




Optimization of the Training Dataset for Numerical Dispersion Mitigation Neural Network

Kirill Gadylshin¹(✉), Vadim Lisitsa¹, Kseniia Gadylshina²,
and Dmitry Vishnevsky¹

¹ Institute of Petroleum Geology and Geophysics SB RAS,
3 Koptug ave., Novosibirsk 630090, Russia

{GadylshinKG,lisitsavv}@ipgg.sbras.ru

² Sobolev Institute of Mathematics SB RAS,
4 Koptug ave., Novosibirsk 630090, Russia

Abstract. We present an approach to construct the training dataset for the numerical dispersion mitigation network (NDM-net). The network is designed to suppress numerical error in the simulated seismic wavefield. The training dataset is the wavefield simulated using a fine grid, thus almost free from the numerical dispersion. Generation of the training dataset is the most computationally intense part of the algorithm, thus it is important to reduce the number of seismograms used in the training dataset to improve the efficiency of the NDM-net. In this work, we introduce the discrepancy between seismograms and construct the dataset, so that the discrepancy between the dataset and any seismogram is below the prescribed level.

Keywords: Deep learning · Seismic modelling · Numerical dispersion

1 Introduction

Seismic modeling is widely used to study elastic wave propagation in complex Earth models. In particular, numerical simulation allows understanding peculiarities of the wavefields in models with small-scale heterogeneities [7], anisotropic [17], viscoelastic [2], and poroelastic [13] media, in models with complex free-surface topology [10, 19]. However, seismic modeling is a computationally intense procedure that requires the use of high-performance computations. In particular, simulation of the wavefield corresponding to one source may take up to several

V.L. developed the algorithm of optimal dataset construction under the support of RSF grant no. 22-11-00004. D.V. performed seismic modeling using NKS-30T cluster of the Siberian Supercomputer Center under the support of RSF grant no. 22-21-00738. Kseniia Gadylshina performed numerical experiments on NDM-net training under the support of the RSF grant no. 19-77-20004. Kirill Gadylshin optimized the NDM-net hyperparameters under the support of the grant for young scientists MK-3947.2021.1.5.

thousand core hours, whereas thousands of shot positions should be simulated. A standard way to reduce the computational error, which is mainly appeared as the numerical dispersion due to the use of symmetric stencils for approximation, is to increase the grid step (reduce the number of grid points). However, it leads to a rapid increase in numerical dispersion. There are various ways to reduce numerical dispersion using a coarse mesh, including dispersion-suppression schemes [12], use of high order finite element, discontinuous Galerkin, and spectral element methods [1, 4, 9, 11, 15]. However, it does not necessarily lead to a reduction of the computational cost of the algorithm, because the number of floating-point operations per degree of freedom (per grid point) increases with the increase of the formal order of approximation.

The other approach to deal with the numerical dispersion is a pre- and post-processing of the emitted and recorded signals [6, 14]. However, the numerical dispersion in the recorded signal depends on the wave's ray path and can hardly be formalized. Thus, it may be treated by the Machine Learning methods. In particular, application of the neural networks to suppress numerical dispersion at the post-processing stage was suggested in [3, 5, 18]. In our previous research, we suggested the approach called the Numerical Dispersion Mitigation network (NDM-net) which is designed to suppress numerical dispersion in already simulated wavefields, recorded at the free surface. We suggest using the peculiarity of the seismic modeling problem; which is the simulation of the wavefield for a high number of right-hand sides (source positions), assuming that the seismograms corresponding to neighboring sources are similar. In this case, the true solution (solution computed on a very fine mesh) corresponding to a relatively small number of source positions can be used as the training dataset. In [3] we illustrated the applicability of the approach to realistic 2D problems. We used as few as 10% of sources equidistantly distributed. However, we have not tried to study the effect of the training dataset on the accuracy of the NDM-net results. In this study, we consider two possible strategies to construct the training dataset. First, we try the different number of equidistantly distributed sources. Second, we construct the training datasets preserving maximal differences between all seismograms and the training dataset.

The remainder of the paper has the following structure. In Sect. 2 we remind the basic concepts of seismic modeling and NDM-net. In Sect. 3 we provide the analysis of the seismograms and introduce the measure in the seismograms space. Different strategies to training dataset construction is presented in Sect. 4.

2 Preliminaries

2.1 Seismic Modelling

Consider a typical statement of seismic modeling problem, where the elastic wave equation is solved in a half-space for a series of right-hand sides. In a short form, the problem can be presented as:

$$L[\mathbf{u}] = \mathbf{f}(t)\delta(\mathbf{x} - \mathbf{x}_j^s), \quad (1)$$

where $\mathbf{f}(t)$ is the time-dependent right-hand side, either external forces or seismic moments, δ is the delta-function, \mathbf{x}_j^s is the location of j -th source. Operator L represents the linear differential operator corresponding to the elastic wave equation with appropriate initial and boundary conditions. The solution of the problem is considered at a free surface, which can be assumed flat for simplicity, for example, $x_3 = 0$. Thus, the solution of a single problem can be represented as

$$\mathbf{u}(\mathbf{x}_j^s, \mathbf{x}^r, t),$$

where \mathbf{x}^r is the receiver positions. Note, that the regular acquisition system follows the source position, which is $\mathbf{x}^r = \mathbf{x}^r(\mathbf{x}_j^s)$. It is convenient to consider an independent parameter, called offset $\mathbf{x}^o = \mathbf{x}_j^s - \mathbf{x}^r(\mathbf{x}_j^s)$. This parameter varies in the same limits for all source positions; thus, two seismograms can be directly compared as functions of (\mathbf{x}^o, t) .

