

УДК 519.712:553.98

© Коллектив авторов

В.И. Галуев, В.Е. Федотов, Д.П. Земцова, А.А. Никитин

ПРОГРАММНО-АЛГОРИТМИЧЕСКОЕ ОБЕСПЕЧЕНИЕ МНОГОАТРИБУТНОГО АНАЛИЗА ГЕОПОЛЕЙ И РЕЗУЛЬТАТЫ ЕГО ПРИМЕНЕНИЯ

Рассматриваемое программно-алгоритмическое обеспечение многоатрибутного анализа геополей, т.е. их комплексного анализа, основано на алгоритме многопризнаковой фильтрации геофизических полей [3] и его программной реализации с использованием систем MatLAB и ГИС ИНТЕГРО.

Алгоритм многопризнаковой фильтрации базируется на разложении Карунена-Лоева, известного в метеорологии как метод естественных ортогональных функций, а в геофизике – как метод главных компонент.

Разложение Карунена-Лоева сводится к линейному преобразованию исходной системы L -случайных величин X_l (в качестве которых в геофизике выступают значения L -полей или L -атрибутов (признаков), измеренных вдоль профиля (трассы, скважины), наблюдаемых по n пикетам), к новой системе L -случайных величин Y_j , называемых главными компонентами, т.е.

$$Y_j = \sum_{l=1}^L a_{lj} X_l + e_l, \quad l = 1, \dots, L, \quad (1)$$

где a_{lj} – коэффициенты перехода от одной системы случайных величин к другой системе той же размерности; e_l – погрешности измерения величин X_l , которыми обычно пренебрегают.

При этом первая главная компонента принимает вид:

$$Y_1 = \sum_{l=1}^L a_{1l} X_l = \vec{a}'_1 X, \quad (2)$$

где \vec{a}'_1 – матрица-строка, а X – матрица исходных (измеренных) геополей размерностью $n \times L$, т.е.

$$X = \begin{pmatrix} X_{11} & \dots & X_{L1} \\ \dots & \dots & \dots \\ X_{1n} & \dots & X_{Ln} \end{pmatrix}. \quad (3)$$

Разложение (1) осуществляется в предположении, что первая главная компонента Y_1 обладает наибольшей дисперсией по отношению к дисперсиям всех остальных главных компонент, т.е.

$$D(Y_1) > D(Y_2) > D(Y_3) \dots \quad (4)$$

Такое разложение проводится с учетом нормировки коэффициентов a_{lj} : $\sum_{l=1}^L a_{lj}^2 = 1$, или

$$\sum_{l=1}^L a_{lj}^2 = \vec{a}'_j I \vec{a}_j = 1, \quad (5)$$

где I – единичная матрица.

Обозначим через R – корреляционную матрицу для системы случайных величин X_l , т.е.

$$R = M X' X, \quad (6)$$

где M – знак математического ожидания.

Тогда дисперсия Y_j будет равна:

$$\begin{aligned} D Y_j &= M (Y_j - M Y_j)^2 = M (Y_j - M Y_j)' (Y_j - M Y_j) = \\ &= M (\vec{a}'_j X - \vec{a}'_j M X) (\vec{a}'_j X - \vec{a}'_j M X) = \vec{a}'_j R \vec{a}_j. \end{aligned}$$

С учетом нормировки коэффициентов a_{lj} получаем отношение:

$$\lambda_j = \frac{\vec{a}'_j R \vec{a}_j}{\vec{a}'_j I \vec{a}_j}. \quad (7)$$

Физический смысл нормировки коэффициентов a_{lj} состоит в том, чтобы масштабы новой системы случайных величин Y_j совпадали бы с масштабом измеренных величин X_l .

Физический смысл выражения (7) соответствует отношению энергии (дисперсии) L -случайных величин к их некоррелированной составляющей, или иначе – величина λ_j представляет энергетическое отношение «сигнал/некоррелируемая помеха» в L -мерном пространстве геополей или атрибутов (признаков).

