

*Абубакиров И.Р.**Камчатский филиал Геофизической службы РАН, г.Петропавловск-Камчатский, air@emsd.iks.ru***Введение**

Работами А.А.Гусева и В.К.Лемзикова 1978-1982 гг. [1-3] на Камчатке был выявлен новый среднесрочный сейсмологический предвестник – аномальные изменения формы огибающих коды (хвостовой части записей) слабых землетрясений, возникающих в зоне подготовки будущего сильного землетрясения. Для описания аномальных свойств коды индивидуального землетрясения Гусевым и Лемзиковым был введен специальный параметр, обозначенный ими через α . Собственно значение α есть аномалия (отклонение от регионального среднего) для производной логарифма текущей амплитуды огибающей записи коды по времени:

$$\alpha = \frac{d}{dt} [\lg A(t) - \lg a(t)], \quad (1)$$

где $A(t)$ – текущая амплитуда огибающей коды, $a(t)$ – реперная или опорная функция, описывающая среднюю региональную форму огибающей записи коды. Иначе говоря, положительные значения параметра α указывают на относительно более пологий спад амплитуд огибающей коды по сравнению с опорной кривой, а отрицательные значения – на относительно более крутой.

Работами 1978-1982 гг. было обнаружено, что в 1963, 1971 и 1978 гг. выраженные бухтообразные отрицательные аномалии параметра α предвещали три сильнейших (с магнитудой около 8) землетрясения Камчатки и Курил, опережая их на 0.8-1.6 года (считая от начала аномалии).

Основываясь на этом опыте, в начале 1982 г. на Камчатке было начато непрерывное слежение за временными вариациями крутизны спада огибающих коды слабых местных землетрясений. Цель этой работы - выявление в оперативном порядке возможных предвестниковых аномалий и выдача, в обоснованных случаях, среднесрочных прогнозов сильных землетрясений, готовящихся вблизи восточного побережья Камчатки.

В настоящее время оперативное слежение за временными вариациями крутизны спада огибающих коды проводится на основе анализа цифровых сейсмограмм, зарегистрированных камчатской сетью радиотелеметрических станций. Процедура обработки этих сейсмограмм состоит из трех этапов:

1. Эмуляция записи прибора ВЭГИК по записи радиотелеметрической станции.
2. Расчет огибающей коды модельной записи прибора ВЭГИК.
3. Оценка крутизны спада огибающей коды модельной записи (оценка параметра α).

Цифровые записи радиотелеметрических станций часто содержат широкополосные электронные помехи большой амплитуды (выбросы), которые варьируют по форме и имеют длительность, достигающую 0.5 с. Подобные выбросы приводят к смещению в оценках параметра α , что может затруднить выделение предвестниковых аномалий. Поэтому желательно избавиться от такого рода смещения. При этом возникают следующие задачи:

1. Разработать алгоритм автоматической идентификации сегментов записей радиотелеметрических станций, зашумленных выбросами.
2. Разработать алгоритм реконструкции сигнала на участках, зашумленных выбросами, позволяющий воспроизвести реалистическую форму огибающих коды модельных записей прибора ВЭГИК.

Решению этих двух задач и посвящена настоящая работа.

1. Идентификация выбросов

Для идентификации выбросов предлагается двухшаговый алгоритм. На первом шаге производится грубая идентификация выбросов, а именно, оценивается число выбросов и производится выделение сегментов сейсмограмм, содержащих выбросы, без точной оценки времен начала и конца выбросов. На втором шаге для каждого из сегментов, выделенных на первом шаге, уточняются моменты вступления и конца выброса.

1.1. Грубая идентификация выбросов

Алгоритм грубой идентификации состоит из следующих шагов.

