

## Сейсмическое моделирование с применением NDM-net, обученной на оптимизированном наборе данных

*К. Г. Гадьльшин<sup>1</sup>, В. В. Лисица<sup>1\*</sup>, Д. М. Вишнеvский<sup>1</sup>, К. А. Гадьльшина<sup>1</sup>*

<sup>1</sup> Институт нефтегазовой геологии и геофизики им. А.А. Трофимука СО РАН,  
г. Новосибирск, Российская Федерация

\* e-mail: LisitsaVV@ipgg.sbras.ru

**Аннотация.** В работе представлен новый подход к конструированию оптимального обучающего набора данных для нейронной сети, применяемой для уменьшения численной дисперсии. Такой обучающий набор данных разработан для того, чтобы сохранить необходимое средне-квадратичное расстояние между любой сейсмограммой из входного набора данных и сейсмограммой из тренировочного набора. Для создания этого обучающего набора данных используется решение, вычисленное на грубой сетке. С применением предлагаемого подхода возможно в четыре раза уменьшить количество источников в обучающем наборе данных, если сравнивать с необходимым набором данных равномерно распределенных источников.

**Ключевые слова:** глубокое обучение, численная дисперсия, сейсмическое моделирование

### NDM-net for seismic modelling: data-based optimization of the training dataset

*K. G. Gadylyshin<sup>1</sup>, V. V. Lisitsa<sup>1\*</sup>, D. M. Vishnevskiy<sup>1</sup>, K. A. Gadylyshina<sup>1</sup>*

<sup>1</sup> Trofimuk Institute of Petroleum Geology and Geophysics SB RAS, Novosibirsk,  
Russian Federation

\* e-mail: LisitsaVV@ipgg.sbras.ru

**Abstract.** This paper presents a new method to construct the optimal training dataset for the numerical dispersion mitigation network. The dataset is designed to preserve the prescribed maximal NRMS distance between any seismogram from the entire dataset and those from the training dataset. So, the training dataset is constructed only using the precomputed coarse-mesh solution; thus, it does not require any additional simulations. Our numerical experiments illustrate that the suggested approach may reduce the number of common-shot gathers in the training set by the factor of four compared to the datasets using uniformly distributed shots. However, the areas with high model and seismograms variability are oversampled by the proposed approach without gaining extra accuracy, which requires further study.

**Keywords:** deep learning, numerical dispersion, seismic modeling

### *Введение*

В рамках сейсмического моделирования появляется возможность изучать характеристики распространения сейсмических волн в сложной среде. Так же сейсмическое моделирование может использоваться при тестировании и для подтверждения корректности сейсмической обработки и интерпретации данных. В то же время крайне высокие требования к объему вычислительных ресурсов снижают применимость методов сейсмического моделирования и являются барьером к их использо-

ванию в качестве стандартного инструмента. В самом деле, проведение полномасштабного моделирования сейсмических данных сеточными вычислительными методами, такими как конечно-разностный метод, метод конечных элементов, разрывный метод Галёркина, занимает тысячи ядрочасов. В этом случае точность численного решения определяется величиной шага вычислительной сетки или количеством степеней свободы: ячеек сетки, вершин ячеек сетки, тестовых функций и т. д. Чем больше учитывается степеней свободы, тем более точным будет численное решение. В итоге сейсмическое моделирование всегда проводится в условиях поиска компромисса между точностью и экономией вычислительных ресурсов. Существуют различные способы уменьшения количества степеней свободы без потери в точности решения. Например, распространенным способом решения этой задачи для миграции в обратном времени является применение конечно-разностных схем с пониженной дисперсией. Однако, это приводит к расширению шаблона схемы, что, в свою очередь, вызывает рост числа арифметических операций на одну степень свободы и снижение дыстродействия.

Другим направлением является применение постобработки данных для уменьшения численной дисперсии в целях повышения точности решения [1], для этого часто применяются инструменты машинного обучения [2, 3]. В [3] мы представили искусственную нейронную сеть NDM-net (numerical dispersion mitigation network), которая значительно уменьшает числовую дисперсию решения при том, что сеть обучается на небольшом количестве общих сейсмограмм, смоделированных на достаточно мелкой сетке. Однако мы строили обучающий набор данных эмпирически, что не является оптимальной стратегией решения поставленной задачи. В этом исследовании мы предлагаем метод построения обучающего набора данных на основе свойств смоделированных данных.

