**ПРОГРАММНО-АППАРАТНЫЙ МОДУЛЬ ПОДДЕРЖКИ ПРИНЯТИЯ
РЕШЕНИЙ О НАЛИЧИИ КАЧЕСТВЕННЫХ АНОМАЛЬНЫХ ИЗМЕНЕНИЙ
В ВЫБОРОЧНЫХ ДАННЫХ НА БАЗЕ ИНФОРМАЦИОННЫХ МЕТРИК**

Ю.Е. Шишкин¹, А.В. Скатков²

¹Институт природно-технических систем, РФ, г. Севастополь, ул. Ленина, 28

²Севастопольский государственный университет,

РФ, г. Севастополь, ул. Университетская, 33

E-mail: iurii.e.shishkin@gmail.com

Предложена информационная технология и реализующий ее программно-аппаратный модуль поддержки принятия решений о наличии качественных аномальных изменений в выборочных данных, которые являются предикторами структурных изменений внутреннего состояния мониторируемых объектов, природно-технических систем или средств контроля. Представлена методика выбора параметрических критериев различия данных мониторинга, использующих численные меры информационной энтропии Шеннона и дивергенции Кульбака-Лейблера. Результирующий компонент системы мониторинга представлен в виде встраиваемого мобильного модуля, интерфейс которого проблемно-ориентирован на интегрирование в существующие информационно-измерительные системы и за счет средств визуализации обеспечивает повышение обоснованности принимаемых решений.

Ключевые слова: статистическая кластеризация, интеллектуальный анализ данных, машинное обучение, метрологическая надежность, выявление аномалий, численное моделирование.

Поступила в редакцию: 24.05.2021.

Введение. Актуальность проблемы выявления качественных аномальных изменений в выборочных данных мониторинга с целью повышения информационной и метрологической надежности измерительных каналов приборов и средств контроля природно-технических систем и объектов (ПТС/О) не оставляет сомнений как в научном, так и прикладном аспекте [1–4]. Особенно остро стоит задача своевременного выявления факта ухудшения ключевых метрологических характеристик точности, сходимости и воспроизводимости измерений, имеющего место при долговременных постановках *in situ* и определении природы аномалий, возникающих в данных мониторинга, с целью повышения адекватности принимаемых решений на их основе [5, 6].

Природа возникновения аномалий данных мониторинга ПТС/О. Выделим и рассмотрим природу возникновения двух доминирующих типов аномалий данных мониторинга, свойственных ПТС/О:

а) единичные аномальные значения в выборке: точечные выбросы, экстре-

мальные члены в выборке, глобальные и локальные минимальные и максимальные элементы, значения, выбивающиеся из временного ряда в заданной окрестности. Как правило, такие ошибки – нетиповые и связаны с техническими причинами, человеческим фактором, искажениями в каналах регистрации, передачи и хранения данных, единичным шумом. Существует развитый математический аппарат, позволяющий относительно легко обнаруживать и восстанавливать данные содержащие подобного рода аномалии [7, 8];

б) аномалии выборочных данных, связанные с изменениями формы распределения и сдвига выборочного распределения, – особый вид аномалий. Источниками таких аномалий являются неопределенности в знаниях, возникающие вследствие того, что данные для обучения и для теста взяты из выборок с различными распределениями, если наблюдаемая система качественно изменила свое состояние, вследствие недостатка объема и высокой вариативности данных обучающей выборки и в целом в

большей степени характеризует свойство модели, а не данных. Интуитивный поиск подобных аномалий в данных, даже представленных в визуальной форме, как правило, затруднен вследствие стохастического характера проявления изменений наблюдаемых величин и слабой выраженности этих изменений в абсолютных показателях [9].

Постановка задачи. В рамках статьи требуется получить информационную технологию для программно-аппаратного модуля поддержки принятия решений о наличии качественных аномалий в выборочных данных, свидетельствующих об изменениях внутреннего состояния мониторируемой ПТС/О или наличии проявлений последствий деградиционных процессов в самих измерительных каналах и средствах регистрации.

