



DOI: 10.5862/JCSTCS/8

УДК 681.51

А.П. Козионов, А.Л. Пяйт, И.И. Мохов, Ю.П. Иванов

ИССЛЕДОВАНИЕ АЛГОРИТМОВ ВОССТАНОВЛЕНИЯ ПРОПУСКОВ В ИЗМЕРЯЕМЫХ СИГНАЛАХ ДЛЯ СИСТЕМЫ МОНИТОРИНГА СОСТОЯНИЯ ДАМБ

A.P. Kozionov, A.L. Pyayt, I.I. Mokhov, Yu.P. Ivanov

RESEARCH ON GAP-FILLING ALGORITHMS FOR DIKE HEALTH MONITORING SYSTEMS

Мониторинг состояния дамб является важной и актуальной задачей. Алгоритмы мониторинга должны в режиме онлайн по сигналам с аппаратуры контроля определять anomalous состояние дамбы. Одна из серьезных проблем системы мониторинга состояния дамб, влияющая на возможность применения алгоритмов мониторинга — качество измерений датчиков. В данных встречаются пропуски, вызванные сбоями, отключениями систем передачи или сбора данных, неправильными настройками систем и другими внешними и внутренними факторами, поэтому необходимы алгоритмы улучшения качества измерений.

Предложен подход, основанный на использовании адаптивных алгоритмов восстановления пропусков в сигналах в условиях априорной неопределенности моделей сигналов. Представлены алгоритмы (на основе модели авторегрессии, метода «Гусеница-SSA», преобразования Фурье), использующие исторические данные для восстановления сигналов. Приведено описание и сравнительный анализ точности восстановления пропусков алгоритмами. Апробация алгоритмов проведена на реальных данных с дамбы, расположенной в городе Бостон, Великобритания. Результаты исследований и алгоритмические решения использованы компанией Siemens в системе мониторинга состояния дамб в рамках проекта Евросоюза UrbanFlood.

ВОССТАНОВЛЕНИЕ ПРОПУСКОВ В СИГНАЛАХ; АВТОРЕГРЕССИЯ; ГУСЕНИЦА; СИНГУЛЯРНЫЙ СПЕКТРАЛЬНЫЙ АНАЛИЗ; ПРЕОБРАЗОВАНИЕ ФУРЬЕ; МОНИТОРИНГ СОСТОЯНИЯ ДАМБ; ИНТЕЛЛЕКТУАЛЬНАЯ ОБРАБОТКА СИГНАЛОВ.

Earthen dam (dike/levee) health monitoring is a challenging task. Monitoring algorithms have to detect anomalies in dike behavior in on-line mode basing on measurement collected from the sensors installed in the dike. One of the important monitoring-related challenges for dike health monitoring is the quality of sensors measurements. There are often gaps that occur due to failures, outages of transmission or data collection systems, incorrect configurations of systems and other internal and external factors, therefore it is necessary to improve the quality measurement using specific algorithms. The proposed approach is based on adaptive algorithms filling in gaps in sensor measurements in the conditions of a priori uncertainty of signals models. The algorithms (based on the autoregressive model, Caterpillar-SSA, Fourier transform) presented in this paper use historical data for signal reconciliation. The description and analysis of the algorithms are also included in the paper. The algorithms have been tested at the Boston dike, Great Britain. The research findings and algorithms are implemented by Siemens in the UrbanFlood Early Warning System.

GAP-FILLING; MEASUREMENT RECONCILIATION; AUTOREGRESSIVE MODEL; SPECTRUM SINGULAR ANALYSIS; FOURIER TRANSFORM; DIKE HEALTH MONITORING; INTELLIGENT SIGNAL PROCESSING.

Мониторинг состояния дамб — важная и актуальная задача. Алгоритмы монито-

ринга состояния дамб должны в режиме онлайн по сигналам с аппаратуры контроля

определять аномальное состояние дамбы. Для оценки состояния дамб применяются алгоритмы машинного обучения, анализирующие частотно-временные свойства сигналов (вейвлет-преобразование, оконное преобразование Фурье и т. д.) [1–4]. Для корректного функционирования алгоритмов необходимы данные, не содержащие пропусков в измерениях, иначе результат обучения алгоритмов и оценки состояния дамбы будут неверными. Одной из серьезных проблем систем мониторинга состояния дамб является качество измерений датчиков, вызванное сбоями, отключениями систем передачи или сбора данных, неправильными настройками систем и другими внешними или внутренними факторами, поэтому необходимы алгоритмы улучшения качества измерений. Восстановление пропусков измерений представляет собой наиболее сложную задачу проблематики улучшения качества измерений. Настоящая проблема актуальна не только для рассматриваемой системы, но и для других систем, имеющих дело с полевыми измерениями: систем водоснабжения [5], метеорологических систем [6] и т. д.

Пропуски в измерениях могут варьиро-

ваться от нескольких минут до нескольких дней. Для обучения алгоритмов мониторинга дамб (алгоритмов на основе машинного обучения) необходимы исторические данные. Алгоритмы восстановления пропусков должны оценивать пропущенные значения сигналов с учетом частотно-временных зависимостей, присутствующих в сигналах. Примеры сигналов (условно, «сигнал А» и «сигнал Б») порового давления (давления воды в порах почвы) с дамбы и пропусков представлены на рис. 1 (дамба в городе Бостон, Великобритания). Периодические колебания соответствуют периодам приливов и отливов.

