера, \hat{y}_l - предсказанное моделью значение.

Шаг 3. Для выбранных базовых моделей определены управляющие параметры, представленные в табл. 2.6.

Шаг 4. Обучение моделей начинается с инициализации параметров модели. Затем модель применяется к обучающим данным, а функция потерь используется для оценки ошибки модели на каждой итерации обучения. Когда функция потерь перестает улучшаться или достигает заданного порогового значения, обучение завершается.

Шаг 5. В результате обучения формируются «итоговые» модель, достигшие в процессе обучения лучших показателей. Полученные модели оцениваются на

ранее сформированном тестовом наборе данных. В табл. 2.7 приведены результаты и параметры для итоговых моделей данного СЭ.

Таблица 2.6 – Управляющие параметры базовых моделей

Базовая модель	Управляющие параметры
CATboost Classifier	<pre>params = { 'iterations': [10, 20,30,40,50], 'Learning rate': [0.01,0.02,0.03,0.04], 'depth': [4,5,6,7,8,9, 10] }</pre>
Xgboost Classifier	<pre>params = { 'min_child_weight': [1, 5, 10], 'learning rate': [0.01,0.02,0.03,0.04], 'gamma': [0.5, 1, 1.5, 2, 5], 'subsample': [0.6, 0.8, 1.0], 'colsample_bytree': [0.6, 0.8, 1.0], 'max_depth': [4,5,6,7,8,9, 10] }</pre>
Logistic Regression	<pre>params = { 'classifier__penalty': ['l1', 'l2'], 'classifier__C': np.logspace(-4, 4, 20)}</pre>

Таблица 2.7 – Оценка качества итоговых моделей

Итоговая модель	Precision (точность)	Recall (полнота)	F1 - метрика
CATboost Classifier	0,86	0,89	0,88
Xgboost Classifier	0,88	0,97	0,92
Logistic Regression	0,61	0,86	0,71

Таким образом, в результате обучения для СЭ ВС «Управляющий сервер» определена итоговая модель, основанная на Xgboost Classifier со следующими параметрами:

```
XGBClassifier: (base_score=0.5, booster='gbtree', colsample_bylevel=1, gamma=5,
    colsample_bytree=0.6, learning_rate=0.02, max_delta_step=0 max_depth=5,
    min_child_weight=5, missing=None, n_estimators=100, n_jobs=1, nthread=1,
    objective='binary:logistic', random_state=0, reg_alpha=0, reg_lambda=1,
    scale_pos_weight=1, seed=None, silent=True, subsample=1.0)
```

2.3 Адаптация непараметрического метода Розенблатта – Парзена для оценки показателей надежности

Постановка задачи. Как отмечают авторы [13], чтобы информация в процессе обработки не была утеряна, для малых выборок применяется «индивидуальный подход», заключающийся в присвоении каждой реализации некоторой элементарной плотности.

В этом случае для полных наработок x_1, \dots, x_N эмпирическая функция распределения времени до отказа может быть представлена в виде:

$$F(t) = \int_{-\infty}^t \frac{1}{N} \sum_{i=1}^N \delta(u, x_i) du, \quad (2.2)$$

где $\delta(u, x_i)$ – дельта-функция Дирака [17].

В результате замены дельта-функции некоторой функцией V , удовлетворяющей условиям $V \geq 0$ и $\int V dt = 1$, будет иметь место выражение, вполне соответствующее традиционному подходу к построению оценки функции распределения времени до отказа:

$$F(t) = \int_{-\infty}^t \frac{1}{N\sigma} \sum_{i=1}^N V\left(\frac{u - x_i}{\sigma}\right) du, \quad (2.3)$$

где под интегралом представлено выражение для суммы оценок плотности распределения; σ – управляющий параметр локальности.

Вместе с тем в проанализированных источниках отсутствует представленное в явном виде аналитическое выражение для функции распределения времени до отказа, объединяющее как полные наработки, так и все виды цензурированных данных. Приведенный способ расчета для включенных в оценку цензурированных справа данных демонстрирует хорошие результаты лишь при малой их доле, так как он приводит к нарушению нормировки. Способ расчета параметра локальности приведен в общем виде, и требуется разработка более подробного алгоритма. Данные обстоятельства обуславливают необходимость проведения

дополнительных исследований и совершенствования метода Розенблатта – Парзена в целях его практического использования для оценки показателей безотказности СЭ ВС.

Алгоритм расчета параметра локальности. Одна из ключевых задач, требующих решения при применении метода Розенблатта – Парзена является определение подходящего параметра локальности. Установлено, что оценки являются состоятельными [14] при выполнении условий для σ :

$$\lim_{N \rightarrow \infty} \sigma(N) = 0, \quad \lim_{N \rightarrow \infty} \sigma(N) \cdot N = \infty, \quad (2.4)$$

где N – объем выборки.

С учетом (2.4) параметр σ представим в виде [59]:

$$\sigma = \frac{\mu}{N^\nu}, \quad (2.5)$$

где $0 < \nu < 1$, μ – коэффициент размерности для случайной величины.

Выражение (2.5) допускает выбор σ в достаточно широких пределах, что приводит к постановке задачи поиска его оптимального значения. На основе (2.25) предлагается следующая эмпирическая формула для нахождения начального приближения σ , которая обеспечивает необходимый порядок его задания:

$$\sigma = \frac{\tau}{x_\tau \cdot \sqrt[5]{N}}, \quad (2.6)$$

где τ – некоторый момент времени, который соответствует числу известных наработок p (при числе наблюдаемых элементов N). Он, например, может соответствовать \bar{t} – границе цензурирования справа; x_τ – квантиль свернутого нормального распределения $\Phi_s(t)$ такой, что $\Phi_s(x_\tau) = \frac{p}{N}$.

В дополнении к [14] авторы работы [10] предлагают определять параметр локальности на основе принципа невязки аналогично нахождению оптимальной константы регуляризации при решении некорректно поставленных задач.

Суть метода нахождения оптимальной константы регуляризации – выбрать константу σ , чтобы обеспечивалось выполнение равенства:

$$\rho_{E_1}(F_\sigma, F_N) = \rho_{E_2}(F, F_N) = \delta, \quad (2.7)$$

где $\rho_{E_2}(F, F_N)$ – близость оценки $F_N(x)$ к истинной $F(x)$, $F_\sigma(x) = Af_\sigma$, f_σ – решение, полученное методом регуляризации с константой σ , A – линейный оператор.

Для приближенной реализации метода невязки метрика $\rho_{E_1}(F_\sigma, F_N)$ заменяется, например, метрикой $L_2(F)$. Тогда, задав нужный квантиль θ распределения ω^2 , параметр σ определяется из равенства:

$$N \int (F_\sigma(x) - F_N(x))^2 dF_\sigma(x) = \theta. \quad (2.8)$$

Автор работы [9] предлагает для вычисления ω^2 – статистики общую формулу, допускающую ситуацию, когда не все N выборочных наблюдений различны, а сосредоточены в нескольких точках в количествах $n_j, j = \overline{1, J}$, которые могут быть отличны от 1. Данная формула отвечает случаю цензурирования интервалами:

$$N \int (F_\sigma(x) - F_N(x))^2 dF_\sigma(x) = \frac{1}{3} + \frac{1}{N} \left\{ \sum_{j=1}^J n_j \left[F_\sigma^2(x_j) - \frac{F_\sigma(x_j)}{N} \left(2 \sum_{i=1}^j n_i - n_j \right) \right] \right\}. \quad (2.9)$$

Стоит отметить, что поиск значения параметра локальности σ по уравнению (2.9) представляет собой сложную задачу, поэтому, в рамках диссертации, для адаптируемого метода Розенблатта – Парзена, нахождение σ предлагается осуществить с помощью алгоритма Хука – Дживса и вычисления функции правдоподобия для кросс-проверочных плотностей [48]:

Выражение для функции правдоподобия применительно к цензурированной выборке имеет вид:

$$L(\sigma) = \prod_{c=1}^I \left\{ \prod_{l=1}^{p_c} f^{cross}(x_{cl}) \cdot \prod_{l=1}^{J_c} f^{cross} \left((j-1) \cdot \Delta_c + \frac{\Delta_c}{2} \right) \cdot f^{cross} \left(t_{nc} - \bar{t}_c + \frac{(t_{nc} - \bar{t}_c)}{2} \right) \right\}, \quad (2.10)$$

для логарифмической функции правдоподобия:

$$\ln L(\sigma) = \sum_{c=1}^I \left\{ \sum_{l=1}^{p_c} \ln [f^{cross}(x_{cl})] + \sum_{l=1}^{J_c} \ln \left[f^{cross} \left((j-1) \cdot \Delta_c + \frac{\Delta_c}{2} \right) \right] + \ln \left[f^{cross} \left(t_{nc} - \bar{t}_c + \frac{(t_{nc} - \bar{t}_c)}{2} \right) \right] \right\}, \quad (2.11)$$

где $f^{cross}(x_{cl})$, $f^{cross}(x_{cl}^{\Delta})$, $f^{cross}(x_c^{\Pi})$ – кросс-проверочные плотности, которые строятся для каждого из значений выборки x_{cl} (для $c = \overline{k, I}, l = \overline{1, p_c}$), $x_{cl}^{\Delta} = (l-1)\Delta_c + \frac{\Delta_c}{2}$ (для $c = \overline{k, I}, l = \overline{1, J_c}$), $x_c^{\Pi} = t_{nc} - \bar{t}_c + \frac{(t_{nc} - \bar{t}_c)}{2}$ (для $c = \overline{k, I}$).

Шаги алгоритма для расчета оптимального значения параметра локальности:

1. Расчет базисного значения σ_6 по выражению (2. 6), шаг изменения значения параметра $d = \sigma_6/2$, выбираем значение e останова алгоритма. Переход на п.2.
2. Расчет значения функции правдоподобия $L(\sigma_6)$, переход на п.3.
3. Расчет $\sigma^+ = \sigma_6 + d$ и $L(\sigma^+)$, переход на п. 4.
4. Если $L(\sigma^+) > L(\sigma_6)$, то $\sigma_6 = \sigma^+$ на п.3, иначе переход на п.5.
5. Расчет $\sigma^- = \sigma_6 - d$, переход на п. 6.
6. Если $\sigma^- < 0$, то переход на п. 9, в противном случае переход на п.7.
7. Расчет функции правдоподобия (логарифмической) $L(\sigma^-)$.
8. Если $L(\sigma^-) > L(\sigma_6)$, то $\sigma_6 = \sigma^-$ и переход на п. 3, в противном случае переход на п. 9.
9. Сравнение значений d и e . Если $d < e$, то искомое значение σ_6 найдено, иначе $d := d/2$ и переход на п.3.

Схема алгоритма для расчета оптимального значения параметра локальности приведена на рис. 2.6.

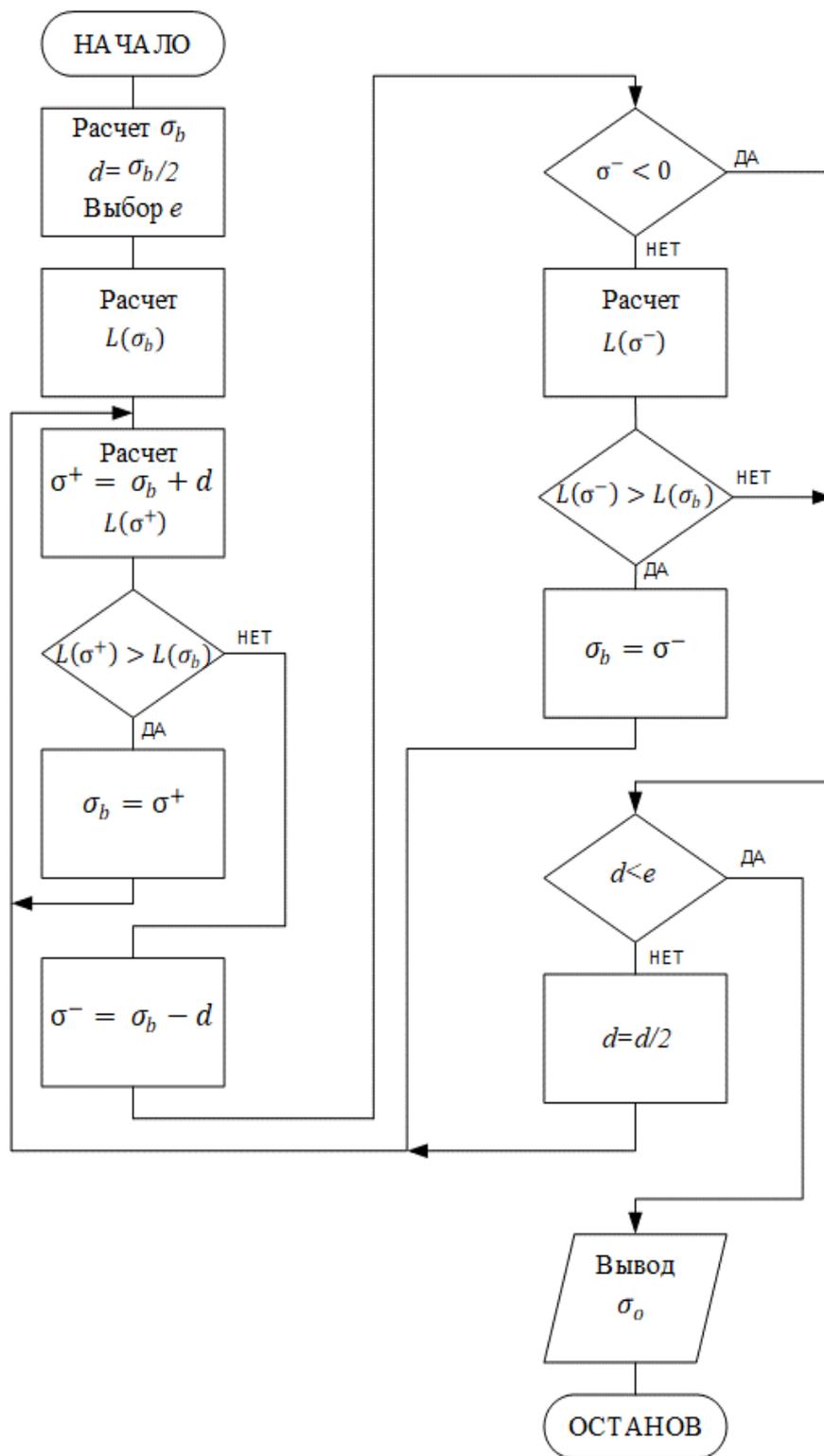


Рисунок 2.6 – Схема алгоритма расчета оптимального значения параметра локальности

Проведена экспериментальная оценка влияния параметра локальности на точность аппроксимации плотности распределения. В рамках эксперимента проведено сравнение параметров локальности, полученных адаптированным методом Хука – Дживса (рис. 2.6) и методом Сильвермана (2.6). Схема эксперимента выглядит следующим образом.

1. Моделирование набора данных по закону распределения Вейбулла с параметрами $m = 1,1$ и $\theta = 1000$:

$$f(t) = \frac{m}{\theta} \cdot t^{m-1} \cdot e^{-\frac{t^m}{\theta}},$$

2. Из набора данных сформированы выборки объемом $n=15$ и $n=30$.

3. Адаптированным методом Хука – Дживса (рис. 2.6) и методом Сильвермана (2.26) восстанавливается исходная плотность распределения.

4. Выводы о точности оценок основываются на вычислении ошибок оценивания e_n в метрике L_1 -пространства по формуле:

$$e_n = \int_0^{\infty} |f_t(t) - f_s(t)| dt,$$

где $f_t(t)$ – плотность распределения модельных данных, $f_s(t)$ – восстановленная плотность методом Розенблатта – Парзена.

В табл. 2.8 представлены максимальные значения ошибок оценивания функции e_n при различных объемах случайных выборок и их параметров локальности.

Таблица 2.8 – Значения ошибок оценивания функции e_n

	n = 5	n = 10	n = 30	n = 50	n = 100	n = 130
$\sigma_b(silverman)$	0,033	0,022	0,021	0,019	0,017	0,016
$\sigma_o(optimal)$	0,031	0,019	0,018	0,015	0,011	0,009

На рис. 2.7 и рис. 2.8 приведены результаты аппроксимации плотности распределения с разными параметрами локальности, где *silverman* (σ_b) рассчитан методом Сильвермана, а *optimal_value* (σ_o) получен адаптированным методом Хука – Дживса.

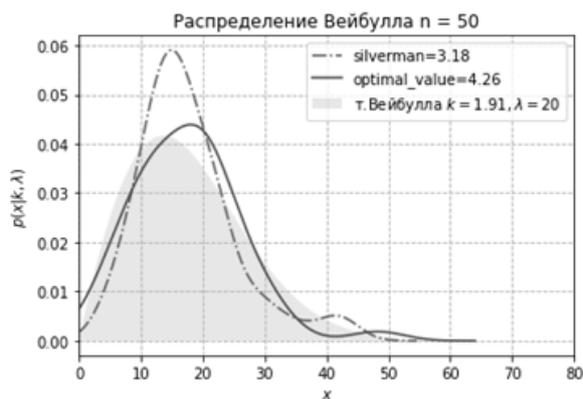


Рисунок 2.7 – Оценка плотности распределения при $n = 50$

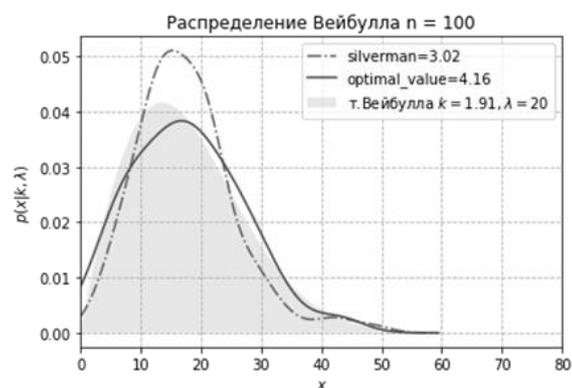


Рисунок 2.8 – Оценка плотности распределения при $n = 100$

Результаты оценки подтверждают тенденцию роста дисперсии при малых значениях σ и увеличения систематической ошибки при больших значениях σ , а также позволяют сделать выводы об улучшении точности оценивания при использовании оптимального параметра локальности и объема выборки. На рис. 2.9 приведены результаты восстановления плотности по малым выборкам, при различных значениях числа отказов.

