

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ НАУЧНОЕ УЧРЕЖДЕНИЕ
ИНСТИТУТ ПРИРОДНО-ТЕХНИЧЕСКИХ СИСТЕМ

На правах рукописи

ШИШКИН ЮРИЙ ЕВГЕНЬЕВИЧ

**МЕТОДИЧЕСКОЕ, АЛГОРИТМИЧЕСКОЕ И ПРИБОРНОЕ
ОБЕСПЕЧЕНИЕ ПРОЦЕДУР ОБНАРУЖЕНИЯ АНОМАЛИЙ
В МОНИТОРИНГОВЫХ НАБЛЮДЕНИЯХ ПРИРОДНОЙ СРЕДЫ
НА ОСНОВЕ МОДУЛЬНО-КОМПОНЕНТНОГО ПОДХОДА**

05.11.13 – Приборы и методы контроля природной среды, веществ,
материалов и изделий

ДИССЕРТАЦИЯ

на соискание ученой степени
кандидата технических наук

Научный руководитель
доктор технических наук, профессор
Скатков Александр Владимирович

Севастополь – 2022

ОГЛАВЛЕНИЕ

Список сокращений и условных обозначений	7
Введение	8
Глава 1. Анализ априорной информации об аномалиях в мониторинговых наблюдениях природной среды и выбор перспективных направлений развития методов их обнаружения	25
1.1 Типизация и классификация потенциально возможных аномалий в данных мониторинга природной среды.....	26
1.1.1 Двойственная интерпретация аномалий в данных мониторинговых наблюдений природной среды: N-аномалии моделей знаний и M-аномалии данных, оценивание вероятностей правильных и ошибочных решений....	27
1.1.2 Формализация и спецификация постановок задач обнаружения потенциально возможных аномалий в данных мониторинговых наблюдений природной среды	33
1.2 Системный анализ существующих методов обнаружения аномалий в данных статистических исследований природной среды.....	43
1.2.1 Особенности процедур обнаружения аномалий в данных мониторинга природной среды аналитическими методами	45
1.2.2 Алгоритмизация процессов обнаружения аномалий в данных мониторинга природной среды	48
1.2.3 Специализированные интеллектуальные информационные технологии процедур обнаружения аномалий для реализации в современных системах мониторинга природной среды	53
1.3 Спецификация требований к методической, технической и программной реализации систем мониторинга	56
1.3.1 Требования к методам обнаружения аномалий в данных мониторинга природной среды.....	58

1.3.2 Требования к технической реализации модулей обнаружения аномалий	59
1.3.3 Пользовательские требования к программной реализации процедур обнаружения аномалий.....	60
1.4 Выбор перспективных направлений развития процедур модульно-компонентного подхода для обнаружения аномалий в данных мониторинга природной среды	60
1.5 Выводы по главе 1	61
Глава 2. Методическое и алгоритмическое обеспечение модульно-компонентного подхода обнаружения аномалий в данных мониторинга.....	63
2.1 Иерархический метод аналитического восстановления нерегулярных и пропущенных данных мониторинга в реальном времени	64
2.2 Метод структурирования данных мониторинга сложного и периодического формата с повышенными требованиями к вычислительным ресурсам	68
2.3 Метод и алгоритм адаптивного выбора метода представления данных мониторинга природных систем с использованием векторных критериев	73
2.4 Метод и алгоритм обнаружения аномалий данных многопараметрического мониторинга природной среды с кластеризацией состояний	76
2.5 Метод и алгоритм численной оценки точности бинарного классификатора данных мониторинговых наблюдений в общем виде	80
2.6 Метод и алгоритм обнаружения аномалий в скалярных полях с многомерными данными мониторинга природной среды.....	83
2.7 Метод обнаружения аномалий в многослойных регулярных полях данных высокой плотности.....	88
2.8 Метод обнаружения аномалий в данных мониторинга при нежестких ограничениях на объем обучающих выборок с использованием агентного подхода.....	90

2.9 Модульно-компонентный подход при реализации полимодельного комплекса обнаружения аномалий различных типов	94
2.10 Выводы по главе 2.....	102
Глава 3. Исследование функциональных свойств основных модулей системы обнаружения аномалий в данных мониторинга природной среды.....	104
3.1 Анализ свойств функциональных модулей обнаружения аномалий в данных мониторинга.....	105
3.1.1 Описание модуля лабораторного стенда тестовой среды	106
3.1.2 Результаты анализа свойств функционального модуля оценки отдельных характеристик методами обнаружения М-аномалий в данных мониторинга природной среды	111
3.1.3 Результаты анализа свойств функционального модуля оценки совокупности характеристик методами обнаружения N-аномалий в нерегулярных и пропущенных данных	123
3.2 Модули системы поддержки принятия решений для обнаружения аномалий	125
3.2.1 Модули системы поддержки принятия решений, оценивающие степень аномальности данных мониторинга природной среды.....	126
3.2.2 Модуль повышения качества принятия решений за счет минимизации риска и оценивания вероятностей ошибок первого и второго рода.....	129
3.3 Выводы по главе 3.....	135
Глава 4. Модульное программно-техническое обеспечение и его практическое применение для поддержки принятия решений о наличии аномалий в данных мониторинга природной среды на основе модульно-компонентного подхода.....	137
4.1 Архитектура модульного программно-технического комплекса	139
4.1.1 Аппаратный универсальный 16-канальный модуль лабораторного стенда для регистрации аналоговых сигналов и потоков цифровых данных,	

и обнаружения в них аномалий, на основе численных метрик, в режиме реального времени	140
4.1.2 Модификация аппаратного модуля программно-технического комплекса для обнаружения аномалий с возможностью использования дополнительных ресурсов на базе агентного подхода.....	141
4.1.3 Модификация аппаратного модуля автономного обнаружения аномалий при долговременном мониторинге	142
4.1.4 Программные модули автономного обнаружения аномалий в данных мониторинга.....	144
4.1.5 Реализация принципов архитектуры комплекса в соответствии с сформированными требованиями к методике обнаружения аномалий	148
4.2 Модули визуализации и анализа данных	149
4.3 Программный модуль выбора сценариев решения задач обнаружения аномалий и оценки адекватности СППР	151
4.4 Модельные экспериментальные исследования процессов распространения аномалий в пространственном и временном измерениях, и их обнаружения	152
4.5 Экспериментальные испытания модульной системы обнаружения аномалий, проведенные в лабораторной установке	154
4.6 Повышение уровня интеллектуальности модуля СППР и анализ его эффективности на основе векторных критериев	157
4.7 Выводы по главе 4.....	160
Заключение	162
Список использованных источников	167
Приложение А. Примеры аномалий природной среды обнаруживаемых системой	188

Приложение Б. Визуальное представление интегрального показателя состояния природной среды и отдельных его параметров для набора данных содержащего N-аномалии	192
Приложение В. Архитектура модульно-компонентной СППР обнаружения М-аномалий и N-аномалий.....	198
Приложение Г. Синтез понятия аномалии в результатах мониторинга природной среды.....	215
Приложение Д. Результаты лабораторных испытаний аппаратного детектора М-аномалий на специализированном лабораторном стенде	218
Приложение Е. Патент и свидетельства о регистрации программ для ЭВМ	221

СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

А-задача	задача обнаружения аномальных значений в данных мониторинга
А-событие	событие присутствия аномальных значений в данных мониторинга
АП	аномальное пятно
БД	база данных
БЗ	база знаний об объекте мониторинга
ДПФ	дискретное преобразование Фурье
ДРМ	данные результатов мониторинга
ИМ	имитационная модель
ЛПР	лицо принимающее решение
МКП	модульно-компонентный подход
МН	мониторинговые наблюдения
МПТО	модульное программно-техническое обеспечение
ОС	операционная система
ПО	программное обеспечение
ПС	природная среда
ПТС	природно-техническая система
СБ	случайные блуждания
СМО	система массового обслуживания
СМ	система мониторинга
САПР	система автоматизированного проектирования
СППР	система поддержки принятия решений
СПИ	сеть первичных измерителей
ЭВМ	электронная вычислительная машина
CAD	система автоматизированного проектирования (Computer Aided Design)
GUI	графический интерфейс пользователя (Graphical User Interface)
PDCA-цикл	планирование, выполнение, проверка, воздействие (Plan Do Check Act)
KPI	ключевые показатели эффективности (Key Performance Indicators)
web-сервис	идентифицируемая web-адресом мобильная программная система

ВВЕДЕНИЕ

Актуальность исследования. Обеспечение рационального природопользования, оптимизация хозяйственной деятельности человека, увеличение безопасности объектов критической инфраструктуры, повышение качества природной среды (ПС) и жизни человека в целом требуют комплексного изучения методов аналитического и неразрушающего контроля ПС в части обнаружения скрытых аномалий в скалярных и векторных полях мониторинга ПС и природно-технических систем (ПТС).

Современные требования к оперативности и надежности систем экологического мониторинга природных объектов непрерывно растут, что обуславливает необходимость как разработки новых проблемно-ориентированных методов, так и усовершенствования существующей системы методов анализа данных результатов мониторинга (ДРМ) ПС. Комплексное решение задачи обнаружения аномальных значений в ДРМ (А-задачи) включает разработку реализующего эти методы алгоритмического и программно-технического обеспечения, использующего новые достижения в области вычислительной техники и программного обеспечения.

Специфика анализа ДРМ ПС включает в себя необходимость учета совместной пространственно-временной изменчивости ее показателей как полей данных. Развитие методов анализа данных скалярных и векторных полей мониторинговых наблюдений (выявления информативных сигналов) ПС представляет несомненный интерес для специалистов различных областей научных исследований (океанологии, биологии, экологии, климатологии и др.) и практической деятельности (рыбопромышленной, нефтегазодобывающей, транспортно-логистической и др.).

Сложность логической и структурной организации системы «природная среда – приборы и методы ее контроля», в равной степени, как и моделей ПТС приводит к объективным трудностям при решении вопросов обнаружения аномалий в данных их мониторинга. Крайне актуальными остаются задачи

повышения эффективности применения приборов и методов контроля ПС, в контексте повышения их информационной надежности, путем обнаружения аномалий в ДРМ, разнообразие формулировок которых обусловлено сложностью ПС и субъективностью восприятия аномалий лицами, принимающими решения (ЛПР).

Исторически сложилось, что термин «аномалия» имеет широкий контекст употребления. Поэтому чтобы исключить неоднозначность здесь и далее под аномалией, в контексте ДРМ ПС, будут пониматься понятия данные в энциклопедическом словаре Ф.А. Брокгауза и И.А. Ефрона «Аномалия – отступление или уклонение от правила, поэтому аномальным называют все отступающее или уклоняющееся от правильного или нормального. Аномальными в области природы считаются такие явления, которые вопреки законам природы представляются исключениями; но несмотря на это, все аномальное не лишено законности; более глубокое созерцание законов природы приводит к тому, что на кажущихся аномалиях замечается отпечаток общей законности. Каждый поэтому может ожидать встретить тем более аномалий, чем менее он знаком с общими законами явлений природы...» [12, 17] как общепринятое и наиболее близкое в контексте решаемой задачи.

Аномалии ДРМ ПС в общем случае будем интерпретировать в двух формах: как появление данных содержащих ошибки (М-аномалии) или как возмущения в данных, вызванные изменениями в состоянии наблюдаемой ПС, которые могут качественно интерпретироваться как новые знания (N-аномалии). Такое разделение обусловлено тем, что в процессе эксплуатации систем мониторинга ПС, и анализа данных, получаемых с помощью таких систем, экспертам приходится решать задачи обнаружения двух типовых классов аномалий:

– М-аномалии, источниками которых является неопределенность в самих данных, не связанная с возникновением качественных изменений в самой ПС, а возникающая из-за ряда других объективных причин (ошибок в средствах мониторинга, шумов, вариативности данных, человеческого фактора и пр.) и в целом характеризуемая свойством самих данных, а не модели. Для их обнаружения

используются подходы: графовые, статистические, иерархические и таксономические алгоритмы кластеризации, нейронная сеть Кохонена, метод k -средних, автоассоциатор, глубокая сеть доверия, сеть радиально-базисных функций, метод главных компонент, анализ независимых компонент, многомерное шкалирование, EM-алгоритм [36–44];

– N-аномалии, источниками которых являются неопределенности в знаниях о ПС, возникающие вследствие возникновения качественных изменений состояния наблюдаемой ПС (или некорректности данных для обучения математической модели) и в целом характеризует свойство модели, а не данных. Такие аномалии вызваны естественными изменениями в самой среде и по сути могут качественно интерпретироваться как новые знания о ПС. Для их обнаружения используются подходы: k -ближайших соседей, опорных и релевантных векторов, линейной и логистической регрессии, наивный байесовский классификатор, дерево решений, перцептрон, ансамбли моделей [45–53].

Возможен пограничный вариант – когда на основе экспертных знаний строится модель нормального состояния ПС и в случае несоответствия экспериментальных данных теоретическим происходит или коррекция модели, или принимается решение о том, что поступившие данные являются аномальными. В таких случаях включают подходы: на основе графов, разделение низкой плотности, генеративные модели, эвристические, интерактивное взаимодействие с оператором [54–63].

Проблемами обнаружения аномальных значений в данных (в широком смысле, без учета контекста природы их возникновения) занимались многие выдающиеся отечественные и зарубежные ученые, получившие основополагающие результаты: К.Ф. Гаусс, П.Л. Чебышев, К. Пирсон, У. Госсет (Стьюдент), В.И. Романовский, Р. Фишер, А.Н. Колмогоров, Н.В. Смирнов, В.С. Королук. Некоторые обобщения результатов их исследований, направленных на решение А-задач содержатся в работах [29–35]. Разработаны различные эмпирические, статистические параметрические и непараметрические математические модели и методы, направленные на обнаружение аномальных

значений как реализаций случайных величин подчиняющихся отличающимся от эталонной выборки законам распределения [25]. Эти теоретические модели могут быть адаптированы для решения А-задач для ПС и ПТС, однако они строятся, исходя из предположения о независимости измерений как элементов случайной выборки, реализации случайного процесса с неизменными дисперсией и математическим ожиданием, и при условии асимптотической сходимости результатов измерений, в то время как в действительности эти данные являются реализацией случайного процесса переменной структуры, как правило нестационарного. Это приводит к смещению статистических оценок, значительному снижению реактивности классификаторов, повышению вероятностей возникновения ошибок I и II рода. Так системы, построенные на основе классических методов анализа данных довольно статичны и направлены на обнаружение известных и точно описанных состояний ПТС, но зачастую оказываются не в состоянии обнаружить их модификации или изменения в ПС антропогенной природы, что делает их использование малоэффективным.

Эмпирические и статистические модели, позволяющие определять численную меру различия между эталонной и эмпирической выборками одной или нескольких наблюдаемых величин, появились задолго до имитационных численных методов в рамках физико-статистического подхода, и существуют, и применяются до сих пор. В основе нового предлагаемого направления лежит комплекс модульных нелинейных аналитических и численных моделей, которые стали результатом усложнения (по сравнению с традиционными статистическими) процессов построения аналитических зависимостей: вводятся различные способы расчёта коэффициентов линейных комбинаций, а также рассматриваются варианты нелинейных связей, численно моделируются процессы включая все более глубокие уровни абстракции. Примерами таких подходов являются искусственные нейронные сети и другие интеллектуальные подходы, параметрические и непараметрические статистические методы, численное и имитационное моделирование. Значительный вклад в становление и развитие двух последних групп методов внесли: А.Н. Колмогоров, Н.В. Смирнов, В.С. Королук,

В.Г. Бондур, результирующие положения работ которых были использованы как основа при написании диссертации.

Значительный вклад в развитие интеллектуальных технологий анализа ДРМ внесли Охтилев М.Ю., Соколов Б.В., Юсупов Р.М. Предложены методы анализа ДРМ с целью поддержки принятия решений по действиям в кризисных ситуациях, в условиях неполноты, неопределенности, неточности и противоречивости информации о складывающейся обстановке, а также классификации информационных ситуаций, при наличии неустранимого порогового ограничения времени на цикл формирования и реализации решений по предотвращению возможных катастроф, чрезвычайных и аварийных ситуаций за счет проактивного управления [64–71].

Существующее аналитическое решение А-задачи базируется на исследованиях академика А.Н. Колмогорова, и использует математическую модель, предполагающую асимптотическую сходимость результатов мониторинга при условии наличия обучающих выборок больших объемов. При том, что сформулированные А.Н. Колмогоровым теоремы нашли широкое применение в качестве критериев согласия их использование в практических вопросах не всегда достаточно обоснованно, так как все они носят характер предельных соотношений. Полученные результаты предполагают наличие выборок неограниченно больших объемов, что в практических мониторинговых наблюдениях практически никогда не имеет место, а характер асимптотической сходимости не исследован и исследованию аналитическими методами, судя по всему, не подлежит.

Важные результаты в области развития методов решения А-задач, выполняемых с целью повышения обоснованности принимаемых решений при осуществлении аналитического контроля ПС, были получены коллективом ученых под руководством вице-президента РАН академика В.Г. Бондура. Ими опубликованы работы, в которых предложены решения практических аспектов задачи обнаружения аномалий ДРМ ПС: аномальной изменчивости пространственно-временных распределений явлений природного и техногенного характера, аномалий взволнованной морской поверхности, аномалий возмущения

состояния геофизических полей, аномальных зон и критериев аномальности в этих зонах, тепловых аномалий как контраста температуры в локализованных участках среды, аномалий как эмиссий вредных примесей по данным космического мониторинга, аномалий как распространения загрязненных вод из подводного источника в стратифицированной среде прибрежной акватории [26–28]. В работах Бондура В.Г. отмечается эффективность совместного использования данных модульных аппаратно-программных решений, сети первичных измерителей (СПИ), данных спутникового мониторинга и методов численного моделирования предметной области для решения задач обнаружения произвольных типов аномалий, в т.ч. новых, а также воздействий антропогенного характера, в полях ДРМ ПС.

Анализ научных работ авторов, посвященных вопросам обнаружения аномалий в ДРМ, равно как и других отечественных и зарубежных работ, показал, что современный уровень развития приборов и методов контроля ПС, программно-технического обеспечения позволяет в полной мере решать лишь ограниченный круг задач обнаружения аномалий в ретроспективных и текущих ДРМ ПС и не отвечает всем запросам практической деятельности современного социума. Не существует функционально полного замкнутого решения, позволяющего производить обнаружение аномалий в ДРМ М и N типов, и принимать однозначно верные решения на их основе. Прикладные задачи анализа результатов мониторинга ПС и ПТС выявляют острую востребованность решения задач обнаружения аномалий произвольных типов в условиях реального времени, малых объемов выборок, пропущенных и искаженных данных, которые до настоящего времени исчерпывающе и в полном объеме не разрешены. Остро стоит также задача повышения достоверности статистических выводов, увеличения точности прогнозов, своевременного предупреждения о возможных экстремальных явлениях. Таким образом, развитие интеллектуальных методов повышения информационной надежности приборов и методов контроля ПС, в равной степени, как и реализующих их программно-технических модулей, являющихся основой для

осуществления поддержки принятия решений о наличии М-аномалий и N-аномалий в ДРМ ПС и ПТС, является продуктивным.

Объект исследования – данные многопараметрических мониторинговых наблюдений природной среды в виде пространственно-временных скалярных и векторных полей.

Предмет исследования – методическое, алгоритмическое и приборное обеспечение процедур обнаружения аномалий в факторном пространстве данных мониторинга природной среды, реализованное на основе модульно-компонентного подхода.

Цель диссертационной работы состоит в повышении информационной надежности модульного методического, алгоритмического и программного обеспечения приборов и средств контроля природной среды за счет разработки модулей поддержки принятия решений о наличии аномалий в факторном пространстве скалярных и векторных полей данных мониторинга.

В соответствии с поставленной **целью** сформулированы и решены следующие **научные задачи**:

1. Осуществить классификацию информационных ситуаций, возникающих при мониторинге природной среды, и разработать проблемно-ориентированные модули их обнаружения.

2. Исследовать и развить методико-алгоритмический инструментарий выявления аномальной изменчивости данных экологического мониторинга природной среды в факторном пространстве скалярных и векторных полей. Обеспечить повышение реактивности принятия решений по данным биосенсорных измерителей, соизмеримое с реактивностью самих измерителей.

3. Исследовать и развить проблемно-ориентированные модульные методы и алгоритмы, адаптивно осуществляющие поддержку принятия решений о наличии аномалий различных типов в данных мониторинга природной среды, на основе векторных критериев качества.

4. Исследовать и развить модульное программно-техническое обеспечение, предназначенное для обработки информативных сигналов, для выявления аномалий в данных мониторинга в реальном времени, включающее связку интегрируемого непосредственно в средство измерения аппаратного модуля выявления аномалий и внешнего программного серверного модуля.

Соответствие диссертации паспорту научной специальности. Диссертация соответствует паспорту научной специальности 05.11.13 – «Приборы и методы контроля природной среды, веществ, материалов и изделий» по пунктам 4. «Разработка методического, технического, приборного и информационного обеспечения для локальных, региональных и глобальных систем экологического мониторинга природных и техногенных объектов» (пункт №2 научной новизны), 6. «Разработка алгоритмического и программно-технического обеспечения процессов обработки информативных сигналов и представление результатов в приборах и средствах контроля, автоматизация приборов контроля» (пункт №3 научной новизны) и 7. «Методы повышения информационной и метрологической надежности приборов и средств контроля в процессе эксплуатации, диагностика приборов контроля» (пункт №1 научной новизны), отрасль наук – технические науки.

Теоретическая значимость работы заключается в постановке и комплексном решении, на базе модульного методического, алгоритмического и приборного обеспечения, новой обобщенной и многокритериальной задачи обнаружения аномалий в данных факторного пространства скалярных и векторных полей мониторинга, ориентированного на обнаружение аномалий в данных мониторинга природной среды. Решение задачи осуществлено за счет модульной организации системы, компоненты которой образуют функционально замкнутую структуру, позволяющую повысить информационную надежность приборов и средств контроля природной среды.

Практическая значимость работы обусловлена тем, что полученные результаты позволят организовать работу многих систем мониторинга, за счет

реализации модульных решений созданного программно-технического комплекса, в котором реализованы модули для обнаружения аномалий в данных, учитывающие особенности их применения в системах мониторинга природной среды: обнаружение аномалий в условиях реального времени, неполной информации, малых объемов выборок, нерегулярных и пропущенных данных. Полученные результаты служат повышению информационной надежности средств контроля природной среды за счет решения задачи обнаружения аномальностей пространственно-временной изменчивости мониторинговых наблюдений природной среды, и представляют практический интерес для широкого круга специалистов в отраслях промысловой океанографии, экологии, морской биологии, нефтегазодобывающей, транспортно-логистической, военной и др., а также других изменчивых во времени пространственно-распределенных характеристик, независимо от их природы.

Методы исследования. Для решения поставленных задач использовались: математическое и компьютерное моделирование, численные методы, методы двумерной и многомерной сплайн-интерполяции, дискретное преобразование Фурье, векторный анализ градиентов скалярного поля, ортогональная цифровая фильтрация, математическая статистика, кластерный анализ, вейвлет преобразования. При реализации программно-технического модуля и в экспериментальной части применялись агентное и дискретно-событийное имитационное моделирование, принципы объектно-ориентированного программирования, интерактивные технологии визуализации данных, вычислительный эксперимент.

Достоверность научных результатов и выводов диссертационной работы обеспечены корректностью применяемого математического аппарата, подтверждается результатами вычислительных экспериментов, полученными с использованием проблемно-ориентированного лабораторного стенда для целенаправленного моделирования с целью выявления уровней достоверности

обнаружения аномалий, при принятии гипотез об ошибках I и II рода, апробацией на научно-технических конференциях.

Диссертация выполнялась в рамках реализации плановых исследований по базовым темам государственных заданий ФГБНУ ИПТС «Научные основы разработки методов, технологий и средств исследования в реках, морях и океанах» (2016 г., АААА-А17-117010900029-0), «Разработка оптических, акустических методов, технологий и средств исследования гидросферы и их оснащение возобновляемыми источниками энергии» (2017 г., 2018 г., АААА-А17-117021310008-4), «Разработка новых средств и измерительных информационных технологий исследований природных вод» (2019 г., 2020 г., № 0012-2019-0003), при поддержке грантов РФФИ «Исследование методов анализа больших данных в конвергентных информационных системах и инфраструктурах для задач экологического мониторинга Севастопольского региона» (2018–2021 гг., № 18-47-920005 p_a), «Комплексная система экологического мониторинга водной среды интеллектуальными автономными роботами» (2018–2021 гг., № 18-48-920018 p_a), «Адаптивные нейросетевые методы обнаружения уязвимостей интерфейсов беспилотных транспортных средств на основе искусственных иммунных систем» (2019–2021 гг., № 19-29-06015). Результаты работ по теме диссертации вошли составной частью в отчеты по этим темам и грантам. Результаты диссертационной работы использованы при выполнении НИОКР по заказу АО «Концерн «Океанприбор».

На защиту выносятся следующие положения (основные научные результаты), являющиеся новыми или содержащими элементы новизны:

1. Получила дальнейшее развитие классификация информационных ситуаций, возникающих при мониторинге природной среды, за счет введения в рассмотрение факторного пространства скалярных и векторных полей мониторинга и критериального пространства точности, полноты, реактивности, рисков и вероятностей ошибок первого и второго рода.

2. Разработано новое модульное методическое, техническое и информационное обеспечение систем экологического мониторинга природной среды за счет нового метода качественной дифференциальной оценки состояния природной среды, установки для его осуществления и программного модуля консолидации и структуризации данных мониторинга большого объема и модуля снижения избыточности данных мониторинга.

3. Разработано новое модульное программно-алгоритмическое обеспечение поддержки принятия решений о наличии аномалий в данных мониторинга природных систем, адаптирующееся к классу выявляемых аномалий в реальном времени за счет динамически переключаемых проблемно-ориентированных моделей и векторного критерия качества.

Внедрение результатов диссертационной работы.

Результаты диссертационной работы и разработанное автором программное обеспечение были успешно использованы:

1. При выполнении ОКР по заказу АО «Концерн «Океанприбор» – «Разработка и испытания опытного образца автоматизированного биосенсорного комплекса раннего оповещения для экологического мониторинга водной среды». ОКР осуществлена в соответствии с задачами национального проекта «Экология» и национального проекта «Наука» по развертыванию в акватории пляжей города Севастополя элементов «Автоматизированной биосенсорной системы раннего оповещения для экологического мониторинга водной среды» (2021 г., № 5/SIMBF/2020) [168].

2. При выполнении НИР в рамках государственных заданий по темам «Научные основы разработки методов, технологий и средств исследования в реках, морях и океанах» (2016 г., АААА-А17-117010900029-0), «Разработка оптических, акустических методов, технологий и средств исследования гидросферы и их оснащение возобновляемыми источниками энергии» (2017 г., 2018 г., АААА-А17-117021310008-4), «Разработка новых средств и измерительных информационных технологий исследований природных вод» (2019 г., 2020 г., № 0012-2019-0003).

3. При выполнении проектов РФФИ № 18-47-920005 р_а «Исследование методов анализа больших данных в конвергентных информационных системах и инфраструктурах для задач экологического мониторинга Севастопольского региона» (2018–2021 гг.), № 18-48-920018 р_а «Комплексная система экологического мониторинга водной среды интеллектуальными автономными роботами» (2018–2021 гг.) и № 19-29-06015 мк «Адаптивные нейросетевые методы обнаружения уязвимостей интерфейсов беспилотных транспортных средств на основе искусственных иммунных систем» (2019–2021 гг.).

4. В лаборатории гидроакустических и оптических методов и средств контроля состояния окружающей среды Центра экологического приборостроения и экоэнергетики Института природно-технических систем в составе автоматизированной системы измерения скорости течения для морских и речных условий.

Результаты работ по теме диссертации вошли составной частью в отчеты по этим темам и грантам.

Основные компоненты созданного комплекса программных модулей: «Модуль консолидации структурированных данных большого объема в режиме реального времени» (2018 г., АС РФ № 2018613184), «Модуль снижения избыточности данных мониторинга» (2017 г., АС РФ № 2017664038), «Имитационная модель маршрутизатора самоорганизующейся сети передачи данных мониторинга» (2018 г., АС РФ № 2018617880), «Программа синхронизации и коррекции измерений вектора скорости течения» (2020г., АС РФ № 2018666181), «Программа для автономного модуля буйкового акустического измерителя скорости течения» (АС РФ № 2020664093), «Программа GPS навигатора для автономного буйкового акустического измерителя скорости течения» (2020г., АС РФ № 2020664094), «Программа распознавания аномальностей в данных мониторинговых наблюдений» (2020г., АС РФ № 2020662839) зарегистрированы в Реестре программ для ЭВМ. Патент на «Способ контроля качества воды и двухканальная биосенсорная установка для его осуществления» (2021г., № 2020141487). Патент и авторские свидетельства представлены в приложении Е.

Личный вклад автора в совместных работах. Материалы диссертации являются обобщением работ автора, выполненных в период с 2015 по 2021 год, и отражают его личный вклад в настоящее исследование. Основные научные результаты получены автором самостоятельно. Постановка цели исследования и решаемых задач, обсуждение полученных результатов и подготовка материалов к печати выполнена совместно с научным руководителем. Разработка общей схемы и детализация этапов исследования, создание интеллектуальной технологии обнаружения аномалий в мониторинговых наблюдениях, разработка численных методов определения контрольно-предупредительных границ параметров факторного пространства, создание программных комплексов реализующих эту технологию, разработка лабораторного стенда для экспериментальной оценки её эффективности, и анализ результатов вычислительных экспериментов проводились автором самостоятельно.

Апробация результатов диссертации. Результаты, полученные в диссертации, докладывались и обсуждались на **18** конференциях: международных научно-технических конференциях «2018 International Russian Automation Conference (RusAutoCon)», «2019 International Russian Automation Conference (RusAutoCon)» и «2021 International Russian Automation Conference (RusAutoCon)» (г. Сочи, 9–16 сентября 2018 г., 8–14 сентября 2019 г., 5–11 сентября 2021 г.); 26-й международной конференции Тихоокеанского конгресса морских наук и технологий «Marine Science and Technology for Sustainable Development (PACON-2019)» (г. Владивосток, 16–19 июля 2019 г.); международных научных конференциях «2019 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon)» и «2020 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon)» (г. Владивосток, 1–4 октября 2019 г., 6–9 октября 2020 г.); международных научно-технических конференциях «Информационные технологии и информационная безопасность в науке, технике и образовании ИНФОТЕХ – 2015», «Информационные технологии и информационная безопасность в науке, технике и образовании ИНФОТЕХ – 2017» и «Информационные технологии и информационная безопасность в науке,

технике и образовании ИНФОТЕХ – 2019» (г. Севастополь, 7–11 сентября 2015 г., 18–20 сентября 2017 г., 18–20 сентября 2019 г.); Всероссийских научно-технических конференциях «Интеллектуальные системы, управление и мехатроника – 2015» и «Интеллектуальные системы, управление и мехатроника – 2016» (г. Севастополь, 13–15 мая 2015 г., 19–21 мая 2016 г.); IV и V Всероссийских научных конференциях молодых ученых «Комплексные исследования Мирового океана (КИМО-2019)» и «Комплексные исследования Мирового океана (КИМО-2020)» (г. Севастополь, 18–21 сентября 2019 г., г. Калининград, 18–22 мая 2020 г.); 3-й всероссийской конференции молодых ученых «Техносфера XXI века» (г. Севастополь, 22–26 апреля 2019 г.); молодежной научной конференции «Морские исследования и рациональное природопользование» (г. Севастополь, 19–23 сентября 2018 г.); научно-практической школы-конференции «Наземные и морские экосистемы Причерноморья и их охрана» (г. Новороссийск, 23–27 апреля 2018 г.); международных научно-технических конференциях «Автоматизация и приборостроение: проблемы, решения» (г. Севастополь, 5–9 сентября 2016 г., 11–15 сентября 2017 г.); Всероссийской научно-практической конференции «Science in Progress» (г. Новосибирск 20 октября 2016 г.); I междисциплинарной Всероссийской научно-практической конференции «Развитие методологии современной экономической науки и менеджмента» (г. Севастополь, 4–5 мая 2017 г.); международной научной конференции «Информатизация непрерывного образования – 2018» (г. Москва, 14–17 октября 2018 г.); IV и V научно-практических конференциях «Экобиологические проблемы Азово-Черноморского региона и комплексное управление биологическими ресурсами» (г. Севастополь, 2–4 октября 2017 г., 8–11 октября 2018 г.); X Всероссийской научно-технической конференции «Робототехника и искусственный интеллект» (г. Железногорск, 8 декабря 2018 г.); I Всероссийской научно-практической конференции «Проблемы оценивания и прогнозирования качества жизни» (г. Севастополь 24 – 28 сентября 2018 г.); I региональной научно-практической конференции «Векторы развития информационных технологий: перспективы и направления» (пгт. Гурзуф 03–05 марта 2017 г.); международной научно-практической конференции

«Экологическая, промышленная и энергетическая безопасность – 2017» (г. Севастополь, 11-15 сентября 2017 г.); международных научно-технических конференциях «Системы контроля окружающей среды – 2017», «Системы контроля окружающей среды – 2018», «Системы контроля окружающей среды – 2019» и «Системы контроля окружающей среды – 2020» (г. Севастополь, 6–9 ноября 2017 г., 5–9 ноября 2018 г., 12–13 сентября 2019 г., 9–12 ноября 2020 г.).

Исследования поддержаны стипендией Президента РФ для аспирантов, обучающихся по образовательным программам, соответствующим приоритетным направлениям модернизации и технологического развития российской экономики (2017–2018 уч. год).

Предложенные автором методы, алгоритмы и программные средства показали свою эффективность при обработке реальных данных, относящихся к акваториям Черного и Азовского морей.

Публикации. По теме диссертации опубликовано **20** научных работ, в том числе: **5** статей в ведущих рецензируемых научных журналах и изданиях, индексируемых в международной наукометрической базе Scopus, **15** статей в ведущих рецензируемых научных журналах и изданиях, рекомендованных ВАК РФ; **7** свидетельств о государственной регистрации программы для ЭВМ (номера свидетельств: № 2020662839, № 2018613184, № 2017664038, № 2018617880, № 2018666181, № 2020663317 и № 2020664094); **1** патент РФ (№ 2020141487); **32** статьи в материалах всероссийских и международных научных и научно-технических конференций. По результатам исследований, представленных в диссертации опубликована монография.

Структура и объем диссертации. Диссертация состоит из введения, четырех глав основной части, заключения, списка литературы из **194** наименований и **6** приложений. Каждая глава разбита на параграфы, включая обзор к главе и выводы к ней. Объем работы **146** страниц текста, в том числе **33** рисунка и **2** таблицы.

Во введении приведена общая характеристика работы, включающая обоснование актуальности темы, определение цели и задач исследования,

излагаются полученные автором основные результаты проведенных исследований, показывается их научная новизна, теоретическая и практическая значимость, отражаются основные положения, выносимые на защиту.

В первой главе представлен анализ процессов формирования и обнаружения аномалий ДРМ ПС. Рассматриваются виды аномалий, присущих таким системам, из которых выделяются два типовых класса: М-аномалии и N-аномалии. Источниками М-аномалий является неопределенность в данных, возникающая из-за ошибок в средствах мониторинга, шумов, вариативности данных и в целом характеризуется свойством самих данных, а не модели. Источниками N-аномалий являются неопределенности в знаниях, возникающие вследствие некорректности данных для обучения или вследствие качественных изменений состояния наблюдаемой системы и в целом характеризует свойство модели, а не данных. Эти аномалии вызваны естественными изменениями в самой среде.

Во второй главе представлены методы и алгоритмы поддержки принятия решений, адаптированные для решения задач по обнаружению аномалий в ДРМ ПС и ПТС. В частности, при использовании обучающей выборки ограниченного объема, задающей скалярное поле системы обнаружения аномалий, вводится этап повышения плотности данных за счет сплайн интерполяции и нелинейной регрессионной модели.

В третьей главе осуществлен синтез модульной системы обнаружения аномалий и представлены результаты исследования функциональных свойств, численного моделирования, валидации и тестирования модулей обнаружения аномалий ДРМ как с использованием данных реальных измерений, так и ИМ с управляемо-сгенерированными аномалиями.

В четвертой главе представлены структуры всех ключевых компонентов программно-технического обеспечения, реализующего разработанные проблемно-ориентированные методы и результаты его практического применения для поддержки принятия решений о наличии аномалий на реальных ДРМ ПС.

В заключении приведены основные результаты и выводы по диссертационному исследованию.

При нумерации формул, рисунков и таблиц первая цифра указывает номер главы, а вторая – порядковый номер формулы, рисунка или таблицы в разделе.

Благодарности. Автор выражает искреннюю признательность своему научному руководителю, д-р техн. наук, проф. А.В. Скаткову. Автор благодарит научного руководителя ИПТС, чл.-корр. РАН, д-р геогр. наук, проф., А.Б. Полонского за поддержку выполненных прикладных исследований. Автор глубоко благодарен коллективу Центра экологического приборостроения и экоэнергетики ИПТС и лично гл. науч. сотр., д-р техн. наук. проф. Н.А. Грекову и гл. науч. сотр., д-р техн. наук. проф. В.А. Гайскому за внимательное отношение и полезные обсуждения результатов работы. Автор признателен сотрудникам кафедры ИТиКС ФГАОУ ВО «СевГУ», заведующему кафедрой канд. техн. наук. доц. А.А. Брюховецкому, за систематическое обсуждение полученных результатов на семинарах и конференциях.

ГЛАВА 1.

АНАЛИЗ АПРИОРНОЙ ИНФОРМАЦИИ ОБ АНОМАЛИЯХ
В МОНИТОРИНГОВЫХ НАБЛЮДЕНИЯХ ПРИРОДНОЙ СРЕДЫ
И ВЫБОР ПЕРСПЕКТИВНЫХ НАПРАВЛЕНИЙ РАЗВИТИЯ
МЕТОДОВ ИХ ОБНАРУЖЕНИЯ

В данной главе содержится анализ современного состояния решения проблем обнаружения аномалий в данных ДРМ ПС, а также оценки их состояния. Анализ выполнен целью типизации возникающих аномалий ПС и постановки задачи построения адаптивно переключаемого комплекса проблемно-ориентированных модулей для обнаружения аномалий каждого рассмотренного типа.

Приведены результаты рассмотрения особенностей существующих решений задач выявления аномалий и результаты проведенных исследований, приводится классификация наиболее распространенных аномалий, возникающих в ДРМ. Особое внимание уделяется классификации аномалий пространственно-временных данных в скалярной и векторной формах, возникающих как в самой среде (N-аномалии), так и в средствах измерения (M-аномалии), и описанию их возможных проявлений по характеристикам ДРМ.

Рассмотренные ниже классические подходы позволяют обнаружить многие типы аномалий в оперативных и ретроспективных ДРМ. Кроме того, частично их функциональность перекрывается. Но, к сожалению, даже все они вместе взятые не позволяют полностью решить проблему целиком. Более того, качество работы систем анализа данных, построенных на их основе, не удовлетворяет актуальным требованиям, предъявляемым к современным СМ природных систем, например, подавляющее большинство из них способны обнаруживать только известные типы аномалий для сбалансированных классов. В то же время, реализованные в некоторых из них алгоритмы обнаружения неизвестных аномалий дают высокий процент ошибок I и, что особенно важно для СМ ПС, II рода.

Показывается фрагментарный характер решения проблемы обнаружения аномалий ДРМ ПС, связанный с отсутствием законченных методик, демонстрирующих возможность их фактического воплощения в автоматизированных системах экологического мониторинга.

Осуществлена спецификация требований к комплексной технологии, ее методической, технической и программной реализации, осуществляющей интеллектуализацию процессов принятия решений о наличии скрытых аномалий в ДРМ.

1.1 Типизация и классификация потенциально возможных аномалий в данных мониторинга природной среды

В процессе мониторинга ПС и ПТС ученые, исследователи и лица принимающие решения по организации экологического контроля и промышленного природопользования сталкиваются с большим разнообразием их потенциально существующих аномальных состояний, более 30% из которых возникают впервые. К источникам таких типов аномалий относятся изменения физико-химических характеристик среды, вызванные факторами естественного, природного и антропогенного характеров, технические и программно-аппаратные сбои приборов и средств контроля ПС и ошибки, обусловленные человеческим фактором.

Среди потенциально существующих типов аномалий ДРМ ПС наибольшую информативность представляют: точечные аномалии, контекстуальные аномалии, аномалии временных рядов, коллективные аномалии, возмущения скалярного и векторного полей, выброс, сдвиг, изменение формы распределения ДРМ, отклонение от «повседневного» (для ДРМ, связанных с сезонностью), поведенческие, совместные аномалии.

Это множество аномалий может быть разделено на классы по степени значимости воздействия на ПС или ПТС, и, как следствие, на человека в целом. При том что само понятие аномалия широко используется в литературе контекст его употребления не имеет четко выраженной системы и требует уточнения. Далее

рассмотрим множество наиболее характерных аномалий, свойственных ДРМ ПС в контексте известных общепринятых определений понятия аномалии из различных источников более подробно. Выделим типовые классы аномалий мониторинга ПС и рассмотрим типовые методы анализа ДРМ, в контексте повышения качества обнаружения таких аномалий.

1.1.1 Двойственная интерпретация аномалий в данных мониторинговых наблюдений природной среды: N-аномалии моделей знаний и M-аномалии данных, оценивание вероятностей правильных и ошибочных решений

Анализируя работы [29–35], посвященные обнаружению аномалий ДРМ ПС, установлены свойства таких аномалий, в соответствии с которыми выделена устойчивая двойственная интерпретация аномалий двух типовых классов, которые назовем условно M-аномалии и N-аномалии.

Источниками аномалий, интерпретируемых как M-аномалии являются неопределенности в данных, возникающие из-за ошибок в средствах мониторинга, сложности структуры данных, шумов и вариативности и в целом характеризуются свойствами самих данных, а не модели. Причинами этих аномалий также являются технические причины, ошибки технических и программных средств, человеческие ошибки, ошибки средств доставки, хранения, преобразования и регистрации данных [36–44].

Источниками аномалий, интерпретируемых как N-аномалии являются неопределенности в знаниях, возникающие вследствие того, что данные для обучения и для теста взяты из выборок с различными распределениями. Такие аномалии могут проявляться если наблюдаемая система качественно изменила свое состояние, вследствие недостатка объёма и вариативности данных обучающей выборки, и в целом характеризуют свойство модели, а не данных. Эти аномалии вызваны естественными изменениями в самой среде [45–53].

Определение типа аномалий в рядах наблюдений, а также их интерпретацию осуществляет ЛПП. Методическое, алгоритмическое и приборное обеспечение обнаружения аномалий для обоих типов задач во многом совпадают [54–63].

Иногда высказывается мнение, по недоразумению, что в природных (естественных) средах не могут присутствовать аномальные значения поскольку эти системы – естественны. Доминирующая точка зрения, принятая в настоящей работе проиллюстрирована содержанием рисунка 1.1, где под обнаружением аномальных значений понимается поиск в ряде наблюдений над естественными средами аномальных значений ДРМ, т.е. речи о «противоестественных» состояниях быть не может.



Рисунок 1.1. Логическая структура проблемы обнаружения аномалий антропогенной и естественной природы по ДРМ ПС

Осуществим формализованную постановку задачи, иллюстрированной структурой на рисунке 1.1.

Оценка вероятностей правильных и ошибочных решений при принятии статистической гипотезы о наличии или отсутствии аномалии в данных мониторинга ПС согласно принципу максимума правдоподобия

Рассмотрим задачу обнаружения аномалий как проверку простой гипотезы против простой альтернативы. При ее решении изначально задан X – набор ДРМ ПС и известно что ему соответствует одна из двух функций правдоподобия $W(X | A)$ или $W(X | 0)$, связанных с взаимоисключающими событиями A – наличия аномалий, и 0 – отсутствия аномалий. Поскольку ДРМ x_1, x_2, \dots, x_n – случайные величины, то и сам набор X должен быть случайным и характеризоваться плотностью вероятностей $W(X)$.

Обозначим через H_1 и H_0 гипотезы о том, что набор ДРМ X определяется функциями правдоподобия $W(X | 0)$ и $W(X | A)$ соответственно, а через γ_1 и γ_0 – решения, состоящие в принятии или отклонении гипотезы H_1 . Гипотеза H_0 является простой альтернативой H_1 .

Тогда правило принятия решения согласно принципу максимума правдоподобия примет вид

$$\Lambda = \frac{W(X | A)}{W(X | 0)} \begin{matrix} > \\ < \end{matrix} \begin{matrix} H_1 \\ H_0 \end{matrix} C(x_1, x_2, \dots, x_n) \begin{matrix} \rightarrow \gamma_1 \\ \rightarrow \gamma_0 \end{matrix}, \quad (1.1)$$

где $C(x_1, x_2, \dots, x_n) = const$ – уравнение поверхности в пространстве наблюдений X , разделяющей это пространство на две области Γ_1 и Γ_0 .

В том случае, если элемент X , характеризующийся функцией правдоподобия $W(X | A)$, попадает в область Γ_1 пространства наблюдений, то принимается

решение γ_1 о наличии аномалии в ДРМ ПС (истинности гипотезы H_1). Если элемент X , характеризующийся функцией правдоподобия $W(X|0)$, попадает в область Γ_0 пространства наблюдений, то принимается решение γ_0 об отсутствии аномалии в ДРМ ПС (отклонении гипотезы H_1).

Выражение (1.1) представляет собой правило принятия решения, в левой части которого указывается оптимальная метрика оценки набора данных при условиях наличия возмущения, а в правой части неравенства определяется принимаемое решение в соответствии с выбранной стратегией поведения. Применение любой заранее установленной стратегии принятия решения в силу случайной природы ДРМ неизбежно приводит к возникновению ошибок I и I рода.

Исходя из приведенных выше обозначений получим возможные ошибки двух видов.

Ошибка I рода, или ложная тревога, возникает если элемент X , характеризующийся функцией правдоподобия $W(X|0)$, попадает в область Γ_1 пространства наблюдений и тем самым принимается гипотеза H_1 , т.е. решение γ_1 об возникновении аномалии когда в действительности верна гипотеза H_0 . Данная ошибка характеризуется как возмущения в состоянии ПС с точки зрения модели, принятой наблюдателем, но являющаяся естественной для ПС и вызванная скорее недостатком данных о среде.

Ошибка II рода, или ложный отбой, возникает если элемент X , характеризующийся функцией правдоподобия $W(X|A)$, попадает в область Γ_0 пространства наблюдений и тем самым принимается гипотеза H_0 , т.е. решение γ_0 об отсутствии аномалии когда в действительности верна гипотеза H_1 . Данная ошибка характеризуется событием пропуска нетипового аномального состояния ПС N-аномалия, когда он в действительности имеет место. Подобное может быть вызвано недостаточным объемом данных для обучения классификатора,

неправильно подобранным математическим аппаратом или ошибками в самих средствах измерения М-аномалий.

В случае наличия некоторого объема размеченных ДРМ ПС (обучающей выборки), в которой имеются прецеденты нормальных и аномальных данных выражения для условных вероятностей ошибок принятия решений и правильных решений примут вид.

Условная вероятность ошибки I рода, вероятность ложной тревоги, уровнем значимости равны

$$\alpha = P(\gamma_1 | H_0) = \int_{\Gamma_1} W(X | 0) dX. \quad (1.2)$$

Условная вероятность ошибки II рода, вероятность ложного отбоя, вероятность пропуска равны

$$\beta = P(\gamma_0 | H_1) = \int_{\Gamma_1} W(X | A) dX. \quad (1.3)$$

Условная вероятность правильного принятия решения, состоящего в принятии гипотезы H_1 об отсутствии аномалии, мощность принятия решения равны

$$P(\gamma_1 | H_1) = \int_{\Gamma_1} W(X | A) dX = 1 - \beta. \quad (1.4)$$

Условная вероятность правильного принятия решения, состоящего в принятии гипотезы H_0 о наличии аномалии равна

$$P(\gamma_0 | H_0) = \int_{\Gamma_0} W(X | 0) dX = 1 - \alpha. \quad (1.5)$$

Предлагаемый в работе проблемно-ориентированный лабораторный стенд, позволяющий на фоне типового набора данных генерировать управляемые возмущения по сути дает инструментарий, позволяющий задавать априорные

вероятности наличия P_0 и отсутствия P_1 аномалий в ДРМ ПС, тогда становится возможным найти априорные вероятности принятия решений по формулам

$$P(\gamma_1) = P_1 P(\gamma_1 | H_1) + P_0 P(\gamma_1 | H_0) = P_1(1 - \beta) + P_0\alpha, \quad (1.6)$$

$$P(\gamma_0) = P_0 P(\gamma_0 | H_1) + P_1 P(\gamma_0 | H_0) = P_0(1 - \alpha) + P_1\beta. \quad (1.7)$$

Эти вероятности определяют частоты появления отдельных решений в повторяющемся процессе принятия решений по ДРМ. Таким образом для интеллектуальной системы становится возможным автоматически выбирать оптимальное правило принятия решения из имеющегося пула моделей, наилучшее для каждого конкретного случая, путем осуществления их полного перебора в ИМ по критерию минимума нормы ошибок принятия решения

$$P_i = P_0\alpha + P_1\beta. \quad (1.8)$$

Иногда, тем не менее, аномальные события остаются своевременно необнаруженными. В частности, это бывает, когда они кратковременны или СМ недостаточно эффективна по ряду других объективных причин. В связи с этим по-прежнему актуальными являются работы, направленные на повышение достоверности обнаружения аномалий, возникающих кратковременно в реальном времени и тем самым снижению вероятностей ошибочных принятий решений, т.е. снижению уровня ошибок I и II рода.

Повышение оперативности и снижение среднего числа ошибок можно достигнуть различными методами. В первую очередь это можно обеспечить за счет увеличения объема измерительной информации, этот путь не всегда эффективен и как правило затратен. Вторым направлением является принятие эффективных организационно-методических мероприятий по их достижению. Третий осуществляется на основе совершенствования математического аппарата и информационных технологий, в том числе и на основе развития специальных подходов, например, за счет совершенствования оценки градиентов, применении методов цифровой фильтрации, разложения функций в ортогональных базисных,

вейвлет-преобразований, использования других интеллектуальных подходов, вовлечения ЛПР.

Ввиду важности и комплексности вышеуказанных методических и прикладных задач обнаружения аномалий в ДРМ в научном и прикладном, экономическом и экологическом плане становится очевидным необходимость их дальнейшего изучения. Особенно следует отметить необходимость учета вариативности задачи обнаружения аномалий в зависимости от целей ЛПР, природы решаемой задачи, в соответствии с которыми пул эффективно применяемых моделей и методов может меняться. Детальное рассмотрение типовых аномалий ДРМ и адаптация интеллектуальных методов анализа, многокритериального оценивания данных, адаптивной цифровой фильтрации, повышения размерности данных в рамках комплексного проблемно-ориентированного системного решения позволит повысить степень обоснованности при принятии решений о наличии аномалий в ДРМ сложных систем различной природы.

В рамках выделенных типовых классов М-аномалий и N-аномалий при практических реализациях СМ ПС, несмотря на огромное разнообразие потенциально существующих аномалий в них, тем не менее, можно выделить типовые аномалии и разработать для них специальных проблемно-ориентированный математический аппарат, адаптивно переключая который можно построить систему поддержки принятия решений, наилучшим образом обнаруживающую каждый из рассмотренных типов аномалий.

1.1.2 Формализация и спецификация постановок задач обнаружения потенциально возможных аномалий в данных мониторинговых наблюдений природной среды

Как правило, современные тенденции решения задач обнаружения аномалий в мониторинговых наблюдениях окружающей среды, ПС и ПТС предусматривают активное применение методологического аппарата математической статистики, численного имитационного моделирования, и интеллектуальных обучаемых

методов. Среди новых направлений следует выделить интеллектуальные адаптивные решения, осуществляющие декомпозицию с последующей спецификацией задач обнаружения аномалий в целом, на подзадачи обнаружения аномалий конкретных видов используя в каждом специфичном случае наиболее подходящую для этого, ориентированную математическую модель. Решение части из этих задач тривиально и рассмотрено в соответствующей литературе, в частности, что касается детектирования точечных выбросов, в то время как значительная часть задач обнаружения аномалий более специфических видов, свойственных ДПР ПС, еще не рассматривались или не имеет замкнутого решения, или же должна быть рассмотрена для полноты.

В контексте прикладного использования приборов и средств контроля ПС часто возникают ситуации, когда возмущения в скалярных полях мониторинга ДРМ, при определенных условиях, могут являться как новыми знаниями о наблюдаемом объекте (N-аномалии), так и ошибками изменений (M-аномалии). Задачи обнаружения таких аномалий в ДРМ ПС подчиняются определенной системе, для определенности сформулируем названия этих задач в соответствии с видами аномалий, которые в них выявляются: порядковая, ранговая, формы распределения, положения, волатильности, трендов, градиентов, дивергенций, моделей.

Формализация постановок задач обнаружения аномалий

Специальная постановка задач поиска аномалий в ДРМ частично схожа с классической задачей кластеризации, с тем отличием, что размеры кластеров несимметричны, имеется малый объем данных о аномальных прецедентах, формулировка аномалии субъективна зависит от ЛПР, возможны возмущения как в самих данных M-аномалии, так и моделях данных среды N-аномалии.

Осуществим оригинальную постановку задач для ряда типовых видов аномалий ПС построив детекторы, которые непротиворечиво образуют комплекс, лежащий в основе СППР. Система строится так, чтобы наилучшим образом обнаруживать каждый из рассмотренных типов аномалий, путем адаптивного переключения ведущей модели, в зависимости от типа обнаруживаемой аномалии.

Дано:

X – пространство ДРМ ПС объемом N ;

X^* – пространство ДРМ ПС в котором отсутствуют М-аномалии объемом N

;

$Y \in [X_N, X_A, X_B]$ – множество возможных решений классификатора по ДРМ ПС (подмножество вариантов нормы, подмножество N-аномалий, подмножество М-аномалий);

$\rho: X \times X \rightarrow [0, \infty)$ – мера различия между записями x_i и x_j , удовлетворяющая условию метрик, ориентированных на измерение межклассовых расстояний

$$\left\{ \begin{array}{l} \rho(x_i, x_j) \geq 0 \\ \rho(x_i, x_j) = 0 \Rightarrow x_i = x_j \\ \rho(x_i, x_j) = 0 \Leftarrow x_i = x_j \\ \rho(x_i, x_j) = \rho(x_j, x_i) \\ \rho(x_i, x_j) \leq \rho(x_j, x_k) + \rho(x_k, x_j) \end{array} \right. ; \quad (1.9)$$

$\nu^* \in \nu: X \rightarrow Y$ – обучаемая модель, осуществляющая отображение каждой записи x_i на одно из состояний множества Y , выбираемая из семейства параметрических функций \mathcal{V} .

Найти:

1. Порядковая аномалия – возникновение возмущений в упорядоченной выборке $sort(X)$ в виде крайних элементов $sort(X)_0 = x_{\min}$ и $sort(X)_n = x_{\max}$ принадлежащих или не принадлежащих генеральной совокупности N , которые подозреваются на аномальность (для квартилей имеет вид)

$$M_1 = \begin{cases} \nu^*: X \rightarrow \begin{cases} X_N \in x^*_i \mid \left| i - \frac{N}{2} \right| \leq \frac{N}{4} \\ X_{A,B} \in x^*_i \mid \left| i - \frac{N}{2} \right| > \frac{N}{4} \end{cases} \\ \text{sort} : X \rightarrow X^* \mid \forall \rho(x^*_{i-1}, x^*_i) \geq 0, X^* \in X \end{cases} . \quad (1.10)$$

При такой постановке в задаче под подозрение на аномальность попадают в первую очередь крайние элементы вариационного ряда, сама задача редуцируется до рассмотрения только крайних элементов выборки из первого и четвертого квартиля как возможно содержащих аномалии. Такой подход даст меньше ошибок II рода для симметричных распределений близких к нормальному (эксцесс которых меньше либо равен нулю).

2. Ранговая аномалия – при построении гистограммы ДРМ $hist(X)$ из k кластеров, каждый разряд которой имеет ранг по частоте попаданий в этот разряд, наибольшая частота соответствует максимальному рангу, а наименьшая – возможной аномалии

$$M_2 = \begin{cases} \nu^*: X \rightarrow \begin{cases} X_N \in x^*_i \mid x_i \in P^{\max} \\ X_{A,B} \in x^*_i \mid x_i \in P^{\min} \end{cases} \\ \text{step} = \frac{x_{\max} - x_{\min}}{k} \\ \text{hist} : X \rightarrow [P_0, P_1, \dots, P_k], P_k = \frac{|X^*_k|}{N} \mid \forall (k-1)\text{step} \leq x^*_k < k \cdot \text{step}, X_k \in X \end{cases} . \quad (1.11)$$

Такая постановка задачи, в отличие от предыдущей, обладает значительно большей адекватностью для смешанных многомодальных распределений, что также встречается в мониторинговых исследованиях сложных систем с несколькими устойчивыми состояниями. Кроме того, позволяет относительно невысокими вычислительными ресурсами осуществлять проверку адекватности измерительных каналов средства мониторинга что в совокупности повышает информационную надежность приборов и средств контроля ПС в целом.

3. Аномалия формы распределения – тип аномалий, связанный с предыдущим, но по существу это аномалия функции распределения. Например, при сравнении двух функций распределения, полученных по двум выборкам, проверяется гипотеза их принадлежности одной генеральной совокупности. Различие функций распределения является признаком аномальности данных второй выборки относительно первой – контрольной выборки (в случае использования информационной метрики Шеннона имеет вид)

$$M_3 = \begin{cases} \nu^*: X \rightarrow \begin{cases} X_N \mid IS(hist(X), hist(X^*)) \geq 1 \\ X_{A,B} \mid IS(hist(X), hist(X^*)) < 1 \end{cases} \\ IS(p, p') = \left| \sum_{i=1}^n S\left(\frac{p_i}{N}\right) - \sum_{i=1}^n S\left(\frac{p'_i}{N}\right) \right| \\ S(p) = -\sum_{i=1}^n \frac{p_i}{\log p_i} \end{cases} . \quad (1.12)$$

Использование информационной метрики (в частности изменения энтропии Шеннона или дивергенции Кульбака-Лейблера) по сравнению с классическим статистическим подходом позволяет получить значительно более выпуклую оценку изменения формы распределения что позволит на значительно более ранних этапах определять граничные предаварийные состояния ПТС до того, как негативные изменения начнут приобретать системный характер.

4. Аномалия положения – выражается в присутствии сдвига математического ожидания выборок \bar{X} и \bar{X}^* при неизменном законе распределения (в случае t-критерия для двух независимых выборок для критерия значимости $\alpha = 0.95$ имеет вид)

$$M_4 = \begin{cases} \nu^*: X \rightarrow \begin{cases} X_N | H_0: \bar{X} = \bar{X}^*, t(X, X^*) \geq t_{(2N-2)} | \alpha = 0.95 \\ X_{A,B} | H_1: \bar{X} \neq \bar{X}^*, t(X, X^*) < t_{(2N-2)} | \alpha = 0.95 \end{cases} \\ t(X, X^*) = \frac{\bar{X} - \bar{X}^*}{\sqrt{\frac{(N-1) \sum_{i=1}^N (x_i - \bar{X})^2 + \sum_{i=1}^N (x_i^* - \bar{X}^*)^2}{2N-1}}} \succ t_{(2N-2)}. \end{cases} \quad (1.13)$$

Данная формулировка задачи аномалий является классической и рассматривается для полноты последующего полимодельного комплекса. Использование t-критерия Стьюдента или же F-критерия Фишера с высокой достоверностью позволяет осуществлять определение факта того что контрольная и тестовая выборки принадлежат одной и той же генеральной совокупности. К недостаткам такого подхода стоит отнести необходимость сбора достаточно большой статистики для принятия решения с высокой долей уверенности. Помимо этого, как правило критерии такого типа строятся в предположении о нормальном распределении сравниваемых величин, что на практике не всегда имеет место.

5. Аномалия волатильности – выражается в наличии значимого различия выборок по критерию дисперсии

$$M_5 = \begin{cases} \nu^*: X \rightarrow \begin{cases} X_N | H_0: \sigma(X) = \sigma(X^*) \\ X_{A,B} | H_1: \sigma(X) \neq \sigma(X^*) \end{cases} \\ \sigma(X) = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{X})^2}{N-1}} \end{cases} \quad (1.14)$$

При такой формулировке задачи можно получить удовлетворительное решение задачи обнаружения аномалий, например, в случае исследования полезного сигнала на фоне помех. Любое изменение волатильности параметров для естественных систем как правило является значимым предиктором аномальных изменений либо в самой системе, либо, что более вероятно, в самих средствах измерений.

Для более сложных многопараметрических процессов, характеризующихся не простой выборкой, а матрицей, описывающей ДРМ имеющие временной или пространственный характер, речь идет о полях мониторинга и случайных функциях, когда рассматриваются аномалии вида:

6. Аномалии трендов – выражается в изменении сезонности показателей, свойственных ПС без явных на то предпосылок (значимому изменению коэффициентов уравнения регрессии при замене обучающей выборки на тестовую). По сути проверяется гипотеза о инвариантности регрессионной модели, выбранной из пространства параметрических функций к изменению вектора параметров контрольного множества ДРМ с тестовым

$$M_6 = \begin{cases} v^*: X \rightarrow \begin{cases} X_N | D(X, X^*) < D_{крит} \\ X_{A,B} | D(X, X^*) \geq D_{крит} \end{cases} \\ D(X, X^*) = \sum_{i=1}^t f^*(i, a_1, a_2, \dots, a_t) - f(i, a_1, a_2, \dots, a_t) \rightarrow \min. \\ f(i, a_1, a_2, \dots, a_t) = x_i \\ f^*(i, a^*_1, a^*_2, \dots, a^*_t) = x^*_i \end{cases} \quad (1.15)$$

При том что тренд ДРМ в большей степени уникален, обладает пространственной и временной изменчивостью, в целом он как правило описывается линейной, степенной или периодической функцией с дополнительными ограничениями на минимальное и максимальное значения описываемого этой функцией поля. В случае если данные контрольной и тестовой выборок взяты из одной генеральной совокупности, то описываемые их модели должны быть инвариантны к замене вектора параметров из одной модели в другую.

7. Аномалии градиентов – проявляются в отклонении градиентных полей наблюдаемой среды относительно ее типового состояния и являющейся достаточно инертной характеристикой (пусть $X_{i,j}$ определен в двумерном пространстве, и градиентом дифференцируемой в области определения I, J скалярной функции

$$X(i, j) = X_{i,j}, \text{ координат } i, j \text{ будет векторная функция с компонентами } \frac{\partial x}{\partial i} \text{ и } \frac{\partial x}{\partial j}$$

$$M_7 = \begin{cases} v^*: X \rightarrow \begin{cases} X_N | \iint_{i,j} (\text{grad}(X) - \text{grad}(X^*)) d(i,j) \approx 0 \\ X_{A,B} | \iint_{i,j} (\text{grad}(X) - \text{grad}(X^*)) d(i,j) \neq 0 \end{cases} \\ \text{grad}(X) = \frac{\partial x}{\partial i} di + \frac{\partial x}{\partial j} dj \end{cases} \quad (1.16)$$

Такая постановка задачи позволяет получить значительно большую сходимость результатов в данных пространственного мониторинга при сохранении высокой чувствительности классификатора. Например, по сравнению со скалярными полями данных температуры, их градиенты как правило обладают гораздо большей стабильностью показателей. С другой стороны, точное построение градиентного поля накладывает повышенные требования к данным скалярного поля, получения которых на практике сопряжены с высокими стоимостными затратами, поэтому для их построения используют объединенные данные спутникового мониторинга с измерениями сетью плавучих и стационарных зондов.

8. Аномалии дивергенций – возникновение значимого изменения интегрального информационного параметра, отображающего векторное поле ДРМ на скалярное, т.е. оценки изменения мощности источника векторного поля

$$M_8 = \begin{cases} v^*: X \rightarrow \begin{cases} X_N | \text{div}(X) - \text{div}(X^*) \approx 0 \\ X_{A,B} | \text{div}(X) - \text{div}(X^*) \neq 0 \end{cases} \\ \text{div}(X) = \frac{\partial \text{grad}(X)_i}{\partial i} + \frac{\partial \text{grad}(X)_j}{\partial j} dj \end{cases} \quad (1.17)$$

При том, что градиентные поля являются достаточно информативными при поддержке принятия решений о наличии аномалий в данных пространственного мониторинга, решение задачи многопараметрического сравнения полей осложнено ограничениями когнитивных способностей человека работать с многомерными данными. В связи с чем, оценка дивергенции скаляра позволяет свести меру различия полей к единому численному критерию, что увеличивает возможности ЛПР по принятию обоснованных управленческих решений на основе ДРМ.