Формальная постановка вопроса рассмотрена в статье В.С. Королюка [10] для оценки уклонений между эмпирической и теоретической кривыми распределения, с использованием критериев согласия А.Н. Колмогорова и Н.В. Смирнова. При том, что сформулированные в статье 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 выборки объемом $n \rightarrow \infty$, что не может удовлетворять при практических приложениях. А характер асимптотической сходимости не исследован и в общем случае исследованию аналитическими методами, судя по всему, не подлежит.

Численное оценивание аномальности выборочных данных информационными метриками. Как альтернатива аналитическому решению в статье предлагается перейти к использованию численных оценок степени аномальности выборочных данных, полученных по информационным соотношениям.

Поэтому выдвигается **Гипотеза 1.** Контроль формы и контроль смещения статистического распределения можно

свести к одному главному информационному признаку. Оба указанных искажения изменяют информацию о выборке. На данный момент нет работ, в которых указанная проблема аналитически исследуется как задача в математической постановке. Поэтому для исследования точности и полноты критерия используются численные методы, в частности целенаправленный эксперимент и вероятностное моделирование [11, 12]. Будем использовать две наиболее часто встречающиеся информационные численные меры различия между опорной и возмущенной (содержащей изменения формы или сдвиг статистического распределения) выборками: информационную энтропию Шеннона (1) и дивергенцию Кульбака-Лейблера (2) [13, 14]:

$$S(p) = - \sum_{i=1}^n \frac{p_i}{\log p_i}, \quad (1)$$

$$K(p, p^*) = \sum_{i=1}^n p_i \log \frac{p_i}{p_i^*}, \quad (2)$$

где n – число интервалов разбиения выборки, p_i – i -й элемент опорной выборки, p_i^* – i -й элемент возмущенной выборки.

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 (1) и (2) в нулевых точках в виде (3) и (4) соответственно, следующим образом:

$$S(p_i) = \begin{cases} -\frac{p_i}{\log p_i}, & \text{если } p_i > 0 \\ 0, & \text{иначе} \end{cases}, \quad (3)$$

$$K(p_i, p_i^*) = \begin{cases} p_i \log \frac{p_i}{p_i^*}, & \text{если } p_i \cdot p_i^* > 0 \\ 0, & \text{иначе} \end{cases}. \quad (4)$$

Для удобства дальнейших выкладок и унификации методики будем использовать нормированные информационные метрики как функционалы над множествами, тогда информационные метрики Шеннона (5) и Кульбака-Лейблера (6) соответственно примут вид:

$$IS(p, p^*) = \left| \sum_{i=1}^n S\left(\frac{p_i}{N}\right) - \sum_{i=1}^n S\left(\frac{p_i^*}{N}\right) \right|, \quad (5)$$

$$IK(p, p^*) = \left| \sum_{i=1}^n K\left(\frac{p_i}{N}, \frac{p_i^*}{N}\right) \right|, \quad (6)$$

где N – объем выборки p и p^* .

Исследуем зависимость критериев (5) и (6) от величины сдвига статистического распределения выборки и изменения формы распределения выборки. Анализ результатов целенаправленного эксперимента осуществляется с целью минимизации факторов, приводящих к снижению метрологических характеристик, представленных в виде ошибок первого и второго рода.

Структура программного обеспечения моделирования аномалий данных мониторинга ПТС/О. Одно из принципиальных требований к программному обеспечению, осуществляющему вероятностное моделирование, является его параметрическая гибкость и наличие точек планирования эксперимента: выбор функции распределения, выбор параметров распределения, выбор объема выборки, выбор разрядности представления, выбор кратности репликаций (число повторов экспериментов) и др.

На рис. 1 представлена структура разработанного программного обеспечения для моделирования аномалий данных ПТС/О и принятия решений на их основе.

Особенностью структуры является наличие в ней модулей генерирования наборов выборок с параметрами, изменяющимися с заданным шагом, попарного их сравнения и оценки границ уверенного распознавания факта аномального изменения формы и сдвига статистического распределения, а также связей между модулями и графическим интерфейсом пользователя, осуществляющим проактивное вовлечение оператора в процесс выбора целевых критериев.