1. Исходный сигнал $\mathbf{x} = (x_1, \dots, x_n)$ пропускается через фильтр, на выходе которого формируется сигнал $\mathbf{K} = (K_1, \dots, K_n)$, содержащий значения выборочного эксцесса в скользящем временном окне длительностью d_m секунд (m отсчетов)

$$K_i = \begin{cases} 0, & \text{при } i \leq m_1 \\ \frac{1}{m} \sum_{j=i-m_1}^{j=i+m_2} (x_j - \hat{a}_i)^4, & \text{при } m_1 < i \leq n - m_2 \\ \hat{\sigma}_i^4, & \\ 0, & \text{при } n - m_2 < i \leq n \end{cases}, \quad (2)$$

где $\hat{a}_i = \frac{1}{m} \sum_{j=i-m_1}^{j=i+m_2} x_j$ - выборочное среднее, $\hat{\sigma}_i = \sqrt{\frac{1}{m} \sum_{j=i-m_1}^{j=i+m_2} (x_j - \hat{a}_i)^2}$ - выборочное стандартное

уклонение, $m_1 = \begin{cases} (m-1)/2, & \text{при нечетном } m \\ m/2, & \text{при четном } m \end{cases}$, $m_2 = \begin{cases} (m-1)/2, & \text{при нечетном } m \\ m/2 - 1, & \text{при четном } m \end{cases}$.

Длительность d_m подбирается равной удвоенной максимальной ожидаемой длительности выброса. В рассматриваемом нами случае сейсмограмм радиотелеметрических станций $d_m = 1$ с (128 отсчетов).

2. Определяются сегменты сигнала, на которых значение выборочного эксцесса превышает заданное пороговое значение $K_{\text{нор}}$.

3. Производится процедура «склейки» соседних сегментов. В ходе этой процедуры соседние сегменты, разделенные интервалами менее d_m секунд, объединяются. При этом за левую границу объединенного сегмента принимается первый отсчет левого сегмента, а за правую границу – последний отсчет правого сегмента. Далее сегменты, длительность которых не превышает d_m , отбрасываются. Они идентифицируются как сегменты, содержащие вступления прямых волн. Оставшиеся сегменты интерпретируются как сегменты, зашумленные выбросами.

Следует отметить, что в отличие от традиционного определения эксцесса, согласно которому эксцесс нормального распределения равен 0, используется дефиниция, при которой эксцесс нормального распределения равен 3.

1.2. Уточненная идентификация выбросов

Уточненная идентификация проводится с использованием статистического подхода [6]. Допустим, что сегмент сейсмограммы $(x_{n_s}, \dots, x_{n_f})$, содержащий выброс, представляет собой кусочно-постоянный сигнал, возмущенный нестационарным белым шумом. Другими словами, пусть $(x_{n_s}, \dots, x_{n_f})$ – последовательность независимых гауссовских случайных величин с кусочно-постоянными средним и дисперсией, причем скачки среднего и дисперсии (начало и конец выброса) происходят в неизвестные моменты времени $t_1 = n_1^* \Delta t, t_2 = n_2^* \Delta t$, где Δt - шаг дискретизации,

$$x_i = \begin{cases} a_1 + \sigma_1 \varepsilon_i, & \text{при } n_s \leq i \leq n_1^* \\ a_2 + \sigma_2 \varepsilon_i, & \text{при } n_1^* < i \leq n_2^* \\ a_3 + \sigma_3 \varepsilon_i, & \text{при } n_2^* < i \leq n_f \end{cases}, \quad (3)$$

где (ε_i) , $i = n_s, \dots, n_f$ - последовательность независимых гауссовских случайных величин с нулевым средним и единичной дисперсией.

Для получения оценок \hat{n}_1^* и \hat{n}_2^* решается задача минимизации целевой функции $G(x_{n_s}, \dots, x_{n_f}; n_1, n_2)$, определенной на множестве значений пар (n_1, n_2) , удовлетворяющих условию $n_s < n_1 < n_2 < n_f$,

$$\left(\hat{n}_1^*, \hat{n}_2^* \right) = \arg \min_{(n_1, n_2)} \left[G(x_{n_s}, \dots, x_{n_f}; n_1, n_2) \right]. \quad (4)$$

Функция $G(x_{n_s}, \dots, x_{n_f}; n_1, n_2)$ строится с использованием логарифмической функции правдоподобия. Обозначим через $l_k(x_{n_{k-1}+1}, \dots, x_{n_k}; a_k, \sigma_k)$ логарифмическую функцию правдоподобия для фрагмента сигнала с номером k ($1 \leq k \leq 3$)

$$l_k(x_{n_{k-1}+1}, \dots, x_{n_k}; a_k, \sigma_k) = \frac{(n_k - n_{k-1})}{2} \ln \left(\frac{2\pi}{\sigma_k^2} \right) - \frac{1}{2\sigma_k^2} \sum_{i=n_{k-1}+1}^{n_k} (x_i - a_k)^2, \quad (5)$$

где $n_0 = n_s, n_3 = n_f$.

