

Вестник Евразийской науки / The Eurasian Scientific Journal <https://esj.today>

2018, №1, Том 10 / 2018, No 1, Vol 10 <https://esj.today/issue-1-2018.html>

URL статьи: <https://esj.today/PDF/29ITVN118.pdf>

Статья поступила в редакцию 07.02.2018; опубликована 02.04.2018

**Ссылка для цитирования этой статьи:**

Буторин А.В., Краснов Ф.В., Михеенков А.В. Восстановление коэффициентов отражения среды по сейсмическим данным при помощи методов машинного обучения // Вестник Евразийской науки, 2018 №1, <https://esj.today/PDF/29ITVN118.pdf> (доступ свободный). Загл. с экрана. Яз. рус., англ.

**For citation:**

Butorin A.V., Krasnov F.V., Mikheenkov A.V. (2018). Reconstruction of medium's reflectivity coefficients based on seismic data through machine learning. *The Eurasian Scientific Journal*, [online] 1(10). Available at: <https://esj.today/PDF/29ITVN118.pdf> (in Russian)

УДК 316.452

**Буторин Александр Владимирович**

ООО «Газпромнефть НТЦ», Санкт-Петербург, Россия  
Эксперт

E-mail: [Butorin.AV@gazpromneft-ntc.ru](mailto:Butorin.AV@gazpromneft-ntc.ru)

ORCID: <http://orcid.org/0000-0002-6074-1439>

РИНЦ: [https://elibrary.ru/author\\_profile.asp?id=877389](https://elibrary.ru/author_profile.asp?id=877389)

SCOPUS: <http://www.scopus.com/authid/detail.url?authorId=56370048400>

**Краснов Федор Владимирович**

ООО «Газпромнефть НТЦ», Санкт-Петербург, Россия  
Эксперт

Кандидат технических наук

E-mail: [Krasnov.FV@Gazprom-Neft.ru](mailto:Krasnov.FV@Gazprom-Neft.ru)

ORCID: <http://orcid.org/0000-0002-9881-7371>

РИНЦ: [https://elibrary.ru/author\\_profile.asp?id=855886](https://elibrary.ru/author_profile.asp?id=855886)

Researcher ID: <http://www.researcherid.com/rid/C-6518-2018>

**Михеенков Андрей Витальевич**

ФГБУН «Институт физики высоких давлений им. Л.Ф. Верещагина Российской Академии Наук», Москва, Россия  
Ведущий научный сотрудник

Доктор физико-математических наук, доцент

E-mail: [mikheen@bk.ru](mailto:mikheen@bk.ru)

Researcher ID: <http://www.researcherid.com/rid/L-2103-2013>

**Восстановление коэффициентов  
отражения среды по сейсмическим данным  
при помощи методов машинного обучения**

**Аннотация.** Сейсморазведка один из немногих универсальных геофизических методов получения информации о недрах. Отраженный сигнал, содержащийся в сейсмических данных несет информацию о свойствах среды через которую он прошел. Коэффициенты отражения определяются изменением упругих свойств среды и являются основой для интерпретации сейсмических данных, а также прогноза геологического строения. В настоящем исследовании авторы разработали новый метод обработки сейсмических данных, который позволяет определить положение отражающих плоскостей и значения коэффициентов отражения с высокой точностью. Для решения задачи авторы использовали методику частичного обучения (Semi-supervised learning). Применение методов машинного обучения позволило создать

математическую модель, оптимизировать ее параметры на синтетических данных для дальнейшего применения модель на размеченных сейсмических данных. Основным новшеством стала разработка алгоритма обучения использующего свертку сигнала и регуляризацию коэффициентов отражения. Полученная авторами модель показала высокую точность на синтетических сейсмических данных с высокой плотностью отражающих плоскостей ( $10^2$  плоскостей на секунду трассы). Достигнутый таким образом низкий уровень ошибки позволяет существенно улучшить количественное понимание структуры недр на основании сейсмических данных и является надежной основой для построения геологических моделей.

**Ключевые слова:** сейсморазведка; методы машинного обучения; оптимизационная проблема; регуляризация

### Введение

Современные методы поиска коэффициентов разложения сигналов с известными составляющими базируются на подходе Matching Pursuit, предложенным в работе [1]. В настоящее время существует множество усовершенствований алгоритма Matching Pursuit, например, Batch Orthogonal Matching Pursuit [2], Stabilized Orthogonal Matching Pursuit [3] и Hierarchical Matching Pursuit [4].