9. Аномалия статистических характеристик адекватности показателей результатов имитационного моделирования по численным регрессионным моделям и данных фактических измерений

$$M_9 = \begin{cases} v^*: X \rightarrow \begin{cases} X_N | D(X, X^*) < D_{1крит} \\ X_{A,B} | D(X, X^*) \geq D_{1крит} \end{cases} \\ D_1(X, X^*) = \sum_{i=1}^t (stat(X)_i - stat(X^*)_i)k_i \rightarrow \min \\ stat(X) = [\bar{X}, \sigma(X), M(X), V(X), As(X), Ex(X), \dots] \end{cases}, \quad (1.18)$$

где $\bar{X}, \sigma(X), M(X), V(X), As(X), Ex(X)$ – среднее выборочное, среднее квадратичное отклонение, медиана, коэффициент вариации, коэффициент асимметрии и коэффициент эксцесса для выборки X соответственно, k_i – коэффициент значимости статистических характеристик для итоговой метрики.

В такой постановке задачи осуществление непрерывного сравнения объекта мониторинга с результатами имитационного моделирования этого же объекта позволяет как определять аномальные возмущения в состоянии объекта, так и осуществлять прогнозирование пессимистичных, оптимистичных и реалистичных последствий принятия того или иного решения по объекту мониторинга и осуществлять оценку возникающих при этом рисков.

Каждая из приведенных выше математически моделей проблемно-ориентирована на обнаружение конкретных типов аномалий ДРМ, свойственных ПС, но не является исчерпывающей, в связи с чем можно предположить, что совокупное их использование позволит комплексно решить задачу исследования для представленных выше типов аномалий.

Решение задачи обнаружения аномалий ДРМ принципиально неоднозначно, и тому есть несколько причин:

– не существует однозначно наилучшего критерия качества обнаружения аномалий. Известен целый ряд эвристических критериев, а также ряд алгоритмов, не имеющих чётко выраженного критерия, но осуществляющих достаточно

разумное обнаружение аномалий в ДРМ «по построению» и все они могут давать разные результаты,

– число возможных классов аномалий данных, как правило, неизвестно заранее и устанавливается в соответствии с некоторым субъективным критерием,

– результат принятия решений о наличии аномалии существен и зависит от метрики, выбор которой, как правило, также субъективен и определяется экспертом.

В связи с этим, необходимо системно исследовать предлагаемые в литературе понятия аномалии и в соответствии с выработанным типовым общепринятым определением, осуществлять строгую постановку задачи обнаружения аномалий ДРМ.

Анализ понятий аномалий в данных мониторинга

Само понятие аномалии имеет достаточно большое разнообразие различных понятий в литературе и зависит, в том числе, от области применения данного термина. Некоторые, наиболее значимые и общепринятые, из таких понятий, которые были даны в различных справочниках и энциклопедиях [12–24], сгруппированные по степени схожести в соответствии с решаемой задачей приведены в приложении Г. Особенно стоит отметить понятие аномалий в природе, приведенное в энциклопедическом словаре Ф.А. Брокгауза и И.А. Ефрона: «Аномалия, (греч. *anomalía*) так называется отступление или уклонение от правила, поэтому аномальным называют все отступающее или уклоняющееся от правильного или нормального. Аномальными в области природы считаются такие явления, которые вопреки законам природы представляются исключениями; но несмотря на это, все аномальное не лишено законности; более глубокое созерцание законов природы приводит к тому, что на кажущихся аномалиях замечается отпечаток общей законности. Каждый поэтому может ожидать встретить тем более аномалий, чем менее он знаком с общими законами явлений природы. В грамматике аномальными называются те формы слов, которые уклоняются в своей флексии от формы, считающейся правильной. Древняя грамматика была очень богата

аномальными, т. е. неправильными, формами. Но после того, как языки были исследованы по их фонетическим законам и для уяснения аномальных форм явилось на помощь сравнительное языкознание, понятие о неправильности стало крайне ограниченным. Многие, что раньше считалось аномальным, оказалось в большинстве случаев лишь скудно удержавшимся остатком первоначальной, совершенно правильной формы, или фонетическими законами выяснилась необходимость и правильность формы, на первый взгляд отступающей от правила» [12].

Исходя из множества существующих понятий аномалии, в контексте задачи обнаружения аномалий ДРМ ПС, можно выделить доминирующие тенденции. Понятие аномалии данных определяется: с одной стороны, в виде статистически различимой ошибки, выброса, несвойственного поведения системы, и с другой стороны, как исключения – события отличающегося от ожидаемого в рамках математической модели (и принятой ЛПП картиной мира), но тем не менее при этом точно отражающего изменение фактического состояния природной среды, и поэтому, возможно, представляющего собой новое знание.

В связи с вышесказанным в математической постановке задачи аномалиями ДРМ являются возмущения в факторном пространстве скалярных и векторных полей, вызванных как ошибками в средствах измерения, так и изменениями состояния исследуемой среды. Анализ численных и аналитических методов обнаружения аномалий в такой формулировке представляет наибольший интерес в контексте решаемой задачи исследования.

1.2 Системный анализ существующих методов обнаружения аномалий в данных статистических исследований природной среды

В настоящий момент существует несколько теоретико-практических подходов для решения задачи обнаружения аномалий ДРМ. Особенностью таких данных является их пространственная или временная изменчивость, и как правило многопараметрический характер. Кроме того, в контексте решаемой задачи,

актуальным становится обнаружение аномалий градиентов полей данных, расширения гибкости средств визуализации для повышения обоснованности принимаемых решений и автоматизация решения задачи понятия типа аномалий для выбора наиболее подходящей математической модели, направленной на её решение.

Исследования, посвященные численному анализу состояний, динамики и периодичности показаний ДРМ ПС насчитывают достаточно продолжительную историю, которые берут свое начало одновременно с появлением первых инструментальных метеорологических наблюдений на постоянной основе. В России инструментальный мониторинг физических параметров атмосферы на регулярной основе начал проводиться в Петербурге в 1725 году академиком Ф.Х. Майером. На сегодняшний день существуют БД, содержащие ДРМ температуры воздуха в Москве начиная с 1775 года. В 1834 году в Петербурге под руководством академика А.Я. Купфера была создана Нормальная магнитно-метеорологическая обсерватория, на базе которой в 1849 году была учреждена Главная физическая обсерватория, в функции которой входило руководство всеми метеорологическими и магнитными наблюдениями России по единым методикам и программам [72].

Специалистами в области мониторинговых наблюдений, занимающихся вопросами обнаружения аномалий, достигнуты значительные результаты при решении подобных задач, наиболее часто применяются подходы, использующие:

1. Аналитические, статистические и вероятностные методы.

- скрытые марковские модели [86, 87, 7, 8],
- анализ плотности и законов распределений [2],
- обнаружение выбросов на основе корреляции [3].

2. Численные и модельные методы.

- ассоциативные правила [10],
- опорные векторы [4],
- кластерный анализ [88, 9].

3. Интеллектуальные, адаптивные и обучаемые методы.

- искусственные нейронные сети [5],

- байесовские сети [6],
- нечёткая логика [11].

Часть из них (статистические параметрические, непараметрические, вейвлет преобразования, аналитические, модельные и численные) нашли свое применение в современных системах анализа ДРМ ПС, в отношении других ведутся работы по поиску возможностей их применения для обнаружения аномалий ПС и ПТС.

1.2.1 Особенности процедур обнаружения аномалий в данных мониторинга природной среды аналитическими методами

Известны различные попытки аналитические решения задач анализа ДРМ с целью обнаружения в них аномалий (М-аномалий) и новизны (N-аномалий). Необходимо в первую очередь выделить фундаментальные исследования проведенные такими учеными как Колмогоров А.Н. и Бондур В.Г. [9, 25–28]. Академиком Колмогоровым А.Н. предложена математическая модель, направленная на обнаружение аномалий в наборах данных, предполагающая асимптотическую сходимость результатов мониторинга при условии, что объемы выборок стремятся к бесконечности. В ряде работ Смирнова Н.В. и Королюка В.С. решается схожая задача, при этом анализ ДРМ рассматривается как частный случай непараметрической задачи статистики.

Задачи обнаружения аномалий в выборках данных успешно решаются методами математической статистики только в простейших случаях. Аномалии мониторинговых наблюдений природной среды гораздо шире, потому как элементами таких выборок могут быть векторы, матрицы, изображения, записи нерегулярной структуры и т.д.

Теоремы, установленные авторами, дают решение важных задач, направленных на обнаружение аномалий и нашли широкое применение на практике в качестве критериев согласия [25]:

1. Имеются результаты независимых испытаний x_1, x_2, \dots, x_n над случайной величиной ξ , имеющей непрерывное распределение вероятностей. Ставится задача поиска такого критерия согласования теоретического предположения, что распределение ξ равно $F(x)$, с полученными опытными данными, который бы не зависел от частных предположений об аналитическом характере функции $F(x)$.

2. Имеются две последовательности результатов независимых испытаний

$$\begin{matrix} x_1, x_2, \dots, x_n \\ y_1, y_2, \dots, y_m \end{matrix}, \quad (1.19)$$

произведенных над случайными величинами ξ и η . Ставится задача поиска такого критерия проверки того, что обе величины ξ и η имеют одно и то же непрерывное распределение вероятностей $F(x)$.

Отметим основные результаты, полученные в указанных направлениях, с использованием обозначений, введенных Королюком [25]:

$$D = \sup_{-\infty < x < +\infty} |S_n(x) - F(x)|, \quad (1.20)$$

$$D_n^+ = \sup_{-\infty < x < +\infty} [S_n(x) - F(x)], \quad (1.21)$$

$$D^+(n, m) = \sup_{-\infty < x < +\infty} [S_n(x) - T_m(x)], \quad (1.22)$$

$$D(n, m) = \sup_{-\infty < x < +\infty} |S_n(x) - T_m(x)|, \quad (1.23)$$

где $S_n(x)$ – эмпирическая функция распределения для последовательности x_1, x_2, \dots, x_n ; $T_m(x)$ – эмпирическая функция распределения для последовательности y_1, y_2, \dots, y_m ; $F(x)$ – функция распределения x_1, x_2, \dots, x_n .

1. Теорема В.И. Гливленко. Какова бы ни была функция распределения $F(x)$ случайной величины ε , при $n \rightarrow \infty$

$$P\{D_n \rightarrow 0\} = 0. \quad (1.24)$$

2. Теорема А.Н. Колмогорова. Какова бы ни была непрерывная функция распределения $F(x)$ при $n \rightarrow \infty$

$$P\{\sqrt{n}D_n < z\} \rightarrow K(z) = \begin{cases} 0 & \text{при } z \leq 0 \\ \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 z^2} & \text{при } z > 0 \end{cases}$$

3. Теорема 1 Н.В. Смирнова. Какова бы ни была непрерывная функция распределения $F(x)$, имеют место соотношения:

$$P\{\sqrt{n}D_n < z\} = 1 - \left(1 - \frac{z}{\sqrt{z}}\right)^n - z\sqrt{n} \sum_{k=r+1}^{n-1} \frac{C_n^k (k - z\sqrt{n})^k (n - k + z\sqrt{n})^{n-k-1}}{n^n},$$

где $r = [z\sqrt{n}]$, $0 < z \leq \sqrt{n}$,

$$P\{\sqrt{n}D_n^+ < z\} \rightarrow K(z) = \begin{cases} 0 & \text{при } z \leq 0 \\ 1 & \text{при } z > \sqrt{n} \end{cases}$$

$$\lim_{n \rightarrow \infty} P\{\sqrt{n}D_n^+ < z\} \rightarrow K(z) = \begin{cases} 1 - e^{-2z^2} & \text{при } z > 0 \\ 0 & \text{при } z \leq 0 \end{cases}$$

Теорема 2 Н.В. Смирнова. Если функция $F(x)$ непрерывна и отношение $\frac{m}{n} = \tau$ остается постоянным ($0 < \tau < +\infty$), то при $n \rightarrow \infty$:

$$\lim P \left\{ \sqrt{\frac{nm}{n+m}} D(n, m) < z \right\} = K(z),$$

$$\lim P\{\sqrt{n}D^+(n, m) < z\} = \begin{cases} 1 - e^{-2z^2} & \text{при } z > 0 \\ 0 & \text{при } z \leq 0 \end{cases}$$

Сформулированные выше теоремы нашли широкое применение в качестве критериев согласия. Однако их использование в практических вопросах не всегда достаточно обоснованно, так как все они носят характер предельных соотношений.

Полученные результаты предполагают выборки неограниченно большого объема что в мониторинговых наблюдениях природной среды никогда не имеет место ввиду высокой стоимости получения таких данных и требований к оперативности принятия решений, а характер асимптотической сходимости не исследован и исследованию аналитическими методами, судя по всему, не подлежит. Поэтому, ввиду вышесказанного, при решении А-задач, использование численных и интеллектуальных методов имеет большой потенциал и позволит принимать решения в условиях, в которых функционируют реальные системы сбора и анализа ДРМ ПС.

В частности, в работах Бондура В.Г. отмечается эффективность использования методов численного моделирования для решения задачи обнаружения аномалий в ДРМ ПС [26–28].

Анализ научных работ авторов, посвященных вопросам обнаружения аномалий в ДРМ, равно как и других отечественных и зарубежных работ показал, что на сегодняшний день задачи анализа ДРМ и обнаружения в них аномалий ДРМ ПС аналитическими методами исчерпывающе и в полном объеме не разрешены. В связи с этим, перспективным видится разработка новых и развитие существующих математических моделей, и в особенности численных методов построения адаптивных интеллектуальных информационных систем обнаружения аномалий в ДРМ.

1.2.2 Алгоритмизация процессов обнаружения аномалий в данных мониторинга природной среды

Существуют различные системы обнаружения рассмотренных выше типов аномалий. Их особенностью является жесткая структура и нацеленность на обнаружение аномалий в больших объемах ретроспективных данных. При этом, по принципу своего действия, они не приспособлены для обнаружения быстропротекающих аномалий в реальном масштабе времени. Ниже представлены наиболее типичные представители класса численных методов и критериев

обнаружения аномалий в ДРМ, используемые в современных системах с указанием основных их особенностей в контексте решаемой задачи исследования:

- пороговые классификаторы и статистические тесты [73–75],
- кластеризация (как задача обучения без учителя) и поиск кластеров минимальной ширины [76–78],
- информационные критерии и приращение энтропии [79],

а также комбинации этих методов при участии ЛПР:

- методы подмены задачи,
- ансамбли алгоритмов,
- обучение с учителем (регрессия, классификация),
- решающие деревья.

Статистические и вероятностные методы

Исследования, посвященные численному анализу состояний, динамики и периодичности показаний ДРМ ПС насчитывают достаточно продолжительную историю. Первые исследования А-задач численными методами были посвящены анализу ДРМ в отрыве от природы их возникновения. Так например в работах [73–75] были предложены математические модели для оценивания степени аномальности ДРМ по значениям отдельных полей данных. Предложенные модели статистической обработки подразумевают расчет информационных характеристик ДРМ, на основе которых вычисляется уровень аномальности, с последующим его сравнением с пороговым уровнем, заданным ЛПР.

Согласно выводам исследователей, предложенные методы дают высокие показатели эффективности обнаружения аномалий ДРМ, но имеют ряд недостатков:

- Требуется предварительно очищенная от аномалий обучающая выборка данных достаточного объема.
- Пороговое значение должно определяться ЛПР. Учитывая, что значения параметров, используемых для вычисления аномалии, сильно зависят от конкретной исследуемой ПС то таких пороговых значений должно быть

определено достаточно большое количество что затрудняет применение таких методов.

– Осуществляется оценка уровня аномальности отдельных точечных ДРМ, которая может быть использована только для обнаружения М-аномалий, в то время как аномалии данных, имеющих пространственный или временной характер, останутся необнаруженными.

Существуют работы в которых решение А-задачи осуществляется методами кластерного анализа, представляя ДРМ как векторы в многомерном пространстве [76–78]. Исследуется эффективность методов кластеризации по разделению набора ДРМ на нормальные и аномальные в условиях отсутствия априорной информации о классах. Предлагаемая авторами методика строится в предположениях:

- объем аномальных ДРМ значительно меньше объема нормальных данных,
- аномальные данные качественно отличаются от нормальных.

Наборы данных в методе представляются векторами в многомерном пространстве, а кластеризация выполняется так чтобы минимизировать сумму внутриклассовых расстояний (между всеми парами векторов одного класса). Таким образом большие кластеры относят к нормальным, а малые к аномальным.

При выполнении непрерывного анализа ДРМ в реальном времени алгоритм предлагается упростить и вычислять расстояния между новыми векторами и существующими центрами кластеров, относя новые вектора либо к уже существующим кластерам, либо образовывать для них новый кластер. В качестве критерия аномальности в данном случае выступает ширина кластера. Основным недостатком такого подхода является сильная зависимость точности классификатора от правильности выбора критерия ширины кластера, при этом методики выбора этого критерия нет. Следует отметить что методы кластеризации при правильно подобранной метрике различия между наборами ДРМ позволяют обнаруживать возмущения в данных при отсутствии априорной информации о характере аномалий.

Оригинальная методика решения А-задач рассмотрена в работе [79], где в качестве критерия аномальности предложено использовать информационные

критерии, в частности информационной энтропии. Выдвинуто предположение что возрастание степени неопределенности мониторируемых параметров и в соответствии с этим энтропии с большой вероятностью может быть связано с возникновением аномальных событий и как следствие служить индикатором возникновения аномалии. Подобные предположения имеют смысл в случае простых ошибок, вызванных неисправностями или отказами в самих средствах мониторинга. Более того, расчет информационных критериев нетребователен к вычислительным ресурсам и может выполняться на базе проблемно-ориентированных микроконтроллеров, интегрируемых в само средство мониторинга обеспечивая тем самым повышение его информационной надежности.

Главной положительной особенностью данного класса методов является их высокая эффективность при решении задач обнаружения М-аномалий в случае наличия достаточного объема ДРМ. Другой их положительной чертой является высокая степень формализации при описании использования, пригодная для реализации в виде проблемно-ориентированного программного обеспечения.

При высокой статистической сходимости получаемых результатов эти методы подходят скорее для решения задачи кластеризации и классификации, при которых аномальные классы заранее известны и определены. В таком случае, системы, построенные на базе вышеуказанных принципов, в комплексе, решают отдельные частные задачи обнаружения аномалий в данных ПС, но не дают исчерпывающего решения.

Несмотря на достаточно обширную методологическую базу, важные методические и прикладные задачи обнаружения аномалий ДРМ ПС существующими численными методами в полном объеме, комплексно, с учетом возможностей современных информационных технологий, не разрешены, в частности:

– недостаточен уровень интеллектуализации систем мониторинга при принятии решений о наличии аномалий,

- недостаточно проработаны методы обобщения слабоструктурированных данных,
- не получили широкого распространения по существу адаптивные методы выбора различных сценариев процесса мониторинга,
- не поставлена и не решена задача построения комплексной характеристики отклонения физико-химических и/или экологических показателей в ДРМ ПС,
- ложная тревога при мониторинге (ошибки I рода),
- пропуск N-событий при мониторинге (ошибки II рода),
- в техническом обеспечении недостаточно представлены автономные и автоматизированные средства мониторинга полей физико-химических и экологических показателей способные осуществлять самопроверку адекватности измеряемых параметров,
- отсутствуют специализированные проблемно-ориентированные программные комплексы, для обнаружения аномальных участков данных при мониторинге,
- мало представлены мобильные приложения, направленные на повышение оперативности принятия решений ЛПР за счет совместного использования мощных визуальных средств и облачных технологий,
- недостаточно развиты средства визуализации слабоструктурированных данных, предназначенные для поддержки принятия решений о наличии аномалий в ДРМ.

В связи с вышесказанным перспективным направлением решения А-задачи видится использование эмерджентной системы, состоящей из комплекса численно-аналитических методов, используемых совместно с интеллектуальными подходами.

1.2.3 Специализированные интеллектуальные информационные технологии процедур обнаружения аномалий для реализации в современных системах мониторинга природной среды

Существует большое разнообразие технологий обнаружения аномалий в данных простых выборок, однако их применение для данных мониторинга ПС ограничено и порождает специфические требования, которые состоят в необходимости анализа данных сложной нерегулярной структуры, представленных в векторной, матричной, графической формах и т.д.

Ввиду того, что отдельными аналитическими и численными методами вся полнота задачи обнаружения аномалий в ДРМ не решается, в современных системах нашли свое применение интеллектуальные информационные технологии, имеющие в своей основе базу знаний (БЗ). В первую очередь они направлены на обнаружение N-аномалий, т.к. M-аномалии успешно обнаруживаются более простыми методами, в особенности численным методом расчета дивергенции Кульбака-Лейблера между типовым профилем показаний измерительного канала и тестовой выборкой.

Базы знаний со статическими паттернами аномальных состояний

Обнаружение N-аномалий может осуществляться по характерным наборам признаков ДРМ ПС записи о которых содержатся в БЗ, содержащие статические паттерны аномалий. Характерным может являться сочетание параметров данных находящихся в конкретных типовых диапазонах. Очевидным недостатком такого метода является способность обнаруживать только известные виды аномалий, отмеченные экспертом как аномальные. Таким образом новые типы аномалий будут оставаться незамеченными. В результате точность обнаружения аномалий системами, реализующими подобные подходы, напрямую зависят от полноты и качества БЗ.

К сожалению, разнообразие потенциально существующих аномалий, возникающих в ПС, достаточно велико что не позволяет гарантировать что полнота

БЗ позволит обнаруживать все возникающие аномалии. Подобный эффект становится более выраженным в связи с непрекращающимся усилением антропогенной нагрузки на ПС, инертность и обилие скрытых внутренних процессов, происходящих в среде увеличивающих долю неопределенности при оценке состояния.

Одним их принципиальных недостатков систем основанных на сравнении характерных признаков данных является высокий уровень ложных срабатываний. При практическом использовании систем экологического мониторинга высокое число ложных срабатываний делает сложной работу эксперта, который вынужден вручную анализировать огромные массивы сообщений о обнаруженных аномалиях, многие из которых таковыми не являются.

Базы знаний с динамическими паттернами нормальных состояний

Альтернативная технология обнаружения N-аномалий осуществляет выявление в ДРМ признаков не свойственных нормальным процессам, протекающим в ПС. При этом под аномалией в таком случае понимается отклонение от нормального состояния среды.

Основной идеей представленной технологии является обучение экспертом системы на множестве входных параметров данных среды, характеризующих процессы, протекающие в ее нормальном состоянии. Каждый такой набор называется паттерном нормального состояния. Отклонение этих параметров является значимым признаком возникновения аномалии.

Выбор параметров паттерна нормального состояния во много определяется экспертом субъективно в зависимости от имеющегося пула данных измерительных каналов и применяемых к ним математических методов.

Эффективность такого подхода определяется такими показателями:

- адекватность выбранной математической модели при описании объекта ПС,
- точность описания выбранной модели, обусловленная наличием достаточного объема данных для корректного ее описания,

- адаптационная гибкость паттернов математической модели к новым, динамически изменяющимся условиям ПС,
- корректностью решающих правил, осуществляющих поддержку принятия решения о наличии аномалии в ПС.

По сравнению с отдельными технологиями, использующими отдельно паттерны аномальных состояний комплексное использование паттернов нормальных и аномальных состояний для ПС является предпочтительной. При том что это увеличивает вычислительную сложность подобных алгоритмов, что успешно компенсируется применением современной высокопроизводительной вычислительной техники, достигается значительное снижение вероятностей ошибок I и II рода.

Обнаружение аномалий данных мониторинга природной среды алгоритмами искусственных иммунных систем

За последние 10–15 лет бурно развивались системы искусственного интеллекта, использующие искусственные иммунные системы либо в своей основе, либо в качестве ключевых узлов. Искусственные иммунные системы успешно применяются для решения задач оптимизации, классификации и идентификации, помимо этого, системы применяются для сжатия информации, кластеризации, обнаружения аномалий, машинного обучения, обработки неструктурированных данных и извлечения информации, компьютерной безопасности и адаптивного контроля [80–84]. В ряде работ отмечается высокий потенциал решения А-задач ПС с применением математических моделей иммунной системы человека [85, 89–91].

Технически, область применения подобных моделей можно расширить на решение А-задачи в ПС, поскольку основным предназначением иммунной системы человека является поиск и реагирование на патогенные частицы, которые до этого не встречались или похожие по своей сигнатуре на антигены с которыми иммунная система боролась ранее, что схоже с постановкой А-задачи.

Наибольший интерес представляют работы [85, 91] в которых авторы рассматривают модель иммунной системы человека в контексте ее адаптации к задаче контроля параметров ПС. Исследование [85] предлагает физическую и концептуальную архитектуру системы обнаружения аномалий, реализующую иммунные принципы обнаружения. Предложенная в [91] модифицированная модель искусственной иммунной системы обладает способностью обучаться новой информации, запоминать ранее полученную информацию и осуществлять распознавание образов и анализ данных на основе принципов биомолекулярного узнавания в высоко распределенной манере.

На основе разработанных математических моделей авторами в работах было показано что адаптация механизмов искусственных иммунных систем за счёт модификации классических математических моделей в соответствии со спецификой искусственных иммунных систем значительно повышает их эффективность при выполнении прогнозирования состояния ПС, решении задачи идентификации, которая заключается в нахождении мощностей источников загрязнения по имеющимся экспериментальным ДРМ ПС.

1.3 Спецификация требований к методической, технической и программной реализации систем мониторинга

Специфика решаемой А-задачи ПС накладывает ряд требования на ее функциональные возможности и техническую реализацию.

Все эти требования представляют собой три группы:

- требования к методам обнаружения аномалий в ДРМ ПС,
- требования к технической реализации модулей решения А-задачи,
- пользовательские требования программной реализации модулей обнаружения аномалий.

В основе предлагаемого решения лежат:

- комплекс моделей (основанных на новой предложенной классификации типов аномалий) направленных на обнаружение аномалий в ДРМ ПС,

- комплекс проблемно-ориентированных программ обнаружения аномалий в ДРМ и адаптивного выбора ведущем модели,
- синтезированный полимодельный комплекс, который объединяет модели, программы и интеллектуальную систему в единое решение, предназначенное для обоснованного обнаружения аномалий в ДРМ.

Разрабатываемая система должна являться новым проблемно-ориентированным инструментальным средством поднятия качества функционирования мониторинговых систем за счет использования в них интеллектуальных информационных комплексов, которые реализуют элементы интеллектуального подхода к адаптивному обучению процессов распознавания. В ней должна использоваться развитая методология на основе применения сценарного подхода, который включает в себя сценарии: минимизации рисков, прямого голосования, минимаксный сценарий с относительными оценками. Комплексное использование предложенного подхода позволит повысить степень обоснованности ЛППР и обеспечить структурную гибкость предложенного решения.


Представим вариант реализации комплексной информационной технологии использующей разрабатываемое решение на рисунке 1.2, с основополагающими компонентами: РСД – распределенная система датчиков (включает в себя сенсоры, СПИ и спутниковые наблюдения), ЛППР – лицо, принимающее решения (оператор), КИТ – комплексная информационная технология,  – мобильная сеть передачи данных (связующее звено всех компонентов).



Рисунок 1.2. Концептуальная схема варианта организации комплексной информационной технологии (информационная система в целом)

Разработанная система должна обладать функциональной полнотой: список задач, возникающих при обнаружении ДРМ ПС полностью покрываться списком методов, технологий и проблемно-ориентированных программ, рассмотренных в дальнейших главах диссертации.

1.3.1 Требования к методам обнаружения аномалий в данных мониторинга природной среды

Требования к методу обнаружения N-аномалий и M-аномалий связаны с требуемыми функциональными возможностями и включают в себя:

- **Требование поддержки полимодельности.** Разрабатываемая методика должна реализовывать пул механизмов обнаружения аномалий, ориентированный на различные математические модели – формализованные представления аномалий различного типа. Реализация этого позволит адаптивно подстраивать

систему для обнаружения аномалий различных в том числе ранее неизвестных типов.

– Требование к адаптивности и функциональной гибкости к динамически изменяющимся условиям ПС и процессам, протекающим в ней. Объективные изменения могут быть связаны с антропогенной деятельностью человека и циклическим характером изменений, протекающих в ПС, которые не были учтены в модели изначально, но обладающих существенным вкладом в мониторируемые параметры ПС.

– Требование к оперативности поддержки принятия решений подразумевает что процесс анализа данных должен генерировать рекомендации до того, как они потеряют актуальность. Это может быть достигнуто на уровне практической реализации путем непрерывного выполнения анализа по мере поступления новых данных в соответствующую модель обнаружения.

1.3.2 Требования к технической реализации модулей обнаружения аномалий

– Модульность системы. Обеспечит относительную независимость компонентов облегчающую интеграцию в уже существующие системы, функциональную масштабируемость и структурную отказоустойчивость.

– Минимизация времени анализа. Выполнение данного требования позволит ЛПР своевременно принимать меры по предотвращению и ликвидации последствий возможных негативных событий.

– Минимизация объемов хранимых ДРМ ПС с минимальными потерями информативности. Соблюдение данного требования позволит снизить избыточность данных, устранить дублирующийся данные и ускорит процесс обнаружения аномалий и уменьшит ограничения к необходимым ресурсам дискового пространства.

– Структурная гибкость и расширяемость. Позволят осуществлять расширение доступного пула моделей и методов обнаружения аномалий в случае возникновения значимых изменений в ПС или расширения количества и разнообразия источников данных в дальнейшем.

1.3.3 Пользовательские требования к программной реализации процедур обнаружения аномалий

– Возможность автоматического и ручного выбора математической модели решения А-задачи. Это сделает процесс обнаружения аномалий и принятия решений на их основе более понятным и обоснованным ЛПР.

– Возможность автоматического обнаружения типовых случаев М-аномалий. Позволит автоматически сообщать ЛПР, когда возмущения в ДРМ ПС явно вызваны некорректностью показаний, передаваемых измерительным прибором, чтобы снизить вероятности совершения ошибок I рода при принятии решений за счет исключения типовых М-аномалий.

– Простота установки и настройки. Позволит упростить процесс внедрения системы в инфраструктуры, обеспечивающие решение А-задач как основной или дублирующий компонент.

– Модульность. Позволит интегрировать предлагаемое программно-техническое обеспечение в виде модуля в уже существующие решения мониторинга ПС и непосредственно в измерительные устройства тем самым увеличить их информационную надежность.

1.4 Выбор перспективных направлений развития процедур модульно-компонентного подхода для обнаружения аномалий в данных мониторинга природной среды

Перспективными направлениями решения А-задач ПС является кластеризация задач обнаружения аномальных значений с последующей целью разработки типовых решений для выделенных классов.

В выделенных классах необходимо предусмотреть ситуации с различными объемами данных, наличием априорной информации о типах аномалий, требованиями к обработке в реальном времени.

Для выделенных классов разработать типовые методы и приборное обеспечение проблемно-ориентированное на решение А-задач.

По результатам проведенных исследований можно выделить следующие выводы о перспективных направлениях дальнейших исследований решения А-задач:

– Необходима адаптация существующих моделей обнаружения аномалий к применению в системах анализа данных с учетом особенностей ДРМ ПС.

– Требуется объединение существующих результатов в единый полимодельный комплекс для формирования целостной методологии обнаружения аномалий ДРМ ПС в отличие от фрагментарных подходов, существующих на сегодняшний день.

– Существует потребность в поиске моделей обнаружения N-аномалий и M-аномалий, поскольку проблема разработки моделей и методов обнаружения аномалий с учетом специфики ПС пока мало изучена.

Осуществим разбиение на кластеры типовых ситуаций и на этой основе представим разработанные методы, направленные на решение А-задач в них:

- наличие ДРМ ПС большого объема,
- нерегулярные и пропущенные данные,
- аномалии известных типов,
- априорно неизвестные аномалии,
- однопараметрический мониторинг,
- многопараметрический мониторинг,
- мониторинг данных распределенных во времени,
- мониторинг данных распределенных в пространстве.

1.5 Выводы по главе 1

В данной главе получены следующие основные результаты:

1. Обоснована актуальность выбранной темы исследования развития процедур обнаружения аномалий в мониторинговых наблюдениях природной среды на основе модульно-компонентного подхода.

2. Предложена типизация аномалий, как N-аномалий моделей знаний и M-аномалий данных, потенциально существующих в ДРМ ПС и выделены их специфические особенности.

3. Введено понятие аномалий ДРМ ПС. Отмечены особенности N-аномалий и M-аномалий, показана значимость решения A-задач.

4. Проведен анализ результатов исследований отечественных и зарубежных ученых. В работах отмечено, что для существующих методов характерен ряд жестких ограничений и ограниченность применения для широкого круга задач обнаружения аномалий в ДРМ ПС в функционально сложных случаях многомерности, нестационарности и недостатка информации:

- как правило существующие результаты ориентированы на выявление аномалий в одномерных выборках и матричных наблюдениях регулярной структуры;

- отсутствует единое представление о сущности аномалий ДРМ ПС, о объекте анализа и параметрах, необходимых для решения A-задач;

- существующие аналитические методы не дают исчерпывающего решения A-задачи так как строятся при достаточно строгих ограничениях, в т.ч. предполагающих наличие выборок больших объемов для достоверного принятия решений, необходимость сбора которых идет в разрез с требованием оперативности.

- многие предлагаемые методы решения A-задач носят скорее эвристический характер и не имеют четких математических обоснований и позволяют обнаруживать только ограниченные типы аномалий известных видов.

5. Математически обоснована эффективность реализации проблемно-ориентированных модулей обнаружения типовых случаев M-аномалий в измерительных приборах на основе информационных и статистических мер, обеспечивающих значительное снижение вероятностей ошибок I рода при принятии решений на основе ДРМ ПС.

ГЛАВА 2.

МЕТОДИЧЕСКОЕ И АЛГОРИТМИЧЕСКОЕ ОБЕСПЕЧЕНИЕ
МОДУЛЬНО-КОМПОНЕНТНОГО ПОДХОДА ОБНАРУЖЕНИЯ АНОМАЛИЙ
В ДАННЫХ МОНИТОРИНГА

Как было сказано в главе 1 на сегодняшний день не существует единой методологии обнаружения аномалий в ДРМ ПС. Так, существующие системы не обладают удовлетворительными способностями по обнаружению известных аномалий и являются неприспособленными к обнаружению ранее неизвестных аномалий. Кроме того, нет проблемно-ориентированных методов, направленных на обнаружения аномалий ДРМ, имеющих пространственное и временное распределение. Анализ исследований отечественных и зарубежных авторов демонстрирует фрагментарный характер решения данной проблемы.

В данной главе предлагается модульная интеллектуальная информационная система решения А-задач. Модели разработанных модулей направлены на решение частных задач, возникающих при принятии решений на основе ДРМПС. Полимодельный комплекс решает задачу обнаружения аномалий в целом, при этом его модульная структура обеспечивает его структурную и параметрическую гибкость. За счет использования предложенного подхода обеспечивается возможность как обнаружения известных аномалий, так и обнаружение новых.

Другой важной особенностью является реализация принципа адаптивности СППР к динамически изменяющимся состояниям ПС и широкого разнообразия возможных интерпретаций аномалий как ошибок или новизны.

Предложенная модульная система включает в себя следующие компоненты:

- обнаружения аномалий нерегулярных и пропущенных ДРМ с использованием регрессионных уравнений и сплайн интерполяции [169, 170],
- обнаружения аномалий данных многопараметрического мониторинга методами параметрической статистики [171, 172],

- обнаружения аномалий данных скалярных и векторных полей мониторинга методами непараметрической статистики [173, 174],
- обнаружения аномалий многослойных полей данных с использованием сверточных вейвлет функций [175–177],
- обнаружения аномалий в данных о состоянии системы при наличии обучающей выборки большого объема на базе агентного подхода [178–180],
- обнаружения аномалий ДРМ как оценки дивергенции с использованием сценарного подхода [181–183],
- полимодельный динамически конфигурируемый комплекс обнаружения аномалий произвольного типа [184–187].

2.1 Иерархический метод аналитического восстановления нерегулярных и пропущенных данных мониторинга в реальном времени

Метод решения А-задачи при нерегулярных и пропущенных данных включает в себя иерархическую процедуру подбора такого регрессионного или сплайн-интерполяционного уравнения, которая по неполным, разрозненным, пространственным данным точечных измерений ПС позволяет воссоздавать все пространственное поле данных измерений. Такое поле должно быть корректным не только с точки зрения обучающей выборки, но и оставаться таковым в рамках допустимого диапазона параметров ПС (для ДРМ водной среды это океанографический диапазон). Модель отличается от известных возможностью автоматизированного переключения регрессионного или сплайн-интерполяционного уравнения наилучшим образом описывающего, с точки зрения выбранных ЛПР метрик качества, ДРМ ПС.

Осуществление процесса эко- и биомониторинга ПС по неполным, разрозненным, пространственным данным точечных измерений с целью своевременного обнаружения аномальных и критических состояний для принятия ответных мер по выходу из этих состояний является актуальным и необходимым в современных реалиях. Зачастую исследователь не располагает данными

достаточно плотной СПИ с требуемым временным разрешением, достаточной длительности наблюдений и требуемого качества [92].

Построение математических моделей осуществляется с использованием аналитически заданного функционального поля изучаемых параметров по априорно заданным точкам. Предполагается, что заданы область $D \subset R_2$ и правило, задающее отображение $M(x, y) \in D \xrightarrow{f} w \in R_1$, согласно которому каждой записи ДРМ ПС M по правилу f ставится в соответствие число w такое, что определяет скалярное поле $w = f(M) = f(x, y)$. Специфика А-задачи подразумевает что значения параметров мониторинга w заданы лишь в некоторых точках области наблюдения D в связи с высокими стоимостными и временными затратами на проведение замеров, поэтому расстояния между ними Δx и Δy могут быть достаточно большими, кроме того, вследствие естественных причин, данные могут быть получены на нерегулярной сетке, что ограничивает применение классических методов линейной и сплайновой интерполяции, где требуется их регулярность. Зададим множества X и Y такие, что определяют правило $M(x \in X, y \in Y) \xrightarrow{f} w$ в n произвольно заданных точках M , а в остальных точках $M(x, y) \in D$ функция не определена, тогда мы имеем дело со скалярным дискретно заданным полем. Осуществим замену скалярного поля его функциональным представлением в виде правила f_m такого, что удовлетворяет условию

$$f(M(x, y)) = f_m(M(x, y)) \forall x \in X, y \in Y, \quad (2.1)$$

при этом отметим, что f_m в отличие от f определена уже для всей области определения $M(x, y) \in D$.

Функция f_m , в зависимости от размерности заданного пространства n и среднеквадратического отклонения значений точек w , может быть выражена в виде уравнения линейной, билинейной, квадратной, биквадратной, кубической, бикубической или регрессии более высокого порядка соответственно [94]

$$f_m(x, y) = a_0 + a_1x + a_2y, \quad (2.2)$$

$$f_m(x, y) = a_0 + a_1x + a_2y + a_3xy, \quad (2.3)$$

$$f_m(x, y) = a_0 + a_1x + a_2y + a_3xy + a_4x^2 + a_5y^2, \quad (2.4)$$

$$f_m(x, y) = a_0 + a_1x + a_2y + a_3xy + a_4x^2 + a_5y^2 + a_6x^2y + a_7xy^2 + a_8x^2y^2, \quad (2.5)$$

$$f_m(x, y) = a_0 + a_1x + a_2y + a_3xy + a_4x^2 + a_5y^2 + a_6x^2y + a_7xy^2 + a_8x^3 + a_9y^3, \quad (2.6)$$

$$f_m(x, y) = a_0 + a_1x + a_2y + a_3xy + a_4x^2 + a_5y^2 + a_6x^2y + a_7xy^2 + a_8x^2x^2 + a_9x^3 + a_{10}y^3 + a_{11}x^3y + a_{12}xy^3 + a_{13}x^3y^3 + a_{14}x^3y^2 + a_{15}x^2y^3. \quad (2.7)$$

Вектор коэффициентов $a_i \in A$ для представленных уравнений находится путем решения задачи минимизации функции невязки как суммы квадратов отклонения значений регрессионного уравнения от заданных исходных значений

$$F(a) = \sum_{i=0}^n (f(x_i, y_i) - f_m(x_i, y_i))^2 \rightarrow \min, \quad x_i \in X, y_i \in Y, \quad (2.8)$$

где $f(x_i, y_i)$ – фактическое значение параметра в точке (x_i, y_i) , $f_m(x_i, y_i)$ – значение точки на поверхности, заданной регрессионным уравнением f_m .

Графическое представление генерируемого пула реализаций моделей в виде поверхностей, заданных регрессионными уравнениями (2.2–2.7), представлено на рисунке 2.1 и 2.2, где точками обозначены заданные изначально значения параметра (температура). Оси x, y задают факторное пространство параметров (широта и долгота) исходного точечно заданного поля распределения температуры в поверхностном слое бухты г. Севастополя в зимний период [93].

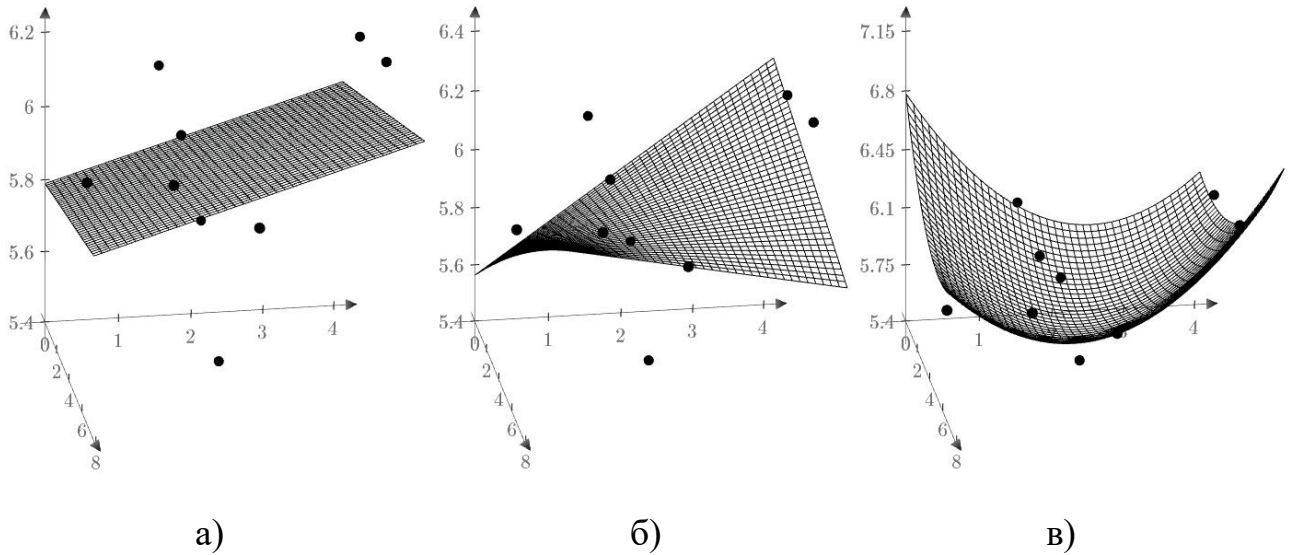


Рисунок 2.1. Поверхность параметров среды, заданная в виде а) линейной, б) билинейной, в) квадратичной регрессии

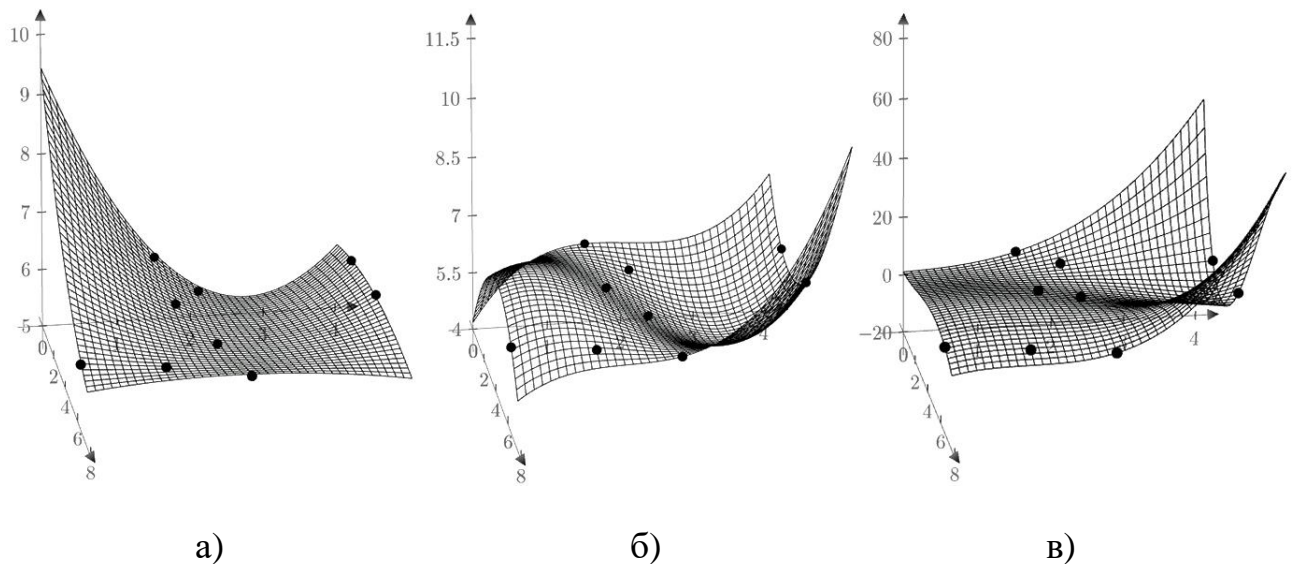


Рисунок 2.2. Поверхность параметров среды, заданная в виде а) биквадратичной, б) кубической, в) бикубической регрессии

Указанный выше подход применения пула параметрических регрессионных моделей является классическим для такого класса задач, имеет низкую вычислительную сложность. Подбор параметров осуществляется известными методами (градиентный спуск, симплекс метод, методы многомерной оптимизации) и для регрессионных уравнений высокого порядка дает приемлемые

показатели среднеквадратической ошибки, связанной с разбросом значений точек фактических наблюдений относительно функционально заданной поверхности.

Область применимости сгенерированных таким образом моделей ограничена. Как будет показано далее, в том числе с использованием разработанной модели на основе численных метрик оценки адекватности, при попытках снижения уровня невязки быстро возникает переобучение модели. В меньшей степени данной проблеме подвержены приближенные сплайновые модели, которые будут более подробно рассмотрены в разделе 2.2 .

Ограничения возникают при определении полей параметров наблюдений морской среды по той причине, что дисперсия случайной выборки из поверхности, заданной регрессионным уравнением, значительно возрастает по сравнению с дисперсией исходного множества, и она тем выше, чем больше степень регрессионного уравнения. Кроме того, локальные экстремумы поверхностей, заданных регрессионными уравнениями, могут выходить за пределы океанографического диапазона, что значительно повышает вероятности ошибок второго рода при обнаружении аномалий в данных, заданных таким образом. Вследствие этого возникает задача получения комплексной численной оценки адекватности математической модели конкретной предметной области, решение которой осуществляется набором проблемно-ориентированных метрик в специальном модуле проверки адекватности, модель которого приложена в разделе 2.3 .

2.2 Метод структурирования данных мониторинга сложного и периодического формата с повышенными требованиями к вычислительным ресурсам

Метод основан на использовании ресурсоемких численных подходов приближенного представления данных с повышенными требованиями к вычислительным ресурсам: пороговой фильтрации, итеративного сглаживания и сплайн интерполяции.

Учитывая противоречивость требований, предъявляемых к современным СМ ПС, которые в работе нашли свое отражение в виде комплексного критерия K , в случаях, когда исключительно аналитическим математическим аппаратом найти удовлетворительное решение не представляется возможным, продуктивно использовать одно из трех, представленных ниже, численно-аналитических решений.

Данные решения, в отличие от чисто регрессионных уравнений, более универсальны: с большей адекватностью осуществляют описание периодических данных, данных многопараметрического мониторинга, данных, описание которых элементарными функциями или простыми их комбинациями не представляется возможным.

К недостаткам подобных решений следует отнести значительно большую вычислительную сложность как для построения таких моделей, так и для дальнейшей их эксплуатации в системах анализа ДРМ ПС. Требуются большие объемы памяти для хранения таких структур и вычислительные ресурсы для оценки достигнутого качества модели при осуществлении итерационных матричных алгоритмов во время ее синтеза.

Модель представления данных на основе метода пороговой фильтрации

Зададим функциональное поле наблюдений биквадратного регрессионного уравнения с пороговой фильтрацией f_l , которое формируется как функционал от f_m (2.5), где при выходе f_m за пределы $\max f(x, y)$ или $\min f(x, y)$ происходит сжатие функции к ее ближайшей границе:

$$f_l(x, y) = \begin{cases} \max f(x, y) + \lambda, & \text{если } f_m(x, y) \geq \max f(x, y) \\ \min f(x, y) - \lambda, & \text{если } f_m(x, y) \leq \min f(x, y) \\ f_m(x, y), & \text{иначе} \end{cases}, \quad (2.9)$$

где $\lambda = Fm(f_m(x, y) - \max f(x, y))$ — возрастающая функция, асимптотически приближающаяся к некоторой константе, выбираемой на усмотрение ЛПР,

удовлетворяющая требованию $Fm(0)=0$. Например, этому условию удовлетворяют нормализованные функции класса сигмоид $Fm_1 \dots Fm_5$ (корневая сигмоида, гиперболический тангенс, арктангенс, рациональная сигмоида, функция Ферми – Дирака):

$$\begin{aligned}
 Fm_1(\varepsilon) &= \frac{\varepsilon}{\sqrt{1+\varepsilon^2}} \\
 Fm_2(\varepsilon) &= \tanh(\varepsilon) \\
 Fm_3(\varepsilon) &= \frac{2}{\pi} \arctan\left(\frac{\pi}{2}\varepsilon\right). \\
 Fm_4(\varepsilon) &= \frac{\varepsilon}{1+|\varepsilon|} \\
 Fm_5(\varepsilon) &= \frac{1}{1+e^{-2\varepsilon}}
 \end{aligned} \tag{2.10}$$

Введение пороговой фильтрации позволяет не ухудшая параметр невязки (метрику Y) улучшить метрики $W \rightarrow 1$, $Q \rightarrow 1$. Алгоритмы, основанные на данном решении, обладают наибольшим быстродействием (сравнимым с аналитическими решениями) по причине минимально возможного увеличения сложности путем добавления того одного порога в классическую аналитическую модель.

Модель представления данных на основе метода итерационного сглаживания

Зададим функциональное поле наблюдений осуществляя итерационное сглаживания гауссовым фильтром исходной матрицы GM , в которой все значения изначально инициализируются нулями, а в точках, где исходные значения заданы, они остаются неизменными во время всех итераций процесса сглаживания, т.е. осуществляется восстановление $GM_{x,y} = f(x, y) \forall x \in X, y \in Y$

$$GM(x, y) = \begin{cases} M(x, y), \text{ если } M(x, y) \neq \emptyset \\ GM(x \pm 1, y \pm 1), \text{ если } \begin{cases} x = 0 \\ y = 0 \end{cases}, \\ Mean(Neighbor(GM(x, y))) \\ \Delta dist(GM) \leq limit \end{cases} \quad (2.11)$$

где функция $Mean(Neighbor(GM(x, y)))$ возвращает среднее арифметическое значений всех точек, лежащих в окрестности $GM(x, y)$, умноженное на гауссову весовую матрицу свертки. Очевидно, что на каждой из итераций вычислений значения матрицы GM будут меняться до тех пор, пока приращение среднеквадратического отклонения $\Delta dist(GM)$ не станет достаточно мало $\Delta dist(GM) \leq limit$. Параметр $limit$ задается ЛППР и определяет вычислительную сложность алгоритма и точность генерируемой модели.

Модель представления данных на основе сплайн интерполяции

Зададим функциональное поле наблюдений с использованием кусочно-полиномиальной сплайновой интерполяции, выполнив интерполяцию кубическими сплайнами функцию двух переменных $f(x, y) \forall x \in X, y \in Y$. Проведем через сетку узлов поверхность, составленную из кубических полиномов, от x до y таким образом, что первые и вторые производные окажутся непрерывными в каждом узле сетки.

Вследствие того, что исходные данные были заданы на нерегулярной сетке значения, в узлах сетки необходимо интерполировать на все множество промежуточных значений. Интерполяция производится по трем точкам: если f_A, f_B, f_C – значение функции $f(x, y)$ в вершинах A, B, C некоторого треугольника, то вычислить приближенное значение функции внутри него можно с помощью функции $f_{\Delta} \approx a_0x + a_1y + a_2$, находя коэффициенты a_0, a_1, a_2 из условий:

$$a_0x_A + a_1y_A + a_2 = f_A, \quad (2.12)$$

$$a_0x_B + a_1y_B + a_2 = f_B, \quad (2.13)$$

$$a_0x_C + a_1y_C + a_2 = f_C, \quad (2.14)$$

где $\{x_A, y_A\}, \{x_B, y_B\}, \{x_C, y_C\}$ координаты вершин A, B, C .

Графическое представление генерируемого пула реализаций моделей (2.9-2.13) в виде функционально заданного пространственного поля заданного в трехмерном пространстве представлено на рисунке 2.3. Набор исходных данных идентичен тому, что бы использован для построения визуализации на рисунке 2.2.

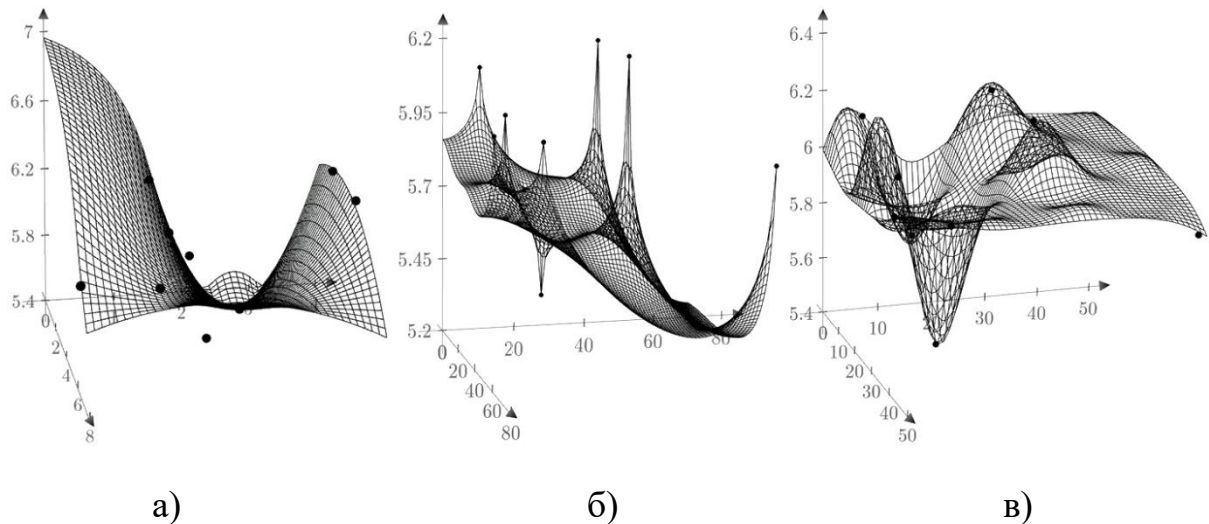


Рисунок 2.3. Поверхность восстановленного функционального поля а) заданного в виде биквадратной функции с пороговой фильтрацией б) полученного с использованием итерационного сглаживания гауссовым фильтром в) кусочно-полиномиальной сплайновой интерполяции

Таким образом, в результате замены исходного, частично определенного поля ДРМ ПС его параметрической моделью, у исследователя появляется возможность использовать классические подходы принятия решения, в том числе для слабоструктурированных данных.

Использование приближенных численных моделей, описанных в данном разделе, вычислительно более затратный подход поэтому их использование

продуктивно в том случае если альтернативные модели (2.2–2.8) в соответствии с численными метриками (2.15–2.17) дают неудовлетворительный результат. Выбор наилучшей математической модели осуществляется по тем их реализациям, которые имеют наибольший ранг, ранжирование моделей осуществляется по метрике (2.18).

В результате выполнения этого этапа ДРМ ПС приводятся к единой унифицированной форме, которая служит исходной в применяемых методах обнаружения аномалий, использующих параметрические, непараметрические, сценарные и полимодельные подходы.

2.3 Метод и алгоритм адаптивного выбора метода представления данных мониторинга природных систем с использованием векторных критериев

Синтез алгоритма обоснованного выбора метода представления ДРМ ПС в виде численно-аналитической математической модели, задающей функциональные поля изучаемых параметров, осуществляется на основе векторных критериев, учитывающих специфику возникающих аномалий ДРМ ПС и характера входных данных ДРМ.

Алгоритм выбора метода представления ДРМ ПС представлен следующим образом:

1. Строятся математические модели, задающие частично определенное поле в факторном пространстве двух переменных: линейной, билинейной, квадратной, биквадратной, кубической, бикубической регрессии. Возможно использование более сложных моделей: биквадратной функции с пороговой фильтрацией, матричного итерационного сглаживания гауссовым фильтром и кусочно-полиномиальной сплайновой интерполяции.

2. Вводятся численные метрики оценки качества математических моделей, построенных по данным точечных измерений, косвенно оценивающие вероятности ложного срабатывания детектора при обнаружении аномалий.

3. Осуществляется расчет численных метрик качества заданного пула математических модели и осуществляется их ранжирование по совокупности критериев качества для заданного обучающего множества данных.

Определим три наиболее репрезентативные метрики Y, W, Q оценки качества модели данных мониторинга ПС, заданных аналитически (множество реализаций которой обозначено как F_m), требующих наличие обучающей выборки F объемом N :

Первая метрика Y – минимум среднеквадратической ошибки, связанной с разбросом значений точек фактических наблюдений относительно поверхности

$$Y = \sum_{n=1}^N (F^n - F_m^n)^2 \rightarrow 0, \quad (2.15)$$

данная метрика традиционно используется при проверке качества регрессионных моделей и обладает хорошей сходимостью, но только в случаях, когда в данных обучающей выборки отсутствуют выбросы, что не всегда имеет место.

Вторая метрика W строится из положений, что параметрам измерений ПС свойственна пространственная автокорреляция, вследствие чего среднеквадратическое отклонение выборки из поверхности, заданной регрессионным уравнением, не должно превышать таковое для ДРМ ПС

$$W = \sqrt{\frac{D[F_m^n]}{D[F^n]}} \rightarrow 1, \quad (2.16)$$

где $F_m^n = \{f_m(R(x), R(y))_0, f_m(R(x), R(y))_1, \dots, f_m(R(x), R(y))_{n-1}\}$,

$R(x) \in X, R(y) \in Y$,

$F^n = \{M_0, M_1, \dots, M_{n-1}\}$,

$D[F]$ – дисперсия множества F ,

$R(x) \in X$ – случайное число, распределенное равномерно распределенное в диапазоне $X_{\min} \leq x \leq X_{\max}$.

Вероятностная природа данной метрики позволяет снизить влияние эффекта переобучения аналитической модели, так как реализация F_m каждый раз случайна и строится в области определения F , что делает корректным сравнение их статистических характеристик.

Третья метрика Q оценивает расстояние между минимальным и максимальным значениями исходной выборки и соответствующей ей реализации, полученной в результате решения регрессионного уравнения

$$Q = \frac{\max(F_m^n) - \min(F_m^n)}{\max(F^n) - \min(F^n)} \rightarrow 1, \quad (2.17)$$

в которой значительное отклонение метрики в большую от 1 сторону свидетельствует о возможном выходе значений за океанографический диапазон и впоследствии к ошибочному детектированию аномалии на последующих этапах обработки, а отклонение в меньшую сторону увеличивает вероятности пропуска аномалии.

Таким образом, итоговый критерий принятия произвольной аналитической модели примет вид

$$K(a, b, c) = a\sqrt{Y} + b|W - 1| + c|Q - 1| \rightarrow 0, \quad (2.18)$$

где a, b, c – коэффициенты значимости отдельных метрик выбираемых ЛПР в зависимости от приоритетов: точности, адекватности и вероятности ошибок I или II рода.

В том случае, когда ДРМ ПС имеют настолько сложную структуру, что классические регрессионные модели не удовлетворяют критерию K , продуктивно

применять более сложный математический аппарат численных, приближенных и итеративных методов описания модели.

2.4 Метод и алгоритм обнаружения аномалий данных многопараметрического мониторинга природной среды с кластеризацией состояний

В классической постановке, для ПС с ограниченным числом состояний, А-задача эффективно решается как частный случай задачи кластеризации данных при малом числе кластеров и статистически различимых наборах признаков среды. Данное допущение справедливо, в частности, для исследования ДРМ гидробионтов в экосистемах с ограниченным видовым разнообразием, например, Черного и Азовского морей.

Модель решения А-задачи обнаружения N-аномалий в такой постановке основана на кластеризации ДРМ ПС за счет использования методов параметрической статистики. Данные многопараметрического мониторинга заданы в виде наборов образов, представленных как скаляры признакового пространства малой размерности при заданном числе кластеров.

В общем виде осуществим оценку вероятности возникновения ошибок I и II рода для случая бинарного классификатора, и обобщим ее на многомерный случай, с целью получения численной оценки адекватности данной модели конкретному рассматриваемому случаю.

При использовании классического подхода в качестве меры сходства между образами (записями о состоянии ПС) $q \in Q$ используется их расстояние в n-мерном признаковом пространстве Z .

Ошибка классификации в моделях параметрической статистики вызвана несоответствием положения образа $q \in Q$ и соответствующего ему класса X в пространстве $Z \in R^n$. Данная ошибка обусловлена стохастической природой объектов исследуемой среды. С другой стороны, в результате применения классификационного алгоритма, может оказаться, что образ принадлежит сразу

нескольким классам $q \in X_1$ и $q \in X_2$, при этом вероятность этой принадлежности для различных классов неодинакова.

Статистический подход кластеризации состояний ПС как задача логистической регрессии

Статистический подход предусматривает оценку частоты появления образа в каждом из классов. При большом количестве образов (достаточном объеме выборок) в k -м классе эта частота стремится к значению условной вероятности $p_k(q|z \in Z)$ появления образа q в данном классе, где z – множество признаков образа. Под $p_k(q|z)$ будем понимать вероятность того, что образ q , характеризуемый параметрами z , относится к классу X_k . Гистограмму распределения $p_k(q|Z)$ можно рассматривать как дискретную аппроксимацию функции плотности распределения $f_k(q)$, определяющей вероятность того, что очередной поступивший в систему образ будет принадлежать k -му классу. Таким образом, обратная функция $F_k: Z \rightarrow p_k(q) \forall k \in X$ будет представлять собой формальное описание модели логистической регрессии, однозначно определяющей вероятность принадлежности образа q , описываемого набором параметров Z , каждому из рассматриваемых классов $X_k \forall k \in X$.

Конкретная реализация функции F_k , как задача машинного обучения, реализуется на основе принципа максимума правдоподобия ML. Этот алгоритм представляют собой классификатор с обучением, в котором ЛПР заранее задает конечное множество классов, по которым имеются достаточные объемы обучающих выборок. Эти данные позволяют получить выборки данных состояния по всем классам и оценить вероятность появления состояния ПС в каждом классе для всего множества Q .

Сущность применяемого метода для рассматриваемого случая A -задачи состоит в следующем: пусть есть выборка X_1, \dots, X_n из распределения p . Пусть $L(X|z): Z \rightarrow R$ – функция правдоподобия, где $z \in Z$ – неизвестные параметры.

Точечная оценка, при использовании ML метода, примет вид

$$ML(X_1, \dots, X_n) = \arg \max_{z \in Z} L(X_1, \dots, X_n | z). \quad (2.19)$$

Необходимо найти такое правило разделения классов X_1, \dots, X_n , при котором обеспечивается максимизация ML критерия. Если X_1, \dots, X_n – независимая выборка из непрерывного равномерного распределения на отрезке $[0, z]$, где $z > 0$ – неизвестный параметр, функция правдоподобия примет вид

$$f(x | z) = \begin{cases} \frac{1}{z^n}, & x \in [0, z]^n \subset \mathfrak{R}^n \\ 0, & x \notin [0, z]^n \end{cases}. \quad (2.20)$$

Адаптируя предложенный подход к А-задаче в бинарной постановке задачи двухпараметрического классификатора наличия или отсутствия аномалий в ДРМ ПС, получим решение в общем виде.

Бинарный классификатор наличия аномалии по данным мониторинга ПС на основе вероятностной модели

Построим двухпараметрический бинарный классификатор для определения нормального или аномального стояния ПС по ДРМ, обозначим их как Ω_1 и Ω_2 соответственно.