Для получения максимума дисперсии Y_j следует продифференцировать выражение (7) по \vec{a}'_j , а производную приравнять к нулю, т.е.

$$\frac{\partial}{\partial \vec{a}'_j} (\vec{a}'_j R \vec{a}_j - \lambda_j \vec{a}'_j I \vec{a}_j) = 0.$$

В результате дифференцирования получаем:

$$R \vec{a}'_j - \lambda_j I \vec{a}'_j = 0,$$

или

$$\vec{a}'_j (R - \lambda_j I) = 0. \quad (8)$$

Из выражения (8) следует, что λ_j представляет собственное значение исходной матрицы R , а вектор является собственным вектором той же матрицы для λ_j .

Соответственно, дисперсия главной компоненты:

$$DY = \vec{a}_j' R \vec{a}_j = \vec{a}_j' \lambda_j \vec{a}_j = \lambda_j,$$

ввиду ортогональности собственных векторов, совпадает с собственным значением матрицы R . Таким образом, последовательность вычисления первой главной компоненты, в данном случае как комплексного параметра для L -полей вдоль профиля, сводится к реализации следующих процедур:

- построение корреляционной матрицы R для L измеренных геополей (или атрибутов одного поля);
- нахождение максимального собственного значения λ_{\max} матрицы R ;
- вычисление координат собственного вектора \vec{a}_{i1} , соответствующего λ_{\max} ;
- нахождение значений первой главной компоненты $Y_1 = \sum_{i=1}^L a_{i1} X_i$.

Аналогичные процедуры реализуются при нахождении последующих главных компонент при $\lambda_2 < \lambda_{\max}$, $\lambda_3 < \lambda_2$ и т.д.

Первая главная компонента имеет прозрачный физический смысл, поскольку она обладает наибольшей энергией (дисперсией), т.е. с ее помощью выделяются наиболее энергетически выраженные области в пространстве L -полей (атрибутов).

На основе второй главной компоненты, как правило, оконтуриваются области околорудных изменений.

Физическая и геологическая интерпретация последующих главных компонент часто затруднена или вообще невозможна.

На базе разложения (1), т.е. метода главных компонент, реализована двумерная адаптивная энергетическая фильтрация наблюдаемого поля, которая обеспечивает его разложение на региональную (тренд) и локальные составляющие по уменьшению их энергии [1, 2].

Комплексный анализ геополей и их атрибутов методом главных компонент используется практически во всех современных геоинформационных системах и технологиях [4]. Однако его реализация ограничивается расчетом первой главной компоненты по комплексу полей для каждого профиля сети наблюдений в отдельности; для комплексного анализа L -полей по сети наблюдений ($n \times N$) точек (n – число пикетов, N – число профилей) этот метод остается до сих пор нереализованным из-за сложности

учета корреляционных связей между полями по всей площади наблюдений.

Алгоритм многопризнаковой фильтрации или многоатрибутного анализа по сети профилей в работе [3] сводится к двухэтапной процедуре метода главных компонент. На первом этапе осуществляется вычисление первой главной компоненты по комплексу атрибутов для каждого профиля сети в отдельности, т.е. как при традиционном варианте метода. На втором этапе проводится расчет первой главной компоненты для уже полученных первых компонент по каждому профилю, что позволяет учесть корреляционные связи между атрибутами по всей площади наблюдений. При этом образуется матрица размерности $n \times N$ из таких главных компонент:

$$\begin{pmatrix} Y_{111} & Y_{112} & \dots & Y_{11N} \\ Y_{121} & Y_{122} & \dots & Y_{12N} \\ \dots & \dots & \dots & \dots \\ Y_{1n1} & Y_{1n2} & \dots & Y_{1nN} \end{pmatrix}. \tag{9}$$

В матрице (9) первый индекс указывает на значение первой главной компоненты, второй индекс – на номер пикета, а третий индекс – на номер профиля.