Гибкость структуры обеспечивается за счет наличия точек планирования экспериментов, доступ к которым осуществляется посредством графического интерфейса, адаптивного выбора метрик выявления факта наличия качественных изменений с учетом величины сдвига выборочного распределения и скорости сходимости результата. Повышение обоснованности принимаемых решений обеспечивается за счет возможности автоматического сравнения результатов с аналитическим решением, в тех случаях где оно имеется. Открытость интерфейса обеспечивается возможностью импорта данных с неявно заданной структурой, что позволяет осуществлять анализ слабоструктурированных данных независимо от их природы. Возможность накопления данных обеспечивается возможностью использовать как внутренние накопители, так и удаленные ресурсы хранения данных. Интерактивные режимы обеспечиваются возможностью циклического взаимодействия пользователя с модулями оценки границ уверенного распознавания, выбора требуемой степени уверенности классификатора, выбора порогов принятия решений, оценки сходимости и получения итогового набора управляющих правил классификатора.

Выполним численное моделирование в соответствии с предложенной методикой, производя попарное сравнение эффективности критериев для ряда видов распределений и смещений.

Рис. 1. Структура ПО программно-аппаратного модуля для проведения целенаправленного эксперимента, численного статистического моделирования аномалий данных ПТС/О и формирования управляющих правил поддержки принятия решений на их основе

Fig. 1. The structure of the software and hardware module for conducting a targeted experiment, numerical statistical modeling of NTS/O data anomalies and the formation of control rules for decision support based on them

Целевой вычислительный эксперимент 1. Целью эксперимента является установление зависимости информационной энтропии Шеннона в случае, представленном на рис. 2, когда опорная выборка представляет собой реализацию случайной величины, распределенной с усеченным нормальным распределением с нулевым средним и границами $\pm 3\sigma$, и возмущенная выборка также подчинена нормальному распределению и перемещается относительно опорной в диапазоне $0 \dots 3\sigma$, при объеме выборок $N=100$, числе интервалов разбиения выборки $n=10$ и границе доверительного интервала $\theta=0,95$.

График зависимости изменения энтропии Шеннона IS от силы возмущения типа сдвиг статистического распределения, рассчитанный по (5) представлен на рис. 3. Зависимость показывает, что существует три зоны: зона I – уверенное принятие гипотезы о том, что смещение не произошло, зона III – уверенное принятие о том, что смещение произошло, зона II – неуверенное распознавание. Для данного случая зона II (зона неопре-

деленности) относительно велика и ограничена диапазоном $0,55 \dots 0,95$ для смещений соответственно $(2,1 \dots 2,7)\sigma$.

На рис. 3–5, 7 «точкой возникновения КАИ» обозначены моменты возникновения качественных аномальных изменений в соответствии с метриками.

Рис. 2. Гистограммы распределения опорной выборки (серая), график ее плотности распределения рассчитанной аналитически (оранжевый), и возмущенной выборки (красная)

Fig. 2. Histogram of the reference sample distribution (gray), the graph of its distribution density calculated analytically (orange), and the perturbed sample histogram (red)

Рис. 3. График зависимости изменения энтропии Шеннона от силы возмущения типа сдвиг (оранжевый); нижняя граница 95% доверительного интервала (красный), верхняя граница (синий)

Fig. 3. The plot of the change dependence in Shannon's entropy on the strength of the shear-type perturbation (orange); lower limit of 95% confidence interval (red), upper limit (blue)

Целевой вычислительный эксперимент 2. Целью эксперимента является установление зависимости информационной метрики дивергенции Кульбака-Лейблера при начальных условиях, аналогичному эксперименту 1.

График зависимости изменения дивергенции Кульбака-Лейблера ИК от силы возмущения типа сдвиг, рассчитанный по (6) представлен на рис. 4.

Зависимость показывает, что существует три зоны: зона I – уверенное принятие гипотезы о том, что смещение не произошло, зона III – уверенное принятие о том, что смещение произошло, зона II – неуверенное распознавание. Для данного случая нижняя граница зоны II существенно ниже (лучше) чем аналогичная граница у метрики Шеннона и ограничена диапазоном $0,1 \dots 0,4$ для сдвига статистического распределения соответственно $(0,8 \dots 2,4)\sigma$. При этом стоит отметить, что при смещении выборки более $2,4\sigma$ метрика (6) не дает однозначного ответа относительно наличия возмущения в отличие от метрики (5) для которой верхняя граница при аналогичных начальных условиях составила $2,7\sigma$.

