

УДК 681.51

ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЕ И ОДНОКЛАССОВАЯ КЛАССИФИКАЦИЯ ДЛЯ МОНИТОРИНГА СОСТОЯНИЯ ДАМБ

А. П. Козионов^{а, б}, инженер, аспирант

А. Л. Пяйт^{а, в}, ведущий инженер, аспирант

И. И. Мохов^а, эксперт-исследователь

Ю. П. Иванов^б, канд. техн. наук, доцент

^аООО «Сименс», департамент корпоративных технологий, Санкт-Петербург, РФ

^бСанкт-Петербургский государственный университет аэрокосмического приборостроения, Санкт-Петербург, РФ

^вУниверситет Амстердама, Амстердам, Нидерланды

Постановка проблемы: мониторинг состояния дамб является важной и актуальной задачей. Алгоритмы мониторинга должны в режиме онлайн по сигналам с аппаратуры контроля определять аномальное состояние дамбы. Под аномальным состоянием (поведением) подразумевается отклонение от нормального состояния объекта, которое может привести к неисправности или разрушению дамбы. Для построения подобных алгоритмов, как правило, требуется знание о нормальном и аномальном режимах поведения дамбы. Зачастую сбор данных, соответствующих аномальному поведению, является высокочастотным, а моделирование аномалий — трудоемким. Цель настоящей работы — исследование и разработка алгоритмов, которые бы обучались на исторических измерениях с дамбы и позволяли определять ее текущее состояние. **Результаты:** представлены подход и алгоритм мониторинга состояния насыпных дамб, а также алгоритм восстановления пропусков в измеряемых сигналах. Идея подхода к обнаружению аномального состояния заключается в применении методов машинного обучения к признакам, извлеченным из сигналов методами частотно-временного анализа последних. Для восстановления пропусков в измерениях, вызванных сбоями системы сбора данных и другими факторами, используется метод «Гусеница»-SSA. Предлагаемый алгоритм обнаружения аномального состояния основан на классификации вейвлет-разложения многомерных измерений. Одноклассовый классификатор «нейронные облака» используется для оценки нелинейной, нечеткой функции принадлежности результатов вейвлет-разложения сигналов к области нормального поведения. Обучение классификатора «нейронные облака» производится на исторических данных о нормальном поведении дамбы, полученных с установленных в ней датчиков. Нечеткий выход классификатора, изменяющийся от 0 до 1, позволяет давать оценку близости текущего состояния дамбы к аномальному. Апробация алгоритмов проводилась на дамбе, расположенной в Бостоне, Великобритания. Была продемонстрирована высокая эффективность алгоритма. **Практическая значимость:** результаты исследований и алгоритмические решения использованы компанией «Сименс» в компоненте на основе искусственного интеллекта при разработке системы мониторинга состояния дамб в рамках проекта Евросоюза UrbanFlood.

Ключевые слова — обнаружение аномалий, мониторинг состояния дамб, интеллектуальная обработка сигналов, вейвлеты, «нейронные облака», одноклассовая классификация.

Введение

В настоящее время более чем две трети европейских городов постоянно находятся под риском затопления. Число зарегистрированных наводнений увеличилось в четыре раза по сравнению с 80-ми годами XX века [1]. Изменение климата и быстрая урбанизация еще сильнее усугубляют проблему.

Большое число наводнений вызвано прорывом систем защиты от наводнений — дамб и плотин [2]. Например, в Нидерландах с 1134 по 2006 год было зарегистрировано примерно 1735 случаев разрушения дамб [3]. В Европе существует несколько проектов, направленных на исследование и разработку систем контроля защиты от наводнений: FLOODsite [4], FloodControl 2015 [5], UrbanFlood [6]. Одним из важнейших элементов подобных систем является система мониторинга состояния дамб.

Исторически мониторинг осуществляется инспекторами, которые занимаются обходом тысяч

километров дамб. Каждый участок проверяется один раз в несколько лет. Современное развитие техники позволяет проводить непрерывный мониторинг состояния дамбы с использованием технологий дистанционных измерений.

Проблема мониторинга состояния земляных дамб заключается в сложном поведении дамбы и, как следствие, измеряемых параметров системы, что требует детального исследования объекта мониторинга и моделирования большого числа возможных сценариев разрушения дамбы. Это весьма затратно и не всегда возможно.

В настоящей работе мы представляем новый подход и алгоритм мониторинга состояния дамб. Предлагаемый подход основан на методах машинного обучения и обработки сигналов и не требует детального исследования объекта мониторинга. Обучение алгоритмов проходит на исторически измеренных данных нормального состояния дамбы. Подход и алгоритм разрабатывались

с участием авторов системы раннего предупреждения UrbanFlood [6]. Апробация алгоритма проводилась на данных с дамбы, расположенной в Бостоне (Великобритания).

Подход к обнаружению аномального поведения дамб

Мониторинг состояния дамбы осуществляется путем анализа измерений, собранных с сети установленных в дамбу различных датчиков, таких как датчики порового давления (давления воды в порах почвы P_p) и температуры, инклинометров. Задача системы мониторинга — обнаружить аномальное состояние дамбы. Под аномальным состоянием (поведением) подразумевается отклонение от нормального состояния объекта, которое может привести к неисправности или разрушению дамбы [7]. Датчики устанавливаются в каждую кросс-секцию дамбы. Такое деление объекта мониторинга позволяет локализовать аномальное состояние и предоставляет избыточность измерений. Применение сетей датчиков как части системы раннего предупреждения в рамках проекта UrbanFlood описано в работах [8, 9].