Рассмотрим правило принятия решений с использованием алгоритма статистической классификации на примере двух классов $k = 2$ в двумерном пространстве признаков Z^2 , где признаки заданы действительными значениями измерений образа $q \rightarrow Z^2$. Осуществляется разделение ДРМ Ω_1 и Ω_2 по двум признакам x_1 и x_2 . В контексте введенных обозначений справедливо следующее $X = \langle \Omega_1, \Omega_2 \rangle$, $Z^2 = \langle x_1, x_2 \rangle$, где Ω_1 соответствует записям класса нормальные, Ω_2 – записям класса аномальные.

Стохастическая природа признаков классифицируемых объектов ПС, будет описана математической моделью в виде совокупности функций случайных величин, заданных в виде двумерного нормального распределения

$$f(x_1, x_2, a_1, a_2, \sigma_1, \sigma_2, \sigma_{12}) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\sigma_{12}^2}} \exp\left\{-\frac{1}{2(1-\sigma_{12}^2)} \times \left[\frac{(x_1-a_1)^2}{\sigma_1^2} - \frac{2\sigma_{12}(x_1-a_1)(x_2-a_2)}{\sigma_1\sigma_2} + \frac{(x_2-a_2)^2}{\sigma_2^2}\right]\right\}, \quad (2.21)$$

где σ_{12} – коэффициент парной корреляции между параметрами x_1 и x_2 ;

a_1, σ_1 – среднее и стандартное отклонение переменной x_1 соответственно;

a_2, σ_2 – среднее и стандартное отклонение переменной x_2 соответственно.

Если случайные величины x_1 и x_2 независимы, то корреляция принимается равной 0, $\sigma_{12} = 0$. Статистическая гипотеза в таком случае может быть представлена визуально в виде графиков двух поверхностей (рис. 2.4), соответствующих классам Ω_1 и Ω_2 .

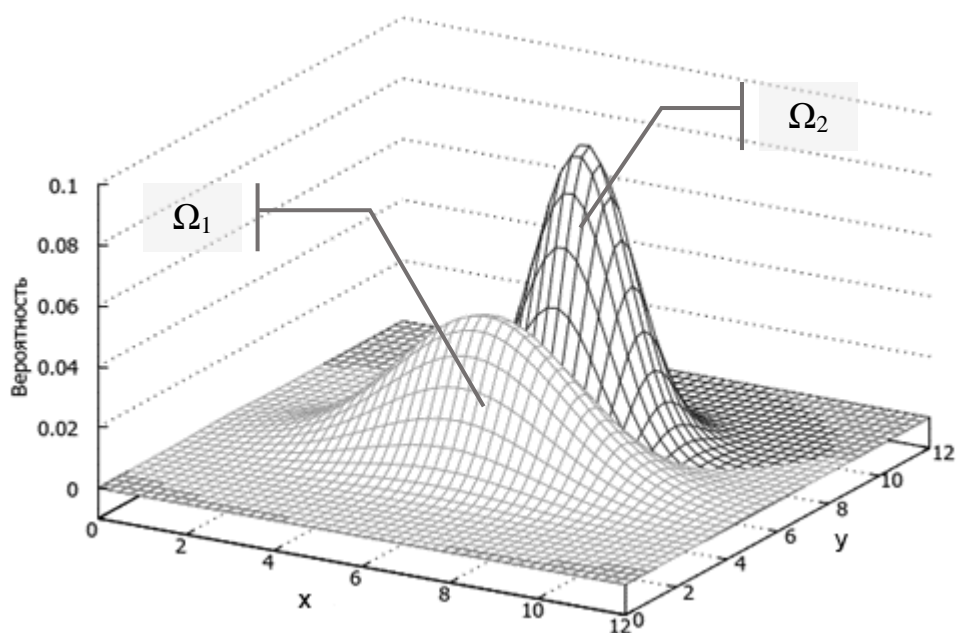


Рисунок 2.4. Статистическая гипотеза для случая двух классов Ω_1 и Ω_2 в двумерном пространстве признаков x, y

Классификатор осуществляет определение интервалов Ω_1 и Ω_2 , где $\Omega = \langle x_1^{\min}, x_1^{\max}, x_2^{\min}, x_2^{\max} \rangle$, на которых будут приниматься решения в пользу первого и II класса соответственно. Пусть получен достаточный объем

статистической информации о классах Ω , а именно функции плотности статистического распределения $f_1(x_1, x_2)$ и $f_2(x_1, x_2)$, и априорные вероятности $P(x \in \Omega_1|z)$ и $P(x \in \Omega_2|z) \forall z \in Z$.

Проведем гиперплоскость $G: x_2 = k_1 \cdot x_1 + k_2$ разделяющую факторное пространство признаков Z на два подпространства, одно из которых $ML: x_2 < k_1 \cdot x_1 + k_2$ будет соответствовать области решений в пользу класса Ω_1 и $x_2 \geq k_1 \cdot x_1 + k_2$ области решений в пользу класса Ω_2 . Полученная гиперплоскость G и является границей, определяющей классификационное правило ML .

Анализ достоинств и недостатков модели. К достоинствам можно отнести простоту интерпретации классификатора, коэффициент независимой переменной напрямую связан с весом, который можно интерпретировать как прогнозируемое значение независимой переменной для целевой переменной. Высокая скорость алгоритма позволит его использовать его в онлайн режиме.

К недостаткам модели отнесем высокие требования к объему выборок, соответствующих каждой категории в целевой переменной. В случае возникновения выбросов ДРМ, они должны быть предварительно удалены, т.к. их наличия вызывает сильные искажения модели. Модель не может обрабатывать пропущенные значения, поэтому их необходимо получать численно-аналитическими методами, рассмотренными в разделах 2.1 и 2.2. Модель представляет собой линейный классификатор, который не всегда адекватно позволяет решить A -задачу.

2.5 Метод и алгоритм численной оценки точности бинарного классификатора данных мониторинговых наблюдений в общем виде

Численная оценка качества принятия решения при кластеризации ДРМ ПС на нормальные и аномальные может быть получена на основе двух групп методов, использующих внешние и внутренние меры:

– внешние меры основаны на сравнении результата кластеризации с априори известным разделением на классы, и требуют наличие предварительно размеченной выборки ДРМ, что имеет место при анализе ретроспективных ДРМ ПС,

– внутренние меры отображают качество кластеризации только по информации в данных, что имеет место при обнаружении неизвестных видов аномалий или при отсутствии априорной информации о выборке.

Рассмотрим первый случай, когда имеется априорная информация о выборке. В случае, когда набор признаков классифицируемых объектов ПС подчиняется известному закону распределения, параметры которого можно определить, качество классификатора может оцениваться по аналитически рассчитанным численным значениям вероятности ошибки I и II рода.

За нулевую гипотезу H_0 примем, что объект с заданными параметрами x_1 и x_2 относится к кластеру нормальные данные $ML(x_1, x_2) \rightarrow \Omega_1 | q \in X_1$. Альтернативная гипотеза H_1 соответственно $ML(x_1, x_2) \rightarrow \Omega_2 | q \in X_2$ – объект q класса аномальные данные.

Ошибка I рода есть вероятность ошибочного принятия гипотезы H_1 когда верна H_0 $\alpha = P(H_1 | H_0)$, которая может быть найдена численно как объем, ограниченный графиком поверхности Ω_2 , горизонтальной плоскостью, проходящей через начало координат и разделяющей гиперплоскостью G . Аналогично определяется ошибка II рода $\beta = P(H_0 | H_1)$. Будем считать, что классы Ω_1 и Ω_2 равноценны и $\alpha = \beta$, тогда путем подбора соответствующих коэффициентов k_1 и k_2 добьемся равенства α и β при условии $V(\Omega_1) - \beta \rightarrow \max$ и $V(\Omega_2) - \alpha \rightarrow \max$, $V(\Omega_1) = V(\Omega_2) = 1$.

Решение задачи расчета объема фигуры с функционально заданными поверхностями тривиально и имеет вид

$$V = \iiint_{\Omega} f(x_1, x_2, z) dx_1 dx_2 dz. \quad (2.22)$$

Составим систему ограничений и найдем решение для α и β в общем виде

$$\left\{ \begin{array}{l} V = \int_0^T dx_1 \int_0^G dx_2 \int_0^H dF(x_1, x_2) \\ T = -\frac{k_2}{k_1} \\ G = k_1 x_1 + k_2 \\ H = \max F(x_1, x_2) \forall x_1, x_2 \in Z^2 \end{array} \right. \quad (2.23)$$

Подставим (2.21) в (2.23) и найдем решение для расчета ошибки II рода в общем виде

$$\beta = \int_0^T dx_1 \int_0^G dx_2 \int_0^H d \max \left\{ \begin{array}{l} \frac{1}{2\sigma_1\sigma_2\pi} e^{-\frac{1}{2}\left(\frac{(x_1-a_1)^2}{\sigma_1^2} + \frac{(x_2-a_2)^2}{\sigma_2^2}\right)} \\ \frac{1}{2\sigma_1'\sigma_2'\pi} e^{-\frac{1}{2}\left(\frac{(x_1-a_1')^2}{\sigma_1'^2} + \frac{(x_2-a_2')^2}{\sigma_2'^2}\right)} \end{array} \right. \quad (2.24)$$

Графически решение системы проиллюстрировано на рисунке 2.5.

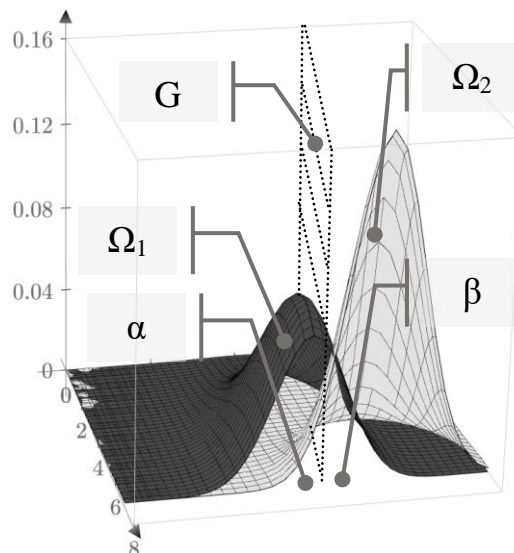


Рисунок 2.5. Сечение графиков поверхностей классов Ω_1 и Ω_2 и разделяющая их гиперплоскость G

Аналогичным образом, получив решения α и β , можно оценить точность P_r и полноту R_c классификатора

$$P_r = \frac{P(H_1 | H_1)}{P(H_1 | H_1) + P(H_0 | H_1)} = \frac{TP}{TP + \beta}, \quad (2.25)$$

$$R_c = \frac{P(H_1 | H_1)}{P(H_1 | H_1) + P(H_1 | H_0)} = \frac{TP}{TP + \alpha}, \quad (2.26)$$

$$TP = P(H_1 | H_1) = \int_0^T dx_1 \int_0^G dx_2 \int_0^H d \frac{1}{2\sigma_1\sigma_2\pi} e^{-\frac{1}{2}\left(\frac{(x_1-a_1)^2}{\sigma_1^2} + \frac{(x_2-a_2)^2}{\sigma_2^2}\right)}. \quad (2.27)$$

Данные метрики используются при оценке большей части алгоритмов извлечения информации как сами по себе, так и в качестве базиса для производных метрик, выбираемых на усмотрение ЛПР, таких как F-мера и R-мера.

Следует отметить что в случае ситуации, когда не удастся найти такую разделяющую гиперплоскость, которая удовлетворяет $\alpha \leq p_{\min} | \beta \leq p_{\min}$ (выбираемую ЛПР), это может свидетельствовать о том что объект ПС, характеризуемый парой x_1, x_2 , не принадлежит ни одному из имеющихся альтернативных вариантов кластеров Ω . Это может являться основой для принятия решения о возможном возникновении аномалии нового типа, и как следствие, необходимости применения более сложных моделей, в частности, моделей непараметрической статистики.

2.6 Метод и алгоритм обнаружения аномалий в скалярных полях с многомерными данными мониторинга природной среды

ДРМ ПС как правило представляют собой записи многомерных массивов, имеющие пространственную привязку, которая вносит существенный вклад в повышение обоснованности процесса их интерпретации. Если объемы этих данных достаточно велики или если к данным применены методы регуляризации и

повышения размерности рассмотренные подробно в 2.1 и 2.2 , то полученные массивы можно рассматривать как данные скалярных и векторных полей мониторинга и применять соответствующие методы решения А-задач в них.

Существующие модели обнаружения аномалий при мониторинге, как правило, строятся в предположении того, что исходные данные распределены по нормальному закону, что обусловлено удобством выполнения аналитических расчетов [95]. Однако в практике мониторинга физических величин часто имеют дело с аномальными моделями распределения, у которых параметры асимметрии и эксцесса значительно отличаются от нуля [96–105].

Осуществление обнаружения аномалий в данных скалярных и векторных полей мониторинга ПС как деятельности оперативного характера в общем случае может быть описано циклом PDCA, включающим ряд этапов [106]:

- определение факторов, имеющих наибольшую ценность при мониторинге, и выбор для каждого из них пороговых значений, полученных в результате измерений, оценок или прогнозов,
- выбор методов и средств измерений для осуществления сбора, анализа, интерпретации и прогнозирования ПС,
- определение оптимальных пространственных и временных интервалов между точками проведения замеров в соответствии с уровнем изменчивости признаков,
- осуществление прямых или косвенных измерений интересующих параметров объектов или процессов с заданными пространственно-временными интервалами,
- проверка результатов замеров с применением установленных методов, обнаружение аномальных значений и определение их природы [107],
- корректировка пороговых значений показателей с учетом поступающих в реальном времени новых и имеющихся ретроспективных данных [108].

Выбор пороговых значений стратегии принятия решений определяет риски возникновения ошибок I и II рода при обнаружении аномалий, вызванных сдвигом наблюдения и повышением дисперсионности.

При распознавании аномалий скалярных и векторных полей ДРМ ПС определим их математическую природу как искажение поля исходного параметра с применением линейных и нелинейных операций.

Примером линейной операции может служить преобразование поля T^0 факторного пространства Fs вида $L: T_{i,j}^0 \rightarrow T_{i,j}^1 + C$, где C – некоторая константа. Нелинейное преобразование подразумевает изменение значений, составляющих поле факторного пространства с применением квадратичных, тригонометрических и других нелинейных преобразований вида $T_{i,j}^1 = N(T_{i,j}^0)^M \quad \forall T_{i,j}^0 \in Fs$ [109].

Осуществляется контроль факта выхода интересующего параметра за контрольные границы, будем считать аномальными те значения T , которые выходят за диапазон $\Delta_1 - \Delta_2$. Функция обнаружения аномалий, в таком случае,

$$\text{примет вид } F(T_{i,j}) = \begin{cases} 0, & \Delta_1 \leq T_{i,j} \leq \Delta_2 \\ 1, & \text{иначе} \end{cases}.$$

Для любого распределения всегда найдется критический уровень, при котором будет совершена ошибка I или II рода. В случае нормального распределения вероятность выхода значения множества T за пределы заданных границ a, b можно выразить из функции распределения случайной величины, которая определяется в виде

$$F(T_i \leq a) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^a \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt, \quad (2.28)$$

что в общем виде можно представить в следующей форме [110]

$$P(a \geq T_i | b < T_i) = F(a) + 1 - F(b), \quad (2.29)$$

которая справедлива в общем случае для законов распределения общего вида.

Решение обратной задачи в численной для экспериментальных данных или в аналитической для теоретических исследований форме любым методом

математического программирования (симплекс-метода, метод градиентного спуска), определяет контрольно-предупредительные границы классификатора, где U – верхняя и D – нижняя границы соответственно. Границы соответствуют обратному кумулятивному нормальному распределению с заданным процентилем. Для нормального распределения с параметрами $\mu=0$, $\sigma=1$, пример результатов имитационного моделирования и аналитического представления, распределения вероятностей и кумулятивного распределения вероятностей, а также контрольно-предупредительных границ в виде прямых 5–95%, 10–90% и 25–75% представлены на рисунке 2.6.

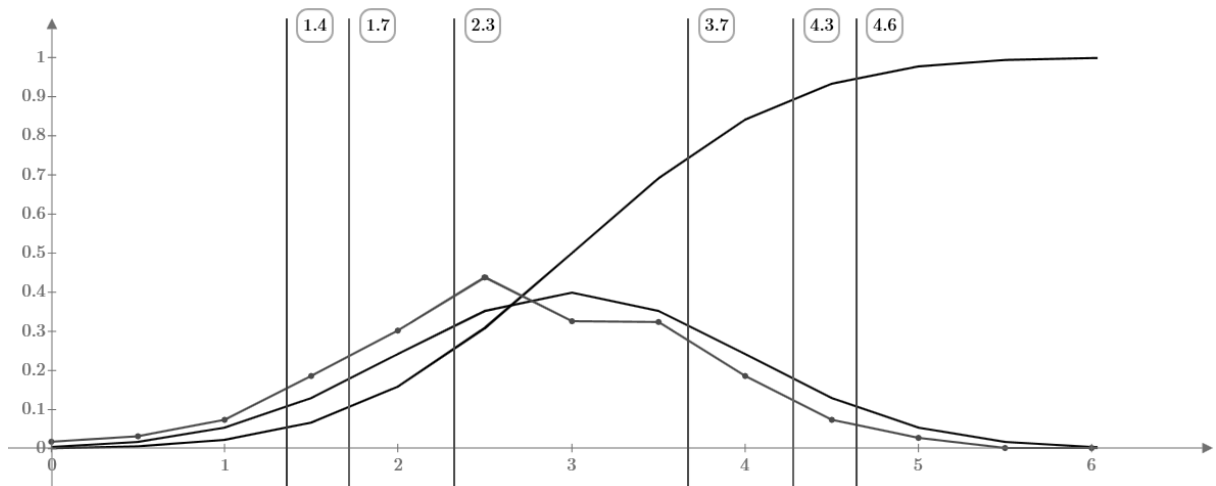


Рисунок 2.6. График эмпирического распределения мониторируемой величины, его аппроксимация нормальным распределением и контрольно-предупредительные границы классификатора

Ситуация совершенно изменяется, если аномалия вызвана изменением дисперсионности рассматриваемой случайной величины, когда распределение становится отличным от нормального, вследствие чего резко возрастает риск ошибок I и II рода. Для этого следует рассматривать ряд возможных реализаций аномалий ПС, например, описываемых следующими законами распределения: экспоненциальным $f_{\text{Э}}$, Вейбулла $f_{\text{В}}$, гамма $f_{\text{Г}}$:

– распределение аномалии случайной величины X задается плотностью вероятности $f_{\text{Э}}(x)$, имеющей вид:

$$f_{\Xi}(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0 \end{cases}, \quad (2.30)$$

где $\lambda > 0$ – обратный коэффициент масштаба.

– распределение аномалии случайной величины x задается плотностью вероятности $f_B(x)$, имеющей вид:

$$f_B(x) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}, & x \geq 0, \\ 0, & x < 0 \end{cases}, \quad (2.31)$$

где $\lambda > 0$ – коэффициент масштаба, $k > 0$ – коэффициент нормы.

– распределение аномалии случайной величины x задается плотностью вероятности $f_{\Gamma}(x)$, имеющей вид:

$$f_{\Gamma}(x) = \begin{cases} x^{k-1} \frac{e^{-x/\theta}}{\theta^k \Gamma(k)}, & x \geq 0, \\ 0, & x < 0 \end{cases}, \quad (2.32)$$

где $\Gamma(k)$ – гамма-функция Эйлера, $k > 0$, $\theta > 0$ – коэффициенты масштаба.

При использовании предложенного метода предполагается осуществлять целенаправленный эксперимент на моделирующем проблемно-ориентированном лабораторном стенде с целью определения степени достоверности для оценки рисков возникновения аномалий при мониторинге. Также следует отметить следующее:

– выбор оптимального доверительного интервала U и D происходит в пространстве компромиссов. При малых значениях сдвига \mathcal{E} оптимальным решением будет выбор узкого доверительного интервала U и D , а при больших, соответственно, широкого;

– выбор границ доверительного интервала U и D однозначно определяет коэффициент ошибок I рода α и поэтому может быть рассчитан заранее;

- при детерминированном законе распределения и границах доверительного интервала U и D параметр сдвига ε определяет коэффициент ошибок II рода β ;
- сужение доверительного интервала U и D приводит к значительному увеличению числа ошибок I рода α до 25% и 50% и практически пропорциональному снижению числа ошибок II рода соответственно.

Учет ошибок α и β при различных значениях сдвига и дисперсионности позволяет повысить достоверность при обнаружении аномалий. Таким образом значительно повышается качество решения А-задачи, если принять меры по компенсации этих ошибок, вследствие чего надежность принятия решений увеличивается.

2.7 Метод обнаружения аномалий в многослойных регулярных полях данных высокой плотности

Специфика современных мониторинговых исследований решения А-задачи зачастую предусматривает необходимость анализа данных, представленных в виде многослойных регулярных полей высокой плотности: спутниковый мониторинг и другие данные которые могут быть интерпретированы в виде графических изображений. Решение задачи в такой постановке требует применения специализированных методов анализа данных, представленных в виде регулярных матриц.

Задача распознавания аномалий на изображениях, получаемых от систем мониторинга ПС, традиционно считается областью, с которой человек справляется лучше машины и ее решение на протяжении многих лет является одной из наиболее критически важных проблем в научном и прикладном плане. Существующие методы распознавания обладают высокой эффективностью только лишь в ограниченных классах задач со строго заданными ограничениями и зачастую требуют большие объемы вычислительных ресурсов и обучающих выборок [111-115]. Именно в таких случаях будет предпочтительным применение методологии компьютерного зрения и системы инвариантных метрик.

Для повышения эффективности мониторинга ПС учеными были достигнуты существенные успехи в разработке целого класса устройств, позволяющих получать и анализировать фото- и видеоизображения объектов морской среды непосредственно *in situ*. Среди наиболее известных следует выделить Video Plankton Recorder (VPR) [116], Underwater Video Profiler (UVP) [117], Shadowed Image Particle Profiling Evaluation Recorder (SIPPER) [118], Zooplankton Visualization System (ZOOVIS) [119], Scripps Plankton Camera (SPC) [120], Imaging FlowCytobot (IFCB) [121], In Situ Ichthyoplankton Imaging System (ISIS) [122], ZooScan [123], которые позволяют собирать значительные объемы данных за удовлетворительное время [124]. Однако вопрос автоматического распознавания объектов ПС с целью получения численной оценки их количественного и видового разнообразия исчерпывающе и в полном объеме не разрешен [125–128].

Обнаружение аномалий многослойных регулярных полей данных эффективно решается путем адаптации механизмов компьютерного зрения, а именно особой реализации метода поиска по шаблону, нечувствительной к аффинным преобразованиям исходных изображений, применительно к решению А-задачи как изображений гетерогенных объектов в условиях малого объема обучающей выборки и малого числа классов [129]. Ограничение объема обучающей выборки является одним из наиболее существенных ограничений для реальных систем распознавания изображений и сильно ограничивает эффективность применения конкурирующих подходов, а именно статистических моделей и нейронных сетей [130]. С другой стороны, ограничение числа классов является оправданным для локальных акваторий, в частности морей и для промышленных систем экологического контроля, оценивающих количественный состав доминирующих видов планктона, когда повышение критерия точности классификатора является предпочтительнее повышения критерия его полноты [131].

Обозначенная проблема выбора, разработки и адаптации численных методов, предназначенных для распознавания гетерогенных объектов в условиях малых обучающих выборок, ввиду экономических причин, является одной из наиболее

часто встречаемых в прикладных системах решения А-задач. И в математическом плане трудноформализуемой из-за вероятностной природы объектов, характеризующих данные в процессе обучения и распознавания.

Метод обнаружения аномалий многослойных регулярных полей данных высокой плотности решает поставленную задачу за счет применения взаимодополняющих подходов: выделения текстурных признаков, инвариантных метрик и сверточных фильтров. Метод является проблемно-ориентированным и позволяет значительно снизить трудоемкость процесса распознавания за счет того, что работа специалиста, которая ограничивается необходимостью обучения системы, значительно снижена за счет уменьшения объема обучающей выборки. Кроме того, присутствует возможность адаптивного задания пороговых значений степени уверенности распознавания что обеспечивает выбор баланса критериев качества классификатора: точности и полноты.

2.8 Метод обнаружения аномалий в данных мониторинга при нежестких ограничениях на объем обучающих выборок с использованием агентного подхода

Информационные технологии и методологическая база, существующие на сегодняшний день, позволяют хранить, анализировать и использовать большие объемы интенсивно меняющихся и сложно организованных данных. Достаточно подробно описаны в литературе методы структуризации, кластеризации и визуализации неструктурированных и слабоструктурированных данных. Использование агентного подхода при решении А-задачи позволяет эффективно применять накопленные ДРМ, полученные за продолжительных период. В случае решения А-задачи при нежестких ограничениях на объем обучающих выборок, описанный подход позволяет продуктивно использовать такие данные в случае кластеризации множества аномалий большей мощности, чем в классической постановке.

Практическая полезность полученных результатов, заключается в поиске закономерностей среди проявлений изучаемой ПС, на основании которых строится

предметная модель, адекватно отражающая связи внутри ПС, что позволяет производить прогнозирование динамики изменения ПС и строить для нее проблемно-ориентированные системы поддержки принятия решений.

Решается задача проектирования метода позволяющего, на основе большого объема слабоструктурированных данных, максимально автоматизировано строить управляющие матрицы агента, осуществляющего мониторинг ПС. Производится оценка сходимости показателей эффективности алгоритма агента к области аттрактора и на основе результирующей поведенческой матрицы, в зависимости от объемов обучающей выборки, сравнивается эффективность применения конкурирующих алгоритмов. На схеме взаимодействия агента и окружающей среды (рис. 2.7) показано, что состояние агента в каждый следующий момент времени однозначно определяется состоянием среды и вознаграждением, полученным в результате действий агента на предыдущей итерации.



Рисунок 2.7. Схема взаимодействия агента и природной среды

Математическая модель приведенной схемы взаимодействия широко представлена в литературе, и в дискретном виде описывается уравнением Белмана

$$V^\pi(s) = \sum_a \pi(s, a) \sum_{s'} P_{ss'}^a \left[R_{ss'}^a + \gamma V^\pi(s') \right], \quad (2.33)$$

где: s – текущее состояние, принадлежащее к множеству S ;

a – действие принадлежащее к множеству $A(s)$, допустимых в состоянии s ;

$\pi(s, a)$ – стратегия выбора a в состоянии s ;

$P_{ss'}^a$ – вероятность перехода из состояния s в состояние s' в случае реализации сценария a .

$R_{ss'}^a$ – вознаграждение при переходе из состояния s в состояние s' в случае реализации сценария a ;

Типовым подходом реализации системы подсчета вознаграждений для цепи итераций является введение функции кумулятивного вознаграждения

$$V^\pi(s) = \sum_a \pi(s, a) Q^\pi(s, a), \quad (2.34)$$

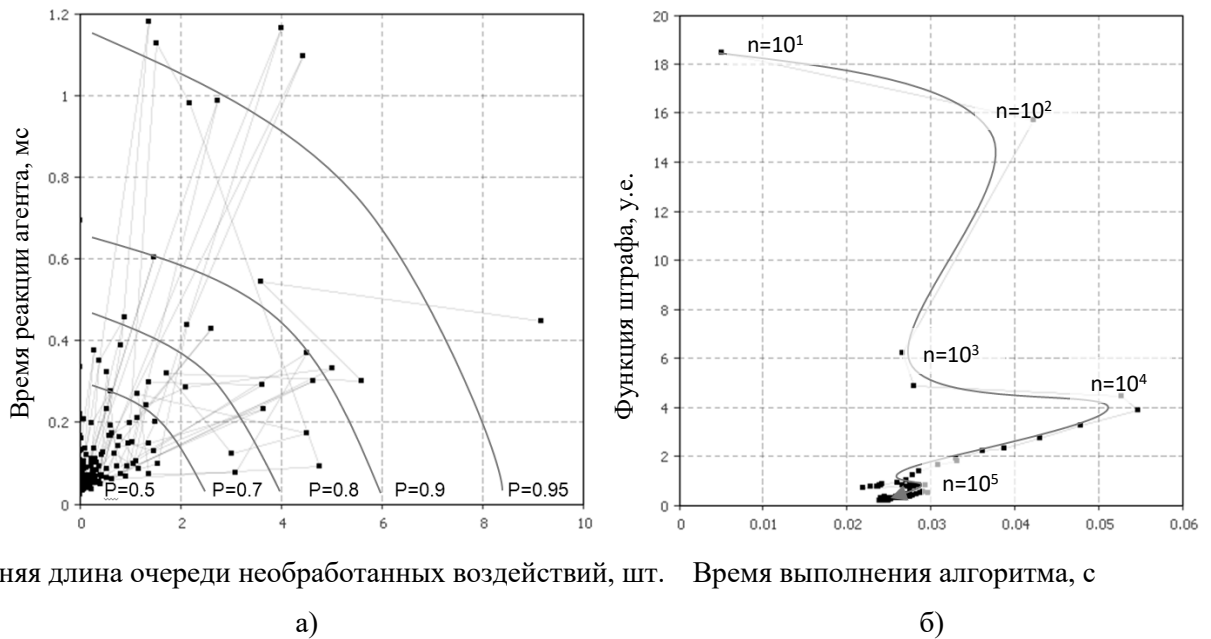
$$Q^\pi(s, a) = \sum_{s'} P_{ss'}^a \left[R_{ss'}^a + \gamma V^\pi(s') \right], \quad (2.35)$$

которая при объединении с функцией оценки образует уравнение Белмана в виде

$$Q^\pi(s, a) = \sum_{s'} P_{ss'}^a \left[R_{ss'}^a + \gamma \sum_{a'} \pi(s', a') Q^\pi(s', a') \right], \quad (2.36)$$

решение данного уравнения позволяет получить оптимальную стратегию принятия решений о наличии или отсутствия аномалий в ДРМ ПС агентом.

Применение таких моделей имеет практическую значимость вследствие того, что позволяет получить достаточно большой объем репрезентативной выборки, необходимый для корректного обучения агента. Для иллюстрации процесса обучения и применения метода (рис. 2.8) применим проблемно-ориентированную ИМ, описывающую состояние внешней среды и обеспечивающую возможность взаимодействия среды с агентом.



Средняя длина очереди необработанных воздействий, шт. Время выполнения алгоритма, с

а)

б)

Рисунок 2.8. Иллюстрация процесса обучения: а — зависимость времени реакции агента на изменение состояния внешней среды и средней длины очереди необработанных воздействий в зависимости от загрузки системы P ; б — зависимость функции штрафа и времени выполнения поведенческого алгоритма от размера обучающей выборки n

При рассмотрении модели в динамике можно произвести учет ограничений аппаратных ресурсов агента, в частности на графике (рис. 2.8а) показаны результаты моделирования системы агент-среда, свидетельствующие о значительном увеличении времени реакции и средней длины очереди необработанных воздействий по мере увеличения интенсивности поступления внешних воздействий при неизменной производительности аппаратных ресурсов агента. На графике зависимости функции штрафа и времени выполнения поведенческого алгоритма от размера обучающей выборки (рис. 2.8) хорошо видно, что по мере увеличения объема выборки функция штрафа монотонно снижается, а время выполнения алгоритма стремится к точке аттрактора.

Применение предложенного подхода становится возможным в том случае, когда модель среды достаточно хорошо определена большим объемом статистических данных. Для этого должны быть известны или получены методами Data Mining матрицы состояний окружающей среды и вознаграждений. Для формирования такой матрицы может использоваться ИМ. Использование данного подхода позволяет значительно снизить требования к аппаратному обеспечению

агента делегируя собственно процесс обучения внешнему программно-техническому решению, включающему ИМ и алгоритм, который может быть представлен в виде нейронных сетей, генетических алгоритмов, обучения с подкреплением и других.

Наиболее сложным вопросом представляется обеспечение корреляции данных, полученных на ИМ, представляющей с некоторой степенью достоверности агента и окружающую среду, с фактическими значениями, полученными на реальном объекте. При построении модели имеют место быть некоторые допущения и погрешности, точная оценка которых позволит оценить степень достоверности полученных результатов, однако это положение вряд ли может быть строго доказано ввиду отсутствия аналитических решений таких систем для произвольных распределений вероятности наступления воздействий в системе агент-среда. Отсюда можно сделать вывод о том, что реализация агента и его алгоритмическое обеспечение, описываемое поведенческими матрицами, полученными на имитационной модели, служит основой повышения информационной надежности приборов и средств контроля окружающей среды.

2.9 Модульно-компонентный подход при реализации полимодельного комплекса обнаружения аномалий различных типов

Как результат осуществленного анализа типовых ситуаций, возникающих при решении А-задач ПС, предложенных в диссертации проблемно-ориентированных методов их решения, учитывая весь спектр методов, область эффективных решений которых пересекается лишь частично, синтезирован полимодельный динамически конфигурируемый комплекс, адаптивно подстраивающийся к задаче обнаружения аномалий различных типов.

Основой интеллектуализации системы поддержки принятия решений по выбору оптимального набора методов анализа выборок ДРМ ПС является ввод точек принятия решений, дающих возможность ЛПР определять приоритеты: чувствительность или вероятность ложного срабатывания, в случае, когда Парето-оптимальное решение отсутствует. Эти точки помещаются в обучаемую

интеллектуальную технологию, где в соответствии с принятыми пользователями решениями системой осуществляется выбор весовых коэффициентов по критериям мощности или чувствительности. По существу, такой комплекс реализует элементы искусственного интеллекта с автоматизацией принятия решения, что является конечной задачей многих проблем искусственного интеллекта [134–141].

Результирующий проблемно-ориентированный динамически конфигурируемый комплекс обнаружения аномалий включает в себя пулы методов, направленных на решение следующих задач:

- приведение ДРМ ПС к универсальному виду, регуляризации и повышения их размерности,
- обнаружения аномалий детерминированных типов и неизвестных типов,
- обнаружение аномалий как возмущений в скалярных и векторных полях данных,
- получения численных оценок эффективности каждого из методов в имеющихся пулах для каждого конкретного случая их применения и выбора наилучшей комбинации методов анализа.

Оптимальный набор методов зависит не только от характера входных данных, но и от выбранного сценария использования ЛПР. Во многом именно ЛПР определяет ключевые критерии выбора порогов классификаторов в зависимости от допустимого для себя уровня риска. В частности, требования допустимой вероятности ложной тревоги и ложного срабатывания несимметричны, и, в зависимости от критичности объекта мониторинга и возможных последствий ошибочного принятия решения, определяют используемую модель представления ДРМ ПС. Требования к оперативности принятия решения определяют использование или более точных, но требующих большие объемы выборок данных или наоборот, способных работать в условиях недостатка времени на сбор выборки данных большого объема.

На рисунке 2.9 представлена структура интеллектуальной системы, которая в зависимости от характеристик выборок ДРМ позволяет выбрать подмножество критериев различия, применение которых будет наиболее продуктивно для обнаружения аномалий с имеющимся входным набором данных.



Рисунок 2.9. Структура конфигурируемого комплекса выбора критерия обнаружения дивергенции данных мониторинга природной среды

При том, что полимодельный комплекс является параметрически гибким и конфигурируемым ЛПР, и включает в себя все возможные комбинации методов представленных в разделах 2.1 – 2.8, функционал которых взаимодополняем и частично пересекается, в зависимости от конкретизации целевой задачи, принцип комбинирования этих методов следует выполнять из следующих соображений.

Параметрические критерии обладают большей статистической мощностью и меньшей вероятностью совершить ошибку II рода в сравнении с непараметрическими. Это означает, что в случае существования какой-либо закономерности между выборками с наибольшей вероятностью эта закономерность будет выявлена с помощью параметрических критериев. К числу параметрических критериев относят критерии Стьюдента, Фишера, Кохрена, Бартлетта, хи-квадрат и др. [142–152].

К ограничениям параметрических методов следует отнести то, что их можно применять только при одновременном соблюдении двух условий: количественном характере вариации и предположении что данные исходной выборки взяты из генеральной совокупности, имеющей конкретное распределение, как правило, нормальное. Однако, как показывает практика, часть данных, получаемых в результате мониторинга окружающей среды, не распределена нормально, поэтому применение параметрических критериев при анализе результатов таких исследований может привести к ошибкам в статистических выводах. В таком случае, непараметрические критерии могут стать более мощными, т.е. способными с большей достоверностью отвергать нулевую гипотезу.

Каждый из критериев различия имеет свою специфику, различаясь между собой по накладываемым ограничениям на входные наборы данных: тип измерительной шкалы, максимальный объем выборки, количество выборок, качество выборок. Разнообразие критериев обнаружения различия позволяет:

- выбирать критерий, адекватный типу шкалы, в которой получены экспериментальные данные;
- производить сравнение данных, характеризующихся связными и несвязными выборками;
- производить сравнение данных с неравными по объему выборками;
- в зависимости от целей исследования выбирать подходящие по мощности критерии.

В общем виде модель полимодельного комплекса решения А-задачи в ПС, при условии $\forall C \in \Omega$, представлена выражением

$$M : C_i, S(G, m), \alpha, F(L) \rightarrow R, \quad (2.37)$$

где Ω – весь допустимый набор критериев различия, C_i – i -й критерий различия, $S(G, N)$ – выборка из генеральной совокупности G объемом N , α – допустимая для данной задачи вероятность ошибки I рода (на практике часто полагают

$\alpha = 0,05$), $R = \langle true, false \rangle$ – бинарный результат применения критерия различия C , $F(L)$ – оператор выбора алгоритма.

Выбор алгоритма $F(L)$ осуществляется на основе подхода информационных сценариев и описывается выражением

$$F : L \langle SR, O(S), W, N \rangle \rightarrow A, \quad (2.38)$$

где $L = \langle SR, S, W \rangle$ – приоритеты, заданные пользователем, SR – выбор приоритета чувствительности или надежности, O – оператор выбора сценариев, S – список сценариев использования информационной системы, W – весовые коэффициенты критериев различия, N – адаптивное правило, A – выбранный алгоритм. Таким образом, интеллектуальное правило выбора приоритетов $F(L)$ осуществляет работу с учетом соотношения априорной и апостериорной информации. В настоящее время технологии сценариев активно используются для решения задач в условиях высокого дефицита априорной информации и дает возможность компенсировать отсутствие первоначальных знаний у персонала, предлагая ряд альтернативных типовых стандартных решений, где каждому такому решению соответствует некоторая алгоритмическая схема.

Результатом применения критерия различия C для выборок S_1, \dots, S_m является численная оценка статистической однородности выборок, для некоторых критериев (Стьюдента, Фишера, хи-квадрат) существует ограничение $m = 2$. Для удобства i -ю выборку $S(G, N)$ далее будем обозначать S_i .

При исследовании мощности критериев C в такой постановке выдвигается гипотеза H_0 – сравниваемые выборки однородны и извлечены из одной и той же генеральной совокупности. Таким образом, основная гипотеза состоит в том, что для любых выборок S_i , их функции распределения равны между собой $F(S_a) = F(S_b) | a \neq b$. В случае принятия гипотезы H_0 следует вывод об отсутствии изменений в состоянии объекта мониторинга или о незначительности этих изменений.

Конкурирующая гипотеза H_1 – выборки $S_a(G_1, N)$, $S_b(G_2, N)$ извлечены из разных генеральных совокупностей, описываемых различными законами распределения $F(G_1) = F(G_2)$. Конкурирующая гипотеза H_1 принимается в случае потенциально опасного изменения параметров, характеризующих состояние сложного объекта или среды. В рассматриваемом случае H_0 и H_1 – простые параметрические гипотезы, то есть полностью определяют все множество параметров распределения случайной величины.

Функция блока оценки качества мониторинга – подсчет частоты использования сценариев, отрицательных и положительных эффектов их использования. При адаптации выбора сценария изначально выбор равновероятностный, затем предпочтение отдается тому сценарию, который дает наибольший относительный эффект. Подсчитываются веса, критерии эффективности, и выбирается тот сценарий, который дает положительный результат чаще.

Для того чтобы отсеять заведомо бесперспективные варианты C для каждого конкретного случая следует учитывать объемы выборок, число выборок (две или более), равные или не равные они по объему, подчинены ли известному закону распределения. Выбор конкретной реализации C недетерминирован и осуществляется в соответствии с предпочтениями пользователя $L = \langle SR, S, W \rangle$ и интеллектуальной системы M .

В качестве критериев для проверки отсутствия дивергенции по средним значениям можно использовать критерии Фишера, Кохрена и Бартлетта. Для проверки отсутствия дивергенции по дисперсии можно использовать критерии Стьюдента, Фишера и приближенный критерий.

Известно, что в теории принятия решений наибольшую сложность вызывает формирование оценочной матрицы, процесс которого до сих пор до конца не формализован и не имеет конкретного алгоритма решения. Поэтому для формирования оценочной матрицы используется ЛПР, на основе технологии информационных сценариев, которые сейчас достаточно успешно применяются в разных областях. У ЛПР имеется некоторая свобода в выборе сценариев, а в рамках

сценария эта свобода ограничена, это ограничение можно снять путем расширения списка сценариев. Облегчить процедуру выбора сценариев можно путем введения интеллектуальных решений на основе технологий искусственного интеллекта и механизмов адаптации. Механизм адаптации необходим для компенсации недостаточности априорной информации. Для иллюстрации этого подхода в работе рассмотрим следующие сценарии: минимаксный подход с относительными оценками, прямое голосование и минимизация рисков.

В интеллектуальной системе M принятие решений осуществляется на основании оценочного функционала или управляющей матрицы $P = P(f_{r,j})$, характеризующей выигрыш или проигрыш при принятии решения $r \in R$ если система или среда находится в состоянии j . Для формирования оценочной матрицы в качестве оценочной характеристики ЛПР принимает решение использовать один из перечисленных сценариев.

Минимаксный сценарий с относительными оценками. Элементы оценочной матрицы P_1 задаются как отношение эмпирического значения критерия к его критическому значению $P_1 = C_{emp} / C_{crit}$, тогда если $P_1 < 1$, то наблюдаемое значение критерия принадлежит области принятия гипотезы и гипотеза согласуется с экспериментальными данными. В противном случае будем считать, что наблюдаемое значение критерия принадлежит критической области и нулевую гипотезу отвергают. Применяется минимаксных критерий в соответствии с которым система M решает функцию поиска такого $C \in \Omega$, который обеспечивает $\min(P_1)$.

Сценарий прямого голосования. Все критерии различия имеют одинаковые единичные веса $P(f_{r,j}) = 1 \forall j \in J$, в таком случае система M при определении $r \in R$ руководствуется принципом прямого голосования, когда каждый критерий $C \in \Omega$ «голосует» за определенную реализацию R , а система M сообщает ЛПР результат голосования.

Сценарий минимизации рисков. В качестве оценочной характеристики используется численное выражение степени риска – вероятность ошибки II рода β

(что будет принята неправильная нулевая гипотеза) умноженная на потери, заданные ЛПР. Система M будет находить такое решение, которое достигается при минимально возможном риске.

Каждый из сценариев имеет свои преимущества и недостатки, которые проявляются или не проявляются в зависимости от конкретной ситуации, заранее определить эффективность какого-либо сценария формальными методами не представляется возможным.

Предлагаемая интеллектуальная информационная система адаптивного выбора сценариев параметрического оценивания и обнаружения дивергенции ДРМ ПС обладает свойствами обучения и адаптации, в зависимости от эффективности применяемых сценариев, и направлена на повышение обоснованности при выборе критерия различия, что позволяет ЛПР выбирать оптимальный сценарий анализа данных с учетом допустимых рисков и потерь.

Дивергенция данных рассматривается как совокупность количественных оценок их аномальности, аномальности, сдвигов центров распределений, неоднородности дисперсий, эффектов значительной асимметричности. Дивергенция как комплексная оценка, которая в целом определяет качество имеющихся данных с точки зрения наличия в них отмеченных эффектов. В англоязычной литературе этот термин *data divergence* встречается, однако он больше соответствует описанию ошибки вследствие устаревания данных, ошибки отклика, кодирования или ввода данных, различия в кодировании и пр.

Данная система сценариев является открытой для пополнения другими параметрическими методами выявления различий и эвристическими подходами, позволяющими ЛПР более обоснованно осуществлять выбор информационного сценария.

Выбор конкретного сценария изначально осуществляется в соответствии с предпочтениями ЛПР, а затем, по мере накопления опыта использования в данной проблемной области и, учитывая частоту успешных и неуспешных использований сценариев, можно построить интеллектуальный регулятор с адаптивным управлением по автоматизации выбора.

Предложены сценарии: минимаксный с относительными оценками, прямого голосования, минимизации рисков, что позволило повысить вовлеченность ЛПР в процесс принятия решений по наличию аномалий ПС, и поднять уровень интеллектуализации в целом. Предложенный список сценариев не является исчерпывающим и может быть расширен в зависимости от задачи мониторинга и наблюдаемой среды.

2.10 Выводы по главе 2

В ходе разработки методики решения А-задачи ПС были получены следующие основные результаты:

1. Предложена методика обнаружения аномалий ДРМ ПС различных типов. Разработаны принципы представления ДРМ в виде аналитических функций с целью анализа нерегулярных и пропущенных данных, сформулированы правила организационного характера по контролю и управлению данным процессом. Методика упорядочивает и конкретизирует идеи, связанные с современным представлением о технологии обнаружения аномалий в данных пространственного мониторинга ПС. Она удовлетворяет предъявляемым к ней требованиям (см. 1.3.1).

2. Осуществлена кластеризация задач обнаружения аномалий мониторинга ПС и разработаны адаптивные проблемно-ориентированные решения для обнаружения аномалий выделенных классов. В особенности предложены решения для ситуаций с различными объемами данных, наличием априорной информации о типах аномалий, требованиями к обработке в реальном времени.

3. Определены наиболее информативные наборы критериев оценки эффективности отдельных реализаций методов представления данных и обнаружения аномалий, определяемых для каждой конкретной постановки А-задачи и особенностей ее возникновения.

4. Предложен пул аналитических и приближенных методов представления ДРМ ПС, представляющий собой набор правил, осуществляющих отображение выборки нерегулярных и пропущенных данных в описывающую их поверхность в

n-мерном пространстве, являющуюся входной для методов обнаружения аномалий. Она обобщает и существенно расширяет разнообразие методов, адаптируя их для решения А-задачи.

5. Разработаны принципы структуризации ДРМ ПС как полей данных высокой плотности, позволяющие осуществлять свертку отдельных объектов ПС в описывающие их векторы в n-мерном пространстве, обладающие высокой сходимостью с точки зрения решения А-задачи. Произведенные эксперименты показали, что предложенный способ комбинирования сверток разной природы позволяет трансформировать входной поток в данные, размер которых составляет 1% от первоначального объема, и осуществлять сравнение этих полей данных по характеризующим их векторам свертки.

6. Предложен полимодельный динамически конфигурируемый комплекс обнаружения аномалий произвольного типа, который позволил реализовать принцип адаптивности системы поддержки принятия решений к объективно нормальным изменениям в состоянии ПС. Этого удалось достичь за счет адаптивной подстройки и обучения классификатора в процессе эксплуатации, а также за счет ввода сценариев с различными целевыми критериями точности и полноты, выбираемых ЛПР.

ГЛАВА 3.

ИССЛЕДОВАНИЕ ФУНКЦИОНАЛЬНЫХ СВОЙСТВ ОСНОВНЫХ МОДУЛЕЙ
СИСТЕМЫ ОБНАРУЖЕНИЯ АНОМАЛИЙ В ДАННЫХ МОНИТОРИНГА
ПРИРОДНОЙ СРЕДЫ

В данной главе рассматриваются алгоритмическое и программно-техническое обеспечение модульной системы, служащее для поддержки принятия решений, определяющее математический аппарат решения А-задачи на основе проблемно-ориентированных ИМ, анализа многомерных скалярных и векторных полей ДРМ, характеризующих состояние ПС. Исследования направлены на повышение информационной надежности приборов и средств мониторинга ПС при обнаружении N-аномалий, а также диагностики отдельных измерительных каналов приборов контроля ПС при обнаружении M-аномалий.

Модульность предложенной системы обнаружения аномалий ПС обеспечивает функциональную гибкость, относительную независимость компонентов, облегчающую интеграцию в уже существующие системы, легкость масштабирования и структурную гибкость.

Предложенные алгоритмы основаны на многомерных вероятностно-статистических, информационных, аналитических и ИМ, определяющих нормальное и аномальное состояния ПС и задающих критерии, рассчитываемые по ДРМ, и свидетельствующих о наличии значимых отклонений в ПС или средствах ее контроля. Их автоматизация позволит повысить эффективность и качество проводимых исследований в областях экологического контроля и оптимизации природопользования.

Предложен специализированный проблемно-ориентированный лабораторный стенд, осуществляющий численное моделирование, оценку свойств моделей, численный расчет для исследования свойств моделей, численную валидацию и тестирование моделей, обнаружение свойств моделей, оценивание ошибок I и II рода.

Глава посвящена методам тонкого анализа свойств моделей на основе общих положений теории тестирования, валидации, целью которых является определение рабочих точек, запасов чувствительности, областей устойчивости, когда заданы допустимые вероятности ошибок I и II рода.

3.1 Анализ свойств функциональных модулей обнаружения аномалий в данных мониторинга

Построение СППР осуществляется на основе:

- тщательного изучения связей типовых аномалий ПС, формальная постановка задач обнаружения которых представлена в разделе 1.1.2 ,
- кластеризации общих направлений решения А-задач, представленных в 1.1.2 ,
- синтеза конкретного пула методов в полимодельный комплекс, представленный в разделе 2.9 и обладающий динамически конфигурируемой структурой, адаптивной к конкретной целевой задаче и характеру поступающих ДРМ ПС.

С этой целью было спроектировано программное обеспечение, позволяющее осуществлять функции захвата, структуризации ДРМ ПС, накопления структурированных данных, визуального анализа данных и обнаружения возмущений в скалярных и векторных полях мониторинговых наблюдений и автоматизированного обнаружения N-аномалий ПС и M-аномалий приборов и средств мониторинга ПС. Принципы его работы и архитектура рассматриваются в главе 4 диссертации.

Для решения задачи исследования свойств методов решения А-задачи было разработано программно-техническое обеспечение лабораторного стенда, позволяющее накапливать структурированную информацию о ДРМ ПС, и генерировать управляемые возмущения в этих данных с целью получения численных оценок эффективности совокупностей методов решения А-задачи. Совокупность методов получения численных оценок, эффективности которых, для

конкретного типового случая окажутся максимальными, выступает основой, на которой осуществляется синтез оптимальной проблемно-ориентированной СППР.

Использование предложенной технической реализации подхода позволяет автоматизировано строить СППР, наилучшим образом адаптированные для конкретной целевой задачи исследования, и реализует принципы расширяемости и адаптивности, предъявленные в разделе 1.3.2 .

3.1.1 Описание модуля лабораторного стенда тестовой среды

Программно-техническое обеспечение лабораторного стенда включает:

– предметные модули имитационного моделирования ПС, откалиброванные по данным многолетнего мониторинга пространственных и временных полей океанографических характеристик Севастопольской бухты [93]. Система обладает возможностью генерировать управляемые возмущения, на фоне типовых ДРМ, характерные для ПС, что осуществляется добавлением аддитивной или мультипликативной составляющей одной или нескольких типовых аномалий распределенных во времени и пространстве (точечные выбросы, волновые флуктуации, сдвиг распределения, появление шума, снижение энтропии),

– структурно и параметрически гибкий модуль имитационного моделирования процессов, протекающих в приборах и средствах мониторинга ПС, учитывающий особенности информационного обмена в таких системах [159] и сетевой модели OSI/ISO [160],

– аппаратный встраиваемый модуль, реализованный на базе микроконтроллера, осуществляющий захват профиля эталонного канала или сенсора и сравнение его с тестовым с использованием информационного критерия в режиме реального времени с целью обнаружения М-аномалий [161],

– программно-аппаратное решение для круглосуточного сбора, структуризации, визуализации, нормализации данных, получаемых в реальном времени от автономного биосенсорного комплекса экологического мониторинга, осуществляющий расчет и контроль информационных метрик отдельных каналов

для обнаружения М-аномалий. В данных, очищенных от М-аномалий, модуль осуществляет расчет профиля текущей активности биосенсоров и сравнение его с данными моделирования типовой динамической временной модели активности биосенсоров для обнаружения N-аномалий среды. Модуль предоставляет ЛПР возможность выбора метода расчета метрик степени аномальности и визуального контроля корректности выбранной совокупности методов [161].

Основной технологией, применяемой в лаборантом стенде, является ИМ, как традиционно активно и успешно применяющейся для решения широкого задач: оптимизации бизнес-процессов [153], инженерно-технических областей [154], критических объектов и процессов, когда цена ошибки велика [155], мониторинга окружающей среды, распространения аномалий при деградиационных воздействиях [156] и других фундаментальных и прикладных областей человеческой деятельности [157]. В литературе активно освещается использование подхода ИМ в CAD/CAM системах для оптимизации внутренней структуры сложных объектов, формирования управляющих правил в СППР, в системах ИМ общего назначения для оценки эффективности систем управления ПТС [158]. Вслед за которыми ИМ применяется для повышения обоснованности СППР при решении А-задач ПС.

Повышение обоснованности принятия решений по выбору методов при синтезе СППР во многом достигается за счет визуализации моделирования. Должным образом построенная визуализация также повышает степень информированности и обоснованности при принятии решений о возникновении аномалий в ДРМ ПС и оценки динамики протекания деградиационных процессов во времени. Наглядной и достаточно универсальной моделью распространения аномалии, распределенной во времени и пространстве, как деградиационного процесса, может послужить процесс независимого движения множества источников аномалий A в факторном пространстве M , оставляющих за собой «аномальное пятно» (АП) в среде E . На рисунке 3.1 представлена структурная схема моделирующего стенда, реализующего модель протекания деградиационных процессов.

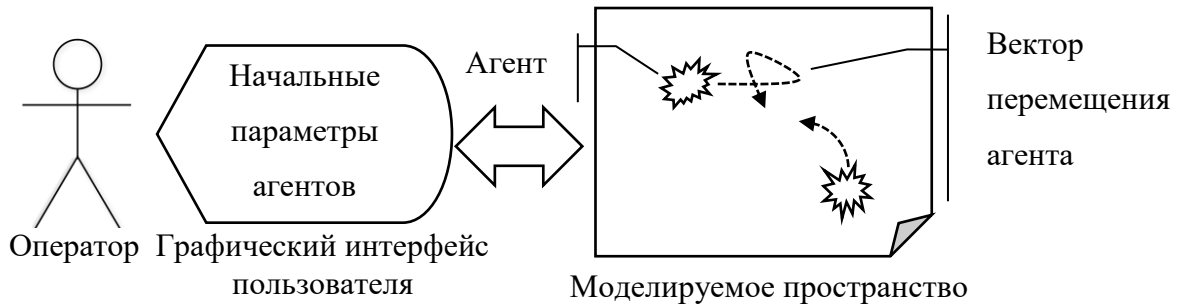


Рисунок 3.1. Структурная схема стенда моделирования управляемых возмущений ПС

Зададим математическую модель агентов моделирования A , распространяющихся в непрерывном пространстве M , отражающем физическое евклидово пространство. Здесь и далее будем рассматривать движение агентов на плоскости, описываемом в пространстве параметров x, y, t – широты, долготы и времени соответственно. Рассмотрим целочисленную решетку \mathcal{R}^2 , в которой задано множество агентов $a_i \in A \mid i \in [0; |A|]$, где $|A| = N$ – параметрически заданное число агентов. Для каждого агента a_i существует начальная точка в пространстве M , заданная парой $\langle x_{i,t}, y_{i,t} \rangle$, такая что определяет расположение i -го агента в момент $t = t_0$ и задается отображением $C: a_i, t \rightarrow c_{i,t} \mid \forall i \in [0; |A|], t \in [0, T]$, где T – общее модельное время. Для упрощения дальнейших выкладок $C(a_i, t)$ будем обозначать как $c_{i,t}$. Любая точка физического непрерывного пространства M с некоторыми допущениями может быть отображена в соответствующую точку дискретного моделируемого пространства \mathcal{R}^2 , $M \rightarrow \mathcal{R}^2$ причем масштаб целочисленной решетки \mathcal{R}^2 определяет погрешность дискретизации расположения агента. Любой агент a_i , стартуя из точки $c_{i,t}$ за время Δt , переходит в точку $c_{i,t+\Delta t}$, причем правило перехода определяется компонентами вектора перемещения $d_{i,t} = \langle l_{i,t}, r_{i,t} \rangle$ – длиной и направлением. Для упрощения процесса функционирования модели будем считать t целочисленным и $\Delta t = 1$, таким образом получим модель с дискретным временем, что решает проблему

неопределенности $\lim_{\Delta t \rightarrow 0} \frac{c_{i,t+\Delta t} - c_{i,t}}{\Delta t}$, обусловленной стохастической природой компонентов вектора перемещения $d_{i,t}$. В условиях принятых допущений и ограничений, движение агента a_i будет определяться некоторым функционалом $D: c_{i,t}, d_{i,t} \rightarrow c_{i,t+1}$, причем реализация D задается исследователем в соответствии с его представлениями о предметной области и зависит от физической природы агентов и среды.

В отсутствии априорной информации о движении АП в пространстве будем считать его случайным. Одной из эффективных часто используемых моделей движения объектов в n -мерном факторном пространстве является модель случайных блужданий (СБ). На рисунке 3.2 представлена визуализация трека движения агента в пространстве.

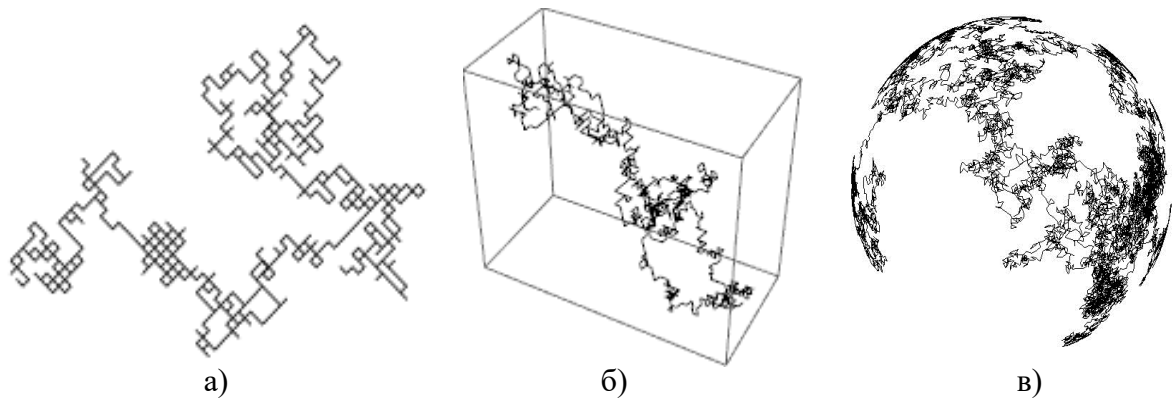


Рисунок 3.2. Визуализация трека движения агента АП: а) на двумерной плоскости, б) в трехмерном пространстве, в) на поверхности сферы

Реализация функционала случайного движения D для агента в евклидовом пространстве в случае, когда его перемещения описываются моделью СБ, выполняется в соответствии с соотношением (3.1). В процессе движения агентов a_i допускается самопересечение их траекторий.

$$\begin{aligned}
D: c_{i,t} \langle x_{i,t}, y_{i,t} \rangle, d_{i,t} \langle l_{i,t}, r_{i,t} \rangle &\rightarrow c_{i,t+1} \langle x_{i,t+1}, y_{i,t+1} \rangle \\
\begin{cases} x_{i,t+1} = x_{i,t} + l_{i,t} \cos(r_{i,t}) \\ y_{i,t+1} = y_{i,t} + l_{i,t} \sin(r_{i,t}) \end{cases} &, \\
r_{i,t} \sim U[0; 2\pi] & \\
l_{i,t} \sim U[0; l_{\max}] &
\end{aligned} \tag{3.1}$$

где функция $U[0; 2\pi]$ задает случайную величину, заданную равномерным случайным распределением на отрезке $[0; 2\pi]$, l_{\max} – максимальная длина шага, на который может сместиться агент a_i за время $\Delta t = 1$.

В процессе моделирования любой агент a_i в каждый момент времени $t \in [0, T]$ генерирует в среде E в окрестности ξ АП радиусом R , заданным пользователем. Выбор R обуславливается природой решаемой задачи, но в общем случае адекватный реальности результат моделирования можно достичь при $R \gg l_{\max}$.

Параметрическое представление лабораторного стенда, построенного по представленной выше модели, будет иметь вид:

$$L \langle N, C \{c_{1,0} \dots c_{n,0}\}, T, R, l_{\max} \rangle, \tag{3.2}$$

где N – число агентов, $C \{c_{1,0} \dots c_{n,0}\}$ – начальные положения агентов, T – внутримодельное время, R – радиус АП агента, l_{\max} – максимальный шаг агента за время Δt .

С точки зрения удобства пользователя, лабораторный стенд позволяет параметрически задавать неограниченное число агентов системы моделирования, на графическом интерфейсе (GUI) курсором указывать их начальное положение, поддерживает масштабирование скорости течения внутримодельного времени, и, в отличие от альтернативных моделей, позволяет задать радиус воздействия агента на среду и вероятности возникновения АП при повторном прохождении агента в области, на которой уже было произведено деградиационное воздействие.

Визуальное анимированное представление процесса моделирования дает возможность развить качественные представления о динамике деградиционных воздействий и порождаемых ими процессов. Это позволит ЛПР выбрать оптимальные совокупности методов решения А-задач для каждого конкретного типа процессов, обладающие наибольшей адекватностью и достоверностью.

3.1.2 Результаты анализа свойств функционального модуля оценки отдельных характеристик методами обнаружения М-аномалий в данных мониторинга природной среды

Целью данных исследований являлось определить типовые тенденции и закономерности проявления М-аномалий в ДРМ при нормальном функционировании ПС при отсутствии N-аномалий и значимых антропогенных воздействий.

Известно, что во многих современных промышленных приборах и средствах контроля ПС, и, в частности, экологического контроля, присутствуют решения, позволяющие практически полностью исключить одну из разновидностей М-аномалий, а именно – ошибки, возникающие при передаче данных в каналах связи за счет использования контрольных сумм и избыточного кодирования.

При этом, слабо представлены методы обнаружения случаев М-аномалий. Когда показания отдельных измерительных каналов становятся неадекватными состоянию ПС, сигнал становится слаборазличим на фоне шумов или заменяется шумом, физической деградации элементов измерительного канала и их периодическим отказам. Особенно остро эта проблема касается критических средств долговременного экологического контроля, измерительные каналы которых могут искажать свои показания из-за агрессивных условий среды и приборов, использующих биосенсоры, которые с одной стороны гораздо более устойчивы к условиям ПС, но обладают трудноформализуемой передаточной характеристикой состояния окружающей ПС.

Осуществлена разработка вычислительно простого подхода, предназначенного для выполнения на локальном интегрируемом модуле автоматизированного самоконтроля приборов и средств контроля ПС, детектирования выбросов ключевых параметров и М-аномалий, характеризующих состояние средства мониторинга с целью предупреждения возникновения ошибок I и II рода (ложной тревоги или пропуска аварийного состояния ПС [162]).

К подходам, которые можно адаптировать для решения вышеуказанной задачи и удовлетворяющим предъявленные требования, относятся:

- кластеризация и основанная на ней адаптивная пороговая фильтрация,
- расчет информационной меры для данных измерительного канала как источника случайного процесса.

Задачи обнаружения аномалий и кластеризации имеют схожую формулировку и метрики качества, такие как чувствительность и полнота алгоритма классификации [163]. При этом в А-задаче имеет место значительный дисбаланс классов, когда аномальные состояния ПС относительно редки, вследствие чего сбор статистики ключевых параметров при А-событиях становится затруднительным.

Проверяется гипотеза зависимости чувствительности системы бинарного классификатора к возникающим аномалиям (управляемым возмущениям) во входном потоке объектов с использованием ортогонального фильтра в зависимости от порога α и размера обучающей выборки n . В качестве метрики качества используется соотношение между долей объектов от общего количества носителей признака, верно классифицированных как несущих признак.

Для иллюстрации работы предложенного подхода в качестве исходных данных используется выборка данных зондирования, выполненного *SVP* зондом в акватории вод г. Севастополя, при этом заданы наборы данных: *A* – заведомо содержащий аномальные значения и *D* – содержащий немаркированные как аномальные, так и корректные данные.

Базовый подход СППР заключается в обнаружении значимых отклонений между ДРМ реального объекта наблюдения его математической моделью,

адаптивно подстраивающейся под изменения состояния ПС, проявляющиеся в статистически значимом изменении параметров наблюдений ПС.

Каждая запись в наборе данных, в общем случае, представляет собой объект в многомерном пространстве поля наблюдений, которое может характеризоваться координатой и временем. Например, при анализе значения температуры во временном ряду выбросами можно считать, как аномально большие и малые значения, так и области с нехарактерными изменениями температурного градиента во времени. В это время, новизна обычно характеризуется принципиально новым поведением объекта и оценивается как мера соответствия очередного поступившего в СППР объекта множеству уже имеющихся объектов [164–166].