В настоящей статье мы представляем подход к решению задачи восстановления пропусков в измеряемых сигналах. Предлагаемый подход основан на использовании адаптивных алгоритмов восстановления пропусков в сигналах в условиях априорной неопределенности моделей сигналов. Представленные алгоритмы используют исторические данные для восстановления сигналов. Настоящий подход исследовался в [1–3] для системы раннего предупреждения UrbanFlood [7]. Нами также представлен алгоритм на основе модели авторе-

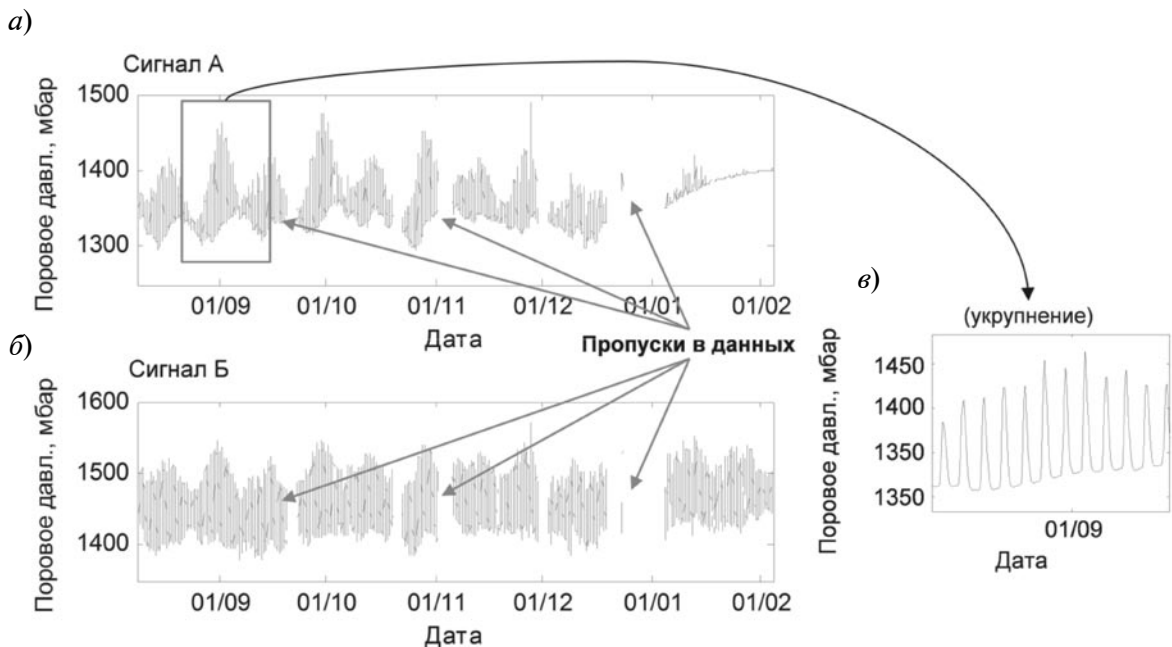


Рис. 1. Примеры сигналов и пропусков в сигналах, формат даты ДД/ММ, 2011-2012 гг.: а – «сигнал А»; б – «сигнал Б»; в – укрупненный участок «сигнала А»

грессии для восстановления пропусков в сигналах. Апробация алгоритмов проводилась на основе данных мониторинга дамбы, расположенной в городе Бостон.

Существует несколько способов восстановления пропусков в сигналах: сплайны, линейная интерполяция, интерполяционные фильтры [8]. Общий недостаток этих методов в том, что они выполняют интерполяцию без восстановления частотно-временных свойств сигналов на участках пропусков. В работе [9] представлен алгоритм восстановления на основе разложения по эмпирическим модам, которое не является стабильным и может привести к неудовлетворительному результату еще на этапе разложения сигнала [10]. При использовании алгоритмов на основе нейронных сетей возникает проблема обучения модели и потери стабильности результатов восстановления [6]. Стоит также отметить алгоритмы 3d-var и 4d-var [11], используемые в метеорологии, для которых требуются априорно заданные модели сигналов и помех. В случае применения алгоритмов многомерного восстановления сигналов накладываются ограничения, связанные с зависимостью процедуры восстановления от доступности других данных (сигналов). Одним из наиболее интересных алгоритмов адаптивного восстановления данных является алгоритм на основе метода «Гусеница-SSA», представленный в статье [12]. Также интересные результаты показаны при использовании алгоритма на основе преобразования Ломба–Скаргла (преобразование Фурье для сигналов с неравномерной дискретизацией)

[13]. Эти алгоритмы могут применяться к одномерным сигналам и имеют высокую точность восстановления.

Алгоритм синхронизации и восстановления пропусков в измеряемых сигналах

Алгоритм синхронизации и восстановления пропусков в измеряемых сигналах состоит из четырех основных этапов (рис. 2).

Этап 1 – проверка измерений (валидация) и обнаружение пропусков. Под ошибками измерений подразумеваются выбросы в измерениях, различные специфические ошибки, например, последовательности нулей, специальные значения измерений, соответствующие сбоям датчиков и т. д. Алгоритм обнаружения пропусков проверяет измерения на наличие пропущенных значений путем сравнения временных дискретов $\Delta_i = t_i - t_{i-1}$ каждого измерения с задаваемой величиной дискретизации ЕВС. Если Δ_i превышает величину дискретизации ЕВС, то отмечается пропуск между измерениями.

Этап 2 – интерполяция сигналов на единую временную сетку (ЕВС). На этом этапе выполняется интерполяция измерений с разной частотой дискретизации на единую временную сетку. Требуемая величина дискретизации ЕВС задается исходя из условий эксплуатации.

Этап 3 – восстановление пропусков. Алгоритмы восстановления пропусков оценивают пропущенные значения измеряемых сигналов в местах пропусков, которые могут варьироваться от одного пропущенного

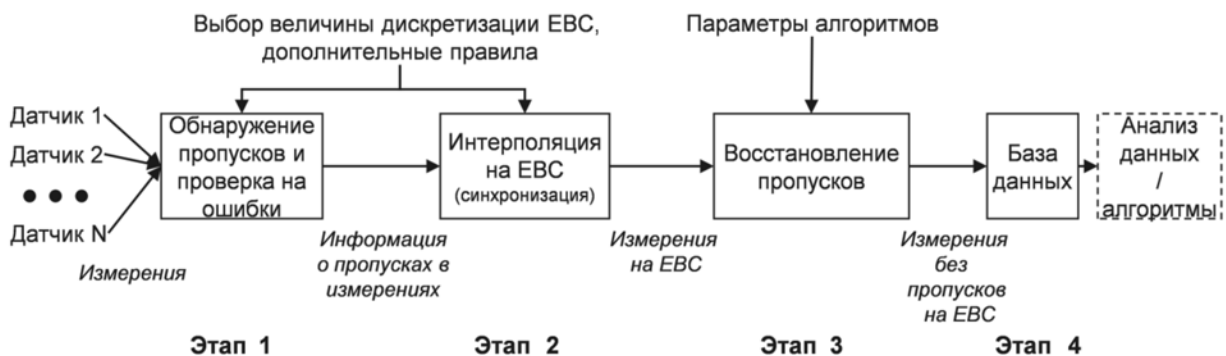


Рис. 2. Алгоритм синхронизации измерений и восстановления пропусков в измеряемых сигналах

измерения до сотен, в зависимости от типа измерений. Этот этап является наиболее сложным, и далее в статье будут исследоваться алгоритмы этого этапа.