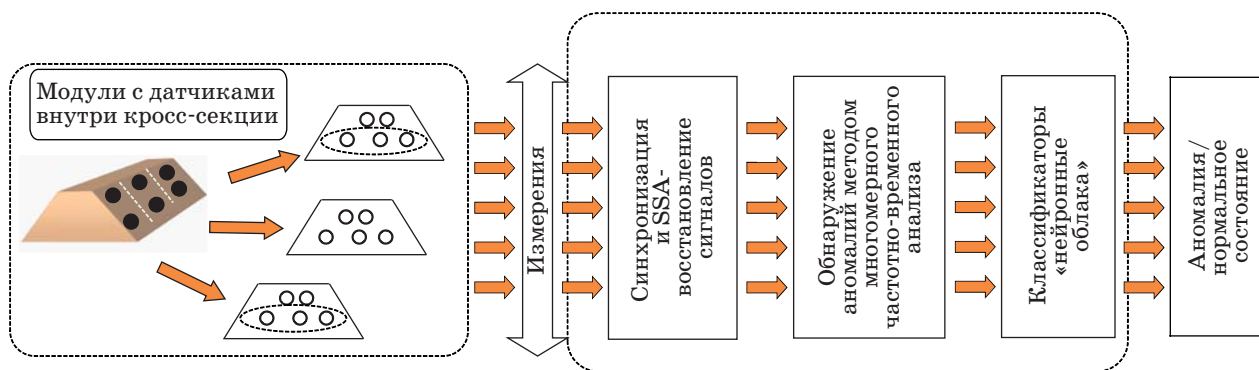
Различные типы измеряемых параметров обрабатываются в соответствии с их свойствами. Например, в случае данных датчиков P_p необходимо учитывать зависимость между измерениями P_p в разных точках дамбы. Для анализа и мониторинга зависимостей между сигналами можно использовать передаточные функции или методы, которые позволяют оценивать границы областей нормального поведения. В первом случае существует проблема устойчивости и идентификации передаточной функции [10]. Алгоритмы, основанные на оценивании допусков для параметров, не имеют проблем, связанных с устойчивостью и идентификацией, но не позволяют учитывать частотно-временные свойства сигналов, а лишь показывают превышение ими

критических значений, поэтому аномальные состояния, проявляющиеся в динамике сигналов, не будут обнаружены.

Из сказанного следует, что необходим алгоритм, который, с одной стороны, позволял бы учитывать частотно-временные зависимости между сигналами, с другой стороны, не имел бы проблем, связанных с устойчивостью и идентификацией.

Основная идея предлагаемого подхода заключается в применении методов обнаружения аномального поведения, которые обучаются на «сырых» и (или) предварительно обработанных данных [11]. При этом данные системы контроля состояний дамб могут содержать пропуски, вызванные сбоями системы сбора данных и другими факторами. Такие пропуски могут иметь протяженность от нескольких минут до нескольких дней. Обнаружение аномального состояния состоит из трех этапов (рис. 1). Первый этап — сбор данных с датчиков. Второй этап — их предварительная обработка, в нашем случае — восстановление пропусков в данных алгоритмом на основе метода «Гусеница»-SSA (Singular Spectrum Analysis — спектрально-сингулярный анализ) [12]. На третьем этапе с помощью алгоритма обнаружения аномального состояния извлекается информация из сигналов методами частотно-временного анализа. Далее алгоритмы классификации, основанные на методах машинного обучения, определяют состояние дамбы.

Поскольку для дамбы зачастую доступны лишь данные о нормальном состоянии, классификацию состояния дамбы логично производить, оценивая степень принадлежности сигналов датчиков к некоей области нормального состояния. Для решения такой задачи применимы алгоритмы одноклассовой классификации, основанные на оценке принадлежности сигнала к одному заданному классу. В нашем случае это область «исторических» измерений, соответствующих нормальному состоянию дамбы.



■ Рис. 1. Этапы обнаружения аномального состояния дамбы

Дискретное вейвлет-преобразование

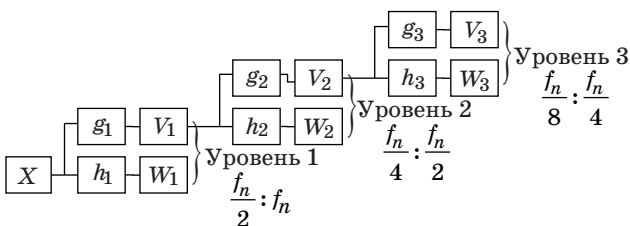
Вейвлет-преобразование (wavelet transform) — это метод частотно-временного (масштабно-временного) — в терминологии теории вейвлетов) разложения сигнала. В отличие от преобразования Фурье вейвлет-преобразование позволяет анализировать только в частотной области [13]. Вейвлет-преобразование делят на два вида — непрерывное (НВП) и дискретное (ДВП).

Непрерывное вейвлет-преобразование задано непрерывными вейвлет-функциями. Масштабы и время, на которых происходит разложение, изменяются непрерывно. Данное преобразование получило широкое применение в научных исследованиях [13].

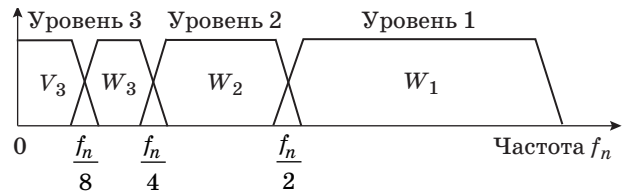
Дискретное вейвлет-преобразование задано семейством дискретных вейвлет-функций и применяется к дискретным сигналам. В случае ДВП масштабы (уровни разложения — в терминологии ДВП), для которых происходит разложение, увеличиваются пропорционально степени двойки, а дискретизация каждого уровня разложения уменьшается в 2 раза по сравнению с предыдущим более высокочастотным уровнем разложения [13]. У ДВП много приложений в естественных науках, инженерном деле, математике (включая прикладную).