Информационная метрика оценки параметров адаптивного фильтра. Для расчета меры соответствия могут использоваться различные методы статистического анализа, такие как: метод χ^2 , регрессионный анализ, пороговый анализ и др. Одним из наиболее часто используемых является метод среднеквадратических отклонений, в котором на основе подмножества предыдущих значений $[d_{i-n}; d_i]$ размером n множества данных $[d_0; d_{\max}] \in D$ строится доверительный интервал $[m - k \cdot s; m + k \cdot s]$ для каждой $d_i \in D$, где m – математическое ожидание; s – среднеквадратическое отклонение; k – коэффициент ширины интервала. За счет использования скользящего окна, ширина которого – n определяет скорость перестроения СППР при естественных изменениях в объекте наблюдения, проявляется адаптационная способность метода, позволяющая игнорировать данные о новом состоянии объекта, не носящие системный характер.

Такой подход позволяет детектировать широкий класс аномальных состояний объекта мониторинга, проявляющихся в виде резких перепадов наблюдаемых параметров относительно их стационарного положения. При этом, в случае постепенного изменения параметров во время перехода объекта наблюдения в предаварийное, а затем в аварийное состояние или их целенаправленном изменении со стороны третьих сил, возникает ситуация

переобучения СППР, характеризующаяся ее функциональной неспособностью к детектированию аномалий. Устранение указанной уязвимости метода возможно следующими способами:

- увеличить размер n обучающей выборки, что в свою очередь приведет к снижению адаптационных свойств СППР;
- определить защищенную область репрезентативной выборки данных системы, заведомо находящейся в стационарном состоянии.

Сбор обучающей выборки, содержащей все возможные N-аномалии ПС крайне трудоемкий, и, зачастую, нереализуемый процесс, что обусловлено низкой частотой их возникновения. Поэтому, на практике применяется обучение на выборке данных достоверно характеризующей нормальное множество состояний ПС.

Определим основные причины возникновения аномалий в наборах данных:

- ошибки в исходных данных, погрешности измерений, неверное округление, дискретизация аналогового сигнала, некорректность формата представления данных;
- наличие источника шума, неверно классифицированные объекты;
- наличие объектов сторонних выборок (в частности, записи вышедших из строя средств мониторинга, датчиков).

Аномалиями обычно являются экстремальные значения параметров мониторинга, но это не всегда так. Приведем пример визуализации участка данных зондирования, выполненного в водах акватории г. Севастополя. Построение производится в пространстве, характеризуемом двумя признаками, температурой и скоростью звука в воде, для глубины 0–20 м. Из рисунка 3.3 видно, что данные содержат как небольшие шумы, расположенные в окрестности линии тренда, так и большие возмущения, модуль которых доходит до 10 м/с.

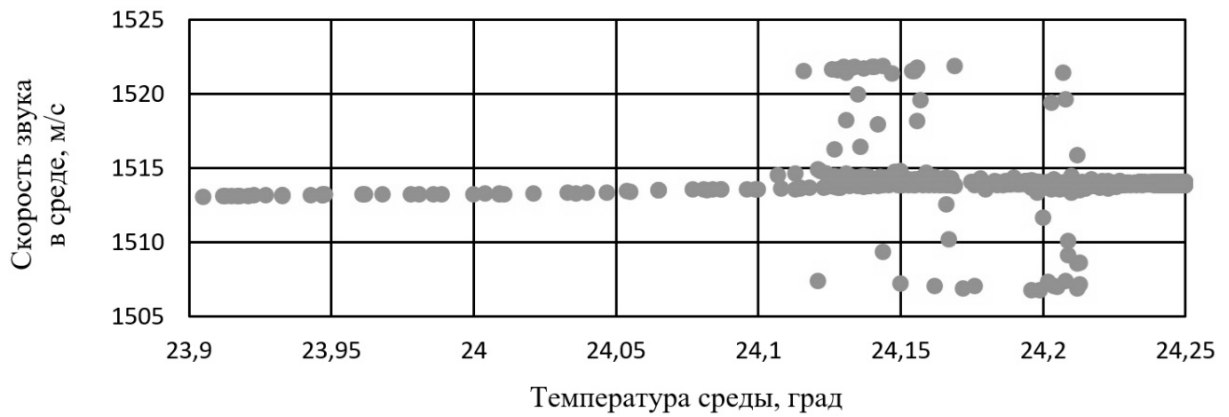


Рисунок 3.3. Точечная диаграмма зависимости скорости звука от температуры по данным 96-го рейса НИС «Профессор Водяницкий», замер выполнен 18.07.2017 в 20:30, координаты станции 44°29'15.4" с.ш. 33°30'37.4" в.д.

Если относительно слабые выбросы не сильно искажают границы наблюдаемого кластера и на физическом уровне обуславливаются наличием тонкой микроструктуры, стратификации поверхностного слоя, то сильные выбросы действительно являются аномальными, что легко проверить с использованием упрощенного термодинамического уравнения состояния морской воды TEOS-10 [167]:

$$S(C, T, P) = \frac{w_1 + w_2 T - w_3 T^2 + w_4 T^3 + w_5 |P| + w_6 P^2 - w_7 |P^3| - w_8 T |P^3| + w_9 T^2 |P| - C}{-w_{10} + w_{11} T - w_{12} T^2 - w_{13} |P|}$$

где C – скорость звука; T – температура; P – давление; значения коэффициентов w составляют: $w_1 = 1402.5$; $w_2 = 5$; $w_3 = 0.0544$; $w_4 = 2.1 \cdot 10^{-4}$; $w_5 = 1.56 \cdot 10^{-2}$; $w_6 = 2.55 \cdot 10^{-7}$; $w_7 = 7.3 \cdot 10^{-12}$; $w_8 = 9.5 \cdot 10^{-13}$; $w_9 = 3 \cdot 10^{-7}$; $w_{10} = 1.33$; $w_{11} = 1.23 \cdot 10^{-2}$; $w_{12} = 8.7 \cdot 10^{-5}$; $w_{13} = 1.43 \cdot 10^{-5}$.

Согласно представленному выше уравнению, область, соответствующая всплескам пары скорости звука и температуры возможна только при уровне солёности, значительно выходящем за рамки океанографического диапазона, доходя в абсолютных значениях до 10‰ и с перепадом до 7‰, что несвойственно для региона.

В силу вышесказанного можно сделать вывод что возникновение точечных возмущений, рассмотренных в данном примере не связано с изменениями температуры и давления, а является М-аномалией, предположительно обусловленной наличием взвесей в водной среде. После обнаружения М-аномалии исследователь либо осуществляет дальнейший ее анализ для определения причины возникновения А-события, например, определения типа взвеси (рассмотрению этого вопроса посвящена НИР АААА-А17-117021310008-4), либо осуществляет устранение обнаруженной аномалии ДРМ ПС.

Одним из способов устранения таких аномалий является выполнение линейного преобразования солености так, чтобы оно оставалось в допустимом интервале и восстановления, с использованием симплекс метода, соответствующего ему пространства температуры и скорости звука. Очевидно, что применение такого преобразования продуктивно производить не для всей области данных, что в противном случае приведет к значительному искажению результата, а с использованием порогового фильтра. Учитывая нелинейность возникающих искажений и недетерминированный характер вызывающих их аномалий, оптимальным будет применение ортогонального адаптивного фильтра.

Предположим, что входной поток данных D , задан на интервале $[d_0; d_{\max}]$, аномалии a_i в данных являются случайными не коррелированными с d_i величинами, принимающими значение $a_j = 0$ с вероятностью $p \geq 0.99$ на всей области определения и нулевым средним значением $M[a_j] = 0$.

Ортогонализация предназначенного для обработки потока данных выполняется на основе дискретного преобразования Фурье (ДПФ), но может производиться и на основе других ортогональных преобразований. Предположим, что входящий поток данных может быть описан функцией с ограниченным спектром, другими словами $F(\nu) = 0$, $|\nu| > \nu_0$ при этом выполняется неравенство Найквиста

$$\Delta_x \leq \frac{1}{2\nu_0}, \quad (3.3)$$

где Δ_x – шаг дискретизации по времени.

Выбирается длина периодической последовательности n , используемая в качестве обучающей выборки, при этом будет рациональным принять $n = 2^m$. Такое допущение позволит использовать алгоритм быстрого ДПФ, что значительно уменьшит число необходимых вычислительных операций при выполнении фильтрации длинных последовательностей $n \geq 10^3$.

Выполним спектральный анализ заданной последовательности данных

$$F_p(l) = \frac{1}{n} \sum_{j=0}^{n-1} f_p(j) e^{-\frac{2\pi i}{n} jl}, \quad (3.4)$$

где $l = 0, \dots, n-1$, $i = \sqrt{-1}$, $F_p(l)$ имеет период n , каждому l соответствует $\omega_l = \Delta_\omega l$, $\Delta_\omega = \frac{2\pi}{\Delta_x n}$ – шаг дискретизации.

Введем функцию фильтрующего множителя, определяющего ошибку фильтрации, уменьшая коэффициенты ДПФ, обусловленные аномалиями

$$W_a(\omega) = \frac{1}{1 + \omega^{2p}}, \quad (3.5)$$

где p – коэффициент крутизны фильтрации для всех $\omega > \omega_0$, ω_0 – граница частотного интервала информационных составляющих входного потока данных.

С использованием функции (3.5) получим вектор фильтрованных коэффициентов ДПФ,

$$F_{filtered}(l) = W_a(l) F_p(l), \quad l \in [0, n-1], \quad (3.6)$$

по значениям которого произведем спектральный синтез, выполнив обратное ДПФ

$$f_{filtered}(j) = \sum_{l=0}^{n-1} F_{filtered}(l) e^{-\frac{2\pi i}{n} jl}, \quad (3.7)$$

где $j \in [0, n-1]$.

Разобьем множество входного потока D на два непересекающихся подмножества D_{normal} и $D_{anomaly}$ такие, что

$$\left\{ \begin{array}{l} d_i \in D_{anomaly} \forall i \in [0, n-1], \text{ где } 1 - \frac{f_{filtered}(i)}{d_i} \geq \alpha \\ d_i \in D_{normal} \forall i \in [0, n-1], \text{ где } 1 - \frac{f_{filtered}(i)}{d_i} < \alpha \end{array} \right., \quad (3.8)$$

где α – порог распознавания аномалии.

Высокая скорость сходимости в численных методах, вообще говоря, является преимуществом, вследствие того, что оптимум рабочей функции достигается при минимальном числе итераций. Следует отметить, что для СППР это справедливо, когда рабочая функция, на которой происходит поиск оптимума, заранее известна, но на практике это не всегда так. Построим точечную диаграмму по отфильтрованным данным D_{normal} (рис. 3.4).

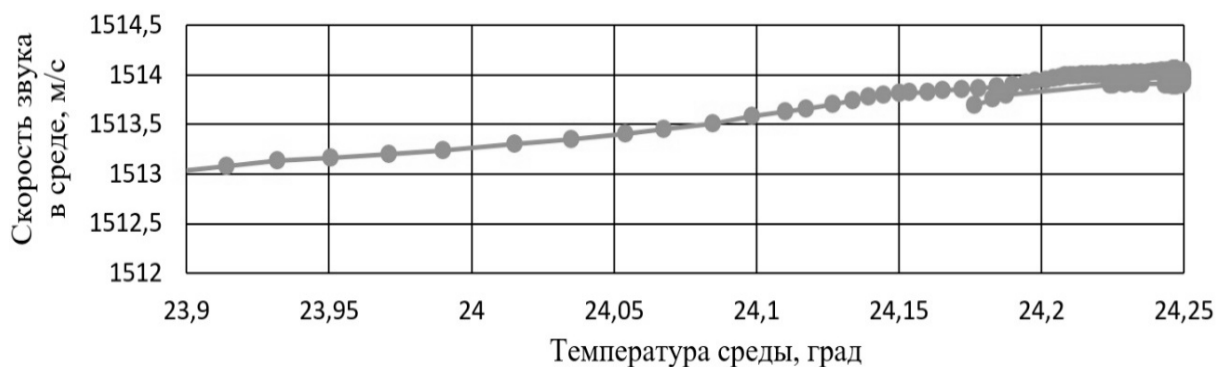


Рисунок 3.4. Точечная диаграмма зависимости скорости звука от температуры после фильтрации по данным 96-го рейса НИС «Профессор Водяницкий», замер выполнен 18.07.2017 в 20:30,, координаты станции 44°29'15.4" с.ш., 33°30'37.4" в.д.

Функция вычисляется приближенно на основе входного потока данных и постоянно дополняется и адаптируется при поступлении новых данных. В таком случае, при снижении скорости сходимости имеет место фильтрация входного сигнала, которая снижает шумовую составляющую.

Произведем анализ зависимости размера обучающей выборки n на характер поверхности функции отклонения фактических данных от прогнозируемых (функция аномалии f), в силу (3.7) и (3.8), для $n = 20$ получим (рис. 3.5).

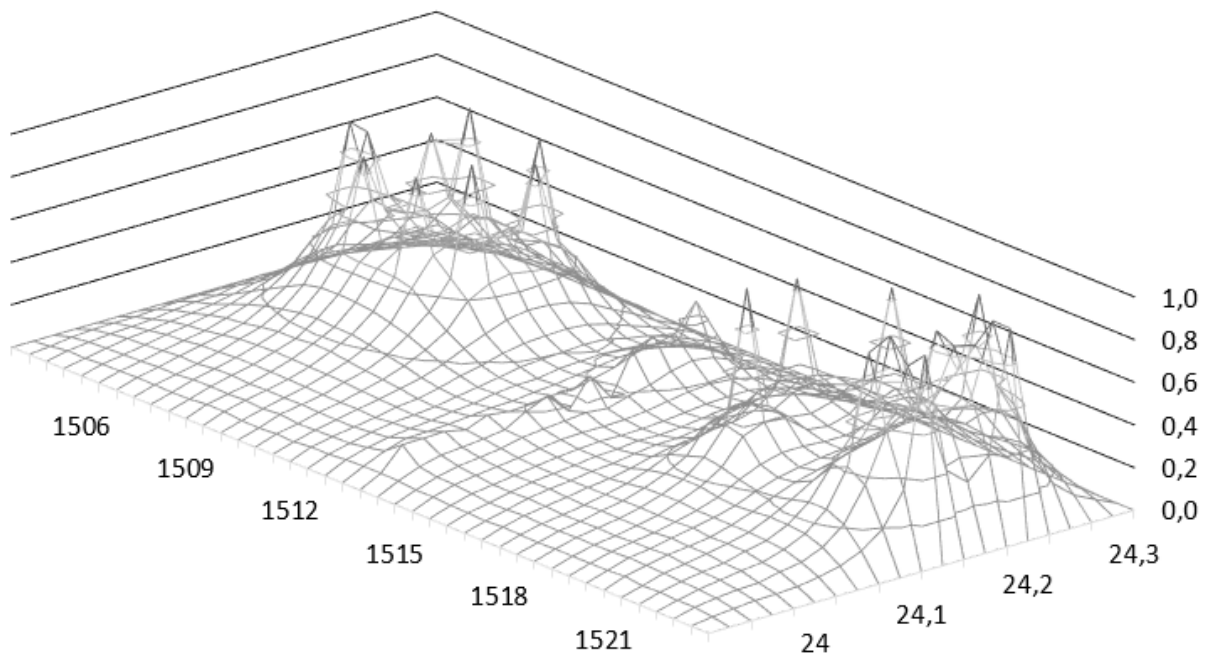


Рисунок 3.5. График поверхности функции аномалии f в пространстве скорости звука и температуры при размере обучающей выборки $n = 20$, где $f = 0$ соответствует минимальной вероятности аномалии в данной области и $f = 1$ – максимальной вероятности

При малых значениях n достаточно быстро наблюдается эффект переобучения, проявляющийся в наличии пиковых значений $f = 1$ в областях, где уже были обнаружены аномалии.

Выбор порогового уровня отклонения α крайне важный процесс, по причине того, что он определяет какие именно данные будут признаны аномальными. Выбор основывается на экспертных решениях и имеющейся априорной информации об аномалиях. В соответствие с выбранной моделью аномалий, была

принята вероятность того, что очередной поступивший в СППР объект не является аномалией с вероятностью $p \geq 0.99$. Это позволяет построить управляющую матрицу бинарного классификатора при низком $\alpha = 0.2$ (рис. 3.6) и высоком $\alpha = 0.6$ (рис. 3.7) порогах.

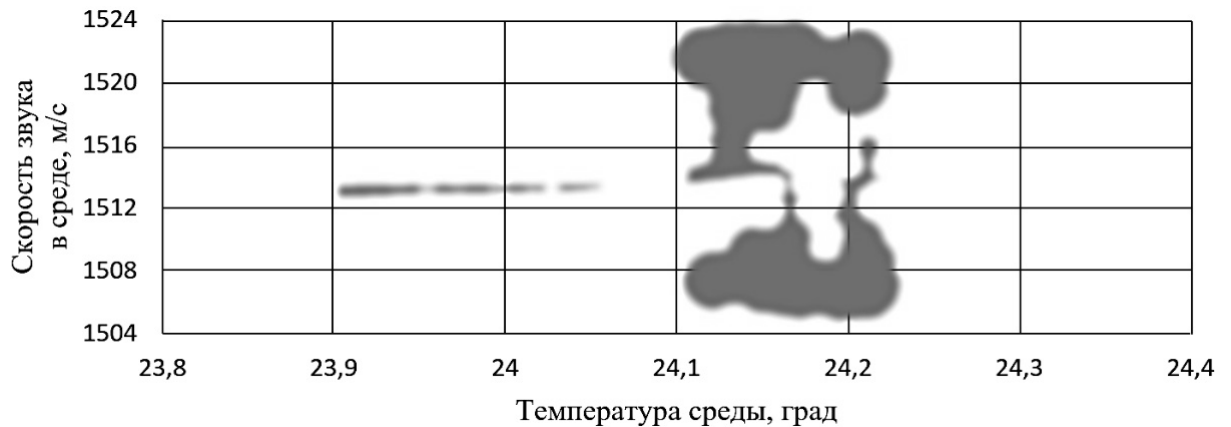


Рисунок 3.6. Диаграмма управляющей матрицы бинарного классификатора, где темным пятном обозначены области, данные в которых признаются аномальными, при пороге $\alpha = 0.2$



Рисунок 3.7. Диаграмма управляющей матрицы бинарного классификатора, где темным пятном обозначены области, данные в которых признаются аномальными, при пороге $\alpha = 0.6$

В результате решения задачи оптимизации получим результирующую для СППР диаграмму управляющей матрицы бинарного классификатора (рис. 3.8).

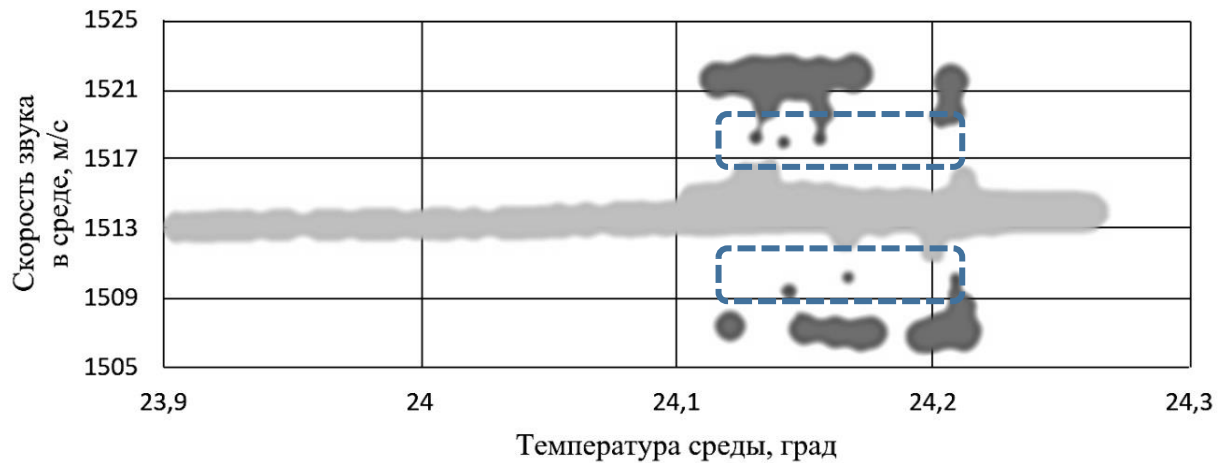


Рисунок 3.8. Результирующая диаграмма управляющей матрицы бинарного классификатора, где темно-серым пятном обозначены области, данные в которых признаются аномальными, светло-серым пятном – истинные данные, пунктиром – зона неопределенности

Предложенный, и продемонстрированный на конкретном примере, подход к решению задачи оперативного обнаружения М-аномалий в ДРМ ПС с использованием адаптивной ортогональной цифровой фильтрации и математического моделирования ориентирован на работу в режиме реального времени. Особенности подхода проиллюстрированы на примере разложения с использованием ДПФ, но этот же подход может быть использован и при других ортогональных преобразованиях. Продемонстрирована методика построения бинарного классификатора нормальных и аномальных зон на его основе.

Для упрощения графической интерпретации результата, здесь и в реализуемой программе, на диаграммах управляющих матриц выделяются три области «нормальная», «аномальная» и «зона неопределенности». Построение разделяющих поверхностей для этих зон служит основой СППР по обнаружению аномалий ПС.

Работа бинарного классификатора проиллюстрирована на выборке данных зондирования скорости звука и температуры, полученных с использованием *SVP* зонда в акватории вод г. Севастополя.

Исследование зависимости чувствительности системы бинарного классификатора к возникающим аномалиям (управляемым возмущениям) во входном потоке объектов с использованием ортогонального фильтра, в

зависимости от размера обучающей выборки n , показало, что малые значения $n \leq 16$ обеспечивают высокую скорость реакции на возникновение аномалий, но при этом быстро возникает эффект переобучения и относительно большие вероятности ложных срабатываний. Размеры обучающей выборки $n \geq 256$ значительно в меньшей степени подвержены переобучению, но обладают меньшей чувствительностью, что, в свою очередь, накладывает повышенные требования к точности выбора порога α . Выбор параметра n остается открытым, он остается на усмотрение ЛПР и должен быть достаточно большим для того, чтобы обеспечить достаточную чувствительность фильтра, обеспечивая возможность адаптивной настройки порога α и во многом определять вычислительную сложность алгоритма.

Предложенный способ построения адаптивной управляющей матрицы бинарного классификатора при заданном пороге α , позволяющей оперативно принимать решения о возникновении аномалии в данных наблюдений, их отсутствии или зоны неопределенности. Показан пример использования модуля для отделения М-аномалий от выбросов ДРМ, которые можно либо подвергнуть дальнейшему анализу с целью обнаружения новых знаний об их источнике, либо устранить применив предложенный подход ортогональной цифровой фильтрации.

Представленный модуль разработан для использования в СППР, реализован в виде бинарного классификатора, блоков задания порога α и размера обучающей выборки n . Указанные блоки используются для определения минимально необходимого объема выборки ДРМ, обеспечивающего достижение допустимых вероятностей ошибок I и II рода. Модуль предназначен для нахождения пограничных зон неуверенного распознавания М-аномалий, с целью определения границ применения СППР, построенной на его основе, и повышения реактивности обнаружения критических А-событий ПС.

3.1.3 Результаты анализа свойств функционального модуля оценки совокупности характеристик методами обнаружения N-аномалий в нерегулярных и пропущенных данных

Выполним расчеты метрик качества F , W , Q , описанных в 2.3 для предложенных регрессионных моделей, описанных в 2.1 и 2.2, представим сводные результаты в табл. 3.1. Численные расчеты и ранжирование моделей выполнены для исходных данных зондирования профиля температуры, полученных с использованием *SVP* зонда в акватории вод г. Севастополя. Отметим, что линейная и билинейная модели имеют достаточно высокое значение невязки, что ограничивает их использование теми параметрами морской среды, которые имеют невысокую дисперсию. Ранжирование способов построения регрессионных уравнений осуществлено отдельно по каждой метрике, интегральной оценкой служит сумма полученных ими рангов.

Произведенный анализ моделей, задающих функциональные поля параметров водной среды в условиях нерегулярно заданной сетки и пропущенных данных, показал, что наилучшую интегральную оценку получили методы кусочно-полиномиальной сплайновой интерполяции, итерационного сглаживания гауссовым фильтром и линейной регрессионной модели. Модели, использующие регрессионные уравнения более высоких порядков, несмотря на достаточно низкие величины параметра невязки, крайне неправдоподобно описывают поле наблюдений между заданными точками регрессии, что недопустимо для предметной области.

Таблица 3.1. Результаты расчета метрик качества регрессионных моделей

№	Регрессионная модель \ Метрика	F	W	Q	Итоговый ранг
1	Линейная	0.477	0.357	0.392	11
2	Билинейная	0.437	0.540	0.883	13
3	Квадратная	0.199	1.014	1.213	17
4	Биквадратная	$3.23 \cdot 10^{-18}$	8.215	20.527	19
5	Кубическая	$4.84 \cdot 10^{-8}$	7.289	20.449	20
6	Бикубическая	$8.61 \cdot 10^{-13}$	17.782	38.454	23
7	Биквадратная с пороговой фильтрацией	$3.23 \cdot 10^{-18}$	1.591	1.914	16
8	Матричное итерационное гауссово сглаживание	0	0.698	1.195	9
9	Кусочно-полиномиальная сплайновая интерполяция	0	0.425	1.014	7

Предложенные численные метрики качества показали свою работоспособность и могут быть использованы для проверки других моделей, задающих функциональные поля параметров в контексте предметной области мониторинга водной среды, в частности, метрика Q оценивает степень выхода поверхностей, заданных регрессионными уравнениями, за пределы океанографического диапазона.

В рамках предложенной интеллектуальной системы, по историческим данным полей наблюдений параметров водной среды осуществляется выбор оптимальной регрессионной модели, наилучшим образом описывающей функциональное поле наблюдений в заданном регионе. Данная интеллектуальная система служит основой для системы поддержки принятия решений о выборе оптимальных интервалов проведения замеров в различных областях наблюдений в соответствии с вероятностью обнаружения в них аномалий.

Низкая вычислительная сложность пула алгоритмов позволяет реализовать его в виде встраиваемого атомарного модуля в приборы и средства контроля ПС,

носимые устройства, смартфоны и планшеты, которые обычно всегда с пользователем, что дает возможность предоставления сервиса по обнаружению N-аномалий и передачи уведомлений в реальном времени не только во время непосредственного взаимодействия с рабочей станцией, а непрерывно и по требованию.

3.2 Модули системы поддержки принятия решений для обнаружения аномалий

Результаты проведенных исследований моделей предложенных в главе 2, рассмотренные в разделах 3.1.2 и 3.1.3 позволили выявить общие свойства и особенности ДРМ ПС содержащих аномалии. Удалось выявить, что M-аномалии можно обнаруживать методами ортогональной цифровой фильтрации и ассиметричными информационными метриками, рассчитываемыми в проблемно-ориентированных модулях, интегрируемых в сами приборы и средства мониторинга ПС. При этом, методы обнаружения N-аномалий требуют использование значительно более сложного и разнообразного математического аппарата, структура которого динамически перестраивается в зависимости от характера исходных ДРМ ПС и выбранного ЛПР сценария допустимых рисков возникновения ложных срабатываний или пропуска аномалий.

Для оценки степени аномальности данных, характеризующих состояние ПС в некотором пространстве, предложено определять пороговые значения СППР как меры схожести векторов, описывающих типовые аномалии и отображений многослойных регулярных полей данных в n-мерное векторное пространство, где компонентами вектора служат наборы текстурных, инвариантных и сверточных метрик. Пороговые значения определяют границы зон аномальности. Введение нескольких пороговых значения позволит определять степень аномальности и зону неопределенности. В результате ЛПР, использующий СППР для анализа состояния ПС, сможет выявить степень аномальности данных характеризующих ключевые показатели ПС и более обоснованно принять соответствующие управленческие решения.

3.2.1 Модули системы поддержки принятия решений, оценивающие степень аномальности данных мониторинга природной среды

Рассмотрим укрупненно основные идеи разработанной СППР. Она определяет аппарат, состоящий из эмерджентного ансамбля моделей М и N, каждая из которых ориентирована на обнаружение М-аномалий и N-аномалий соответственно. Модели базируются на противоположных, но взаимодополняющих методах. Такое разбиение выполнено в соответствии двойственной природой самих аномалий, рассмотренной в разделе 1.1.1, в том же разделе осуществлена математическая формальная постановка задачи СППР в общем виде.

Модуль обнаружения М-аномалий базируется на использовании типовых профилей нормального состояния ПС, при этом отклонение показателей от этого профиля может свидетельствовать о возникновении аномалий в самих приборах и средствах мониторинга, а своевременное обнаружения таких аномалий приведет к повышению информационной надежности этих приборов и средств.

Модуль обнаружения N-аномалий, напротив, базируется на использовании типовых профилей аномалий, возникающих в ПС, и их проявлений в ДРМ. Такой подход позволит обнаруживать наличие аномалий достаточно сложной формы, распределенных в пространстве и времени, отслеживать протекание аномальных процессов и прогнозировать их дальнейшее течение.

Модели в СППР используются последовательно и взаимодополняют друг друга перекрывая потребность обнаружения всех типовых аномалий мониторинга, рассмотренных подробнее в разделе 1.1.2. Отметим что обнаружение одновременного появления N-аномалий при условии наличия М-аномалий не представляется возможным из-за сложности внутренних связей ПС, и не имеет смысла исходя из принятого определения аномалии как редкого явления. Исходя из этого, события появления М-аномалий и N-аномалий можно считать

несовместными т.е. $P(M + N) = P(M) + P(N) - P(M \cdot N) | P(M \cdot N) \approx 0$, а их совместное появление является редким событием более высокого порядка.

В модуле обнаружения М-аномалий типовые профили сигналов измерительных каналов записываются либо самим ЛПР по требованию, либо периодически по расписанию.

В качестве метрики отклонения текущего профиля от типового была выбрана дивергенция Кульбака-Лейблера, поскольку, в отличие от других метрик (информационная энтропия по Шеннону, t-критерий Стьюдента, U-критерий Манна-Уитни, А/В-тестирования, критерия согласия Колмогорова), она ассиметрична и обладает большей чувствительностью, что позволяет учесть как изменение абсолютных значений характеристик измерительных каналов, так и формы их распределения. Так, мера аномальности профилей измерительных каналов как дивергенции Кульбака-Лейблера, вычисляется по следующей формуле

$$K(p, p') = \sum_{i=1}^n p_i \log \frac{p_i}{p'_i}, \quad (3.9)$$

где n – число интервалов разбиения выборки, p_i – i -й элемент опорной выборки, p'_i – i -й элемент возмущенной выборки.

Использование дивергенции Кульбака-Лейблера в качестве информационной меры расстояния между выборками позволяет частично решить задачу отделения аномалии от нового знания. Аномалии ДРМ могут нести в себе информацию о новом знании, и информационный критерий позволяет поставить этот вопрос (но не решить исчерпывающе), т.к. обнаружение нового знания — это не формальная задача, а творческая, решение которой во многом зависит от ЛПР.

При численном моделировании возможна ситуация, когда некоторые интервалы разбиения не содержат в себе ни одной реализации случайной величины, т.е. для отдельных интервалов $p_i = 0$ или $p'_i = 0$. Это может возникнуть вследствие малого объема выборок – N , большого числа интервалов разбиения – n , особенностей формы распределения и случайной природы процесса

моделирования. Чтобы не рассматривать этот случай отдельно, доопределим области определения функций (3.9) в нулевых точках в виде (3.10) следующим образом:

$$K(p_i, p'_i) = \begin{cases} p_i \log \frac{p_i}{p'_i}, & \text{если } p_i \cdot p'_i > 0 \\ 0, & \text{иначе} \end{cases} \quad (3.10)$$

Для осуществления возможности сравнения профилей различных каналов и унификации методики, продуктивно использовать нормированные информационные метрики как функционалы над множествами, тогда дивергенция Кульбака-Лейблера примет вид

$$IK(p, p') = \left| \sum_{i=1}^n K\left(\frac{p_i}{N}, \frac{p'_i}{N}\right) \right|, \quad (3.11)$$

где N – объем выборок p и p' .

В целом, общие шаги работы СППР по обнаружению N-аномалий заключаются в следующем:

1. Если исходная выборка ДРМ содержит нерегулярные или пропущенные данные, осуществить ее отображение в виде аналитической модели методом из раздела 2.1 так, чтобы комплексный критерий раздела 2.3 был минимален.

2. Если значение комплексного критерия не удовлетворительно, использовать дополнительные ресурсы и осуществить ДРМ отображение в виде численной модели методом из раздела 2.2 .

3. Если имеются априорные данные о типах аномалий, то с использованием метода из раздела 2.4 построить параметрический классификатор, заданный в области определения модели полученной на предыдущем этапе.

4. Если объем ДРМ достаточно большой, число потенциально существующих аномалий велико или отсутствуют априорные данные о них,

использовать метод обнаружения возмущений в полях мониторинговых наблюдений из раздела 2.4 .

Границы зон аномальности задаются значениями метрик дивергенции Кульбака-Лейблера для М-аномалий и евклидового расстояния между векторами n -мерного пространства для N-аномалий, определяемых по доверительным вероятностям, определяемых эмпирически и использованием тестовой среды лабораторного стенда.

По динамике изменения этих метрик ЛПР может делать выводы о состоянии ПС и, в том числе, оценивать адекватность корректность синтезированной СППР. Обоснование модели проводилось на основе экспериментальных данных результаты которых, представлены в разделе 3.1.3 .

3.2.2 Модуль повышения качества принятия решений за счет минимизации риска и оценивания вероятностей ошибок первого и второго рода

Специфика А-задачи ПС подразумевает необходимость принятия решений в условиях неполной информации, как следствие этого возникают риски появления ошибок принятия решений, последствия которых охарактеризуем неотрицательной функцией потерь. В данном разделе описывается модуль выбора пользовательских сценариев, определяющих процесс синтеза правил принятия решений и минимизации функции потерь, направленных на предоставление возможности ЛПР выбирать допустимые риски принятия решений.

В разделе 1.1.1 осуществлена строгая постановка задачи синтеза правила принятия решения $\gamma = \gamma(X)$ по ДРМ. Здесь используются те же обозначения: X – набор ДРМ ПС объемом N , которому соответствует одна из двух функций правдоподобия $W(X|A)$ или $W(X|0)$, связанных с взаимоисключающими событиями A – наличия аномалий, и 0 – отсутствия аномалий. ДРМ x_1, x_2, \dots, x_n – случайные величины, вследствие чего сам набор X случайный и характеризуется плотностью вероятностей $W(X)$. H_1 и H_0 гипотезы о том, что набор ДРМ X

определяется функциями правдоподобия $W(X|0)$ и $W(X|A)$ соответственно, а γ_1 и γ_0 – решения, состоящие в принятии или отклонении гипотезы H_1 . Гипотеза H_0 является простой альтернативой H_1 .

В общем случае пусть функцией потерь является такая неотрицательная функция $L(X^*, \gamma)$, где γ – элемент множества решений, X^* – набор априорных данных о состоянии ПС, который находит свое отражение в наборе эмпирически наблюдаемых данных X . Математическое ожидание функции потерь, в теории решающих функций – функция риска примет вид:

$$R[X^*, \gamma(X)] = \frac{\sum L[X^*, \gamma(X)]}{N} = \int_{\Omega(X)} L[X^*, \gamma(X)] W(X|A) d\Omega(X). \quad (3.12)$$

Стратегия принятия решений, выбираемая ЛПР, определяет соответствующий статистический критерий качества, который минимизирует функцию риска. Выбор конкретной стратегии зависит от конкретизации целевой задачи и характера исходных данных. Сущность рассмотренных сценариев, предложенных в разделе 2.9 заключается в определении такого уравнения поверхности $C(x_1, x_2, \dots, x_N) = const$, разделяющего области Γ_0 и Γ_1 при которой достигаются минимум негативных последствий с точки зрения ЛПР.

Сценарий на основе критерия минимума среднего риска

Подход к решению основан на предположении что имеется набор фактических параметров ПС – X^* , который считается случайной величиной с известной априорной плотностью вероятности $w(X^*)$, тогда средний риск определяется соотношением

$$\begin{aligned}
S_1 &= \int_{\Omega(X^*)} R[X^*, \gamma(X)] W(X^*) d\Omega(X) = \\
&= \int_{\Omega(X^*)} \int_{\Omega(X)} L[X^*, \gamma(X)] W(X | X^*) W(X^*) d\Omega(X) d\Omega(X^*) = . \\
&= \int_{\Omega(X^*)} \int_{\Omega(X)} L[X^*, \gamma(X)] W(X, X^*) d\Omega(X) d\Omega(X^*)
\end{aligned} \tag{3.13}$$

Применительно к постановке А-задачи, приведенной в разделе 1.1.1 , в которой рассматриваются взаимоисключающие события А – наличие аномалий, и 0 – отсутствия аномалий, γ_1 и γ_0 – решения, состоящие в принятии или отклонении гипотезы H_1 соответственно, где априорная вероятность появления аномальных данных равна P_1 , а нормальных данных P_0 , функция потерь преобразуется в матрицу потерь

$$L(x^*_i, \gamma_i) = \begin{pmatrix} L(0, \gamma_0) & L(0, \gamma_1) \\ L(A, \gamma_0) & L(A, \gamma_1) \end{pmatrix}, \tag{3.14}$$

где $L(0, \gamma_0)$ стоимость поощрения за принятие правильного решения γ_0 т.е. принимается гипотеза H_0 об отсутствии аномалии, когда это действительно имеет место; $L(0, \gamma_1)$ стоимость штрафа за ошибочное решение γ_1 при истинности гипотезы H_0 об отсутствии аномалии; $L(A, \gamma_0)$ стоимость штрафа за ошибочное решение γ_0 при истинности гипотезы H_1 о наличии полезного сигнала; $L(A, \gamma_1)$ стоимость поощрения за правильное принятие решения γ_1 т.е. принятие гипотезы H_1 о наличии полезного аномалии.

Таким образом, учитывая априорные вероятности наличия P_1 и отсутствия P_0 аномалии и матрицы потерь (3.14) минимум среднего риска (3.13) $S_1 \rightarrow \min$ будет достигнут в случае если подынтегральная функция – неотрицательная

$$\frac{W(X | A)}{W(X | 0)} > \frac{P_0 [L(0, \gamma_1) - L(0, \gamma_0)]}{P_1 [L(A, \gamma_0) - L(A, \gamma_1)]}. \tag{3.15}$$

Таким образом, уравнение поверхности, разделяющей в этом случае области Γ_0 и Γ_1 , а, следовательно, и определяющей размеры этих областей, имеет вид

$$C(x_1, x_2, \dots, x) = \frac{P_0[L(0, \gamma_1) - L(0, \gamma_0)]}{P_1[L(A, \gamma_0) - L(A, \gamma_1)]}. \quad (3.16)$$

Следовательно, правило принятия решения, основанное на критерии минимума среднего риска, можно сформулировать следующим образом

$$\Lambda = \frac{W(X | A)}{W(X | 0)} \underset{H_0}{\overset{H_1}{>}} C = \frac{P_0[L(0, \gamma_1) - L(0, \gamma_0)]}{P_1[L(A, \gamma_0) - L(A, \gamma_1)]} \begin{matrix} \rightarrow \gamma_1 \\ \rightarrow \gamma_0 \end{matrix}. \quad (3.17)$$

т.е. принимается решение γ_1 (отвергается гипотеза H_0), если отношение правдоподобия больше величины C и принимается решение γ_0 (отвергается гипотеза H_1), если отношение правдоподобия меньше указанной величины.

Сценарий на основе минимаксного критерия качества

В отличие от предыдущего сценария, если априорное распределение $w(x^*)$ неизвестно, то для установления критерия качества принятия решения будет использоваться только функция риска $L[X^*, \gamma(X)]$. Суть критерия состоит в минимизации максимального значения функции риска что является оправданным в критических системах решения А-задач ПС, например, в системах экологического контроля, когда цена ошибки ложного пропуска аномалии выше чем ложной тревоги.

Для определенной выше постановки А-задачи средний риск зависит от априорных вероятностей наличия P_1 и отсутствия аномалии $P_0 = 1 - P_1$. Значение P_1 , при котором достигается наибольшее значение среднего риска S_{\max} , определяется из функции вида

$$\frac{\partial S}{\partial P_1} = L(A, \gamma_0) - L(0, \gamma_0) - \frac{\partial S}{\partial P_1} \int_{\Gamma_1} \{P_1[L(A, \gamma_0) - L(A, \gamma_1)]W(X | A) - P_0[L(0, \gamma_1) - L(0, \gamma_0)]W(X | 0)\} d\Omega(X) = 0 \quad (3.18)$$

Средний риск функционально зависит от P_1 как в явном виде, так и через размеры областей Γ_0 и Γ_1 , которые определяются поверхностью $C(x_1, x_2, \dots, x_N)$. Дифференцирование приводит к уравнению относительно искомого значения P_1

$$L(0, \gamma_0)[1 - \alpha(P_1)] + L(0, \gamma_1)\alpha(P_1) = L(A, \gamma_0)\beta(P_1) + L(A, \gamma_1)[1 - \beta(P_1)], \quad (3.19)$$

где $\alpha(P_1)$ и $\beta(P_1)$ соответственно условные вероятности ошибок I и II рода, как функции от неизвестной априорной вероятности наличия аномалии. Таким образом, правило принятия решения при минимаксном критерии качества будет определяться соотношением

$$\Lambda = \frac{W(X | A)}{W(X | 0)} \underset{H_0}{>} \underset{H_1}{<} C = \frac{(1 - P_1')[L(0, \gamma_1) - L(0, \gamma_0)]}{P_1'[L(A, \gamma_0) - L(A, \gamma_1)]} \begin{matrix} \rightarrow \gamma_1 \\ \rightarrow \gamma_0 \end{matrix}, \quad (3.20)$$

где P_1' – значение априорной вероятности которой соответствует абсолютный максимум среднего риска

$$S_{\max} = (1 - P_1')L(0, \gamma_0) + P_1'L(A, \gamma_0) + (1 - P_1')[L(0, \gamma_1) - L(0, \gamma_0)]\alpha(P_1') - P_1'[L(A, \gamma_0) - L(A, \gamma_1)][1 - \beta(P_1')], \quad (3.21)$$

при этом, в отличие от предыдущей стратегии $S_{\max} - S_1 \geq 0$ из-за отсутствия априорной информации о наличии или отсутствии аномалии в обучающей выборке.

Сценарий на основе информационного критерия качества

Сценарий принятия решений основе информационного критерия качества в качестве наилучшего правила принимает такое, при котором достигается максимум

частного количества информации. Частные количества информации, содержащиеся в наблюдении X относительно передаваемых состояний z_i (аномалия) и z_k (нормальное состояние), определяются соответственно соотношениями

$$I(X | z_i) = \log W(z_i | X) - \log W(z_i), \quad (3.22)$$

$$I(X | z_k) = \log W(z_k | X) - \log W(z_k), \quad (3.23)$$

разность частных количеств информации будет равна

$$I(X, z_i) - I(X, z_k) = \log \frac{W(z_i | X)W(z_k)}{W(z_k | X)W(z_i)}, \quad (3.24)$$

$$I(X, z_i) - I(X, z_k) = \log \frac{W(X | z_i)}{W(X | z_k)}. \quad (3.25)$$

Если разность частных количеств информации положительная то принимается решение о состоянии ПС в состоянии z_i . Следовательно, правило принятия решения о наличии или отсутствии аномалии при информационном критерии качества примет вид

$$\Lambda = \frac{W(X | A)}{W(X | 0)} \underset{H_0}{\overset{H_1}{>}} 1 \begin{matrix} \rightarrow \gamma_1 \\ \rightarrow \gamma_0 \end{matrix}. \quad (3.26)$$

Использование сценария на основе информационного критерия позволяет осуществлять расчеты в реальном времени при относительно небольших объемах выборок в условиях ограниченных вычислительных ресурсах с высокой реактивностью.

Предложенный набор сценариев является открытым для пополнения и позволяет ЛПР более обоснованно осуществлять поддержку принятия решений о наличии аномалий ДРМ ПС в условиях неполной информации с учетом возможных рисков. Разработка данного модуля позволила повысить вовлеченность ЛПР в процесс решения А-задачи ПС, и поднять уровень интеллектуализации системы в

целом. Предложенный список сценариев не является исчерпывающим и может быть расширен в зависимости от конкретизации решаемых задач и особенностей ДРМ ПС.

3.3 Выводы по главе 3

В данной главе получены следующие основные результаты:

1. Предложено алгоритмическое обеспечение работы СППР по обнаружению аномалий в ДРМ ПС, на основе проблемно-ориентированных ИМ, и анализа многомерных скалярных и векторных полей ДРМ, характеризующих состояние ПС. Это достигается в СППР путем анализа дивергенции поступающих данных и типового профиля, а также по метрике расстояния между векторами n -мерного пространства характеризующих вновь поступающие и типовые данные соответственно. В качестве меры отклонений для M -аномалий используется дивергенция Кульбака-Лейблера, обладающая высокой чувствительностью и позволяющая учесть как изменения абсолютных значений характеристик измерительных каналов, так и формы их распределения, и для N -аномалий используется экстремум скалярного поля-результата свертки поля данных с ядром, заданным профилем аномалии являющийся инвариантным аффинным преобразованиям профиля аномалии.

2. Проведены исследования статистических свойств классификатора, направленного на решение A -задачи, с учетом характеристик приборов и средств контроля состояния ПС на основе данных зондирования профиля температуры, полученных с использованием *SVP* зонда в акватории вод г. Севастополя.

3. Предложен специализированный проблемно-ориентированный лабораторный стенд, осуществляющий численное моделирование, генерирование управляемых возмущений в данных, оценку свойств моделей, численный расчет для исследования свойств моделей, оценивание ошибок I и II рода.

4. Оценена эффективность СППР на основе критериев точности, полноты, вероятностей ошибок I и II рода. Полученные оценки показали высокую

сходимость результатов для экспериментальных данных и целенаправленных экспериментов на лабораторном стенде.

5. Предложены принципы применения СППР для автоматизации процесса принятия решения по обнаружении аномалий ДРМ ПС.

ГЛАВА 4.

МОДУЛЬНОЕ ПРОГРАММНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ И ЕГО
ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ ДЛЯ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ
О НАЛИЧИИ АНОМАЛИЙ В ДАННЫХ МОНИТОРИНГА ПРИРОДНОЙ СРЕДЫ
НА ОСНОВЕ МОДУЛЬНО-КОМПОНЕНТНОГО ПОДХОДА

В данной главе детально рассмотрена предложенная в диссертационной работе методика решения А-задачи, определяющая конкретные методы и алгоритмы анализа, рассмотренные в главах 2 и 3, которые легли в основу автоматизированного модульного программно-технического обеспечения СППР [188–193]. Данная система органично дополняет функциональность систем приборов и средств контроля ПС и служит повышению их информационной надежности.

Модульное представление предложенного МПТО служит обеспечению его универсальности, облегчающей интеграцию в уже существующие приборы и средства контроля ПС для обнаружения аномалий ДРМ, в том числе в реальном времени. Обеспечивается легкость масштабирования предлагаемых решений, возможность адаптации к активно изменяющимся внешним условиям за счет структурной и параметрической гибкости.

Задачи, решаемые МПТО, можно разбить на две группы:

- анализ данных поступающих от приборов и средств контроля ПС, обнаружение М-аномалий,
- оценка состояния ПС, обнаружение N-аномалий.

Данные задачи решаются в МПТО за счет:

- приведения ДРМ к общему виду путем регуляризации и устранения пропусков, расчета информационных метрик, позволяющих оценить различные аспекты работоспособности средств измерений и наличия М-аномалий,
- автоматизированного обнаружения N-аномалий в ПС, основанного на модульной СППР, приведенной в 3 главе данной работы.

Таким образом, МПТО является инструментом, позволяющим ЛПР обнаруживать А-события ПС, определять их причины, оценивать корректность ДРМ средств измерений и отслеживать динамику протекания аномальных процессов в ПС во времени. Информация, полученная с помощью МПТО, позволяет ЛПР эффективно осуществлять управленческие решения по обеспечению процессов экологического контроля и рационального природопользования.

МПТО способно обнаружить А-событие либо в момент его возникновения (например, фиксацией факта наличия нетипичного возмущения в скалярном поле физико-химических параметров ПС), либо по его дальнейшему проявлению (например, нетипичной дивергенции сигналов отдельных измерительных каналов или их совокупности в самих приборах и средствах контроля ПС).

В настоящий момент реализован прототип МПТО, проходящий этап опытной эксплуатации в рамках выполняемого стенда автономного биосенсорного комплекса ИПТС [194]. Он реализует функции захвата потока ДРМ ПС, его структуризации, построения типовых суточных профилей данных для каждого отдельного канала и их совокупности в реальном времени, предупреждения о возможном отказе отдельных каналов и исключение их показаний из обучающей выборки, накопления данных и их профилей в БД, и позволяет осуществить визуальный анализ динамики изменения состояния ПС.

СППР для обнаружения М-аномалий реализована в виде аппаратного модуля на базе микроконтроллера, интегрируемого в измерительное устройство, а обнаружение N-аномалий осуществляется специальным программным модулем, исполняемым на сервере и транслирующем результаты анализа на рабочую станцию оператора, которые в настоящий момент интегрируются в МПТО.

Архитектура модульной интеллектуальной СППР обнаружения М-аномалий и N-аномалий включает в себя методы, описанные в главе 2, реализованные в виде модулей согласно методологии, представленной в главе 3, представлена подробно в приложении В.

4.1 Архитектура модульного программно-технического комплекса

Для обеспечения функциональности, описанной в главе 2 и главе 3, и требований к технической реализации предъявленных в разделе 1.3.2, была спроектирована архитектура МПТО, состоящая из следующих компонент:

- модуль обнаружения М-аномалий (М-модуль) [189], пример результатов лабораторного исследования представлен в приложении Д.
- модуль обнаружения N-аномалий (N-модуль) [191], примеры детектируемых аномалий представлены в приложении А.
- модуль выбора сценариев и оценки адекватности СППР [188],
- система визуализации и анализа данных (СВАД) [190].
- агент 1...К – модифицированные приборы или средства контроля ПС,
- оператор 1...У – пользователи или лица, принимающее решение.

На рисунке 4.12 изображена взаимосвязь между этими компонентами.

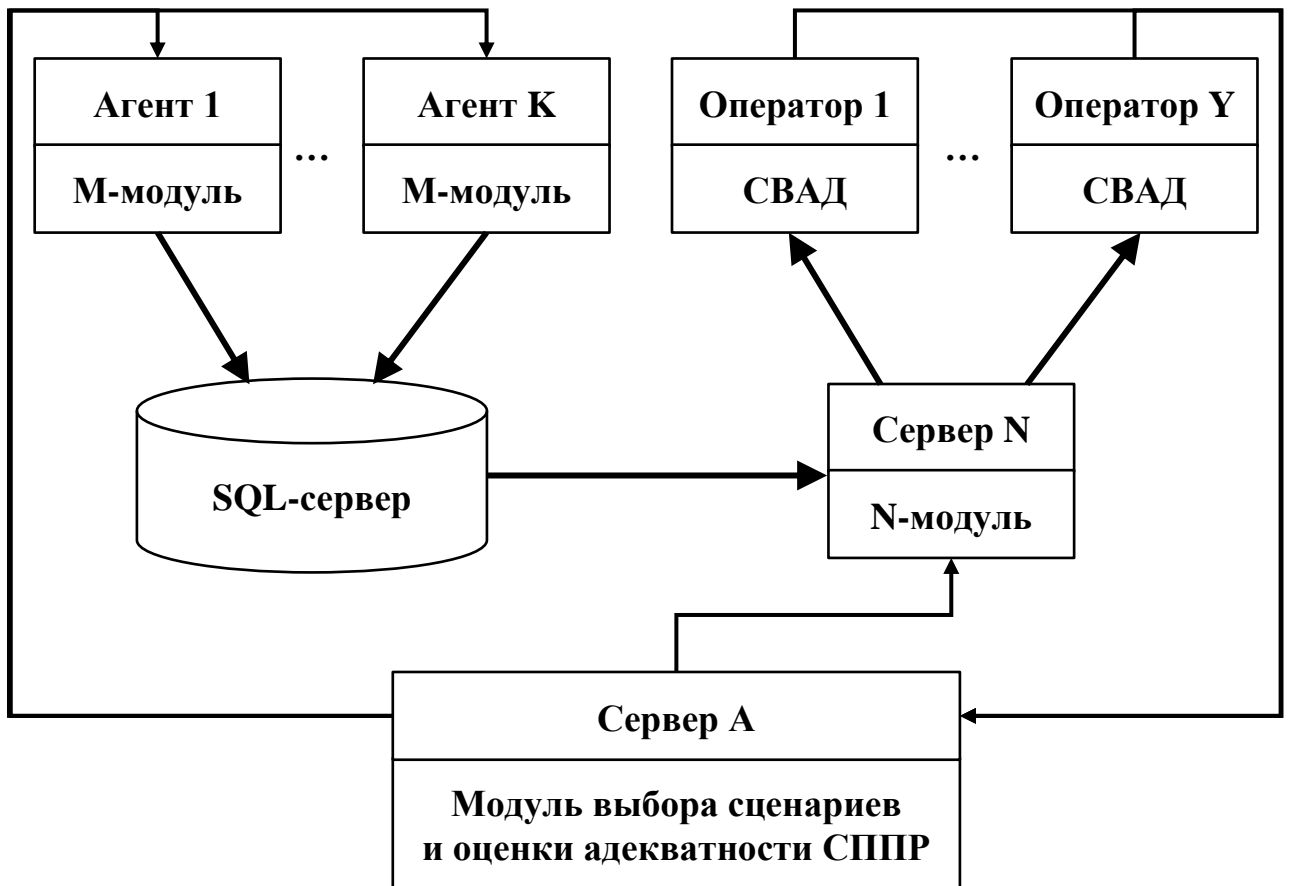


Рисунок 4.1. Общая архитектура МПТО

4.1.1 Аппаратный универсальный 16-канальный модуль лабораторного стенда для регистрации аналоговых сигналов и потоков цифровых данных, и обнаружения в них аномалий, на основе численных метрик, в режиме реального времени

Для проведения экспериментальных исследований осуществлена разработка и изготовление модуля лабораторного стенда на базе платформы Arduino Uno с микроконтроллером ATmega328 включающая универсальный аналого-цифровой 16-канальный комплекс мониторинга состояния измерительных каналов (до 6 аналоговых каналов), представленный на рисунке 4.2.

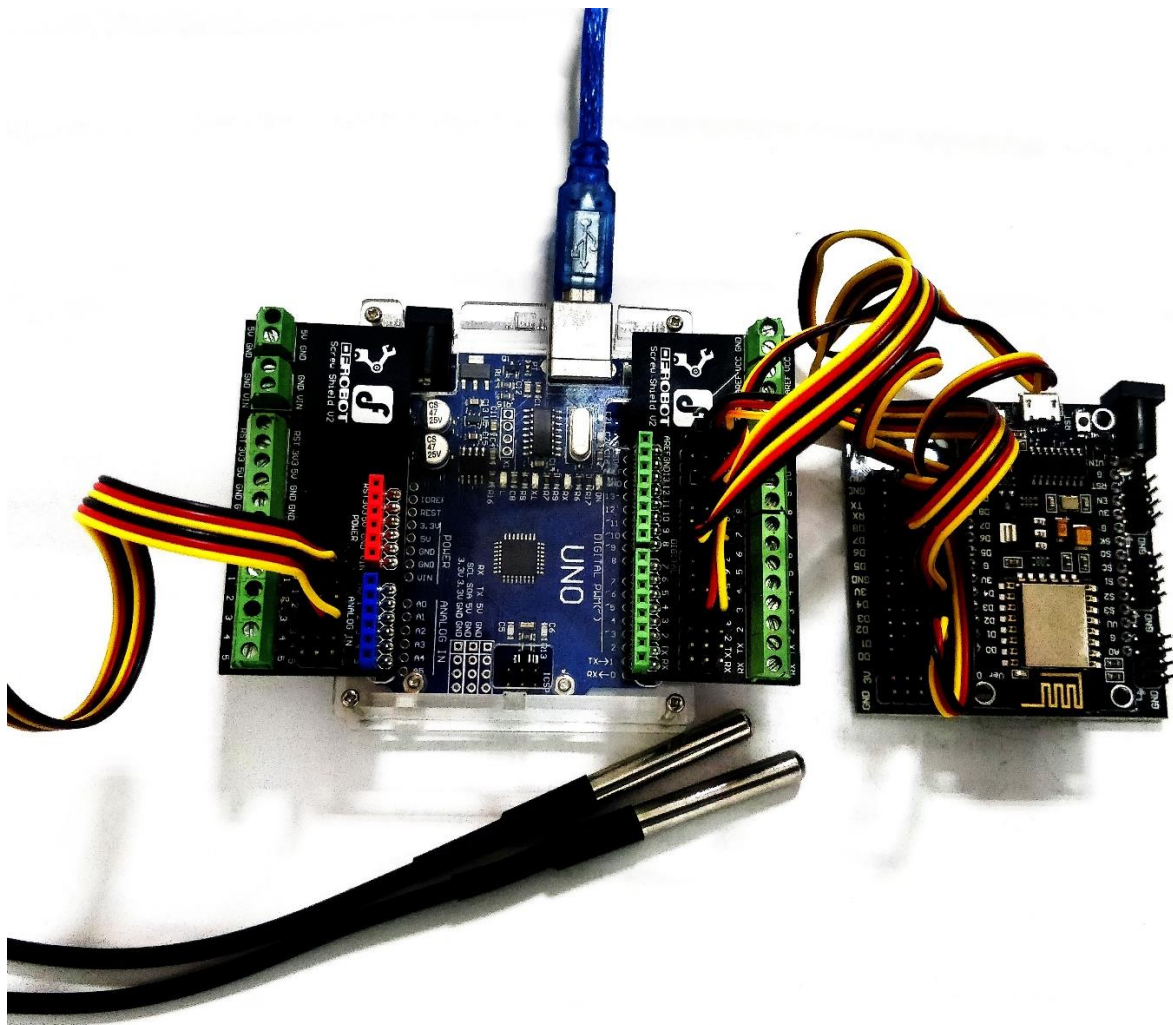


Рисунок 4.2. Общий вид универсального 16-канального аналого-цифрового модуля лабораторного стенда для регистрации, обработки и трансляции данных

На рисунке 4.2 изображены подключенные датчики измерения температуры DS18B20 в герметичном корпусе с интерфейсом 1-Wire, подключенных в качестве контрольного и тестового источников сигнала.

Данный комплекс совместим информационно и логически с автономным биосенсорным комплексом экологического мониторинга водной среды, по данным натурных экспериментов которого осуществлялись практические испытания лабораторном стенде.

Измерительный модуль с подключенным к нему модулем обнаружения М-аномалий вместе представляют собой Агента, генерирующего вместе с выходным сигналом состояния биосенсоров в лабораторной среде численную и качественную оценку адекватности этих данных. Исходная интегральная схема измерительного модуля и модифицированная добавлением модуля обнаружения М-аномалий будут подробнее рассмотрены в соответствующих разделах.

4.1.2 Модификация аппаратного модуля программно-технического комплекса для обнаружения аномалий с возможностью использования дополнительных ресурсов на базе агентного подхода

Агент представляет собой прибор контроля ПС, который в лабораторной установке представлен в виде универсального 16-канального аналого-цифрового регистратора данных, модифицированного за счет последовательного подключения к выходному каскаду его шины передачи данных аппаратного модуля обнаружения М-аномалий реализованного на базе платформы NodeMCU с микроконтроллером ESP8266 и интегрированным WiFi модулем и возможностью трансляции потока данных измерительных каналов в сеть Internet.

Обе используемые платформы оснащены возможностью передавать данные по UART интерфейсу, реализовано аппаратное подключение USB разъемом посредством USB/SERIAL преобразователя на базе микросхемы CH340.

Агент, исходная и модифицированная интегральные схемы которого показаны на рисунке 4.3 работают совместно.

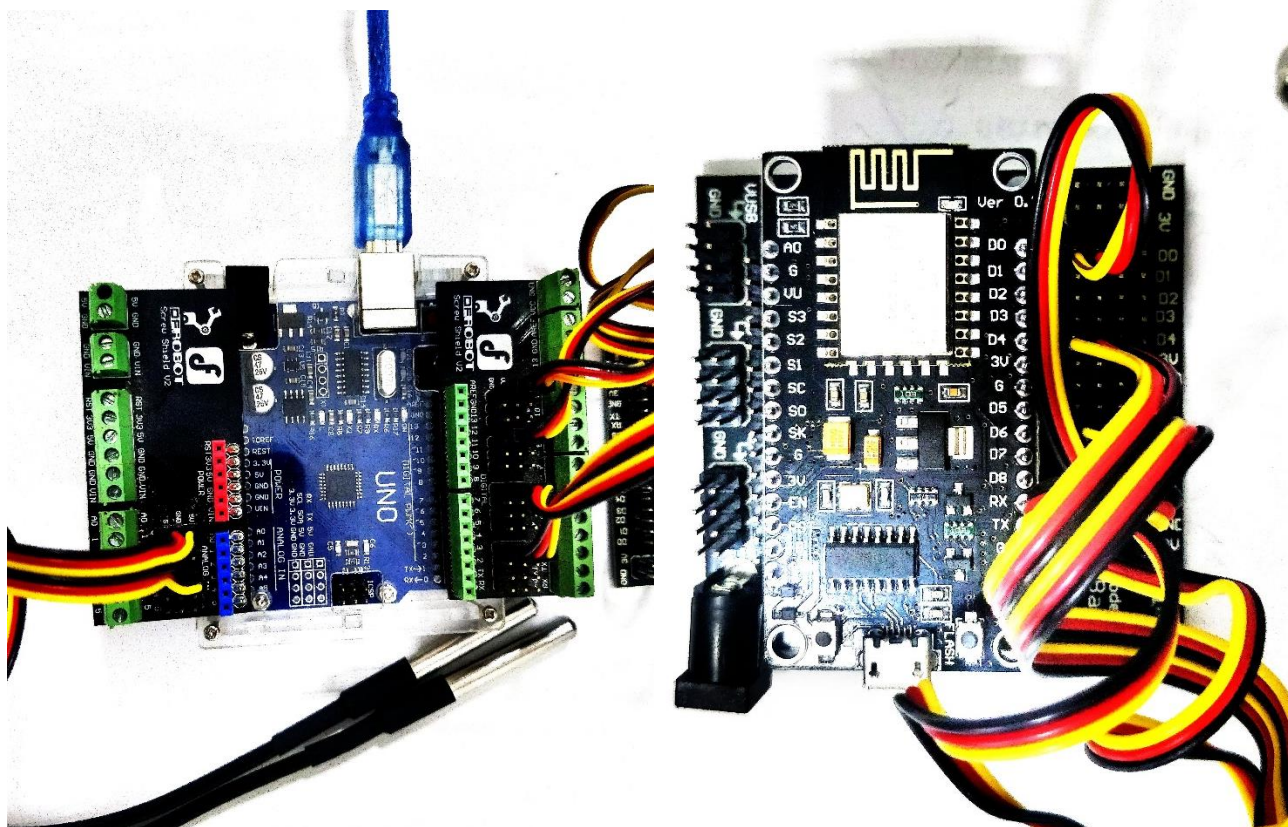


Рисунок 4.3. Общий вид интегральных схем исходного измерительного канала (слева) и доработанного модулем обнаружения М-аномалий (справа)

Принципиальной особенностью модифицированного модуля является техническая реализация агентного подхода, когда, в случае возникновения неопределенности при обнаружении аномалии, применяются более точные, но вычислительно более сложные метрики, рассчитываемые на удаленном сервере. Технически это становится возможным за счет использования, интегрированного в микроконтроллер ESP8266, WiFi модуля, обеспечивающего двустороннюю связь к подключаемым по требованию внешним ресурсом.

4.1.3 Модификация аппаратного модуля автономного обнаружения аномалий при долговременном мониторинге

Для осуществления долговременного мониторинга состояния измерительных каналов с привязкой ко времени реализован дополнительный подключаемый модуль лабораторного стенда, показанный на рисунке 4.4, выполненный на базе платформы

Arduino Leonardo с микроконтроллером ATmega32u4 отличающегося наличием двух последовательных портов, позволяющих как осуществлять отладку при подключении к ПК так и записывать данные измерительных каналов на съемный flash носитель.