Этап 4 – запись результатов в базу данных. Восстановленные и синхронизированные сигналы записываются в базу данных, места восстановленных значений помечаются специальным флагом в базе данных.

Постановка задачи восстановления пропусков измеряемых сигналов

После этапа синхронизации сигналов формируются измерения с постоянной частотой дискретизации, описываемые временными рядами $X = \{X(t), t = T\{0 \dots N\}\}$, где T – множество отсчетов времени, $X \in R$.

Значения в некоторых отсчетах времени t отсутствуют (пропуски). Необходимо найти оценку $\hat{X}(t)$ значений сигнала $X(t)$ в местах пропусков. Априорно модель сигнала неизвестна. Имеются исторические записи сигналов (временных рядов).

Алгоритм восстановления пропусков на основе метода «Гусеница-SSA»

«Гусеница-SSA» [14] – это адаптивный непараметрический метод разложения временных рядов (SSA – singular spectrum analysis, сингулярный спектральный анализ). Алгоритм восстановления на основе метода «Гусеница-SSA» рассмотрен в работах [12, 15]. Идея алгоритма восстановления пропусков схожа с алгоритмом прогнозирования на основе метода «Гусеница-SSA» и заключается в оценке сигнала в местах пропусков моделями извлеченных компонент разложения.

Базовый алгоритм «Гусеницы-SSA» раскладывает сигнал на набор компонент $X = \sum_{k=1}^K X^{(k)}$ по базису собственных векторов лаговой ковариационной матрицы сигнала C_X . Для расчета матрицы C_X задается длина окна преобразования L . Матрица C_X может быть рассчитана через траекторную матрицу D временного ряда размерности $K \times L$ (где $K = N - L + 1$), столбцы которой сформированы из исходного временного ряда следующим образом:

$$(X(i), \dots, X(i + L - 2))^T, \quad 1 \leq i \leq K \quad (1)$$

$$C_X = \frac{1}{K} D^T D. \quad (2)$$

Далее рассчитываются собственные векторы E_k ковариационной матрицы C_X . Обозначим $\lambda_1 \geq \dots \geq \lambda_L \geq 0$ собственные числа матрицы C_X . Сумма всех собственных чисел матрицы дает полную дисперсию временного ряда X .

Проецирование временного ряда на каждый собственный вектор дает главные компоненты A_k :

$$A_k(t) = \sum_{j=1}^L X(t + j - 1) E_k(j). \quad (3)$$

Частотные свойства характеризуются парами почти равных собственных чисел и собственных векторов, соответствующих одной частоте. Длина окна L определяет наименьшую частоту, которую будет учитывать разложение SSA.

На последнем этапе алгоритма «Гусеница-SSA» проводится восстановление временного ряда из компонент разложения R_k , которые рассчитываются как линейная комбинация главных компонент и собственных векторов с последующим диагональным усреднением:

$$R_k(t) = \frac{1}{M_t} \sum_{k \in \kappa} \sum_{j=L_t}^{U_t} A_k(t - j + 1) E_k(j), \quad (4)$$

где κ – набор собственных векторов, на основе которых идет восстановление; M_t – фактор нормализации; L_t и U_t – верхние и нижние границы суммирования. M_t , L_t и U_t изменяются в зависимости от положения отсчета времени t относительно начального и конечного отсчетов времени временного ряда $X(t)$ [12].

Алгоритм восстановления пропусков на основе метода «Гусеница-SSA» учитывает временные зависимости во временных рядах для заполнения пропусков. Алгоритм итеративно рассчитывает и переоценивает корреляционную матрицу сигнала C_X и собственные векторы E_k .

Восстановление пропусков происходит следующим образом: сначала сигнал центрируется путем оценки его среднего и заполнения пропусков нулевыми значениями.

После начинается итеративная процедура восстановления. Во внутреннем цикле выполняется разложение сигнала методом «Гусеница-SSA» (вычисляются C_X и E_k). Далее, временной ряд восстанавливается с учетом только первой компоненты R_1 : в местах пропусков записываются новые значения временного ряда $X(t) = R_1(t)$. Эта процедура выполняется до тех пор, пока не будет достигнута сходимость в местах пропусков. Иными словами, результат предыдущего восстановления (внутренний цикл) не будет отличаться от текущего результата.

В качестве критерия сходимости, к примеру, можно выбрать дисперсию разности результатов между текущим и предыдущим итерациями цикла, отнесенную к дисперсии результата восстановления на текущей итерации. После того как во внутреннем цикле будет достигнута сходимость, во внешнем цикле добавится еще одна компонента — R_2 и внутренний цикл восстановления повторится уже для компонент R_1 и R_2 , и так далее по всем компонентам.

Восстановление сигнала происходит за счет итеративной переоценки ковариационной матрицы C_X и адаптивности базиса собственных векторов, как следствие последовательной оценки сигнала в местах пропусков. Подставив выражение (3) в (4), получим:

$$R_k(t) = \frac{1}{M_i} \sum_{k \in K} \sum_{j=L_i}^{L} X(t) E_k(j) E_k(j). \quad (5)$$

Видно, что $R_k(t)$ представляет собой свертку временного ряда с фильтром f_n и является промежуточной оценкой $X^*(t)$ значений $X(t)$:

$$X^*(t) = R_k(t) = \sum_{n=-(L-1)}^{L-1} X(i+n) f_n. \quad (6)$$

С каждой итерацией алгоритм приближает промежуточные оценки $X^*(t)$ к конечной оценке $\bar{X}(t)$. Ширина окна преобразования «Гусеница-SSA» выбирается в два раза большей, чем величина пропуска, чтобы с запасом оценить диапазон частот в пропуске. Также стоит отметить, что преобразование «Гусеница-SSA» для заполнения пропуска использует информацию слева и справа от пропуска, за счет операции диа-

гонального усреднения (4).