Тогда,

$$\begin{aligned} G(x_{n_s}, \dots, x_{n_f}; n_1, n_2) &\equiv -(n_f - n_s + 1)[1 + \ln(2\pi)] - 2 \sum_{k=1}^3 l_k \left(x_{n_{k-1}+1}, \dots, x_{n_k}; \hat{a}_k, \hat{\sigma}_k^2 \right) = \\ &= \sum_{k=1}^3 (n_k - n_{k-1}) \ln \left(\hat{\sigma}_k^2 \right), \end{aligned} \quad (6)$$

где

$$\hat{a}_k = \frac{1}{(n_k - n_{k-1})} \sum_{i=n_{k-1}+1}^{n_k} x_i, \quad \hat{\sigma}_k^2 = \frac{1}{(n_k - n_{k-1})} \sum_{i=n_{k-1}+1}^{n_k} \left(x_i - \hat{a}_k \right)^2$$

выборочные значения среднего и дисперсии, соответственно.

Для построения оценок \hat{n}_1^* и \hat{n}_2^* используется глобальный поиск, при котором производится полный перебор всех возможных значений пар (n_1, n_2) . В целях ускорения расчета значений целевой функции в процедуре глобального поиска, сначала производится расчет кумулятивных сумм $\mathbf{C} = (C_{n_s}, \dots, C_{n_f})$, $\mathbf{D} = (D_{n_s}, \dots, D_{n_f})$

$$C_i = \sum_{j=n_s}^i x_j, \quad D_i = \sum_{j=n_s}^i x_j^2, \quad i = n_s, \dots, n_f \quad (7)$$

При известных \mathbf{C} и \mathbf{D} выборочная дисперсии рассчитывается по формуле

$$\hat{\sigma}_k^2 = \frac{D_{n_k} - D_{n_{k-1}}}{(n_k - n_{k-1})} - \frac{(C_{n_k} - C_{n_{k-1}})^2}{(n_k - n_{k-1})^2}. \quad (8)$$

2. Реконструкция сегментов сейсмограмм, зашумленных выбросами

Реконструкция отсчетов сигнала $(x_{n_1}, \dots, x_{n_2})$, зашумленных выбросами, производится на основе двух авторегрессионных моделей. Первая из этих моделей строится для фрагмента сигнала, который предшествует выбросу; вторая – для фрагмента сигнала, следующего за выбросом [4, 5].

2.1. Схема реконструкции

Схема реконструкции отсчетов сигнала, зашумленного выбросами, состоит из следующих шагов:

1. Оценка коэффициентов (a_1^f, \dots, a_N^f) авторегрессионной модели N -го порядка для фрагмента сигнала, предшествующего выбросу.

2. Линейное предсказание (экстраполяция вперед) отсчетов сигнала на участке, зашумленном выбросами,

$$x_i^f = \sum_{k=1}^N a_k^f x_{i-k}, \quad i = n_1, \dots, n_2. \quad (9)$$

3. Оценка коэффициентов (a_1^b, \dots, a_N^b) авторегрессионной модели для сегмента сигнала, следующего за выбросом.