Рис. 4. График зависимости дивергенции Кульбака-Лейблера от силы возмущения типа сдвиг (коричневый); нижняя граница 95% доверительного интервала (желтый), верхняя граница (черный)

Fig. 4. The plot of the Kullback-Leibler divergence dependence on the strength of the shear type disturbance (brown); lower limit of 95% confidence interval (yellow), upper limit (black)

Целевой вычислительный эксперимент 3. Целью эксперимента является оценка сходимости, нахождение точки безразличия и выявление границ эффективности двух конкурирующих информационных метрик изменения энтропии Шеннона и дивергенции Кульбака-Лейблера при начальных условиях, аналогичному эксперименту 1 но с переменным объемом выборки $N = 10 + 10\sigma$.

Графики и 95% доверительные интервалы зависимости дивергенции Кульбака-Лейблера ИК и изменения энтропии Шеннона IS от силы возмущения типа сдвиг и объема выборки, рассчитанные по (6) и (5) соответственно представлены на рис. 5.

Результаты целенаправленных экспериментов показали, что существует пять зон на оси сдвига статистического распределения:

– зона I $(0 \dots 0,4)\sigma$ – ни одна из рассмотренных метрик не может достоверно детектировать факт наличия смещения выборки при объеме выборки заданного объема;

– зона II $(0,4 \dots 1,3)\sigma$ – уверенное принятие о том, что смещение произошло метрикой дивергенции Кульбака-Лейблера но не изменения энтропии Шеннона;

Рис. 5. График и 95% доверительный интервал зависимости дивергенции Кульбака-Лейблера (красный) и изменения энтропии Шеннона (зеленый) от силы возмущения типа сдвиг статистического распределения и объема выборки

Fig. 5. Graph and 95% confidence interval of the dependence of the Kullback-Leibler divergence (red) and the change in Shannon entropy (green) on the strength of a disturbance such as a shift in the statistical distribution and sample size

– зона III $(1,3...2,2)\sigma$ – уверенное принятие обеими метриками решения о том, что смещение произошло, при этом метрика дивергенции Кульбака-Лейблера является более предпочтительной как обладающая большей доверительной вероятностью для данного диапазона и начальных условий;

– зона IV $(2,2...4,2)\sigma$ – уверенное принятие решения обеими метриками о том, что смещение произошло, при этом метрика изменения энтропии Шеннона является более предпочтительной как обладающая большей доверительной вероятностью;

– зона V $(4,2... \infty)\sigma$ – уверенное принятие того, что смещение произошло метрикой изменения энтропии Шеннона но не дивергенции Кульбака-Лейблера.

Граница между зонами III и IV $4,2\sigma$ является точкой безразличия, когда точность обоих рассмотренных информационных критериев одинакова.

Скорость схождения метрики изменения энтропии Шеннона при увеличении объема выборки выше чем у конкурирующей, но сам доверительный интервал в окрестности нулевого смещения распределения значительно шире.

Целевой вычислительный эксперимент 4. Целью эксперимента является установление зависимости информационных метрик с начальными аналогичному эксперименту 1, но в случае, представленном на рис. 6, когда возмущенная выборка изменила форму статистического распределения и распределена по равномерному закону.

Рис. 6. Гистограммы распределения опорной выборки (серая), график ее плотности распределения рассчитанной аналитически (оранжевый), и возмущенной выборки с равномерным распределением (красная)

Fig. 6. Histograms of the distribution of the reference sample (gray), the graph of its distribution density calculated analytically (orange), and a disturbed sample with a uniform distribution (red)

Графики и 95% доверительные интервалы зависимости дивергенции Кульбака-Лейблера IK и изменения энтропии Шеннона IS от силы возмущения типа сдвиг при изменении формы распределения возмущенной выборки с нормального на равномерное, рассчитанные по (6) и (5) соответственно представлены на рис. 7.