Алгоритм восстановления пропусков на основе преобразования Фурье

Алгоритм восстановления предложен в статье [13] для заполнения пропусков в измерениях параметров атмосферы. Алгоритм применим для сигналов, имеющих ярко выраженные периодические составляющие. Идея алгоритма заключается в том, чтобы с использованием спектра Фурье оценивать в сигнале базовые частоты и восстанавливать с их помощью пропущенные значения. В оригинальной статье авторами применялся алгоритм Ломба-Скаргла, т. к. предполагалась неравномерная дискретизация сигнала. В нашем случае мы будем использовать дискретное преобразование Фурье, поскольку сигнал уже синхронизирован и имеет постоянную частоту дискретизации.

Алгоритм состоит из следующих шагов: сначала происходит центрирование временного ряда путем оценки и вычитания среднего и заполнение пропусков нулевыми значениями. К временному ряду применяется прямое дискретное преобразование Фурье:

$$S(k) = \sum_{t=0}^{N-1} X(t) e^{-\frac{2\pi i}{N} kt}, \quad (7)$$

где $S(k)$ — коэффициенты преобразования Фурье; k — индекс частоты.

Далее производится модификация спектра Фурье с целью фильтрации базовых частот сигнала. Выполняется эта процедура путем сравнения амплитуд частот с порогом и приравнивания к нулю частот, которые ниже порога:

$$\hat{S}(k) = \begin{cases} 0, & |S(k)| < Thr \\ S(k), & |S(k)| \geq Thr \end{cases} \quad (8)$$

где Thr — величина порога.

Порог оценивается как средняя мощность спектра плюс три среднеквадратических отклонения разброса мощности спектра. На третьем шаге выполняется обратное преобразование Фурье и заполнение временного ряда в местах пропусков результатами обратного преобразования Фурье и прибавление среднего:

$$\widehat{X}(t) = \frac{1}{N} \sum_{k=0}^{N-1} S(k) e^{\frac{2\pi i}{N} kt}. \quad (9)$$

Стоит отметить, что для восстановления пропуска используется информация обо всем сигнале, за счет того, что коэффициенты Фурье рассчитываются как свертка базиса по всей длине сигнала.

Алгоритм восстановления пропусков на основе модели авторегрессии

Настоящий алгоритм предложен авторами в качестве альтернативы двум предыдущим алгоритмам. Авторегрессионным называется процесс, в котором значение временного ряда находится в линейной зависимости от предыдущих значений временного ряда. Если значение временного ряда зависит от значений, отстоящих от 1 до p лагов (отсчетов времени) назад, то мы наблюдаем авторегрессионный процесс порядка p . Отсюда следует, что авторегрессионная модель (АР) представляет собой модель, в которой моделируемые значения задаются линейной функцией предыдущих наблюдений. По существу, авторегрессионная модель есть аналог линейной многофакторной модели (множественной регрессии) с той лишь разницей, что в качестве независимых переменных выступают предыдущие значения временного ряда. Авторегрессионный временной ряд можно описать следующим образом [16]:

$$X(t) = a_1 X(t-1) + a_2 X(t-2) + \dots + a_p X(t-p) + e(t), \quad (10)$$

где a – коэффициенты авторегрессии; e – белый шум; p – порядок модели авторегрессии.

Для оценки коэффициентов авторегрессии используется метод наименьших квадратов (МНК) [16].

Модель авторегрессии сигнала позволяет вычислять оптимальный прогноз на заданное число шагов вперед и давать оценку ошибки прогноза. Согласно [16] дисперсия $V(l)$ ошибки прогноза на l шагов вперед для модели авторегрессии порядка p [16] оценивается следующим образом:

$$V(l) = (1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{l-1}^2) \sigma_e^2, \quad (11)$$

где σ_e^2 – дисперсия шума модели; ψ – веса, рассчитываемые рекуррентно [16]:

$$\psi_j = a_1 \psi_{j-1} + a_2 \psi_{j-2} + \dots + a_p \psi_{j-p}. \quad (12)$$

Идея алгоритма восстановления пропусков на основе модели авторегрессии заключается в прогнозировании значений временного ряда как вперед во времени, так и назад, с последующим оптимальным усреднением результатов. Таким образом, необходимы две модели временного ряда: прямая во времени M_{AR}^f (прогнозирующая), описываемая выражением (10), и обратная M_{AR}^b (интерполяционная):

$$X(t-p) = a_1^b X(t-p+1) + a_2^b X(t-p+2) + \dots + a_p^b X(t-1) + e(t). \quad (13)$$

Модель, прогнозирующая вперед (прямая), будет давать лучшие оценки в начале пропуска, а модель, прогнозирующая назад (обратная), – лучшие оценки в конце пропуска, т. к. точность прогноза падает с увеличением количества шагов прогнозирования, в соответствии с выражением (11).

Зная оценки значений в местах пропусков прямой и обратной модели $\widehat{X}^f(t)$ и $\widehat{X}^b(t)$ соответственно и дисперсии ошибки оценки прогноза прямой и обратной моделей, формула (11), для каждого момента времени V_t^f и V_t^b соответственно, можно применить оптимальное усреднение результатов прогноза методом наилучшей линейной несмещенной оценки, чтобы получить оценку пропущенных значений:

$$\widehat{X}(t) = \frac{V_t^f \widehat{X}^b(t) + V_t^b \widehat{X}^f(t)}{V_t^f + V_t^b}. \quad (14)$$

Алгоритм восстановления состоит из следующих шагов: как в предыдущих алгоритмах сначала исходный временной ряд центрируется путем вычитания среднего и заполнения пропусков нулевыми значениями. Далее оцениваются прямая M_{AR}^f и обратная M_{AR}^b модели. Оценка модели авторегрессии проводится методом перекрестной проверки, заключающимся в искусственном создании пропусков разной величины в сигналах, восстановления пропусков разными моделями и выбора такой модели, которая давала бы наименьшую ошибку на

всех пропусках [17]. На третьем шаге выполняется оценка значений в местах пропусков обеими моделями. На последних шагах выполняется оптимальное усреднение результатов и прибавление среднего.