4. Линейное предсказание (экстраполяция назад) отсчетов сигнала на участке, зашумленном выбросами,

$$x_i^b = \sum_{k=1}^N a_k^b x_{i+k}, \quad i = n_2, \dots, n_1. \quad (10)$$

5. Расчет реконструированного сигнала $(\hat{x}_{n_1}, \dots, \hat{x}_{n_2})$, как взвешенной суммы сигналов, экстраполированных вперед и назад,

$$\hat{x}_i = w_i x_i^f + (1 - w_i) x_i^b, \quad i = n_1, \dots, n_2, \quad (11)$$

где w - весовая функция, которая имеет вид [5]:

$$w_i = \begin{cases} 1 - \frac{1}{2}(2u_i)^\beta, & \text{при } u_i \leq 0,5 \\ \frac{1}{2}(2 - 2u_i)^\beta, & \text{при } u_i > 0,5 \end{cases}, \quad i = n_1, \dots, n_2, \quad (12)$$

где $u_i = (i - n_1)/(n_2 - n_1)$, β – параметр, контролирующий крутизну спада весовой функции.

Приведенную схему восстановления сигнала иллюстрирует рис. 1.

2.2. Оценка параметров авторегрессионной модели

Авторегрессионный сигнал $\{x_k\}$ N -го порядка

$$x_k = \sum_{i=1}^N a_i x_{k-i} + e_k \quad (13)$$

формируется путем пропускания дискретного стационарного белого шума $\{e_k\}$ через чисто рекурсивный фильтр N -го порядка с коэффициентами (a_1, \dots, a_N) . Элемент e_k в уравнении (13) можно интерпретировать как погрешность (ошибку) линейного предсказания отсчета x_k .

Предположим, что имеется фрагмент $[x_1, x_2, \dots, x_M]^T$ авторегрессионного сигнала N -го порядка, состоящий из M отсчетов ($M > 2N$). Тогда на основании формулы (13) можно записать следующее матрично-векторное уравнение

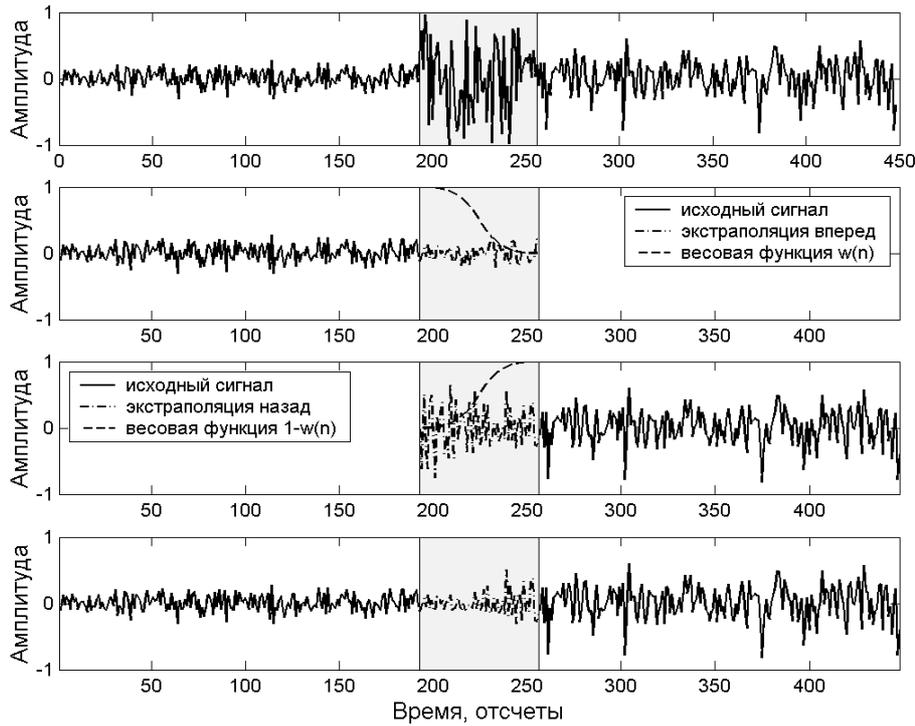


Рис. 1. Схема реконструкции сигнала на участке, зашумленном выбросом. Зашумленный участок ограничен вертикальными линиями и залит серым цветом. Реконструированный сигнал рассчитывается как взвешенная сумма сигналов, экстраполированных вперед и назад.