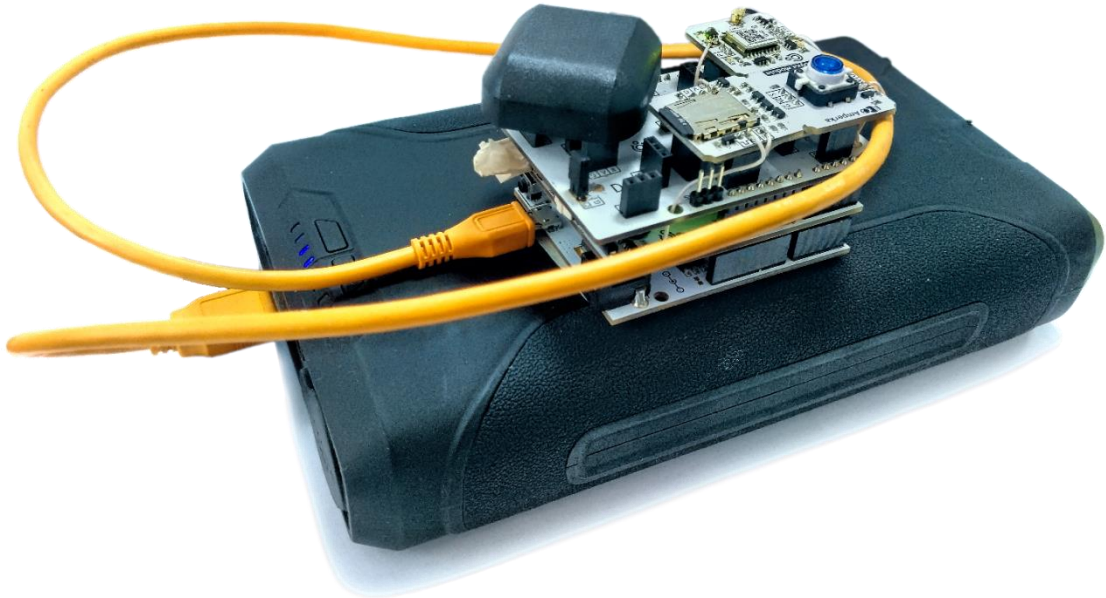


Рисунок 4.4. Модуль автономного долговременного (до полугода) мониторинга данных измерительного канала с привязкой ко времени и записью на съемный flash носитель

Реализованная модификация модуля проблемно-ориентирована для осуществления долговременного развертывания с одновременным анализом регистрируемых данных.

Осуществлена успешная долговременная постановка модуля, в течение 9 месяцев, осуществляющего анализ данных, получаемых от GPS+ГЛОНАСС датчика.

Агент выполняет следующие шаги алгоритма:

- захват и трансляция всех поступающих на входной тракт потоков данных и их кэширование во внутренней ОЗУ,
- расчет информационной метрики (дивергенции) по каждому из каналов в реальном времени,

- сравнение рассчитанных метрик с записанными ранее эталонами, расчет доверительных интервалов,
- маркировка кэшированных данных результатами расчета количественной метрики аномальности и качественного критерия,
- трансляция маркированных данных.

4.1.4 Программные модули автономного обнаружения аномалий в данных мониторинга

Система визуализации и анализа данных

СВАД реализует отдельные функции СППР обнаружения А-событий, рассмотренные в главе 3, и предоставляет пользовательский интерфейс, внешний вид которого, отображаемый у оператора, представлен на рисунке 4.5.

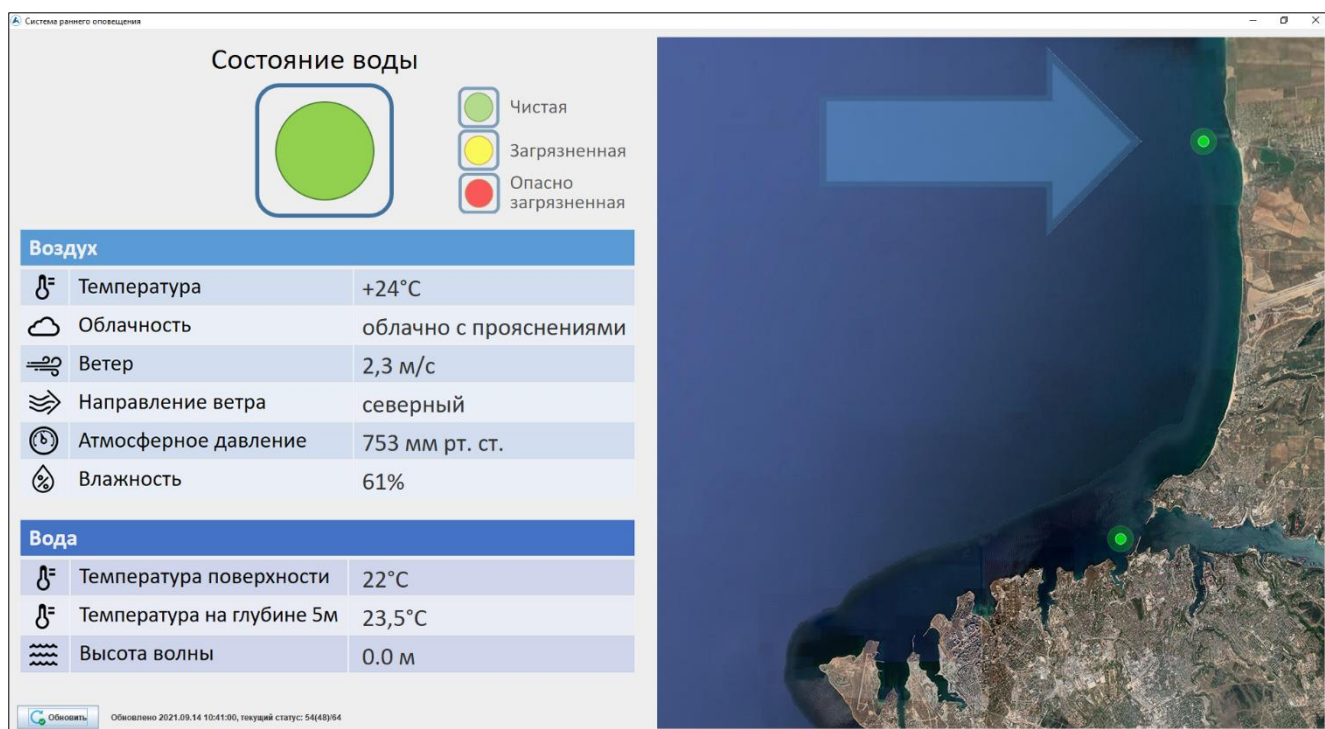


Рисунок 4.5. Внешний вид пользовательского интерфейса оператора системы визуализации и анализа данных

СВАД позволяет оператору управлять процессом анализа и получать доступ к ДРМ за произвольно выбранный пользователем интервал времени для осуществления его подробного анализа, как показано на рисунке 4.6.

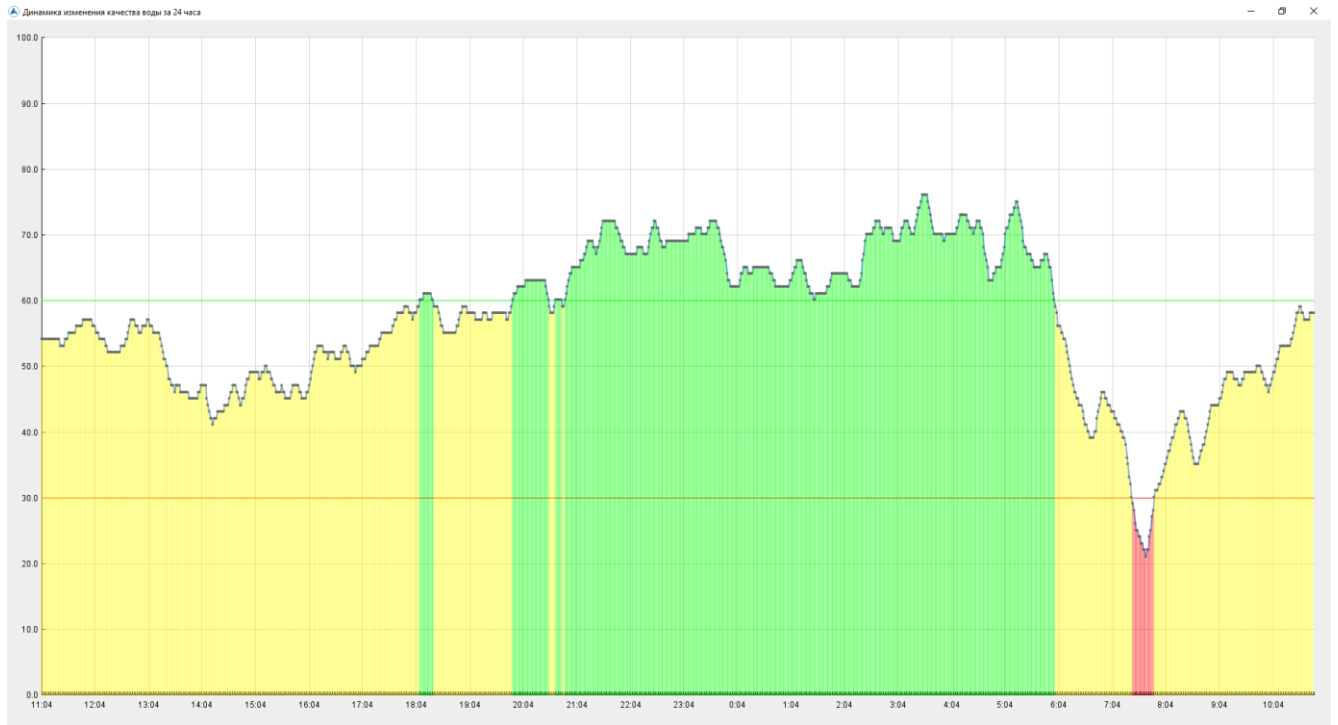


Рисунок 4.6. Внешний вид отображения временной динамики интегральных показателей пользовательского интерфейса оператора системы визуализации и анализа данных

Повышение процесса интеллектуальности СППР осуществляется за счет привлечения ЛПР в процесс принятия решения за счет визуализации в реальном времени как исходных ДРМ, так и интегральных характеристик, а также расширенной системы фильтров, позволяющих визуализировать данные исключив М-аномалии.

Модуль выбора сценариев и оценки адекватности СППР

Данный модуль предназначен для исполнения на сервере, решает задачу оценки адекватности выбранной структуры СППР характеру поступающих данных, учитывает выбранные пользователями сценарии оценки рисков, подбирает контрольно-предупредительные границы классификаторов N-аномалий так, чтобы обеспечить оптимальные значения вероятностей ошибок I и II рода.

Вследствие специфики решаемых задач, программа имеет упрощенный графический интерфейс отображения типовых профилей измерительных каналов, представленный на рисунке 4.7.



Рисунок 4.7. Внешний вид отображения типовых временных профилей показателей измерительных каналов модуля выбора сценариев и оценки адекватности СППР

Программа является утилитарный модулем, хранит в своей локальной БД профили типовых сигналов (суточные, сезонные и др.) и профили типовых аномалий (распределенные во времени и пространстве), непрерывно актуализирует имеющиеся профили согласно изменению структуры поступающих данных или по распоряжению уполномоченных операторов.

Программный модуль обнаружения N-аномалий

Особенностью модуля является его универсальность к характеру входных данных, которые могут быть представлены как виде точечных измерений, содержащих пропуски и определенные на нерегулярной сетке, так и в виде многослойных регулярных полей. Такая универсальность достигается за счет применения в его основе полимодельного динамически конфигурируемого комплекса, предусматривающего различные постановки А-задачи.

Для исключения необходимости дублирования производимых вычислений, расчеты выполняются на отдельном сервере, а результаты расчетов передаются пользователям в готовом виде как web-сервис. Таким образом значительно снижается уровень системных требований к пользовательским устройствам, позволяя, в том числе, осуществлять эксплуатацию системы с мобильных портативных устройств. Графический интерфейс пользователя сервиса представлен на рисунке 4.8.

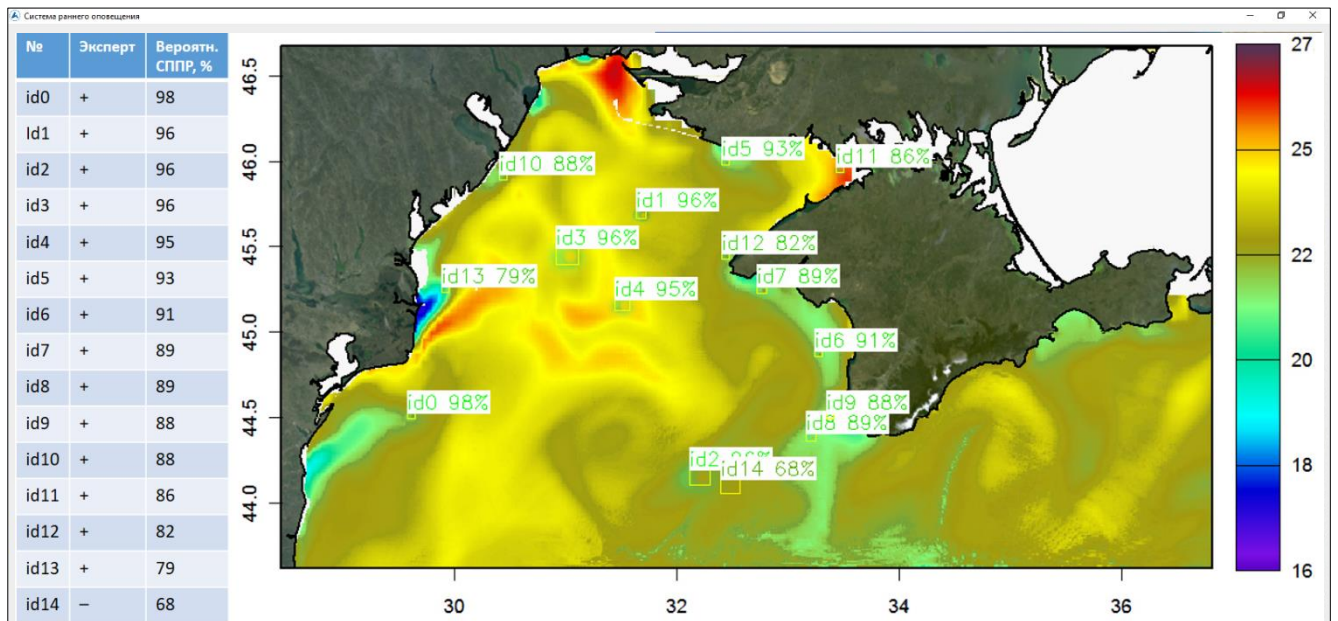


Рисунок 4.8. Внешний вид пользовательского интерфейса модуль обнаружения N-аномалий в данных пространственного мониторинга

Интеллектуальность модуля заключается в возможности осуществлять обучение его пользователем указывая ошибочно распознанные области как аномалии или наоборот, пропуск аномальных областей. Для обеспечения указанных функциональных возможностей модуля пользователь, имеющий соответствующие полномочия, посредством обращения к модулю выбора сценариев и оценки адекватности СППР, осуществляет коррекцию профиля.

Графический интерфейс пользователя сервиса, при указании областей, ошибочно распознанных как аномальные или пропущенных аномальных областей, представлен на рисунке 4.9.

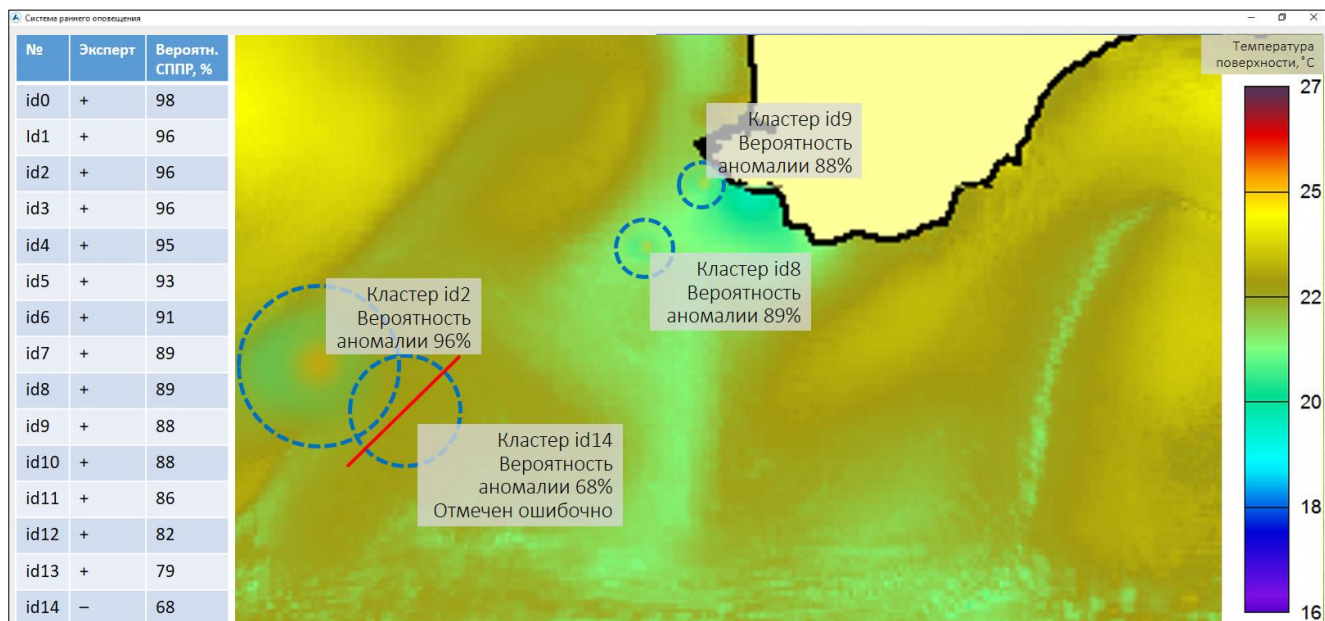


Рисунок 4.9. Внешний вид пользовательского интерфейса обучения модуля обнаружения N-аномалий в данных пространственного мониторинга

За счет реализации такого механизма, система способна обучаться в процессе эксплуатации, тем самым оперативно реагируя на изменения в ПС, подстраиваясь под них и осуществляя поддержку принятия решения требуемого качества.

4.1.5 Реализация принципов архитектуры комплекса в соответствии с сформированными требованиями к методике обнаружения аномалий

Требование модульности достигнуто на структурном уровне за счет унификации интерфейсов информационного обмена. Агенты передают ДРМ ПС в БД вместе с информацией о качественных характеристиках этих данных для конкретного измерительного устройства. Этот подход позволяет децентрализованно решать А-задачу для больших географически распределенных сетей мониторинга ПС поддерживать возможность подключения большого числа пользователей и реализовать расширение функционала системы в виде атомарных web-сервисов, не затаривающих основное распределенное ядро системы.

Требование минимизации времени анализа достигнуто за счет реализации предварительного этапа обнаружения М-аномалий нетребовательными к объему

данных информационными метриками и использованием проблемно-ориентированного интегрируемого в средство измерения модуля на базе микроконтроллера. Ускорение принятия решения по поводу возникновения N-аномалий осуществляется за счет последовательно усложнения используемого алгоритма по мере усложнения характера поступающих ДРМ ПС, а также за счет распределения вычислений на различных серверах, разделенных по функциональному признаку.

Достижение требования минимизации объемов хранимых ДРМ ПС с минимальными потерями информативности достигнуто с использованием специализированного программного модуля, позволяющего осуществлять фильтрацию слабоструктурированных ДРМ по критерию их информационного вклада в описание состояния ПС. Таким образом, даже для достаточно «агрессивных» сценариев в плане прореживания ДРМ, будет достигаться максимальная информативность по информационному критерию.

Требование структурной гибкости и расширяемости МПТО достигается использованием при ее разработке и создании объектно-ориентированных принципов проектирования и программирования, позволивших реализовать модульную структуру для отдельных компонентов, что позволит в дальнейшем подключать к МПТО дополнительные функциональные модули. Совместное использование и программных и аппаратных решений позволяет на разных уровнях осуществлять расширение доступного пула моделей и методов обнаружения аномалий в случае возникновения значимых изменений в ПС или расширения количества и разнообразия источников данных в дальнейшем.

4.2 Модули визуализации и анализа данных

Графический интерфейс системы визуализации и анализа, показанный на рисунке 4.10 служит для повышения обоснованности принятия решения о состоянии ПС по ДРМ за счет вовлечения ЛПР в процесс принятия решения.



Рисунок 4.10. Визуальное представление на графическом интерфейсе интегрального показателя состояния ПС и отдельных его компонентов распределенные во времени

Графически выделяется три зоны аномальности: зеленая, желтая и красная. Соответствующие нормальным данным, зоне неопределенности и наличием аномалий в данных, соответственно. Одновременно с этим, у пользователя имеется возможность получения поступающих в систему оцифрованных сигналов и наборов данных в графической форме. Представленный пример отображает параметры регистрируемые автономным биосенсорным комплексом экологического мониторинга.

Более подробно список регистрируемых и визуализируемых параметров представлен в приложении Б.

4.3 Программный модуль выбора сценариев решения задач обнаружения аномалий и оценки адекватности СППР

Модуль выбора сценариев и оценки адекватности СППР осуществляет подбор оптимального набора методов из доступного пула, который наилучшим образом решает А-задачу с учетом выбранного ЛПП сценария допустимых рисков.

Структура модуля вместе с наборами методов и решающих правил изображена на рисунке 4.11.

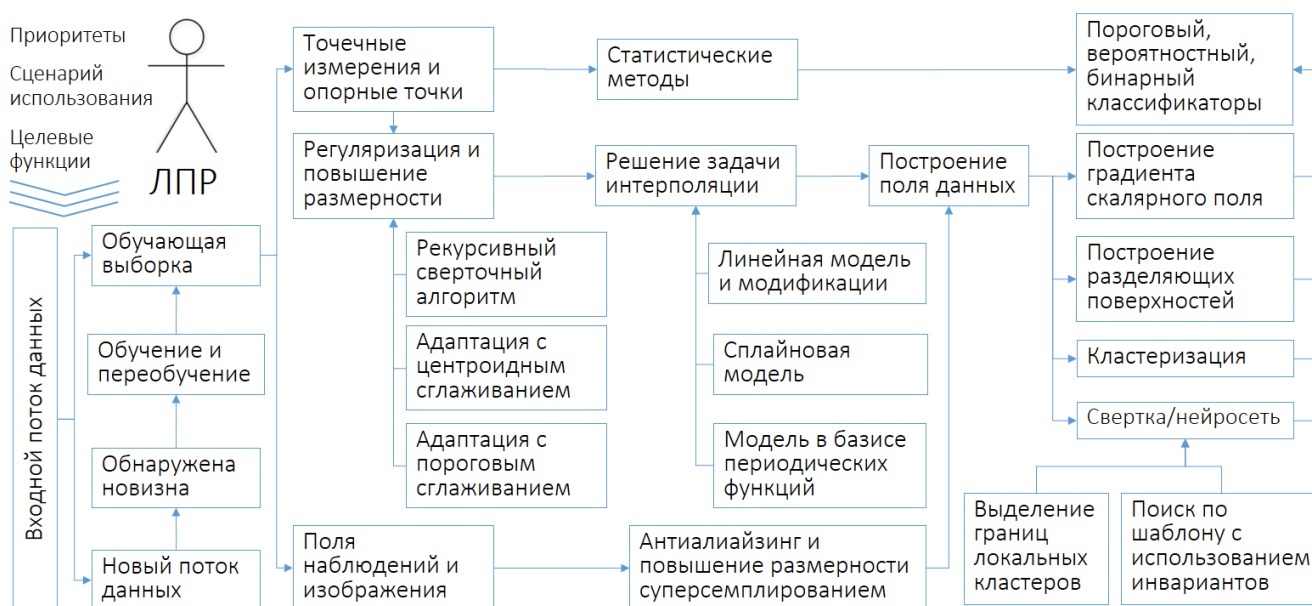


Рисунок 4.11. Структура интеллектуального модуля выбора сценариев и оценки адекватности СППР

Модуль реализован в виде серверного приложения на базе языка Java для исполнения в качестве web-сервиса, позволяющего осуществлять доступ к нему в том числе с низкопроизводительных и мобильных платформ, осуществляя вычисления по запросу на центральном сервере.

Подробное описание реализованных методов, областей, где их использование оптимально и описание решающих правил представлены соответственно в главе 2.

4.4 Модельные экспериментальные исследования процессов распространения аномалий в пространственном и временном измерениях, и их обнаружения

Произведя моделирование одного деградиационного процесса, протекающего в среде E , на лабораторном стенде, с параметрами $N=1$, $C=\{0,0\}$, $T=1000$, $R=10$, $l_{\max}=1$ получим результат показанный на рисунке 4.12. Агенты обозначены черными маркерами.

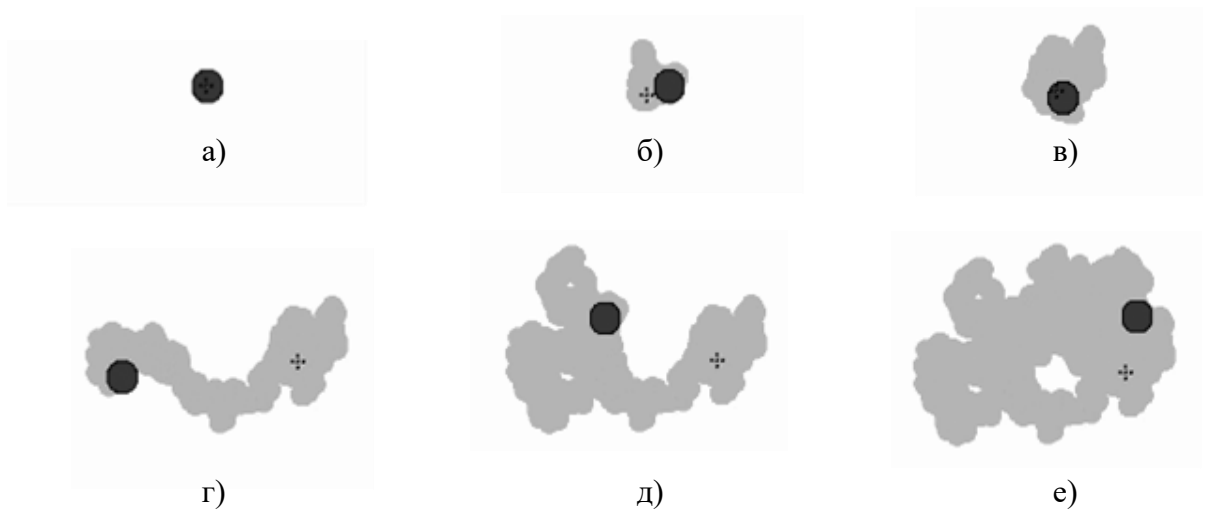


Рисунок 4.12. Визуализация результата моделирования распространения АП единичного агента: а) $t=0$, б) $t=20$, в) $t=50$, г) $t=200$, д) $t=400$, е) $t=1000$

Представленный вариант использования модели может иметь физическую интерпретацию в виде распространения загрязнителя, генерируемого некоторым подвижным агентом в стационарной среде. Моделирование с такими начальными условиями весьма специфично, потому что на практике распространение происходит по нескольким направлениям одновременно. Если осуществить моделирование распространения нескольких агентов в модели $N=5$, при этом остальные параметры, в частности начальная точка моделирования для каждого из

агентов останутся неизменными $c_i = \{0,0\} \forall i \in [0; N]$, реализация результатов моделирования примет вид представленный на рисунке 4.13.

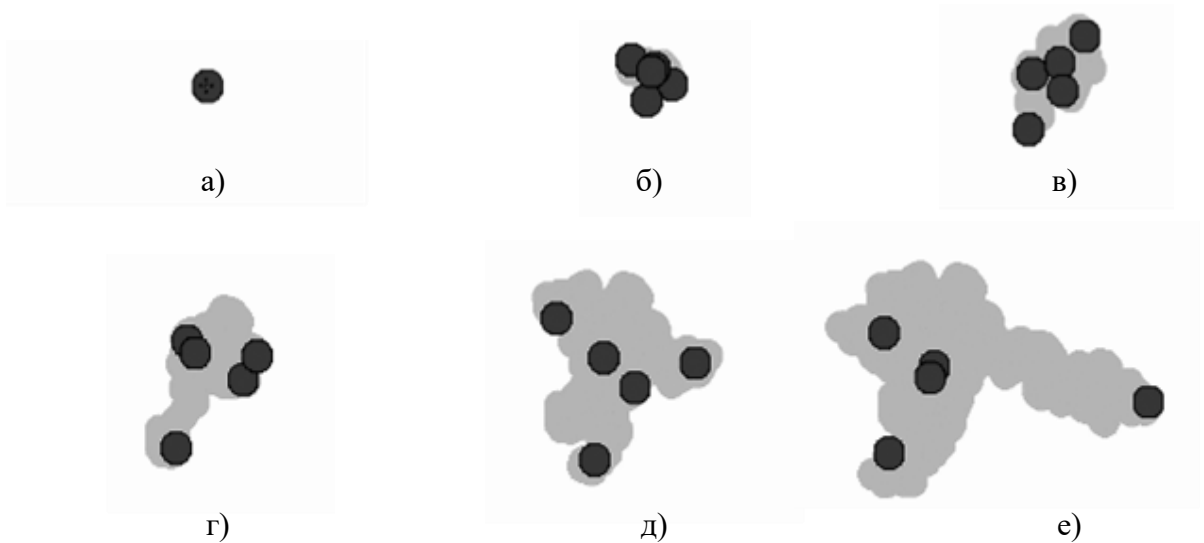


Рисунок 4.13. Визуализация результата моделирования распространения АП нескольких агентов: а) $t=0$, б) $t=20$, в) $t=50$, г) $t=200$, д) $t=400$, е) $t=1000$

Данный вариант использования моделирующего стенда может иметь физическую интерпретацию модели распространения коррозии в объекте из единой начальной точки в результате дефекта этого объекта.

Второй вариант представления результата моделирования основан на расширенной модели, где факт воздействия агента на среду имеет вероятностную природу, в таком случае при повторном прохождении агента в некоторой области пространства E , которую обозначим $P(A, C, E)$, вероятность возникновения АП – $P(АП)$ возрастает. Если предположить, что события появления агента в области $P(A, C, E)$ независимы, то вероятность возникновения АП в области C примет вид $P(АП) = 1 - (1 - P(A, C, E))^k$, где k – число появлений агента в области C . Обозначив $P(АП) \in [0; 1]$ как параметр интенсивности цвета, для параметров $P(A, C, E) = 5\%$, $N = 1$, $C = \{0, 0\}$, $T = 1000$, $R = 10$, $l_{\max} = 1$ получим визуализацию, представленную на рисунке 4.14.

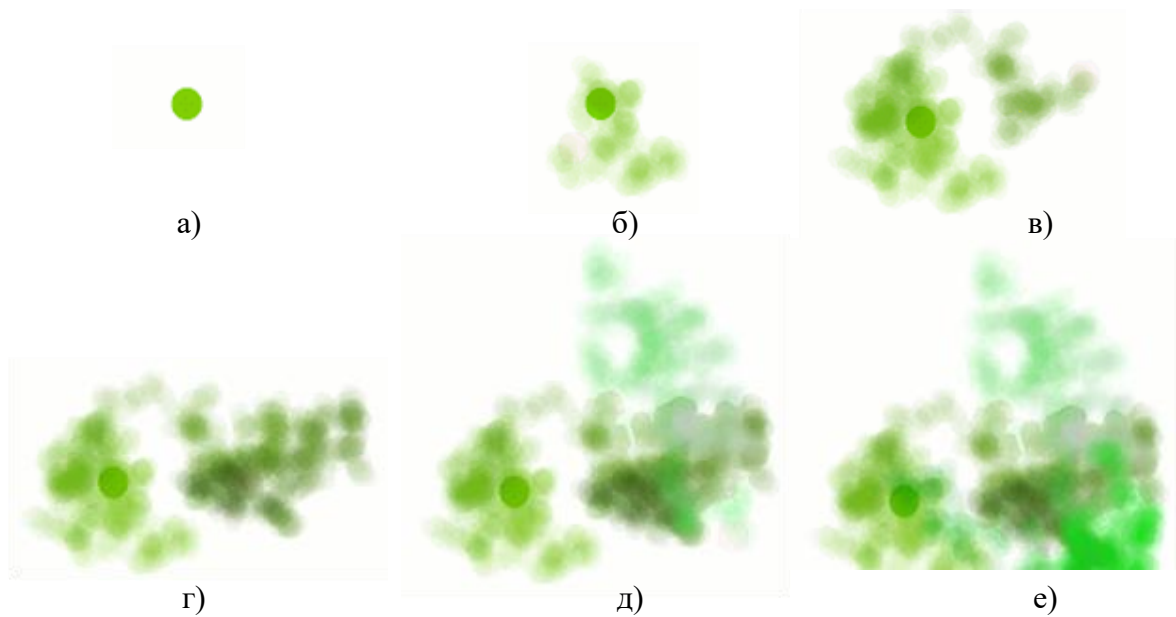


Рисунок 4.14. Визуализация результата моделирования распространения вероятностного АП единичного агента: а) $t=0$, б) $t=20$, в) $t=50$, г) $t=200$, д) $t=400$, е) $t=1000$

С использованием моделирующего имитационного стенда генерирования управляемых возмущений на фоне типового сигнала осуществляется мультиагентное имитационное моделирование распространения аномалии при деградиционных воздействиях. На базе разработанного лабораторного стенда была произведена серия целенаправленных экспериментов, направленных на разработку аппарата проверки качества методов обнаружения аномалий ПС различных видов. В представленных примерах рассматривались решения задач в евклидовом пространстве, но область применения предложенной системы расширена путем дополнения ее необходимым числом измерений создавая факторное пространство требуемой размерности.

4.5 Экспериментальные испытания модульной системы обнаружения аномалий, проведенные в лабораторной установке

Эффективность модульно-компонентной системы обнаружения аномалий оценена на основе методики описанной в разделе 2.5 для следующих критериев:

– точность обнаружения А-событий. Этот показатель определяет количественную оценку качества вывода, сделанного СППР в отношении состояний ПС. Он характеризуется отношением количества состояний, соответствующих правильно обнаруженным А-событиям, и количества всех состояний, охарактеризованных СППР как А-событий.

– полнота обнаружения А-событий. Полнота обнаружения определяется как количественное отношение правильно обнаруженных А-событий и общего количества всех А-событий, действительно имевших место.

– величина ошибки 1-го рода – вероятность ложного срабатывания.

– величина ошибки 2-го рода – вероятность пропуска аномалии.

Все теоретически возможное множество событий, характеризующее разные состояния ПС, разбито на два подмножества: включающие в себя нормальную и аномальную работу ПС. Результат работы СППР за некоторый промежуток времени также представляется в виде двух множеств, содержащих нормальные и аномальные (с точки зрения СППР) события, которые, в свою очередь (в силу несовершенства алгоритмов обнаружения), могут частично войти как в множество нормальных, так и в множество аномальных событий.

В итоге, результаты работы СППР могут быть представлены в виде четырех множеств:

- правильно зафиксированная нормальная работа;
- пропущенные аномалии;
- ложные срабатывания;
- правильно зафиксированные аномалии.

Предложенные выше критерии оценки эффективности могут быть описаны следующими критериями:

- P – точность обнаружения аномалий,
- R – полнота обнаружения аномалий,
- α – вероятность ошибки 1-го рода,
- β – вероятность ошибки 2-го рода.

В качестве исходных данных для натурального эксперимента использованы данные результатов мониторинга состояния водной среды автономным биосенсорным комплексом экологического мониторинга, установленным в акватории в районе г. Севастополя. Графическое представление исходных данных приведено в приложении Б.

Следует понимать, что границы принятия решения о возникновении А-события по отдельным параметрам ПС и их совокупности зависит от конкретизации целевой задачи и их нельзя обобщать. Вместо этого следует для каждой конкретной области использования осуществлять синтез по приведенной в работе методике.

Значения критериев аномальности, рассчитываемые в СППР, служат для ЛПР ориентиром. Следует понимать, что они носят рекомендательный характер. Именно поэтому для упрощения восприятия рассчитываемых метрик в программе оператора предложена графическая интерпретация областей аномальности:

- зеленая – чистая,
- желтая – загрязненная,
- красная – опасно загрязненная.

Отнесение последовательности данных к одной из этих областей позволяет ЛПР сделать предварительный вывод о состоянии ПС и в случае необходимости визуально оценить динамику изменения отдельных параметров ПС или интегральных показателей. Так степень аномальность данных есть отклонение метрик аномальности текущих данных от рассчитанных для нормального профиля.

Таблица 4.1. Оценка эффективности обнаружения А-события

Критерий оценки	Зоны аномальности		
	Зеленая (>60)	Желтая (20–60)	Красная (<20)
P	<0,62	0,76	$\geq 0,83$
R	>0,92	0,90	$\leq 0,81$
α	>0,12	0,08	$\leq 0,06$
β	<0,10	0,13	$\geq 0,18$

Полученные значения оценок критериев эффективности демонстрируют приемлемое качество обнаружения А-событий. Они достаточно хорошо согласуются с данными, полученными в лабораторных условиях при наличии искусственно созданной аномалии.

Под полнотой обнаружения А-событий также понимается как способность системы обнаруживать существующие (известные) типы аномалий, так и возможность обнаруживать новые. В таком контексте, полнота обнаружения зависит от чувствительности интегральных показателей ПС к А-событиям. В п. 2.9 было показано, что все их множество покрывает известные типы А-событий. Это указывает на то, что известные типы А-событий могут быть зафиксированы на данном наборе интегральных показателей. Кроме того, на рассматриваемых наборах характеристик будут зафиксированы и новые типы А-событий, если их проявление будет отражаться на характеристиках аналогично известным типам. Это обусловлено тем, что используемая в диссертации метрика, определяющая степень аномальности ПС, зависит не только от абсолютных значений интегральных показателей и зависимостей между ними, но и учитывает конкретный тип обнаруживаемых аномалий.

4.6 Повышение уровня интеллектуальности модуля СППР и анализ его эффективности на основе векторных критериев

Принимая во внимание результаты, полученные в разделах 3.1 и 3.2 , рассмотрим принципы автоматизации СППР в программном комплексе. Автоматизация СППР предполагает конкретизацию трех ключевых функций методики решения А-задачи:

- расчет интегральных показателей ДРМ ПС,
- модификация структуры проблемно-ориентированной СППР для ее актуализации в соответствии с характером входных данных и выбранным ЛПР сценарием использования,

– оценка степени аномальности данных по критериям М-аномалий и N-аномалий, визуализация результатов расчета в динамике.

Каждая из функций представляет собой самостоятельный непрерывный вычислительный процесс, управляемый ЛПР. Модификация структуры СППР предполагает оценку объема входных данных, степени их изменчивости, структуры, регулярности и наличия пространственной или временной привязки. Результаты анализа загружаются в тестовую среду лабораторного стенда в которой осуществляется численное моделирование аномалий различных типов и расчет интегральных метрик качества доступного пула методов. Совокупность методов, получившая максимальное значение комплексного критерия, объединяется в последовательность этапов, выполняемых в СППР. Оценка степени аномальности ДРМ ПС заключается в определении отклонения метрик различия шаблонов аномального и нормального состояний ПС и данных оперативного мониторинга. Результаты обнаружения А-событий представляются графически в виде маркеров на графическом пространственном поле ДРМ с указанием степени сходства отмеченной области и аномального паттерна.

Результаты проведенного анализа показали факт повышения информационной надежности приборов и средств контроля ПС за счет обнаружения М-аномалий свойственных самим приборам и средствам мониторинга ПС и новых типов N-аномалий при использовании СППР в составе программно-технического обеспечения. Очевидно, что предложенная в СППР методика не дает абсолютной гарантии обнаружения аномалий, соответствующие ошибки были оценены в разделе 3.2.2 . Поэтому проанализируем теоретически возможные ограничения, присущие предложенной СППР по решению А-задачи.

Так, если обобщить принципы работы СППР, то критерием возникновения А-события является факт увеличения информационной меры различия нормального паттерна и вновь поступивших данных, а также увеличения экстремума скалярного поля-результата свертки поля данных с ядром заданным профилем аномалии.

Таким образом, теоретически, А-события, мало влияющие на мониторируемые параметры, или профили которых не могут быть сведены с помощью аффинных преобразований к уже имеющимся аномальным профилям, могут быть не обнаружены СППР. С другой стороны, возможны ошибки, связанные с ошибочным принятием нормальных ДРМ за аномальные, в то время как они являются нормальными. Это связано с появлением возмущений в полях мониторинговых наблюдений по причинам, не связанным с нарушениями в самих приборах и средствах мониторинга или процессами, происходящими в ПС. Причины этого могут быть самыми различными начиная от циклических сезонных изменений, естественных для ПС и, заканчивая антропогенной деятельностью не несущей негативного влияния на ПС.

Повышение интеллектуализации СППР осуществляется за счет вовлечения ЛПР в принятие решения на основе средств оперативной визуализации результатов анализа ДРМ ПС и предоставления возможности переключать сценарии использования СППР с различными весовыми коэффициентами обобщенного критерия эффективности. Далее сформулируем обобщенные правила работы ЛПР в программной системе СППР, функционирующей в СМ ПС:

- визуальный анализ диаграмм, графически отображающих области данных пространственного мониторинга, которые подозреваются на наличие аномалий, и фиксация факта устойчивого состояния аномального процесса в ПС,

- анализ архивных ДРМ ПС с целью поиска предпосылок, которые могли привести к возникновению такого процесса или возможных прецедентов в прошлом,

- определение типа возникшей аномалии методом обнаружения М-аномалий, и в случае отрицательного результата методами обнаружения N-аномалий, а также сопоставление результатов с данными из других источников,

- в случае обнаружения устойчивой М-аномалии, вызванной отказами в измерительном канале прибора или средства мониторинга принять решение о продуктивности ремонта или замены такого прибора или продолжить его эксплуатацию,

– в случае обнаружения устойчивой N-аномалии принять меры по уведомлению уполномоченных лиц, либо овестествить дополнительную проверку фактического состояния ПС более точными средствами, либо осуществить комплекс мер по ликвидации или снижению возможных негативных последствий от предстоящей критической ситуации.

4.7 Выводы по главе 4

В ходе разработки МПТО были решены следующие задачи:

1. Предложены принципы реального воплощения методики обнаружения М-аномалий и N-аномалий ДРМ ПС, реализующей алгоритмы обнаружения А-событий в соответствии с введенными требованиями, в виде распределенной программно-технической системы МПТО, позволяющей обнаруживать аномалии данных сетей мониторинга ПС произвольного масштаба.

2. Разработана архитектура МПТО, представляющая собой распределенную модульную программно-техническую систему, включающую модуль обнаружения М-аномалий (М-модуль), модуль обнаружения N-аномалий (N-модуль), модуль выбора сценариев и оценки адекватности СППР и модуль визуализации и анализа данных (СВАД). Важными особенностями предложенной архитектуры являются: 1) разделение решаемых задач между компонентами по функциональному признаку, позволяющему осуществлять гибкую интеграцию в существующие системы контроля ПС и 2) возможность децентрализованного масштабирования системы при контроле множества территориально удаленных сетей мониторинга ПС, что позволит эффективно применять МПТО в сетях мониторинга ПС как регионального, так и государственного масштабов.

3. Произведено испытание отдельных компонент с использованием проблемно-ориентированного лабораторного стенда, позволяющего создавать управляемые возмущения в лабораторной среде. Это позволило на этапе разработки провести анализ эффективности архитектурных решений МПТО, выявить наиболее уязвимые в плане скорости работы решения. Так были

предложены оптимизированные реализации алгоритма качественной оценки данных на предмет М-аномалий на языке Си, выполняемые на базе микроконтроллера.

4. Реализованы в виде программных и аппаратных модулей: модуль обнаружения М-аномалий, модуль обнаружения N-аномалий, модуль выбора сценариев и оценки адекватности СППР и модуль визуализации и анализа данных. Модуль обнаружения М-аномалий реализован на языке Си для выполнения на базе микроконтроллера, что обеспечит высокую реактивность системы при работе в реальном времени и возможность интегрировать модуль отдельно от всей СППР. Основные функциональные модули N-модуль и СВАД реализованы на языке Java, что обеспечивает их легкую переносимость и возможность использования на различных программных платформах. Это важно при интеграции МПТО в системы с большим числом пользователей, использующих различные программные платформы. При создании модуля выбора сценариев и оценки адекватности СППР, и части графического интерфейса пользователя были использованы такие web-технологии как Js, Xml, Ajax, Php, Python, что позволило не только значительно снизить требования к вычислительным устройствам пользователей реализуя вычисления по требованию на сервере, но и реализовать возможность работы с системой как с web-сервисом с использованием мобильного устройства в любом удобном пользователю месте.

5. Реализованы модули автоматического обнаружения М-аномалий и автоматизированного обнаружения N-аномалий, реализующие СППР и выполняется их интеграция в МПТО. Эта работа ведется в настоящий момент и будет продолжена в дальнейшем.

6. МПТО успешно внедрена в опытную эксплуатацию в Лаборатории гидроакустических и оптических методов и средств контроля состояния окружающей среды Центра экологического приборостроения и экоэнергетики Института природно-технических систем.

ЗАКЛЮЧЕНИЕ

В результате проведенных исследований решен ряд важных задач, составляющих проблему повышения информационной надежности систем мониторинга природной среды (ПС), на основе модульно-компонентного подхода, интеллектуализации процессов принятия решений о наличии скрытых аномалий контролируемых характеристик в факторном пространстве скалярных и векторных полей данных мониторинга (ДРМ), за счет перехода к адаптивным полимодельным комплексам, использования нового проблемно-ориентированного программно-технического обеспечения, классификации видов аномалий ПС и проактивного вовлечения ЛПР в процесс принятия решений. Основные результаты работы заключаются в следующем:

1. В современных системах мониторинговых наблюдений природной среды аномалии, в большинстве случаев, обладают сложной структурой, что в значительной степени затрудняет их обнаружение традиционными методами математической статистики. В современных представлениях аномалии в мониторинговых наблюдениях – это аномалии полей данных, аномалии градиентов, аномалии при пропущенных и нерегулярных данных, где использование традиционных методов – неэффективно. Развитие этих направлений в рамках модульно-компонентного подхода, включающее классификацию информационных ситуаций и анализ потенциально существующих видов аномалий в данных мониторинга природной среды, и поиск наиболее оптимальных проблемно-ориентированных методов их обнаружения, с целью реализации в виде функционально-замкнутого набора модулей, осуществлено в диссертационной работе. Применение предложенного подхода динамически переключаемых модулей позволило на 10-25% уменьшить вероятности ошибок I и на 7-15% уменьшить вероятности ошибок II рода при принятии статистических гипотез о наличии и отсутствии аномалий в ДРМ по сравнению со стационарными решениями.

2. Специфика современных систем мониторинга природной среды требует осуществление обнаружения аномалий по множеству критериев: точности, ошибок

I и II рода, реактивности. Ранее, при решении простых задач, осуществлялась оценка по одному скалярному критерию. На данный момент, для осуществления достоверной оценки среды, этого недостаточно, требуется производить анализ комплекса критериев (зачастую противоречивых), на основе Парето подхода. Разработанные новые модульные решения на основе методического, алгоритмического и приборного обеспечения, реализованные в соответствии с модульно-компонентным подходом, направлены на отображение множества возможных решений по обнаружению аномалий в ДРМ на множество оптимальных решений значительно меньшей мощности. Разработаны программно-аппаратные решения для получения дифференциальной оценки состояния природной среды. За счет применения, в системах экологического мониторинга, разработанных решений консолидации достигается уменьшение Парето множества решений о состоянии среды в 4-6 раз, по сравнению классическим набором однокритериальных подходов.

3. Непрерывное повышение требований к скорости принятия решений, реактивности систем мониторинга и обоснованности решений на основе анализа данных многопараметрического мониторинга неизбежно сводится к задаче принятия решений в условиях неполной информации на многомерных полях данных. Решение такого класса задач лежит в плоскости, многомерного анализа скалярных и векторных полей ДРМ, обнаружения аномальной динамики, и снижения избыточности и размерности за счет продвинутых средств визуализации. Для построения замкнутого решения задач такого класса разработано новое комплексное программно-алгоритмическое обеспечение, в рамках модульно-компонентного подхода, обеспечивающее поддержку принятия решения о наличии аномалий в ДРМ, которое позволяет в диалоговом режиме осуществлять выбор ожидаемых типов аномалий и формировать набор критериальной функций в соответствии с типовыми паттернами нормальных и аномальных данных. Динамически подключаемые проблемно-ориентированные модуля осуществляют, как регуляризацию и повышение размерности исходных данных, так и снижение их избыточности с заданным уровнем информационного критерия, позволяя

осуществлять свертку поступающего потока данных до 1-10% от его исходного объема при сохранении максимально возможного уровня информативности.

Проведена всесторонняя апробация предлагаемого методологического аппарата, методико-алгоритмического и программно-технического обеспечения, выполненного в рамках модульно-компонентного подхода, направленного на осуществление информационного обеспечения процессов обнаружения аномалий в данных мониторинга природной среды, выполненная как в рамках плановых исследований по базовым темам государственных заданий ФГБНУ «Институт природно-технических систем», так и при осуществлении научно-исследовательской и опытно-конструкторской работы по заказу АО «Концерн «Океанприбор».

При осуществлении поддержки принятия решений, с использованием динамически адаптируемых к характеру поступающих данных мониторинга природной среды проблемно-ориентированных модулей, реактивность увеличивается от 1,5 до 3-3,5 раз за счет применения выпуклого отображения в ортогональном базисе, ассиметричных информационных индексов, с применением компьютерного моделирования и баз знаний, ориентированных на реализацию в рамках модульно-компонентного подхода.

Применение разработанного программно-технического обеспечения для анализа данных в системах экологического мониторинга и лабораторного моделирующего стенда для адаптивного переключения ведущего модуля в соответствии с характером ДРМ позволило достичь уменьшения вероятности ложных срабатываний на 10-15% в зависимости от набора регистрируемых показателей интегрального критерия загрязненности природной среды.

Внедрение разработанных в ходе диссертационного исследования методов и алгоритмов позволяет осуществлять синтез информационного обеспечения для комплексных сетей мониторинговых наблюдений сложной гетерогенной структуры с уменьшением затрат на хранение данных на 20-25% в зависимости от структуры данных мониторинга, перечня и изменчивости их показателей.

Результаты научной работы позволяют повысить эффективность и надежность информационного и научно-методического обоснования уровней

контрольно-предупредительных границ отдельных приборов и средств контроля с техническими средствами принятия комплексных решений наличием аномалий в данных мониторинга, и минимизировать требования к объему данных для принятия решений на 14-28% за счет применения методов регуляризации, повышения размерности и имитационного моделирования.

Осуществлена кластеризация задач обнаружения аномалий мониторинга ПС и разработаны адаптивные проблемно-ориентированные решения для обнаружения аномалий выделенных классов. В особенности, предложены решения для ситуаций с различными объемами данных, наличием априорной информации о типах аномалий, требованиями к обработке в реальном времени. Предложены модули, реагирующие пул аналитических и приближенных методов представления ДРМ ПС, представляющий собой набор правил, осуществляющих отображение выборки нерегулярных и пропущенных данных в описывающую их поверхность в n -мерном пространстве, являющуюся входной для методов обнаружения аномалий. Он обобщает и существенно расширяет разнообразие доступных методов обнаружения аномалий, адаптируя уже существующие к решению А-задачи ПС.

Определены наиболее информативные наборы критериев оценки эффективности отдельных реализаций методов представления данных и обнаружения аномалий, определяемых для каждой конкретной постановки А-задачи и особенностей ее возникновения. Сформулирован комплексный критерий, дающий возможность ЛПР осуществлять выбор сценария использования за счет варьирования коэффициентов значимости отдельных метрик в зависимости от имеющихся приоритетов: точности, адекватности, вероятности ошибок I и II рода. Разработана модульная система поддержки принятия решений, включающая в себя алгоритмы автоматизированного обнаружения М-аномалий и N-аномалий в ДРМ ПС. На данных реального мониторинга ПС и имитационных примерах лабораторного стенда тестовой среды генерирующего управляемые возмущения оценена эффективность СППР на основе критериев точности, полноты, вероятностей ошибок I и II рода. Полученные оценки показали высокую

сходимость результатов для экспериментальных данных и целенаправленных экспериментов на лабораторном стенде на уровне 80–90%.

Современные методы решения А-задач обладают принципиальными недостатками и на сегодняшний день не существует универсального исчерпывающего эффективного решения А-задачи ПС. Сформулированы основные требования к методике решения А-задачи и ее программно-технической реализации, предусматривающие возможность обнаружения аномалий произвольных типов, в том числе профили, которых не заданы в системе, и позволяющие ЛПР гибко использовать соответствующий инструментарий для решения вопросов обнаружения аномалий состояния ПС. Разработаны принципы структуризации ДРМ ПС как полей данных высокой плотности, позволяющие осуществлять свертку отдельных объектов ПС в описывающие их векторы в n-мерном пространстве, обладающие высокой сходимостью с точки зрения решения А-задачи. Предложенный способ комбинирования сверток разной природы позволяет трансформировать входной поток в данные, размер которых составляет 3-7% от первоначального объема, и осуществлять сравнение этих полей данных по характеризующим их векторам свертки.

Разработан прототип модульной программно-технического обеспечения (МПТО), реализующего методику решения А-задач. В ходе проектирования и создания МПТО предложены программные и технические решения по повышению эффективности ее работы. МПТО является системой, которая органично дополняет функциональность приборов и средств контроля ПС и служит повышению их информационной надежности. Прототип МПТО успешно внедрен в опытную эксплуатацию в Лаборатории гидроакустических и оптических методов и средств контроля состояния окружающей среды Центра экологического приборостроения и экоэнергетики Института природно-технических систем. Модульность служит обеспечению его универсальности, облегчающей интеграцию в уже существующие приборы и средства контроля ПС для обнаружения аномалий ДРМ, в том числе в реальном времени. Обеспечивается легкость масштабирования предлагаемых решений, возможность адаптации к активно изменяющимся внешним условиям за счет структурной и параметрической гибкости.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Шишкин Ю.Е., Скатков А.В. Информационные технологии обнаружения аномалий в мониторинговых наблюдениях: монография. Симферополь: ИТ «АРИАЛ», 2019. 368 с. ISBN 978-5-907198-32-6. DOI: 10.33075/978-5-907032-64-4
2. Графкин А.В., Графкин В.В., Дегтярева О.А. Прикладной анализ случайных процессов. Под редакцией С.А. Прохорова. Самара, 2007.
3. Болгов М.В. Современное состояние теории корреляции для гидрологических расчетов и стохастического моделирования. Водное хозяйство России: проблемы, технологии, управление. 2012. № 4. С. 7–20.
4. Попов А.А., Саутин А.С. Построение регрессионных зависимостей с использованием алгоритма опорных векторов с адаптивными функциями потерь. Научный вестник Новосибирского государственного технического университета. 2011. № 1 (42). С. 17–26.
5. Назаров А.В., Лоскутов А.И. Нейросетевые алгоритмы прогнозирования и оптимизации систем. Профи. А. В. Назаров, А. И. Лоскутов. СПб. 2003.
6. Тулупьев А.Л. Ациклические алгебраические байесовские сети: логико-вероятностный вывод. Нечеткие системы и мягкие вычисления. 2006. Т. 1. № 1. С. 57–93.
7. Гулятьева Т.А., Попов А.А. Классификация зашумленных последовательностей, порожденных близкими скрытыми Марковскими моделями. Научный вестник Новосибирского государственного технического университета. 2011. № 3 (44). С. 3–16.
8. Двоенко С.Д., Динь Вьет Ш. Оценка параметров ациклических Марковских моделей при сегментации растровых текстурных изображений. Известия Тульского государственного университета. Технические науки. 2013. № 2. С. 86–95.
9. Бондур В.Г., Крапивин В.Ф., Солдатов В.Ю., Потапов И.И. Прогнозирование фазовых переходов в системе "Океан-атмосфера". Проблемы окружающей среды и природных ресурсов. 2010. № 5. С. 82–93.

10. Замятин А.В. Методы интеллектуального анализа данных в региональных системах аэрокосмического мониторинга. Вестник СГУГиТ (Сибирского государственного университета геосистем и технологий). 2016. № 4 (36). С. 74–88.
11. Веретёхин А.В. Обеспечение эколого-экономической безопасности промышленного предприятия в условиях повышенной неопределенности экзосреды. Вестник НГИЭИ. 2017. № 3 (70). С. 91–101.
12. Энциклопедический словарь Ф.А. Брокгауза и И.А. Ефрона (В 5 тт.) / Ред.: Андреевский И. Е., Арсеньев К. К., Петрушевский Ф. Ф. М.: Аутопан, 1998.
13. Zimek A., Schubert E. Outlier Detection // Encyclopedia of Database Systems. Springer New York, 2017. ISBN 9781489979933. doi:10.1007/978-1-4899-7993-3-1
14. Аномалия – Википедия. Свободная энциклопедия. URL: <https://ru.wikipedia.org/wiki/Аномалия> (дата обращения: 01.09.2020).
15. Большой российский энциклопедический словарь. Дрофа, 2009. 1888 с. ISBN: 978-5-85270-332-3
16. Ожегов С.И., Шведова Н.Ю. Толковый словарь русского языка. / Ред. Трушкин С.А. М.: ИТИ Технологии. Издание 4-е, доп. 2006. 944 с. ISBN: 5-902638-10-0
17. Даль В.И. Толковый словарь живого великорусского языка: (в 4 т.) / [соч.] Владимира Даля. - 3-е изд., испр. и знач. доп., изд. под ред. [и с предисл.] проф. И.А. Бодуэна-де-Куртенэ. Т. 1-4. Санкт-Петербург; Москва: т-во М. О. Вольф, 1903–1911.
18. Энциклопедический словарь медицинских терминов: (в 3 т.). Советская энциклопедия. 1982-1984.
19. Брокгауз Ф.А., Ефрон И.А. Энциклопедический словарь: (в 5 т.) / Ред.: Андреевский И. Е., Арсеньев К. К., Петрушевский Ф. Ф. М.: Аутопан, 1998.
20. Крысин Л.П. Современный словарь иностранных слов. сер. Настольные словари русского языка. АСТ-Пресс, 2020 г. 416 с. ISBN: 978-5-462-01308-9.
21. Ушаков Д.Н. Большой толковый словарь русского языка. Современная редакция. Славянский Дом Книги, 2014. 960 с. ISBN: 978-5-903036-99-8.
22. Абрамов Н.Ф. Словарь русских синонимов и сходных по смыслу выражений. Около 5000 синонимических рядов. / Ред. Гришина Е.А. АСТ, 2008. 672 с. ISBN: 978-5-17-037436-6.

23. Ефремова Т. Ф. Толковый словарь русского языка Ефремовой. Новый словарь русского языка. Толково-словообразовательный. (в 3 т.). Том 1. А-Л. М.: Русский язык, 2005. 1168 с. ISBN: 5-17-029520-0
24. Словарь русского языка: В 4-х т. / РАН, Ин-т лингвистич. исследований; Под ред. А. П. Евгеньевой. 4-е изд., стер. М.: Рус. яз. Полиграфресурсы, 1999.
25. Королюк В. С. О расхождении эмпирических распределений для случая двух независимых выборок // Изв. АН СССР. Сер. матем. № 19(1). 1955. С. 81–96.
26. Бондур В.Г., Цидилина М.Н., Кладов В.Л., Гордо К.А. Аномальная изменчивость пространственно-временных распределений природных пожаров и эмиссий вредных примесей на территории Европы по данным космического мониторинга // Доклады академии наук. Т. 485. № 6. С. 91–95.
27. Бондур В.Г., Серебряный А.Н., Замшин В.В., Тарасов Л.Л., Химченко Е.Е. Обнаружение аномального цуга интенсивных солитоноподобных внутренних волн на шельфе черного моря // Известия РАН, Физика атмосферы и океана. 2019. № 1. С. 114–127. DOI: 10.31857/S0002-3515531114-127
28. Бондур В.Г., Цидилина М.Н., Черепанова Е.В. Космический мониторинг воздействия природных пожаров на состояние различных типов растительного покрова в федеральных округах Российской Федерации // Исследование Земли из космоса. № 3. С. 13–32. DOI: 10.31857/S0205-96142019313-32
29. Ярушкина Н.Г. Основы теории нечетких и гибридных систем. Москва, 2004. 320 с.
30. Барский А.Б. Нейронные сети: распознавание, управление, принятие решений. Прикладные информационные технологии. Москва, 2004. 176 с.
31. Комарцова Л.Г., Максимов А.В. Нейрокомпьютеры. Учебное пособие для вузов. Информатика в техническом университете. Москва, 2004. 399 с.
32. Мыльник В.В., Титаренко Б.П., Волочиенко В.А. Исследование систем управления. Gaudeamus. Москва, 2003. 320 с.
33. Вагин В.Н., Головина Е.Ю., Загорянская А.А., Фомина М.В. Достоверный и правдоподобный вывод в интеллектуальных системах. Москва, 2008. 704 с.

34. Луценко Е.В. Интеллектуальные информационные системы. Учебное пособие для студентов очной и заочной форм обучения по специальности прикладная информатика. Краснодар, 2004. 645 с.
35. Фролов Ю.В. Интеллектуальные системы и управленческие решения. Москва, 2000. 294 с.
36. Lakhina A., Crovella M., Diot C. Mining anomalies using traffic feature distributions. ACM SIGCOMM Computer Communication Review, 35(4), 217. 2005. DOI: 10.1145/1090191.1080118
37. Wenke L., Stolfo, S.J., Chan, P.K. Real time data mining-based intrusion detection // Proceedings DARPA Information Survivability Conference and Exposition II. DISCEX'01. IEEE Comput. Soc. DOI: 10.1109/disceX.2001.932195
38. Вагин В.Н., Головина Е.Ю., Загорянская А.А., Фомина М.В. Достоверный и правдоподобный вывод в интеллектуальных системах // М: ФИЗМАТЛИТ, 2008. 712 с.
39. Дьяконов А.Г., Головина А.М. Выявление аномалий в работе механизмов методами машинного обучения // Аналитика и управление данными в областях с интенсивным использованием данных. Под ред. Л.А. Калиниченко, Я. Манолопулос, Н.А. Скворцова, В.А. Сухомлина. 2017. С. 469–476.
40. Дубнов Ю.А., Булычев А.В. Байесовская идентификация параметров смеси нормальных распределений // Информационные технологии и вычислительные системы. 2017. № 1. С. 101–111.
41. Ковалев С.М., Суханов А.В. Обнаружение особых типов паттернов во временных рядах на основе гибридной стохастической модели // Известия ЮФУ. Технические науки. 2014. № 4 (153). С. 142–149.
42. Фатхи Д.В., Галушка В.В. Методика определения оптимального числа нейронов выходного слоя сети Кохонена при решении задач кластеризации // Информационная безопасность регионов. 2011. № 2 (9). С. 41–44.
43. Кравченко Ю.А., Нацкевич А.Н., Курситыс И.О. Бустинг биоинспирированных алгоритмов для решения задачи кластеризации // Международная конференция по мягким вычислениям и измерениям. 2018. Т. 1. С. 777–780.