Дискретное вейвлет-преобразование получают применением набора фильтров. Сначала сигнал пропускается через низкочастотный (масштабирующий) фильтр. В результате получают коэффициенты аппроксимации (V). Одновременно сигнал пропускается через высокочастотный фильтр (вейвлет-фильтр) — получают вейвлет-коэффициенты (W). Эти два фильтра, называемые квадратурными зеркальными фильтрами, связаны между собой. Так как половина частотного диапазона отфильтрована, то, согласно теореме Котельникова, отсчеты полученных сигналов можно проредить в 2 раза. Это разложение можно повторить несколько раз для коэффициентов аппроксимации каждого последующего разложения. Данный каскадный алгоритм, предложенный Стефаном Малла [13] (рис. 2), можно записать в следующем виде:

$$W_j = (h * V_{j-1}) \downarrow 2, \tag{1}$$



■ Рис. 2. Каскадный алгоритм ДВП: f_n — частота Найквиста



■ Рис. 3. Представление ДВП в частотной области

где W_j — вейвлет-коэффициенты ДВП уровня разложения j ; V_{j-1} — коэффициенты аппроксимации уровня разложения $j - 1$; $*$ — оператор полной свертки; \downarrow — оператор прореживания.

Коэффициенты аппроксимации

$$V_j = (h * V_{j-1}) \downarrow 2, \tag{2}$$

где V_j — коэффициенты аппроксимации ДВП уровня разложения j ; V_{j-1} — коэффициенты аппроксимации уровня разложения $j - 1$; $*$ — оператор полной свертки; \downarrow — оператор прореживания.

Поскольку квадратурные зеркальные фильтры ДВП на каждом уровне разложения разделяют сигнал на низкочастотную и высокочастотную часть, то полоса частот коэффициентов аппроксимации каждого уровня разложения (по отношению к исходному сигналу) будет находиться в области $\left[0; \frac{f_n}{2^j}\right]$, а полоса частот вейвлет-коэффициентов — в области $\left[\frac{f_n}{2^j}; \frac{f_n}{2^{j-1}}\right]$, где f_n в 2 раза ниже частоты дискретизации сигнала; j — уровень разложения. На рис. 3 представлены уровни ДВП в частотной области.

В отличие от НВП, ДВП рассчитывается на масштабах 2^j , где j называется уровнем разложения, что позволяет рассчитывать коэффициенты ДВП быстрее, чем в случае НВП, за $n \log_2(n)$ операций, где n — длина временного ряда. В силу двукратного прореживания длина сигнала должна быть кратна $2^{J_{\max}}$, где J_{\max} — число уровней разложения.

Также стоит отметить ортогональность вейвлет и масштабирующих фильтров к полиномам заданной степени. Это означает, что если во временном ряде присутствует такой полином, то он будет оставаться в последнем уровне аппроксимации.

Алгоритм «нейронные облака»

К одним из наиболее быстроразвивающихся и перспективных методов классификации относятся методы на основе машинного обучения. Для обучения классификатора используются лишь данные о нормальном поведении сигнала, данные о видах аномального поведения недоступны. В качестве такого метода в работе рассматривается классификатор «нейронные облака» [14].

Наиболее часто для задач обнаружения аномалий используют различные метрики расстояний (евклидову, Махаланобиса и др.) [15], методы на основе оценки плотности вероятности данных, например гауссовы смеси (Gaussian Mixture Models) [15], которые оценивают плотность вероятности для данных, соответствующих нормальному состоянию. Общий недостаток этих методов в том, что они сами по себе не являются классификаторами, и их результат все равно нужно каким-то образом классифицировать [16]. Одноклассовый метод опорных векторов (one-class SVM) [15] обладает другим недостатком, заключающимся в том, что он оценивает четкую границу класса, которую нельзя перенастроить (скорректировать) во время использования классификатора, а лишь только при новом переобучении.

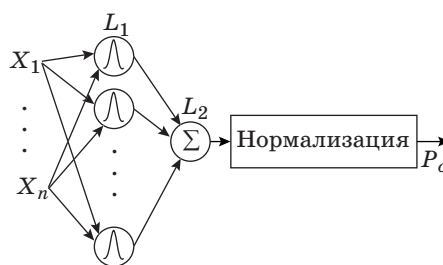
«Нейронные облака» оценивают нелинейную, нечеткую функцию принадлежности (инкапсулятор) данных к области нормального поведения для исходного «нормального» множества признаков. Затем функция принадлежности используется для оценки принадлежности новых экземпляров данных к области нормального поведения. Результаты «нейронных облаков» — значения от 0 до 1. Значения, близкие к 1, отражают нормальное поведение, а близкие к 0 — аномальное. Таким образом, «нейронные облака» позволяют давать оценку близости текущего состояния к аномальному, что может быть использовано при принятии решения.

Оценка функции принадлежности состоит из двух шагов: кластеризации данных методом АКМ (advanced k-means, улучшенный k-средних) и аппроксимации кластеров радиальными базисными функциями (РБФ). Смесь РБФ можно представить в виде нечеткой нейронной сети.

Метод АКМ — это модификация хорошо известного метода k-средних с возможностью адаптивного расчета оптимального числа кластеров при заданном максимальном числе кластеров (центроидов). АКМ состоит из следующих этапов:

- 1) задаются начальные максимальное и минимальное числа центроидов;
- 2) вызывается алгоритм k-средних;
- 3) добавляется или убирается центроид, исходя из следующих условий:
 - если дистанция до данных выше заданной, то создается новый центроид;
 - если кластер состоит из числа точек, меньше заданного, то он убирается;
 - если расстояние между центроидами меньше заданного, то кластеры объединяются.