Рассмотрим подробнее алгоритмы на основе Matching Pursuit применительно к данным сейсморазведки. Алгоритм OMP [2] решает задачу определения коэффициентов разложения заданного сигнала на основании словаря (например, с вейвлетами [5, 9, 10]). При этом алгоритм OMP основывается на понятии остаточной ошибки разложения.

#### *Алгоритм 1. Алгоритм работы OMP [2]*

---

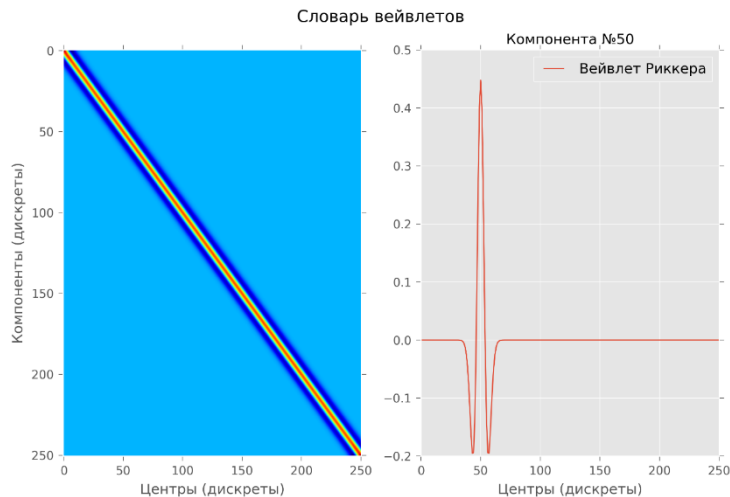
**Algorithm 1** ORTHOGONAL MATCHING PURSUIT

---

```
1: Input: Dictionary  $\mathbf{D}$ , signal  $\underline{x}$ , target sparsity  $K$  or target error  $\epsilon$ 
2: Output: Sparse representation  $\underline{\gamma}$  such that  $\underline{x} \approx \mathbf{D}\underline{\gamma}$ 
3: Init: Set  $I := ()$ ,  $\underline{r} := \underline{x}$ ,  $\underline{\gamma} := \underline{0}$ 
4: while (stopping criterion not met) do
5:    $\hat{k} := \text{Argmax}_k |d_k^T \underline{r}|$ 
6:    $I := (I, \hat{k})$ 
7:    $\underline{\gamma}_I := (\mathbf{D}_I)^+ \underline{x}$ 
8:    $\underline{r} := \underline{x} - \mathbf{D}_I \underline{\gamma}_I$ 
9: end while
```

---

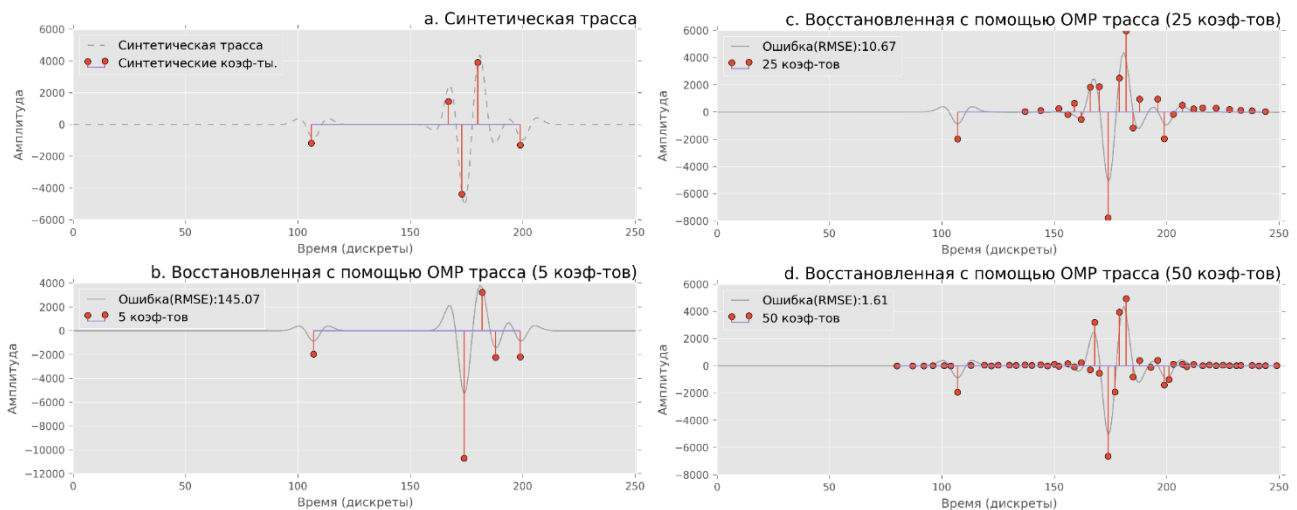
В результате работы алгоритма OMP (алгоритм 1) можно достигнуть очень высокой точности разложения сигнала на разреженные коэффициенты при заданном словаре. Чтобы показать преимущества и недостатки алгоритма OMP применительно к сейсмическим данным создадим словарь из одного вейвлета Риккера с частотой 30 Гц. Полученный словарь изображен на рисунке (рисунок 1).