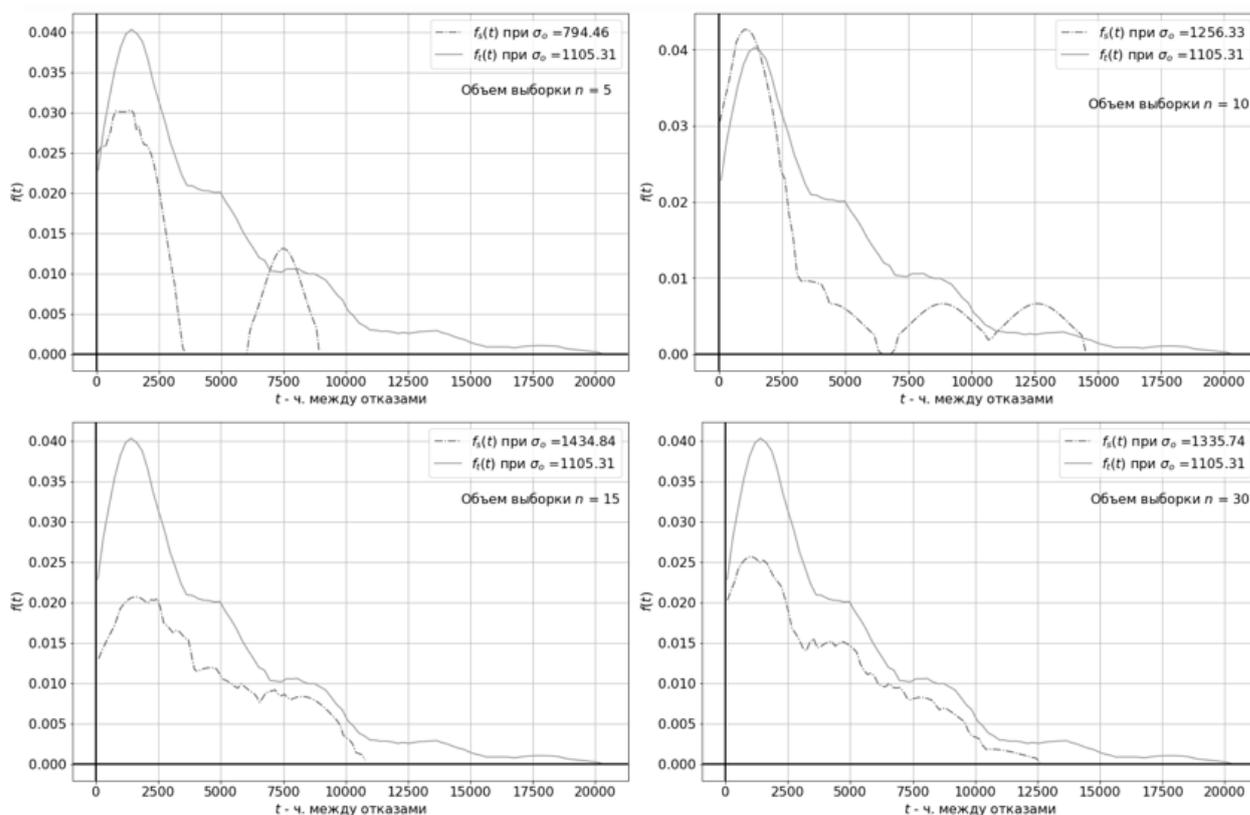


Рисунок 2.9 – Оценка плотности распределения при различных значениях n .

Как видно из графиков, уже при размере выборки, равной 15, и использовании оптимального параметра локальности, полученная оценка плотности достигает заданного уровня точности.

Оценка для полных наработок. Введем обозначения. Пусть от некоторого источника (объекта) имеется следующая статистическая информация по СЭ ВС:

N – число наблюдаемых элементов объекта;

$\vec{X} = (x_1, x_2, \dots, x_p)$ – массив полных наработок;

$\vec{L} = [(l_1, l_2); (l_2, l_3); \dots; (l_{J-1}, l_J)]$ – массив интервалов цензурирования, на которых зафиксировано n_1, n_2, \dots, n_J наработок. Причем $l_1 = 0$; $l_j + \Delta_j = l_{j+1}$, $j = \overline{1, J-1}$, где Δ_j – длина интервала цензурирования. Не нарушая общности, будем рассматривать частный случай – с одинаковым интервалом Δ . Описание массива \vec{L} удовлетворяет условиям периодического контроля с равными интервалами.

Рассмотрим элементарную плотность, соответствующую одному отказу (в формулах вывода только для одной элементарной плотности индексы опущены). При заданной дисперсии нормальное распределение имеет максимальную энтропию, так что априорным данным в виде дисперсии σ^2 центрированной случайной величины соответствует априорная компонента оценки плотности распределения в форме плотности нормального распределения [13]. Поэтому наиболее естественным выбором в качестве элементарной плотности является гауссовское ядро:

$$V(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{x^2}{2\sigma^2}}, \quad (2.12)$$

где σ – параметр локальности.

Наблюдаемая величина, для которой строится функция распределения (время), задано на области определения $[0, +\infty)$. Поскольку выбранная в качестве ядра функция Гаусса является непрерывной на $(-\infty, +\infty)$, то в окрестности точки ноль будет присутствовать относительное смещение (рис. 2.7),

и не будет выполняться условие $F(0) = 0$, подразумевающее работоспособность в момент начала эксплуатации. В связи с этим, необходимо исследование и применение методов компенсации смещения.

Компенсация смещения полных наработок. Для компенсации смещения известны такие методы как усеченные ядра и метод зеркального отображения исходных данных [40]. Исследования данных методов компенсации проведено аналогично предыдущему эксперименты, на модельных данных (рис. 2.10).

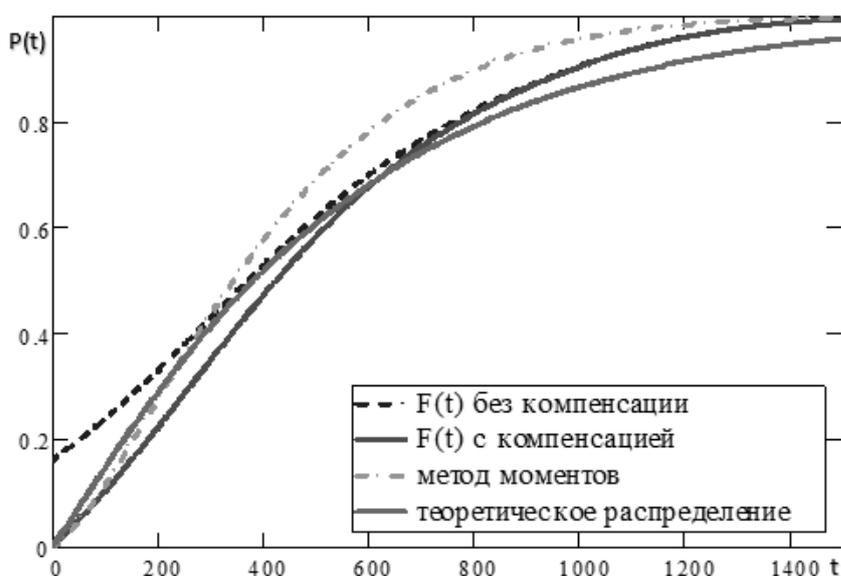


Рисунок 2.10 – Визуализация результатов компенсации смещения методом зеркального отображения данных.

В результате исследования определено, что использование метода зеркального отображения исходных данных уменьшает ошибку теоретической функции распределения.

Суть метода зеркального отображения исходных данных заключается в замене симметричного ядра $V\left(\frac{t-x}{\sigma}\right)$ свернутым нормальным ядром [9]:

$$\varphi_s(t) = \left[V\left(\frac{t-x}{\sigma}\right) + V\left(\frac{t+x}{\sigma}\right) \right] = \frac{1}{\sqrt{2\pi}\sigma} \cdot \left(e^{-\frac{(t-x)^2}{2\sigma^2}} + e^{-\frac{(t+x)^2}{2\sigma^2}} \right) \quad (2.13)$$

с установлением в ноль нижней границы области определения функции распределения. В этом случае для единичной наработки x функция распределения $F(t)$ представима в виде:

$$F(t) = \int_0^t f(u) du = \frac{1}{\sigma} \int_0^t \left[V\left(\frac{u-x}{\sigma}\right) + V\left(\frac{u+x}{\sigma}\right) \right] du = \left[\Phi\left(\frac{t-x}{\sigma}\right) + \Phi\left(\frac{t+x}{\sigma}\right) - 1 \right], \quad (2.14)$$

где $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{v^2}{2}} dv$.

Для последующего единообразного представления по всем компонентам исходных данных будем использовать функцию $erf(x)$:

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt = 2 \Phi(x\sqrt{2}) - 1. \quad (2.15)$$

Тогда из формулы (2.5) для единичной наработки получим:

$$F(t) = \frac{1}{2} \left[erf\left(\frac{t-x}{\sigma\sqrt{2}}\right) + erf\left(\frac{t+x}{\sigma\sqrt{2}}\right) \right], \quad (2.16)$$

где x – значение наработки до отказа.

Таким образом, расчет оценки плотности и функции распределения полных наработок с компенсацией смещения проводится по формулам (2.14) и (2.16).

Оценка при цензурировании интервалом. Априорным данным в виде интервала изменения случайной величины адекватна компонента оценки плотности распределения в форме плотности равномерного распределения. Поэтому, используя гауссовское ядро, плотность распределения $f(t)$ для единичной реализации наработки x с учетом ее равномерной плотности на интервале цензурирования $[l, l + \Delta]$ представима в виде:

$$f(t) = \int_l^{l+\Delta} \frac{1}{\Delta} \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(t-x)^2}{2\sigma^2}} dx = \frac{1}{\Delta} \left(\Phi\left(\frac{t-l}{\sigma}\right) - \Phi\left(\frac{t-l-\Delta}{\sigma}\right) \right). \quad (2.17)$$

Тогда для $F(t)$ с учетом (2.14) будем иметь:

$$\begin{aligned}
 F(t) &= \int_{-\infty}^t f(u) du = \int_{-\infty}^t \int_l^{l+\Delta} \frac{1}{\Delta} \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(u-x)^2}{2\sigma^2}} dx du = \\
 &= \frac{1}{\Delta} \int_l^{l+\Delta} \int_{-\infty}^t \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(u-x)^2}{2\sigma^2}} du dx = \frac{1}{\Delta} \int_l^{l+\Delta} \Phi\left(\frac{t-x}{\sigma}\right) dx.
 \end{aligned} \tag{2.18}$$

При замене переменных $\frac{t-x}{\sigma} = v$, $dx = -\sigma dv$ верхний предел интегрирования в формуле (2.17) равен $\frac{t-l-\Delta}{\sigma}$, а нижний $-\frac{t-l}{\sigma}$. Тогда

$$F(t) = \frac{1}{\Delta} \int_l^{l+\Delta} \Phi\left(\frac{t-x}{\sigma}\right) dx = -\frac{\sigma}{\Delta} \int_{\frac{t-l}{\sigma}}^{\frac{t-l-\Delta}{\sigma}} \Phi(v) dv. \tag{2.19}$$

Используем функцию $erf(x)$ и поменяем местами пределы интегрирования в выражении (2.17):

$$\begin{aligned}
 F(t) &= -\frac{\sigma}{\Delta} \int_{\frac{t-l}{\sigma}}^{\frac{t-l-\Delta}{\sigma}} \Phi(v) dv = \frac{\sigma}{2\Delta} \int_{\frac{t-l-\Delta}{\sigma}}^{\frac{t-l}{\sigma}} \left[1 + erf\left(\frac{v}{\sqrt{2}}\right)\right] dv = \\
 &= \frac{\sigma}{2\Delta} \left[v \Big|_{\frac{t-l-\Delta}{\sigma}}^{\frac{t-l}{\sigma}} + \int_{\frac{t-l-\Delta}{\sigma}}^{\frac{t-l}{\sigma}} erf\left(\frac{v}{\sqrt{2}}\right) dv \right] = \frac{1}{2} + \frac{\sigma}{2\Delta} \int_{\frac{t-l-\Delta}{\sigma}}^{\frac{t-l}{\sigma}} erf\left(\frac{v}{\sqrt{2}}\right) dv,
 \end{aligned} \tag{2.20}$$

сделав замену переменных $z = \frac{v}{\sqrt{2}}$ и учитывая (формула 1.5.1.9 на стр. 30 [25]), что

$$\int erf x dx = x \cdot erf x + \frac{1}{\sqrt{\pi}} e^{-x^2},$$

получим выражение для функции распределения времени до отказа:

$$\begin{aligned}
 F(t) &= \frac{1}{2} + \frac{\sigma}{\sqrt{2}\Delta} \int_{\frac{t-l-\Delta}{\sqrt{2}\sigma}}^{\frac{t-l}{\sqrt{2}\sigma}} erf(z) dz = \frac{1}{2} + \frac{\sigma}{\sqrt{2}\Delta} \left[\left(z \cdot erf z + \frac{1}{\sqrt{\pi}} e^{-z^2} \right) \Big|_{\frac{t-l-\Delta}{\sigma\sqrt{2}}}^{\frac{t-l}{\sigma\sqrt{2}}} \right] = \\
 &= \frac{1}{2} + \frac{\sigma}{2\Delta} \left[\frac{t-l}{\sigma} erf\left(\frac{t-l}{\sigma\sqrt{2}}\right) - \frac{t-l-\Delta}{\sigma} erf\left(\frac{t-l-\Delta}{\sigma\sqrt{2}}\right) + \frac{\sqrt{2}}{\sqrt{\pi}} \left(e^{-\frac{(t-l)^2}{2\sigma^2}} - e^{-\frac{(t-l-\Delta)^2}{2\sigma^2}} \right) \right].
 \end{aligned} \tag{2.21}$$

После проведения аналогичных преобразований для свернутого ядра (2.14) функция распределения для цензурированной интервалом наработки примет вид:

$$\begin{aligned}
 F(t) = \frac{\sigma}{2\Delta} & \left[\frac{t-l}{\sigma} \operatorname{erf}\left(\frac{t-l}{\sigma\sqrt{2}}\right) - \frac{t-l-\Delta}{\sigma} \operatorname{erf}\left(\frac{t-l-\Delta}{\sigma\sqrt{2}}\right) \right. \\
 & + \frac{\sqrt{2}}{\sqrt{\pi}} \left(e^{-\frac{(t-l)^2}{2\sigma^2}} - e^{-\frac{(t-l-\Delta)^2}{2\sigma^2}} \right) + \frac{t+l+\Delta}{\sigma} \operatorname{erf}\left(\frac{t+l+\Delta}{\sigma\sqrt{2}}\right) \\
 & \left. - \frac{t+l}{\sigma} \operatorname{erf}\left(\frac{t+l}{\sigma\sqrt{2}}\right) + \frac{\sqrt{2}}{\sqrt{\pi}} \left(e^{-\frac{(t+l+\Delta)^2}{2\sigma^2}} - e^{-\frac{(t+l)^2}{2\sigma^2}} \right) \right]. \quad (2.22)
 \end{aligned}$$

Таким образом, выведено аналитическое выражение для функции распределения времени до отказа, представляющее цензурированные интервалом данные в ядерных оценках с компенсацией смещения.

Оценка при цензурировании данных справа. Цензурирование справа можно охарактеризовать следующим образом: после некоторого момента последней регистрации наработок \bar{t} осталось m работоспособных элементов, отказы которых когда-нибудь в будущем неизбежно произойдут, но времена их наступления неизвестны.

Если статистическую оценку строить только по полным наработкам и цензурированным интервалами данным, то оценка будет построена по усеченной выборке и надежность будет занижена. График функции распределения пройдет выше теоретического (рис. 2.11). Учет цензурированных справа данных корректирует оценку в сторону приближения к теоретическому распределению.

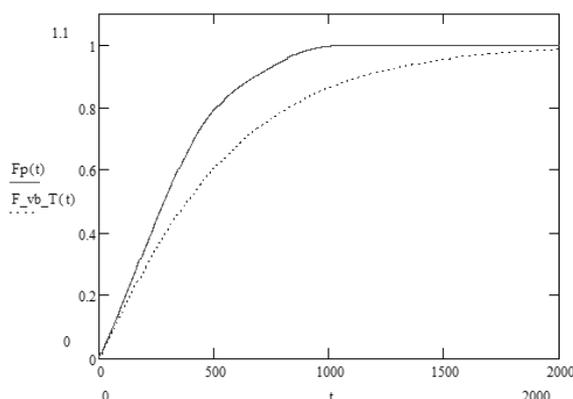


Рисунок 2.11 – Сравнение положения графика статистических оценок, теоретическое распределение (непрерывная линия) и не учитывающих цензурированные справа данные (пунктирная линия)

Для учета цензурирования справа в методе ядерных оценок описан подход, заключающийся в замене случайной величины t , распределенной на интервале $(\bar{t}, +\infty)$, случайной величиной $\tau = \frac{1}{t}$, которая будет уже распределена на интервале $(0, \frac{1}{\bar{t}})$ [7].