### ***Краткое описание искусственной нейронной сети NDM-net***

При проведении наземной съемки источники размещаются на регулярной сетке таким образом, что расстояние между источниками варьируется от 20 до 50 метров в пределах линии, а расстояние между линиями составляет от 100 до 300 метров. В таком случае система регистрации может считаться плотной, и мы предполагаем, что из-за малой изменчивости модели в латеральном направлении соответствующие соседним источникам сейсмограммы подобны. Мы можем ожидать, что репрезентативный набор данных будет получен при гораздо более разреженной сети приемников, которая в дальнейшем будет использоваться в качестве обучающего набора данных.

В итоге алгоритм работы искусственной нейронной сети NDM-net может быть представлен следующим образом:

1. Моделирование волнового поля для всех позиций источников с использованием метода конечных разностей на грубой сетке. На такой сетке будет достигаться относительно высокая скорость расчётов, но и уровень численной дисперсии ожидается большим.

2. Моделирование волнового поля для некоторого репрезентативного набора источников на достаточно мелкой сетке. Полученное решение должно быть достаточно точным, чтобы использоваться в качестве обучающих данных.

3. Обучение искусственной нейронной сети (NDM-net) для построения отображения решения, полученного на грубой сетке, в решение, рассчитанное на мелкой сетке.

4. Применение искусственной нейронной сети для сейсмограммы с полным набором источников для повышения точности решения.

Мы опускаем описание технических деталей NDM-net, с этими подробностями можно ознакомиться в работе [3]. Основной темой данного исследования является построение репрезентативного набора данных для обучения нейронной сети.

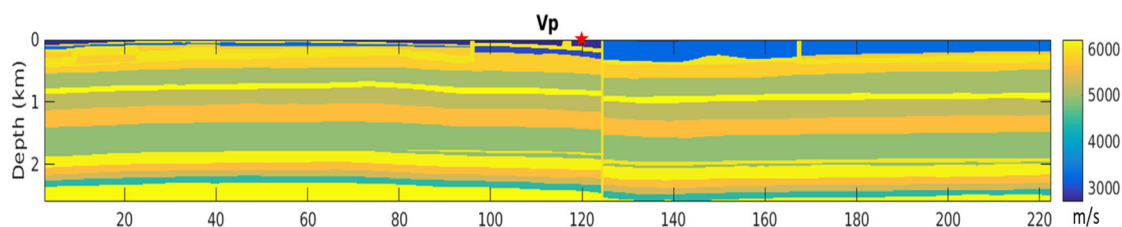


Рис. 1. Модель продольных скоростей Vanavar

### *Построение обучающего набора данных*

Проиллюстрируем метод построения обучающего набора данных на примере двумерной модели. Мы рассматриваем модель, соответствующую формации в Восточно-Сибирском регионе под названием Ванавар, представленную на рис. 1. Модель близка к горизонтально-слоистой и имеет два типа аномалий: это приповерхностные высокоскоростные интрузии (левая часть модели, т.е. 0–100 км) и вертикальные интрузии (первая располагается на уровне  $x=125$  км, вторая – на  $x=167$  км). Вторая интрузия присутствует только в приповерхностной части, она будет иметь определяющее значение для последующего анализа. Мы смоделировали волновые поля для 1901 источника, распределенных в длину на 100 метров. Сейсмограммы записывались с временным шагом 1 мс, время записи 5 секунд. Мы использовали импульс Рикера с центральной частотой 30 Гц. Волновые поля смоделированы с применением схемы четвертого порядка точности по пространству и второго порядка по времени [4]. Мы работали с двумя пространственными сетками с шагом по пространству 5 м (крупная сетка) и 2,5 м (мелкая сетка).