Выход алгоритма АКМ — это центроиды кластеров, представляющих собой нормальное поведение (обучающее множество). После того как все центроиды получены, данные инкапсулируются



■ Рис. 4. «Нейронные облака»: L_1 — РБФ; L_2 — суперпозиции РБФ; P_c — значение функции принадлежности

функцией принадлежности. Для данных задачи используются РБФ

$$R_i = e^{-\frac{|x-m_i|}{2\sigma^2}},$$

где x — входные данные; m_i — среднее (центр) РБФ; σ — среднеквадратическое отклонение (СКО) (ширина) РБФ.

Центроиды, полученные от АКМ, становятся центрами соответствующих РБФ. Сумма всех колоколов дает функцию принадлежности (рис. 4). Нормализация применяется для того, чтобы получить выходное значение функции принадлежности в интервале от 0 до 1.

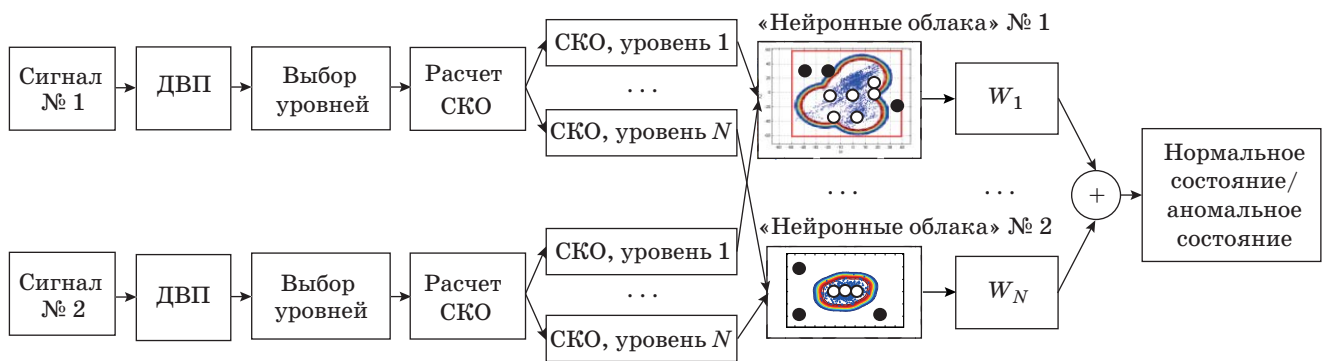
Алгоритм обнаружения аномального состояния дамбы на основе многомерного частотно-временного анализа сигналов

Предлагаемый алгоритм основан на одноклассовой классификации результатов частотно-временного преобразования анализируемых сигналов. Для частотно-временного преобразования (разложения) сигналов было выбрано ДВП. Для одноклассовой классификации используется алгоритм «нейронные облака». Блок-схема алгоритма обнаружения аномального состояния представлена на рис. 5.

На первом этапе к сигналам датчиков, установленных на дамбе, применяется ДВП и производится выбор наиболее мощных (базовых) уровней для анализа с целью сократить число классификаторов. Далее выполняется дополнительная обработка вейвлет-коэффициентов, направленная на расчет локальных СКО во времени, и коэффициентов для каждого уровня разложения, характеризующих мощность каждого уровня.

На втором этапе для одинаковых уровней разложения сигналов строятся совместные функции принадлежности к области нормального поведения для исходных «нормальных» исторических значений локальных СКО уровней разложения с использованием алгоритма «нейронные облака».

На третьем этапе выход «нейронного облака» каждого уровня разложения нормируется в соот-



■ Рис. 5. Алгоритм обнаружения аномального поведения на основе классификации результатов ДВП

ветствии с мощностным вкладом данного уровня разложения в суммарную мощность отобранных для анализа уровней ДВП — $W_1 \dots W_N$ (см. рис. 5). Такая нормировка необходима для комбинирования выходов различных классификаторов. Таким образом, состояние определяется как суммарный взвешенный вклад от каждого уровня разложения. В итоге на выходе алгоритма значение изменяется от 0 до 1, где близость к 0 соответствует аномальному, а близость к 1 — нормальному состоянию.

Восстановление пропусков в измерениях

Данные системы контроля состояний дамб содержат пропуски, вызванные сбоями системы сбора данных и другими факторами. Пропуски могут иметь протяженность от нескольких часов до нескольких дней. Для восстановления данных используется метод «Гусеница» [11]. Он позволяет восстанавливать многие виды сигналов с R^2 до 0,99.

Стоит также отметить такие методы восстановления пропусков в данных, как 3D-VAR и 4D-VAR [17], которые используются для восстановления метеорологических данных. Недостаток этих методов заключается в том, что они требуют предварительно известных моделей измерений в пространстве состояний, в то время как для метода «Гусеница» заранее предопределенная модель не требуется.

Идея метода состоит в постепенном разложении сигнала на компоненты по базису собственных векторов ковариационной матрицы сигнала. Далее над компонентами проводится необходимое преобразование, после которого происходит процедура восстановления сигнала. Процедура восстановления сигнала состоит из следующих шагов.