44. Hu W., Xie D., Tan T., Maybank S. Learning Activity Patterns Using Fuzzy Self-Organizing Neural Network // *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, 34(3), 2004. pp. 1618 – 1626. DOI: 10.1109/tsmcb.2004.826829
45. Camps-Valls, G., Bruzzone, L. (Eds.). *Kernel Methods for Remote Sensing Data Analysis*. John Wiley & Sons, Ltd. 2009. DOI: 10.1002/9780470748992
46. Ярушкина Н.Г. Основы теории нечетких и гибридных систем. М: Финансы и статистика. 2004. 320 с.
47. Бутакова М.А., Климанская Е.В., Янц В.И. Мера информационного подобия для анализа слабоструктурированной информации // *Современные проблемы науки и образования*. 2013. № 6. С. 130–136.
48. Мадера А.Г. Метод определения вероятностей прогнозируемых событий при принятии решений // *Искусственный интеллект и принятие решений*. 2016. № 2. С. 38–45.
49. Котов Э.М., Целых А.Н. Исследование моделей информационного поиска // *Известия ЮФУ. Технические науки*. 2009. № 4 (93). С. 163-168.
50. Разин Н.А., Черноусова Е.О., Красоткина О.В., Моттль В.В. Применение машины релевантных объектов в задачах восстановления числовых зависимостей // *Машинное обучение и анализ данных*. 2013. Т. 1. № 5. С. 641–652.
51. Кропотов Д.А. Вариационный метод релевантных векторов для задач классификации и регрессии с многомерными массивами признаков // *Журнал вычислительной математики и математической физики*. 2011. Т. 51. № 8. С. 1541–1560.
52. Shon, T., Moon, J. A hybrid machine learning approach to network anomaly detection // *Information Sciences*, 177(18), 2007. pp. 3799 – 3821. DOI: 10.1016/j.ins.2007.03.025
53. De la Hoz, E., De la Hoz, E., Ortiz, A., Ortega, J., & Martínez-Álvarez, A. Feature selection by multi-objective optimisation: Application to network anomaly detection by hierarchical self-organising maps // *Knowledge-Based Systems*, 71, 2014. pp. 322–338. DOI: 10.1016/j.knosys.2014.08.013
54. Shabtai A., Tenenboim-Chekina, L., Mimran, D., Rokach, L., Shapira, B., Elovici, Y. Mobile malware detection through analysis of deviations in application network behavior // *Computers & Security*, 43, 2014. pp. 1–18. DOI: 10.1016/j.cose.2014.02.009

55. Nadiammai, G. V., Hemalatha, M. (2014). Effective approach toward Intrusion Detection System using data mining techniques // *Egyptian Informatics Journal*, 15(1), 37–50. DOI: 10.1016/j.eij.2013.10.003
56. Шайтура С.В. Интеллектуальный анализ геоданных // *Перспективы науки и образования*. 2015. № 6 (18). С. 24–30.
57. Бондур В.Г., Килер Р.Н., Старченков С.А., Рыбакова Н.И. Мониторинг загрязнений прибрежных акваторий с использованием многоспектральных космических изображений высокого разрешения // *Исследование Земли из космоса*. 2006. № 6. С. 42–49.
58. Шелухин О.И., Сакалема Д.Ж., Филинова А.С. Обнаружение вторжений в компьютерные сети (сетевые аномалии) // *Учебное пособие для вузов / М: Горячая линия–Телеком*, 2013. 220 с.
59. Гвишиани А.Д., Диаман М., Михайлов В.О. Алгоритмы искусственного интеллекта для кластеризации магнитных аномалий // *Физика Земли*. 2002. № 7. С. 13–28.
60. Полонский А.Б., Башарин Д.В., Воскресенская Е.Н., Ворли С. Североатлантическое колебание: описание, механизмы и влияние на климат Евразии // *Морской гидрофизический журнал*. 2004. № 2. С. 42–59.
61. Ватутин Э.И., Титов В.С. Анализ результатов применения алгоритма муравьиной колонии в задаче поиска пути в графе при наличии ограничений // *Известия ЮФУ. Технические науки*. 2014. № 12 (161). С. 111–120.
62. Чернышев Ю.О., Григорьев Г.В., Венцов Н.Н. Искусственные иммунные системы: обзор и современное состояние // *Программные продукты и системы*. 2014. № 4. С. 136–142.
63. Meng, Y., & Kwok, L. (2012). Intrusion Detection Using Disagreement-Based Semi-supervised Learning: Detection Enhancement and False Alarm Reduction // *Cyberspace Safety and Security* (pp. 483–497). Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-35362-8_36
64. Охтилев М.Ю., Соколов Б.В., Юсупов Р.М. Интеллектуальные технологии мониторинга и управления структурной динамикой сложных технических объектов. М.: Наука, 2006. 410 с.

65. Майданович О.В., Охтилев М.Ю., КуССуль Н.Н., Соколов Б.В., Цивирко Е.Г., Юсупов Р.М. Междисциплинарный подход к оцениванию и анализу эффективности информационных технологий и систем / Приборостроение. 2010. Т.53, №11. С. 7–16.
66. Майданович О.В., Охтилев М.Ю., Соколов Б.В., Юсупов Р.М., Цивирко Е.Г. Количественное и качественное оценивание влияния информационных технологий на эффективность систем управления сложными объектами // Моделирование и анализ безопасности и риска в сложных системах: тр. Международной науч. Шк. МА БР-2010 (Санкт-Петербург, 6–10 июля, 2010 г.). СПб.: ГУАП, 2010. С. 79–84.
67. Охтилев М.Ю., Мустафин Н.Г., Миллер В.Е., Соколов Б.В. Концепция проактивного управления сложными объектами: теоретические и технологические основы // Известия высших учебных заведений. Приборостроение. 2014. Т. 57. № 11. С. 7–15.
68. Охтилев М.Ю., Соколов Б.В., Юсупов Р.М. Теоретические и технологические основы концепции проактивного мониторинга и управления сложными объектами // Известия ЮФУ. Технические науки. 2015. № 1 (162). С. 162–174.
69. Охтилев М.Ю., Соколов Б.В. Теоретические и прикладные проблемы разработки и применения автоматизированных систем мониторинга состояния сложных технических объектов // Труды СПИИРАН. 2002. Т. 1. № 1. С. 167–180.
70. Охтилев М.Ю., Соколов Б.В. Новые информационные технологии мониторинга и управления состояниями сложных технических объектов в реальном масштабе времени // Труды СПИИРАН. 2005. Т. 2. № 2. С. 249–265.
71. Майданович О.В., Соколов Б.В., Охтилев М.Ю. Новый подход к созданию интеллектуальных информационных технологий проектирования систем мониторинга состояния сложных объектов // XI Международная научно-техн. конф. 12–14 мая 2010 г. «Кибернетика и высокие технологии XXI века»: сб. докл. Воронеж: НПФ «Саквое», 2010. С. 601–608.
72. Костяной А.Г. Спутниковый мониторинг климатических параметров океана. часть 1 // Фундаментальная и прикладная климатология. 2017. Т. 2. С. 63–85.

73. Bykova M. Statistical analysis of malformed packets and their origins in the modern Internet. School of Electrical Engineering & Computer Science Ohio University, 2002
74. Krugel C., Toth T., Kirda E. Service Specific Anomaly Detection for Intrusion Detection. Vienna, Austria, 2002
75. Krugel C. Network Alertness Towards an adaptive, collaborating Intrusion Detection System. Dissertation, Vienna, Austria, 2002
76. Chan P.K., Mahoney M.V., Arshad M.H. A Machine Learning Approach to Anomaly Detection // Technical report, Florida Institute of Technology, Melbourne, 2003.
77. Portnoy L., Eskin E., Stolfo S. Intrusion Detection with Unlabeled Data Using Clustering, Department of Computer Science Columbia University, New York, 2001.
78. Shah H., Undercoffer J., Joshi A. Fuzzy Clustering for Intrusion Detection 11 Proceedings of the 12th IEEE International Conference on Fuzzy Systems, 2003.
79. Lee W., Xiang D. Information-Theoretic Measures for Anomaly Detection // Proceedings. IEEE Symposium, 2001.
80. Станкевич Л.А., Казанский А.Б. Иммунологическая система обеспечения безопасности гуманоидного робота // Актуальные проблемы защиты и безопасности: тр. 9-й Всерос. науч.-практ. конф., 2006. № 5. С. 145–152.
81. Garrett S.M. How do we evaluate artificial immune systems? How do we evaluate artificial immune systems? 2005. Vol. 13. P. 145–178.
82. Hunt J.E., Cooke D.E. Learning using an artificial immune system // Journ. of Network Computing Applications, 1996. Vol. 19. P. 189–212.
83. Knight T., Timmis J. Aine: An immunological approach to data mining // IEEE Intern. Conf. on Data Mining, 2001. P. 297–304.
84. Kim J., Bentley P. Towards an artificial immune system for network intrusion detection: An investigation of dynamic clonal selection. In Proc. Congress on Evolutionary Computation, Honolulu, HI, USA, 2002. P. 1244–1252.
85. Скатков А.В., Брюховецкий А.А., Моисеев Д.В. Адаптация механизмов искусственных иммунных систем для контроля параметров окружающей среды / Системы контроля окружающей среды. 2020. № 2 (40). С. 127–133.

86. Скатков А.В., Брюховецкий А.А., Моисеев Д.В. Использование Марковских моделей для оценки состояния окружающей среды // В книге: Системы контроля окружающей среды - 2019. Тезисы докладов Международной научно-технической конференции. 2019. С. 115.
87. Скатков А.В., Брюховецкий А.А., Моисеев Д.В. Обнаружение аномалий экосистем акваторий г. Севастополя на основе Марковской модели // В сборнике: Экологическая, промышленная и энергетическая безопасность - 2019. Сборник статей по материалам международной научно-практической конференции. Под редакцией Л.И. Лукиной, Н.В. Ляминой. 2019. С. 1502–1505.
88. Брюховецкий А.А., Моисеев Д.В., Скатков А.В. Интеллектуальная технология обнаружения аномалий экосистем акватории г. Севастополя на основе кластеризации данных // В сборнике: Экологическая, промышленная и энергетическая безопасность - 2019. Сборник статей по материалам международной научно-практической конференции. Под редакцией Л.И. Лукиной, Н.В. Ляминой. 2019. С. 310-314.
89. Yang H., Li T., Hu X., Wang F., Zou Y. A Survey of Artificial Immune System Based Intrusion Detection // The Scientific World Journal. 2014.
90. Вторый В.Ф., Вторый С.В. Перспективы экологического мониторинга сельскохозяйственных объектов с использованием беспилотных летательных аппаратов // Технологии и технические средства механизированного производства продукции растениеводства и животноводства. 2017. № 92. С. 158-166. DOI: 10.24411/0131-5226-2017-00028
91. Skatkov A.V., Bryukhovetskiy A.A. and Moiseev D.V. 2020 Adaptive vulnerability detection model for unmanned vehicles drugs based on artificial immune systems IOP Conference Series: Materials Science and Engineering 734 012028 iopscience.iop.org/article/ DOI: 10.1088/1757-899X/734/1/012028
92. Intelligent positive computing with mobile, wearable, and IoT devices: Literature review and research directions / Lee U., Han K. [et. al] // Ad Hoc Networks. 2018. Vol. 83. pp 8–24.

93. Атлас океанографических характеристик Севастопольской бухты / С.К. Коновалов, А.С. Романов, О.Г. Моисеенко [и др.] // Севастополь: «ЭКОСИ-ГИДРОФИЗИКА» 2010. 320 с.
94. Düzgün H. Ş., Demirel N. Remote Sensing of the Mine Environment. London: CRC Press, 2017. 220 P.
95. Лепихин А.П., Возняк А.А. Статистические функции распределения гидрохимических показателей качества воды поверхностных водных объектов // Водное хозяйство России: проблемы, технологии, управление. 2012. № 4. С. 21–32.
96. Kind A., Stoecklin M. P., Dimitropoulos X. Histogram-based traffic anomaly detection // IEEE Transactions on Network and Service Management, vol. 6, no. 2. pp. 110–121, 2009. DOI: 10.1109/TNSM.2009.090604
97. Mitchell R., Chen I. Behavior-Rule Based Intrusion Detection Systems for Safety Critical Smart Grid Applications // IEEE Transactions on Smart Grid, vol. 4, no. 3. pp. 1254–1263. 2013. DOI: 10.1109/TSG.2013.2258948
98. Huang L. Communication-Efficient Online Detection of Network-Wide Anomalies // IEEE INFOCOM 2007 – 26th IEEE International Conference on Computer Communications, Barcelona, 2007. pp. 134–142. DOI: 10.1109/INFCOM.2007.24
99. Xu W., Huang L., Fox A., Patterson D., Jordan M. Online System Problem Detection by Mining Patterns of Console Logs // 2009 Ninth IEEE International Conference on Data Mining, Miami, FL, 2009. pp. 588–597. DOI: 10.1109/ICDM.2009.19
100. Sotiris V. A., Tse P. W., Pecht M. G. Anomaly Detection Through a Bayesian Support Vector Machine // IEEE Transactions on Reliability, vol. 59, no. 2. pp. 277–286. 2010. DOI: 10.1109/TR.2010.2048740
101. O'Reilly C., Gluhak A., Imran M. A., Rajasegarar S. Anomaly Detection in Wireless Sensor Networks in a Non-Stationary Environment // IEEE Communications Surveys & Tutorials, vol. 16, no. 3. pp. 1413–1432, Third Quarter 2014. DOI: 10.1109/SURV.2013.112813.00168
102. Hu Y., Wielicki B. A., Yang P., P. Stackhouse W., Lin B., Young D. F. Application of deep convective cloud albedo observation to satellite-based study of the terrestrial

- atmosphere: monitoring the stability of spaceborne measurements and assessing absorption anomaly // *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 11. pp. 2594–2599. 2004. DOI: 10.1109/TGRS.2004.834765
103. Yuan Y., Wang Q., Zhu G. Fast Hyperspectral Anomaly Detection via High–Order 2–D Crossing Filter // *IEEE Transactions on Geoscience and Remote Sensing*. 2015. vol. 53, no. 2. pp. 620–630. DOI: 10.1109/TGRS.2014.2326654
104. Agrawal S., Deb S., Naidu K. V. M., Rastogi R. Efficient Detection of Distributed Constraint Violations // *2007 IEEE 23rd International Conference on Data Engineering*, Istanbul, 2007. pp. 1320–1324. DOI: 10.1109/ICDE.2007.369002
105. Alarcon-Aquino V., Barria J. A. Anomaly detection in communication networks using wavelets // *IEE Proceedings – Communications*. 2001. vol. 148, no. 6. pp. 355–362. DOI: 10.1049/ip-com:20010659
106. A PDCA-based approach to Environmental Value Stream Mapping (E-VSM) / J.A. Garza-Reyes, J.T. Romero, K. Govindan [et al.] // *Journal of Cleaner Production*. 2018. Vol. 180. pp. 335–348. DOI: 10.1016/j.jclepro.2018.01.121
107. Vychuzhanin P., Hvatov A., Kalyuzhnaya A.V. Anomalies Detection in Metocean Simulation Results Using Convolutional Neural Networks // *Procedia Computer Science*. 2018. Vol. 136. pp. 321–330. DOI: 10.1016/j.procs.2018.08.282
108. Comparison of atmospheric particle concentration measurements using different optical detectors: Potentiality and limits for air quality applications / A. Dinoi, A. Donato, F. Belosi [et al.] // *Measurement*. 2017. Vol. 106. pp. 274–282. DOI: 10.1016/j.measurement.2016.02.019
109. Харман Г. Современный факторный анализ. М.: Статистика. 1972. 489 с.
110. Семенов В.А. Теория вероятностей и математическая статистика: Учебное пособие. Стандарт третьего поколения. СПб.: Питер. 2013. 192 с.
111. G. R Bradski and A. Kaehler. *Learning OpenCV: Computer vision with the OpenCV library*. O’Reilly Media, 2013.
112. B. Luca, S. Mario, P. Margherita, V. Roberto, and P. Marcello, “Deep Learning-Based Method for Vision-Guided Robotic Grasping of Unknown Objects,” *ATDE*, vol.

- 7, no. Transdisciplinary Engineering Methods for Social Innovation of Industry 4.0, pp. 281–290, 2018.
113. J. Daugman, “How Iris Recognition Works,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 21–30, Jan. 2004.
114. I. E. Shishkin, and A. N. Grekov, “Analysis of Image Clusterization Methods for Oceanographical Equipment,” 2018 International Russian Automation Conference (RusAutoCon), pp. 1–6, 2018.
115. A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, “CNN Features Off-the-Shelf: An Astounding Baseline for Recognition,” in 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2014.
116. C. S. Davis, F. T. Thwaites, S. M. Gallager, and Q. Hu, “A three-axis fast-tow digital Video Plankton Recorder for rapid surveys of plankton taxa and hydrography,” *Limnology and Oceanography: Methods*, vol. 3, no. 2, pp. 59–74, Feb. 2005.
117. M. Picheral, L. Guidi, L. Stemmann, D. M. Karl, G. Iddaoud, and G. Gorsky, “The Underwater Vision Profiler 5: An advanced instrument for high spatial resolution studies of particle size spectra and zooplankton,” *Limnology and Oceanography: Methods*, vol. 8, no. 9, pp. 462–473, Sep. 2010.
118. S. Samson, T. Hopkins, A. Remsen, L. Langebrake, T. Sutton, and J. Patten, “A system for high-resolution zooplankton imaging,” *IEEE Journal of Oceanic Engineering*, vol. 26, no. 4, pp. 671–676, 2001.
119. H. Zheng, R. Wang, Z. Yu, N. Wang, Z. Gu, and B. Zheng, “Automatic plankton image classification combining multiple view features via multiple kernel learning,” *BMC Bioinformatics*, vol. 18, no. S16, Dec. 2017.
120. J. S. Jaffe, “To sea and to see: That is the answer,” 2016 Meth Oceanogr, pp. 3–20, 2016.
121. R. J. Olson and H. M. Sosik, “A submersible imaging-in-flow instrument to analyze nano-and microplankton: Imaging FlowCytobot,” *Limnology and Oceanography: Methods*, vol. 5, no. 6, pp. 195–203, Jun. 2007.
122. R. K. Cowen and C. M. Guigand, “In situ ichthyoplankton imaging system (ISIIS): system design and preliminary results,” *Limnology and Oceanography: Methods*, vol. 6, no. 2, pp. 126–132, Feb. 2008.

123. G. Gorsky et al., “Digital zooplankton image analysis using the ZooScan integrated system,” *Journal of Plankton Research*, vol. 32, no. 3, pp. 285–303, Feb. 2010.
124. H. Zheng, R. Wang, Z. Yu, N. Wang, Z. Gu, and B. Zheng, “Automatic plankton image classification combining multiple view features via multiple kernel learning,” *BMC Bioinformatics*, vol. 18, no. S16, Dec. 2017.
125. J. Yu et al., “Video-Based Real Time Analysis of Plankton Particle Size Spectrum,” *IEEE Access*, vol. 7, pp. 60020–60025, 2019.
126. P. Gonzalez et al., “Multiclass Support Vector Machines With Example-Dependent Costs Applied to Plankton Biomass Estimation,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 11, pp. 1901–1905, Nov. 2013.
127. Y. E. Shishkin, “Big Data visualization in decision making,” *Science in Progress*, pp. 203–205, 2016.
128. G. Tsechpenakis, C. Guigand, and R. K. Cowen, “Image Analysis Techniques to Accompany a new In Situ Ichthyoplankton Imaging System,” in *OCEANS 2007 Europe*, 2007.
129. X. Qi and L. Miao, “A Template Matching Method for Multi-Scale and Rotated Images Using Ring Projection Vector Conversion,” in *2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)*, 2018.
130. W. C. Lee and C. H. Chen, “A Fast Template Matching Method for Rotation Invariance Using Two-Stage Process,” in *2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2009.
131. B. Liu, X. Shu, and X. Wu, “Fast Screening Algorithm for Rotation Invariant Template Matching,” in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018.
132. R. Y. Wong and E. L. Hall, “Scene matching with invariant moments,” *Computer Graphics and Image Processing*, vol. 8, no. 1, pp. 16–24, Aug. 1978.
133. Ming-Kuei Hu, “Visual pattern recognition by moment invariants,” *IEEE Transactions on Information Theory*, vol. 8, no. 2, pp. 179–187, Feb. 1962.
134. Stepashko V. From Inductive to Intelligent Modeling // *2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)*, Lviv. 2018. pp. 32–35. DOI: 10.1109/STC-CSIT.2018.8526641

135. Mirzaraxmedova A. X., Fozilova M. M. Analysis of Prospects of Technology of Intelligent Monitoring Systems // 2019 International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, Uzbekistan. 2019. pp. 1–8. DOI: 10.1109/ICISCT47635.2019.9011845
136. Korablyov M., Axak N., Barkovska O., Rosinskiy D. Hybrid adaptive decision-making model based on fuzzy logic and artificial immune systems // 2019 IEEE 14th International Conference on Computer Sciences and Information Technologies (CSIT), Lviv, Ukraine. 2019. pp. 62–66. DOI: 10.1109/STC-CSIT.2019.8929826
137. Tianshu W., Shuyu C., Jie Y., Peng W. Intelligent prognostic and health management based on IOT cloud platform // 2019 14th IEEE International Conference on Electronic Measurement & Instruments (ICEMI), Changsha, China. 2019. pp. 1089–1096. DOI: 10.1109/ICEMI46757.2019.9101690
138. Hall D. L., Llinas J. An introduction to multisensor data fusion // Proceedings of the IEEE. 1997. vol. 85, no. 1. pp. 6–23. DOI: 10.1109/5.554205
139. Guerrero J. M., Chandorkar M., Lee T., Loh P. C. Advanced Control Architectures for Intelligent Microgrids-Part I: Decentralized and Hierarchical Control // IEEE Transactions on Industrial Electronics. 2013. vol. 60, no. 4. pp. 1254–1262. DOI: 10.1109/TIE.2012.2194969
140. Islam S. M., Kwak D., Kabir M. H., Hossain M., Kwak K. The Internet of Things for Health Care: A Comprehensive Survey // IEEE Access, vol. 3. pp. 678–708. 2015. DOI: 10.1109/ACCESS.2015.2437951
141. McIlraith S. A., Son T. C., Zeng H. Semantic Web services // IEEE Intelligent Systems, vol. 16, no. 2. pp. 46–53. 2001. doi: 10.1109/5254.920599
142. Sandu L., Oksman J., Fleury G. A. Information criteria for the choice of parametric functions for measurement // IEEE Transactions on Instrumentation and Measurement, vol. 47, no. 4, pp. 920–924. 1998. DOI: 10.1109/19.744643
143. Gajda J., Szyper M. Parametric optimization of measuring systems according to the joint error criterion // IEEE Transactions on Instrumentation and Measurement, vol. 46, no. 4, pp. 769–775. 1997. DOI: 10.1109/19.650770

144. Gajda J., Szyper M. Parametric optimization of measuring systems according to the joint error criterion // *Quality Measurement: The Indispensable Bridge between Theory and Reality*. IEEE Instrumentation and Measurement Technology Conference and IMEKO Tec, Brussels, Belgium. 1996. vol.1. pp. 174–179. DOI: 10.1109/IMTC.1996.507370
145. Heon K., Seo Y. Kyun W. Robust stability analysis of parametric uncertain time-delay systems // *Proceedings of the 37th IEEE Conference on Decision and Control (Cat. No.98CH36171)*, Tampa, FL, USA, 1998, vol.2. pp. 1346–1351. DOI: 10.1109/CDC.1998.758472
146. Pushkarev M. I., Gaivoronsky S. A. Control system robust controller parametric synthesis based on coefficient estimation of stability and oscillation indices // *2012 7th International Forum on Strategic Technology (IFOST)*, Tomsk, 2012, pp. 1–4. DOI: 10.1109/IFOST.2012.6357658
147. Onoda T. Experimental analysis of generalization capability based on information criteria // *Proceedings of International Conference on Neural Networks (ICNN'96)*, Washington, DC, USA. 1996. vol.1. pp. 114–119. DOI: 10.1109/ICNN.1996.548876
148. Chen T. Multiple Criteria Group Decision Making Using a Parametric Linear Programming Technique for Multidimensional Analysis of Preference Under Uncertainty of Pythagorean Fuzziness // *IEEE Access*. 2019. vol. 7. pp. 108–128. DOI: 10.1109/ACCESS.2019.2957161
149. Wu B., Ma L., Perng J., Chin H. Absolute stability analysis in uncertain static fuzzy control systems with the parametric robust Popov criterion // *2008 IEEE International Conference on Fuzzy Systems (IEEE World Congress on Computational Intelligence)*, Hong Kong. 2008. pp. 1325–1330. DOI: 10.1109/FUZZY.2008.4630544
150. Xia L. Optimization of parametric policies of Markov decision processes under a variance criterion // *2016 13th International Workshop on Discrete Event Systems (WODES)*, Xi'an. 2016. pp. 332–337. DOI: 10.1109/WODES.2016.7497868
151. Chimitova E. V., Vedernikova M. A. Goodness of fit criteria for checking adequacy of parametric models for reliability and survivalability // *2012 IEEE 11th International*

- Conference on Actual Problems of Electronics Instrument Engineering (APEIE), Novosibirsk. 2012. pp. 58–65. DOI: 10.1109/APEIE.2012.6629054
152. Padhy P. K., Majhi S. Parametric identification of non-linear systems // Proceedings of the IEEE INDICON 2004. First India Annual Conference, 2004. Kharagpur, India. 2004. pp. 220–224. DOI: 10.1109/INDICO.2004.1497742
153. Боев В.Д. Имитационное моделирование систем: учебное пособие для прикладного бакалавриата. М.: Юрайт. 2018. 253 с.
154. Emiliano C., Daniele P. Handling obstacles in pedestrian simulations: Models and optimization // Applied Mathematical Modelling, № 45. 2017. С. 285–302.
155. Hee-Soo K., Seok-Won L., Dependability-enhanced unified modeling and simulation methodology for Critical Infrastructures // Information and Software Technology, № 102, 2018, pp. 175–192. DOI: 10.1016/j.infsof.2018.06.002
156. Лисов А.А., Чернова Т.А., Горбунов М.С. Системный подход к анализу деградационных процессов в электротехнических устройствах // Труды МАИ. 2018. № 102. С. 23.
157. Горбунов А.Р., Лычкина Н.Н. Парадигмы имитационного моделирования: новое в решении задач стратегического управления (объединенная логика имитационного моделирования) // Бизнес-информатика. 2007. № 2 (2). С. 60–66.
158. Дозорцев В.М. Компьютерные тренажеры для обучения операторов технологических процессов. М.: Синтег. 2009. 365 с.
159. Шишкин Ю.Е. Свидетельство № 2018617880 Российская Федерация. Имитационная модель маршрутизатора самоорганизующейся сети передачи данных «МАРШРУТ»: свидетельство о государственной регистрации программы для ЭВМ / Ю.Е. Шишкин; заявитель и правообладатель Ю.Е. Шишкин; № 2018614503; заявл. 08.05.2018, опубл. 03.07.2018; Бюл. № 7. - 1 с.
160. Кручинин С.В. Семиуровневая модель OSI/ISO и стек протоколов TCP/IP: исследование взаимоотношения и интерпретации // Научно-исследовательские публикации. 2015. № 5 (25). С. 115–120.
161. Шишкин Ю.Е., Скатков А.В. Программно-аппаратный модуль поддержки принятия решений о наличии качественных аномальных изменений в выборочных

- данных на базе информационных метрик / Системы контроля окружающей среды. 2021. № 2 (44). С. 142–151. DOI: 10.33075/2220-5861-2021-2-142-151
162. Анализ данных и процессов: учеб. пособие / А.А. Барсегян, М.С. Куприянов, И.И. Холод [и др.]. 3-е изд., перераб. и доп. СПб.: БХВ-Петербург. 2009. 512 с.
163. Зубриенко Г.А., Лапоница О.Р. Методы оптимизации выборки данных для определения аномального трафика / International Journal of Open Information Technologies. 2016. Т. 4. № 10. С. 1–8.
164. Шишкин Ю.Е. Исследование возможностей систем обнаружения заимствований в методологии Больших Данных // Фундаментальные основы инновационного развития науки и образования: Монография / Под ред. Г.Ю. Гуляева. Пенза: Наука и Просвещение, 2017. С. 55–73.
165. Шишкин Ю.Е., Лей В.А. Организация воспитательной работы в вузе с позиций безопасности жизнедеятельности // Техносфера XXI века: материалы 3-й всероссийской конференции молодых ученых. Под ред. Г.А. Сигора. 2018. С. 120–121.
166. Головкин О.Н., Шишкин Ю.Е. Информационные технологии в диагностике нравственной воспитанности студентов высшей школы // Информатизация непрерывного образования – 2018: материалы Международной научной конференции: в 2 томах. / под общей редакцией В. В. Гриншкун. 2018. С. 23-26.
167. The International Thermodynamic Equation of Seawater 2010 (TEOS-10): Calculation and Use of Thermodynamic Properties / T.J. McDougall, R. Feistel, F.J. Millero [et al]. Intergovernmental Oceanographic Commission IOC of Unesco. 2010. 218 pp.
168. Резолюция № 5/SIMBF/2020 // VIII Международный морской бизнес-форум СИ МБФ 2020. 06-07 октября 2020. 16 с.
169. Шишкин Ю.Е., Греков А.Н. Статистические методы кластеризации изображений гидробионтов // Системы контроля окружающей среды. 2020. № 1 (39). С. 153–159. DOI: 10.33075/2220-5861-2020-1-153-159
170. Скатков А.В., Брюховецкий А.А., Моисеев Д.В., Шишкин Ю.Е. Дискриминантный подход к обнаружению аномалий с использованием марковских последовательностей // Системы контроля окружающей среды. 2019. № 4 (38). С. 43–49. DOI: 10.33075/2220-5861-2019-4-43-49

171. Скатков А.В., Брюховецкий А.А., Моисеев Д.В., Шишкин Ю.Е. Мера Кульбака в задачах динамической кластеризации наблюдений состояния окружающей среды // Системы контроля окружающей среды. 2019. № 3 (37). С. 35–38. DOI: 10.33075/2220-5861-2019-3-35-38
172. Шишкин Ю.Е., Скатков А.В. Интеллектуальная система адаптивного выбора сценариев параметрического обнаружения дивергенций данных мониторинга // Системы контроля окружающей среды. 2019. № 2 (36). С. 37–42. DOI: 10.33075/2220-5861-2019-2-37-42
173. Шишкин Ю.Е., Скатков А.В., Маловик К.Н. Модельные представления для мобильных приложений оценки и прогнозирования аномальных и паталогических состояний // Качество и жизнь. 2019. № 1 (21). С. 55–60.
174. Шишкин Ю.Е., Скатков А.В. Метрики качества для оценки и прогнозирования критических состояний // Качество и жизнь. 2019. № 1 (21). С. 61–66.
175. Скатков А.В., Брюховецкий А.А., Шишкин Ю.Е. Разработка интеллектуальной технологии обнаружения аномалий экосистем акватории г. Севастополя // Системы контроля окружающей среды. 2019. № 1 (35). С. 27–34. DOI: 10.33075/2220-5861-2019-1-27-34
176. Шишкин Ю.Е., Скатков А.В. Повышение достоверности оценок рисков в процессах мониторинга при распределениях общего вида // Системы контроля окружающей среды. 2019. № 1 (35). С. 41–47. DOI: 10.33075/2220-5861-2019-1-41-47
177. Скатков А.В., Шишкин Ю.Е. Кластеризация данных в задачах обнаружения аномалий на основе ортогональных фильтров // Системы контроля окружающей среды. 2018. № 11 (31). С. 36–43.
178. Шишкин Ю.Е., Скатков А.В. Метод распознавания в интерактивном режиме аномалий градиентов скалярных полей наблюдений // Системы контроля окружающей среды. 2018. № 12 (32). С. 30–37.
179. Шишкин Ю.Е., Скатков А.В. Акторная модель мониторинга с использованием мобильных облачных микросервисов // Системы контроля окружающей среды. 2018. № 14 (34). С. 56–62. DOI: 10.33075/2220-5861-2018-4-56-62

180. Шишкин Ю.Е., Греков А.Н. Концепция интеллектуальной системы автоматизированного экологического мониторинга на базе малогабаритных автономных роботов // Системы контроля окружающей среды. 2018. № 14 (34). С. 63–69. DOI: 10.33075/2220-5861-2018-4-63-69
181. Скатков А.В., Шишкин Ю.Е. Модель обнаружения аномалий в наблюдениях параметров полей окружающей среды с использованием систем мониторинга // Системы контроля окружающей среды. 2017. № 10 (30). С. 48–53.
182. Брюховецкий А.А., Скатков А.В., Шишкин Ю.Е. Моделирование процессов обнаружения аномалий в сложноструктурированных данных мониторинга // Системы контроля окружающей среды. 2017. № 9 (29). С. 45–49.
183. Shishkin I.E., Grekov A.N. Development of Image Analysis Methods for Detecting Nonhomogeneity and Anomalies in the Marine Environment // в сборнике: Proceedings - 2021 International Russian Automation Conference, RusAutoCon 2021. 2021. С. 1–6.
184. Shishkin I.E., Grekov A.N. Methods for Recognizing Images of Heterogeneous Objects in Small Training Sample // в сборнике: 2020 International Multi-Conference on Industrial Engineering and Modern Technologies, FarEastCon 2020. 2020. С. 1–6. DOI:10.1109/FarEastCon50210.2020.9271330
185. Shishkin, I. E., Grekov, A. N., Nikishin, V. V. Intelligent Decision Support System for Detection of Anomalies and Unmanned Surface Vehicle Inertial Navigation Correction // в сборнике: Proceedings - 2019 International Russian Automation Conference, RusAutoCon 2019. 2019. С. 1–6. DOI:10.1109/rusautocon.2019.8867601
186. Shishkin I.E., Grekov A.N., Grekov N.A. A multi-model system of intelligent unmanned surface vehicles for environmental monitoring // в сборнике: 2019 International Multi-Conference on Industrial Engineering and Modern Technologies, FarEastCon 2019. 2019. С. 1–6. DOI:10.1109/FarEastCon.2019.8934037
187. Shishkin I.E., Grekov A.N. Analysis of image clusterization methods for oceanographical equipment // В сборнике: 2018 International Russian Automation Conference, RusAutoCon 2018. 2018. С. 1–6. DOI:10.1109/RUSAUTOCON.2018.8501756
188. Шишкин Ю.Е., Кузьмин К.А., Алексеев С.Ю. Свидетельство № 2018666181 Российская Федерация. Программа синхронизации и коррекции измерений вектора

скорости течения «ИНТЕГРАЦИЯ»: свидетельство о государственной регистрации программы для ЭВМ / Ю.Е. Шишкин, К.А. Кузьмин, С.Ю. Алексеев; заявитель и правообладатель ИПТС; № 2018663369; заявл. 26.11.2018, опубл. 13.12.2018; Бюл. № 12. - 1 с.

189. Шишкин Ю.Е. Свидетельство № 2017664038 Российская Федерация. Модуль снижения избыточности данных мониторинга «МОСИДАМ»: свидетельство о государственной регистрации программы для ЭВМ / Ю.Е. Шишкин; заявитель и правообладатель Ю.Е. Шишкин; № 2017660984; заявл. 27.10.2017, опубл. 14.12.2017; Бюл. № 12. - 1 с.

190. Шишкин Ю.Е. Свидетельство № 2018613184 Российская Федерация. Модуль консолидации структурированных данных большого объема в режиме реального времени «МОКОДАБО»: свидетельство о государственной регистрации программы для ЭВМ / Ю.Е. Шишкин; заявитель и правообладатель Ю.Е. Шишкин; № 2018610307; заявл. 18.01.2018, опубл. 05.03.2018; Бюл. № 3. - 1 с.

191. Шишкин Ю.Е. Свидетельство № 2020662839 Российская Федерация. Программа распознавания аномальностей в данных мониторинговых наблюдений «А-КОНСУЛЬТАНТ»: свидетельство о государственной регистрации программы для ЭВМ / Ю.Е. Шишкин, А.В. Скатков; заявитель и правообладатель Ю.Е. Шишкин, А.В. Скатков; № 2020660075; заявл. 07.09.2020, опубл. 20.10.2020; Бюл. № 10. - 1 с.

192. Кузьмин К.А. Свидетельство № 2020664093 Российская Федерация. Программа для автономного модуля буйкового акустического измерителя скорости течения «ИСТ-1МА-SD»: свидетельство о государственной регистрации программы для ЭВМ / К.А. Кузьмин, Ю.Е. Шишкин; правообладатель Федеральное государственное бюджетное научное учреждение «Институт природно-технических систем» (ИПТС); № 2020663317; заявл. 27.10.2020, опубл. 06.11.2020; Бюл. № 11. - 1 с.

193. Шишкин Ю.Е. Свидетельство № 2020664094 Российская Федерация. Программа GPS навигатора для автономного буйкового акустического измерителя скорости течения «А-НАВИГАТОР»: свидетельство о государственной

регистрации программы для ЭВМ / Ю.Е. Шишкин, К.А. Кузьмин; правообладатель Федеральное государственное бюджетное научное учреждение «Институт природно-технических систем» (ИПТС); № 2020663316; заявл. 27.10.2020, опубл. 06.11.2020; Бюл. № 11. - 1 с.

194. Греков А.Н., Шишкин Ю.Е, Кузьмин К.А., Мишуров В.Ж., Рязанов В.А., Трусевич В.В. Патент № 2755407 Российская Федерация. Способ контроля качества воды и двухканальная биосенсорная установка для его осуществления / правообладатель Федеральное государственное бюджетное научное учреждение «Институт природно-технических систем» (ИПТС); № 2020141487; заявл. 15.12.2020, опубл. 15.09.2021; Бюл. № 26-2021. - 1 с.

ПРИЛОЖЕНИЕ А. ПРИМЕРЫ АНОМАЛИЙ ПРИРОДНОЙ СРЕДЫ
ОБНАРУЖИВАЕМЫХ СИСТЕМОЙ

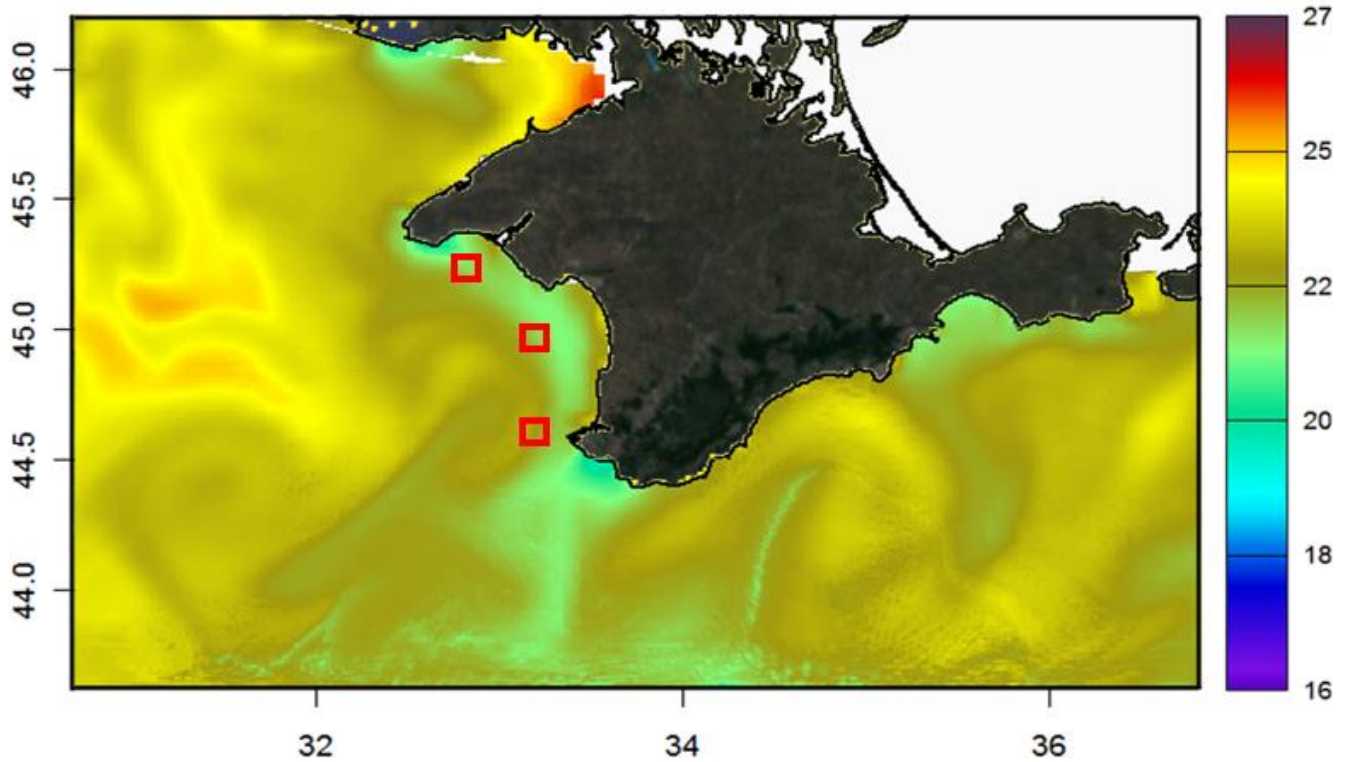


Рис.П.1. Нетиповые волновые возмущения скалярного поля температуры Черного моря

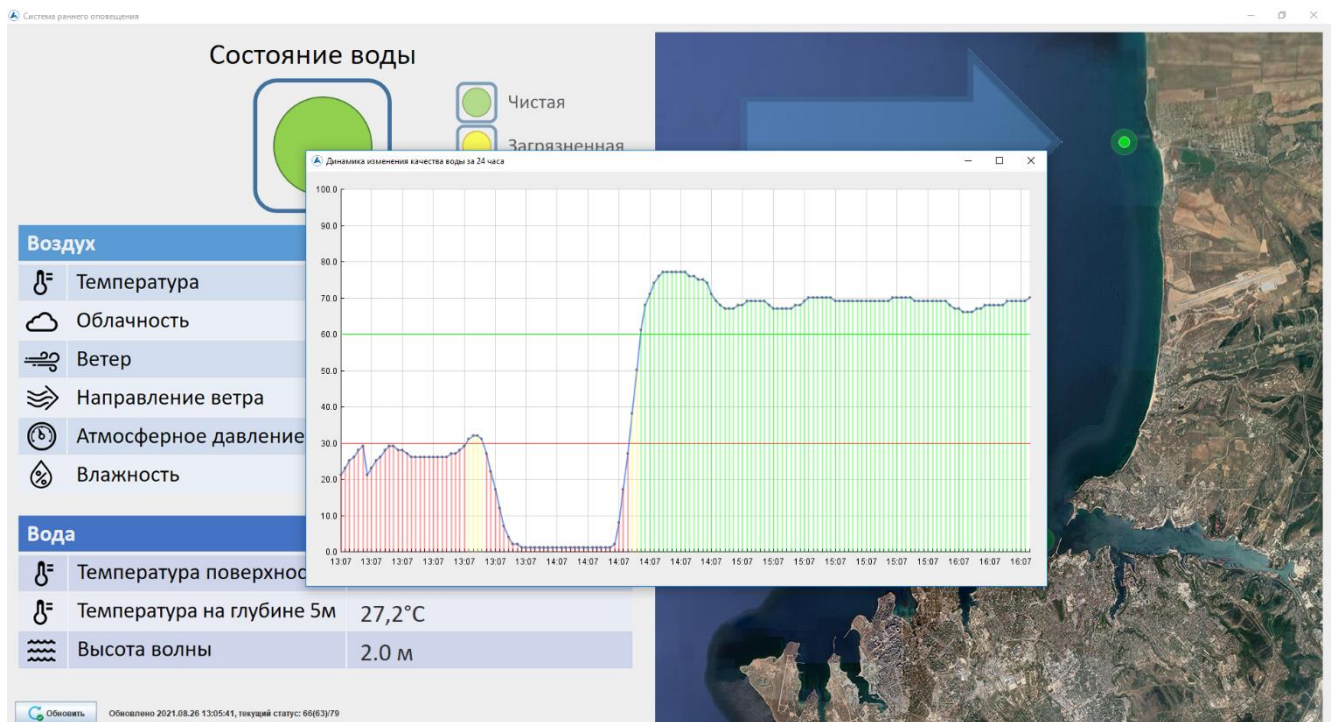


Рис.П.2. Аномалия реакции биосенсоров (угнетение створчатой активности моллюсков),
распределенная во времени

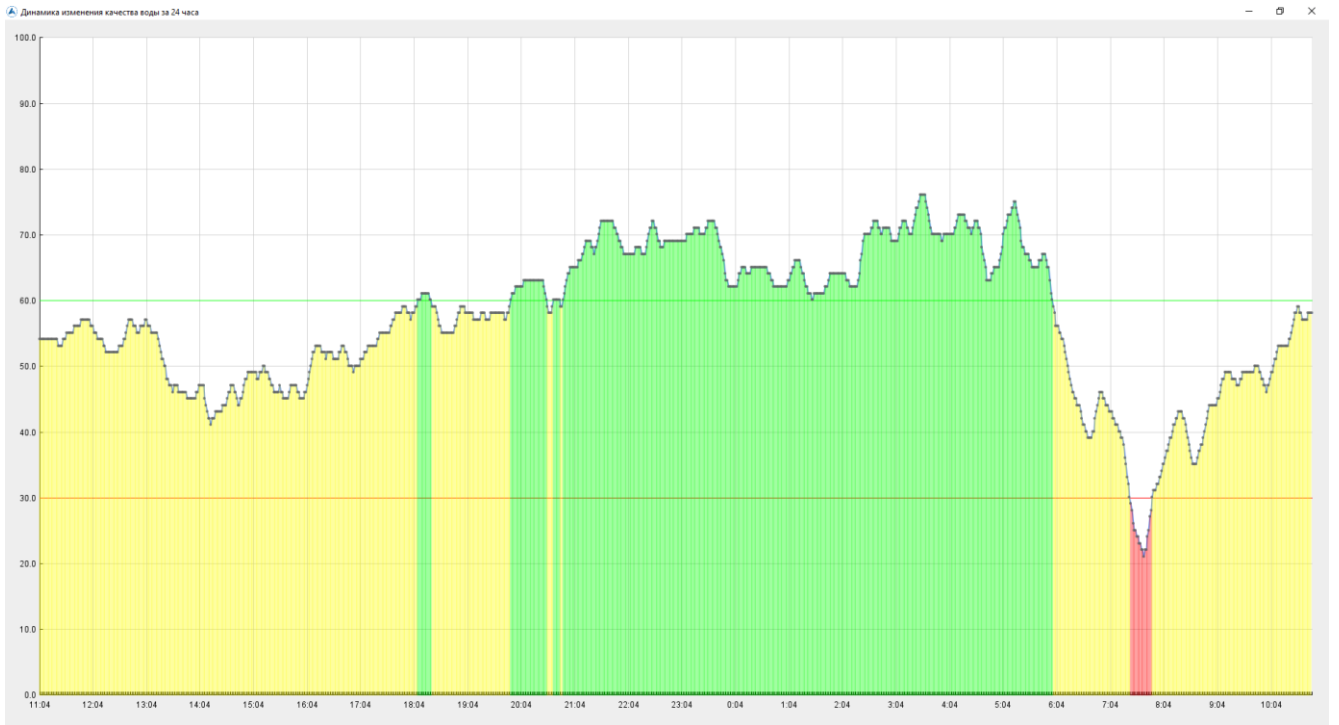


Рис. П.3. Аномалия реакции биосенсоров (угнетение створчатой активности совокупности моллюсков), относительно типовой суточной ритмики, рассчитанной за предыдущие периоды

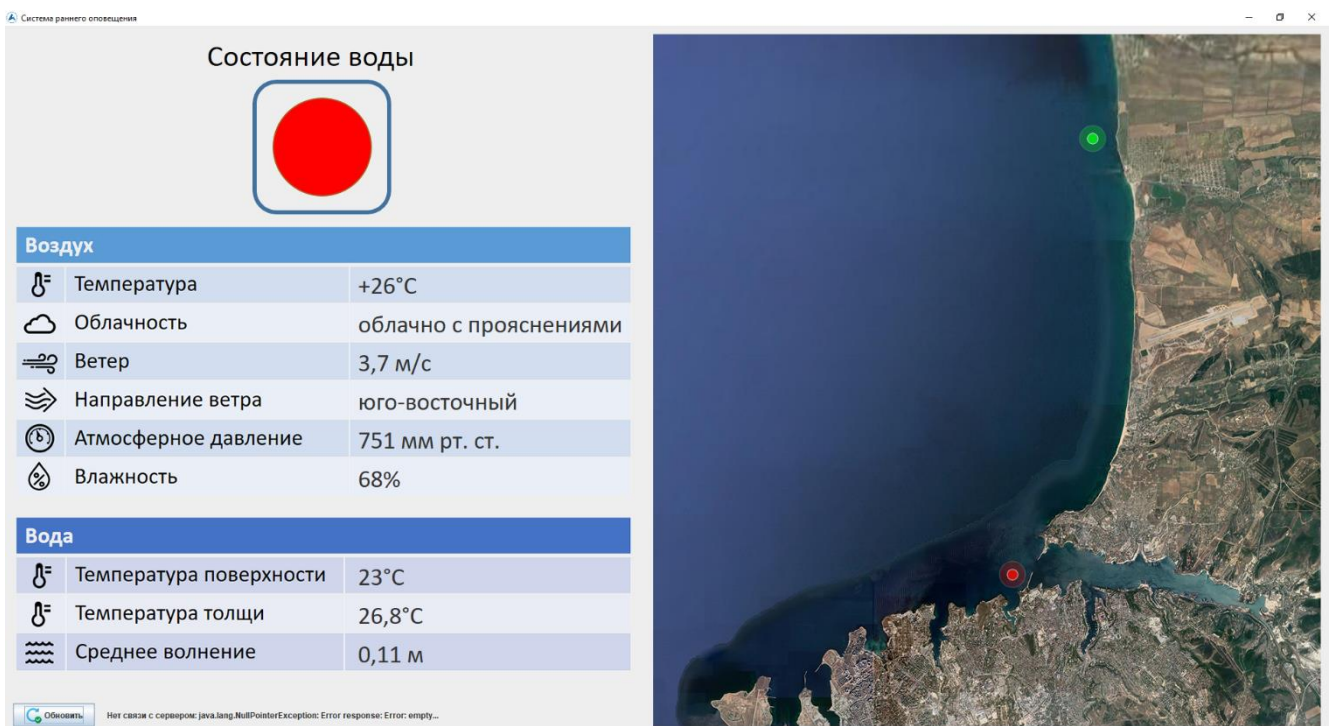


Рис. П.4. Аномалия реакции биосенсоров, вызванная экологическими условиями (искусственно созданное возмущение, в виде внесения глиняной взвеси в рабочий объем автоматизированного биосенсорного комплекса экологического мониторинга)

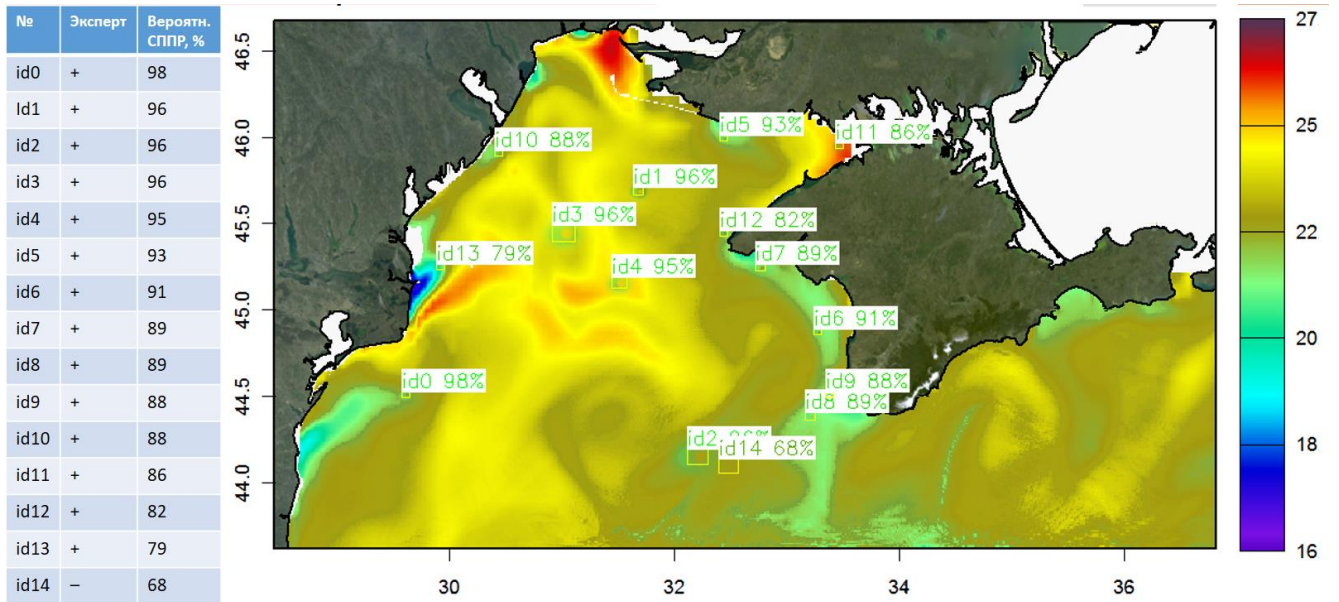


Рис. П.5. Типовые возмущения скалярного поля температуры, обнаруженные системой.

Параметрически заданные возмущения имеют волновую природу и сгенерированы в лабораторном стенде для оценки эффективности пула методов детектирования аномалий

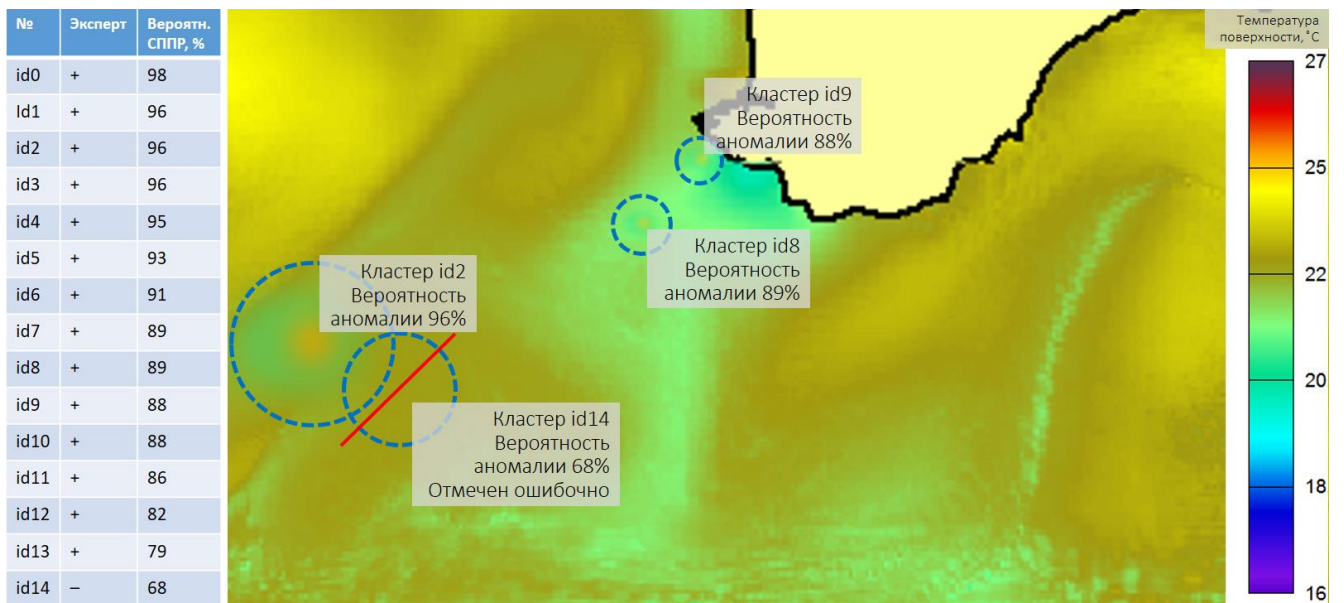


Рис. П.6. Ошибки первого и второго рода при детектировании аномальных возмущений скалярного поля температуры, отмеченные пользователем в графическом интерфейсе

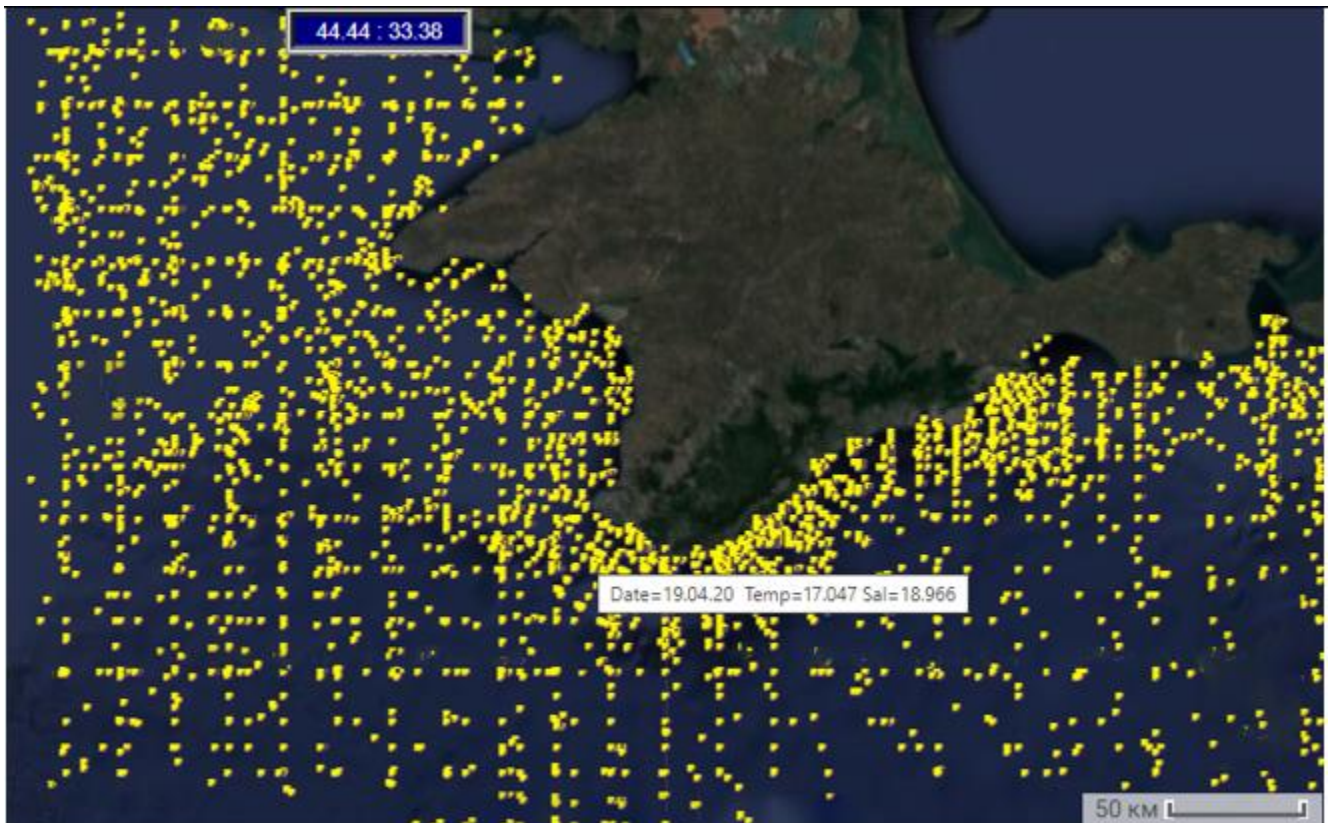


Рис. П.7. Аномальные точечные выбросы в скалярном поле температуры



Рис. П.8. Аномалии суточной ритмики отдельных биосенсоров

ПРИЛОЖЕНИЕ Б. ВИЗУАЛЬНОЕ ПРЕДСТАВЛЕНИЕ ИНТЕГРАЛЬНОГО ПОКАЗАТЕЛЯ СОСТОЯНИЯ ПРИРОДНОЙ СРЕДЫ И ОТДЕЛЬНЫХ ЕГО ПАРАМЕТРОВ ДЛЯ НАБОРА ДАННЫХ СОДЕРЖАЩЕГО N-АНОМАЛИИ



Рис. П.9. Визуальное представление интегрального показателя экологического состояния природной среды и отдельных его компонентов (содержащих аномалии)

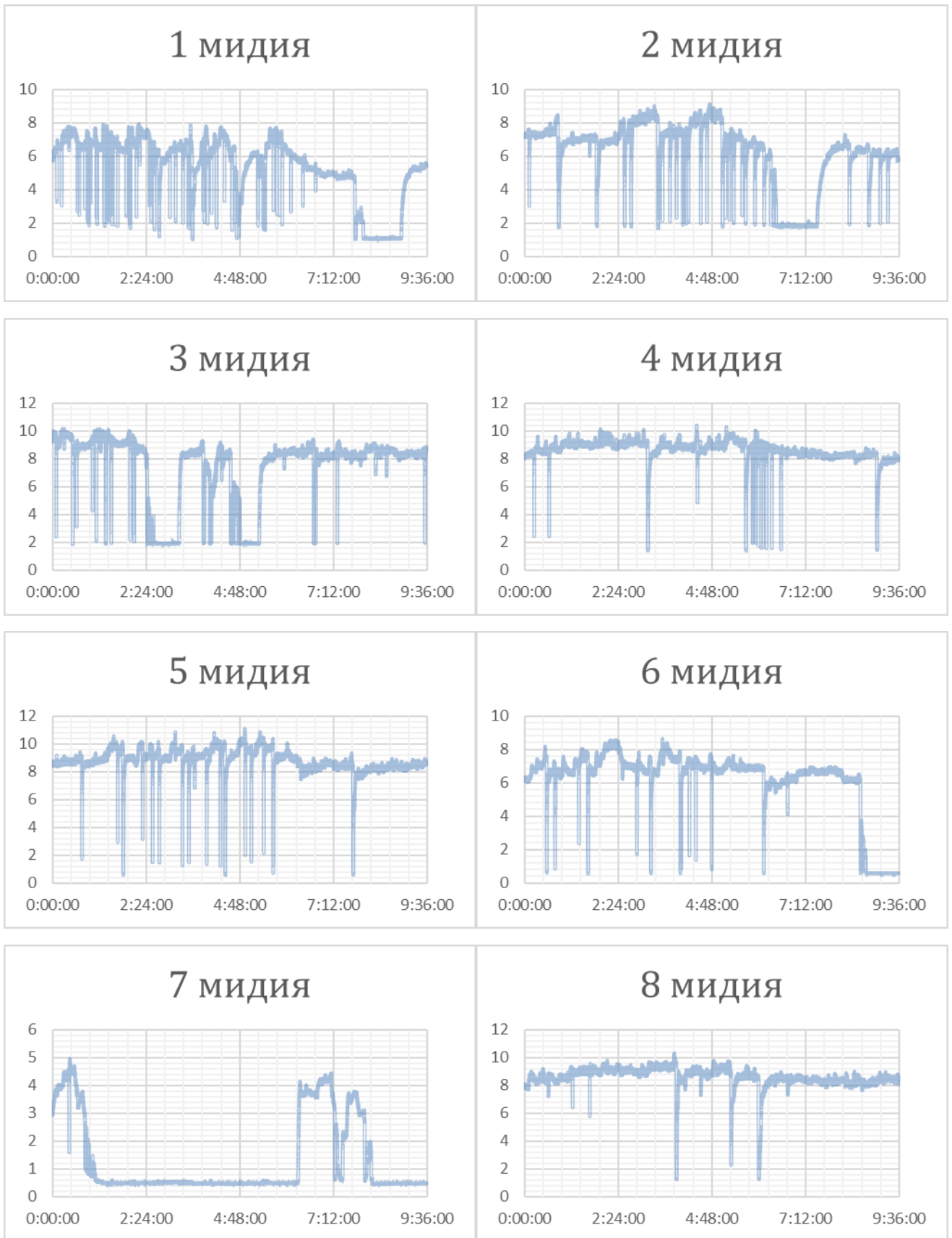


Рис. П.10. Визуальное представление интегрального показателя экологического состояния природной среды и отдельных его компонентов (содержащих аномалии)

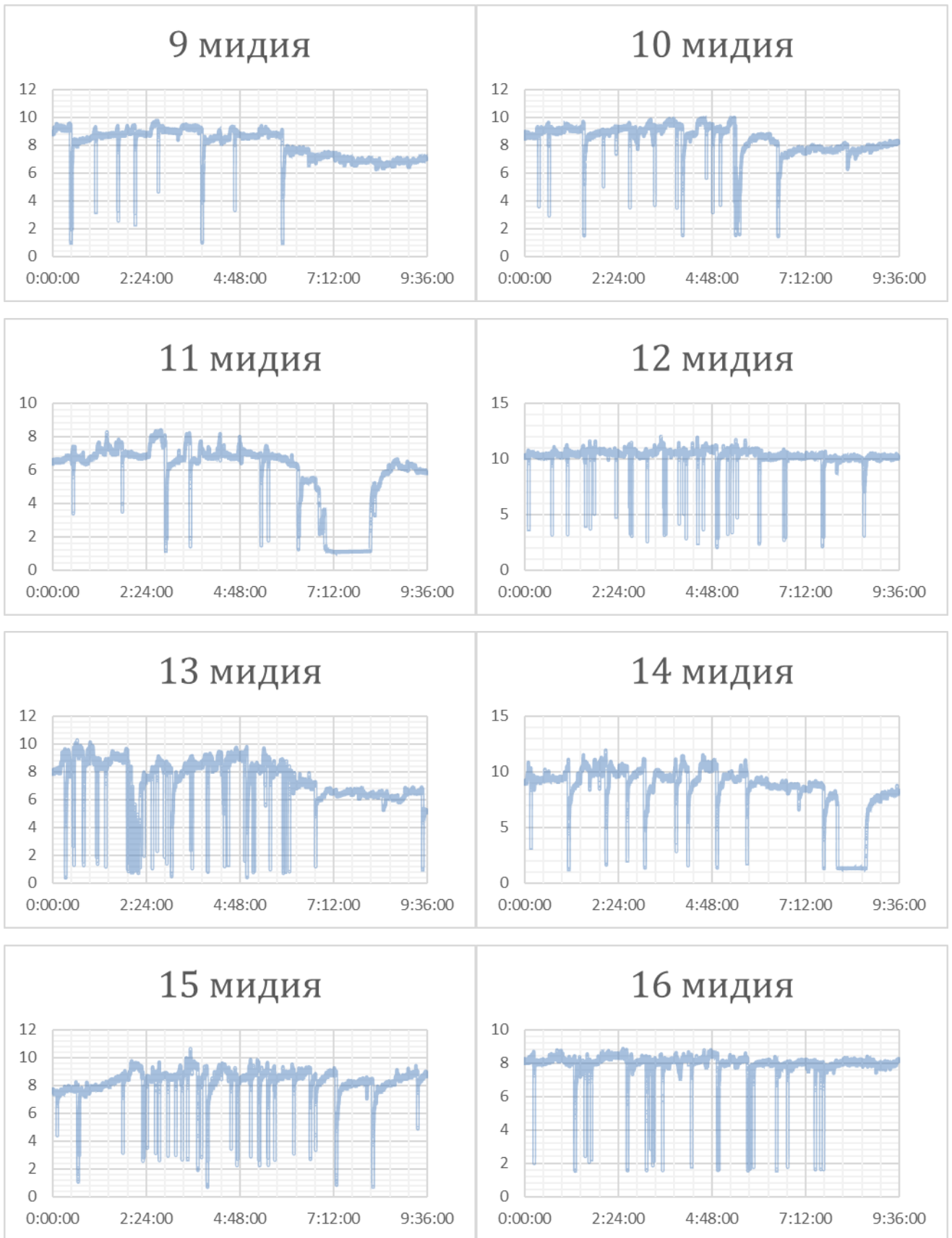


Рис. П.11. Визуальное представление интегрального показателя экологического состояния природной среды и отдельных его компонентов (содержащих аномалии)

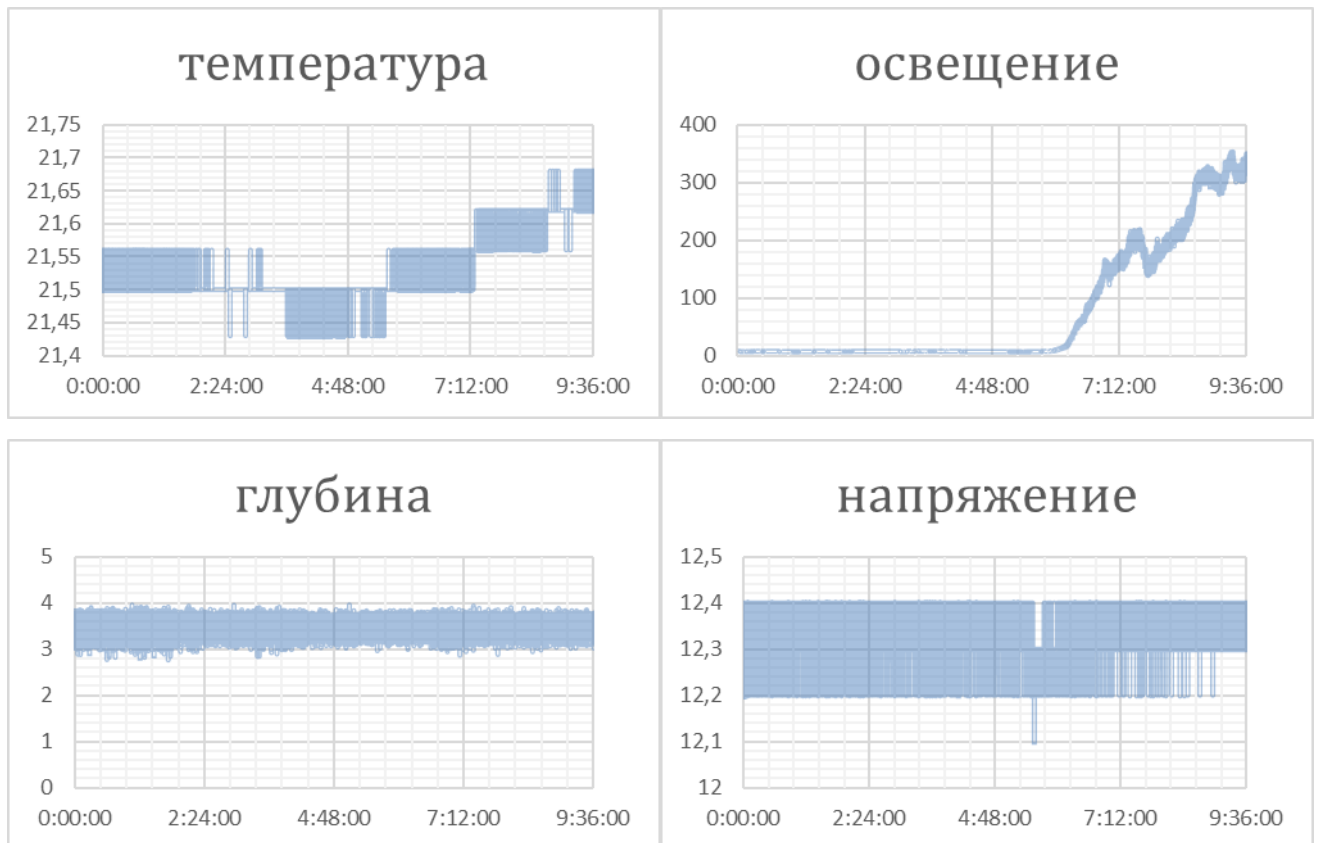
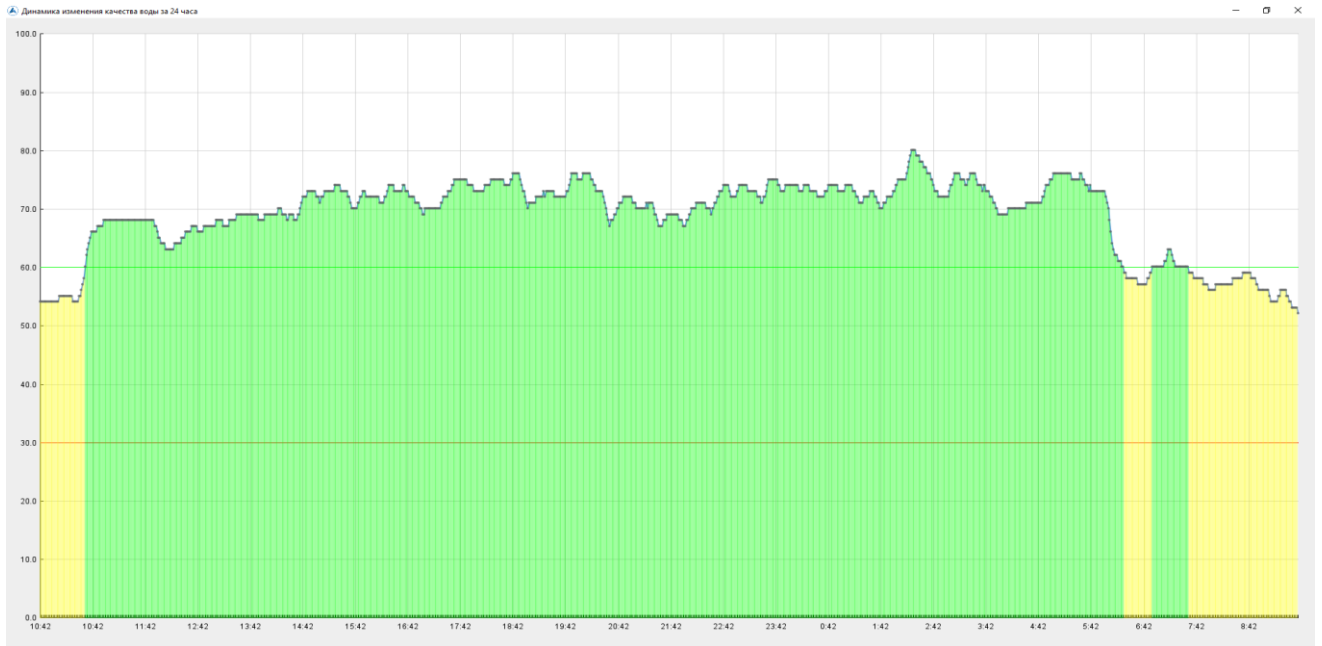


Рис. П.12. Визуальное представление интегрального показателя экологического состояния природной среды и отдельных его компонентов (не содержащих аномалии)



Рис. П.13. Визуальное представление интегрального показателя экологического состояния природной среды и отдельных его компонентов (не содержащих аномалии)



Рис. П.14. Визуальное представление интегрального показателя экологического состояния природной среды и отдельных его компонентов (не содержащих аномалии)

ПРИЛОЖЕНИЕ В. АРХИТЕКТУРА МОДУЛЬНО-КОМПОНЕНТНОЙ СППР ОБНАРУЖЕНИЯ М-АНОМАЛИЙ И N-АНОМАЛИЙ

На базе системного подхода с применением современных визуальных средств разработана и представлена в виде совокупности модулей комплексная интеллектуальная СППР обнаружения М-аномалий и N-аномалий (рис. 15, 16), базирующаяся на результатах рассмотрения и развития пула существующих методов и технологий и адаптации их к новым условиям применения, направленных на обнаружение аномалий в данных мониторинга, ориентированных на применение в системах контроля природной среды.