Сравнение алгоритмов восстановления пропусков

Мы провели апробацию алгоритмов на реальных данных системы мониторинга состояния дамбы города Бостон (Великобритания). В дамбе была установлена сеть многопараметрических датчиков, измеряющих поровое давление (давление воды в порах почвы), температуру почвы и отклонение от вертикали (инклинометры). Дискретизация сигналов составляла 15 мин. Был проведен анализ датчиков порового давления.

Для моделирования было выбрано 14 сигналов, по 7 сигналов из различных точек установки в дамбе. Примеры трех сигналов для моделирования представлены на рис. 3. Для получения точностных характеристик восстановления пропусков в каждом из сигналов в случайных местах создавались искусственные пропуски величиной от 15 мин (один отсчет) до двух недель (1344 отсчета). Для всех сигналов пропуск каждой величины генерировался тысячу раз в случайные моменты времени. Приведенные в настоящей статье алгоритмы восста-

новления также сравнивались с линейной интерполяцией.

Для расчета характеристик точности восстановления были выбраны метрики: R^2 – коэффициент детерминации (p -квадрат) и СКО (среднеквадратическое отклонение) ошибки оценки.

R^2 рассчитывается следующим образом:

$$R^2 = 1 - \frac{\sigma_{\delta}^2}{\sigma_x^2}, \quad (14)$$

где σ_{δ}^2 – дисперсия ошибки оценки пропущенных значений; σ_x^2 – дисперсия сигнала.

С уменьшением дисперсии ошибки оценки R^2 стремится к единице, с увеличением ошибки оценки R^2 стремится к отрицательным значениям. Значение R^2 мы считали минимально допустимым.

На рис. 4 представлены зависимости R^2 и СКО ошибки оценки соответственно для разных величин пропусков от 15 мин (один отсчет) до двух недель (1344 отсчета). Примерно одинаковую точность показывают алгоритмы восстановления на основе «Гусеница-SSA» и AP-модели, далее следует алгоритм на основе преобразования Фурье и наихудшая точность соответствует линейной интерполяции. Алгоритм на основе «Гусеница-SSA» обеспечивает значение

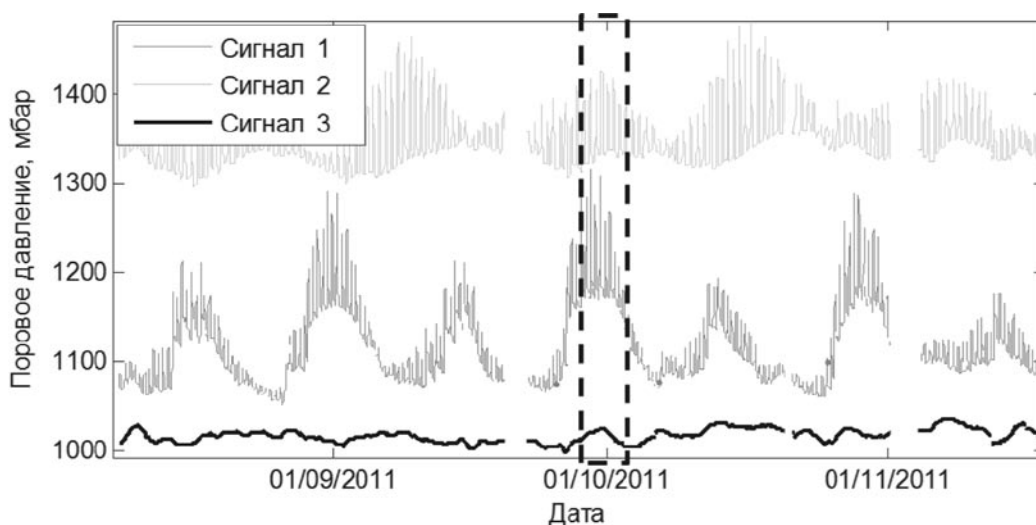


Рис. 3. Сигналы для апробации алгоритмов, формат даты ДД/ММ/ГГГГ, пунктирной рамкой выделен пример места, где создавался искусственный пропуск