Далее проводится расчет коэффициентов корреляции между любыми парами главных компонент и построение по этим коэффициентам корреляционной матрицы R размерностью $N \times N$:

$$[R] = \begin{pmatrix} 1 & R_{12} & \dots & R_{1N} \\ R_{21} & 1 & \dots & R_{2N} \\ \dots & \dots & \dots & \dots \\ R_{N1} & R_{N2} & \dots & 1 \end{pmatrix}, \quad R_{ij} = R_{ji} \tag{10}$$

В матрице (10) индекс для первых главных компонент опущен.

Для матрицы (10) вычисляется максимальное собственное значение λ_{\max} , которое характеризует энергетическое отношение сигнал/помеха уже в многоатрибутном пространстве по всей площади наблюдений.

Нахождение значений собственного вектора $(a_{11}, a_{12}, \dots, a_{1N})$ матрицы (10), соответствующего λ_{\max} , реализуется из системы линейных уравнений (11):

$$\begin{pmatrix} 1 - \lambda_{\max} & R_{12} & \dots & R_{1N} \\ R_{21} & 1 - \lambda_{\max} & \dots & R_{2N} \\ \dots & \dots & \dots & \dots \\ R_{N1} & R_{N2} & \dots & 1 - \lambda_{\max} \end{pmatrix} \begin{pmatrix} a_{11} \\ a_{12} \\ \dots \\ a_{1N} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \dots \\ 0 \end{pmatrix}. \tag{11}$$

Определенность решения системы линейных уравнений (11) при равенстве всех уравнений в правой их части нулю обеспечивается нормировкой

$$\sum_{j=1}^N a_{ij}^2 = 1.$$

Значения собственного вектора $(a_{11}, a_{12}, \dots, a_{1N})$ характеризуют весовые коэффициенты каждого профиля в многоатрибутном пространстве.

Нахождение аналогичных весовых коэффициентов для каждого пикета сети наблюдений осуществляется путем умножения вектора-столбца $(a_{11}, a_{12}, \dots, a_{1N})^T$ на матрицу (9):

$$\begin{pmatrix} Y_{111} & Y_{112} & \dots & Y_{11N} \\ Y_{121} & Y_{122} & \dots & Y_{12N} \\ \dots & \dots & \dots & \dots \\ Y_{1n1} & Y_{1n2} & \dots & Y_{1nN} \end{pmatrix} \begin{pmatrix} a_{11} \\ a_{12} \\ \dots \\ a_{1N} \end{pmatrix} = \begin{pmatrix} b_{11} \\ b_{21} \\ \dots \\ b_{n1} \end{pmatrix}. \quad (12)$$

Результатом перемножения таких матриц являются значения $(b_{11}, b_{n1}, \dots, b_{n1})$, которые характеризуют вес каждого пикета в многоатрибутном пространстве.

Расчет комплексного параметра для L -атрибутов (L -геополей) по сети $n \times N$ точек площади наблюдений осуществляется путем перемножения вектора-столбца весовых коэффициентов для каждого пикета $(b_{11}, b_{n1}, \dots, b_{n1})^T$ на вектор-строку весовых коэффициентов каждого профиля $(a_{11}, a_{12}, \dots, a_{1N})$, т.е.

$$\begin{pmatrix} b_{11} \\ b_{21} \\ \dots \\ b_{n1} \end{pmatrix} (a_{11} \ a_{12} \ \dots \ a_{1N}) = \begin{pmatrix} b_{11}a_{11} & a_{11}a_{12} & \dots & b_{11}a_{1N} \\ b_{21}a_{11} & b_{21}a_{12} & \dots & b_{21}a_{1N} \\ \dots & \dots & \dots & \dots \\ b_{n1}a_{11} & b_{n1}a_{12} & \dots & b_{n1}a_{1N} \end{pmatrix}. \quad (13)$$

При этом для каждого столбца полученной матрицы (13) следует добавить среднее значение соответствующей первой компоненты, поскольку при расчете коэффициентов корреляции в матрице (10) средние значения вычитаются.