В этом случае функции распределения τ и t и их плотности будут взаимосвязаны следующим образом:

$$G(\tau) = p(x \leq \tau) = p\left(\frac{1}{x} \geq \frac{1}{\tau}\right) = 1 - p\left(\frac{1}{x} \leq \frac{1}{\tau}\right) = 1 - F\left(\frac{1}{\tau}\right),$$

$$g(\tau) = \frac{1}{\tau^2} f\left(\frac{1}{\tau}\right).$$

В случае выбора гауссова ядра плотность распределения в соответствии с [33] имеет вид:

$$g(t) = \frac{\bar{t}}{t^2} \left[\Phi\left(\frac{1}{\sigma t}\right) - \Phi\left(\frac{1}{\sigma} \left(\frac{1}{t} - \frac{1}{\bar{t}}\right)\right) \right]. \quad (2.23)$$

На основании (2.23) функция распределения может быть представлена выражением:

$$G(t) = \frac{1}{2} - \frac{\bar{t} \cdot \sigma}{2} \left[\frac{1}{\sigma t} \operatorname{erf}\left(\frac{1}{\sqrt{2}\sigma t}\right) - \frac{1}{\sigma} \frac{1}{t} \operatorname{erf}\left(\frac{1}{\sqrt{2}\sigma} \left(\frac{1}{t} - \frac{1}{\bar{t}}\right)\right) + \sqrt{\frac{2}{\pi}} \left(e^{-\frac{1}{2t^2\sigma^2}} - e^{-\frac{(\frac{1}{t} - \frac{1}{\bar{t}})^2}{2\sigma^2}} \right) \right]. \quad (2.24)$$

При этом, плотность $g(t)$ остается распределенной на $(-\infty, +\infty)$ и $\int_0^\infty g(t) < 1$ в отличие от того, что требуется для наработки.

Таким образом, необходимость применения методов компенсации смещения становится неизбежной. С применением метода зеркального отображения данных для плотности и функции распределения получены формулы:

$$g(t) = \frac{\bar{t}}{t^2} \left[\Phi \left(\frac{\frac{1}{t} + \frac{1}{\bar{t}}}{\sigma} \right) - \Phi \left(\frac{\frac{1}{t} - \frac{1}{\bar{t}}}{\sigma} \right) \right], \quad (2.25)$$

$$G(t) = 1 - \frac{\bar{t} \cdot \sigma}{2} \left[\frac{1}{\frac{1}{t} + \frac{1}{\bar{t}}} \operatorname{erf} \left(\frac{\frac{1}{t} + \frac{1}{\bar{t}}}{\sigma\sqrt{2}} \right) - \frac{1}{\frac{1}{t} - \frac{1}{\bar{t}}} \operatorname{erf} \left(\frac{\frac{1}{t} - \frac{1}{\bar{t}}}{\sqrt{2}\sigma} \right) + \frac{\sqrt{2}}{\sqrt{\pi}} \left(e^{-\frac{(\frac{1}{t} + \frac{1}{\bar{t}})^2}{2\sigma^2}} - e^{-\frac{(\frac{1}{t} - \frac{1}{\bar{t}})^2}{2\sigma^2}} \right) \right]. \quad (2.26)$$

Особенностью и недостатком функций (2.23 – 2.24) является то, что они устанавливают единое поведение для всех цензурированных справа данных от любых источников, так как имеют только два управляющих параметра: значение границы \bar{t} и параметр локальности σ . Они не зависят от поведения всей предыдущей статистики и, как показывают исследования, малочувствительны к изменению управляющих параметров. Одна из возможных альтернатив – использование для цензурированных данных априорной информации. В этом случае ядро представимо в виде:

$$f(x) = \frac{1}{\sigma} \int V \left(\frac{u-x}{\sigma} \right) [F(u)]' du, \quad (2.25)$$

где $[F(u)]'$ – плотность априорного распределения.

Источником априорных данных должны быть сведения о надежности аналогичного оборудования, в частности, некоторое согласованное экспертами конечное значение t_{Π} , на которое может быть заменена бесконечность на правой границе. Если это единственная априорная информация и нет сведений о поведении функции распределения на интервале (\bar{t}, t_{Π}) , то на основании принципа максимума энтропии оценка строится по выражению (2.17) для цензурированных интервалами данных. При отсутствии априорной информации о значении t_{Π} для его

определения предлагается воспользоваться приемом, заключающимся в расчете t_{Π} для каждого из источников, исходя из пропорции относительно числа элементов в источнике и числа отказов, произошедших до наступления цензурирования:

$$t_{\Pi} = \frac{N}{N - r} \cdot \bar{t},$$

где r – число цензурированных справа наработок, N – число элементов в источнике.

Данный прием обладает следующими достоинствами:

- простота оценки;
- поведение функции распределения за пределами границы цензурирования увязано с ее поведением до этой границы;
- достигается единство используемых выражений;
- обеспечивается гладкость функции в точке \bar{t} ;
- снимается неопределенность относительно цензурированной информации, что позволяет в дальнейшем осуществлять прогнозирование.

Оценка поведения распределения в интервале цензурирования справа в зависимости от направленности ошибки может быть заниженной и завышенной. Занижение показателей надежности относительно истинного значения может привести к закладке большего запаса ЗИП на период планирования, чем требовалось бы. Завышение оценки приводит к неэффективному эксплуатированию ВС.

Однако следует иметь в виду, что с набором статистики оценка будет корректироваться. Кроме того, со временем ценность СЭ ВС, даже работоспособных, будет снижаться ввиду их морального старения, в связи с этим оно может подлежать замене уже по другим основаниям. Чем более важна к какому-то сроку работоспособность конкретного образца изделия, тем больше необходимость сохранять резерв образцов именно этого типа для замены. И чем менее важно сохранение работоспособности изделий, тем с большей готовностью принимается решение на использование новых типов образцов (т. е. проведение

модернизации), так что чрезмерное количество ЗИП даже недостаточно надежных изделий также неразумно.

Из перечисленных факторов следуют выводы: срок службы и необходимое количество ЗИП должны регулярно пересматриваться на основании вновь полученной информации. Стратегия пополнения запасов ЗИП имеет преимущество над разовой его закладкой на весь продолжительный срок службы, если стоимость доставки ЗИП не будет высока, а эксплуатируемые изделия представлены в массовом количестве.

Данный вариант оценки функции распределения в наибольшей степени избавлен от нежелательного влияния не находящихся под наблюдением элементов. Из всех приведенных он лучше всего подходит для технологического анализа надежности элементов с целью выявления доминирующих причин отказов.

В качестве недостатка метода Розенблатта – Парзена, например, в диссертации О.А. Дегтяревой указывается, что он, наряду с гистограммными методами, не позволяет представлять информацию в компактном виде, поскольку требует постоянного хранения всего массива исходных данных [39]. Следует отметить, что данный недостаток обусловлен адаптивностью, так как оценка будет изменяться в соответствии с потоком вновь поступающей информации об отказах без перестройки метода и применяемых зависимостей. Также немаловажно, что хранение полного объема данных позволяет в дальнейшем получать статистику для групп, отобранных по какому-нибудь признаку. Например, если требуется оценить надежность, эксплуатирующихся на разных объектах.

Для снижения размерности и упрощения вычислений с ростом объема полных наработок данные могут быть сгруппированы. Временная шкала разбивается на интервалы, и по имеющимся полным наработкам для каждого интервала рассчитываются частоты. Однако после проведения процедуры группировки данных следует осуществить пересчет оптимального значения параметра локальности.

2.4 Оценка точности адаптированного метода Розенблатта – Парзена

В целях экспериментальной проверки адаптированного метода Розенблатта – Парзена была проведена серия экспериментов (приложение Б) с моделированием наработок по заданным законам распределения.

На смоделированных данных были проведены исследования точности оценок. Выводы о точности оценок основывались на вычислении ошибок оценивания e_n по выражению:

$$e_n = \int_0^{\infty} |f_{\text{теор.}}(t) - f_{\text{вост.}}(t)| dt,$$

где $f_{\text{теор.}}(t)$ – плотность закона распределения времени до отказа, по которому осуществлялось моделирование наработок; $f_{\text{вост.}}(t)$ – восстановленная оценка плотности распределения адаптированным методом Розенблатта – Парзена.

Моделируемые законы распределения и их параметры приведены в табл. 2.9. Для удобства моделирования и сравнимости результатов параметры подбирались таким образом, чтобы математическое ожидание было около 500.

Таблица 2.9 – Моделируемые законы распределения времени до отказа

Моделируемый закон плотности распределения	Параметры	Значение МО
Вейбулла: $f(t) = \frac{m}{\theta} \cdot t^{m-1} \cdot e^{-\frac{t^m}{\theta}}$.	$m = 1,1,$ $\theta = 10^3$	515
Гамма-распределение: $f(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \cdot t^{\alpha-1} \cdot e^{-\lambda t}$.	$\alpha = 1,1,$ $\lambda = 0,0022$	500
Логнормальное: $f(t) = \frac{1}{t\sqrt{2\pi\sigma}} \cdot e^{-\frac{[\ln(\frac{t}{m})]^2}{2\sigma^2}}$.	$m = 3,6 \cdot 10^2$ $\sigma = 0,8$	496

Для каждого из законов распределения было сгенерировано 20 серий из 5 выборок по 10 случайных величин. Выборки в сериях имитировали источники

данных. Для них были установлены границы цензурирования справа: 200, 300, 400, 500, 600. Сгенерированные случайные величины, вышедшие за границы цензурирования справа, отбрасывались. В табл. 2.10 приводится процент отбрасываемых наработок по каждому из источников для моделируемых законов распределения.

Таблица 2.10 – Процент отбрасываемых наработок в экспериментах по каждому из источников

Объем моделируемых выборок	10	10	10	10	10
Значение границы цензурирования справа	200	300	400	500	600
Процент цензур. справа наработок:					
Вейбулла –	71%	59%	48%	39%	32%
Гамма-распределение –	69%	56%	46%	38%	30%
Логнормальное –	77%	59%	45%	34%	26%

Общее количество цензурированных справа наработок по всей совокупности источников составило:

для распределения Вейбулла – 50%;

для гамма-распределения – 48%;

для логнормального распределения – 48%.

В табл. 2.11 приведены усредненные ошибки оценивания e_n плотности распределения, полученные методом максимума правдоподобия и методом моментов, а также для цензурированных оценок по вариантам 1 и 2. Значения ошибок оценивания усреднены по всем 20 экспериментальным сериям случайных генераций выборок. Оценки методами максимума правдоподобия и моментов проводились для полных наработок, объединенных по источникам выборок (т. е. для 50 полных наработок в каждом эксперименте).

Таблица 2.11 – Усредненные ошибки оценивания для плотностей распределения времени до отказа

Метод	Вейбулла	Гамма	Логнормальное
Метод максимума правдоподобия (в отсутствии цензур.)	0,137	0,126	0,127
Метод моментов (в отсутствии цензур.)	0,141	0,18	0,128
Адаптированный метод Розенблатта – Парзена	0,219	0,248	0,334

Максимальные значения ошибок оценивания e_n , полученные при моделировании каждого из законов, приведены табл. 2.12.

Таблица 2.12 – Максимальные ошибки оценивания для плотностей распределения времени до отказа

Метод	Вейбулла	Гамма	Логнормальное
Метод максимума правдоподобия (в отсутствии цензур.)	0,275	0,233	0,319
Метод моментов (в отсутствии цензур.)	0,294	0,428	0,315
Адаптированный метод Розенблатта – Парзена	0,364	0,391	0,507

Как видно из табл. 2.11 и табл. 2.12, даже при существенной доле цензурированных справа данных, средние значения отклонения оценок вполне сравнимы с оценками по методам максимума правдоподобия и моментов, полученными с задействованием всех данных выборки, т. е. в условиях отсутствия цензурирования справа. Существенно более высокие погрешности оценок для логнормального распределения объясняются наличием у данного типа распределения «тяжелого хвоста». Функция плотности долго асимптотически приближается к оси абсцисс, что при выбранном способе аппроксимации в условиях неизвестности поведения функции за пределами цензурирования справа внесло значительный вклад в величину погрешности.

2.5 Выводы по главе

Разработана методика оценки эксплуатационной надежности вычислительных систем по малым выборкам с учетом наличия цензурированных данных. В рамках методики предложена адаптация непараметрического метода Розенблатта – Парзена для построения плотности распределения отказов. Адаптация метода заключается в учете цензурированных данных и компенсации смещения плотности распределения отказов составных элементов. Получено аналитическое представление функция плотности распределения отказов, учитывающая полные наработки на отказ, цензурированные интервалом и цензурированы справа данные.

Разработана технология сбора и подготовки статистического набора данных пригодного для последующей оценки эксплуатационной надежности вычислительных систем. В технологии впервые внедрены модели для распознавания отказов в потоке данных, автоматизирован процесс сбора и подготовки данных от различного типа элементов вычислительного комплекса.

Разработаны модели машинного обучения для распознавания отказов. В разработанных моделях учитываются показатели функционирования программных и аппаратных элементов одновременно. Модели позволяют определять отказы составных элементов вычислительных систем в процессе их функционирования, тем самым формируя эксплуатационные данные.

ГЛАВА 3. Разработка программного обеспечения для оценки надежности вычислительных систем. Экспериментальная проверка методики

Третья глава посвящена разработке программного обеспечения для оценки эксплуатационной надежности вычислительных систем (далее – ПО «Модуль СМ») и экспериментальной проверки разработанной методики.

В первом параграфе приведена постановка задачи для разработки ПО «Модуль СМ» с учетом предъявляемых функциональных требований.

Во втором параграфе описан процесс разработки. Новое программное решение выделяется на фоне существующих систем использованием открытых стандартизированных протоколов обмена данными, обеспечивает собственную масштабируемость в реальном времени. Разработанная трехуровневая архитектура позволяет проводить модернизацию и масштабирование компонентов независимо, а также предоставляет возможность создания децентрализованного хранилища и анализа надежности удаленных систем.

Разработанное программное обеспечение позволяет автоматизировать процесс подготовки данных, поступающих из эксплуатации и на основе полученных данных проводить оценку показателей надежности вычислительной системы. Осуществлять контроль за функционированием вычислительной системы и ее составных частей, посредством информирования обслуживающего персонала.

В третьем параграфе проведена экспериментальная проверка разработанной во второй главе методики и ПО «Модуль СМ». В качестве исследуемого объекта представлена вычислительная система, состоящая из 20 однотипных вычислительных узлов Supermicro Server-1041M-T2, введенных в эксплуатацию в три этапа в период с 2011 по 2014 г. В рамках эксперимента, на основе исторических данных получены оценки плотности и функции распределения

наработок времени до отказа, ВБР, интенсивности отказов, плотности распределения с учетом цензурированной информации.

В приложении А представлено «Свидетельство о регистрации программы для ЭВМ в Едином реестре российских программ для электронных вычислительных машин и баз данных № 2020661451».

3.1 Постановка задачи

Потребность в разработке ПО «Модуль СМ» обоснована функциональными требованиями к применению разработанной методики. В частности, разработанное ПО «Модуль СМ» должно обеспечивать:

- хранение значений контролируемых параметров в базе данных в течение всего срока службы или до принудительной очистки базы данных;
- регистрацию управляющих команд от пользователей и выполнение соответствующих действий по управлению отдельными техническими средствами;
- корректное завершение работы и выключение при получении аварийного сигнала от подсистемы бесперебойного;
- визуальный контроль за состоянием технических средств через веб-интерфейс в виде таблиц и графиков;
- информирование обслуживающего персонала о возникновении любой нештатной ситуации при помощи главной панели, а также светозвукового сигнала, установленного в помещении для работы операторов;
- учитывать возможность масштабирования и модифицирования ВС, не оказывая влияния на общую работоспособность ПО «Модуль СМ»;
- развертывание ПО из предоставляемых исходных пакетов, стандартными средствами операционной системы;
- возможность архивирования собранных значений от контролируемых параметров с изменяемым интервалом времени.

3.2 Разработка программного обеспечения «Модуль СМ» для оценки надежности

Анализ применимости нереляционных баз данных для хранения значений контролируемых элементов. Разработанная во второй главе архитектура сбора данных, позволят гибко настраивать ПО «Модуль СМ», например, собирать значения различными коллекторами, хранить значения в децентрализованных базах данных и визуализировать полученные данные в одной оболочке. В процессе сравнения современных БД для реализации хранилища метрик в ЦСОД рассмотрены решения, представленные в табл. 3.1:

Таблица 3.1 - Сравнение характеристик БД

Показатели	InfluxDB	VictoriaMetrics	TimescaleDB
Тип измерения	метрики, события	метрика	метрики, события
Используемые БД	реляционные/ нереляционные	реляционные/ нереляционные	реляционные
Типы данных	int64, float64, bool, string	float64	int64, float64, bool, string
Скорость обработки запросов	Быстрая постоянная	Быстрая постоянная	Средняя снижающаяся по мере наполнения БД
Использование модулей	+	+	-
Объем памяти на 10 000 событий	> 10 мб	> 10 мб	~ 100 мб

Из табл. 3.1 видно, что системы построенные на нереляционных БД такие как InfluxDB [77], VictoriaMetrics [76] обладают преимуществом перед TimescaleDB [78] по скорости обработки запросов и объему расходуемой памяти. Данные характеристики определяют дальнейший выбор в нереляционных БД.

Требования к аппаратной части для реализации ПО «Модуль СМ». В качестве контролируемых устройств и источников данных (первый уровень) предлагается использовать вычислительный узел, сетевой коммутатор, служебные

базы данных. Данный выбор обусловлен необходимостью отобразить возможность управления и контроля как аппаратных, так и программных СЭ ВС.

Трехуровневая архитектура позволяет реализовать второй и третий уровень ПО «Модуль СМ» на одном сервере. В качестве аппаратной части использован физический сервер с характеристиками, представленными в табл. 3.2.