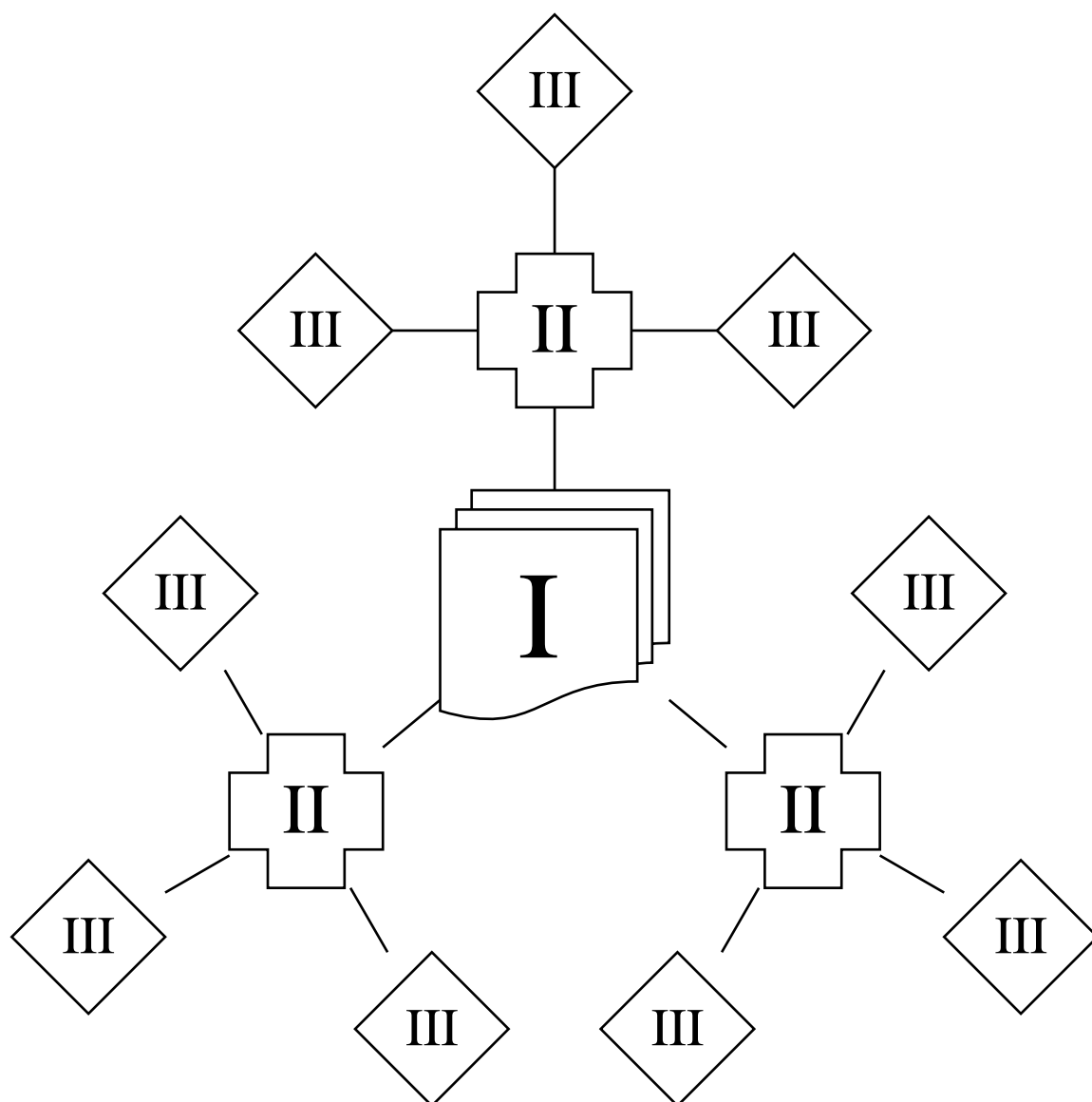


Рис. П.15. Структурная схема модульной интеллектуальной СППР обнаружения М-аномалий и N-аномалий, подсистемы I – III порядка

Декомпозиция осуществляется на подсистемы I – IV порядка, где система I порядка – комплексная информационная технология для обнаружения аномалий, а подсистема IV порядка – программа, модуль, совокупность алгоритмов, ИМ или множество сценариев.

На первом уровне решаются задачи согласования моделей интеллектуальных обучаемых систем, принимающих решения на основе априорной и экспертной информации, комплекса проблемно-ориентированных программ, осуществляющих статистическую обработку данных мониторинга и системы интеллектуализации, повышающей уровень обоснованности принятых решений за счет развитых средств визуализации.

На втором уровне осуществляется композиция различных сценариев использования, подходов и методов к решению конкретных задач, выявления границ применимости и областей наибольшей эффективности имеющегося в распоряжении пула методов с последующим автоматизированным выбором наилучшей решающей модели для каждого рассматриваемого случая.

Ключевым элементом интеллектуализации является ЛПР, который имеет высший первый уровень в иерархии, составляя триаду наравне с комплексной информационной технологией и системой сбора данных мониторинга. Именно ЛПР является той сущностью, которая определяет принципиально что есть аномалия и каким образом ее искать.

Синтез комплексной информационной технологии осуществляется снизу-вверх, где частные прикладные структуры и программные модули объединяются в подсистемы более высокого порядка.

Сущности подсистем низкого порядка в совокупности образуют пул подходов и методов, реализованных в программных подсистемах, направленных на решения задачи обнаружения аномалий в различных начальных условиях: малого объема данных, отсутствия обучающей выборки, многокритериального оценивания данных. На более высоких уровнях выбирается одна модель, наилучшим образом удовлетворяющая условиям ее использования.

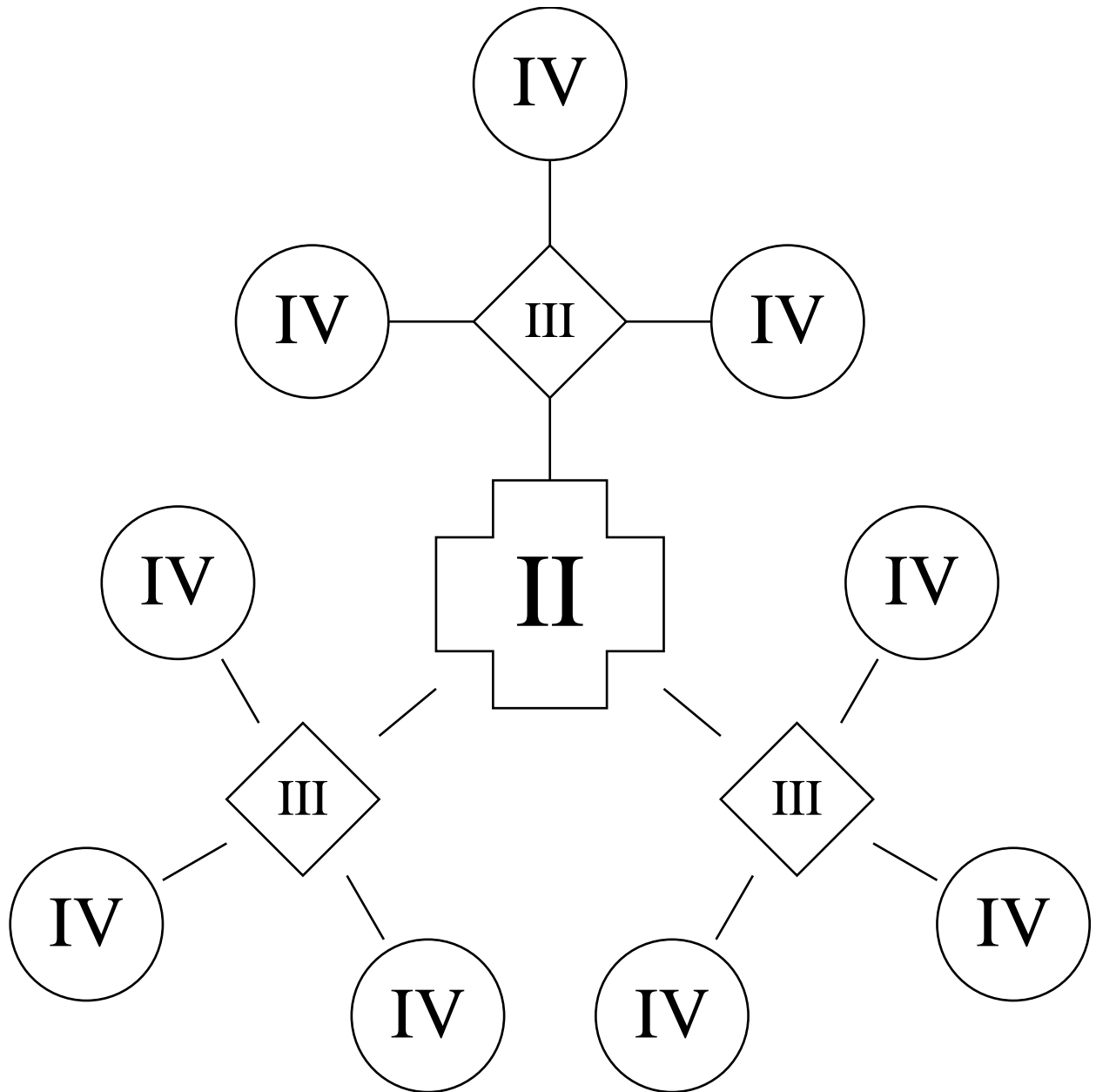


Рис. 16. Структурная схема модульной интеллектуальной СППР обнаружения М-аномалий и N-аномалий, подсистемы I – IV порядка

Каждый из компонентов подсистем, в соответствии с системным подходом, может быть подвергнут декомпозиции на еще более мелкие подсистемы, а итоговый полимодельный комплекс представляет собой единую эмерджентную структуру, ориентированную и подчиненную решению единой задачи обнаружения аномалий в данных мониторинга.

Другой принципиальной возможностью такой системы является ее гибкость и расширяемость, например, у ЛПР на каждом уровне есть возможность добавить

дополнительный решающий модуль, который наилучшим образом решает поставленную задачу.

Структура подсистемы I уровня – комплексная информационная технология для обнаружения аномалий

На основе комплекса моделей интеллектуальных систем обнаружения аномалий в данных мониторинговых наблюдений естественных сред, комплекса проблемно-ориентированных программ обнаружения аномалий в статистических данных и системы интеллектуализации на основе визуализации данных осуществим синтез единой подсистемы I порядка, которая объединяет модели, программы и интеллектуальную систему в единый комплекс и предназначена для обоснованного обнаружения аномалий в данных мониторинговых наблюдений.

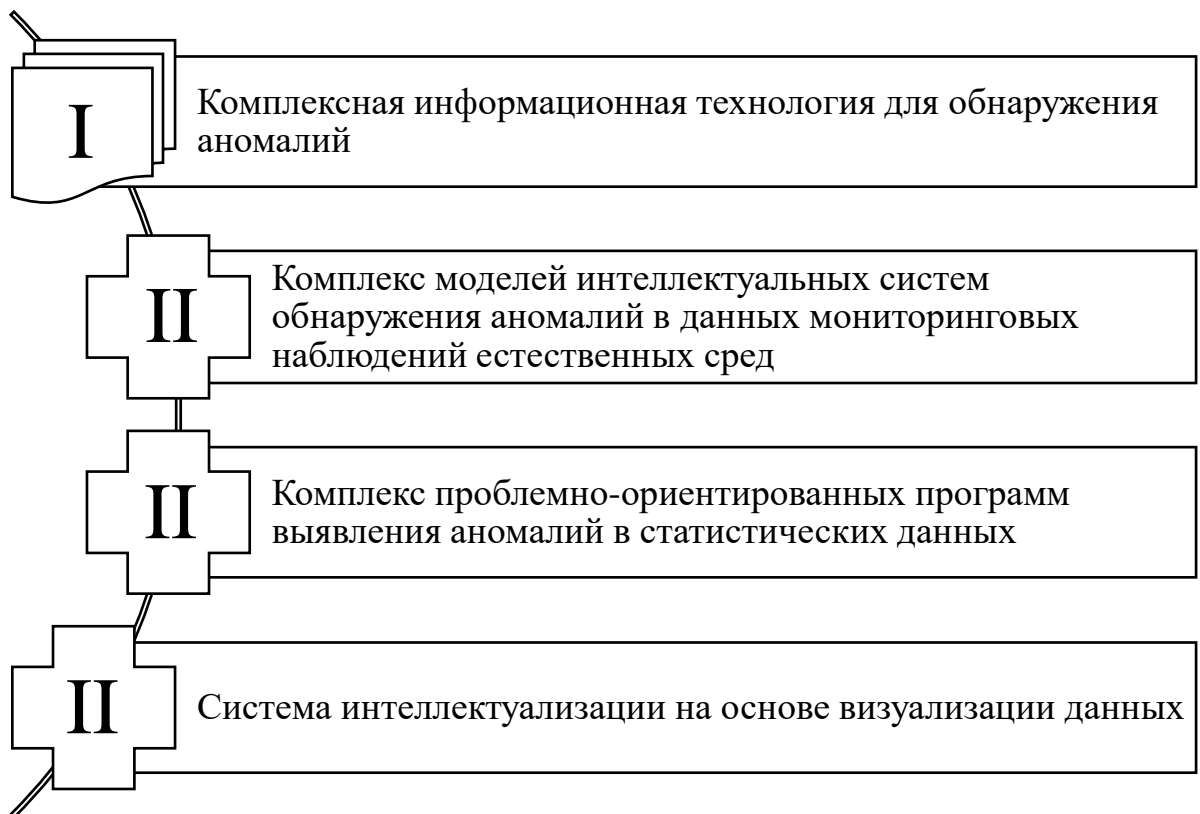


Рис. П.17. Структурная схема комплексной системы для обнаружения аномалий

Представленная комплексная информационная технология, как нам представляется, является новым инструментальным средством поднятия качества функционирования мониторинговых систем за счет использования в них интеллектуальных информационных комплексов, которые реализуют элементы интеллектуального подхода к адаптивному обучению процессов распознавания, в

ней используется развитая методология моделирования на основе применения сценарного подхода, который включает в себя сценарии: минимизации рисков, прямого голосования, минимаксный с относительными оценками. Комплексное использование сценарного подхода позволяет повысить степень обоснованности ЛПР и обеспечить структурную гибкость.


Представим вариант использования комплексной информационной технологии на рисунке 18, с основополагающими компонентами: РСД – распределенная система датчиков (включает в себя сенсоры, СПИ и спутниковые наблюдения), ЛПР – лицо, принимающее решения (оператор), КИТ – комплексная информационная технология (представленная в главе V),  – мобильная сеть передачи данных (связующее звено всех компонентов).



Рис. П.18. Структурная схема варианта организации комплексной информационной технологии (информационная система в целом)

Разработанная система обладает функциональной полнотой: список задач, возникающих при обнаружении аномалий в данных мониторинга природной среды полностью покрывается списком методов, технологий и проблемно-ориентированных программ, рассмотренных в главах диссертации.

Система является ориентированной на обнаружение данных в естественных средах, представлены конкретные ее приложения для решения задач обнаружения аномалий данных мониторинга полей наблюдений в акватории прибрежной и шельфовой зон бухты г. Севастополя. Применяемые методы вносят вклад в решение задач повышения достоверности при обнаружении аномалий, используются интеллектуальные методы распознавания информационных ситуаций, в проблемно-ориентированных программах осуществлена реализация интерактивных режимов для подключения человеческого интеллекта, выполнено повышение уровня интеллектуализации, получили развитие методы работы с данными мониторинга в форме слабоструктурированных данных.

Структура подсистемы II уровня – комплекс моделей интеллектуальных систем обнаружения аномалий в данных мониторинговых наблюдений естественных сред

Основой интеллектуализации является возможность обучения и адаптации системы к существующим и изменяющимся условиям использования. Повышение степени вовлеченности ЛПР в процесс принятия решения осуществляется на основе сценарного подхода, в рамках которого предоставляется возможность выбора приоритетов и допустимых рисков, которые определяют набор целевых характеристик системы обнаружения аномалий.

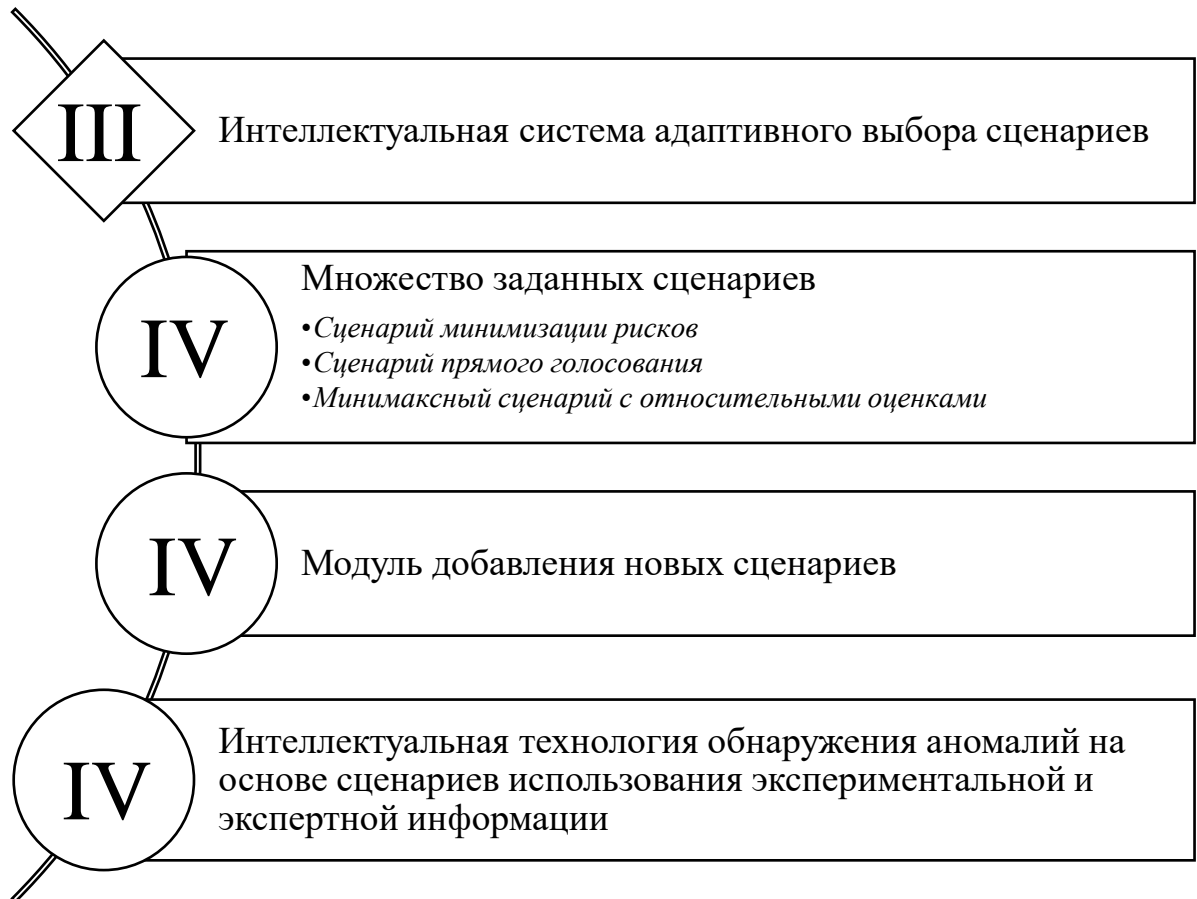


Рис. П.19. Структурная схема интеллектуальной системы адаптивного выбора сценариев

Совокупность сценариев представляет собой интеллектуальный подход, позволяющий ЛПР выбирать не непосредственно метод решения задачи, а совокупность целевых критериев. Например, сценарий минимизации рисков относится к минимаксным стратегиям, которые позволяют снизить число ошибок I или II рода в зависимости от потерь каждой из них, в то время как сценарий прямого голосования позволяет получить усредненную оценку от нескольких моделей что обеспечивает большую универсальность полученной модели принятия решений при меньшей ее специфичности.

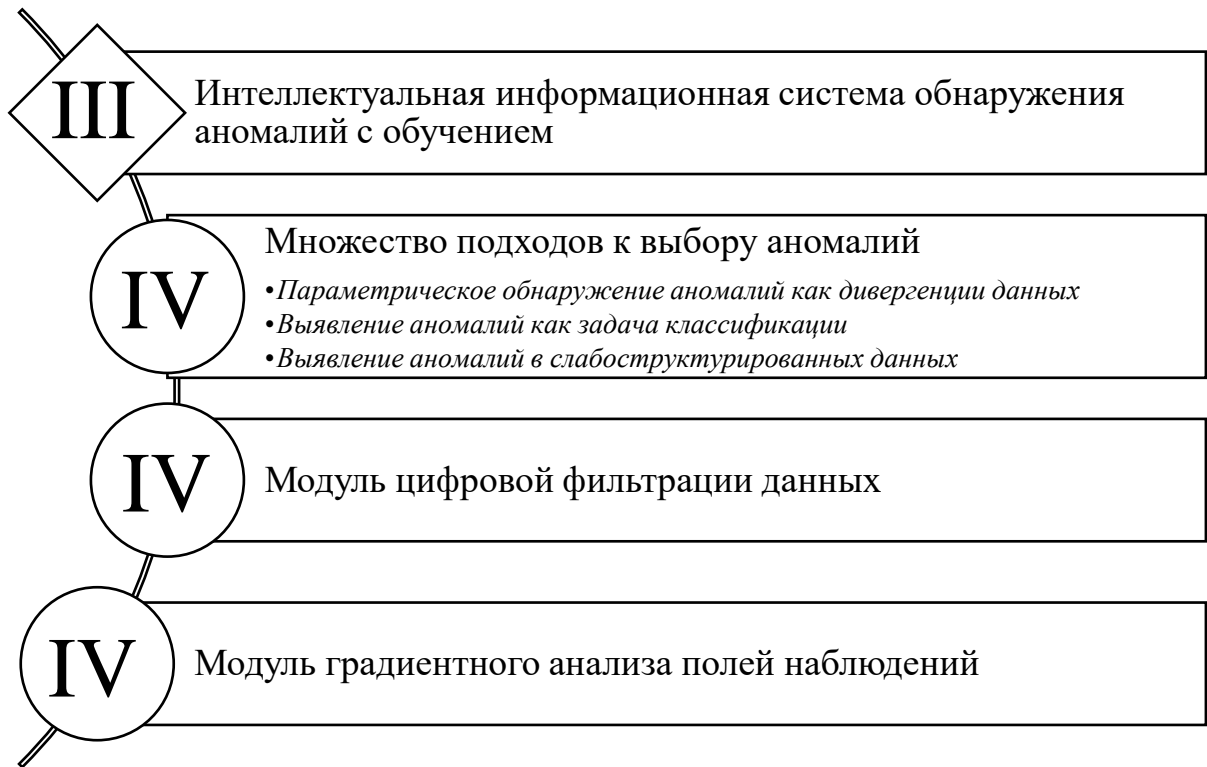


Рис. П.20. Структурная схема интеллектуальной информационной системы обнаружения аномалий с обучением

Использование пула подходов позволяет выбирать способ решения задачи обнаружения аномалий с учетом природы этих аномалий. Аномалия может рассматриваться как кластер данных мониторинга минимального размера, новизна в поведении наблюдаемой системы или дивергенция данных.

Модуль градиентного анализа используется в случае, когда данные заданы для полей наблюдений и имеют внутреннюю автокорреляцию. В таком случае для обнаружения аномалий эффективными являются визуальные средства отображения данных мониторинга и сравнения поверхностей, заданных данными наблюдений с поверхностями, заданными их математическими моделями. Мера расстояния таких поверхностей является основой для обнаружения аномалий в данных заданных указанным образом.

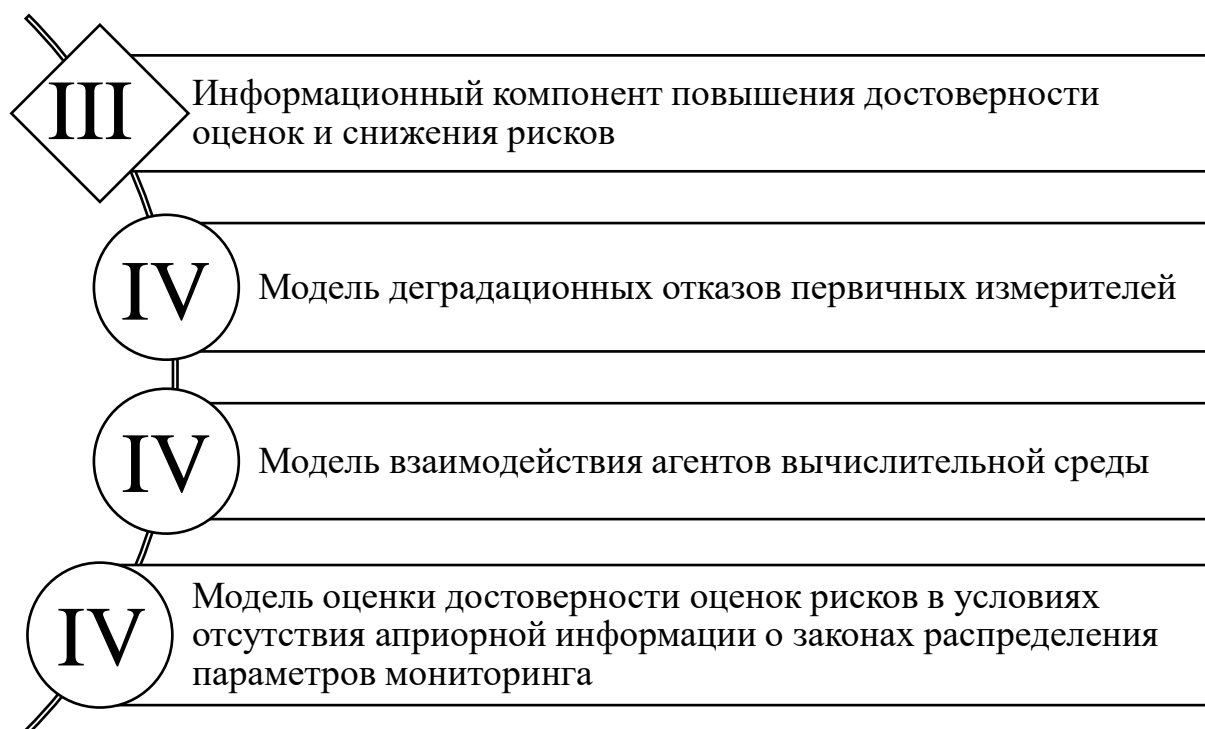


Рис. П.21. Структурная схема информационных компонент повышения достоверности оценок и снижения рисков

На основе интеллектуальной системы адаптивного выбора сценариев, интеллектуальной информационной системы обнаружения аномалий с обучением и информационного компонента повышения достоверности оценок и снижения рисков осуществим синтез подсистемы II порядка в виде комплекса моделей интеллектуальных систем обнаружения аномалий, направленного на повышение степени интеллектуализации при принятии решений за счет совместного использования экспертных и экспериментальных данных.

Основой интеллектуализации для информационной системы является процесс ее обучения. При обучении решаются задачи выбора необходимости дообучения и переобучения в том случае, когда обучающая выборка значительно отличается от ранее поступавших данных.

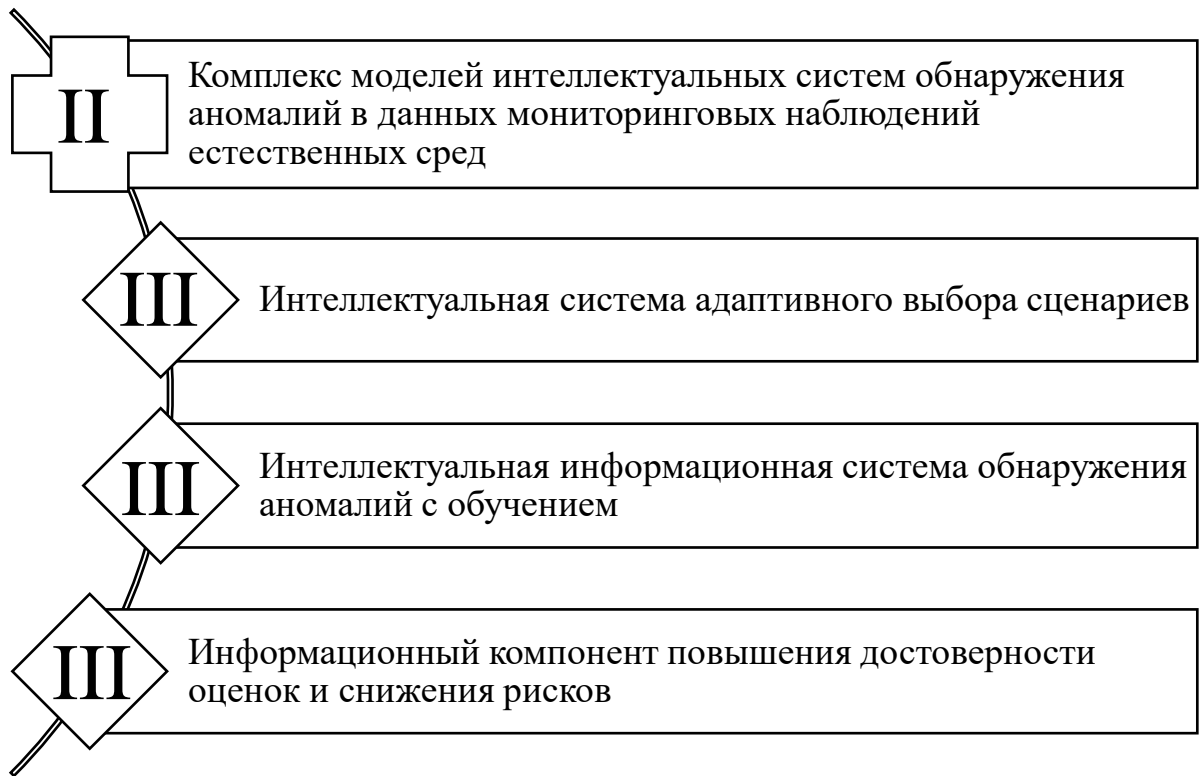


Рис. П.22. Структурная схема комплекса моделей интеллектуальных систем обнаружения аномалий в данных мониторинговых наблюдений природных сред

Снижение рисков достигается за счет активного вовлечения ЛПР в процесс принятия решений и его экспертной оценки критичности возникновения ошибок I и II рода для каждого интересующего его параметра. Комплексный учет потерь, вызванных пропуском аномалии или ложным срабатыванием позволяет оценить эффективность каждого имеющегося в распоряжении метода и в соответствии с полученными оценками выбрать один или несколько оптимальных методов. Принципиальной особенностью обнаружения аномалий именно в естественных средах является в том числе снижение вероятности пропусков аномалий, так как они могут служить предикторами значимых изменений в природной системе в целом.

Структура подсистемы II уровня – комплекс проблемно-ориентированных программ обнаружения аномалий в статистических данных

Модели оптимизации процесса мониторинга включают в себя модели скалярной и векторной оптимизации. Ключевой особенностью моделей

оптимизации является их многоцелевой характер, и в отличие от формально идентичных задач принятия решений в других областях приложений, здесь целевые функции задаются непосредственно исследователем, и их важность как правило не одинаковая. На практике, кроме целевых функций, характеризующих степень аномальности данных мониторинга с разных, обычно противоречивых сторон, у исследователя имеется представление о глобальном качестве различных наборов данных.

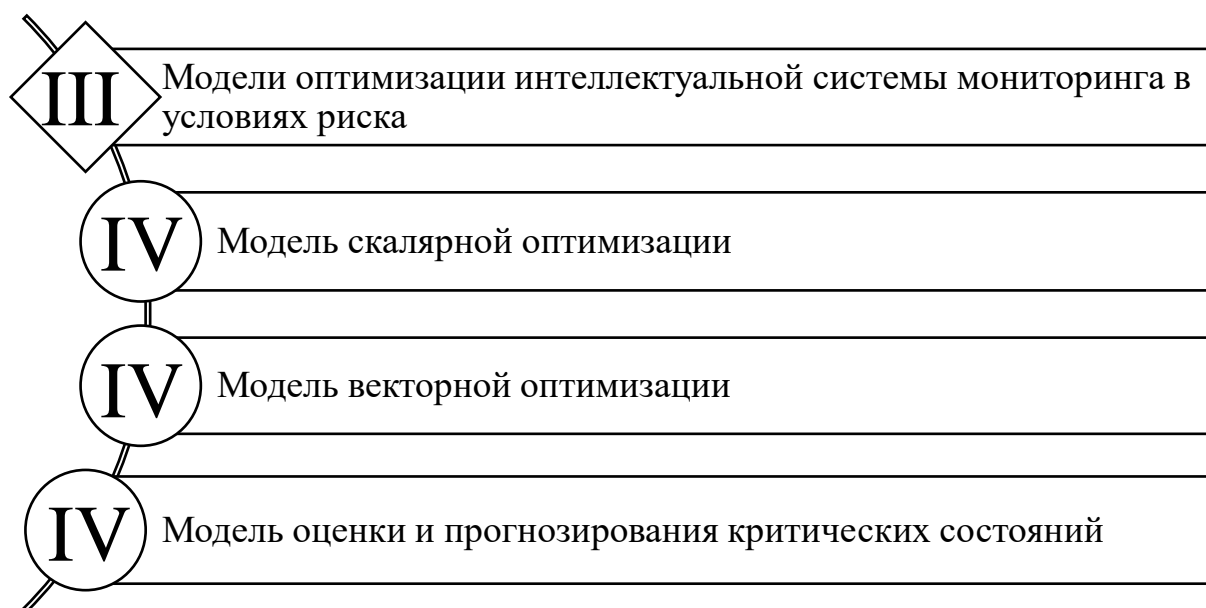


Рис. П.23. Структурная схема модели оптимизации интеллектуальной системы мониторинга в условиях риска

Однако представление об аномалии изменяется в зависимости от конкретного приложения, и трудно выделить какие-либо классы, где оно было бы единым и формализуемым. Вместе с тем, хотя представление о совокупном оптимуме порой весьма расплывчато, ценой дополнительных издержек (или риска) можно осуществить сравнение детектирующих свойств различных моделей и определить, какой из них более приемлем для принимающего решение.

Суть мобильных интеллектуальных систем подразумевает использование детектирующей программы как облачного сервиса. Необходимость использования внешних вычислительных и информационных ресурсов обусловлена порой

принимать решения в условиях, когда у исследователя нет доступа к производительной вычислительной системе, например, во время сбора данных мониторинга.

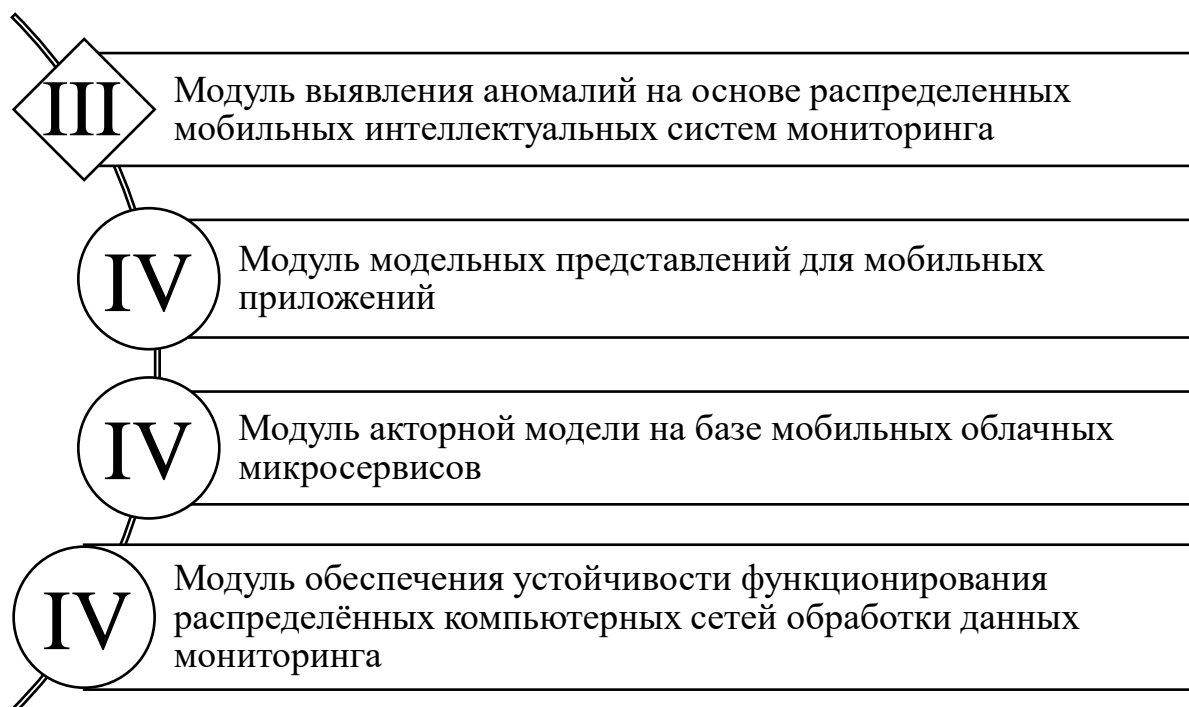


Рис. П.24. Структурная схема модуля обнаружения аномалий, на основе распределенных мобильных интеллектуальных систем мониторинга

ИМ позволяют осуществлять принятие решений о наличии аномалий в условиях малого объема данных и частичной неопределенности. В этом случае ИМ исследуемого объекта выполняется параллельно с осуществлением процесса мониторинга и калибруется при наличии значимых отклонений от реального объекта. Мера отклонения показателей ИМ от реального прототипа может являться численной оценкой вероятности возникновения аномалии в данных. Далее система определяет тип аномалии – единичная или из последовательности нескольких значений, возможно аномалия вызвана изменением состояния объекта мониторинга. Конечное решение по полученным численным оценкам и исходя из собственного опыта принимает ЛПР.

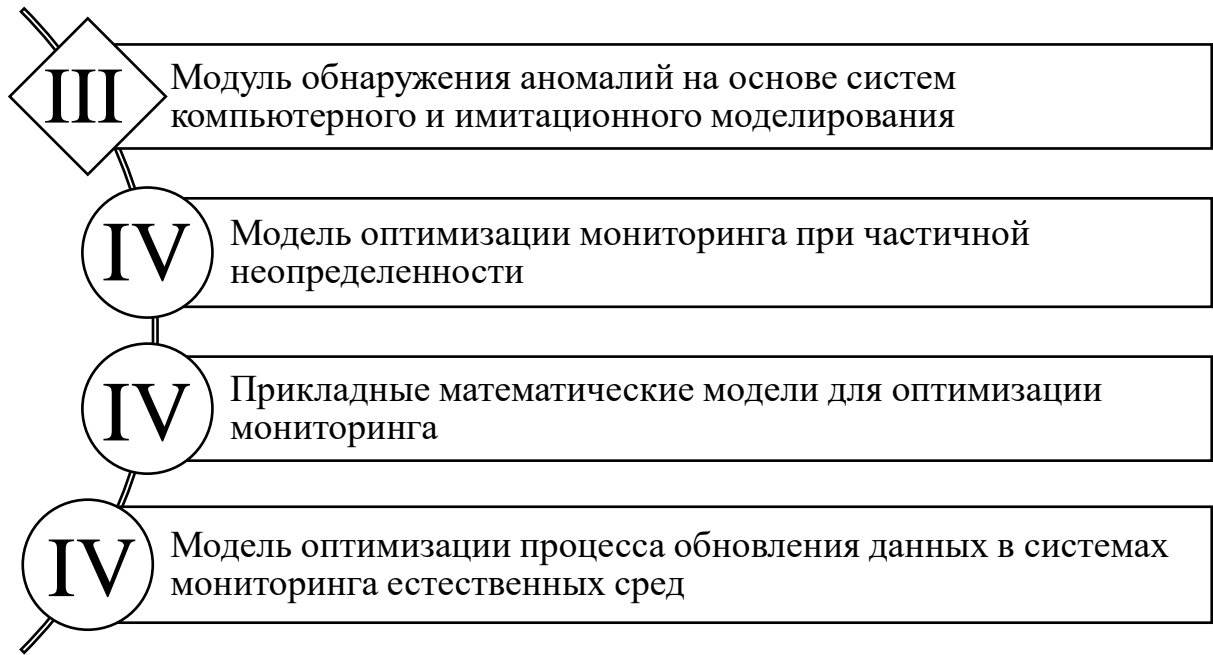


Рис. П.25. Структурная схема модуля обнаружения аномалий на основе систем компьютерного и имитационного моделирования

В результате объединения подсистем III уровня образуется комплекс проблемно-ориентированных программ обнаружения аномалий, который в зависимости от характера имеющихся данных мониторинга, наличия и объема обучающей выборки применяет соответствующую модель на основе набора статистических метрик или предметных проблемно-ориентированных ИМ.

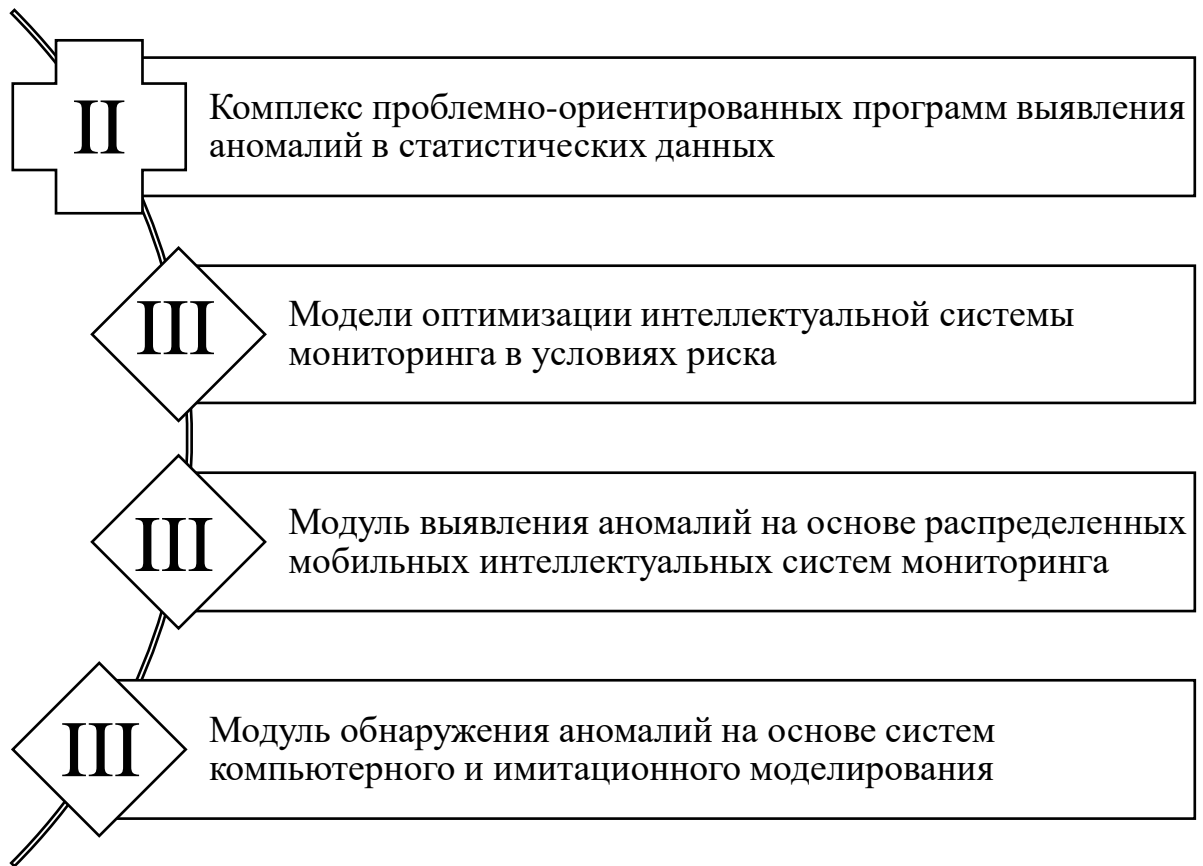


Рис. П.26. Структурная схема комплекса проблемно-ориентированных программ обнаружения аномалий в статистических данных

Учет рисков осуществляется путем последовательного перемножения вероятностей ошибочной классификации на коэффициенты потерь, которые могут возникнуть в случае ошибочной классификации данных как аномальных или наоборот, пропуска аномальных данных.

Структура подсистемы II уровня – система интеллектуализации на основе визуализации данных

Используется интеллектуальная система, осуществляющая самообучение на базе ряда ИМ и лабораторного стенда так как нельзя заранее определить специфику данных обучающей выборки, ее объем и качество. В зависимости от характера исходных данных осуществляется выбор модели, наилучшим образом предназначенной для имеющихся данных. В зависимости от природы аномалий могут возникать точечные всплески в данных, которые устраняются локальными

сверточными фильтрами, аномалии в полях данных, характеризующиеся появлением аддитивной волновой функции в данных и выявляемые сравнением функции распределения наблюдаемых данных с эталонными.

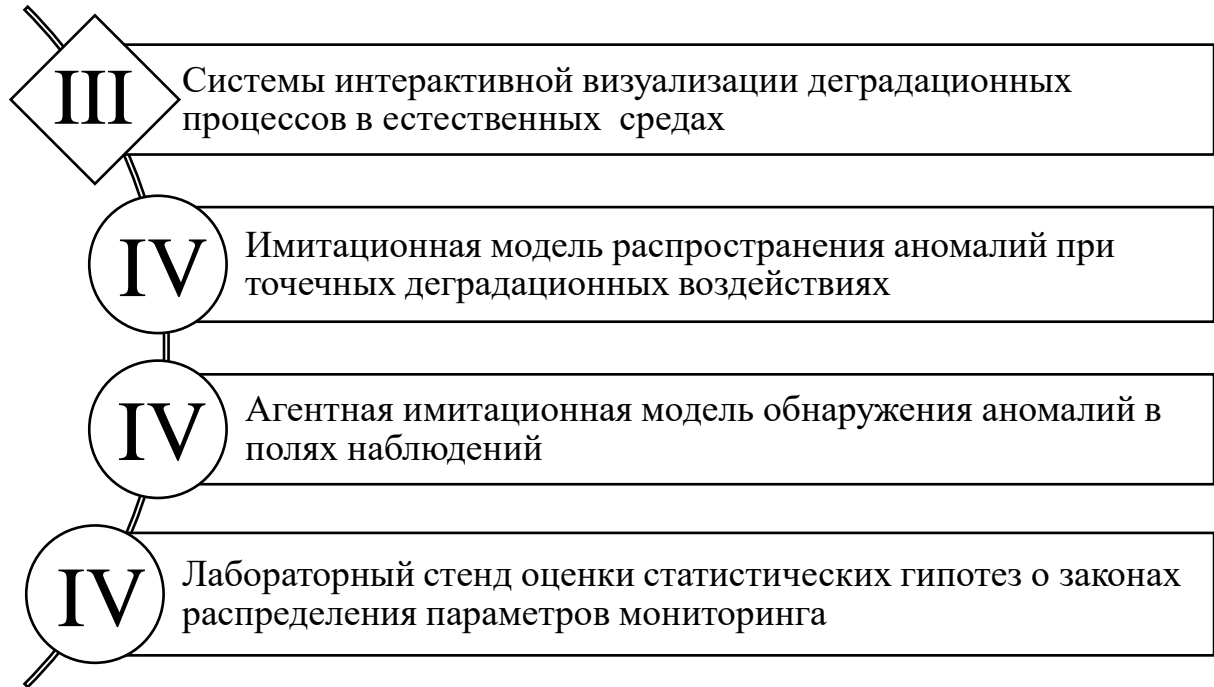


Рис. П.27. Структурная схема системы визуализации деградационных процессов в естественных средах

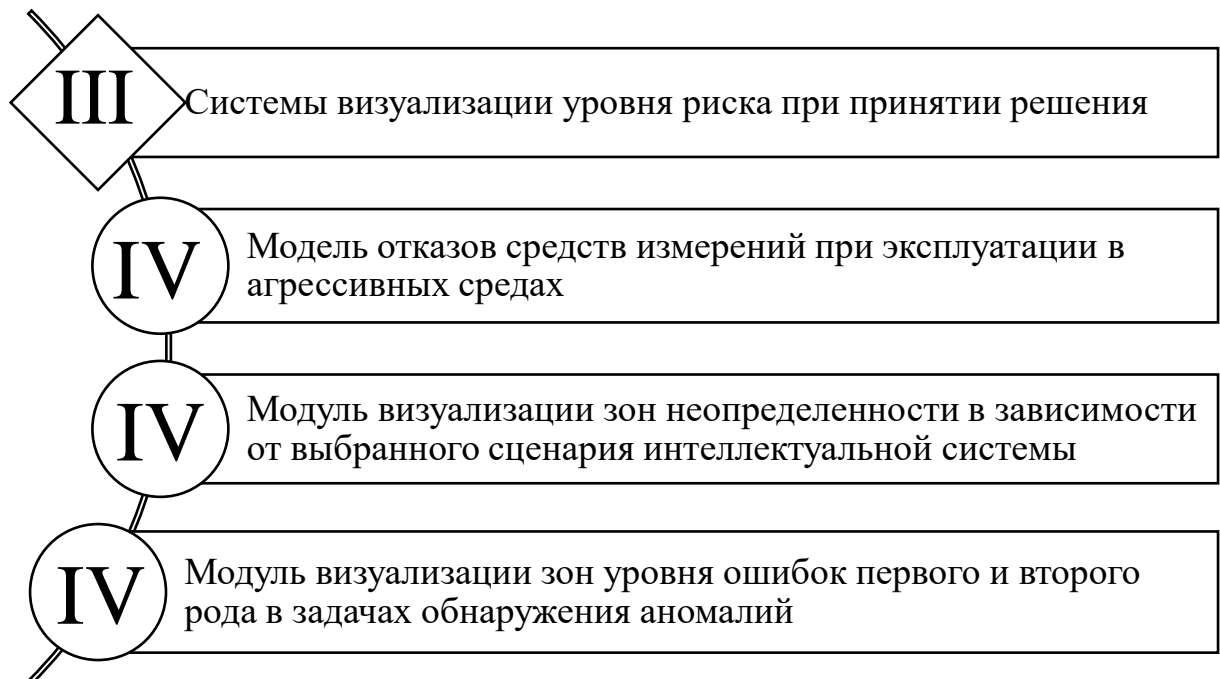


Рис. П.28. Структурная схема системы визуализации уровня риска при принятии решения

Визуализация рисков подразумевает активное вовлечение ЛПР в выбор приоритетов при обнаружении аномалий, в частности учет возможных потерь при пропуске аномалий и последствия в случае ошибочной реакции на неаномальные события. Визуализация зон риска позволяет ЛПР иметь более полную картину о качестве классификатора в условиях имеющихся данных и принимать решения с учетом степени уверенности системы классификатора. Сценарии интеллектуальной системы позволяют заранее выбрать приоритет системы на более оперативное принятие решений, повышая реактивность системы или сбор дополнительной статистики для повторной проверки гипотезы о возникновении аномалии в данных. Кроме того, различные типы возникающих аномалий требуют различного объема данных для принятия решений и используют различные метрики для их обнаружения.

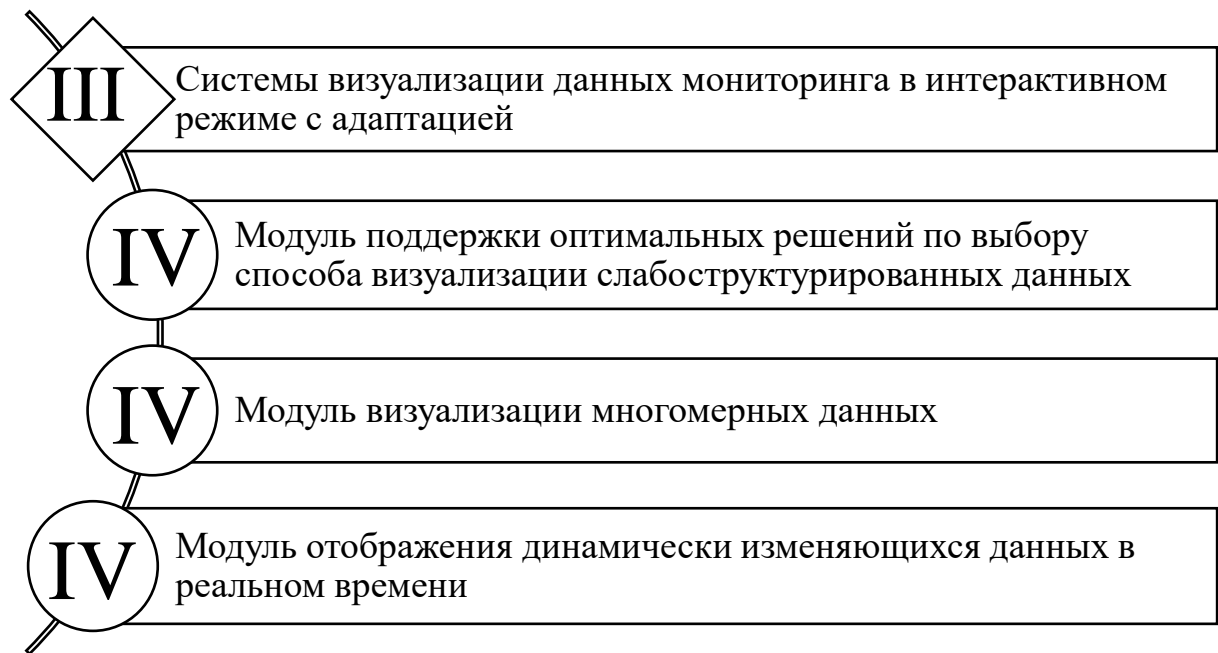


Рис. П.29. Структурная схема системы визуализации данных мониторинга в интерактивном режиме с адаптацией

Система визуализации, в рамках решаемой задачи, осуществляет повышение уровня обоснованности принимаемых решений за счет вовлечения ЛПР в процесс обучения. Специфика человеческого восприятия состоит в том, что очень сложно обнаруживать аномалии в многопараметрических данных с размерностью более

трех измерений. Как правило многомерные данные представляются в виде таблиц и не имеют выразительной графической составляющей. Суть применяемых систем визуализации состоит в снижении размерности поступающих данных, путем поиска наиболее существенных параметров, применении легких алгоритмов, которые могут быть использованы в реальном времени и в анализе структуры имеющихся данных с целью поиска наиболее оптимальной формы их представления и оценки динамики их изменения во времени.