$$\mathbf{e} = \mathbf{x} - \mathbf{B}\mathbf{a}, \quad (14)$$

где

$$\mathbf{e} = [e_{N+1}, e_{N+2}, \dots, e_M]^T, \quad \mathbf{x} = [x_{N+1}, x_{N+2}, \dots, x_M]^T, \quad \mathbf{a} = [a_1, a_2, \dots, a_N]^T,$$

$$\mathbf{B} = \begin{bmatrix} x_N & x_{N-1} & \dots & x_1 \\ x_{N+1} & x_N & \dots & x_2 \\ \vdots & \vdots & & \vdots \\ x_{M-1} & x_{M-2} & \dots & x_{M-N} \end{bmatrix}.$$

Для нахождения вектора коэффициентов $\mathbf{a} = [a_1, \dots, a_N]^T$ используется метод наименьших квадратов, а именно, производится минимизация квадрата ошибки линейного предсказания $\mathbf{e}^T \mathbf{e} = \sum_{i=N+1}^M e_i^2$, что дает оценку

$$\hat{\mathbf{a}} = (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \mathbf{x}. \quad (15)$$

Для удовлетворительной реконструкции сигнала необходима аккуратная авторегрессионная модель. Это обычно подразумевает использование моделей высокого порядка. Тем самым подразумевается, что и длительности фрагментов, предшествующих выбросу и следующих за ним, достаточно велики, поскольку число отчетов сигнала, по которому производится оценка коэффициентов авторегрессионной модели, должно превосходить порядок авторегрессионной модели, по меньшей мере, вдвое.

3. Опробование алгоритма автоматической идентификации выбросов и реконструкции сигнала

Описанный алгоритм автоматической идентификации выбросов и реконструкции отсчетов сигнала, зашумленных выбросами, был опробован на ряде записей слабых близких землетрясений Камчатки, зарегистрированных радиотелеметрическими станциями. Целью опробования было: 1 - убедиться в работоспособности алгоритма; 2 - оценить рабочие значения настроечных параметров, а именно, пороговое значение $K_{пор}$ выборочного эксцесса; величину параметра β , определяющего крутизну спада весовой функции; порядок N авторегрессионных моделей, используемых для реконструкции; длительность M фрагментов сигнала, по которым производится оценка коэффициентов авторегрессионной модели. Проведенное опробование показало, что алгоритм позволяет успешно идентифицировать выбросы; реконструировать сигнал на участках, зашумленных выбросами, и воспроизводить реалистическую форму огибающих коды модельных записей прибора ВЭГИК при следующих значениях настроечных параметров: $K_{пор}=12$; $\beta=3$; $N=2m$, где m – длительность выброса, измеренная в отсчетах; $M=5m$.

На рисунках 2-5 приведены примеры результатов работы алгоритма, полученных при обработке записи слабого близкого землетрясения, зарегистрированного станцией «Русская».

Обсуждение

При идентификации выбросов в работе использована модель нестационарного гауссовского случайного процесса, а при восстановлении сигнала – модель авторегрессии. Здесь можно усмотреть некоторое противоречие. Действительно, в рамках самосогласованного подхода, для поиска времен начала и конца выброса следовало бы использовать выходной сигнал фильтра ошибки предсказания. Однако такой подход привел бы к неоправданному усложнению алгоритма. С другой стороны, целевую функцию, построенную на основе гауссовской логарифмической функции правдоподобия, формально можно использовать и для сигналов, отличных от нестационарного гауссовского шума, в том числе и для авторегрессионного сигнала, возбуждаемого нестационарным белым шумом.