Рис. 7. График и 95% доверительный интервал зависимости дивергенции Кульбака-Лейблера (зеленый) и изменения энтропии Шеннона (синий) от силы возмущения типа сдвиг статистического распределения и изменения формы распределения

Fig. 7. Graph and 95% confidence interval of the dependence of the Kullback-Leibler divergence (green) and the change in Shannon entropy (blue) on the strength of a disturbance such as a statistical distribution shift and the distribution shape changes

Зависимость показывает, что обе рассмотренные метрики способны детектировать наличие возмущения, проявляющиеся в изменении формы распределения, при этом метрика изменения энтропии Шеннона IS для детектирования возмущений такого рода является более предпочтительной. Метрики расположились в двух зонах: метрика изменения энтропии Шеннона расположена в зоне I – уверенное принятие гипотезы о том, что смещение произошло, метрика дивергенции Кульбака-Лейблера расположена в зоне II – неуверенное распознавание.

Структура программно-аппаратного модуля поддержки принятия решений. Реализацию разработанной информационной технологии предлагается осуще-

ствить в виде встраиваемого программно-аппаратного модуля, оснащенного гибким интерфейсом для интегрирования в существующие приборы и информационные-измерительные системы контроля ПТС/О в том числе разработанные в Центре экологического приборостроения и эконергетики ИПТС: ИСТ-1М, ИСТ-1МА, АБКЭМ, ИСЗ-1, ВИАП и др.

Основными компонентами результирующего программно-аппаратного модуля являются:

- вычислительный ресурс модуля осуществляет сравнение параметров опорной (эталонной) выборки и информационные метрики экспериментально полученных данных мониторинга;
- система хранения данных модуля предоставляет образцы нормальных и аномальных выборочных распределений, контрольно-предупредительных границ;
- система адаптивной интеграции осуществляет импорт данных представленных в широком диапазоне форматов, получаемых от измерительных средств, в том числе слабоструктурированных данных с самоописываемой структурой;
- система визуализации осуществляет гибкое взаимодействие с пользователем, отображение рассчитанных информационных метрик для повышения обоснованности принимаемых решений, предоставляет интерактивные возможности планирования серий целевых вычислительных экспериментов для выбора оптимальных информационных метрик обладающих наибольшей чувствительностью к возмущениям для каждого конкретного случая.

Например, если функция распределения измеряемых параметров среды обусловлена наличием загрязняющих примесей (аномалии температурного фронта, плотности среды, физико-химических показателей), то они будут детектированы как аномалии возмущения формы распределения. Аналогично система сообщает о качественных аномальных изменениях, когда одномодальное распределение становится двумодальным, или почти равномерное становится в значительной степени экспоненциальным или гамма-распределением с выраженными асимметриями.

Возможные варианты компоновки прибора по степени автономности:

- встраиваемый компонент в который поступают предварительно обработанные данные, передаваемые по интерфейсной шине в соответствии с показаниями датчиков из среды, в котором собственно происходит реализация системы поддержки принятия решений;

- автономный модуль в котором осуществляется обработка, который связан интерфейсом непосредственно с лицом, принимающим решение (ЛПР);

- полностью автономный модуль, где вместо ЛПР – искусственный интеллект.

Заключение. Анализ результатов целевых вычислительных экспериментов по предложенной информационной технологии и реализующему ее программно-аппаратному модулю поддержки принятия решений о наличии качественных аномальных изменений в выборочных данных показал высокую эффективность выявления возмущений изменения формы распределения и сдвига выборочного распределения.

Предлагаемые критерии согласия проявили свои лучшие свойства, когда сравниваются нормальное распределение в качестве опорного и равномерное в качестве возмущенного, что соответствует случаям отказа датчика или обрыва связи.

Целевые эксперименты показали:

- изменения энтропии Шеннона имеют большую эффективность при выявлении возмущения с t -распределением Стюдента и равномерным распределением,

- метрика дивергенции Кульбака-Лейблера имеет большую эффективность при выявлении возмущения с t -распределением Пуассона и распределением Вейбулла.

Статистическая сходимость результатов обеспечивается при объеме выборки $N > 10$, скорость сходимости изменения энтропии Шеннона выше конкурирующего метода, но изначально при малом сдвиге статистического распределения $(0...0,4)\sigma$ доверительный интервал этой метрики шире.