Таблица 3.2 – Основные технические характеристики сервера мониторинга

Техническая характеристика	Значение
Тип процессора	Intel i7 - 8700
Объем оперативной памяти	16 Гб
Объем физической памяти	1 Тб
Модель платы мониторинга	Supermicro AOC-SIMSO
Дополнительный адаптер Gigabit Ethernet	PCI/PCI-X 10/100/1000Mbps
Интерфейс ввода-вывода сервисной сети	Порт Fast Ethernet (BMC)

Общими требованиями к аппаратной реализации сервера ПО «Модуль СМ» являются наличие сети Ethernet, пропускной способностью не менее 1 Гб/с для сервисной подсети, а также сервер наличие бесперебойного источника питания, обеспечивающего его работу на время аварийного выключения комплексов.

Требования к программной части для реализации ПО «Модуль СМ». Программная реализация макета ПО «Модуль СМ» должна соответствовать общим требованиям, таким как поддержка кроссплатформенности контролируемых устройств и источников данных, развертывание из предоставляемых исходных пакетов с использованием стандартных средств операционной системы, возможность миграции на другой сервер и хранение данных от контролируемых параметров с гибкими интервалами времени.

3.3 Интерфейс управления программным обеспечением «Модуль СМ»

Веб-приложение представляет конечный элемент взаимодействия макета «Модуль СМ» и пользователей системы, позволяя реализовать функциональное требование управления и контроля показателей надежности ВС в режиме реального времени. Элементы взаимодействия макета «Модуль СМ» и пользователей системы, описание визуального наполнения и функциональных особенностей представлены в табл. 3.3.

Таблица 3.3 – Управление и контроль ПО «Модуль СМ»

Наименование страницы: Функциональные особенности	Реакция ПО «Модуль СМ»
<p><i>Авторизация:</i> Наличие полей для ввода логина/пароля пользователя. Авторизация пользователей в ПО «Модуль СМ» в соответствии с ранее заданными ролями. Пользователю с ролью «Гость» доступна возможность наблюдения за текущим состоянием контролируемых устройств и источников данных. Пользователю с ролью «Оператор» доступна дополнительная возможность управления контролируемыми устройствами и источниками данных. Пользователю с ролью «Администратор» дополнительно доступно конфигурирование наполнения модуля СМ.</p>	<p>Роль «Гость»: Доступ к странице «Главная» с функцией просмотра. Роль «Оператор»: Доступ к странице «Главная» с функцией управления и просмотра. Роль «Администратор»: Доступ к странице «Главная» с функцией управления, просмотра и конфигурирования ПО «Модуль СМ».</p>
<p><i>Контроль:</i> Панели — базовые элементы визуализации контролируемых устройств (графики, таблицы, единичные статусы и статус панели). Доступна конфигурация стилей и форматов каждой панели, динамическое перемещение и изменение размера; Рабочая область — набор отдельных панелей, размещенных в сетке с набором переменных (например, имя комплекса, приложения или системы). Изменение переменных, влияет на отображаемые данные в рабочей области (например, данные с двух отдельных серверов). Все рабочие области можно конфигурировать в соответствии с потребностями пользователя;</p>	<p>Роль «Гость»: Доступ к странице «Главная». Роль «Оператор»: Доступ к странице «Главная» с функцией перехода в блок управления. Роль «Администратор»: Доступ к странице «Главная» с функцией перехода в блок управления, конфигурирования ПО «Модуль СМ».</p>

Наименование страницы: Функциональные особенности	Реакция ПО «Модуль СМ»
Блок журнала сообщений (отображение изменения контролируемых параметров); Блок диспетчеризации задач (наблюдение за текущими расчетными задачами пользователей вычислительных комплексов); Переход в блок управления вычислительными комплексами; Переход в блок конфигурации ПО «Модуль СМ».	
<i>Управление:</i> Возможность включения, выключения, перезагрузки контролируемых устройств и источников данных);	Отправка управляющих команд на контролирурующие устройства и источники данных; Оповещение оператора через журнал сообщений о статусе выполнения команды;
<i>Конфигурирование:</i> Масштабирование системы мониторинга (изменение, добавление, удаление количества опрашиваемых устройств и источников данных) Администрирование системы мониторинга (изменение состава контролируемых параметров, настройку пороговых значений, аккаунтов пользователей).	Отправка управляющих команд на контролирующие устройства и источники данных; Оповещение оператора через журнал сообщений о статусе выполнения команды;

На рис. 3.1 представлена структура веб-интерфейса. Структура состоит из следующих страниц: главная страница, страница со статистикой отказов и событий, страница с информацией о состоянии выбранного устройства и диалоговое окно, располагающееся на главной странице системы мониторинга. Отдельным проектом представлена система администрирования системы мониторинга.

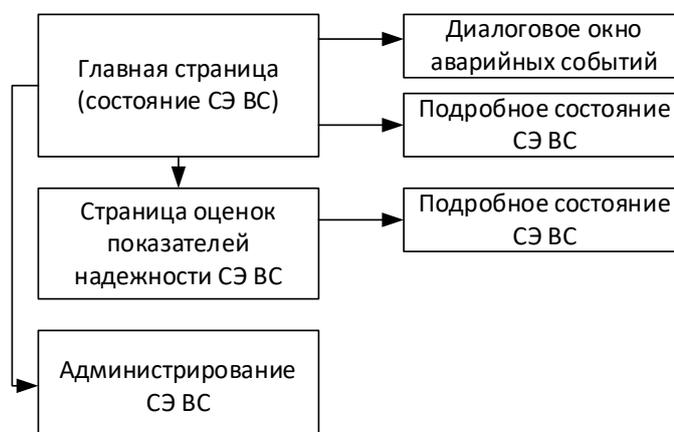


Рисунок 3.1 – Структура веб-интерфейса ПО «Модуль СМ»

Веб-интерфейс разработан с использованием языка *JavaScript*, библиотеки *jQuery*, а также *HTML*, *CSS* и *Bootstrap*. При использовании *JavaScript* был применен стандарт *ES6*. Для функционирования интерфейса в старых браузерах был использован пакет *Babel*, который производит трансляцию с нового стандарта языка в более старый (поддерживаемый старыми версиями браузеров).

Для автоматизации установки и обновления пакетов, как правило, применяются системы управления пакетами или менеджеры. Непосредственно в нашей задаче для этой цели используется пакетный менеджер *yarn* . Также, при написании кода был использован сборщик пакетов – *web pack* .

Обращение к серверной части проекта происходит посредством технологии *AJAX* библиотеки *jQuery*, которая позволяет получать данные с сервера без перезагрузки страницы.

Главная страница ПО «Модуль СМ». Графическое отображение элементов ВС располагается под их названиями и реализовано в виде таблицы. Элементы имеют цветовую кодировку, зависящую от текущего состояния оборудования (зеленый – нормальное состояние, желтый – некритическая неисправность, красный – критическая неисправность или отказ). Диалоговое окно расположено в правом нижнем углу страницы и отображает информацию о последних событиях (перехода параметров из нормальных в критические значения), произошедших в ПО «Модуль СМ». Внизу, в левой части страницы расположены функции отключения звука и светозвукового маячка, а также переход к модулю отказов и событий устройств и модулю администрирования системы. Интерфейс главной страницы представлен на рис. 3.2.

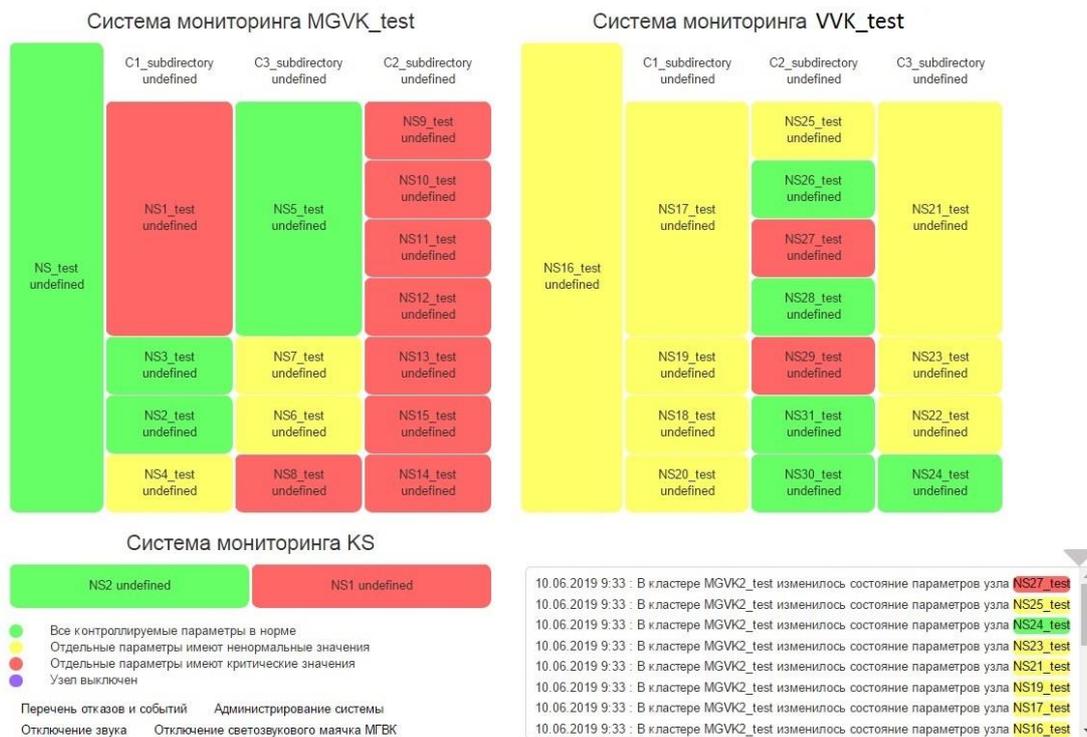


Рисунок 3.2 – Главная страница ПО «Модуль СМ»

Диалоговое окно аварийных событий. Внутри ПО «Модуль СМ» функция сравнивает новые и старые данные из БД в формате JSON, и на выходе возвращает их разницу. Далее происходит формирование событий с их выводом в веб-интерфейс. Диалоговое окно аварийных событий (рис. 3.3) разработано с целью динамического отображения состояния СЭ ВС и измененных показателей.

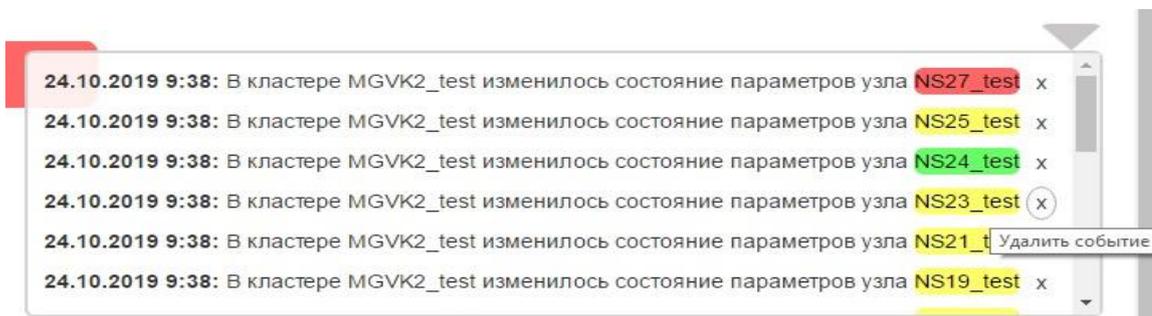


Рисунок 3.3 – Диалоговое окно аварийных событий

Страница с информацией о состоянии выбранного устройства. Переход к данной странице осуществляется при нажатии на ссылку с названием узла на главной странице ПО «Модуль СМ» или на ссылку с названием узла в диалоговом окне системы мониторинга. Окно представляет собой две колонки – контролируемый параметр и его значение (рис. 3.4). Если цвет на строке параметра желтый – значение параметра в пределах допустимых отклонений, красный – значение имеет критические отклонения.

Параметр	Значение
Доступность по внешней сети mgvk.local (1-доступно,0-недоступно)	1
Доступность по сети мониторинга vms-контроллера(0-недоступен,1-доступен)	1
Температура процессора 1	50
Состояние вкл/выкл	1
Напряжение батареи	null
Скорость вращения четвертого вентилятора	null

Рисунок 3.4 – Страница с информацией о состоянии выбранного устройства

При нажатии на ссылку перечня отказов и событий происходит перенаправление на страницу с перечнем отказов и событий устройств (рис. 3.5), где оператор может видеть список произошедших событий на определенных устройствах.

Устройство	Название события	Значение	Начало	Окончание
NS1	Аварийное напряжение на ядре CPU4	0.704	2018-08-22 12:24:02	-----
NS1	Перегрев CPU3	31	2018-05-29 11:43:36	-----
NS1	Перегрев CPU2	31	2018-05-29 11:21:22	-----
WV1	ВМС-контроллер модуля недоступен по сети мониторинга	0	2018-04-18 16:44:22	-----
NG	Модуль недоступен по внутренней сети Ethernet	0	2018-04-18 16:44:22	-----
NG	ВМС-контроллер модуля недоступен по сети мониторинга	0	2018-04-18 16:44:22	-----
NG	Устройство недоступно по внешней сети	0	2018-04-18 16:44:22	-----
NV21	Модуль недоступен по внутренней сети Ethernet	0	2018-04-18 16:44:22	-----
NV21	ВМС-контроллер модуля недоступен по сети мониторинга	0	2018-04-18 16:44:22	-----

Рисунок 3.5 – Страница с перечнем отказов и событий устройств

Страница администрирования. Страница администрирования ПО «Модуль СМ» написана с использованием языка программирования *Python*. Страница управления оборудованием представлена на рис. 3.6, где в левой части расположена информация о СЭ (названия узлов), и ссылка к СМ, а в правой части расположены функции добавления и редактирования оборудования.

Администрирование системы мониторинга

Контролируемое оборудование

- KS
 - KS1
 - NS11
 - NS12
 - KS2
 - NS21
 - NS22
- MGVK KVANT
 - Frame_1 Стойка №1
 - IN1
 - MN1
 - NV1
 - NV10
 - NV11
 - NV12
 - NV2
 - NV3
 - NV4
 - NV5
 - NV6
 - NV7
 - NV8
 - NV9
 - VVU
 - Frame_2 Стойка №2
 - DM1
 - KVM
 - MN2
 - NF1
 - NF2
 - NG
 - NI
 - NM
 - NU
 - SM
 - SU
 - Frame_3 Стойка №3
 - MN3

<< Система мониторинга

Добавление новых объектов в систему мониторинга

Добавить оборудование Добавить раздел системы мониторинга

Подробная информация о «MN1»

Имя:	MN1
Полное наименование:	нет данных
EXAMPLE:	нет данных
Описание:	нет данных
Контролирующий узел:	нет данных
Операционная система:	нет данных
Команда:	оп
Расположение узла:	10

Редактировать **Удалить**

Запросы контролируемых параметров IP-адреса Скопировать оборудование

Рисунок 3.6 – Страница администрирования системы мониторинга

3.4 Экспериментальная проверка методики оценки эксплуатационной надежности вычислительных систем

Постановка задачи. При эксплуатации ВС особое внимание уделяется надежности и безотказному функционированию. Системный анализ является ключевым этапом в обеспечении надежности и безотказного функционирования. Важными аспектами системного анализа надежности ВС являются технологический анализ истории эксплуатации объекта, количественный анализ безотказности и ремонтпригодности.

Количественный анализ позволяет оценить текущий уровень надежности и предоставляет необходимые данные для принятия управленческих решений по обеспечению безопасности, повышению эффективности, оптимизации стратегий технического обслуживания и ремонта, включая перепланирование циклов ремонтных работ, корректировку запасов материально-технических ресурсов и вероятностный анализ безопасности.

Результаты анализа надежности включают оценки плотности распределения, вероятности безотказной работы и интенсивности отказов, построенные с использованием разработанной методики по данным об отказах СЭ ВС. Полученные результаты могут быть использованы:

- при анализе нарушений и аварийных ситуаций при эксплуатации ВС;
- при анализе работоспособности оборудования;
- при разработке и тестировании новых программных элементов;
- при совершенствовании эксплуатационной документации;
- при разработке требований ко вновь внедряемому оборудованию;
- при разработке систем технической поддержки ВС.

Описание вычислительной системы как объекта исследования характеристик надежности. Апробация разработанного ПО «Модуль СМ» проведена на экспериментальной ВС. Исследуемая ВС предназначена для решения задач связанных с нейросетевым моделированием. Исходя из задачи были определены составные элементы, представленные в табл. 3.4 и архитектура ВС, представленная в общем виде на рис. 3.7. Программные элементы предустановлены на аппаратные элементы, в соответствии с условными обозначениями.