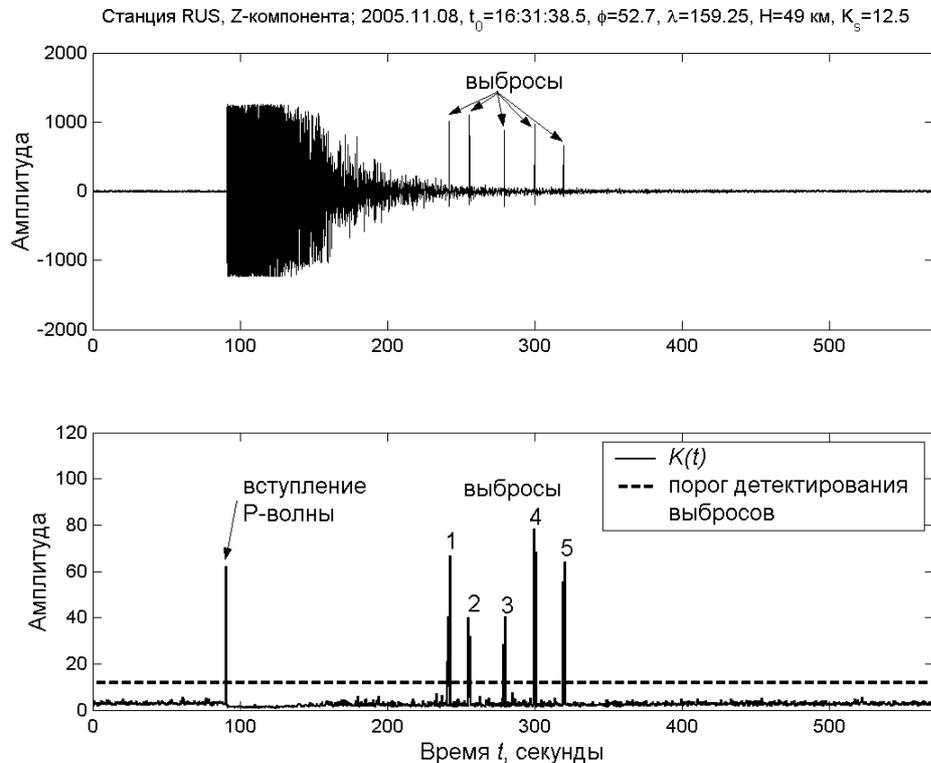


Рис. 2. Пример результата работы алгоритма грубой идентификации выбросов. Верхний график – исходная (зашумленная выбросами) запись землетрясения на радиотелеметрической станции «Русская» (вертикальная компонента). На нижнем графике сплошной линией показана функция $K(t)$, значения которой равны выборочному эксцессу исходной записи в скользящем временном окне длительностью 1 с (128 отсчетов). Горизонтальной штриховой линией отмечен порог детектирования выбросов ($K_{пор}=12$).