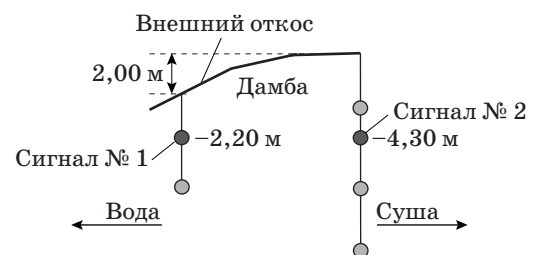
1. Разложение сигнала.
2. Выбор числа компонент, по которым будет производиться восстановление.
3. По каждой компоненте, начиная с 1-й до последней (выбранной на шаге 2), производится

разложение. После производится восстановление с 1-й компоненты по текущую в цикле. Значения, полученные на местах пропусков, вставляются в исходный сигнал. Данная процедура выполняется до тех пор, пока не будет достигнута сходимость в местах пропусков. Иными словами, результат предыдущего восстановления не будет отличаться от текущего (в качестве критерия, к примеру, можно выбрать дисперсию разности результатов между текущим и предыдущим шагом, отнесенную к дисперсии результата восстановления на текущем шаге). После обеспечения сходимости цикл выполняется для следующей компоненты.

Моделирование алгоритма

В настоящей работе мы будем рассматривать обнаружение реальной аномалии дамбы в Бостоне. В дамбу была установлена сеть многопараметрических датчиков GeoBeads компании Alert Solutions [18]. Датчики измеряют поровое давление P_p , температуру почвы и отклонение от вертикали. Расположение датчиков показано на рис. 6. Дискретизация данных составляет 15 мин. Первые результаты анализа данных были представлены в статье [10].

Локальное сползание внешнего откоса дамбы в результате эрозии почвы произошло 12 января 2012 г. Аномалия проявилась в измерениях датчика порового давления — сигнал № 1 (см. рис. 6). Этот датчик установлен примерно



■ Рис. 6. Структура одной из кросс-секций дамбы

на одной глубине с другим датчиком порового давления (сигнал № 2), расположенным дальше от воды, на котором аномальное состояние не проявилось. Далее будет анализироваться зависимость между этими измерениями.

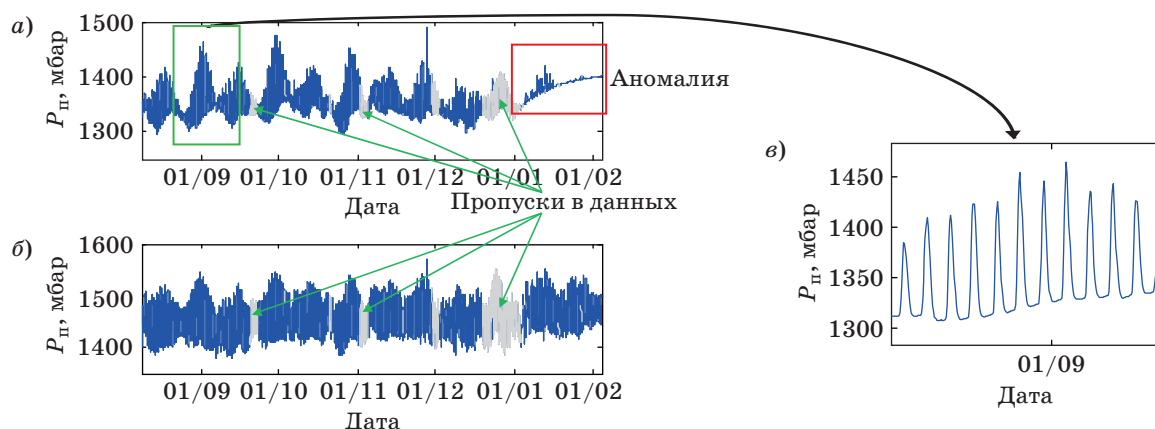
Аномальное поведение проявилось как демпфирование дамбой колебаний порового давления, вызванных колебаниями воды (рис. 7, а). В измерениях присутствуют пропуски длительностью от нескольких минут до нескольких дней (рис. 7, а, б), вызванные сбоями системы сбора данных или другими факторами. Наиболее мощные колебания порового давления вызваны приливами и отливами с периодичностью 12 ч (рис. 7, в).

Для обучения и апробации алгоритма в местах пропусков сигналы были восстановлены методом «Гусеница» [11] (рис. 8, а, б).

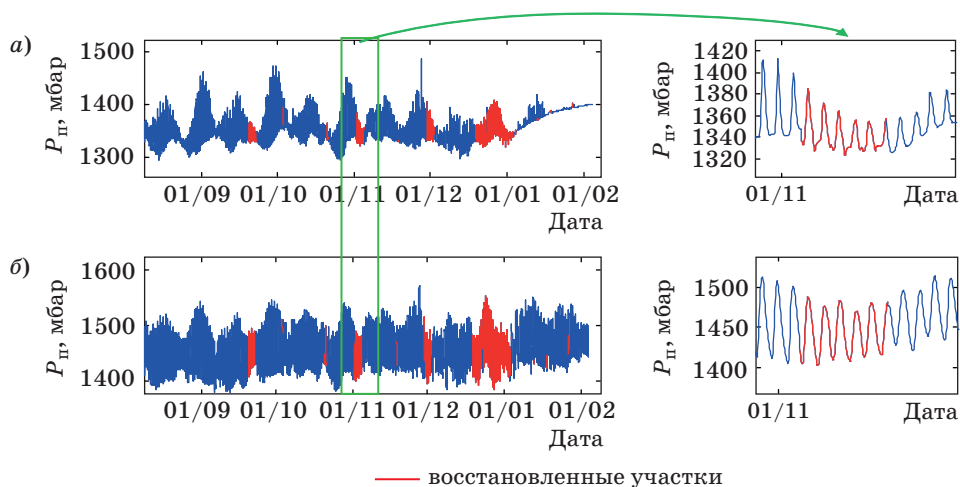
Обучающее множество для алгоритма — первые 50 % отсчетов сигналов с 8 августа 2011 г.