**Рисунок 1.** Словарь вейвлетов с частотой 30 Гц. На рисунке справа показан один вейвлет из компоненты словаря номер 50 (разработано авторами)

Отметим, что данный словарь создан для трасс длиной 250 дискретов и шагом 2 мс. С помощью алгоритма ОМР и такого словаря вейвлет одинаковой формы может быть обнаружен в любом из 250 дискретов трассы. Алгоритм ОМР определяет коэффициент для каждого вейвлета из словаря для конкретной трассы.

Рассмотрим подробнее свойства получаемого ОМР разложения по словарю (рисунок 1) на примере одной сейсмической трассы.



**Рисунок 2.** На рисунке а. отображена синтетическая трасса. На рисунках b, c, d отображены получаемые в результате ОМР разложения коэффициенты и восстановленные по этим коэффициентам трассы (разработано авторами)

На рисунке (рисунок 2а) изображена одна сейсмическая трасса, полученная с помощью свертки коэффициентов отражения среды и вейвлета Риккера с частотой 30 Гц. Далее при разложении синтетической трассы методом ОМР использовалось условие на количество ненулевых коэффициентов разложения и измерялась ошибка root-mean-square error (RMSE). Как мы видим ошибка разложения достаточно велика (145.07) для 5 коэффициентов (рисунок 2b), хотя изначально синтетическая трасса была построена именно на 5 коэффициентах отражения.

В работе [2] рекомендуется использовать количество коэффициентов разложения равное 10 % от длины сигнала. При длине трассы в 250 дискретов – это 25 коэффициентов. При

увеличении числа коэффициентов разложения до 25 (рисунок 2с) ошибка восстановления трассы становится более приемлемой (10.67), но соответствия между коэффициентами разложения и синтетическими коэффициентами отражения не наблюдается.

В результате разложения по 50 ненулевым коэффициентам (рисунок 2с) у алгоритма ОМР получается, что возникает много небольших коэффициентов, которые уменьшают ошибку RMSE, но не изменяют уже выбранные коэффициенты.

Невысокая точность и появление ложных плоскостей отражения делает разложение на коэффициенты по словарю с помощью ОМР неприемлемым для задач поиска коэффициентов отражения. Таким образом, можно сделать предварительное заключение о нефизической природе коэффициентов разложения, получаемых в результате применения алгоритма ОМР.

В данном исследовании авторы сосредоточились на поиске композиции вейвлетов минимизирующих ошибку восстановления трассы при заданном положении коэффициентов.

Авторы выдвинули следующую исследовательскую гипотезу:

***Гипотеза:** Существует алгоритм разложения сейсмической трассы на такие коэффициенты разложения, которые с высокой точностью соответствуют коэффициентам отражения среды, по количеству, амплитуде и положению на трассе.*

Далее авторы исследовали возможности построения алгоритмов с помощью методов машинного обучения, создали методику обучения учитывающую физические особенности среды и провели несколько цифровых экспериментов на синтетических сейсмических трассах.

Данная статья состоит из введения, методической части, результатов эксперимента и заключения.