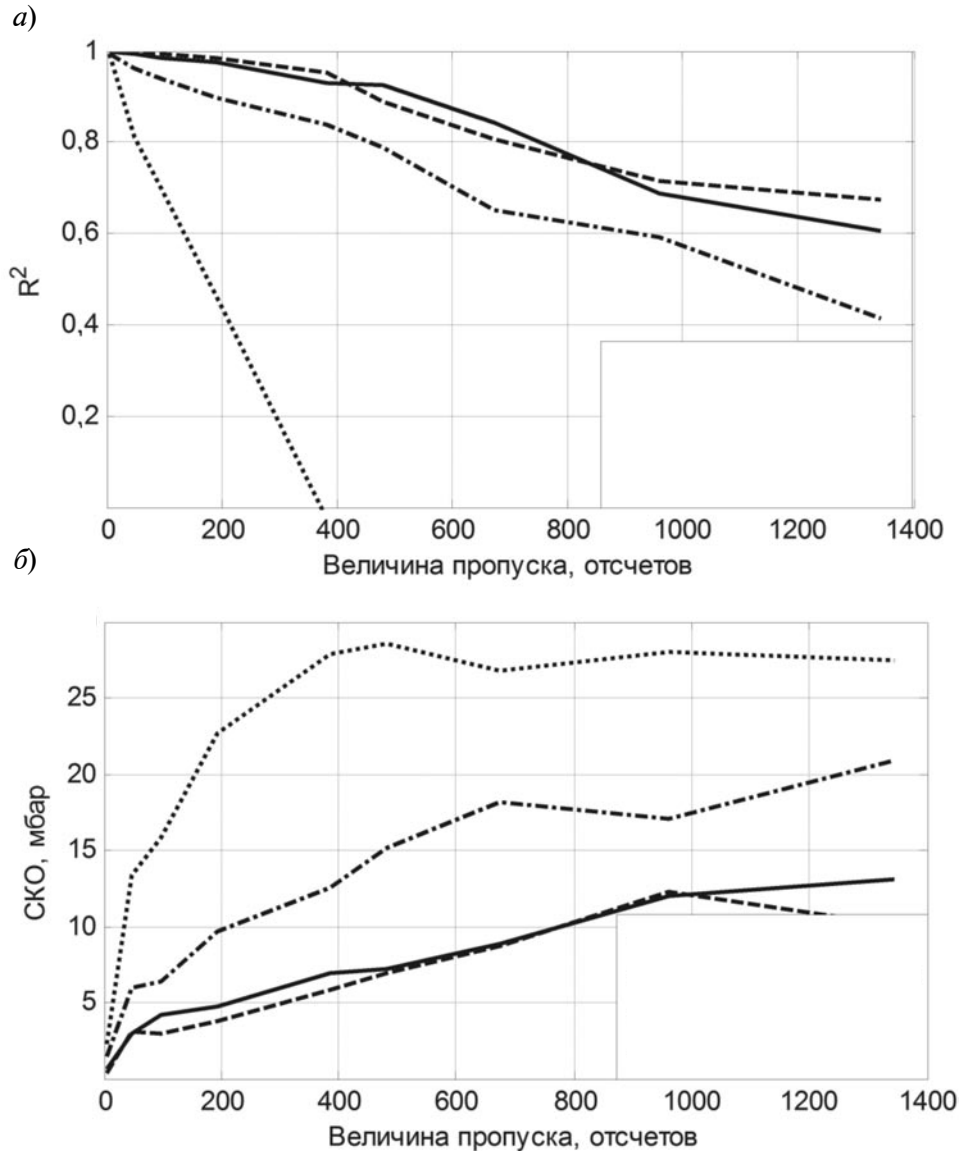


Рис. 4. Анализ точности восстановления:
а – зависимость коэффициента детерминации от величины пропуска для всех алгоритмов;
б – зависимость SKO ошибки оценки от величины пропуска для всех алгоритмов
(.....) лин. интерполяция; (—·—) преобразование Фурье; (----) «Гусеница SSA»; (—) AP

$R^2 = 0,8$ для величины пропуска 755 отсчетов ($\sim 1,12$ недели) с SKO ≈ 10 (мбар). AP-модели обеспечивают значение $R^2 = 0,8$ для величины пропуска 672 (одна неделя) с SKO ≈ 8 (мбар). Для величины пропуска до 10 отсчетов (150 мин) все алгоритмы показывают примерно одинаковую точность восстановления. Это говорит о том, что для восстановления небольших пропусков (до 10 отсчетов) можно использовать линейную интерполяцию на этапе синхрониза-

ции измерений. Значения R^2 ниже нуля не показаны (линейная интерполяция), т. к. соответствуют слишком низкой точности, не интересной для анализа.

На рис. 5 представлены примеры восстановления пропусков величиной 384 отсчета (пять дней) для сигналов 1, 2 и 3. Вырезанный участок сигнала представлен на рис. 3 и отмечен пунктирной рамкой. Как видно, алгоритмы восстановления на основе «Гусеница-SSA», AP-модели и пре-