Мера правдоподобия между сейсмограммами определяется как нормированное среднеквадратичная ошибка (NRMS measure – normalized root mean square measure) [5]. Во-первых, мы рассмотрели парные NRMS всех наборов данных, соответствующих грубой и мелкой сеткам. В результате мы получили две матрицы расстояний для решений с грубой и мелкой сеткой. На графиках (рис. 2) для нескольких столбцов матрицы, соответствующих двадцати позициям источников, проиллюстрированы основные характеристики матриц расстояний. Можно заметить, что NRMS почти линейно возрастает вблизи фиксированного источника. После этого она достигает предельного значения в области, где источники находятся достаточно далеко друг от друга и практически не изменяются. Для рассматриваемой модели предельное значение NRMS между двумя случайными общими выборками составляет примерно от 115 до 150 %.

Очевидный способ построить обучающий набор данных — использовать каждые пятую, десятую или двадцатую общую выборку импульсов для равномерно распределенных источников. После создания набора обучающих данных можно измерить минимальную ошибку NRMS между данными с каждой сейсмограммы и соответствующими данными из обучающего набора. На рисунке 3 минимальные расстояния для трех простых наборов обучающих данных обозначены цифрами 5, 10 и 20. Как и ожидалось, чем плотнее набор обучающих данных, тем меньше расстояние от данных случайной сейсмограммы до данных обучающего набора. Однако результат менее оптимален, если набор обучающих данных становится чересчур плотным (использование каждой десятой части каждой пятой выборки почти неразлично). Кроме того, мы применили сети NDM-net, обученные на разных простых наборах данных, и измерили NRMS между улучшенными данными и точным решением (т. е. решением для модели с мелкой сеткой). Среднее значение NRMS, представленное в таблице 1 для всех сейсмограмм, подтверждает, что избыточная выборка действительно приводит к значительному уменьшению дисперсии, однако снижает скорость вчислений, поскольку большой набор обучающих данных моделируется на мелкой сетке.

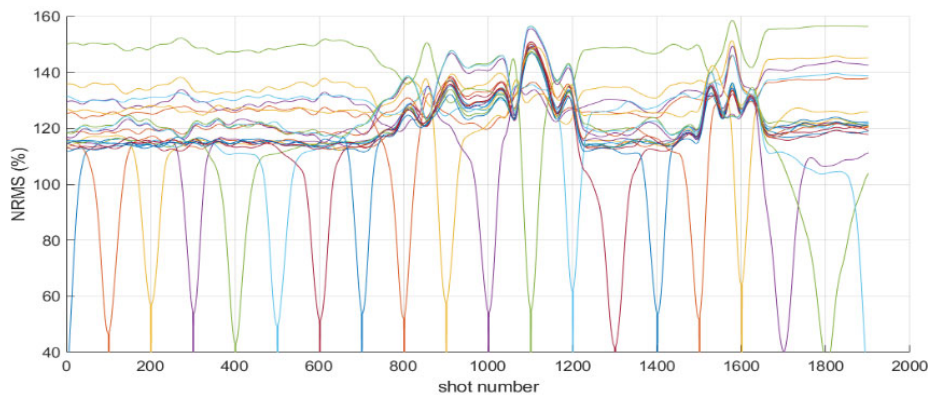


Рис. 2. Среднее значение NRMS между сейсмограммами в зависимости от расположения источника. Разными цветами выделены разные фиксированные значения положения источника

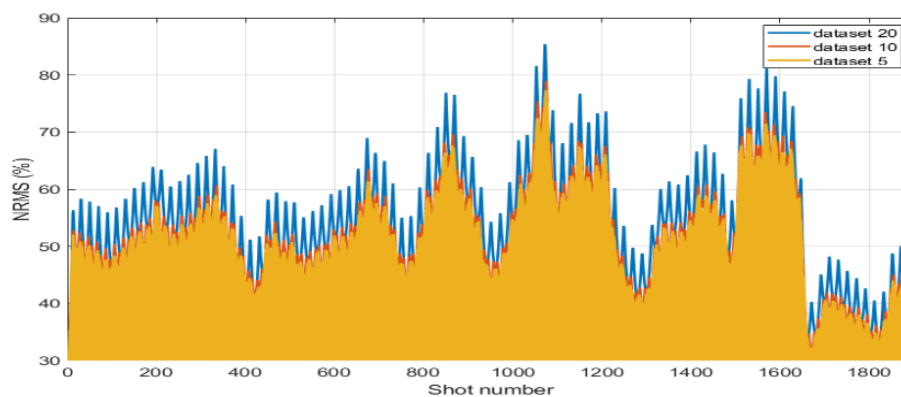


Рис. 3. Минимальное значение NRMS между сейсмограммами общей точки возбуждения и сейсмограммами обучающей выборки для трех различных наборов данных