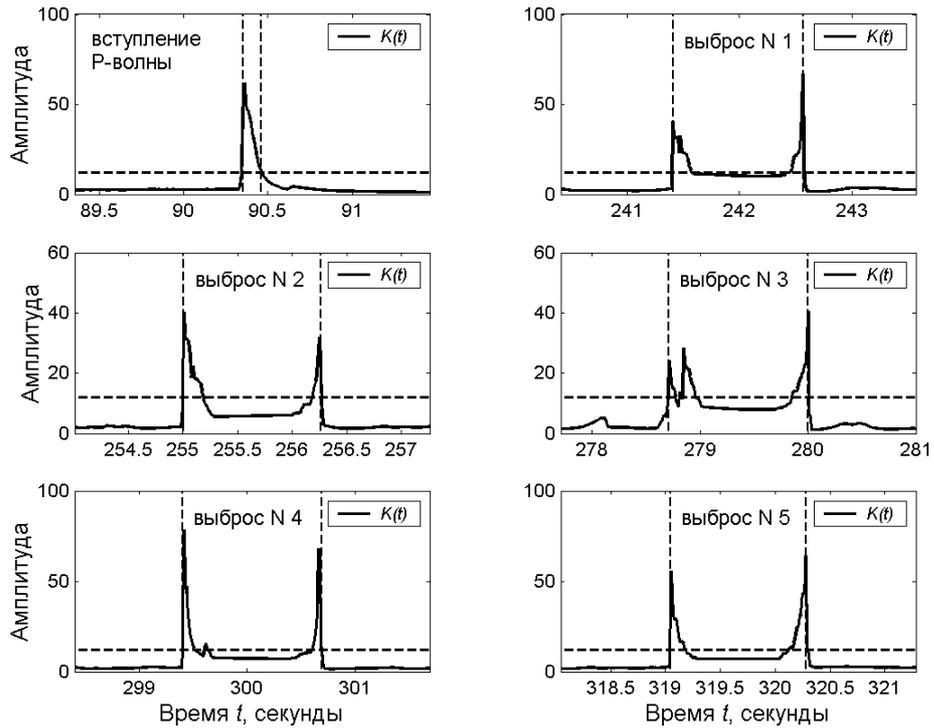


Рис. 3. Пример результата работы алгоритма грубой идентификации выбросов. Детальный вид функции $K(t)$ для каждого из шести сегментов, на которых значение выборочного эксцесса в скользящем временном окне длительностью 1 с превышает пороговое значение $K_{\text{пор}}=12$ (горизонтальная штриховая линия). Эти сегменты ограничены вертикальными штриховыми линиями. Сегмент, длительность которого меньше 1 с (левый верхний график), идентифицируется как фрагмент сигнала, содержащий вступление Р-волны. Оставшиеся сегменты идентифицируются как участки сигнала, содержащие выбросы.

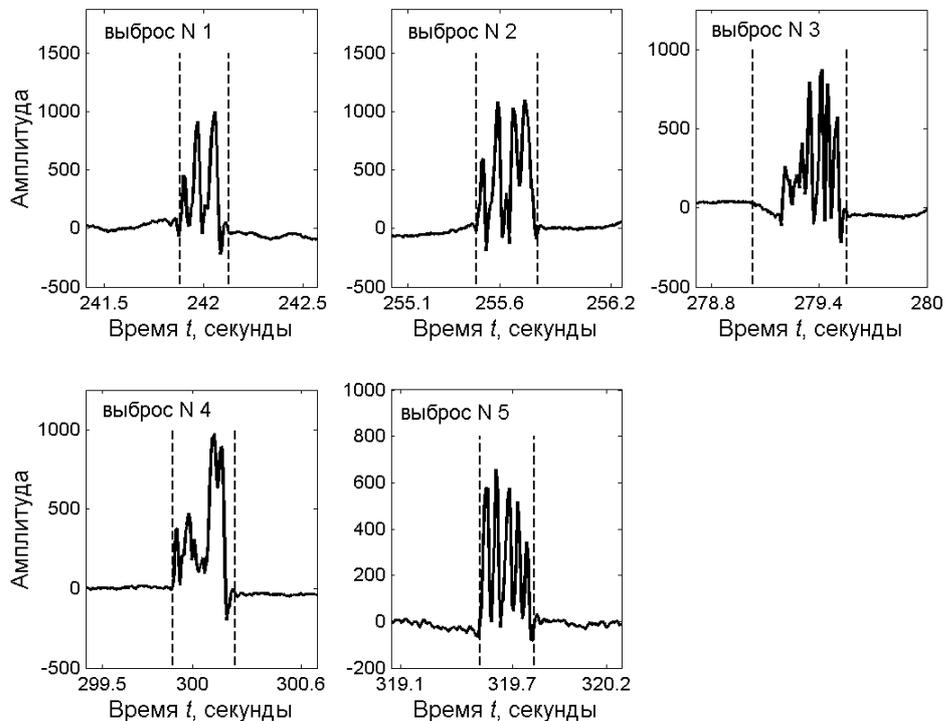


Рис. 4. Пример результата работы алгоритма уточненной идентификации выбросов. На графиках представлены сегменты исходной сейсмограммы, выделенные на этапе грубой идентификации как участки, содержащие выбросы. Вертикальными штриховыми линиями отмечены уточненные моменты вступления и окончания выбросов, полученные путем минимизации целевой функции (6).