### Методика

Согласно [7] задача, решаемая авторами данного исследования, относится к классу обратных коэффициентных задач. Пусть изучаемый в эксперименте процесс моделируется решением задачи

$$L_{\theta}[u] = g(x, \theta), x \in X \subset \mathbb{R}^K \quad (1)$$

с дополнительными условиями

$$l_{\theta}[u] = h(x, \theta), x \in \delta X \quad (2)$$

Здесь:  $x = \{x_1, x_2, \dots, x_k\}$  – набор так называемых контролируемых переменных,  $\theta \in \Omega$  – совокупность некоторых параметров,  $L_{\theta}[\cdot]$  – детерминированный дифференциальный оператор, зависящий от  $\theta$ ,  $\mathbb{R}^K$  – евклидово пространство размерности  $K$ ,  $\delta X$  – граница множества  $X$ .

На практике величины  $\theta$  неизвестны и поэтому возникает следующая обратная задача: оценить исходные параметры  $\theta$  и функцию отклика  $u = (x, \theta)$  для модели (1)-(2) по экспериментальным данным, если в эксперименте наблюдаются некоторые функционалы  $b[u]$  от отклика  $u$ .

В качестве экспериментальных данных в данном исследовании будут доступны только сейсмические трассы и вейвлет, по которым нужно будет определить с приемлемой точностью коэффициенты отражения среды.

Для решения данной задачи авторы применили методологию машинного обучения. Применение машинного обучения для решения инженерных задач уже было использовано

авторами в работе [8]. В качестве размеченных данных используется синтетическая трасса, построенная по заданным синтетическим коэффициентам отражения. Алгоритм  $\mathcal{A}0$  обучается подбирать коэффициенты разложения, соответствующие синтетическим коэффициентам, на основании которых построена синтетическая трасса.

Формальная постановка данной обратной задачи такова; есть дискретный синтетический сигнал (трасса), представленный в виде вектора  $T \in \mathbb{R}^N$ . Трасса  $T$  построена с помощью вейвлета Риккера  $W$  и  $K$  отражающих плоскостей при помощи функции свертки. Каждая отражающая плоскость  $k$  характеризуется номером дискрета  $t \in \mathbb{N}^K$  и определённым коэффициентом отражения  $r \in \mathbb{R}^K$ . Тогда процесс создания трассы может быть представлен формулой (формула 1):

**Формула 1.** Формула получения трассы из вейвлетов и отражающих плоскостей (разработана авторами)

$$T = \sum_i^K W(t_i) * r_i$$

На основании только данных о трассе  $T$  необходимо получить алгоритм  $\mathcal{A}0$  определяющий коэффициенты отражения  $\tau \in \mathbb{N}^M$  и  $\rho \in \mathbb{R}^M$  удовлетворяющие следующим условиям:

1.  $T^{\sim} \approx T$ , где  $T^{\sim} = \sum_i^M W(\tau_i) * \rho_i$ .
2.  $K \approx M$ .
3.  $\tau \approx t$ .
4.  $r \approx \rho$ .

Рассмотрим более подробно условия 1-4 с точки зрения количественной оценки ошибок  $E_{1-4}$ . Ошибка  $E_1$  – мера несовпадений двух трасс вычисляется как нормированная квадратичная сумма отклонений –  $RMSE(T, T^{\sim})$ . Отклонение в количестве исходных и полученных плоскостей ( $E_2$ ) измеряется как модуль их разности –  $abs(K-M)$ . Отклонение в положении отражающих плоскостей ( $E_3$ ) определяется по метрике F1-score. Отличия в амплитудах коэффициентов отражения ( $E_4$ ) вычисляются только для совпадающих по положению отражающих плоскостей как  $RMSE(T[t == \tau], T^{\sim}[t == \tau])$ .

Ошибки  $E_{2-4}$  являются количественными оценками качества работы алгоритма  $\mathcal{A}0$ , но не могут участвовать в оптимизационном. Для поиска оптимального решения необходимо варьировать коэффициенты отражения с целью минимизировать ошибку  $E_1$ .

Помимо условия минимизации ошибки  $E_1$  мы можем учитывать в процессе оптимизации следующие физические условия:

1. Отражающие плоскости не могут быть слишком близко друг к другу ( $U_1$ ).
2. Количество отражающих плоскостей должно быть минимальным ( $U_2$ ).

Количественно физические условия  $U_1$  и  $U_2$  будут выражены в виде функций штрафов  $F_1(U_1)$ ,  $F_2(U_2)$  при определении движения оптимизации. Таким образом мета алгоритм оптимизации можно описать в следующем виде (алгоритм 2):

**Алгоритм 2.** Упрощенный алгоритм оптимизации  $\mathcal{A}0$  (разработано авторами)

- Шаг 1. Выбираем начальные значения для отражающих плоскостей.
- Шаг 2. Вычисляем  $E_1, F_1(U_1), F_2(U_2)$ .