по 10 ноября 2011 г. Для ДВП был выбран вейвлет Добеши 4, так как в уровнях разложения ДВП, полученных с использованием такого вейвлета, исключается тренд, описываемый полиномом порядка ниже 4-го [13]. Число уровней разложения 6 выбрано так, чтобы с запасом перекрыть диапазон наиболее мощных частот сигнала, 12-часовые колебания будут содержаться в 4-м уровне разложения.

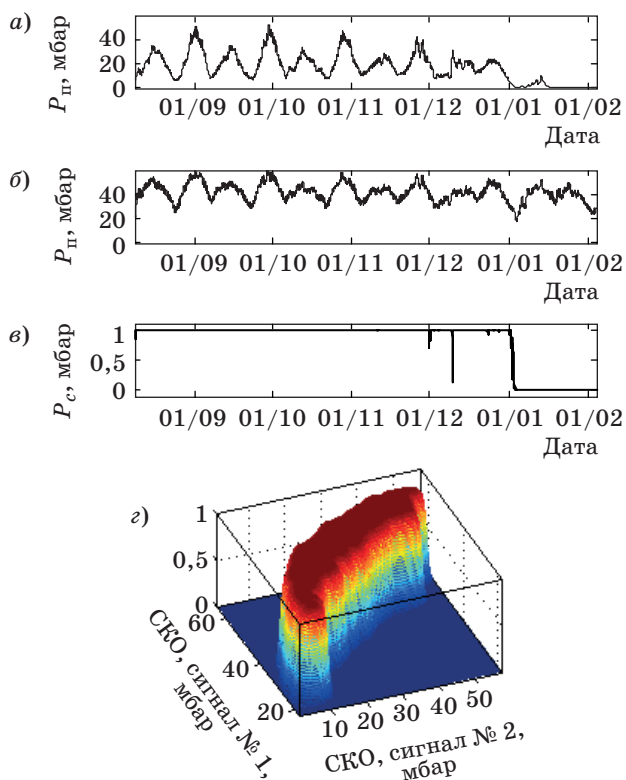
Из шести уровней разложения для анализа в качестве базовых были выбраны 3-, 4- и 5-й уровни, так как на них пришлось 16, 64 и 15 % мощности сигнала. Суммарная мощность трех уровней 86 %. Для каждого уровня разложения были рассчитаны локальные СКО. Далее попарно для каждого уровня разложения были построены функции принадлежности алгоритмом «нейронные облака». В итоге получились три двумерных «нейронных облака» (по двум сигналам) для 3-, 4- и 5-го уровней разложений.



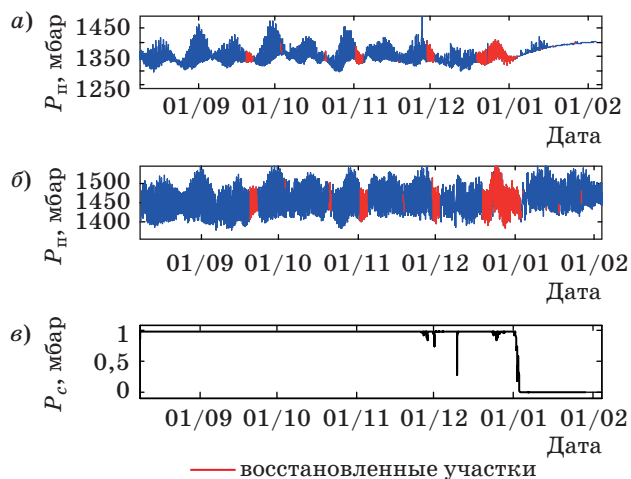
■ Рис. 7. Сигналы для моделирования: а — сигнал № 1; б — сигнал № 2; в — укрупненный участок сигнала № 1



■ Рис. 8. Восстановление сигналов методом «Гусеница»: а — сигнал № 1; б — сигнал № 2; справа показаны участки в укрупненном масштабе



■ **Рис. 9.** Анализ алгоритмом «нейронные облака» уровня разложения ДВП № 4: *a* — локальные СКО, сигнал № 1; *b* — локальные СКО, сигнал № 2; *в* — значение функции принадлежности; *г* — «нейронное облако»



■ **Рис. 10.** Восстановленные сигналы и результирующее значение функции принадлежности: *a* — сигнал № 1; *b* — сигнал № 2; *в* — значение функции принадлежности

Рассмотрим классификацию 4-го уровня разложения — наиболее мощного. Для сигнала № 1 видно демпфирование колебаний порового давления после 1 января и, как следствие, падение СКО, вызванное изменением свойств дамбы в ре-

зультате эрозии (рис. 9, *a*). Для сигнала № 2 anomalous behavior did not manifest (рис. 9, *b*). Значение функции принадлежности на участке после 2 января падает к 0, что соответствует аномальному состоянию на границе функции принадлежности (рис. 9, *в*). Функция принадлежности представлена на рис. 9, *г*. Таким образом, эрозия почвы была обнаружена на 10 дней раньше сползания откоса дамбы. Также присутствует локальный участок 7 декабря, в котором значение функции принадлежности падает до 0,1. Данный участок соответствует ложному срабатыванию, которое вызвано отсутствием подобного поведения в обучающем множестве (см. рис. 9, *a*). Подобное поведение характерно для 3-го и 5-го уровней разложения.

Восстановленные для анализа сигналы № 1 и № 2 представлены на рис. 10, *a* и *б* соответственно.

Суммарный взвешенный вклад всех уровней разложения тоже стремится к 0 после 2 января (рис. 10, *в*). 7 декабря присутствует локальный участок ложного срабатывания, в котором значение функции принадлежности падает к 0,15, что вызвано отсутствием такого поведения в обучающем множестве.