Описанные процедуры расчета комплексного параметра используются для наблюдения L -геофизических полей при их съемке с поверхности по N -профилям, обеспечивая выделение наиболее энергетически выраженной области в пространстве измеренных полей, или иначе – перспективного аномального объекта.

При комплексном анализе сейсмических атрибутов, полученных по временному разрезу, следует ограничить область анализа по времени целевым горизонтом в $0,1 \div 0,25$ с, а по оси наблюдений на поверхности – предполагаемыми размерами нефтегазовой залежи. Нефтегазовая залежь в пространстве L -атрибутов будет подчеркнута максимальными значениями первой главной компоненты, рассчитанной по формуле (13).

Надежное выделение контура самой залежи осуществляется по значению изолинии первой главной компоненты, превышающей ее двойную среднеквадратичную погрешность, т.е. величину $2\sqrt{\lambda_{\max}}$, поскольку λ_{\max} представляет дисперсию первой главной компоненты.

Программное обеспечение многоатрибутного анализа создано в виде скрипта filter_PCA для пакета MatLAB и в виде утилиты для среды ГИС ИНТЕГРО. Для запуска скрипта из консоли MatLAB, нужно создать трехмерную матрицу данных $N \times n \times L$ значений поля, где N – количество профилей съемки, n – количество точек измерения (пикетов) поля на профиле, L – количество расчетных полей. Созданная матрица передается как первый аргумент в функцию filter_PCA, второй аргумент – номер главной компоненты. Результат выполнения – матрица $N \times n$, содержащая значения соответствующих главных компонент.

В качестве примеров использования созданного программного обеспечения рассматривается выделение перспективных нефтегазоносных объектов по комплексному анализу временного разреза (рис. 1), по которому были рассчитаны шесть атрибутов волнового поля целевого горизонта – анизотропия, дисперсия, полный градиент, энергия, частота и энтропия (приведены на рис. 2). Для такого анализа был взят интервал временного разреза в 0,2 с (100 дискретов $\Delta t = 0,002$ с), а по оси X – 4 км (80 дискретов $\Delta x = 25$ м). Для данного интервала, были рассчитаны перечисленные атрибуты в окне, включающем 11 трасс и 21 дискрет по времени. Ни по одному атрибуту перспективный нефтегазоносный объект не выделяется однозначно. На рис. 3 представлен результат многопризнакового анализа по алгоритму двухэтапной процедуры метода главных компонент, на котором максимальные значения первой главной компоненты четко фиксируют местоположение двух нефтегазоперспективных объектов, приуроченных к крыльям антиклинальной структуры (рис. 3).

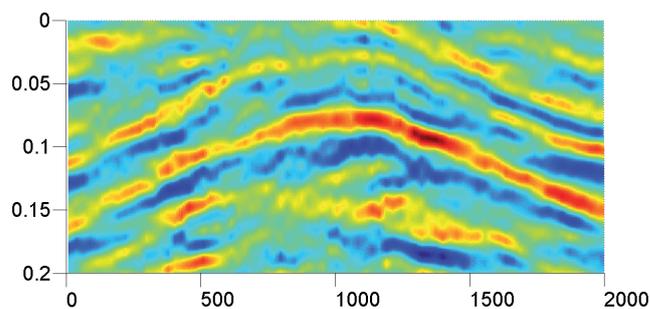


Рис. 1. Исходный временной разрез в интервале целевого горизонта, зафиксированного на времени 2,9 с