- Шаг 3. Вычисляем изменение отражающих плоскостей.
- Шаг 4. Применяем изменение отражающих плоскостей к текущим значениям.
- Шаг 5. Вычисляем  $E_1$ ,  $F_1(U_1)$ ,  $F_2(U_2)$  и сравниваем со значениями на Шаг 2.
- Шаг 6. Проверяем критерий остановки по значению  $E_1$ , если критерий остановки не выполнен, то переходим к Шаг 3.

Алгоритм  $\mathcal{A}0$  использует информацию о невязке трасс  $T^{\sim}$  и  $T$  для определения направления и размера изменения отражающих плоскостей. На этом основании  $\mathcal{A}0$  можно отнести к классу вариативных оптимизационных алгоритмов, нацеленных на поиск глобального минимума. Протестированные авторами эвристические подходы к оптимизационной задаче на основе работы [6] показали себя менее эффективными.

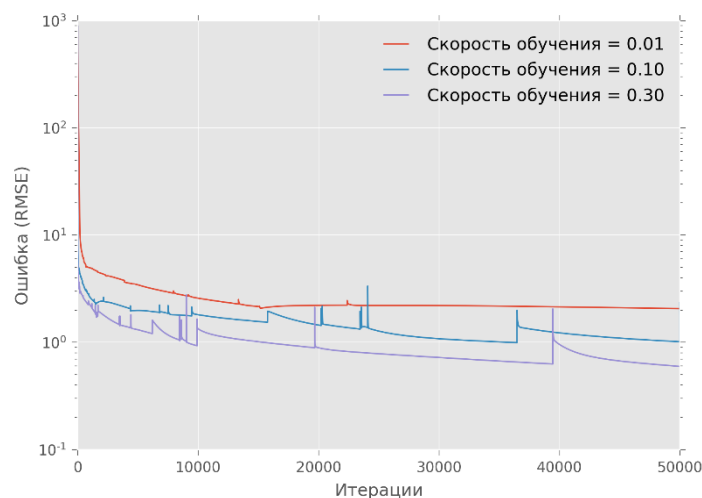
### Результаты эксперимента

Для проверки, описанной выше методики были проведены эксперименты с синтетическими трассами с разной плотностью отражающих плоскостей. По результатам экспериментов были определены оптимальные значения скорости обучения, весов штрафов от  $U_1$  и  $U_2$ .

Отдельно была исследована инициализация вектора плоскостей отражения. Были протестированы три типа инициализации:

1. Инициализация случайными числами по нормальному распределению.
2. Инициализация значениями амплитудами из трассы с масштабирующим коэффициентом.
3. Инициализация амплитудами экстремумов из трассы.

Наиболее быструю сходимость показала инициализация амплитудами экстремумов из трассы. На рисунке (рисунок 3) приведены зависимости функции потерь (loss) от различных значений параметра скорости обучения (learning rate).

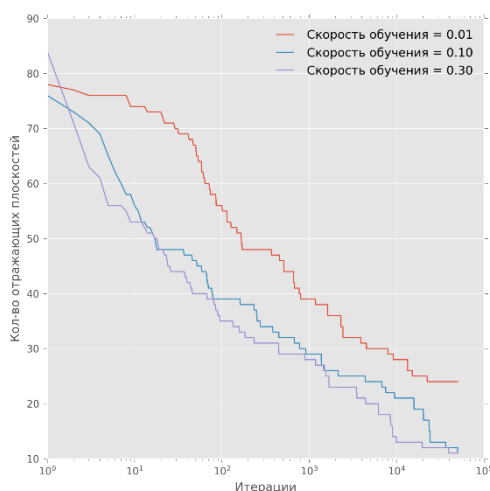


**Рисунок 3.** Функции потерь в зависимости от различных скоростей обучения (разработано авторами)

Зависимости на рисунке 3 показывают стандартное поведение – при увеличении скорости обучения сходимость ускоряется. Применение регуляризации по количеству отражающих плоскостей ( $U_2$ ) можно наблюдать на рисунке 3 в виде выбросов при переходе к