Зависимость времени обучения, NRMS и ускорения в зависимости от выбора обучающего набора данных

Обуч. выборка	$t_{tr}$ (min)	NRMS	Ускорение
5%	32	44%	4.5
10%	40	29%	3.6
20%	43	23%	2.6

С целью уменьшения количества сейсмограмм в обучающей выборке предлагаем строить обучающую выборку таким образом, чтобы заранее ограничить максимальное расстояние между случайной сейсмограммой и выборкой. В частности, мы рассмотрели наборы данных с 60%, 70%, 80% и 90% от максимального NRMS между сейсмограммами. В результате такого подбора получены наборы данных с редко распределёнными источниками в тех областях, где данные слабо зависят от латерального направления, и с плотным расположением в областях, где модель и данные быстро меняются (рис. 4).

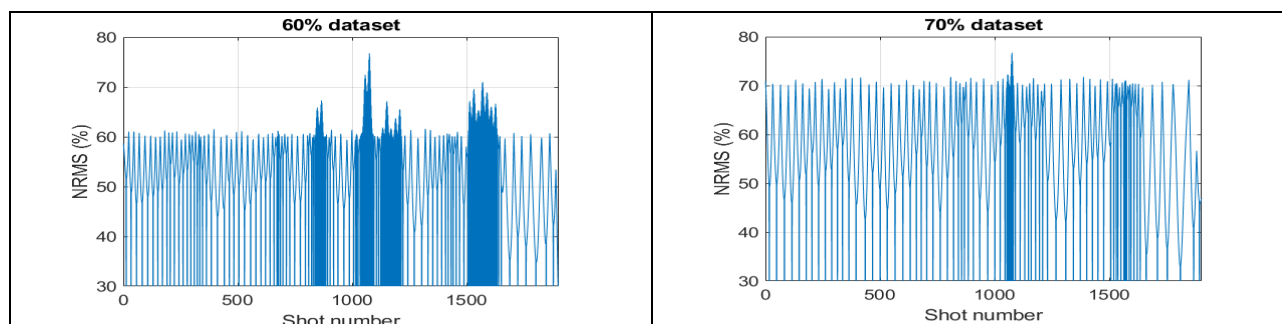


Рис. 4. Минимальное NRMS между каждым источником и обучающей выборкой для двух наборов данных (60% слева и 70% справа)

Набор данных с ограничением расстояния в 60% содержал 240 сейсмограмм из-за резкого расхождения в волновых полях вблизи вертикальных интрузий, где NRMS между соседними сейсмограммами превышал 60%. Набор данных для ограничения расстояния 70% содержал 80 сейсмограмм, набор данных для расстояния в 80% содержал 54 сейсмограммы, а набор данных для расстояния в 90% включал 42 сейсмограммы. Обратите внимание, что уменьшение количества сейсмограмм не имеет решающего значения в том случае, если предельная ошибка высока. Мы обучили NDM-net на построенных наборах данных, применили ее для уменьшения численной дисперсии и измерили NRMS ошибку между прогнозируемыми данными и решением с мелкой сеткой для каждой позиции источника. Результаты представлены на рисунке 5. NDM-net с адаптированным набором данных с ограниченным расстоянием в 60% дает наилучшие общие варианты, за исключением самой правой части модели, где однородный набор данных демонстрирует более высокую точность. Последний набор данных является самым плотным в этой области. Кроме того, большая плотность источников

в областях с высокой изменчивостью сейсмограммы (из-за вертикальных интрузий в модели) не улучшает точность предсказания NDM-net и может быть более разреженной. В то же время использование адаптированного 70-процентного набора данных при относительно небольшом количестве сейсмограмм (80 из 191 для однородного набора данных) позволяет добиться приемлемой точности подавления численной дисперсии.