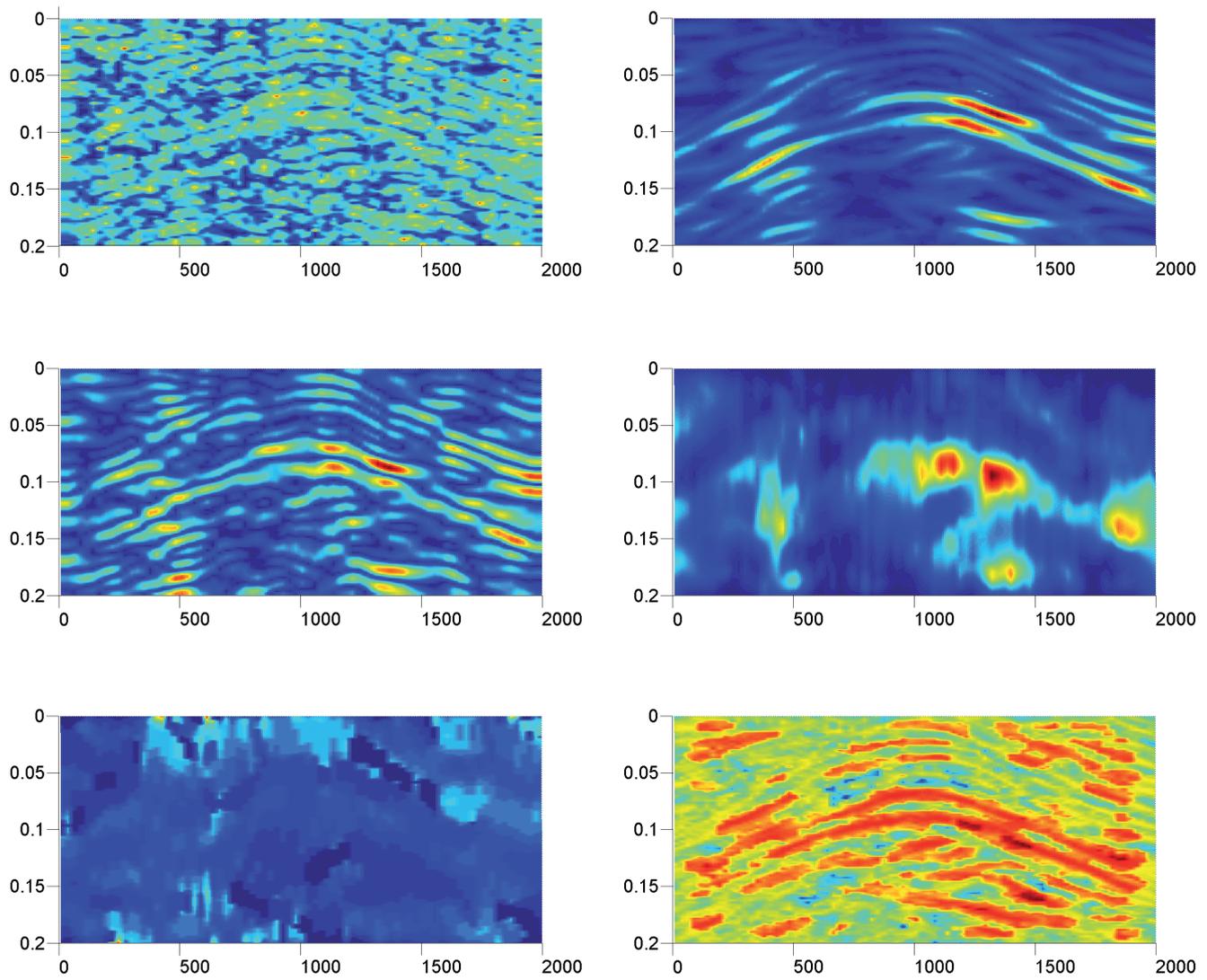


Рис. 2. Анизотропия, дисперсия, полный градиент, энергия, частота, энтропия

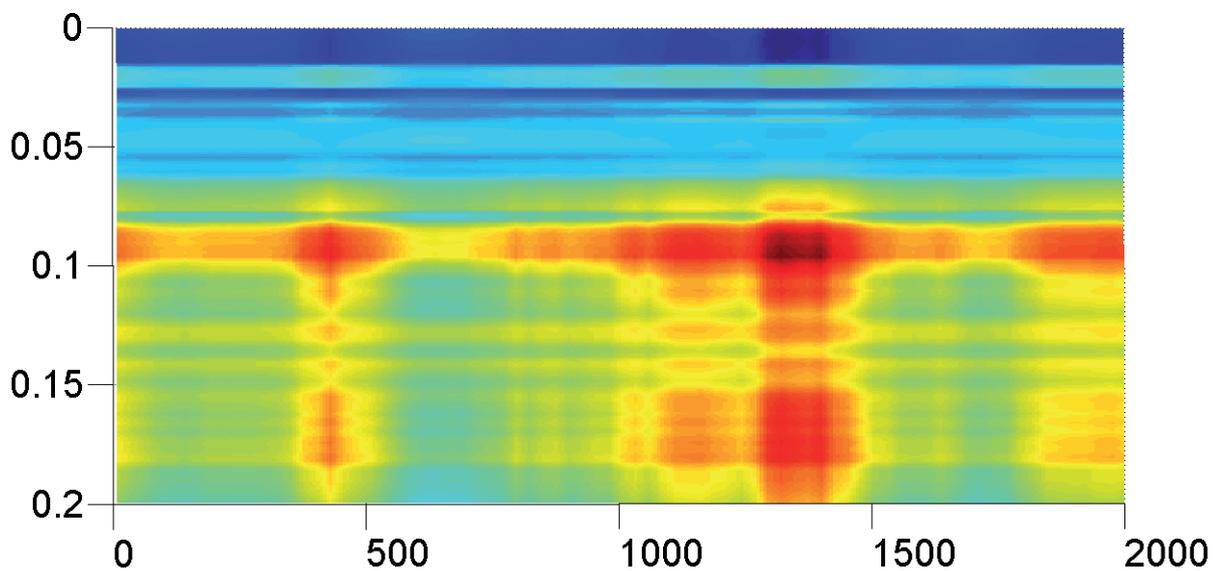


Рис. 3. Результат фильтрации

Аналогичный результат получен при многопризнаковой фильтрации для более глубокого перспективного объекта на времени 4 с, с использованием тех же шести атрибутов волнового поля. На рис. 4 приведен временной разрез целевого горизонта 0,3 с и по оси X – 16 км. На рис. 5 – результат многопризнаковой фильтрации, на котором в двух временных интервалах выделяются перспективные объекты, четко локализованные в пространственно-временной области и также приуроченные к крыльям структуры.

Ключевые слова: метод главных компонент, многопризнаковые фильтрации, нефтегазовые объекты.

ЛИТЕРАТУРА

1. Галуев В.И., Каплан С.А., Никитин А.А. Технология создания физико-геологических моделей

земной коры по региональным профилям на основе геоинформационных систем. – М. : ВНИИгеосистем, 2009. – 231 с.

2. Карасевич А.М., Земцова Д.П., Никитин А.А. Новые технологии геофизических исследований при поисках и прогнозе углеводородного сырья. – М. : Страхование ревью, 2010. – 160 с.

3. Никитин А.А. Алгоритм многопризнаковой фильтрации геофизических полей // Геоинформатика. – 2012. – № 1. – С. 51-53.

4. Никитин А.А, Петров А.В. Теоретические основы обработки геофизической информации. – М. : ООО Центр информационных технологий в природопользовании, 2010. – 114 с.

5. Черемисина Е.Н., Никитин А.А. Геоинформационные системы и технологии. – М : ВНИИгеосистем, 2011. – 375 с.