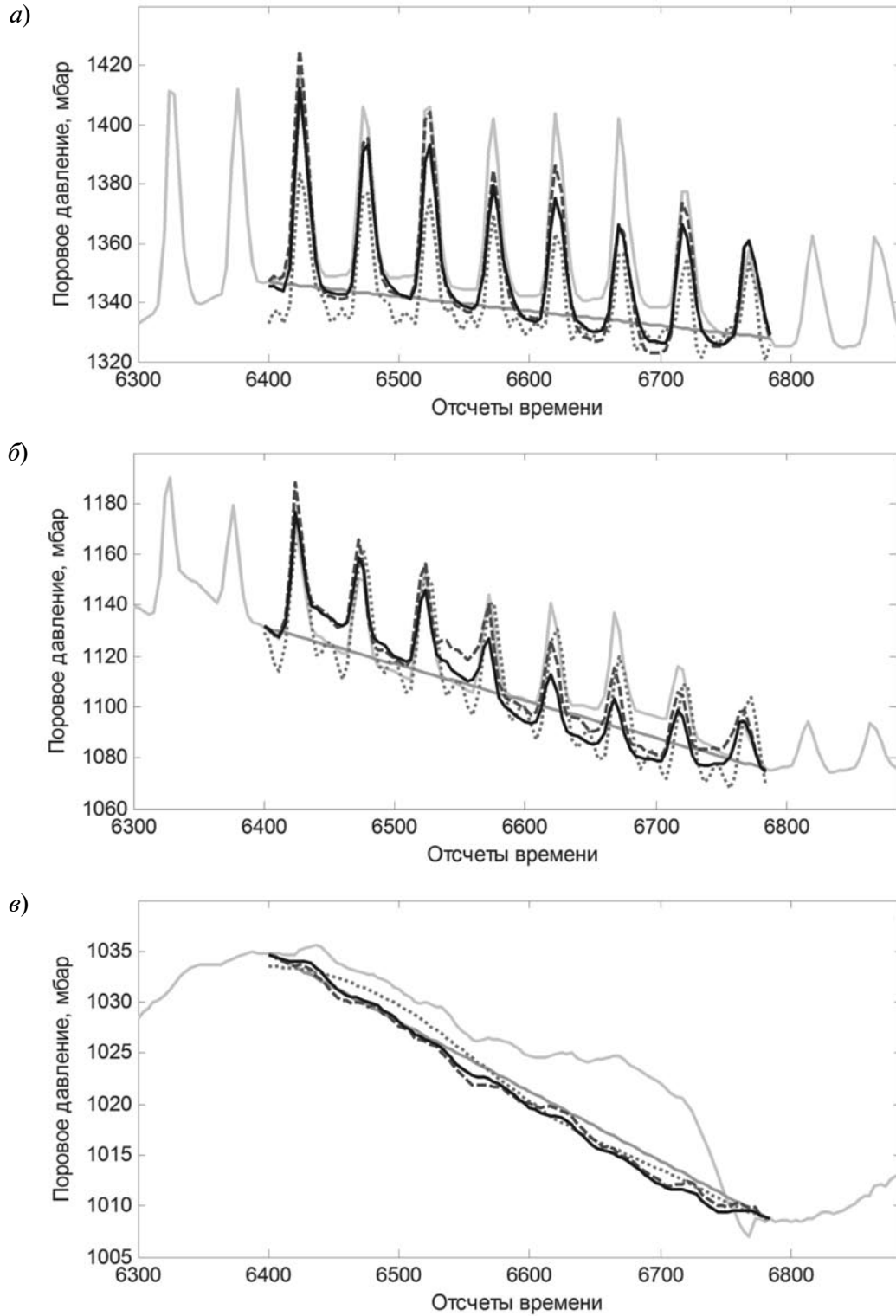


Рис. 5. Примеры восстановления пропусков в сигналах, представленных на рис. 3 пунктирной рамкой:

a – сигнал 1; *б* – сигнал 2; *в* – сигнал 3

(—) реальный сигнал; (---) лин. интерп.; (.....) преобр. Фурье; (-.-.-) «Гусеница SSA»; (—) AP

образования Фурье сохраняют динамику сигнала и показывают примерно одинаковое поведение.

Стоит отметить, что важным фактором в выборе конкретного алгоритма восстановления пропусков является время восстановления пропуска, которое зависит от вычислительной сложности алгоритма. В случае алгоритма на основе «Гусеница-SSA», наиболее сложный этап расчета – сингулярное разложение ковариационной матрицы, вычислительная сложность которого $O(K \cdot L^2)$, где L – величина окна преобразования, $K = N - L + 1$, также необходимо учесть, что для восстановления пропуска вычисления проводятся итерационно и, как следствие, сингулярное разложение вычисляется много раз, что сильно увеличивает общее время восстановления. Вычислительная сложность МНК для авторегрессии $O(p^3)$, где p – порядок модели авторегрессии. Вычислительная сложность для быстрого преобразования Фурье $O(N \cdot \log(N))$, где N – длина временного ряда. Таким образом, наиболее эффективным с точки зрения скорости вычисления является алгоритм на основе преобразования Фурье, затем – алгоритм на основе модели авторегрессии и, наконец, самые высокие вычислительные затраты соответствуют алгоритму на основе метода «Гусеница-SSA».

Предложенный подход к восстановлению пропусков в измеряемых сигналах основан на использовании адаптивных алгоритмов восстановления пропусков в сигналах в условиях априорной неопределенности параметров сигналов. Суть представленных алгоритмов заключается в оценке пропущенных значений с использованием исторических записей измеряемых сигналов. Преимущества такого подхода состоят в том, что алгоритмы адаптируются с учетом поведения сигналов и не требуют

предварительной оценки параметров сигналов.

В данной статье сравнивались четыре алгоритма восстановления пропусков: на основе метода «Гусеница-SSA», преобразования Фурье, линейной интерполяции, а также предложенный нами алгоритм на основе модели авторегрессии.

Сравнение алгоритмов восстановления пропусков было проведено на реальных данных дамбы города Бостон (Великобритания). Для пропусков величиной до 10 отсчетов все алгоритмы показали высокую точность, что говорит о том, что для восстановления небольших пропусков достаточно использовать линейную интерполяцию. Алгоритм на основе преобразования Фурье показал худшие результаты, чем алгоритмы на основе метода «Гусеница-SSA» и модели авторегрессии. Но его преимущество состоит в том, что для его вычисления можно использовать быстрое преобразование Фурье, а это может пригодиться в приложениях, где существуют ограничения на быстроедействие. Алгоритмы на основе метода «Гусеница-SSA» и модели авторегрессии показали высокую точность восстановления пропусков величиной менее семи дней, учитывая частотно-временные свойства измеряемых сигналов. К недостаткам алгоритмов можно отнести сравнительно высокие вычислительные затраты, особенно для алгоритма на основе «Гусеницы-SSA».

Предлагаемый подход к восстановлению пропусков в сигналах, опирающийся на рассмотренные алгоритмы, может быть распространен и на другие системы, такие как системы водоснабжения, метеорологические системы и другие, в которых актуальны проблемы, связанные с улучшением качества измерений.

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

СПИСОК ЛИТЕРАТУРЫ

1. Krzhizhanovskaya V.V. et al. Flood early warning system: design, implementation and computational modules // *Procedia Computer Science*. 2011. No. 4. Pp. 106–115.
2. Козионов А.П. и др. Вейвлет-преобразо-

вание и одноклассовая классификация для мониторинга состояния дамб // Информационно-управляющие системы. 2014. Вып. 71. № 4.

3. Pyayt A. et al. Data-driven modelling for flood defence structure analysis // *Comprehensive*



Flood Risk Management. Rotterdam: Klijn & Schweckendiek, 2013.

4. **Pyayt A.L. et al.** Time-Frequency Methods for Structural Health Monitoring // *Sensors*. 2014. No. 14. Pp. 5147–5173.

5. **Quevedoa J. et al.** Validation and reconstruction of flow meter data in the Barcelona water distribution network // *Control Engineering Practice*. 2012. Vol. 6. No. 18. Pp. 640–651.

6. **Moffata A.M. et al.** Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes // *Agricultural and Forest Meteorology*. 2006. Vol. 3. No. 147.

7. **UrbanFlood project** [Электронный ресурс] / URL: <http://www.urbanflood.eu>

8. **Spath H.** *One-Dimensional Spline Interpolation Algorithms*. MA: AK Peters Wellesley, 1995.

9. **Moghtaderi A., Borgnat P., Flandrin P.** Gap-filling by the empirical mode decomposition // *IEEE Internat. Conf. on Acoustics, Speech and Signal Processing*. 2012. Pp. 3821–3824.

10. **Wei Q. et al.** Analysis of EEG via Multivariate Empirical Mode Decomposition for Depth of

Anesthesia Based on Sample Entropy // *Entropy*. 2013. No. 15. Pp. 3458–3470.