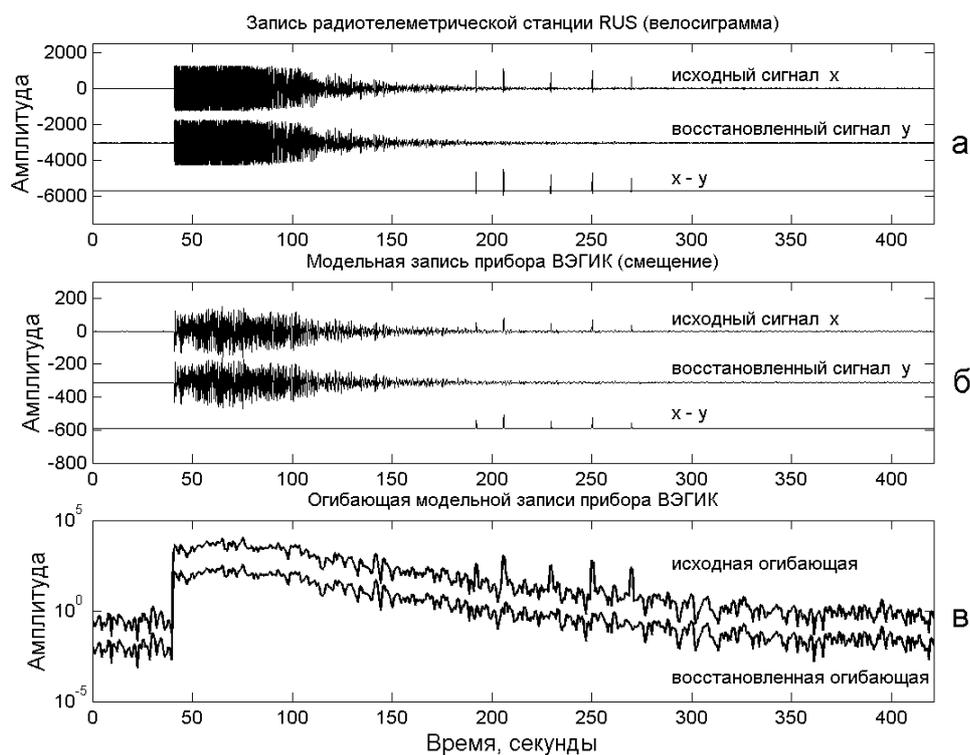


Рис. 5. Пример результатов работы алгоритма реконструкции сигнала, зашумленного выбросами: а - сверху вниз - исходная (зашумленная выбросами) запись землетрясения на радиотелеметрической станции «Русская» (вертикальная компонента), реконструированная запись радиотелеметрической станции, разность между исходной и реконструированной записями; б - сверху вниз – исходная модельная запись прибора ВЭГИК, реконструированная модельная запись прибора ВЭГИК, разность между исходной и реконструированной модельными записями прибора ВЭГИК; в - огибающие исходной и реконструированной модельных записей прибора ВЭГИК.

Заключение

Разработанный алгоритм позволяет успешно идентифицировать выбросы на записях радиотелеметрических станций; реконструировать сигнал на участках, зашумленных выбросами, и воспроизводить реалистическую форму огибающих коды модельных записей прибора ВЭГИК.

Список литературы

1. Гусев А.А., Лемзиков В.К. Предварительные результаты изучения вариаций огибающих коды близких землетрясений перед Усть-Камчатским землетрясением 1971 г. // Вулканология и сейсмология. 1980. № 6. С. 82-93.
2. Гусев А.А., Лемзиков В.К. Аномалии характеристик кода-волн слабых землетрясений перед тремя сильными землетрясениями Курило-Камчатской зоны. // Вулканология и сейсмология. 1984. № 4. С. 76-90.
3. Gusev A.A., Lemzikov V.K. Properties of scattered elastic waves in the lithosphere of Kamchatka: parameters and temporal variations. // Tectonophysics. 1985. V. 112. P. 137-153.
4. Etter W. Restoration of a discrete-time signal segment by interpolation based on the left-sided and right-sided autoregressive parameters. // IEEE Transactions on Signal Processing. 1996. V. 44, № 5. P. 1124-1135.
5. Kauppinen I., Kauppinen, J. Reconstruction method for missing or damaged long portions of audio signal data. // J. Audio. Eng. Soc. 2002. V. 50. № 7/8. P. 594-602.
6. Lavielle M. Using penalized contrasts for the change-point problem. // Signal Processing. 2005. V. 85. № 8. P. 1501-1510.