Заключение

Предложенный подход к мониторингу состояния дамб основан на использовании методов машинного обучения и обработки сигналов. Суть алгоритма заключается в применении одноклассовой классификации алгоритмом «нейронные облака» результатов ДВП сигналов с датчиков системы контроля состояния дамб. Обучение алгоритма «нейронные облака» проводится на исторических данных о нормальном состоянии дамбы.

Преимущества такого подхода состоят в том, что для его применения не требуется подробного описания структуры и аномального поведения объекта. Кроме того, система мониторинга на его основе сама подстраивается и обучается с момента установки на объект.

Моделирование алгоритма обнаружения аномального состояния, построенного на основе предлагаемого подхода, было проведено на реальных данных дамбы Бостона. Сползание откоса дамбы было зафиксировано на 10 дней раньше визуального обнаружения как аномальное поведение измерений датчиков порового давления. Преимуществом алгоритма является устойчивый контроль частотно-временных зависимостей между сигналами. К недостаткам можно отнести то, что при отсутствии в обучающих данных некоторых примеров нормального поведения алгоритм может ложно классифицировать их как аномальное состояние. Увеличение обучающего множества будет вести к улучшению результатов классификации.

Также было проведено моделирование и показана приемлемость алгоритма восстановления измеряемых сигналов на основе метода «Гусеница»-SSA. Восстановление данных необходимо для применения предложенного ДВП к сигналам.

Представленные подход и алгоритм могут быть применены и в других областях, таких как системы контроля состояния конструкций, медицинские приложения, системы вибродиагностики и др.

Работа выполнена при поддержке гранта Евросоюза FP7, проект UrbanFlood, грант № 248767.

Литература

1. **Flood** — Data and Statistics. <http://www.preventionweb.net/english/hazards/statistics/> (дата обращения: 22.01.2014).
2. **Wohl E.** Inland Flood Hazards: Human, Riparian, and Aquatic Communities. — Cambridge, Cambridge University Press, 2000. — 518 p.
3. **Baars S., Kempen M.** The Causes and Mechanisms of Historical Dike Failures in the Netherlands. — Official Publication of the European Water Association, 2009. — 14 p.
4. **FLOODsite** project. <http://www.floodsite.net/> (дата обращения: 28.01.2014).
5. **FloodControl** 2015. <http://www.floodcontrol2015.com/> (дата обращения: 28.01.2014).
6. **UrbanFlood** project. <http://www.urbanflood.eu> (дата обращения: 28.01.2014).
7. **Isermann R.** Fault-Diagnosis Systems: an Introduction from Fault Detection to Fault Tolerance. — Berlin, Berlin Springer, 2006. — 475 p.
8. **Krzhizhanovskaya V. V.** et al. Flood Early Warning System: Design, Implementation and Computational Modules // *Procedia Computer Science*. 2011. N 4. P. 106–115. doi:10.1016/j.procs.2011.04.012
9. **Simm J.** et al. Interpreting Sensor Measurements in Dikes — Experiences from Urbanflood Pilot Sites // *Comprehensive Flood Risk Management / Klijn & Schweckendiek (eds.)*, 2012. P. 327–336. doi:10.1201/b13715-47
10. **Pyayt A. L.** et al. Data-Driven Modelling for Flood Defence Structure Analysis // *Comprehensive Flood Risk Management / Klijn & Schweckendiek (eds.)*, 2013. P. 301–306.
11. **Pyayt A. L.** et al. Machine Learning Methods for Environmental Monitoring and Flood Protection // *World Academy of Science, Engineering and Technology*. 2011. N 5. P. 118–123.
12. **Kondrahov D., Shprits Y., Ghil M.** Gap Filling of Solar Wind Data by Singular Spectrum Analysis // *Geophysical Research Letters*. 2010. N 37. P. 1–6. doi:10.1029/2010gl044138
13. **Mallat S.** Wavelet Tour of Signal Processing. — San Diego, CA: Academic Press, 1999. — 832 p.
14. **Lang B.** et al. Neural Clouds for Monitoring of Complex Systems // *Optical Memory and Neural Networks*. 2008. Vol. 3. N 17. P. 183–192. doi:10.3103/S1060992X08030016
15. **Bishop C.** Pattern Recognition and Machine Learning. — N. Y.: Springer, 2006. — 738 p.
16. **Yousri N. A., Ismal M. A., Kamel M. S.** Fuzzy Outlier Analysis a Combined Clustering-outlier Detection Approach // *Systems, Man and Cybernetics*. 2007. P. 412–418. doi:10.1109/ICSMC.2007.4413873
17. **Lorenc A. C., Rawlins F.** Why Does 4D-Var Beat 3D-Var? // *Quarterly J. of the Royal Meteorological Society*. 2005. N 131(613). P. 3247–3257. doi:10.1256/qj.05.85
18. **Alert Solutions.** <http://www.alertsolutions.nl/?> (дата обращения: 28.01.2014).