В частности, подобная методология может быть применена в медицинских

приложениях при анализе эпидемиологических показателей, в природоохранной тематике, в задачах информационной защиты при выявлении аномальности трафика, при решении задач выявления уязвимостей интерфейсов беспилотных транспортных средств.

Представленный подход является информационной технологией, поэтому необходимо в каждом конкретном случае по предлагаемой методике проводить численное моделирование в рамках тех задач, которые стоят перед исследователем и находить зоны уверенного обнаружения. В статье продемонстрирован метод как это можно сделать. На данный момент подана заявка на регистрацию разработанного ПО на представленный программно-аппаратный модуль.

Работа выполнена при частичной поддержке Российского фонда фундаментальных исследований (грант № 19-29-06015/19, 19-29-06023/19).

СПИСОК ЛИТЕРАТУРЫ

1. *Мониторинг* антропогенных воздействий на прибрежные акватории Черного моря по многоспектральным космическим изображениям / В.Г. Бондур, В.Е. Воробьев, В.В. Замшин [и др.]. // Исследование Земли из космоса. 2017. № 6. С. 3–22.
2. Губанов В.И., Стельмах Л.В., Клименко Н.П. Комплексные оценки качества вод севавтопольского взморья (Черное море) // Экология моря. 2002. Т. 62. С. 76–80.
3. Dunbabin M., Marques L. Robots for Environmental Monitoring Significant Advancements and Applications // IEEE Robot. Automat. Mag. 2012. № 19 (1). P. 24–39. doi: 10.1109/mra.2011.2181683
4. Шишкин Ю.Е., Скотков А.В. Информационные технологии обнаружения аномалий в мониторинговых наблюдениях: монография. Симферополь: ИТ «АРИАЛ», 2019. 368 с.
5. Design of embedded software for in-situ monitoring system of marine radioactivity based on a buoy / B. Wu; Y. Zhang; D. Yuan // 2019 Chinese Control

Conference (CCC). 2019. doi: 10.23919/chicc.2019.8866680

6. 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

7. Ковалев С.М. Методы многошагового предсказания аномалий в темпоральных данных // Известия ЮФУ. Технические науки. 2013. № 7 (144). С. 85–91.

8. Прикладной анализ случайных процессов / А.В. Графкин, В.В. Графкин, О.А. Дегтярева [и др.] / под ред. С.А. Прохорова. Самара. 2007. 582 с.

9. Шайтура С.В. Интеллектуальный анализ геоданных // Перспективы науки и образования. 2015. № 6 (18). С. 24–30.

10. Корольюк В.С. О расхождении эмпирических распределений для случая

двух независимых выборок // Изв. АН СССР. Серия математика. 1955. № 19 (1). С. 81–96.

11. Болгов М.В. Современное состояние теории корреляции для гидрологических расчетов и стохастического моделирования // Водное хозяйство России: проблемы, технологии, управление. 2012. № 4. С. 7–20.

12. 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

13. Кульбак С. Теория информации и статистика. М.: Наука, 1967. 408 с.

14. Hurvich C.M., Tsai C.L. Regression and time series model selection in small samples // Biometrika. 1989. № 76 (2). P. 297–307. DOI: 10.1093/biomet/76.2.297

SOFTWARE AND HARDWARE MODULE TO SUPPORT DECISION-MAKING OF QUALITATIVE ABNORMAL CHANGES PRESENCE IN SAMPLE DATA BASED ON INFORMATION METRICS

Y.E. Shishkin¹, A.V. Skatkov²

¹Institute of Natural and Technical Systems,
RF, Sevastopol, Lenin St., 28

²Sevastopol State University,
RF, Sevastopol, Universitetskaya St., 33

An information technology and a software and hardware module for decision-making support on the presence of qualitative anomalous changes in sample data, which are predictors of structural changes of monitored objects internal state, natural-technical systems or controls, are proposed. A method for choosing parametric criteria for the difference in monitoring data using numerical measures of Shannon information entropy and Kullback-Leibler divergence is presented. The resulting component of the monitoring system in the form of an embedded mobile module is presented. Interface of monitoring system is problem-oriented for integration into existing information-measuring systems and, due to visualization tools, provides an increase in the validity of decisions made.