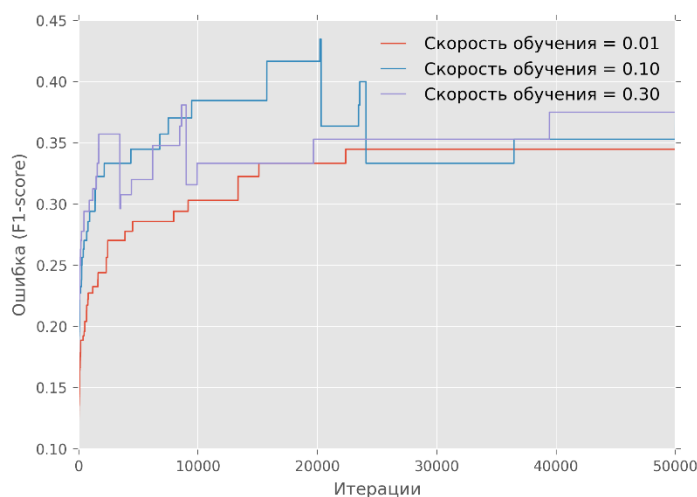
меньшему количеству отражающих плоскостей. Более подробно этот эффект отображен на рисунке (рисунок 4).



**Рисунок 4.** Количество отражающих плоскостей в зависимости от итераций (разработано авторами)

Из зависимости на рисунке 4 мы видим, что при определенном числе итераций количество отражающих плоскостей перестает изменяться. Достижение минимальной постоянной величины служит одним из сигналов к остановке работы алгоритма на ряду с уменьшающейся ошибкой RMSE.

Отклонение в положении отражающих плоскостей ( $E_3$ ) определяется по метрике F1-score. На рисунке 5 отображена зависимость метрики F1-score для различных скоростей обучения алгоритма  $\mathcal{A}0$ .



**Рисунок 5.** Зависимость метрики F1-score для различных скоростей обучения алгоритма  $\mathcal{A}0$  (разработано авторами)

Мы рассмотрели отдельно зависимости для ошибок  $E_1$  (рисунок 3),  $E_2$  (рисунок 4),  $E_3$  (рисунок 5). В таблице (таблица 1) приведено сравнение ошибок  $E_{1-4}$  для алгоритма OMP и  $\mathcal{A}0$ .

Таблица 1

**Сравнение ошибок  $E_{1-4}$  для алгоритма OMP  
и  $\mathcal{A}$  для трассы с пятью отражающими плоскостями**

Алгоритм/Ошибка	OMP, 5 коэф-тов	OMP, 25 коэф-тов	OMP, 50 коэф-тов	$\mathcal{A}$
$E_1$	145.07	10.67	1.61	0.59
$E_2$	0	20	45	0
$E_3$	0.2	0.06	0.04	0.38
$E_4$	829.43	788.61	752.37	10.03

Показательно, что применение алгоритма  $\mathcal{A}$  позволяет достичь существенно меньших ошибок для трасс с пятью отражающими плоскостями. В таблице (таблица 2) приведено сравнение OMP и  $\mathcal{A}$  для трассы с 103 плоскостями (500 дискретов с шагом 2 мс).

Таблица 2

**Сравнение ошибок  $E_{1-4}$  для алгоритма OMP  
и  $\mathcal{A}$  для трассы со 103-мя отражающими плоскостями**

Алгоритм/Ошибка	OMP, 50 коэф-тов	OMP, 103 коэф-та	OMP, 150 коэф-тов	$\mathcal{A}$
$E_1$	39.84	9.40	0.92	0.87
$E_2$	53	0	47	2
$E_3$	0.23	0.25	0.24	0.33
$E_4$	1644.11	1562.16	1517.16	100.23

Как видно из таблицы 2 алгоритм  $\mathcal{A}$  позволяет минимизировать ошибки разложения до уровней ниже чем при OMP. Но при этом алгоритм  $\mathcal{A}$  сохраняет физический смысл коэффициентов разложения.

### Заключение

Авторы разработали алгоритм, позволивший встроить физические законы в методы машинного обучения. Для сравнения эффективности предложенного алгоритма авторами была разработана комплексная метрика точности, состоящая из 4 компонент:

- $E_1$  – мера несовпадений двух трасс вычисляется как нормированная квадратичная сумма отклонений –  $RMSE(T, T^{\sim})$ .
- Отклонение в количестве исходных и полученных в результате работы алгоритма отражающих плоскостей ( $E_2$ ).
- Отклонение в положении отражающих плоскостей ( $E_3$ ) определяется по метрике F1-score.
- Отличия в амплитудах коэффициентов отражения ( $E_4$ ).