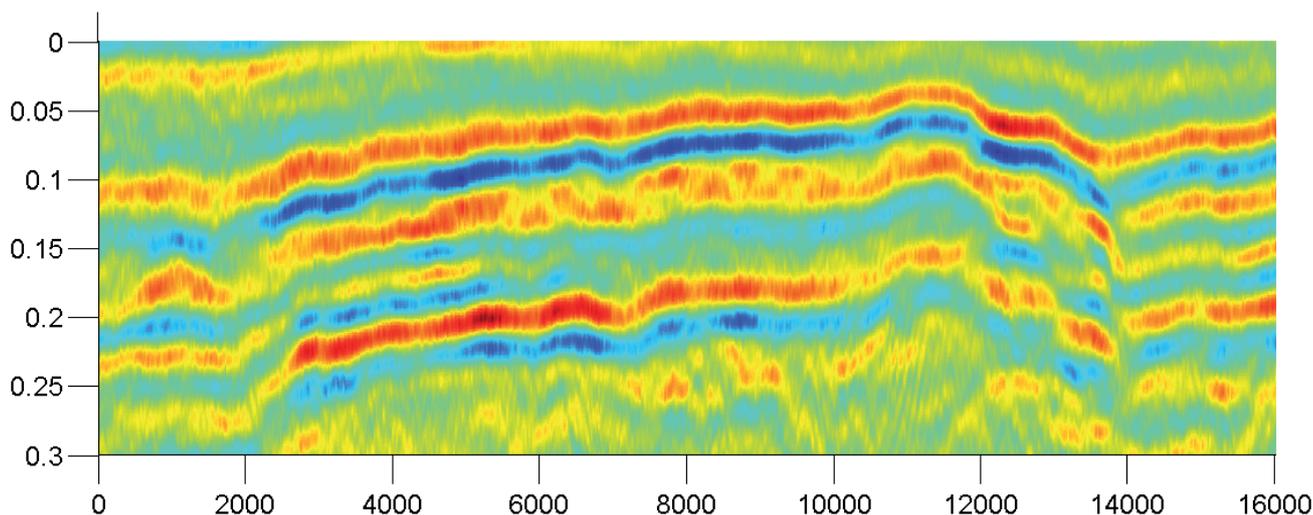


Рис. 4. Временной разрез

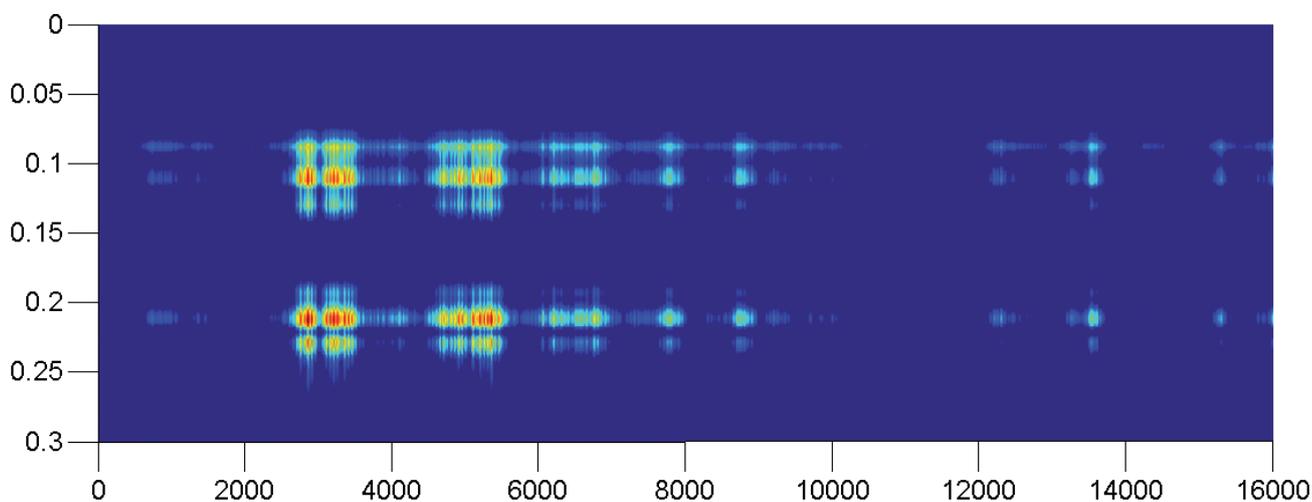


Рис. 5. Результат многопризнаковой фильтрации