UDC 681.51

Wavelet Transform and One-Class Classification for Dike Health Monitoring

A. P. Kozionov^{a, b}, R&D Engineer, Post-Graduate Student, alexey.kozionov@siemens.com

A. L. Pyayt^{a, c}, R&D Leading Engineer, Post-Graduate Student, alexander.pyayt@siemens.com

I. I. Mokhov^a, R&D Expert, ilya.mokhov@siemens.com

U. P. Ivanov^b, PhD, Tech., Associate Professor, upi@mail.ru

^aSiemens, Corporate Technology, 3A, Volynskii St., 191186, Saint-Petersburg, Russian Federation

^bSaint-Petersburg State University of Aerospace Instrumentation, 67, B. Morskaya St., 190000, Saint-Petersburg, Russian Federation

^cUniversity of Amsterdam, Science Park 904, 1098 XH, Amsterdam, Netherlands

Purpose: Dike health monitoring is a challenging task. The monitoring algorithms must detect abnormal behaviour of a dike online, by signals from remote sensors. Abnormal behaviour is a deviation from normal behaviour, able to cause a failure of the dike or its demolition. Development of these algorithms usually involves knowledge about abnormal and normal dike behaviour. Obtaining the data about abnormal behaviour is often expensive, and the simulation is time-consuming. The goal of this paper is studying and developing algorithms which can learn from historical data of the dike behaviour and detect the current dike state. **Results:** The paper presents an approach and algorithm for dike health monitoring, along with a method of filling gaps in the measurements. The gap-filling algorithm is based on Spectrum Singular Analysis (SSA). To detect abnormal behaviour, machine learning methods are applied

to the characteristics extracted from time-frequency decomposition of the signals. The proposed algorithms are based on classification of wavelet decomposition of multidimensional signals. A one-class classifier “Neural Clouds” is used to estimate the nonlinear fuzzy membership function that checks whether the wavelet decomposition features are within the normal behaviour area. The classifier is trained on the historical data about normal dike behaviour. Fuzzy classifier response changing from 0 to 1 gives an estimation of how close the current state of the dike is to abnormal behaviour. The algorithms have been tested on a dike in Boston, UK, and showed high efficiency. **Practical relevance:** the proposed approach and algorithms were used by Siemens in its AI component which is a part of UrbanFlood Early Warning System.

Keywords – Anomaly Detection, Dike Health Monitoring, Intelligent Signal Processing, Wavelets, Neural Clouds, One-Class Classification.

References

1. *Flood — Data and Statistics*. (2014). Available at: <http://www.preventionweb.net/english/hazards/statistics/> (accessed 22 January 2014).
2. Wohl E. *Inland Flood Hazards: Human, Riparian, and Aquatic Communities*. Cambridge, Cambridge University Press, 2000. 518 p.
3. Baars S., Kempen M. *The Causes and Mechanisms of Historical Dike Failures in the Netherlands*. Official Publication of the European Water Association, 2009. 14 p.
4. *FLOODsite project*. (2014). Available at: <http://www.floodsite.net/> (accessed 28 January 2014).
5. *FloodControl 2015*. (2014). Available at: <http://www.floodcontrol2015.com/> (accessed 28 January 2014).
6. *UrbanFlood project*. (2014). Available at: <http://www.urbanflood.eu> (accessed 28 January 2014).
7. Isermann R. *Fault-Diagnosis Systems: an Introduction from Fault Detection to Fault Tolerance*. Berlin, Berlin Springer, 2006. 475 p.
8. Krzhizhanovskaya V. V., Shirshov G. S., Melnikova N. B., Belleman R. G., Rusadi F. I., Broekhuijsen B. J., Gouldby B. P., Lhomme J., Balis B., Bubak M., Pyayt A. L., Mokhov I. I., Ozhigin A. V., Lang B., Meijer R. J. Flood Early Warning System: Design, Implementation and Computational Modules. *Procedia Computer Science*, 2011, vol. 4, pp. 106–115. doi:10.1016/j.procs.2011.04.012
9. Simm J., Jordan D., Topple A., Mokhov I., Pyayt A., Abdoun T., Bennett V., Broekhuijsen J., Meijer R. Interpreting Sensor Measurements in Dikes — Experiences from Urban-flood Pilot Sites. In: *Comprehensive Flood Risk Management*, Klijn & Schweckendiek (eds.), 2012, pp. 327–336. doi:10.1201/b13715-47
10. Pyayt A. L., Mokhov I. I., Kozionov A. P., Kuserbaeva V. T., Lang B., Krzhizhanovskaya V. V., Meijer R. J. Data-Driven Modelling for Flood Defence Structure Analysis. In: *Comprehensive Flood Risk Management*, Klijn & Schweckendiek (eds.), 2013, pp. 301–306.
11. Pyayt A. L., Mokhov I. I., Lang B., Krzhizhanovskaya V. V., Meijer R. J. Machine Learning Methods for Environmental Monitoring and Flood Protection. *World Academy of Science, Engineering and Technology*, 2011, vol. 5, pp. 118–123.
12. Kondrashov D., Shprits Y., Ghil M. Gap Filling of Solar Wind Data by Singular Spectrum Analysis. *Geophysical Research Letters*, 2010, vol. 37, pp. 1–6. doi:10.1029/2010gl044138
13. Mallat S. *Wavelet Tour of Signal Processing*. San Diego, CA, Academic Press, 1999. 832 p.
14. Lang B., Poppe T., Minin A., Mokhov I., Kuperin Y., Mekler A., Liapakina I. Neural Clouds for Monitoring of Complex Systems. *Optical Memory and Neural Networks*, 2008, vol. 3, no. 17, pp. 183–192. doi:10.3103/S1060992X08030016
15. Bishop C. *Pattern Recognition and Machine Learning*. New York, Springer, 2006. 738 p.
16. Yousri N. A., Ismal M. A., Kamel M. S. Fuzzy Outlier Analysis a Combined Clustering-outlier Detection Approach. *Systems, Man and Cybernetics*, 2007, pp. 412–418. doi:10.1109/ICSMC.2007.4413873
17. Lorenc A. C., Rawlins F. Why Does 4D-Var Beat 3D-Var? *Quarterly Journal of the Royal Meteorological Society*, 2005, vol. 131, no. 613, pp. 3247–3257. doi:10.1256/qj.05.85
18. *Alert Solutions*. (2014). Available at: <http://www.alertsolutions.nl/> (accessed 28 January 2014).