Проведенные эксперименты показали, что обученный алгоритм позволяет достигать меньших значений ошибки чем OMP [2] и при этом сохранять физический смысл получаемых коэффициентов разложения.

Дальнейшие исследования в данном направлении целесообразно провести на реальных, не синтетических данных, для определенного месторождения с достаточным количеством исследованных скважин.



## ЛИТЕРАТУРА

1. Mallat S.G., Zhang Z. Matching pursuits with time-frequency dictionaries // IEEE Transactions on signal processing. – 1993. – Т. 41. – № 12. – С. 3397-3415.
2. Rubinstein, R., Zibulevsky, M. and Elad, M., Efficient Implementation of the K-SVD Algorithm using Batch Orthogonal Matching Pursuit Technical Report – CS Technion, April 2008.
3. Saadat S.A., Safari A., Needell D. Sparse reconstruction of regional gravity signal based on Stabilized Orthogonal Matching Pursuit (SOMP) // Pure and Applied Geophysics. – 2016. – Т. 173. – № 6. – С. 2087-2099.
4. Bo L., Ren X., Fox D. Unsupervised feature learning for RGB-D based object recognition // Experimental Robotics. – Springer, Heidelberg, 2013. – С. 387-402.
5. Butorin A.V. et al. Spectral Inversion Methods and its Application for Wave Field Analysis (Russian) // SPE Russian Petroleum Technology Conference. – Society of Petroleum Engineers, 2017.
6. Pedersen M.E.H. Tuning & simplifying heuristical optimization: дис. – University of Southampton, 2010.
7. Мирзаджанзаде А.Х., Хасанов М.М., Бахтизин Р.Н. Моделирование процессов нефтегазодобычи. – Институт компьютерных исследований, 2005.
8. Krasnov F., Glavnov N., Sitnikov A. A Machine Learning Approach to Enhanced Oil Recovery Prediction // International Conference on Analysis of Images, Social Networks and Texts. – Springer, Cham, 2017. – С. 164-171.
9. Буторин А.В., Краснов Ф.В. Возможности использования результатов спектральной инверсии при интерпретации сейсмических данных // Геофизика. – 2017. – № 4. – С. 2-7.
10. Буторин А.В. Изучение спектральных характеристик волнового поля на примере модельных данных по результатам вейвлет-преобразования // Геофизика. – 2016. – № 4. – С. 61-67.

**Butorin Aleksandr Vladimirovich**

LLC "Gazpromneft NTC", Saint-Petersburg, Russia  
E-mail: Butorin.AV@gazpromneft-ntc.ru

**Krasnov Fedor Vladimirovich**

LLC "Gazpromneft NTC", Saint-Petersburg, Russia  
E-mail: Krasnov.FV@Gazprom-Neft.ru

**Mikheenkov Andrey Vital'evich**

Vereshchagin Institute for high pressure physics Russian academy of sciences, Moscow, Russia  
E-mail: mikheen@bk.ru

## **Reconstruction of medium's reflectivity coefficients based on seismic data through machine learning**

**Abstract.** Seismic survey is one of the few universal geophysical methods of obtaining information of the Earth subsurface. A reflected signal as a part of seismic data provides information of the properties of a medium through which it has passed. Reflectivity coefficients are determined by fluctuation of the medium's elastic properties and serve as a basis for interpretation of seismic data as well as for prediction of geological structures. The authors of this research have developed a new method of processing seismic data which enables to locate reflecting planes and compute values of reflectivity coefficients with high degree of precision. To resolve this problem the authors have used the Semi-supervised learning method. The machine learning method made it possible to develop a mathematical model, optimize its parameters for synthetic data in order to further use the model for unmarked-up seismic data. The main novelty is in developing a learning algorithm using signal convolution and reflectivity coefficients' regularization. The model developed by the authors demonstrated high precision for synthetic seismic data with high density of reflecting planes (102 planes per a second on trace). The resulting low level of errors allows significant improving of quantitative understanding of the subsurface structure based on seismic data and is a firm basis for building geological models.

**Keywords:** seismic data; semi-supervised learning; optimization task; regularization