### *Заключение*

В этой статье представлен новый метод построения адаптированного обучающего набора данных для нейронной сети, ориентированной на подавление численной дисперсии. Набор данных предназначен для сохранения предписанного максимального расстояния ошибки NRMS между любой из сейсмограмм из всего набора данных и сейсмограммой из обучающего набора данных. Таким образом, обучающий набор данных строится только с использованием предварительно вычисленного решения на грубой сетке; таким образом, он не требует проведения дополнительных расчетов. Наши численные эксперименты показывают, что предлагаемый подход может уменьшить количество сейсмограмм в обучающей выборке в четыре раза по сравнению с наборами данных, использующими равномерно распределенные сейсмограммы. Однако области с высокой изменчивостью модели и сейсмограммы подвергаются передискретизации предлагаемым подходом без получения дополнительной точности, что требует дальнейшего изучения.

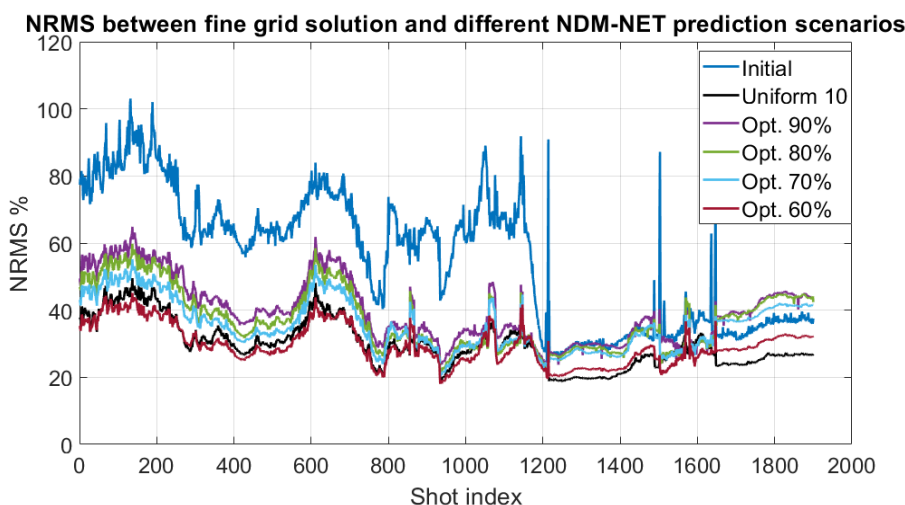


Рис. 5. Средний NRMS между решением на мелкой сетке и предсказанным решением после применения NDM-net для различных сценариев построения обучающей выборки

### *Благодарности*

Вадим Лисица разработал алгоритм построения оптимальной обучающей выборки, а Ксения Гадыльшина выполнила численные эксперименты по обучению NDM-net сети для различных сценариев при поддержке Российского научного фонда грант №22-11-00004. Дмитрий Вишневский выполнил моделирова-

ние сейсмических данных при поддержке Российского научного фонда грант №22-21-00738. Кирилл Гадильшин оптимизировал гиперпараметры нейронной сети NDM-net при поддержке Совета по грантам Президента Российской Федерации грант МК-3947.2021.1.5.

#### БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Koene E., and Robertsson J. Removing numerical dispersion artifacts from reverse time migration and full-waveform inversion // SEG Technical Program Expanded Abstracts. – 2017. – P. 4143–4147. <https://doi.org/10.1190/segam2017-17631589.1>
2. Kaur H., Fomel S., and Pham N. Overcoming numerical dispersion of finite-difference wave extrapolation using deep learning // SEG Technical Program Expanded Abstracts. –2019. –P. 2318–2322. <https://doi.org/10.1190/segam2019-3207486.1>
3. Gadylyshin K., Vishnevsky D., Gadylyshina K., and Lisitsa V. DNN-based approach to seismic modelling: mitigation the numerical dispersion // 82nd EAGE Annual Conference & Exhibition, Amsterdam, 2021 <https://doi.org/10.3997/2214-4609.202112693>
4. Levander A. R. Fourth-order finite-difference P-SV seismograms // Geophysics. – 1998. – 53 (11). – P. 1425–1436. <https://doi.org/10.1190/1.1442422>
5. Kragh E., and Christie P. Seismic repeatability, normalized RMS, and predictability // The Leading Edge. – 2002. – 21 (7). – P. 640–647. <https://doi.org/10.1190/1.1497316>

© К. Г. Гадильшин, В. В. Лисица, Д. М. Вишневский, К. А. Гадильшина, 2022