11. **Lorenc C., Rawlins F.** Why does 4D-Var beat 3D-Var? // *Quarterly Journal of the Royal Meteorological Society*. 2005. Vol. 613. No. 131.

12. **Kondrashov D., Shprits Y., Ghil M.** Gap filling of solar wind data by singular spectrum analysis // *Geophysical research letters*. 2010. Pp. 1–6.

13. **Hocke K., Kämpfer N.** Gap filling and noise reduction of unevenly sampled data by means of the Lomb-Scargle periodogram // *Atmos. Chem. Phys.* 2009. No. 9. Pp. 4197–4206.

14. **Голяндина Н.Э.** Метод «Гусеница»-SSA: анализ временных рядов: Учеб. пособие. СПб.: Изд-во СПбГУ, 2003. 87 с.

15. **Golyandina N., Osipov E.** The “Caterpillar”-SSA method for analysis of time series with missing values // *J. of Statistical Planning and Inference*. 2007. Vol. 8. No. 137. Pp. 2642–2653.

16. **Box G., Jenkins G.** *Time series analysis: Forecasting and control*. San Francisco: Holden-Day, 1970.

REFERENCES

1. **Krzhizhanovskaya V.V. et al.** Flood early warning system: design, implementation and computational modules, *Procedia Computer Science*, 2011, No. 4, Pp. 106–115.

2. **Kozionov A.P. et al.** Veyvlet-preobrazovaniye i odnoklassovaya klassifikatsiya dlya monitoringa sostoyaniya damb [The wavelet transform and odnoklassovaya classification for monitoring dams]. *Informatsionno-upravlyayushchiye sistemy [Information and Control Systems]*, 2014, Vol. 71, No. 4. (rus)

3. **Pyayt A. et al.** Data-driven modelling for flood defence structure analysis, *Comprehensive Flood Risk Management*. Rotterdam: Klijn & Schweckendiek, 2013.

4. **Pyayt A.L. et al.** Time-Frequency Methods for Structural Health Monitoring, *Sensors*, 2014, No. 14, Pp. 5147–5173.

5. **Quevedoa J. et al.** Validation and reconstruction of flow meter data in the Barcelona water distribution network, *Control Engineering Practice*, 2012, Vol. 6, No. 18, Pp. 640–651.

6. **Moffata A.M. et al.** Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes, *Agricultural and Forest Meteorology*, 2006, Vol. 3, No. 147.

7. **UrbanFlood project**. Available: <http://www.urbanflood.eu>

8. **Spath H.** *One-Dimensional Spline Interpolation Algorithms*. MA: AK Peters Wellesley, 1995.

9. **Moghtaderi A., Borgnat P., Flandrin P.** Gap-filling by the empirical mode decomposition, *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2012, Pp. 3821–3824.

10. **Wei Q. et al.** Analysis of EEG via Multivariate Empirical Mode Decomposition for Depth of Anesthesia Based on Sample Entropy, *Entropy*, 2013, No. 15, Pp. 3458–3470.

11. **Lorenc C., Rawlins F.** Why does 4D-Var beat 3D-Var? *Quarterly Journal of the Royal Meteorological Society*, 2005, Vol. 613, No. 131.

12. **Kondrashov D., Shprits Y., Ghil M.** Gap filling of solar wind data by singular spectrum analysis, *Geophysical research letters*, 2010, Pp. 1–6.

13. **Hocke K., Kämpfer N.** Gap filling and noise reduction of unevenly sampled data by means of the Lomb-Scargle periodogram, *Atmos. Chem. Phys.*, 2009, No. 9, Pp. 4197–4206.

14. **Golyandina N.E.** *Metod «Gusenitsa»-SSA: analiz vremennykh ryadov [The Caterpillar-SSA method for analysis of time series]*. St. Petersburg: SPbGU Publ., 2003, 87 p. (rus)

15. **Golyandina N., Osipov E.** The “Caterpillar”-SSA method for analysis of time series with missing values, *Journal of Statistical Planning and Inference*, 2007, Vol. 8, No. 137, Pp. 2642–2653.

16. **Box G., Jenkins G.** *Time series analysis: Forecasting and control*. San Francisco: Holden-Day, 1970.

КОЗИОНОВ Алексей Петрович – аспирант кафедры аэрокосмических измерительно-вычислительных комплексов Санкт-Петербургского государственного университета аэрокосмического приборостроения, инженер ООО «Сименс».

191186, Россия, Санкт-Петербург, Волынский пер., д. 3а.

E-mail: alexey.kozionov@gmail.com

KOZIONOV Alexey P. *St. Petersburg State University of Aerospace Instrumentation. Siemens, Corporate Technology.*

191186, Volynskii Lane 3A, St. Petersburg, Russia.

E-mail: alexey.kozionov@gmail.com

ПЯЙТ Александр Леонидович – ведущий инженер ООО «Сименс».

191186, Россия, Санкт-Петербург, Волынский пер., д. 3а.

E-mail: alexander.pyayt@siemens.com

PYAYT Alexander L. *Siemens.*

191186, Volynskii Lane 3A, St. Petersburg, Russia.

E-mail: alexander.pyayt@siemens.com

МОХОВ Илья Игоревич – ведущий эксперт-исследователь ООО «Сименс».

191186, Россия, Санкт-Петербург, Волынский пер., д. 3а.

E-mail: ilya.mokhov@siemens.com

МОКHOV Илья I. *Siemens.*

191186, Volynskii Lane 3A, St. Petersburg, Russia.

E-mail: ilya.mokhov@siemens.com

ИВАНОВ Юрий Павлович – доцент кафедры аэрокосмических измерительно-вычислительных комплексов Санкт-Петербургского государственного университета аэрокосмического приборостроения, кандидат технических наук.

190000, Россия, Санкт-Петербург, ул. Б. Морская, д. 61.

E-mail: upi@mail.ru

IVANOV Yuri P. *St. Petersburg State University of Aerospace Instrumentation.*

190000, Bolshaya Morskaya Str. 61, St. Petersburg, Russia.

E-mail: upi@mail.ru