Таблица 3.4 – Основные аппаратные и программные элементы ВС

Обозначение	Описание СЭ	Кол-во
Аппаратные элементы		
1	Управляющий сервер Supermicro Server YS-1028GQ-TR 1U Server BB 4x PCI-E 3.0 x16 slots, 2x PCI-E 3.0 x8 (in x16) Комплектация: Модульный блок питания 2000W - 2 шт., Intel Xeon E5-2640 V4 10-core - 1 шт., KSM26RD8 16GB - 8 шт., HDD WD SAS 1Tb, 12 G/s – 2 шт.	1
2	Сервер мониторинга Supermicro Server YS-1028GQ-TR 1U Server BB 4x PCI-E 3.0 x16 slots, 2x PCI-E 3.0 x8 (in x16) Комплектация: Модульный блок питания 2000W - 2 шт., Intel Xeon E5-2640 V4 10-core - 1 шт., KSM26RD8 16GB - 2 шт., HDD WD SAS 1Tb, 12 G/s – 2 шт.	1
3	Сетевой коммутатор управляющей сети Cisco WS-C2960S-48LPS-L 48p – 1 шт.	1
4	Высокоскоростной коммутатор Mellanox SX6036G InfiniBand Switch 36 x 56Gb QSFP Ports	1
5	Сетевой коммутатор сети мониторинга Cisco WS-C2960S-48LPS-L 48p – 1 шт.	1
NV1-NV20	Вычислительный сервер Supermicro Super Server YS-1028GQ-TR 1U Server BB 4x PCI-E 3.0 x16 slots, 2x PCI-E 3.0 x8 (in x16) Комплектация: Модульный блок питания 2000W - 2 шт., Intel Xeon E5-2640 V4 10-core - 2 шт., KSM26RD8 16GB - 16 шт., HDD WD SAS 1Tb, 12 G/s – 2 шт., Tesla K40 GK110BGL – 3 шт.	20
Программные элементы		
NV1 - NV20	NVIDIA Driver for Linux x86_64 v. 340.58	20
NV1 - NV20	NVIDIA Driver CUDA Toolkit: 6.5	20
1	Система управления и постановки задач (СУПЗ)	1
2	Система мониторинга состояния оборудования Telegraph	1
2	ПО «Модуль СМ»	1
NV1- NV20, 1, 2	Ubuntu 12.10	22

На рис. 3.9 приняты следующие условные обозначения: 1 – управляющий сервер, 2 – сервер мониторинга/ПО «Модуль СМ», 3 – сетевой коммутатор управляющей сети, 4 – высокоскоростной коммутатор, 5 – коммутатор сети мониторинга, NV1 – NV20 – вычислительные сервера.

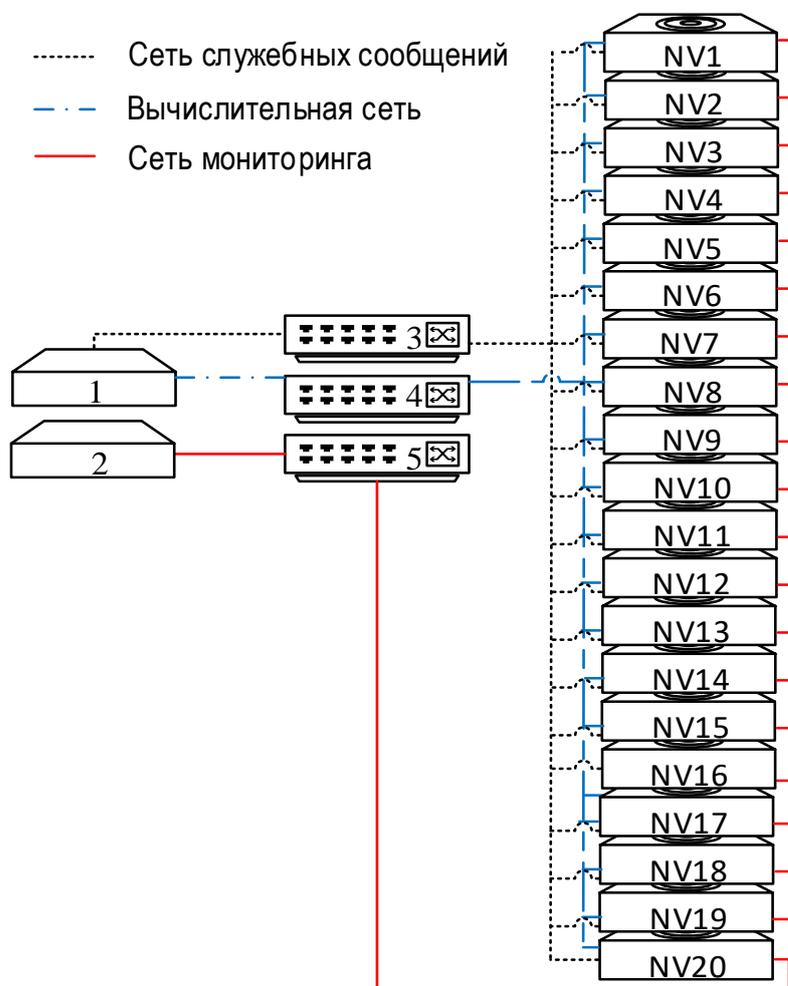


Рисунок 3.7 – схема взаимодействия элементов вычислительной системы и ПО «Модуль СМ»

Из журналов эксплуатации сформирована информация о 29 отказах, цензурированных интервалом. Частичная выборка из таких отказов представлена в табл. 3.5.

Таблица 3.5 – Частичная выборка об отказах СЭ ВС, цензурированная интервалом.

Дата фиксации	Событие	Комментарии
11.11.2018 14:48:00	Обрыв сети электропитания. Аварийное отключение ВС. Включение: 12.11.2015 02:20:00	СЭ NV17 – выход из строя БП.
07.05.2018 07:22:00	Плановое ТО. Выявлен не рабочий модуль RAID на сервере мониторинга.	СЭ 2 – RAID контроллер.
12.08.2019 11:34:00	Обрыв сети электропитания. Аварийное отключение ВС. Включение: 12.08.2016 12:28:00	СЭ NV11, NV12 - некорректный запуск драйвера GPU.

В качестве основного источника информации о полных наработках выступает ПО «Модуль СМ». Интервал формирования данных с 01.01.2017 по 31.12.2021. Сформированные данные об отказах представлены в табл. 3.6. При этом для СЭ с условными обозначениями «1 – 5» отказы не зафиксированы.

Таблица 3.6 – Информация об отказах СЭ ВС, сформированная ПО «Модуль СМ»

NV	$\vec{E} = (\varepsilon_1, \dots, \varepsilon_i)$	Σ
1	114, 4286, 4607, 13007, 21279, 26439, 29732, 30938, 31618, 36970, 37598	11
2	4172, 12137, 14002, 20353, 21290, 26766, 37555, 39935	8
3	560, 2515, 3377, 12120, 12946, 14002, 29113, 30452, 38810, 39935	10
4	204, 2368, 4109, 5310, 8987, 15072, 20361, 21871, 29366, 31928	10
5	2040, 5621, 10643, 12339, 21069, 26980, 38987	7
6	1883, 3146, 12309, 18016, 21275, 31901, 37080, 38987	8
7	475, 4581, 9498, 13333, 21172, 39726	6
8	861, 3183, 6589, 15331, 20514, 22400, 30779, 37223	8
9	4893, 9845, 12423, 26280	4
10	587, 17655, 19371, 24361, 32653, 41757	6
11	2506, 15770, 17623, 19799, 26378, 30778, 32450, 32450, 38068, 40938	10
12	98, 2328, 3106, 3375, 4062, 8860, 9556, 13569, 16198, 19933, 22347, 23980, 28412, 28916, 30196, 31034, 36350, 39880	18
13	485, 2373, 4536, 11292, 13054, 22275, 22846, 31689, 39011	9
14	3875, 3934, 8949, 10789, 12227, 15742, 21103, 22174, 22846, 30424, 36244, 38099, 40361	13
15	136, 2497, 5208, 9634, 10925, 15442, 21156, 22080, 22611, 31398, 29790, 38533	12
16	2778, 6212, 11444, 12686, 13849, 22103, 22766, 26394, 37825	9
17	1639, 2036, 3096, 10836, 12384, 13973, 20454, 21875, 25321, 26992, 29887, 40299	12
18	588, 4579, 12351, 24488, 29200, 39486	6
19	3915, 4851, 6020, 12700, 26408, 28391, 40998	7
20	272, 3688, 4976, 11031, 15250, 23029, 26255, 32421, 37876, 38380	10

где NV – порядковый номер элемента ($NV1, \dots, NV20$), $\vec{E} = (\varepsilon_1, \dots, \varepsilon_i)$ - массив наработок на отказ в часах, Σ - количество отказов.

Порядок проведения эксперимента. Количественный анализ надежности проводится по принципу последовательного обобщения: от индивидуальных показателей СЭ до показателей, усредненных по ВС.

Как известно, зная плотность распределения, можно переходить к определению основных характеристик надежности, таких как функция распределения, ВБР.

Исходя из этого, расчеты характеристик надежности ВС проводились в следующем порядке:

1. Расчет средних наработок между отказами.
2. Построение оценок плотности и функции распределения наработок между отказами в соответствии с методикой, разработанной во второй главе.
3. На основе полученных на предыдущем этапе функциональных характеристик рассчитывались ВБР и интенсивность отказа.

В эксперименте рассчитывается суммарный поток событий, предполагая, что отказы и повреждения событий и составных частей являются неразличимыми. Общие для всех расчетов параметры:

- интервал наблюдения: с 01.01.2017 по 31.12.2021;
- доверительный уровень: 0.95;
- единица времени: часы.

Результаты. При проведении расчетов, адаптированным методом Розенблатта – Парзена в качестве ядра использовалась функция Гаусса с зеркальным отображением исходных данных. По суммарному объему данных об отказах СЭ ВС были получены оценки для следующих показателей: плотности и функции распределения наработки до отказа, ВБР, плотности распределения с учетом цензурированных справа данных (рис. 3.8-3.10).

На рисунках приняты следующие обозначения \hat{f} – оценка плотности распределения наработки до отказа, \hat{f}_H – нижняя, \hat{f}_B – верхняя границы доверительного интервала для оценки плотности, $\hat{f}_{\text{ценз}}$ – оценка плотности

распределения наработки между отказами с учетом цензурированной информации, \hat{F} — оценка функции распределения наработки до отказа, \hat{P} — оценка ВБР, $[\hat{F}_H, \hat{F}_B]$, $[\hat{P}_H, \hat{P}_B]$ — доверительные границы для функции распределения наработки до отказа и ВБР, соответственно.

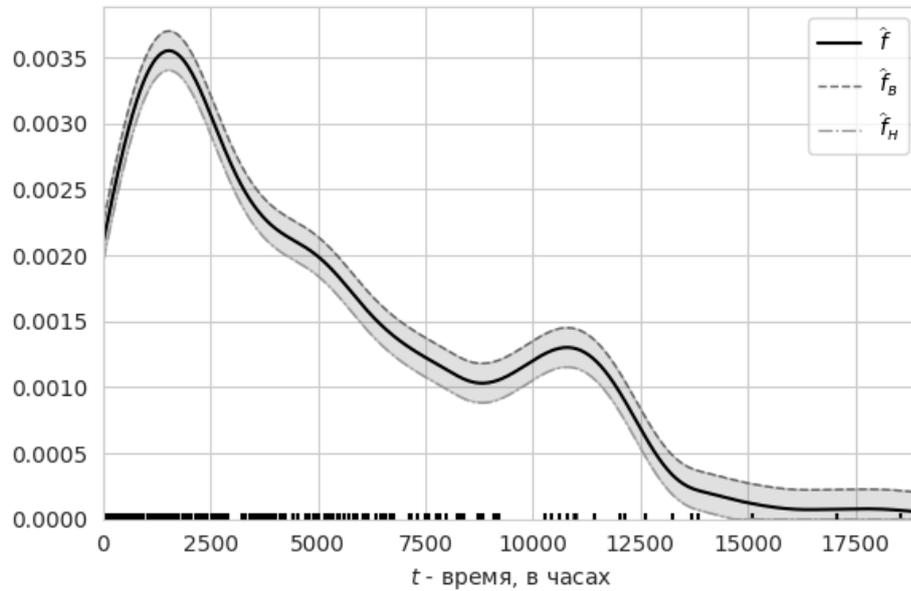


Рисунок 3.8 – оценка плотности распределения наработок между отказами с доверительными интервалами, построенная по полным наработкам $n=p=184$

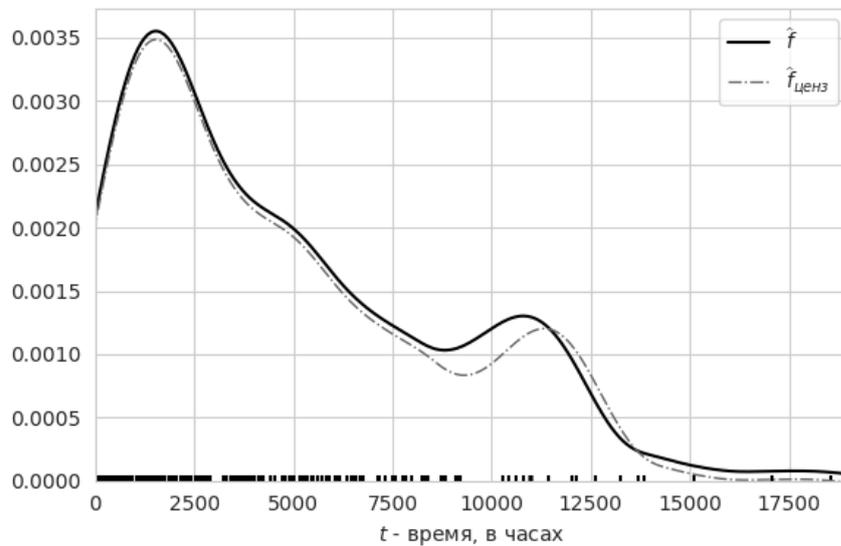


Рисунок 3.9 – оценка плотности распределения наработок между отказами с учетом и без учета цензурированной информации $n=213, p=187, q=29$

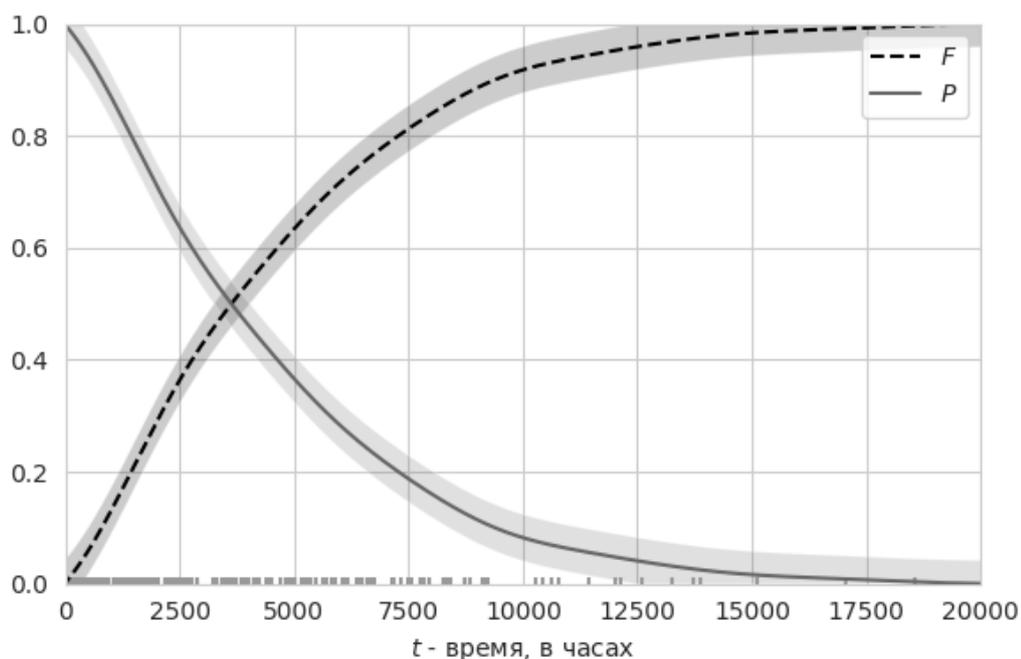


Рисунок 3.10 – оценка функции распределения наработок между отказами (F) и вероятности безотказной работы (P) с доверительными границами по полным наработкам и цензурированным данным $n=213$, $p=187$, $q=29$

В результате анализа плотности распределения наработок между отказами (рис. 3.8) выявлено, что помимо статистического шума каждый пик (горб) функции может быть обусловлен объективными причинами отказов. В рассматриваемом случае функция по оси времени имеет три пика.

Последующий анализ исходных данных (таблиц отказов) позволил определить, что наблюдаемое увеличение количества отказов на исследуемом интервале времени в каждом случае имеет доминирующую причину отказов.

Первый пик наблюдается при наработках в интервале от 1.5 до 2 тысяч часов, характеризующий более 70% отказов, связан с выходом из строя СЭ «NVIDIA Driver for Linux x86_64 v. 340.58» (драйвера). Более детальный анализ опыта эксплуатации позволил выявить недостаток запуска процедуры обучения нейросетевых моделей. В частности, при проведении одной из базовых процедур (прерывание обучения) происходило отключение драйвера, требующий перезапуск вычислительного сервера.

Второй пик, в интервале от 3 до 6 тысяч часов, связан с недостатками технического обслуживания и ремонта. А именно, не регламентирована процедура обновления программных элементов ВС на более новые версии, что вызывает проблему несоответствия зависимостей между СЭ ВС и запускаемыми задачами.

Третий пик, в интервале от 10 до 12 тысяч часов, связан с отказами отдельных СЭ ВС в период неплановых аварийных отключений.

В качестве рекомендаций по снижению числа отказов предлагается:

1. Ввести в эксплуатацию программный модуль контейнеризации расчетных задач. Это позволит создавать требуемое программное окружение для расчетных задач, изолировать воздействие друг на друга и на ВС в целом. Такой подход приведет к явному снижению отказов первого и второго пика.

2. Для уменьшения третьего пика предлагается разработать систему планомерной остановки вычислительных серверов в случае аварийного отключения ВС. Включить в план ТО проверку работоспособности системы аварийного отключения/включения ВС.

3.5 Выводы по главе

1. На основании сформированных требований разработано ПО «Модуль СМ» формирующее статистику по отказам СЭ ВС. Преимущество ПО «Модуль СМ» заключается в возможности модульного расширения СЭ по сравнению с коммерческими аналогами: для добавления нового элемента в подконтрольную систему или создания триггера к новому элементу достаточно указать интерфейс обмена данными, определенный в панели управления. Контролирующая система начинает производить опрос состояния вновь подписанного элемента с момента следующего опроса, не прекращая функционирования.