Keywords: statistical clustering, data mining, machine learning, metrological reliability, anomaly detection, numerical modeling.

REFERENCES

1. Bondur V.G., Voron'ev V.E., and Zamshin V.V. Monitoring antropogennykh vozhdeystviy na pribrezhnyye akvatorii chernogo morya po mnogosppektral'nym kosmicheskim izobrazheniyam (Monitoring of anthropogenic impacts on the coastal waters of the Black Sea using multispectral satellite images). *Issledovanie Zemli iz kosmosa*, 2017, Vol. 6, pp. 3–22.

2. Gubanov V.I., Stelmah L.V., and Klimenko N.P. Kompleksnye ocenki kachestva vod sevastopol'skogo vzmor'ya Chernoe more (Comprehensive assessments of the water quality of the Sevastopol seaside Black Sea). *Ekologiya morya*, 2002, Vol. 62, pp. 76–80.
3. Dunbabin M. and Marques L. Robots for Environmental Monitoring Significant Advancements and Applications, *IEEE Robot. Automat. Mag.*, 2012, Vol. 19, No. 1, pp. 24–39.
4. Shishkin Iu.E. and Skatkov A.V. Informacionnye tekhnologii obnaruzheniya anomalij v monitoringovyh nablyudeniyah (Information technologies for detecting anomalies in monitoring observations), Simferopol': IT "ARIAL", 2019, 368 p.
5. Wu B., Zhang Y., and Yuan D. Design of embedded software for in-situ monitoring system of marine radioactivity based on a buoy. 2019 Chinese Control Conference (CCC), 2019, pp. 15–19.
6. Shishkin Iu.E. and Skatkov A.V. Intellektual'naya sistema adaptivnogo vybora scenariy parametrov obnaruzheniya divergencij dannyh monitoringa (Intelligent system for adaptive selection of scenarios for parametric detection of monitoring data divergences) *Monitoring systems of environment*, 2019, Vol. 2, No. 36, pp. 37–42.
7. Kovalev S.M. Metody mnogoshagovogo predskazaniya anomalij v temporal'nyh dannyh (Methods for multistep prediction of temporal anomalies) *Izvestiya YUFU Tekhnicheskie nauki*, 2013, Vol. 7, No. 144, pp. 85–91.
8. Grafkin A.V., Grafkin V.V., and Degtyareva O.A. Prikladnoj analiz sluchajnyh processov (Applied analysis of stochastic processes) Samara, 2007, 582 p.
9. Shajtura S.V. Intellektual'nyj analiz geodannyh (Intelligent analysis of geodata). *Perspektivy nauki i obrazovaniya*, 2015, Vol. 6, No. 18, pp. 24–30.
10. Korolyuk V.S. O raskhozhdenii empiricheskikh raspredelenij dlya sluchaya dvuh nezavisimyh vyborok (On the discrepancy between empirical distributions for the case of two independent samples). *Izv. AN SSSR. Seriya matematika*, 1955, Vol. 19, No. 1, pp. 81–96.
11. Bolgov M.V. Sovremennoe sostoyanie teorii korrelyacii dlya gidrologicheskikh raschetov i stohasticheskogo modelirovaniya (The current state of the theory of correlation for hydrological calculations and stochastic modeling). *Vodnoe hozyajstvo Rossii: problemy, tekhnologii, upravlenie*, 2012, Vol. 4, pp. 7–20.
12. Shishkin Iu.E. and Grekov A.N. Statisticheskie metody klasterizacii izobrazhenij gidrobiontov (Statistical methods of clustering images of aquatic organisms) *Monitoring systems of environment*, 2020, Vol. 1, No. 39, pp. 153–159.
13. Kulbak S. Teoriya informacii i statistika (Information theory and statistics). Moscow: Nauka, 1967, 408 p.
14. Hurvich C.M., Tsai C.L. Regression and time series model selection in small samples, *Biometrika*, 1989, Vol. 76, No. 2, pp. 297–307.