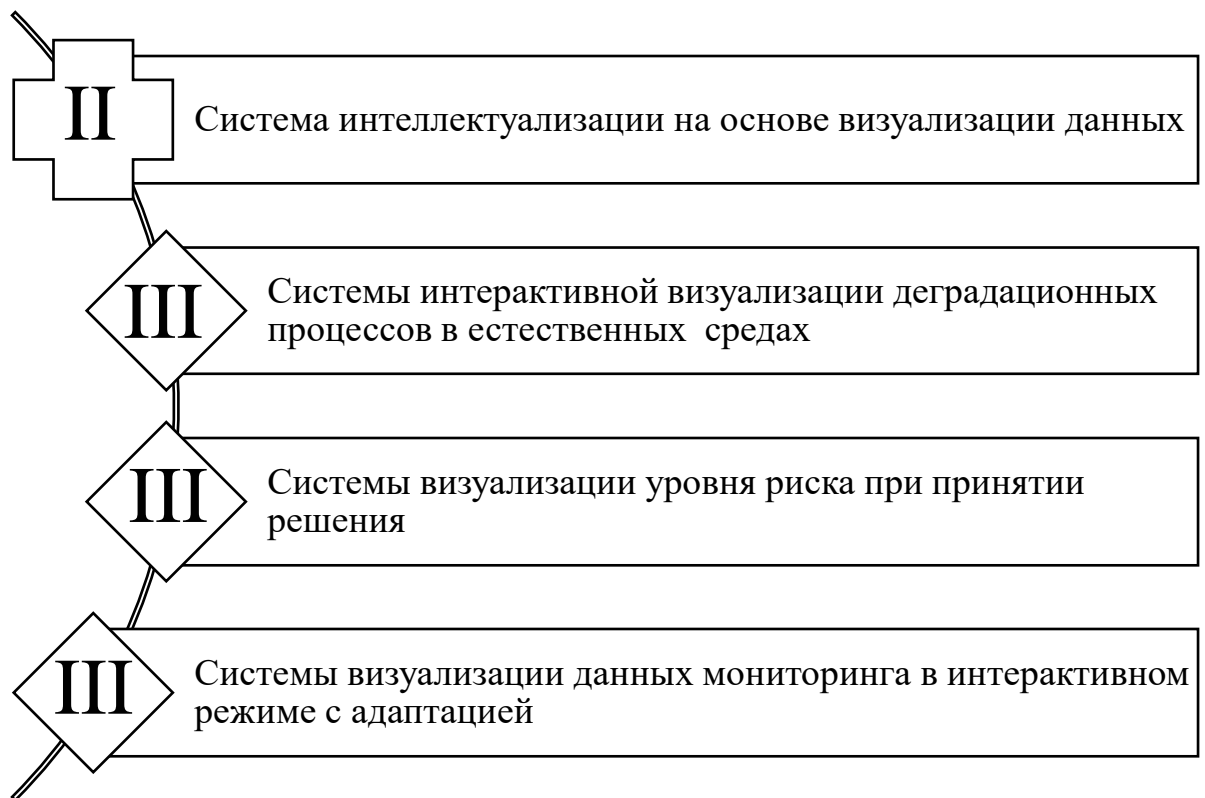


Рис. П.30. Структурная схема системы интеллектуализации на основе визуализации данных

Представленная архитектура модульной интеллектуальной СППР обнаружения М-аномалий и N-аномалий является гибкой с адаптивно перенастраиваемой структурой в зависимости от природы обнаруживаемых аномалий. Учет представленных факторов в совокупности приводит к увеличению реактивности системы и увеличению доступности системы при работе в режиме реального времени.

ПРИЛОЖЕНИЕ Г. СИНТЕЗ ПОНЯТИЯ АНОМАЛИИ В РЕЗУЛЬТАТАХ МОНИТОРИНГА ПРИРОДНОЙ СРЕДЫ

Само понятие аномалии имеет достаточно большое число понятий в литературе. Специфика этих понятий зависит, в том числе, от области применения данного термина. Некоторые их наиболее значимых и общепринятых понятий, которые были даны в различных справочниках и энциклопедиях, сгруппированы по степени схожести в контексте А-задачи и приведены ниже.

В энциклопедии систем баз данных дается понятие: «Аномалия (данных), редкие данные, события или наблюдения, которые вызывают подозрения ввиду существенного отличия от большей части данных» [13].

На сайте «Википедия – свободная энциклопедия» дается следующее понятие: «Аномалия (данных) выбросы, необычности, шум, отклонения или исключения, являющиеся отображением таких эффектов как: структурный дефект, медицинские проблемы, мошенничество в банке или ошибки в тексте. Аномальные объекты часто не являются редкими, но проявляют неожиданную вспышку активности. Это не соответствует обычному статистическому определению выбросов как редких объектов и многие методы обнаружения выбросов терпят неудачу на таких данных, пока данные не будут сгруппированы подходящим образом. Зато алгоритмы кластерного анализа способны заметить микрокластеры, образованные таким поведением» [14].

Большой российский энциклопедический словарь: «Аномалия (греч. *anomalía*), отклонение от нормы, от общей закономерности, неправильность» [15].

Толковый словарь русского языка С.И. Ожегова: «Аномалия -и, ж. (книжн.). Отклонение от нормы, общей закономерности; неправильность» [16].

Толковый словарь живого великорусского языка В.И. Даля: «Аномалия, ж. греч. уклонение от обычного, несходство с обыкновенным, отступление в каком-либо явлении природы; изъятие, исключение, уклонение, причуда, необычайность, странность. Астроном. угол расстояния планеты от афелия или перигелия.

Аномальный, -личный, или аномальный, составляющий изъятие, исключение, уклонение от правила, порядка» [17].

Энциклопедический словарь медицинских терминов: «Аномалия (anomalía; греч. "отклонение"; син. а. развития) в биологии отклонение от структуры и (или) функции, присущей данному биологическому виду, возникшее вследствие нарушения развития организма; к А. относят пороки развития и уродства» [18].

Энциклопедический словарь Ф.А. Брокгауза: «Аномалия, в теории эллиптического движения аномалией называют угол между некоторым радиусом вектором и осью эллипса. Истинная аномалия есть угол между радиусом-вектором данного небесного тела (планеты, кометы, двойной звезды и пр.) и осью его орбиты. Эксцентрическая аномалия есть угол между направлением, проведенным из центра эллипса к точке пересечения круга, построенного на оси эллипса как на диаметре, с ординатою положения небесного тела относительно оси эллипса. Средняя аномалия – есть аномалия воображаемого тела, описывающего данный круг равномерным движением. Аномалистическим месяцем в движении Луны называется промежуток времени между двумя последовательными прохождениями Луны через перигей. Вследствие движения лунного перигея, в течение одного сидерического оборота луны в направлении движения самой Луны, аномалистический месяц длиннее сидерического» [19].

Современный словарь иностранных слов Л.П. Крысина: «Аномалия, [гр. anomalía] - 1) отклонение от нормы, от общей закономерности, неправильность, напр. а. развития; 2) в астрономии - истинная аномалия - угол между радиус-вектором небесного тела и направлением на перигей (перигелий или периастр) его орбиты; 3) в геофизике - магнитная аномалия - резкое увеличение составляющих земного магнетизма на каком-либо, участке земной коры, обычно обусловленное залеганием в ней больших скоплений железных руд (напр., курская магнитная аномалия, криворожская магнитная аномалия); гравитационная аномалия - разность между наблюдаемой силой тяжести и ее теоретическим (нормальным) значением (в той же точке), в которую введена поправка, учитывающая зависимость силы тяжести от высоты точки наблюдения» [20].

Большой толковый словарь русского языка Д.Н. Ушакова: «Аномалия, ж. (греч. *anomalía* - неравенство) (книжн.). Уклонение от закономерности явлений, отступление от существующего положения или порядка, неправильность. Аномалия в физическом развитии» [21].

Словарь русских синонимов и сходных по смыслу выражений Н.Ф. Абрамова: «Аномалия, неправильность, отступление, уклонение, уродливость, странность. Болезненное уклонение, недостаток, неправильность, ошибка» [22].

Толковый словарь русского языка Т.Ф. Ефремовой: «Аномалия, ж., отклонение от нормы, от общей закономерности» [23].

Словарь русского языка А. П. Евгеньевой: «Аномалия, ж., отклонение от нормы; неправильность. Аномалия в росте. Аномалия в физическом развитии» [24].

Исходя из приведенных понятий аномалии, в контексте задачи обнаружения аномалий в данных мониторинга природной среды, понятие аномалии данных определяется: с одной стороны, в виде статистически различимой ошибки, выброса, несвойственного поведения системы, и с другой стороны, как исключения – события отличающегося от ожидаемого в рамках и принятой ЛПП картиной мира, но тем не менее при этом точно отражающего изменение фактического состояния природной среды, и поэтому, возможно, представляющего собой новое знание.

Таким образом аномалиями в результатах мониторинга природной среды являются возмущения в факторном пространстве скалярных и векторных полей, вызванных как ошибками в средствах измерения, так и изменениями состояния исследуемой среды.

ПРИЛОЖЕНИЕ Д. РЕЗУЛЬТАТЫ ЛАБОРАТОРНЫХ ИСПЫТАНИЙ АППАРАТНОГО ДЕТЕКТОРА М-АНОМАЛИЙ НА СПЕЦИАЛИЗИРОВАННОМ ЛАБОРАТОРНОМ СТЕНДЕ

Визуализация телеметрических данных оцифрованных сигналов опорного и тестового измерительных каналов, получаемых с программируемого генератора сигнала на фоне случайной помехи.

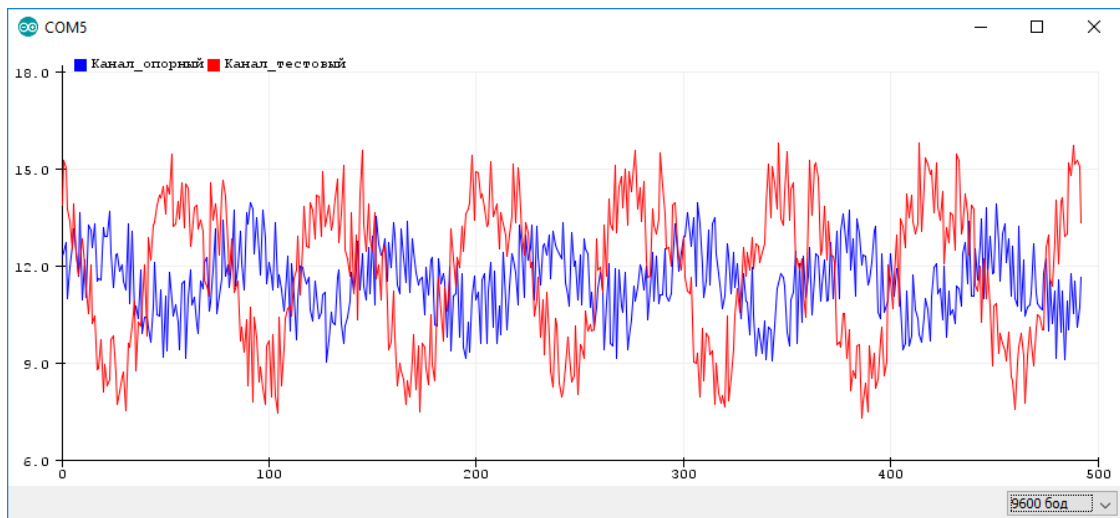


Рис. П.31. Результаты испытаний: опорный и тестовый сигналы

Визуализация листинга типового взаимодействия пользователя с терминалом при калибровке эталонного профиля измерительного канала.

```
COM5
Проверить 25000 [Отправить]
15:50:29.112 -> Система поддержки принятия решений о наличии аномалий в измерительных каналах
15:50:41.066 -> Загрузка профиля эталонного канала осуществлена успешно
15:50:41.160 ->
15:53:07.157 -> Обучить 50000
15:53:07.157 -> Запись нового профиля эталонного канала...
15:53:47.470 -> Запись профиля осуществлена успешно
15:53:47.517 -> Профиль эталонного канала сохранен во FLASH память
15:56:01.877 ->
15:57:01.145 -> Проверить 5000
15:57:01.145 -> Поиск аномалий измерительного канала...
15:58:22.909 -> Поиск аномалий завершен
15:58:22.909 -> Информационный критерий аномальности 0.7
15:58:23.003 -> Граница 95% доверительного интервала 1.1
15:58:23.096 -> Объем данных недостаточен для достоверного принятия решений
15:58:23.190 ->
[ ] Автопрокрутка [x] Показать отметки времени NL (Новая строка) 9600 бод Очистить вывод
```

Рис. П.32. Результаты испытаний (графический интерфейс)

Визуализация листинга типового взаимодействия пользователя с терминалом при проверке измерительного канала.

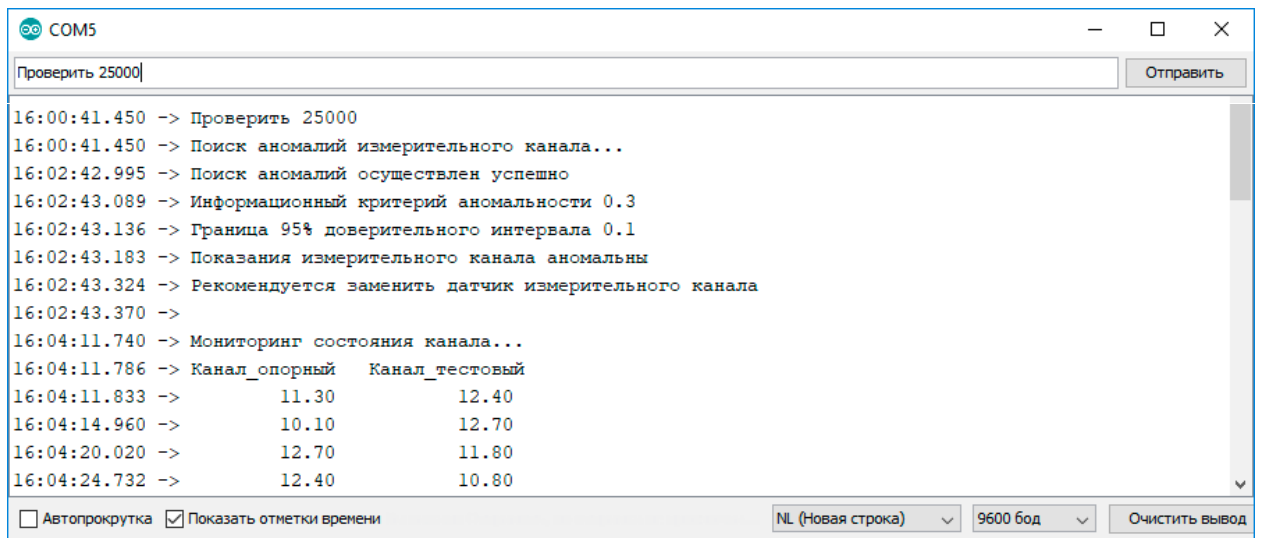


Рис. П.33. Результаты испытаний (графический интерфейс)

Визуализация гистограммы плотности распределения выборочных данных опорного и тестового измерительных каналов.

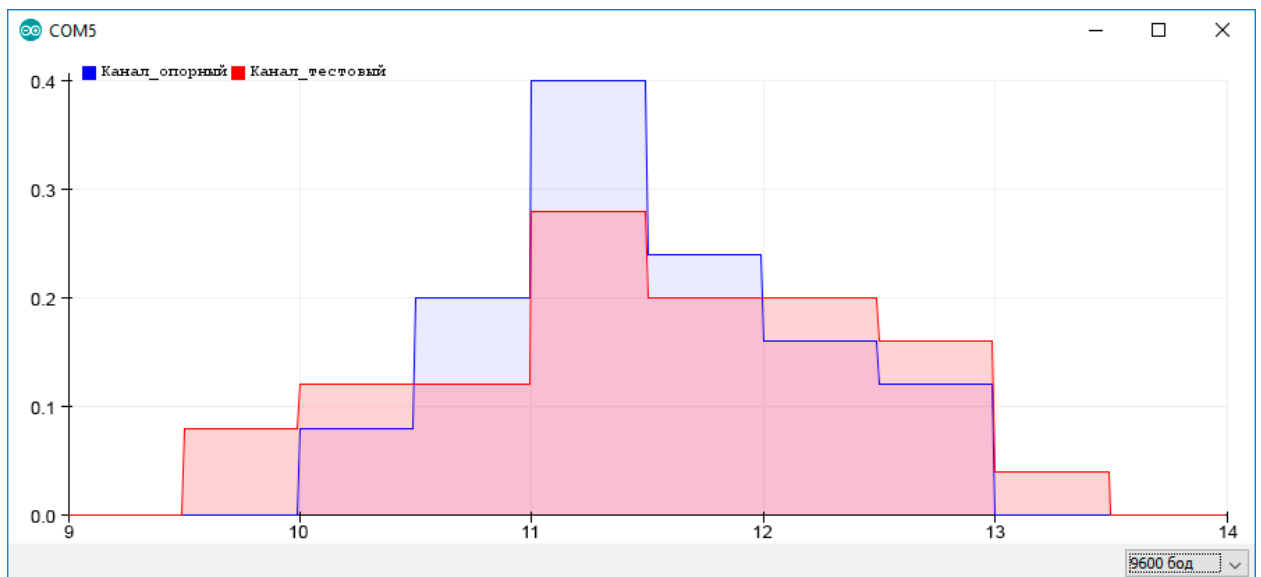


Рис. П.34. Результаты испытаний: гистограммы опорного и тестового сигналов

Результаты лабораторных испытаний детектора М-аномалий на специализированном лабораторном стенде подтвердили способность детектора обнаруживать управляемые возмущения в измерительном канале.

Иллюстрация подключения программного обеспечения модуля вычислительного контроллера к интерфейсу измерительного канала

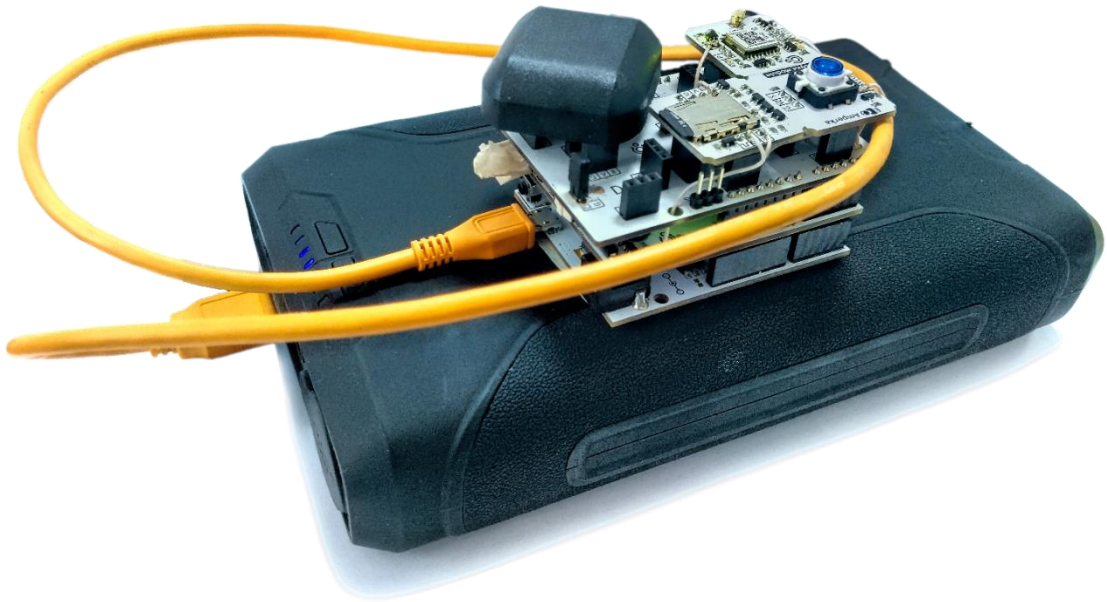


Рис. П.35. Модуль автономного длительного (до полугода) мониторинга данных измерительного канала с привязкой ко времени и записью на flash носитель

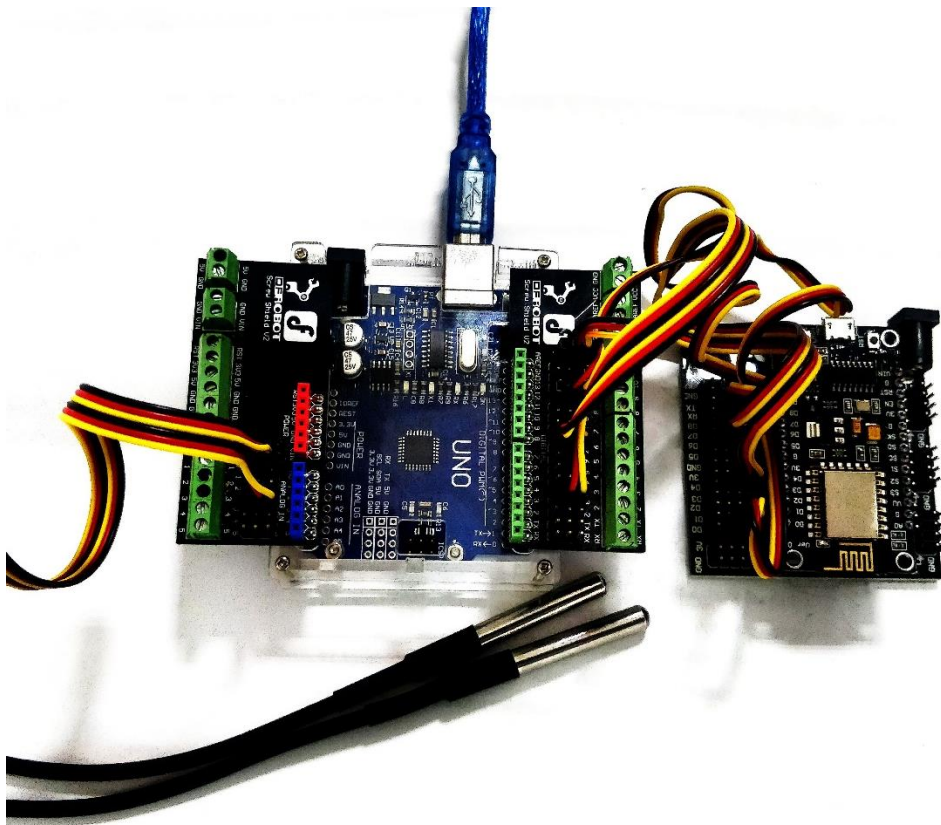


Рис. П.36. 16-канальный модуль фиксации данных аналого-цифровых измерительных каналов и передачи регистрируемого потока данных в сеть Internet

ПРИЛОЖЕНИЕ Е. ПАТЕНТ И СВИДЕТЕЛЬСТВА О РЕГИСТРАЦИИ
ПРОГРАММ ДЛЯ ЭВМ

РОССИЙСКАЯ ФЕДЕРАЦИЯ



ПАТЕНТ

НА ИЗОБРЕТЕНИЕ

№ 2755407

**СПОСОБ КОНТРОЛЯ КАЧЕСТВА ВОДЫ И
ДВУХКАНАЛЬНАЯ БИОСЕНСОРНАЯ УСТАНОВКА
ДЛЯ ЕГО ОСУЩЕСТВЛЕНИЯ**

Патентообладатель: *Федеральное государственное бюджетное
научное учреждение "Институт природно-технических
систем" (ИПТС) (RU)*

Авторы: *Греков Александр Николаевич (RU), Шишкин Юрий
Евгеньевич (RU), Кузьмин Константин Александрович (RU),
Мишуров Василий Жанович (RU), Рязанов Виктор
Алексеевич (RU), Трусевич Валерий Владимирович (RU)*

Заявка № 2020141487

Приоритет изобретения **15 декабря 2020 г.**
Дата государственной регистрации
в Государственном реестре изобретений
Российской Федерации **15 сентября 2021 г.**
Срок действия исключительного права
на изобретение истекает **15 декабря 2040 г.**

*Руководитель Федеральной службы
по интеллектуальной собственности*

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ
Сертификат 0x02AB5CF5C0B1ACEF9A40A2F08992E9A118
Владелец **Ивлиев Григорий Петрович**
Действителен с 15.01.2021 по 15.01.2035

Г.П. Ивлиев



РОССИЙСКАЯ ФЕДЕРАЦИЯ



(19) RU (11)

2 755 407 (13) C1

(51) МПК
G01N 33/18 (2006.01)ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(52) СПК
G01N 33/18 (2021.08)

(21)(22) Заявка: 2020141487, 15.12.2020

(24) Дата начала отсчета срока действия патента:
15.12.2020Дата регистрации:
15.09.2021

Приоритет(ы):

(22) Дата подачи заявки: 15.12.2020

(45) Опубликовано: 15.09.2021 Бюл. № 26

Адрес для переписки:

299011, г. Севастополь, ул. Ленина, 28,
Федеральное государственное бюджетное
научное учреждение "Институт природно-
технических систем" (ИПТС)

(72) Автор(ы):

Греков Александр Николаевич (RU),
Шипкин Юрий Евгеньевич (RU),
Кузьмин Константин Александрович (RU),
Мишуков Василий Жанович (RU),
Рязанов Виктор Алексеевич (RU),
Трусевич Валерий Владимирович (RU)

(73) Патентообладатель(и):

Федеральное государственное бюджетное
научное учреждение "Институт
природно-технических систем" (ИПТС) (RU)

(56) Список документов, цитированных в отчете

о поиске: RU 2570375 C2, 10.12.2015. RU
2437093 C1, 07.08.1984. US 2011028849 A1,
03.02.2011. CN 102012419 A, 13.04.2011.(54) СПОСОБ КОНТРОЛЯ КАЧЕСТВА ВОДЫ И ДВУХКАНАЛЬНАЯ БИОСЕНСОРНАЯ
УСТАНОВКА ДЛЯ ЕГО ОСУЩЕСТВЛЕНИЯ

(57) Реферат:

Использование: для контроля качества воды. Сущность изобретения заключается в том, что автоматическую непрерывную регистрацию в реальном масштабе времени реакций водных тест-объектов и пропускание контролируемой воды через резервуары с тест-объектами осуществляют посредством двухканальной биосенсорной установки, погружаемой в водоем, причем пропускание непрерывное, а поток контролируемой воды проходит в каждый момент времени через оба резервуара с тест-объектами в одинаковых физических условиях, при этом контролируемая вода поступает в первый и второй резервуары с разницей на интервал времени, необходимый для возникновения реакции тест-объектов на изменения химического состава контролируемой воды, а сравнение реакций групп тест-объектов между собой проводят путем вычисления разницы их параметров в интервале времени, равном разнице

времени поступления контролируемой воды в первый и второй резервуар, и при вычислении минимальной разницы параметров численных характеристик двигательной активности тест-объектов в обоих резервуарах делают вывод о неизменном химическом составе контролируемой воды, а при вычислении максимальной разницы делают вывод об измененном химическом составе контролируемой воды. Двухканальная биосенсорная установка включает в себя фильтр, насос с электромотором, делитель потока, два канала одинакового диаметра, но разной длины, при этом один из каналов соединен с одним, а второй - с другим резервуаром из двух одинаковых резервуаров, снабженных дренажными отверстиями, распределенными равномерно, общая площадь которых для каждого резервуара на 4-6% меньше площади сечения канала, кроме этого в каждом резервуаре размещены одинаковые тест-объекты и приборы для регистрации их реакции на

RU 2 755 407 C 1

RU 2 755 407 C 1

RUSSIAN FEDERATION

FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY(19) **RU** (11) **2 755 407**⁽¹³⁾ **C1**(51) Int. Cl.
G01N 33/18 (2006.01)(12) **ABSTRACT OF INVENTION**(52) CPC
G01N 33/18 (2021.08)

(21)(22) Application: 2020141487, 15.12.2020

(24) Effective date for property rights:
15.12.2020Registration date:
15.09.2021Priority:
(22) Date of filing: 15.12.2020

(45) Date of publication: 15.09.2021 Bull. № 26

Mail address:
299011, g. Sevastopol, ul. Lenina, 28, Federalnoe
gosudarstvennoe byudzhetnoe nauchnoe
uchrezhdenie "Institut prirodno-tehnicheskikh
sistem" (IPTS)

(72) Inventor(s):

Grekov Aleksandr Nikolaevich (RU),
Shishkin Yuriy Evgenevich (RU),
Kuzmin Konstantin Aleksandrovich (RU),
Mishurov Vasilij Zhanovich (RU),
Ryazanov Viktor Alekseevich (RU),
Trusevich Valerij Vladimirovich (RU)

(73) Proprietor(s):

Federalnoe gosudarstvennoe byudzhetnoe
nauchnoe uchrezhdenie "Institut
prirodno-tehnicheskikh sistem" (IPTS) (RU)(54) **METHOD FOR MONITORING WATER QUALITY AND A TWO-CHANNEL BIOSENSOR INSTALLATION FOR ITS IMPLEMENTATION**

(57) Abstract:

FIELD: water quality control.

SUBSTANCE: essence of the invention lies in the fact that automatic continuous registration in real time of the reactions of water test objects and the passage of controlled water through tanks with test objects is carried out by means of a two-channel biosensor installation immersed in a reservoir, and the transmission is continuous, and the flow of controlled water passes at each time through both tanks with test objects under the same physical conditions, while due to the difference in the lengths of the installation channels, controlled water enters the first and second tanks with a difference in the time interval necessary for the reaction of test objects to changes in the chemical composition of the controlled water to occur, and the reactions of groups of test objects are compared with each other by calculating the difference in their parameters in a time interval equal to the difference in the time of receipt of controlled water in the first and second tanks, and when calculating the minimum difference in the parameters of the numerical

characteristics of the motor activity of test objects in both tanks, it is concluded that the chemical composition of the controlled water is unchanged, and when calculating the maximum difference, the conclusion is made about the changed chemical composition of the controlled water. The two-channel biosensor installation includes a filter, a pump with an electric motor, a flow divider, two channels of the same diameter, but of different lengths, while one of the channels is connected to one, and the second to another tank of two identical tanks equipped with drainage holes distributed evenly, the total area of which for each tank is 4-6% less than the cross-sectional area of the channel, in addition, similar test objects and devices to register their reaction to changes in the conditions of the aquatic environment are placed in each tank.

EFFECT: invention provides the possibility of rapid detection of changes in the chemical composition of the aquatic environment.

2 cl, 1 dwg

RU 2 7 5 5 4 0 7 C 1

RU 2 7 5 5 4 0 7 C 1

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2020662839

Программа распознавания аномальностей в данных
мониторинговых наблюдений «А-КОНСУЛЬТАНТ»

Правообладатели: *Шишкин Юрий Евгеньевич (RU), Скатков
Александр Владимирович (RU)*

Авторы: *Шишкин Юрий Евгеньевич (RU),
Скатков Александр Владимирович (RU)*

Заявка № 2020660075

Дата поступления 07 сентября 2020 г.

Дата государственной регистрации

в Реестре программ для ЭВМ 20 октября 2020 г.

Руководитель Федеральной службы
по интеллектуальной собственности

 Г.П. Иблиев



РОССИЙСКАЯ ФЕДЕРАЦИЯ



RU2020662839

ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ
ГОСУДАРСТВЕННАЯ РЕГИСТРАЦИЯ ПРОГРАММЫ ДЛЯ ЭВМ

Номер регистрации (свидетельства):

2020662839

Дата регистрации: 20.10.2020

Номер и дата поступления заявки:

2020660075 07.09.2020

Дата публикации и номер бюллетеня:

20.10.2020 Бюл. № 10

Контактные реквизиты:

iurii.e.shishkin@gmail.com

Автор(ы):

Шишкин Юрий Евгеньевич (RU),

Скатков Александр Владимирович (RU)

Правообладатель(и):

Шишкин Юрий Евгеньевич (RU),

Скатков Александр Владимирович (RU)

Название программы для ЭВМ:

Программа распознавания аномальностей в данных мониторинговых наблюдений
«А-КОНСУЛЬТАНТ»

Реферат:

Программа представляет собой полимодельный комплекс, предназначенный для автоматизации процессов выявления аномальностей в данных мониторинга природных, технических и других сложных систем, представленных в текстовом и графическом видах. Программа обеспечивает выполнение следующих функций: цифровая фильтрация и восстановление недостающих данных, поиск аномальных объектов по шаблонам в массивах пространственно-временных данных, визуализация потенциально аномальных объектов и результатов их классификации. В основе алгоритмов программы использованы научные и технологические разработки авторов.

Тип ЭВМ: IBM PC – совместимые ПК.

ОС: Windows XP/Vista/Server/7/8/10, Linux Ubuntu/Red Hat.

Язык программирования:

Java

Объем программы для ЭВМ:

136,5 КБ

**ФЕДЕРАЛЬНАЯ СЛУЖБА ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ
(РОСПАТЕНТ)**

Бережковская наб., 30, корп. 1, Москва, Г-59, ГСП-3, 125993. Телефон (8-499) 240-60-15. Факс (8-495) 531-63-18

На № - от 30.10.2020

Наш № 2020Э10909

При переписке следует ссылаться на наш №
Исходящая корреспонденция от 20.10.2020

299003,
г. Севастополь,
ул. Черноморская
Шишкину Ю.Е.



У В Е Д О М Л Е Н И Е
о положительном результате проверки заявки
на государственную регистрацию программы для ЭВМ

Ставим Вас в известность о том, что программа для ЭВМ: Программа распознавания аномальностей в данных мониторинговых наблюдений «А-КОНСУЛЬТАНТ»

(21) по заявке № 2020660075/69

(22) Дата поступления заявки 07.09.2020

(71) Заявитель(и) Шишкин Юрий Евгеньевич (RU), Скотков Александр Владимирович (RU)

внесена в Реестр программ для ЭВМ, регистрационный № 2020662839 от 20.10.2020

Приложение: 1. заключение на 1 л. в 1 экз.

2. свидетельство о государственной регистрации на 1 л. в 1 экз.

Специалист-эксперт отдела
правового обеспечения
предоставления государственных
услуг

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ

Сведения о сертификате ЭП

Сертификат: 01DFF77700F4AB6EA7439B9DC1D703819F

Владелец: Бражникова Надежда Евгеньевна

Срок действия с 10.07.2020 по 27.07.2033

Бражникова Н.Е.

В результате проверки заявки, проведенной в соответствии со ст. 1262 Гражданского Кодекса Российской Федерации* и п.п. 60 – 82 Административного Регламента** установлено, что заявка на государственную регистрацию программы для ЭВМ удовлетворяет требованиям, установленным Правилами составления*** и Правилами оформления****.

Чайковская Е.А. (499)240-33-42

* Гражданский кодекс Российской Федерации (часть четвертая) от 18 декабря 2006 г. № 230-ФЗ (с изменениями)

** Административный регламент предоставления Федеральной службой по интеллектуальной собственности государственной услуги по государственной регистрации программы для электронных вычислительных машин или базы данных и выдаче свидетельств о государственной регистрации программы для электронных вычислительных машин или базы данных, их дубликатов, утвержден приказом Министерства экономического развития Российской Федерации от 5 апреля 2016 г. № 210, зарегистрированным Министерством юстиции Российской Федерации 5 июля 2016 г. № 42759.

*** Правила составления документов, являющиеся основанием для осуществления юридически значимых действий по государственной регистрации программы для электронных вычислительных машин или базы данных, утверждены приказом Министерством экономического развития Российской Федерации от 5 апреля 2016 г. № 211, зарегистрированным Министерством юстиции Российской Федерации 5 июля 2016 г. № 42758.

**** Правила оформления заявки на государственную регистрацию программы для электронных вычислительных машин или базы данных, утверждены приказом Министерством экономического развития Российской Федерации от 5 апреля 2016 г. № 211, зарегистрированным Министерством юстиции Российской Федерации 5 июля 2016 г. № 42758.

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2020664094

**«Программа GPS навигатора для автономного буйкового
акустического измерителя скорости течения
«А-НАВИГАТОР»**

Правообладатель: *Федеральное государственное бюджетное научное
учреждение «Институт природно-технических систем» (ИПТС)
(RU)*

Авторы: *Шишкин Юрий Евгеньевич (RU),
Кузьмин Константин Александрович (RU)*

Заявка № 2020663316

Дата поступления 27 октября 2020 г.

Дата государственной регистрации

в Реестре программ для ЭВМ 06 ноября 2020 г.

Руководитель Федеральной службы
по интеллектуальной собственности

Г.П. Ивлиев



РОССИЙСКАЯ ФЕДЕРАЦИЯ

**RU2020664094**

ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ
ГОСУДАРСТВЕННАЯ РЕГИСТРАЦИЯ ПРОГРАММЫ ДЛЯ ЭВМ

Номер регистрации (свидетельства):
2020664094

Дата регистрации: 06.11.2020

Номер и дата поступления заявки:
2020663316 27.10.2020

Дата публикации и номер бюллетеня:
06.11.2020 Бюл. № 11

Контактные реквизиты:
ipts-sevastopol@mail.ru

Автор(ы):

Шишкин Юрий Евгеньевич (RU),

Кузьмин Константин Александрович (RU)

Правообладатель(и):

Федеральное государственное бюджетное
научное учреждение «Институт

природно-технических систем» (ИПТС) (RU)

Название программы для ЭВМ:

Программа GPS навигатора для автономного буйкового акустического измерителя скорости течения «А-НАВИГАТОР»

Реферат:

Программа предназначена для эксплуатации ГЛОНАСС/BDS/GPS навигационного модуля для буйкового акустического измерителя скорости течения. Программа обеспечивает выполнение следующих функций: циклический опрос данных геолокации, приведение данных к стандартному виду, цифровая фильтрация аномальных значений, запись на внешний носитель, передача телеметрической информации, управление режимами энергосбережения навигационного модуля и периферийных устройств. В программе использованы научные и технологические разработки авторов.

Язык программирования: C/C++

Объем программы для ЭВМ: 16,9 КБ

**ФЕДЕРАЛЬНАЯ СЛУЖБА ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ
(РОСПАТЕНТ)**

Гиреевский наб., 30, корп. 1, Москва, Г-59, ГСП-3, 125993. Телефон (8-499) 240-60-15. Факс (8-495) 531-63-18

Иа № 10211/1-240

от 19.10.2020

Иаш: № 2020016609

При открытии следует сослаться на иаш №
Исходный корреспонденция от 06.11.2020

299011,
г. Севастополь,
ул. Лешина, 28,
ИПТС



**УВЕДОМЛЕНИЕ
о положительном результате проверки заявки
на государственную регистрацию программы для ЭВМ**

Ставим Вас в известность о том, что программа для ЭВМ: «Программа GPS навигатора для автономного буйкового акустического измерителя скорости течения «А-НАВИГАТОР»

(21) по заявке № 2020663316/69

(22) Дата поступления заявки 27.10.2020

(71) Заявитель(и) Федеральное государственное бюджетное научное учреждение «Институт природно-технических систем» (ИПТС) (RU)

внесена в Реестр программ для ЭВМ, регистрационный № 2020664094 от 06.11.2020

Приложение: 1. заключение на 1 л. в 1 экз.

2. свидетельство о государственной регистрации на 1 л. в 1 экз.

Специалист-эксперт отдела
правового обеспечения
предоставления государственных
услуг

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ

Сведения о сертификате ЭП

Сертификат: 01DFF77720F4A6E6A7430B9DC1D7C3610F

Елазегец: Бражникова Надежда Евгеньевна

Срок действия с 10.07.2020 по 27.07.2033

Бражникова Н.Ф.

В результате проверки заявки, проведенной в соответствии со ст. 1262 Гражданского Кодекса Российской Федерации* и п.п. 60 – 82 Административного Регламента** установлено, что заявка на государственную регистрацию программы для ЭВМ удовлетворяет требованиям, установленным Правилами составления*** и Правилами оформления****.

Чайковская Е.А. (499)240-33-42

* Гражданский кодекс Российской Федерации (часть четвертая) от 18 декабря 2006 г. № 230-ФЗ (с изменениями)

** Административный регламент предоставления Федеральной службой по интеллектуальной собственности государственной услуги по государственной регистрации программы для электронных вычислительных машин или базы данных и выдаче свидетельства о государственной регистрации программы для электронных вычислительных машин или базы данных, их дубликатов, утвержден приказом Министерства экономического развития Российской Федерации от 5 апреля 2016 г. № 210, зарегистрированным Министерством юстиции Российской Федерации 5 июля 2016 г. № 42758.

*** Правила составления документов, являющиеся основанием для осуществления юридически значимых действий по государственной регистрации программы для электронных вычислительных машин или базы данных, утверждены приказом Министерства экономического развития Российской Федерации от 5 апреля 2016 г. № 211, зарегистрированным Министерством юстиции Российской Федерации 5 июля 2016 г. № 42758.

**** Правила оформления заявки на государственную регистрацию программы для электронных вычислительных машин или базы данных, утверждены приказом Министерства экономического развития Российской Федерации от 5 апреля 2016 г. № 211, зарегистрированным Министерством юстиции Российской Федерации 5 июля 2016 г. № 42758.

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2020664093

Программа для автономного модуля буйкового
акустического измерителя скорости течения «ИСТ-1МА-SD»

Правообладатель: *Федеральное государственное бюджетное научное
учреждение «Институт природно-технических систем» (ИПТС)
(RU)*

Авторы: *Кузьмин Константин Александрович (RU),
Шишкин Юрий Евгеньевич (RU)*

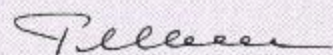
Заявка № 2020663317

Дата поступления 27 октября 2020 г.

Дата государственной регистрации

в Реестре программ для ЭВМ 06 ноября 2020 г.

Руководитель Федеральной службы
по интеллектуальной собственности

 Г.П. Ислюев



РОССИЙСКАЯ ФЕДЕРАЦИЯ



RU2020664093

ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ
ГОСУДАРСТВЕННАЯ РЕГИСТРАЦИЯ ПРОГРАММЫ ДЛЯ ЭВМ

Номер регистрации (свидетельства):
2020664093
Дата регистрации: 06.11.2020
Номер и дата поступления заявки:
2020663317 27.10.2020
Дата публикации и номер бюллетеня:
06.11.2020 Бюл. № 11
Контактные реквизиты:
ipts-sevastopol@mail.ru

Автор(ы):
Кузьмин Константин Александрович (RU),
Шишкин Юрий Евгеньевич (RU)
Правообладатель(и):
Федеральное государственное бюджетное
научное учреждение «Институт
природно-технических систем» (ИПТС) (RU)

Название программы для ЭВМ:

Программа для автономного модуля буйкового акустического измерителя скорости течения
«ИСТ-1МА-SD»

Реферат:

Программа предназначена для обеспечения автономной работы буйкового акустического измерителя скорости течения и его взаимодействия с ПЭВМ. Программа обеспечивает функционирование прибора в устанавливаемых пользователем режимах работы: в автономном режиме (включение и выключение измерительного модуля в соответствии с расписанием, приём и запись измерительных данных в модуль памяти), в телеметрическом режиме (передача данных измерительного модуля на ПЭВМ), в режиме работы с настройками и памятью (обмен данными настроек с ПЭВМ, передача содержимого памяти на ПЭВМ, очистка памяти), а также в режиме энергосбережения и индикации состояния контроллера заряда аккумуляторной батареи. В основе алгоритмов программы использованы научные и технологические разработки авторов.

Язык программирования: C/C++

Объем программы для ЭВМ: 96 КБ

**ФЕДЕРАЛЬНАЯ СЛУЖБА ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ
(РОСПАТЕНТ)**

Бережковская наб., 30, корп. 1, Москва, Г-59, ГСП-3, 125993. Телефон (8-499) 240-60-15. Факс (8-495) 531-63-18

На № 10211/1-239

от 19.10.2020

Наш № 2020Э16608

При переписке следует ссылаться на наш №
Исходящая корреспонденция от 06.11.2020

299011,
г. Севастополь,
ул. Ленина, 28,
ИПТС



У В Е Д О М Л Е Н И Е
о положительном результате проверки заявки
на государственную регистрацию программы для ЭВМ

Ставим Вас в известность о том, что программа для ЭВМ: Программа для автономного модуля буйкового акустического измерителя скорости течения «ИСТ-1МА-SD»

(21) по заявке № 2020663317/69

(22) Дата поступления заявки 27.10.2020

(71) Заявитель(и) Федеральное государственное бюджетное научное учреждение «Институт природно-технических систем» (ИПТС) (RU)

внесена в Реестр программ для ЭВМ, регистрационный № 2020664093 от 06.11.2020

Приложение: 1. заключение на 1 л. в 1 экз.

2. свидетельство о государственной регистрации на 1 л. в 1 экз.

Специалист-эксперт отдела
правового обеспечения
предоставления государственных
услуг

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ

Сведения о сертификате ЭП

Сертификат: 01DFF77700F4A36EA7438B9DC1C703819F

Владелец: Бражникова Надежда Евгеньевна

Срок действия с 10.07.2020 по 27.07.2033

Бражникова Н.Е.

В результате проверки заявки, проведенной в соответствии со ст. 1262 Гражданского Кодекса Российской Федерации* и п.п. 60 – 82 Административного Регламента** установлено, что заявка на государственную регистрацию программы для ЭВМ удовлетворяет требованиям, установленным Правилами составления*** и Правилами оформления****.

Чайковская Е.А. (499)240-33-42

* Гражданский кодекс Российской Федерации (часть четвертая) от 18 декабря 2006 г. № 230-ФЗ (с изменениями)

** Административный регламент предоставления Федеральной службой по интеллектуальной собственности государственной услуги по государственной регистрации программы для электронных вычислительных машин или базы данных и выдаче свидетельства о государственной регистрации программы для электронных вычислительных машин или базы данных, их дубликатов, утвержден приказом Министерством экономического развития Российской Федерации от 5 апреля 2016 г. № 216, зарегистрированным Министерством юстиции Российской Федерации 5 июля 2016 г. № 42759.

*** Правила составления документов, являющихся основанием для осуществления юридически значимых действий по государственной регистрации программы для электронных вычислительных машин или базы данных, утверждены приказом Министерством экономического развития Российской Федерации от 5 апреля 2016 г. № 211, зарегистрированным Министерством юстиции Российской Федерации 5 июля 2016 г. № 42758.

**** Правила оформления заявки на государственную регистрацию программы для электронных вычислительных машин или базы данных, утверждены приказом Министерством экономического развития Российской Федерации от 5 апреля 2016 г. № 211, зарегистрированным Министерством юстиции Российской Федерации 5 июля 2016 г. № 42758.

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2018666181

Программа синхронизации и коррекции измерений вектора скорости течения «ИНТЕГРАЦИЯ»

Правообладатель: *Федеральное государственное бюджетное научное учреждение «Институт природно-технических систем» (ИПТС) (RU)*

Авторы: *Шишкин Юрий Евгеньевич (RU), Кузьмин Константин Александрович (RU), Алексеев Сергей Юрьевич (RU)*

Заявка № 2018663369

Дата поступления 26 ноября 2018 г.

Дата государственной регистрации

в Реестре программ для ЭВМ 13 декабря 2018 г.

Руководитель Федеральной службы
по интеллектуальной собственности

Г.П. Ивлиев



РОССИЙСКАЯ ФЕДЕРАЦИЯ



RU 2018666181

ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

ГОСУДАРСТВЕННАЯ РЕГИСТРАЦИЯ ПРОГРАММЫ ДЛЯ ЭВМ

Номер регистрации (свидетельства):
2018666181

Дата регистрации: 13.12.2018

Номер и дата поступления заявки:
2018663369 26.11.2018Дата публикации и номер бюллетеня:
13.12.2018 Бюл. № 12Контактные реквизиты:
ipts-sevastopol@mail.ru

Автор(ы):

Шишкин Юрий Евгеньевич (RU),
Кузьмин Константин Александрович (RU),
Алексеев Сергей Юрьевич (RU)

Правообладатель(и):

Федеральное государственное бюджетное
научное учреждение «Институт
природно-технических систем» (ИПТС) (RU)

Название программы для ЭВМ:

Программа синхронизации и коррекции измерений вектора скорости течения «ИНТЕГРАЦИЯ»

Реферат:

Программа предназначена для сведения данных мониторинга, полученных с использованием измерителя скорости течений ИСТ-1М и GPS трекера с целью коррекции результатов зондирования на основании интеграционного метода и их оперативной визуализации. Программа обеспечивает выполнение следующих функций: поиск соответствий в наборах данных зондирования по временным маркерам; восстановление исходного вектора скорости и направления течений на каждом горизонте наблюдений; формирование выборок и визуализация отчетов в виде диаграмм и электронных таблиц. Программа используется в системах мониторинга водной среды, природных и технических систем.

Язык программирования: Java

Объем программы для ЭВМ: 77,5 Кб

**ФЕДЕРАЛЬНАЯ СЛУЖБА ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ
(РОСПАТЕНТ)**

Бережковская наб., 30, корп. 1, Москва, Г-59, ГСП-3, 125993. Телефон (8-499) 240-60-15. Факс (8-495) 531-63-18

На № 10211/1-381

от 15.11.2018

299011,

Наш № 2018Э16563

Севастополь,

При переписке следует ссылаться на наш №

ул. Ленина, 28,

Исходящая корреспонденция от 13.12.2018

ИПТС

*Копия
в дело*
Э.И. Терещенко
к сведениям
25.12.18 М.И.В.П.



**УВЕДОМЛЕНИЕ
о положительном результате проверки заявки
на государственную регистрацию программы для ЭВМ**

Ставим Вас в известность о том, что программа для ЭВМ: Программа синхронизации и коррекции измерений вектора скорости течения «ИНТЕГРАЦИЯ»

(21) по заявке № 2018663369/69

(22) Дата поступления заявки 26.11.2018

(71) Заявитель(и) Федеральное государственное бюджетное научное учреждение «Институт природно-технических систем» (ИПТС) (RU)

внесена в Реестр программ для ЭВМ, регистрационный № 2018666181 от 13.12.2018

Приложение: 1. заключение на 1 л. в 1 экз.

2. свидетельство о государственной регистрации на 1 л. в 1 экз.

Заместитель начальника
Управления организации
предоставления государственных
услуг

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ

Сведения о сертификате ЭП

Сертификат: 25155CA55794A0A3E811CB9ED545A365

Владелец: Терещенко Даниил Геннадьевич

Срок действия с 13.08.2018 по 08.02.2027

Терещенко Д.Г.

В результате проверки заявки, проведенной в соответствии со ст. 1262 Гражданского Кодекса Российской Федерации* и п.п. 60 – 82 Административного Регламента** установлено, что заявка на государственную регистрацию программы для ЭВМ удовлетворяет требованиям, установленным Правилами составления*** и Правилами оформления****.

Чайковская Е.А., (499)240-33-42

* Гражданский кодекс Российской Федерации (часть четвертая) от 18 декабря 2006 г. № 230-ФЗ (с изменениями)

** Административный регламент предоставления Федеральной службой по интеллектуальной собственности государственной услуги по государственной регистрации программы для электронных вычислительных машин или базы данных и выдаче свидетельства о государственной регистрации программы для электронных вычислительных машин или базы данных, их дубликатов, утвержден приказом Министерства экономического развития Российской Федерации от 5 апреля 2016 г. № 210, зарегистрированным Министерством юстиции Российской Федерации 5 июля 2016 г. № 42758.

*** Правила составления документов, являющихся основанием для осуществления юридически значимых действий по государственной регистрации программы для электронных вычислительных машин или базы данных, утверждены приказом Министерства экономического развития Российской Федерации от 5 апреля 2016 г. № 211, зарегистрированным Министерством юстиции Российской Федерации 5 июля 2016 г. № 42758.

**** Правила оформления заявки на государственную регистрацию программы для электронных вычислительных машин или базы данных, утверждены приказом Министерства экономического развития Российской Федерации от 5 апреля 2016 г. № 211, зарегистрированным Министерством юстиции Российской Федерации 5 июля 2016 г. № 42758.

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2018617880

Имитационная модель маршрутизатора
самоорганизующейся сети передачи данных «МАРШРУТ»

Правообладатель: *Шишкин Юрий Евгеньевич (RU)*Автор: *Шишкин Юрий Евгеньевич (RU)*

Заявка № 2018614503

Дата поступления 08 мая 2018 г.

Дата государственной регистрации

в Реестре программ для ЭВМ 03 июля 2018 г.

Руководитель Федеральной службы
по интеллектуальной собственности

Г.П. Ислюев



РОССИЙСКАЯ ФЕДЕРАЦИЯ

**RU 2018617880**

ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

ГОСУДАРСТВЕННАЯ РЕГИСТРАЦИЯ ПРОГРАММЫ ДЛЯ ЭВМ

Номер регистрации (свидетельства):
2018617880

Дата регистрации: 03.07.2018

Номер и дата поступления заявки:
2018614503 08.05.2018

Дата публикации и номер бюллетеня:
03.07.2018 Бюл. № 7

Контактные реквизиты:
yurii.shishkin@gmail.com; +7(978)861-55-69

Автор(ы):
Шишкин Юрий Евгеньевич (RU)

Правообладатель(и):
Шишкин Юрий Евгеньевич (RU)

Название программы для ЭВМ:

Имитационная модель маршрутизатора самоорганизующейся сети передачи данных «МАРШРУТ»

Реферат:

Программа реализует дискретно-событийную модель маршрутизатора самоорганизующейся частично децентрализованной одноранговой сети, расширяет функциональные возможности системы имитационного моделирования AnyLogic. Обеспечивает выполнение следующих функций: автоматическое динамическое построение оптимального диверсифицируемого маршрута произвольных пользовательских агентов, разрешение коллизий в сети, визуализация динамики возникновения маршрута. Предназначена для оценки эффективности методов вероятностного анализа сетевого трафика с целью осуществления адаптивного выбора вариантов конфигурирования сети. В основе программы лежат научные и технологические разработки автора.

Язык программирования: Java

Объем программы для ЭВМ: 163,5 Кб

**ФЕДЕРАЛЬНАЯ СЛУЖБА ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ
(РОСПАТЕНТ)**

Бережковская наб., 30, корп. 1, Москва, Г-59, ГСП-3, 125993. Телефон (8-499) 240-60-15. Факс (8-495) 531-63-18

На № _____ от _____
Наш № 2018Э05457
При перепишке следует ссылаться на наш №
Исходящая корреспонденция от 03.07.2018

299003,
г. Севастополь,
ул. Черноморская
Шишкину Ю.Е.



У В Е Д О М Л Е Н И Е
о положительном результате проверки заявки
на государственную регистрацию программы для ЭВМ

Ставим Вас в известность о том, что программа для ЭВМ: Имитационная модель маршрутизатора самоорганизующейся сети передачи данных «МАРШРУТ»

(21) по заявке № 2018614503/69

(22) Дата поступления заявки 08.05.2018

(71) Заявитель(и) Шишкин Юрий Евгеньевич (RU)

внесена в Реестр программ для ЭВМ, регистрационный № 2018617880 от 03.07.2018

Приложение: 1. заключение на 1 л. в 1 экз.

свидетельство о государственной регистрации на 1 л. в 1 экз.

Главный специалист-эксперт отдела
правового обеспечения
предоставления государственных
услуг

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ

Сведения о сертификате ЭП

Сертификат: 5FCA7035000200001748

Владелец: Круглов Вадим Эдуардович

Срок действия с 16.01.2018 по 01.04.2030

Круглов В.Э.

В результате проверки заявки, проведенной в соответствии со ст. 1262 Гражданского Кодекса Российской Федерации* и п.п. 60 – 82 Административного Регламента** установлено, что заявка на государственную регистрацию программы для ЭВМ удовлетворяет требованиям, установленным Правилами составления*** и Правилами оформления****.

Чайковская Е.А. (499)240-33-42

* Гражданский кодекс Российской Федерации (часть четвертая) от 18 декабря 2006 г. № 230-ФЗ (с изменениями)

** Административный регламент предоставления Федеральной службой по интеллектуальной собственности государственной услуги по государственной регистрации программы для электронных вычислительных машин или базы данных и выдаче свидетельств о государственной регистрации программы для электронных вычислительных машин или базы данных, их дубликатов, утвержден приказом Министерства экономического развития Российской Федерации от 5 апреля 2016 г. № 210, зарегистрированным Министерством юстиции Российской Федерации 5 июля 2016 г. № 42759.

*** Правила составления документов, являющихся основанием для осуществления юридически значимых действий по государственной регистрации программы для электронных вычислительных машин или базы данных, утверждены приказом Министерства экономического развития Российской Федерации от 5 апреля 2016 г. № 211, зарегистрированным Министерством юстиции Российской Федерации 5 июля 2016 г. № 42758.

**** Правила оформления заявки на государственную регистрацию программы для электронных вычислительных машин или базы данных, утверждены приказом Министерства экономического развития Российской Федерации от 5 апреля 2016 г. № 211, зарегистрированным Министерством юстиции Российской Федерации 5 июля 2016 г. № 42758.

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2018613184

Модуль консолидации структурированных данных
большого объема в режиме реального времени
«МОКОДАБО»

Правообладатель: **Шишкин Юрий Евгеньевич (RU)**

Автор: **Шишкин Юрий Евгеньевич (RU)**

Заявка № **2018610307**

Дата поступления **18 января 2018 г.**

Дата государственной регистрации

в Реестре программ для ЭВМ **05 марта 2018 г.**

Руководитель Федеральной службы
по интеллектуальной собственности

Г.П. Ивлиев



РОССИЙСКАЯ ФЕДЕРАЦИЯ

**RU 2018613184**ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ**ГОСУДАРСТВЕННАЯ РЕГИСТРАЦИЯ ПРОГРАММЫ ДЛЯ ЭВМ**Номер регистрации (свидетельства):
2018613184

Дата регистрации: 05.03.2018

Номер и дата поступления заявки:
2018610307 18.01.2018Дата публикации и номер бюллетеня:
05.03.2018 Бюл. № 3Контактные реквизиты:
yurii.shishkin@gmail.com, +7(978)8615569Автор(ы):
Шишкин Юрий Евгеньевич (RU)Правообладатель(и):
Шишкин Юрий Евгеньевич (RU)

Название программы для ЭВМ:

Модуль консолидации структурированных данных большого объема в режиме реального времени «МОКОДАБО»**Реферат:**

Программа предназначена для консолидации больших структурированных массивов данных мониторинга, полученных из различных источников в режиме реального времени, с динамически определяемым множеством ключевых полей. Программа обеспечивает выполнение следующих функций: преобразование семантически корректных данных в единый формат и их запись в структуры данных, оптимизированные для быстрого чтения и поиска; разрешение коллизий путем поиска однозначных соответствий записей по временным маркерам; обеспечение гибкости управления и согласованности данных. Программа переназначена для использования в системах мониторинга объектов облачной инфраструктуры, природных и технических систем.

Язык программирования: Java**Объем программы для ЭВМ:** 55,7 Кб

**ФЕДЕРАЛЬНАЯ СЛУЖБА ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ
(РОСПАТЕНТ)**

Бережковская наб., 30, корп. 1, Москва, Г-59, ГСП-3, 125993. Телефон (8-499) 240-60-15. Факс (8-495) 531-63-18

На №
Наш № 2018Э00480
При переписке следует ссылаться на наш №
Исходящая корреспонденция от 05.03.2018

от

299003,
г. Севастополь,
ул. Черноморская
Шишкину Ю.Е.



**У В Е Д О М Л Е Н И Е
о государственной регистрации**

Ставим Вас в известность о том, что программа для ЭВМ: Модуль консолидации структурированных данных большого объема в режиме реального времени «МОКОДАБО»

(21) по заявке № 2018610307/69

(22) Дата поступления заявки 18.01.2018

(71) Заявитель(и) Шишкин Юрий Евгеньевич (RU)

внесена в Реестр программ для ЭВМ, регистрационный № 2018613184 от 05.03.2018

Приложение: свидетельство о государственной регистрации на 1 л. в 1 экз.

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ

Сведения о сертификате ЭП

Сертификат: 5F4390CD000200001745

Владелец: Шевырева Жанна Игоревна

Срок действия с 16.01.2018 по 01.04.2030

Консультант

Шевырева Ж.И.

Мамонтов А.В. (499)240-33-42

В результате проверки заявки, проведенной в соответствии со ст. 1262 Гражданского Кодекса Российской Федерации* и п.п. 60 – 82 Административного Регламента** установлено, что заявка на государственную регистрацию программы для ЭВМ удовлетворяет требованиям, установленным Правилами составления*** и Правилами оформления****.

Чайковская Е.А. (499)240-33-42

* Гражданский кодекс Российской Федерации (часть четвертая) от 18 декабря 2006 г. № 230-ФЗ (с изменениями)

** Административный регламент предоставления Федеральной службой по интеллектуальной собственности государственной услуги по государственной регистрации программы для электронных вычислительных машин или базы данных и выдаче свидетельства о государственной регистрации программы для электронных вычислительных машин или базы данных, их дубликатов, утвержден приказом Министерством экономического развития Российской Федерации от 5 апреля 2016 г. № 210, зарегистрированным Министерством юстиции Российской Федерации 5 июля 2016 г. № 42759.

*** Правила составления документов, являющиеся основанием для осуществления юридически значимых действий по государственной регистрации программы для электронных вычислительных машин или базы данных, утверждены приказом Министерством экономического развития Российской Федерации от 5 апреля 2016 г. № 211, зарегистрированным Министерством юстиции Российской Федерации 5 июля 2016 г. № 42758.

**** Правила оформления заявки на государственную регистрацию программы для электронных вычислительных машин или базы данных, утверждены приказом Министерством экономического развития Российской Федерации от 5 апреля 2016 г. № 211, зарегистрированным Министерством юстиции Российской Федерации 5 июля 2016 г. № 42758.

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2017664038

Модуль снижения избыточности данных мониторинга
«МОСИДАМ»

Правообладатель: *Шишкин Юрий Евгеньевич (RU)*

Автор: *Шишкин Юрий Евгеньевич (RU)*

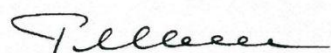
Заявка № 2017660984

Дата поступления 27 октября 2017 г.

Дата государственной регистрации

в Реестре программ для ЭВМ 14 декабря 2017 г.

Руководитель Федеральной службы
по интеллектуальной собственности

 Г.П. Ивлиев



РОССИЙСКАЯ ФЕДЕРАЦИЯ

**RU 2017664038**

ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

ГОСУДАРСТВЕННАЯ РЕГИСТРАЦИЯ ПРОГРАММЫ ДЛЯ ЭВМ

Номер регистрации (свидетельства):
2017664038

Дата регистрации: 14.12.2017

Номер и дата поступления заявки:
2017660984 27.10.2017

Дата публикации: 14.12.2017

Контактные реквизиты:
yurii.shishkin@gmail.com, тел:
+7(978)8615569

Автор:

Шишкин Юрий Евгеньевич (RU)

Правообладатель:

Шишкин Юрий Евгеньевич (RU)

Название программы для ЭВМ:

Модуль снижения избыточности данных мониторинга «МОСИДАМ»

Реферат:

Программа предназначена для уменьшения объема данных мониторинга, представленных в структурированной и слабоструктурированной форме, за счет удаления статистически незначимых записей. Программа обеспечивает выполнение следующих функций: детектирование факта выхода наблюдаемой величины за статистически допустимые пределы; удаление записей мониторинга наблюдаемой величины, значения которых статистически предсказуемы. Программа может использоваться в качестве дополнительного модуля в системах мониторинга компьютерных сетей, биологических объектов, автоматических или автоматизированных системах поддержки принятия решений по управлению облачными сервисами, а также в виде web-приложения.

Тип реализующей ЭВМ: IBM PC-совмест. ПК

Язык программирования: Java

Вид и версия операционной системы: Windows XP/Vista/7/8/10

Объем программы для ЭВМ: 23,7 Кб

**ФЕДЕРАЛЬНАЯ СЛУЖБА ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ
(РОСПАТЕНТ)**

Бережковская наб., 30, корп. 1, Москва, Г-59, ГСП-3, 125993. Телефон (8-499) 240-60-15. Факс (8-495) 531-63-18

На № Наш № 2017Э13323 <i>При переписке следует ссылаться на наш № Исходящая корреспонденция от 14.12.2017</i>	от	299003, г. Севастополь, ул. Черноморская Шишкину Ю.Е.
---	----	--



**У В Е Д О М Л Е Н И Е
о государственной регистрации**

Ставим Вас в известность о том, что программа для ЭВМ: Модуль снижения избыточности данных мониторинга «МОСИДАМ»

(21) по заявке № 2017660984/69

(22) Дата поступления заявки 27.10.2017

(71) Заявитель(и) Шишкин Юрий Евгеньевич (RU)

внесена в Реестр программ для ЭВМ, регистрационный № 2017664038 от 14.12.2017

Приложение: свидетельство о государственной регистрации на 1 л. в 1 экз.

Ведущий специалист-эксперт

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ

Сведения о сертификате ЭП

Сертификат: 1736355F000200001286
Владелец: Шевырева Жанна Игоревна
Срок действия с 19.01.2017 по 01.04.2030

Шевырева Ж.И.

Астафьева Е.Н. (499)240-33-42

В результате проверки заявки, проведенной в соответствии со ст. 1262 Гражданского Кодекса Российской Федерации* и п.п. 60 – 82 Административного Регламента** установлено, что заявка на государственную регистрацию программы для ЭВМ удовлетворяет требованиям, установленным Правилами составления*** и Правилами оформления****.

Чайковская Е.А. (499)240-33-42

* Гражданский кодекс Российской Федерации (часть четвертая) от 18 декабря 2006 г. № 230-ФЗ (с изменениями)

** Административный регламент предоставления Федеральной службой по интеллектуальной собственности государственной услуги по государственной регистрации программы для электронных вычислительных машин или базы данных и выдаче свидетельств о государственной регистрации программы для электронных вычислительных машин или базы данных, их дубликатов, утвержден приказом Министерства экономического развития Российской Федерации от 5 апреля 2016 г. № 210, зарегистрированным Министерством юстиции Российской Федерации 5 июля 2016 г. № 42759.

*** Правила составления документов, являющихся основанием для осуществления юридически значимых действий по государственной регистрации программы для электронных вычислительных машин или базы данных, утверждены приказом Министерства экономического развития Российской Федерации от 5 апреля 2016 г. № 211, зарегистрированным Министерством юстиции Российской Федерации 5 июля 2016 г. № 42758.

**** Правила оформления заявки на государственную регистрацию программы для электронных вычислительных машин или базы данных, утверждены приказом Министерства экономического развития Российской Федерации от 5 апреля 2016 г. № 211, зарегистрированным Министерством юстиции Российской Федерации 5 июля 2016 г. № 42759.