2. Экспериментальная проверка методики оценки эксплуатационной надежности вычислительных систем позволила выявить интервалы времени, на

которых наблюдается увеличение количества отказов. Более глубокий анализ исходной информации показал, что в ряде данных временных интервалов имеют место доминирующие причины отказов, а именно:

- в интервале 1.5 – 2 тысяч часов отказы связаны с выходом из строя СЭ «NVIDIA Driver for Linux x86_64 v. 340.58»;

- в интервале 3 – 6 тысяч часов отказы связаны с недостатками технического обслуживания и ремонта;

- в интервале 10 – 12 тысяч часов отказы СЭ ВС возникали в период неплановых аварийных отключений.

3. Предложены рекомендации для снижения числа возникающих отказов. В качестве рекомендаций предлагается:

- Ввести в эксплуатацию программный модуль контейнеризации расчетных задач. Это позволит создавать требуемое программное окружение для расчетных задач, изолировать воздействие друг на друга и на ВС в целом. Такой подход приведет к снижению отказов первого и второго пика;

- Для уменьшения третьего пика предлагается разработать систему плановой остановки вычислительных серверов в случае аварийного отключения ВС. Включить в план ТО проверку работоспособности системы аварийного отключения/включения ВС.

ЗАКЛЮЧЕНИЕ

Разработана методика оценки эксплуатационной надежности вычислительных систем по малым выборкам с учетом наличия цензурированных данных. В рамках методики предложена адаптация непараметрического метода Розенблатта – Парзена для построения плотности распределения отказов. Адаптация метода заключается в учете цензурированных данных и компенсации смещения плотности распределения отказов составных элементов. Получено аналитическое представление функции плотности распределения отказов, учитывающее полные наработки на отказ, наработки цензурированные интервалом и наработки цензурированы справа.

Разработана технология сбора и подготовки статистического набора эксплуатационных данных пригодного для формирования набора отказов по каждому элементу вычислительной системы. В технологии впервые внедрены модели машинного обучения для распознавания отказов в потоке данных, а также автоматизирован процесс сбора и подготовки данных от различного типа элементов вычислительного комплекса.

Разработаны модели машинного обучения для распознавания отказов. В разработанных моделях учитываются показатели функционирования программных и аппаратных элементов одновременно. Модели позволяют определять отказы составных элементов вычислительных систем в процессе их функционирования, тем самым формируя эксплуатационные данные.

Разработан программный комплекс для статистической оценки эксплуатационной надежности вычислительных систем, включающий реализацию технологии сбора и подготовки данных. Его достоинством является возможность получения оценок в режиме реального времени, на основании которых обслуживающий персонал получает возможность своевременного принятия решений для корректировки режима эксплуатации и функционирования ВС.

Список литературы

1. ГОСТ 27.002. Надежность в технике. Основные понятия. Термины и определения: утв. постановлением Госстандарта СССР от 15 ноября 1989 г.
2. ГОСТ 27.310. Надежность в технике. Анализ видов, последствий и критичности отказов: утв. Межгосударственным Советом по стандартизации, метрологии и сертификации от 26 апреля 1995 г.
3. ГОСТ 51901.5. Руководство по применению методов анализа надежности: утв. приказом Федерального агентства по техническому регулированию и метрологии от 30 сентября 2005 г.
4. РД 50-690-89. Методические указания. Надежность в технике. Методы оценки показателей надежности по экспериментальным данным: утв. Государственным комитетом СССР по управлению качеством продукции и стандартам от 01 января 1991 г.
5. Александровская Л.Н. Современные методы обеспечения безотказности сложных технических систем: учеб. пособие / Л.Н. Александровская, А.П. Афанасьев, А.А. Лисов. – М.: Логос, 2001. – 208 с.
6. Андреев А.В. Теоретические основы надежности технических систем: учеб. пособие / А.В. Андреев, В.В. Яковлев, Т.Ю. Короткая. – СПб.: Изд-во Политехн. ун-та, 2018. – 164 с.
7. Антонов А.В. Статистические модели в теории надежности: учеб. пособие / А.В. Антонов, М.С. Никулин. – М.: Абрис, 2012. – 390 с.
8. Богданов Ю.И. Основная задача статистического анализа данных: корневой подход / Ю.И. Богданов. – М.: МИЭТ, 2002. – 96 с.
9. Бостанджиян В.А. Пособие по статистическим распределениям / В.А. Бостанджиян. – Черноголовка: Редакционно-издательский отдел ИПХФ РАН, 2013. – 1060 с.

10. Вапник В.Н. Восстановление зависимостей по эмпирическим данным / В.Н. Вапник. – М.: Наука, 1979. – 448 с.
11. Вигерс К. Разработка требований к программному обеспечению / К. Вигерс, Д. Битти. – 3. – М.: «Русская редакция», 2014. – 736 с.
12. Викторова В.С. Анализ надежности отказоустойчивых вычислительных систем / В.С. Викторова, Н.В. Лубков, А.С. Степанянц. – М.: ИПУ РАН, 2016. – 117 с.
13. Викторова В.С. Модели и методы расчета надежности технических систем / В.С. Викторова, А.С. Степанянц. – 2, испр. – М.: Издательская группа URSS, ООО «ЛЕНАНД», 2016. – 256 с.
14. Воеводин В.П. Эволюция понятия и показателей надежности вычислительных систем / В.П. Воеводин. – Протвино: Препринт ИФВЭ, 2012. – 24 с.
15. Вьюгин В.В. Математические основы машинного обучения и прогнозирования / В.В. Вьюгин. – М.: МЦНМО, 2022. – 399 с.
16. Гаскаров Д.В. Малая выборка / Д.В. Гаскаров, Шаповалов В.И. – М.: Статистика, 1978. – 248 с.
17. Деврой Л. Непараметрическое оценивание плотности. L1-подход / Л. Деврой, Л. Дьерфи. – М.: Мир, 1988. – 408 с.
18. Карякин М.И. Технологии программирования и компьютерный практикум на языке Python / М.И. Карякин, К.А. Ватульян, Р.М. Мнухин. – Ростов-на-Дону: Южный федеральный университет, 2022. – 242 с.
19. Кокс Д.Р. Анализ данных типа времени жизни / Д.Р. Кокс, Д. Оукс. Перевод: О. Селезнев. – М.: Финансы и статистика, 1988. – 192 с.
20. Колмогоров А.Н. Элементы теории функций и функционального анализа / А.Н. Колмогоров, С.В. Фомин. – М.: Наука, 1976. – 543 с.
21. Литтл Р. Статистический анализ данных с пропусками / Р. Литтл, Д. Рубин. – М.: Финансы и статистика, 1991. – 336 с.

22. Матвеевский В.Р. Надежность технических систем: учеб. пособие / В.Р. Матвеевский. – М.: Московский государственный институт электроники и математики, 2002. – 113 с.
23. Пампел Ф. Логистическая регрессия / Ф. Пампел, А. Груздев, Д. Цвиркун. – М.: ДМК ПРЕСС, 2022. – 218 с.
24. Половко А.М. Основы теории надежности / А.М. Половко, С.В. Гуров. – 2-е изд., прераб. и доп. – СПб.: БХВ-Петербург, 2006. – 704 с.
25. Порутчиков М.А. Анализ данных: учеб. пособие / М.А. Порутчиков. – Самара: Издательство Самарского университета, 2016. – 29 с.
26. Приемы объектно-ориентированного проектирования. Паттерны проектирования / Э. Гамма [и др.]. – СПб.: Питер, 2017. – 366 с.
27. Прохоров С.А. Аппроксимативный анализ случайных процессов / С.А. Прохоров. – Самара: Самарский государственный аэрокосмический университет, 2001. – 329 с.
28. Прудников А.П. Интегралы и ряды. Специальные функции / А.П. Прудников, О.И. Брычков, О.И. Маричев. – М.: Наука, 1983. – 753 с.
29. Родзин С.И. Отказоустойчивые вычислительные системы / С.И. Родзин. – Таганрог: ТРТУ, 2021. – 274 с.
30. Рябинин И.А. Надежность и безопасность структурно сложных систем / И.А. Рябинин. – СПб.: Политехника, 2000. – 248 с.
31. Скрипник В.М. Анализ надежности технических систем по цензурированным выборкам / В.М. Скрипник. – М.: Радио и связь, 1988. – 184 с.
32. Тихонов В.И. Марковские процессы / В.И. Тихонов, М.А. Миронов. – М.: Сов. радио, 1977. – 488 с.
33. Чубукова И.А. Data Mining: учеб. пособие / И.А. Чубукова. – 2. – М.: БИНОМ. Лаборатория знаний, 2008.
34. Шубинский И.Б. Функциональная надежность информационных систем. Методы анализа / И.Б. Шубинский. – Ульяновск: Надежность, 2012. – 295 с.

35. Эфрон Б. Нетрадиционные методы многомерного статистического анализа / Б. Эфрон. – М.: Финансы и статистика, 1988. – 261 с.
36. Principles of Data Wrangling: Practical Techniques for Data Preparation / J. Hellerstein, C. Carreras, T. Rattenbury [et al.]. – California: O'Reilly Media, 2017. – 82 p.
37. Антонов А.В. Статистический анализ эксплуатационной надежности электронасосных агрегатов ЦН 60-180 реакторов ВВЭР-1000 методами ядерного оценивания / А.В. Антонов // Известия вузов. Ядерная энергетика. – 2009. – № 3. – С. 5-14.
38. Базылев В.В. Сравнение возможностей логистической регрессии и искусственных нейронных сетей в прогнозировании результатов исследования на малой выборке / В.В. Базылев, В.А. Карнахин // Health, Food & Biotechnology. – 2019. – Т. 1. – № 3. – С. 11-20.
39. Береснева Н.М. Методология подготовки данных для вычислительных экспериментов в исследованиях энергетической безопасности России / Н.М. Береснева, О.В. Курганская // Вестник Иркутского государственного технического университета. – 2017. – Т. 21. – № 9. – С. 45-47.
40. Боровиков С.М. Оценка ожидаемой надежности прикладных программных средств для компьютерных информационных систем / С.М. Боровиков // Информатика. – 2021. – Т. 18. – № 1. – С. 84-95.
41. Викторова В.С. Динамические деревья отказов / В.С. Викторова, А.С. Степанянц // Надежность. – 2011. – № 3. – С. 20-32.
42. Вожов С.С. Сравнительный анализ алгоритмов построения непараметрической оценки функции распределения по интервальным данным / С.С. Вожов, Е.В. Чимитова // Обработка информации и математическое моделирование Рос. науч.-техн. конф. – Новосибирск: НГТУ, 2015. – С. 242-246.

43. Галимов Р.Г. Основы алгоритмов машинного обучения - обучение без учителя / Р.Г. Галимов // Научно-практический электронный журнал «Аллея Науки». – 2017. – № 14. – С. 807-809.

44. Дегтерева О.А. Программный комплекс аппроксимативного анализа законов распределения случайных процессов ортогональными функциями : дис. канд. тех. наук : 05.13.18 / О.А. Дегтерева. – Самара: Самарский государственный аэрокосмический университет, 2006. – 134 с.

45. Захаров Д.Н. Анализ методов статистической оценки эксплуатационной надежности вычислительных комплексов / Д.Н. Захаров, В.С. Никулин // Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. Информатика. Телекоммуникации. Управление. – 2020. – Т. 12. – № 1. – С. 64-69.

46. Зюляева Н.Г. Определение характеристик надежности оборудования АЭС непараметрическими методами : дис. канд. тех. наук : 05.13.01 / Н.Г. Зюляева. – Обнинск: Обнинский государственный технический университет атомной энергетики, 2009. – 177 с.

47. Карлов И.А. Восстановление пропущенных данных при численном моделировании сложных динамических систем / И.А. Карлов // Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. Информатика. Телекоммуникации. Управление. – 2013. – № 186. – С. 137-144.

48. Карлов И.А. Методы восстановления пропущенных значений с использованием инструментария Data Mining / И.А. Карлов // Вестник Сибирского гос. аэрокосмического ун-та им. академика М.Ф. Решетнева. – 2011. – № 7(40). – С. 29-33.

49. Клименко Т.М. Обзор методов обнаружения распределенных атак типа «Отказ в обслуживании» на основе машинного и глубокого обучения / Т.М. Клименко, Р.Р. Акжигитов // International Journal of Open Information Technologies. – 2023. – № 6. – С. 21-27.

50. Козлов Д.С. Нейросетевые методы обнаружения отказов датчиков и приводов летательного аппарата / Д.С. Козлов, Ю.В. Тюменцев // Труды МАИ. – 2012. – № 52. – С. 2-14.
51. Кузовлев В.И. Метод выявления аномалий в исходных данных при построении прогнозной модели решающего дерева в системах поддержки принятия решений / В.И. Кузовлев // Наука и образование: науч. изд. МГТУ им. Н. Э. Баумана. – 2012. – № 9. – С. 16-20.
52. Лубков Н.В. Влияние характеристик контроля на показатели надежности систем / Н.В. Лубков, И.Б. Спиридонов, А.С. Степанянц // Труды МАИ. – 2016. – Т. 85. – С. 1-27.
53. Николаев Д.А. Параметрический метод обработки результатов наблюдений с учетом пропущенных данных / Д.А. Николаев // Надежность. – 2017. – Т. 17. – № 1. – С. 53-58.
54. Никулин В.С. Адаптация метода Розенблатта – Парзена для экспериментальной оценки надежности вычислительной системы / В.С. Никулин // Прикладная дискретная математика. Приложение. – 2021. – № 14. – С. 148-153.
55. Никулин В.С. Методика подготовки данных для интеллектуального анализа надежности вычислительных комплексов / В.С. Никулин // Вестник СибГУТИ. – 2020. – № 3 (51). – С. 26-37.
56. Никулин В.С. Методический подход к оценке эксплуатационной надежности вычислительного комплекса / В.С. Никулин // Наука. Технологии. Инновации. – 2022. – Т. 2. – С. 42-46.
57. Никулин В.С. Создание автоматизированной системы сбора сведений о качестве функционирования вычислительных комплексов / В.С. Никулин, А.С. Павлова // Сборник научных статей по материалам X Всероссийской научно-практической конференции. – 2017. – С. 541-545.
58. Никулин В.С. Аппроксимация плотности распределения времени работы между отказами вычислительной системы непараметрическим методом

Розенблатта – Парзена / В.С. Никулин, А.И. Пестунов // Вестник ВГТУ. – 2022. – Т. 18. – № 1. – С. 36-41.

59. Никулин В.С. Оценка характеристик эксплуатационной надежности вычислительных комплексов адаптированным методом Розенблатта – Парзена / В.С. Никулин, А.И. Пестунов // Тезисы XXI Всероссийской конференции молодых уче

ных по математическому моделированию и информационным технологиям. Тезисы докладов. – 2020. – С. 39.

60. Никулин В.С. Проектирование программного комплекса для статистической оценки эксплуатационной надежности вычислительных систем / В.С. Никулин // Математическое моделирование и информационные технологии. – 2020. – С. 51.

61. Никулин В.С. Реализация метода обработки статистических данных полученных в ходе эксплуатации вычислительных комплексов / В.С. Никулин // Сборник научных трудов XV-ой Международной научно-практической конференции. – 2020. – С. 214-216.

62. Никулин В.С. Сравнительный анализ СУБД для реализации подсистемы хранения событий мониторинга вычислительных комплексов / В.С. Никулин // Наука. Технологии. Инновации. Сборник научных трудов. – 2019. – Т. 2. – С. 46-49.

63. Павский В.А. Анализ функционирования вычислительных и сложных технологических систем : дис. д-ра тех. наук: 05.13.15 / В.А. Павский. – Новосибирск, 2007. – 316 с.

64. Рузанов П.А. Методика создания приложений для работы с базами данных с использованием высокоуровневого языка Python / П.А. Рузанов // Цифровые технологии и информационная безопасность бизнес-процессов. Сборник научных статей по итогам научно-практической конференции с международным участием. – 2022. – С. 272-277.

65. Стратонович Л.Р. Быстрота сходимости алгоритмов оценки плотности распределения вероятностей / Л.Р. Стратонович // Изв. АН СССР, сер. техн. киберн. – № 6. – С. 3-15.
66. Тарасов А.Г. Система мониторинга вычислительного кластера расширенной функциональности : дис. канд. тех. наук : 05.13.11 / А.Г. Тарасов. – Хабаровск: РАН ДО ВЦ, 2011. – 145 с.
67. Тарасов А.Г. Трехуровневая система мониторинга расширенной функциональности / А.Г. Тарасов // Параллельные вычислительные технологии. Челябинск. – 2008. – С. 464-469.
68. Успенский М.Б. Разработка и исследование методов и моделей обработки диагностической информации для обнаружения и локализации неисправностей в системах хранения данных : дис. канд. тех. наук : 05.13.01 / М.Б. Успенский. – ФГАОУ ВО «Санкт-Петербургский политехнический университет Петра Великого», 2020.
69. Филиппов Б.И. Определение характеристик надежности изготовленных образцов радиоэлектронных систем / Б.И. Филиппов, Ю.В. Замятина // Надежность. – 2017. – Т. 17. – № 1 (60). – С. 27-31.
70. Ченцов Н.Н. Оценка неизвестной плотности распределения по наблюдениям / Н.Н. Ченцов // Доклады АН СССР. – 1962. – Т. 147. – № 1. – С. 45-48.
71. Чепурко В.А. Об однородном методе обнаружения неоднородности потока отказов оборудования АЭС / В.А. Чепурко, С.В. Чепурко // Известия вузов. Ядерная энергетика. – 2012. – № 2. – С. 65-73.
72. Шмидт О.О. Обобщенная модель процесса восстановления в теории надежности использования информационных технологий : дис. канд. ф.-м. наук : 05.13.17 / О.О. Шмидт. – Красноярск: СФУ, 2008. – 125 с.

73. A Data pre-processing method to increase efficiency and accuracy in data mining / Artificial Intelligence in Medicine: Lecture Notes in Computer Science // A.R. Razavi, S. Miksch, J. Hunter [et al.] – Berlin, Heidelberg: Springer, 2005. – pp. 434-443.
74. Chapman P. CRISP-DM 1.0: Step-by-step data mining guide / P. Chapman, et al. – 2000. – Vol. 9. – no 13. – pp. 1-73.
75. Foster I. Globus Toolkit Version 4: Software for Service-Oriented Systems / I. Foster // Journal of Computer Science and Technology. – 2006. – Vol. 21. – no 4. – pp. 513-520.
76. Massie B. The ganglia distributed monitoring system: Design, implementation and experience / B. Massie, N. Chun, E. Culler // Parallel Computing. – 2004. – Vol. 7. – no 30. – pp. 817-840.
77. Multisensory data-driven health degradation monitoring of machining tools by generalized multiclass support vector machine / Y. Cheng [et al.] // IEEE Access. – 2019. – Vol. 7. – pp. 102-113.
78. Nikulin V. Adaptation of non-parametric Parzen-Rosenblatt window method for estimating operational reliability characteristics of computing systems / V. Nikulin, A. Pestunov // Proceedings of the International Conference on Applied Science and Emerging Technology (STECH-2020). 2024. AIP Conference Proceedings (ISSN 0094-243X). Vol. 2436. Melville.
79. Padgett W. Nonparametric density estimation from censored data / W. Padgett, D. McNichols // Commun. Statist. Theory and Meth. – 1984. – Vol. 13. – no 13. – pp. 1581-1611.
80. Parzen E. On estimation of a probability density function and mode / E. Parzen // Annals of Mathematical Statistics. – 1962. – no 33. – pp. 1065-1076.
81. Prokhorenkova L. Catboost: Unbiased boosting with categorical features / L. Prokhorenkova, G. Gusev, A. Vorobev // Advances in Neural Information Processing Systems. – 2018. – no 32. – pp. 6638-6648.

82. Reineke D. Estimation of hazard, density and survival functions for randomly censored data / D. Reineke, J. Crown // *Journal of Applied Statistics*. – 2004. – Vol. 31. – no 10. – pp. 1211-1225.
83. Rozenblatt M. Remark on some nonparametric estimates of a density function / M. Rozenblatt // *Annals of Mathematical Statistics*. – 1956. – no 27. – pp. 832-837.
84. Wang C. Imbalance-XGBoost: leveraging weighted and focal losses for binary label-imbalanced classification with XGBoost / C. Wang, C. Deng, S. Wang // *Pattern Recognition Letters*. – 2020. – no 136. – pp. 190-197.
85. Wei C. Application of data mining on the development of a disease distribution map of screened community residents of taipei county in taiwan / C. Wei, S. Su, M. Yang // *Journal of Medical Systems*. – 2012. – pp. 2021-2027.
86. Yusuf I. Reliability modeling and analysis of client-server using Gumbel–Hougaard family copula / I. Yusuf, A.L. Ismail, M.A. Lawan // *Life Cycle Reliab Saf Eng*. – 2021. – no 4. – pp. 235-248.
87. Zhang L. Adaptive kernel density-based anomaly detection for nonlinear systems / L. Zhang, J. Lin, R. Karim // *Knowledge-Based Systems*. – 2018. – no 139. – pp. 50-63.
88. Zhang S. Data preparation for data mining / S. Zhang, C. Zhang, Q. Yang // *Appl. Artif. Intell*. – 2003. – pp. 37-381.
89. АРБИТР - компьютерная программа [Электронный ресурс]. – URL: <http://szma.com/arbitr/> (дата обращения: 02.09.2022).
90. Модуль Relex Reliability Prediction [Электронный ресурс]. – URL: <https://pro-technologies.ru/components/windchill-prediction.pdf> (дата обращения: 22.08.2022).
91. Система АСОНИКА-К-РЭС [Электронный ресурс]. – URL: <https://asonika-k.ru> (дата обращения: 01.09.2022).

92. An open-source framework for real-time anomaly detection using [Электронный ресурс]. – URL: <https://github.com/MentatInnovations/datastream.io>.
93. Anomaly Detection and Monitoring Service [Электронный ресурс]. – URL: <https://anomaly.io/> (дата обращения: 02.12.2021).
94. Favoweb. Failure analysis for preventive action [Электронный ресурс]. – URL: <https://www.favoweb.com> (дата обращения: 01.09.2022).
95. Grafana: The open observability platform [Электронный ресурс]. – URL: <https://grafana.com/> (дата обращения: 14.03.2021).
96. Prometheus - Monitoring system & time series database [Электронный ресурс]. – URL: <https://prometheus.io/> (дата обращения: 14.04.2021).
97. Riskspectrum вероятностный анализ безопасности [Электронный ресурс]. – URL: <http://riskspectrum.ru/> (дата обращения: 20.08.2022).
98. VictoriaMetrics: Simple & Reliable Monitoring for Everyone [Электронный ресурс]. – URL: <https://victoriametrics.com/> (дата обращения: 03.10.2023).
99. InfluxDB Times Series Data Platform | InfluxData [Электронный ресурс]. – URL: <https://www.influxdata.com/> (дата обращения: 03.10.2023).
100. Timescale is PostgreSQL++ for time series and event data | Timescale [Электронный ресурс]. – URL: <https://www.timescale.com/> (дата обращения: 03.10.2023).
101. Wolski R. Developing a Dynamic Performance Information Infrastructure for Grid Systems [Электронный ресурс]. – URL: <http://dast.nlanr.net/GridForum/Perf-WG/white.PDF> (дата обращения: 02.01.2020).
102. Никулин В.С. Свидетельство о государственной регистрации программы для ЭВМ 2020661451. Программный комплекс для статистической оценки эксплуатационной надежности вычислительных систем / В.С. Никулин, А.И. Пестунов. – ФГБОУ ВО «НГУЭУ «НИНХ», 2020.

ПРИЛОЖЕНИЕ А. Свидетельство о государственной регистрации программы для ЭВМ

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2020661451

Программный комплекс для статистической оценки эксплуатационной надежности вычислительных систем

Правообладатель: **Федеральное государственное бюджетное образовательное учреждение высшего образования «Новосибирский государственный университет экономики и управления «НИНХ» (RU)**

Авторы: **Пестунов Андрей Игоревич (RU),
Никулин Владимир Сергеевич (RU)**

Заявка № **2020616499**

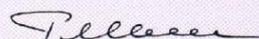
Дата поступления **23 июня 2020 г.**

Дата государственной регистрации

в Реестре программ для ЭВМ **24 сентября 2020 г.**



Руководитель Федеральной службы
по интеллектуальной собственности

 **Г.П. Ивлиев**

ПРИЛОЖЕНИЕ Б. Исследование свойств адаптированного метода Розенблатта – Парзена

Общие условия для всех экспериментов:

Генерация СВ, распределенных по Вейбулла с параметрами (масштаба $\theta = \beta^m = 1000$, формы $m = 1,1$); на графиках $\Sigma_{core}(t)$ – восстановленная адаптированным методом Розенблатта – Парзена, $\Sigma_{vb_T}(t)$ – теоретическая ФР, $\Sigma_{vb_E}(t)$ – восстановленная методом моментов.

Эксперимент 1 – Исследование масштаба параметра локальности. Полные наработки. С отбросом правой границы.

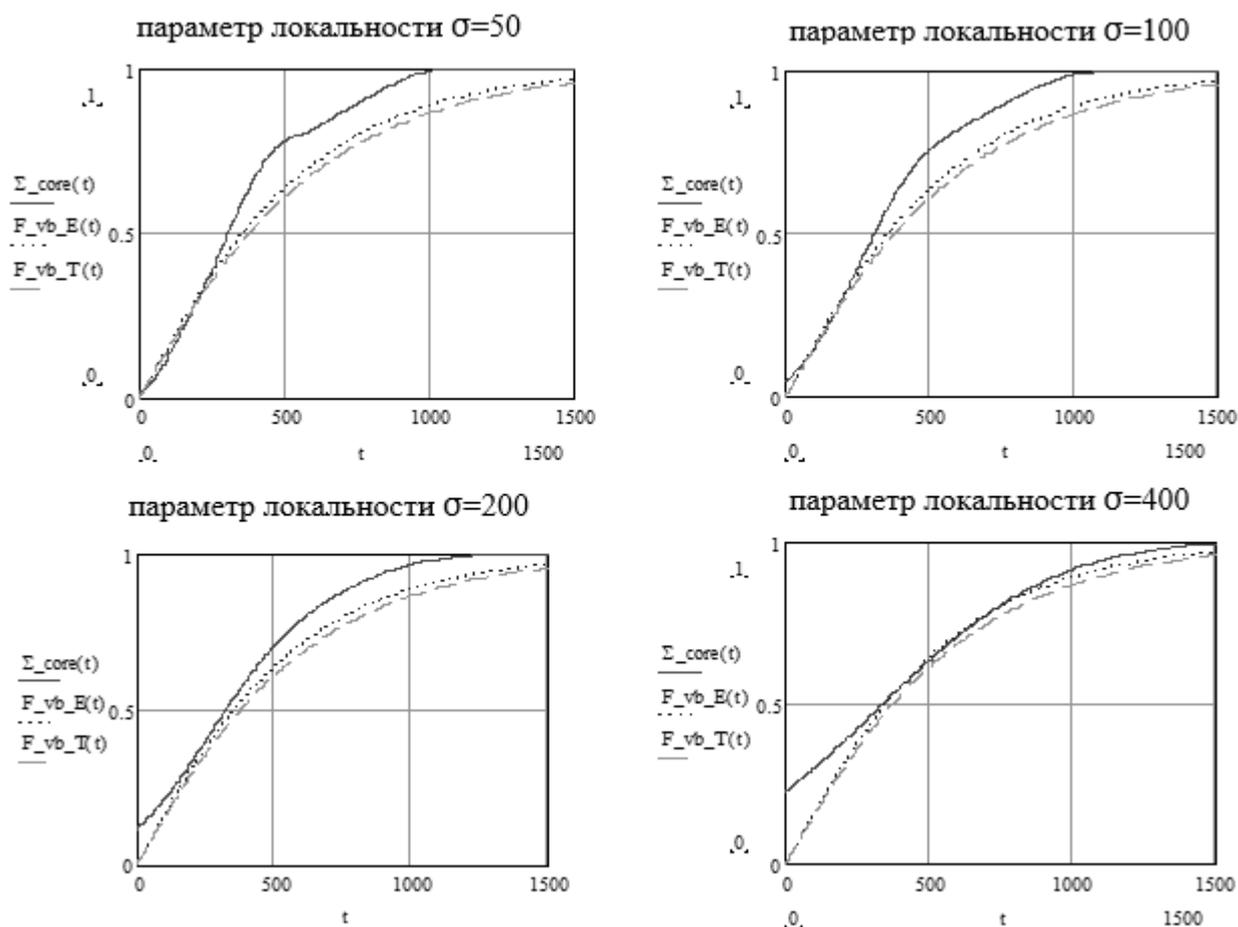
Условия проведения: после генерации выходящие за пределы t СВ отбрасываются (имитируется цензурирование справа).

Исходные данные:

момент последнего контроля $t=1000$;

число сгенерированных СВ: $n=100$ (в расчет $p_q=88$, отброшено $r=12$);

число СВ, отобранных для генерации интервалами равно нулю ($n_q=0, p=88$).



Вывод:

при увеличении параметра локальности σ сказывается отсутствие усечения;
при уменьшении σ влияние усечения уменьшается, но страдает гладкость.

Эксперимент 2 – Исследование масштаба параметра локальности. Полные наработки. Без отброса правых элементов.

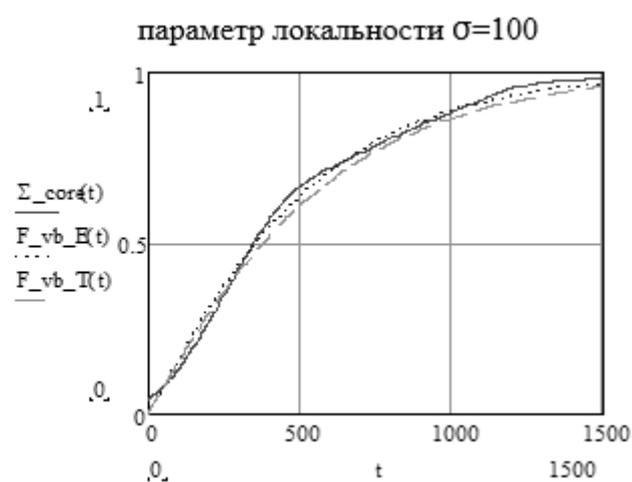
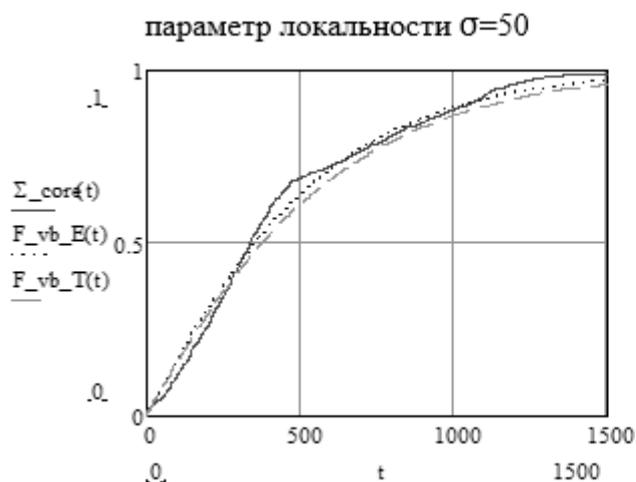
Условия проведения: в рамках исходных данных 1 эксперимента увеличение момента последнего контроля с целью исключить отбрасываемые правые значения.

Исходные данные:

момент последнего контроля $t=3000$;

число сгенерированных СВ: $n=100$ (в расчет $p_q=100$, отброшено $r=0$);

число СВ, отобранных для генерации интервалами равно нулю ($n_q=0$, $p=100$).



Эксперимент 3 – Исследование масштаба параметра локальности. Полные наработки. Малая выборка. Без отброса правых элементов.

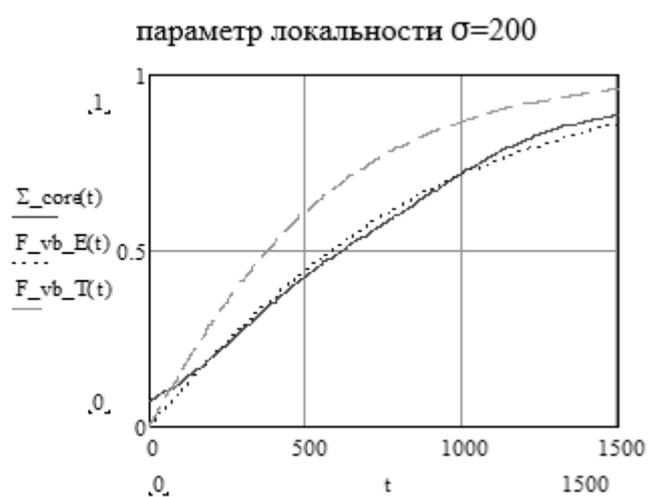
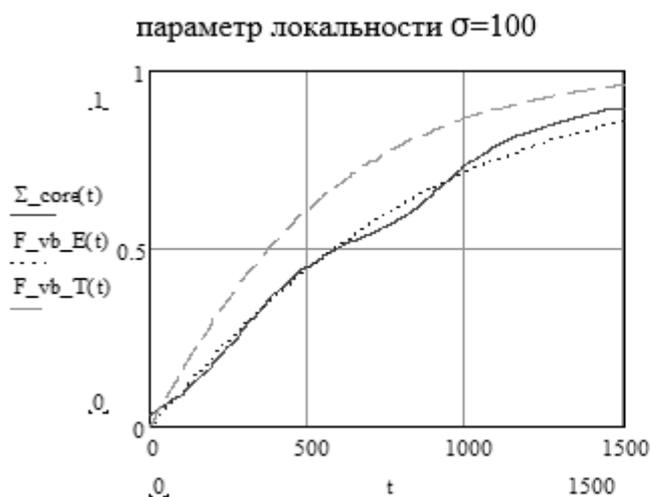
Условия проведения: в рамках исходных данных 1 эксперимента уменьшение объема выборки.

Исходные данные:

момент последнего контроля $t=3000$;

число сгенерированных СВ: $n=20$ (в расчет $p_q=20$, отброшено $r=0$);

число СВ, отобранных для генерации интервалами равно нулю ($n_q=0$, $p=20$).



Эксперимент 4 – Исследование масштаба параметра локальности.

Только цензурированные интервалом. Малая выборка. Без отброса правых элементов

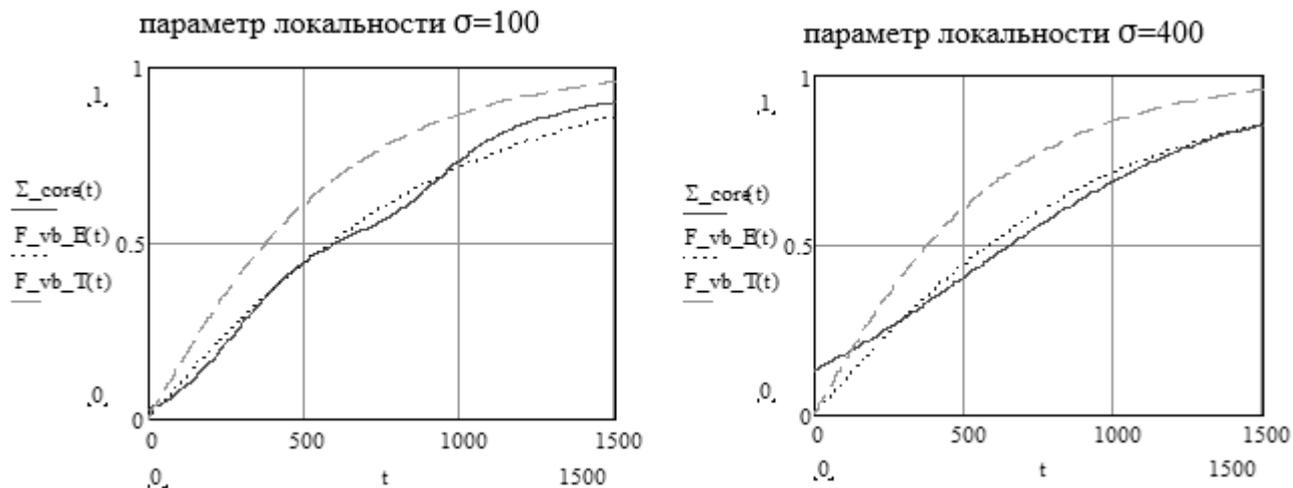
Условия проведения: Все данные цензурируются интервалом.

Исходные данные:

момент последнего контроля $t=3000$; интервалы одинаковые $\Delta=100$;

число сгенерированных СВ: $n=20$ (в расчет $p_q=20$, отброшено $r=0$);

число СВ, отобранных для генерации интервалами – все ($n_q=20$, $p=0$).



Эксперимент 5 – Только цензурирование интервалом. Малая выборка. Без отброса правых элементов. Широкие интервалы

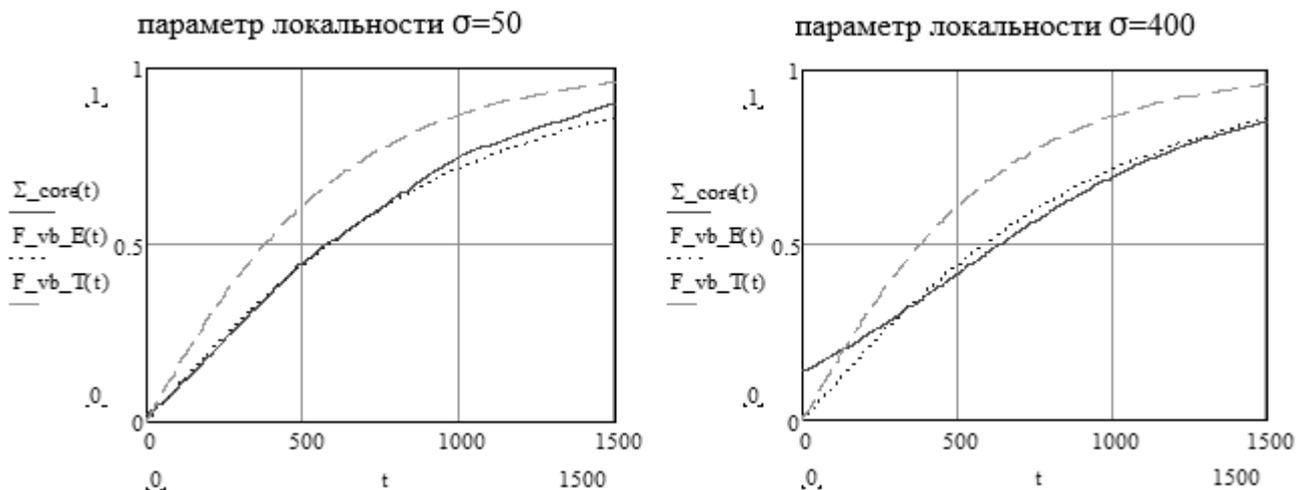
Условия проведения: Все данные цензурируются интервалом. Интервал наблюдения увеличен.

Исходные данные:

момент последнего контроля $t=3000$; интервалы одинаковые $\Delta=500$;

число сгенерированных СВ: $n=20$ (в расчет $p_q=20$, отброшено $r=0$);

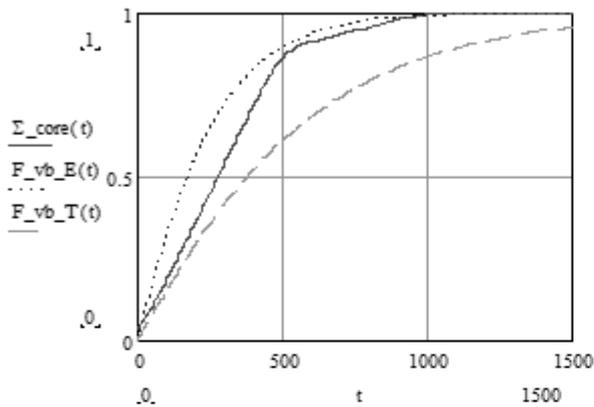
число СВ, отобранных для генерации интервалами – все ($n_q=20$, $p=0$).



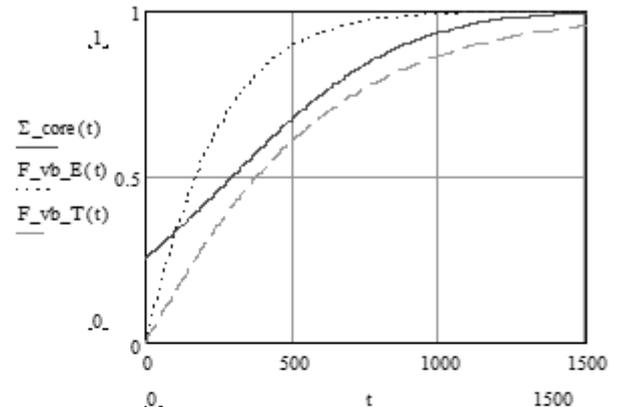
Эксперимент 6 – Только цензурирование интервалом. Очень малая выборка. Без отброса правых элементов

Исходные данные: момент последнего контроля $t=3000$;
 число сгенерированных СВ: $n=9$ (в расчет $p_q=9$, отброшено $r=0$);
 число СВ, отобранных для генерации интервалами – все ($n_q=9$, $p=0$).

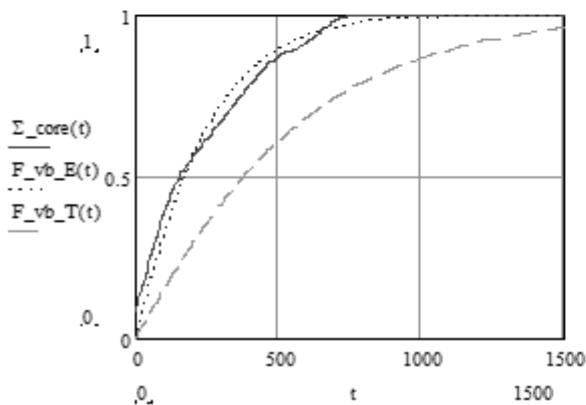
параметр локальности $\sigma=50$
 интервалы одинаковые $\Delta=500$
 массив интервалов: (8, 1, 0, 0, 0)



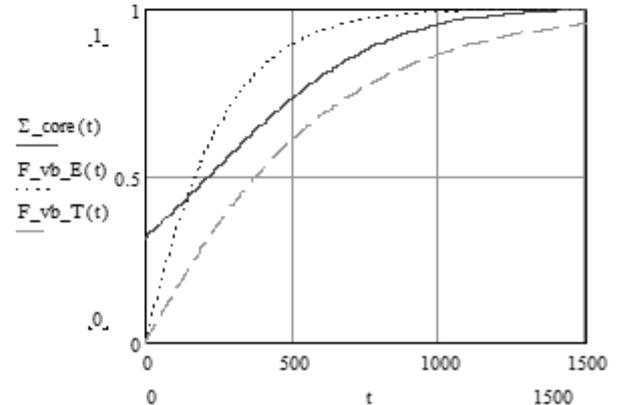
параметр локальности $\sigma=400$
 интервалы одинаковые $\Delta=500$
 массив интервалов: (8, 1, 0, 0, 0)



параметр локальности $\sigma=50$
 интервалы одинаковые $\Delta=100$
 массив интервалов: (4, 1, 1, 1, 1, 0, 1, 0, 0, 0)



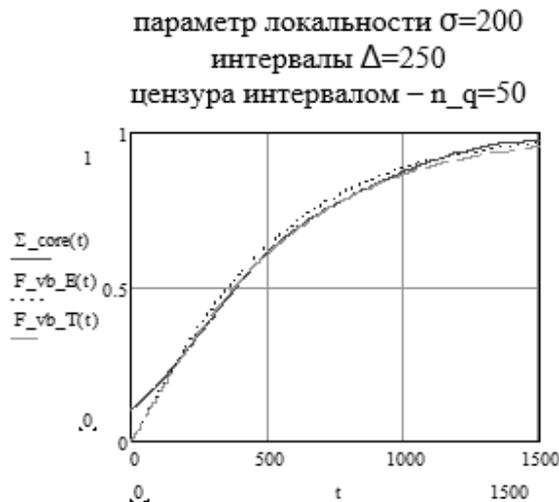
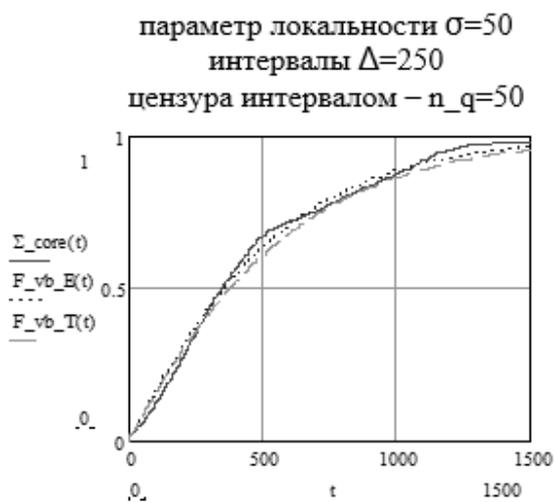
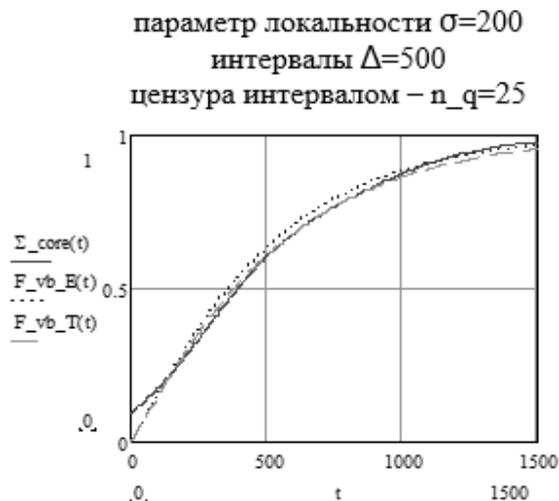
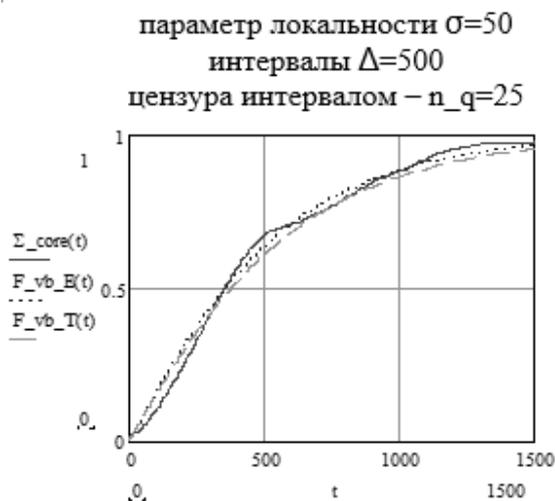
параметр локальности $\sigma=400$
 интервалы одинаковые $\Delta=100$
 массив интервалов: (4, 1, 1, 1, 1, 0, 1, 0, 0, 0)



Эксперимент 7 – Полные наработки и цензурирование интервалом. Большой массив данных. Разные параметры локальности и интервалы.

Исходные данные:

момент последнего контроля $t=3500$; число сгенерированных СВ: $n=100$ (в расчет $p_q=100$, отброшено $r=0$); число СВ, отобранных для генерации интервалами – 50 ($n_q=50, p=50$).

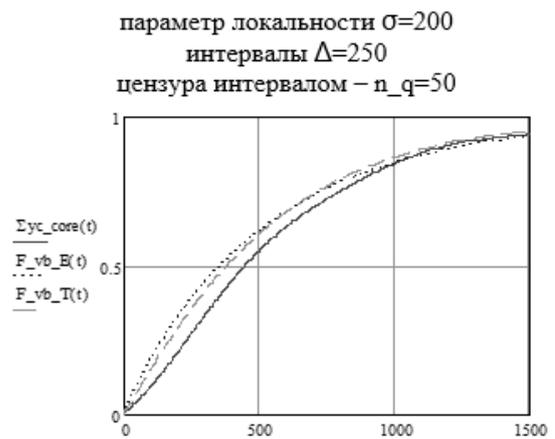
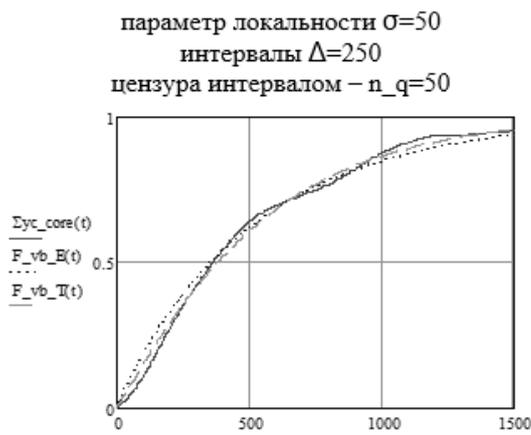
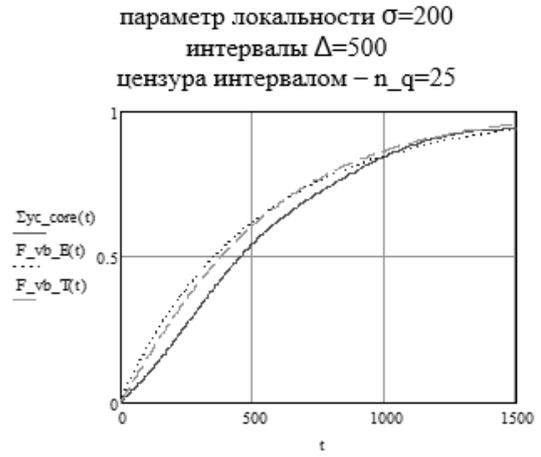
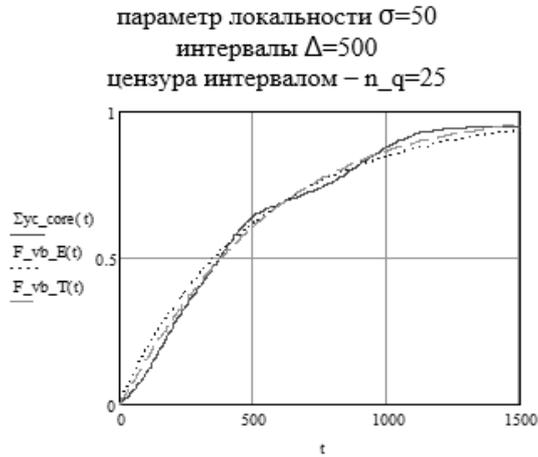


Эксперимент 8 – Усеченные ядра. Полные наработки и цензурирование интервалом. Большой массив данных. Разные параметры локальности и интервалы.

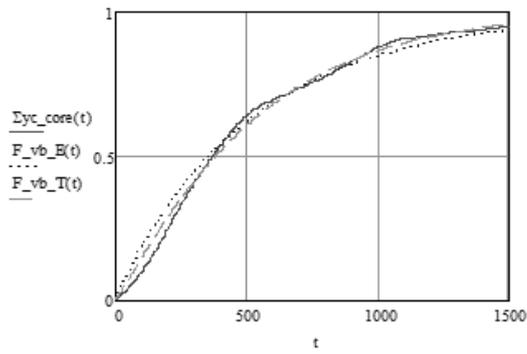
Условия проведения:

Исходные данные:

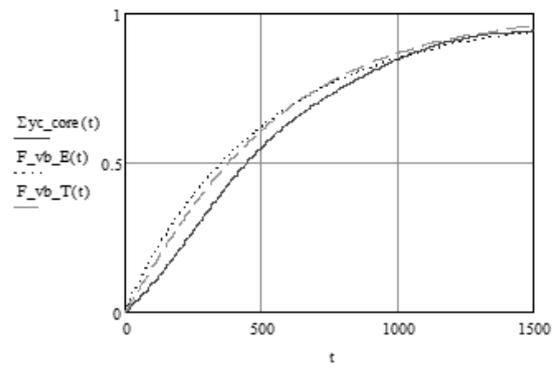
момент последнего контроля $t=3500$; число сгенерированных СВ: $n=100$ (в расчет $p_q=100$, отброшено $r=0$); число СВ, отобранных для генерации интервалами – 50 ($n_q=50$, $p=50$).



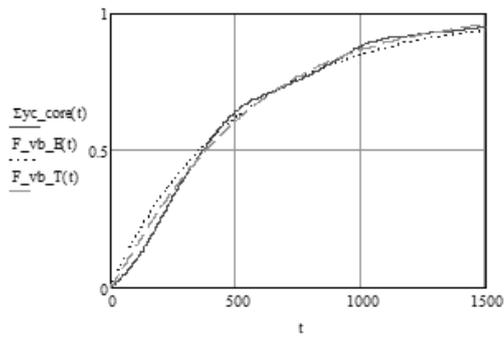
параметр локальности $\sigma=50$
интервалы $\Delta=250$
цензура интервалом $-n_q=75$



параметр локальности $\sigma=200$
интервалы $\Delta=250$
цензура интервалом $-n_q=75$



параметр локальности $\sigma=50$
интервалы $\Delta=500$
цензура интервалом $-n_q=75$



параметр локальности $\sigma=200$
интервалы $\Delta=500$
цензура интервалом $-n_q=75$