If the wavefield is simulated using a numerical method, in this study, we focus on finite-differences with the fourth-order of approximation in space and second-order in time [8], it can be written as

$$L_h[\mathbf{u}_h] = \mathbf{f}_h(t)\delta(\mathbf{x} - \mathbf{x}_j^s), \quad (2)$$

where L_h is the finite-difference approximation of the original differential operator L , \mathbf{f}_h is the approximation of the right-hand side, and \mathbf{u}_h is the finite-difference solution corresponding to the grid with the step h . Due to the convergence of the finite-difference solution to that of the differential problem one gets the estimate:

$$\|\mathbf{u}(\mathbf{x}_j^s, \mathbf{x}^o, t) - \mathbf{u}_h(\mathbf{x}_j^s, \mathbf{x}^o, t)\| = \varepsilon_h \leq C_1 h^4 + C_2 \tau^2 \leq Ch^2, \quad (3)$$

we assume that $\tau \approx C_0 h$ due to the Courant stability criterion. Parameters C_0 , C_1 , C_2 , and C are constants independent of grid steps.

2.2 NDM-Net

The error estimate (3) means that the finer the grid step the lower the error ε_h ; which is

$$\varepsilon_{h_1} \leq \varepsilon_{h_2} \quad \text{if} \quad h_1 \leq h_2.$$

However, the reduction of the grid step leads to a significant increase in the computational resources demand and the computational intensity of the algorithm. We suggested recently [3] using machine learning to map coarse-grid solution to the fine-grid solution:

$$\mathcal{N}[\mathbf{u}_{h_2}(\mathbf{x}_j^s, \mathbf{x}^o, t)] = \tilde{\mathbf{u}}_{h_2}(\mathbf{x}_j^s, \mathbf{x}^o, t),$$

so that

$$\|\tilde{\mathbf{u}}_{h_2}(\mathbf{x}_j^s, \mathbf{x}^o, t) - \mathbf{u}_{h_1}(\mathbf{x}_j^s, \mathbf{x}^o, t)\| \leq \varepsilon_{21} \ll \varepsilon_{h_2}$$

for all source positions \mathbf{x}_j^s . If we manage to construct the map, which ensures that ε_{21} is small enough, we get

$$\begin{aligned} & \|\tilde{\mathbf{u}}_{h_2}(\mathbf{x}_j^s, \mathbf{x}^o, t) - \mathbf{u}(\mathbf{x}_j^s, \mathbf{x}^o, t)\| \\ \leq & \|\tilde{\mathbf{u}}_{h_2}(\mathbf{x}_j^s, \mathbf{x}^o, t) - \mathbf{u}_{h_1}(\mathbf{x}_j^s, \mathbf{x}^o, t)\| + \|\mathbf{u}_{h_1}(\mathbf{x}_j^s, \mathbf{x}^o, t) - \mathbf{u}(\mathbf{x}_j^s, \mathbf{x}^o, t)\| \\ & \leq \varepsilon_{21} + \varepsilon_{h_1} < \varepsilon_{h_2}. \end{aligned}$$

In particular, we use the special case of Convolutional Neural Network (CNN) - a U-Net [16], however other types of neural networks, for example, generative adversarial networks (GANs) have also been applied to suppress the numerical dispersion [5, 18]. The main problem in the NDM-net implementation is the training dataset construction. In our previous study [3] we suggested that due to low model variability in the horizontal direction, the seismograms corresponding to neighboring sources are similar. Thus, we may compute a small number of seismograms corresponding to a small number of sources using a fine enough grid to use them as the training dataset. However, we provided no quantitative analysis of the assumption and effect of wavefield similarity on the NDM-net accuracy.

3 Analysis of Seismogramms

In this section, we provide the study of the seismogram's similarities in dependence on the distance between the sources. Consider a standards 2D acquisition system with sources placed in points \mathbf{x}_j^s , $j = 1, \dots, J_s$. In this case, the entire set of seismograms can be represented as The entire dataset is a union of the solutions corresponding to all source positions

$$U = \bigcup_{j=1, \dots, J_s} \mathbf{u}(\mathbf{x}_j^s, \mathbf{x}^o, t) = \bigcup_{j=1, \dots, J_s} \mathbf{u}(\mathbf{x}_j^s).$$

we further omit the variables t and \mathbf{x}^o assuming that they get the same values for all seismograms. To compare the seismograms and measure their similarity we suggest using the repeatable measure - normalized root mean square (NRMS). The NRMS between two traces a_t and b_t at point t_0 using a window size dt is the RMS of the difference divided by the average RMS of the inputs, and expressed as a percentage:

$$NRMS(a_t, b_t, t_0) = \frac{200 \times RMS(a_t - b_t)}{RMS(a_t) + RMS(b_t)}$$

where the RMS is defined as:

$$RMS(x_t) = \sqrt{\frac{\sum_{t_0-dt}^{t_0+dt} x_t^2}{N}}$$

and N is the number of samples in the interval $[t_0 - dt, t_0 + dt]$. We introduce the distance $d(\mathbf{u}(\mathbf{x}_j^s), \mathbf{u}(\mathbf{x}_k^s))$ as an average NRMS between $\mathbf{u}(\mathbf{x}_j^s)$ and $\mathbf{u}(\mathbf{x}_k^s)$ seismograms.

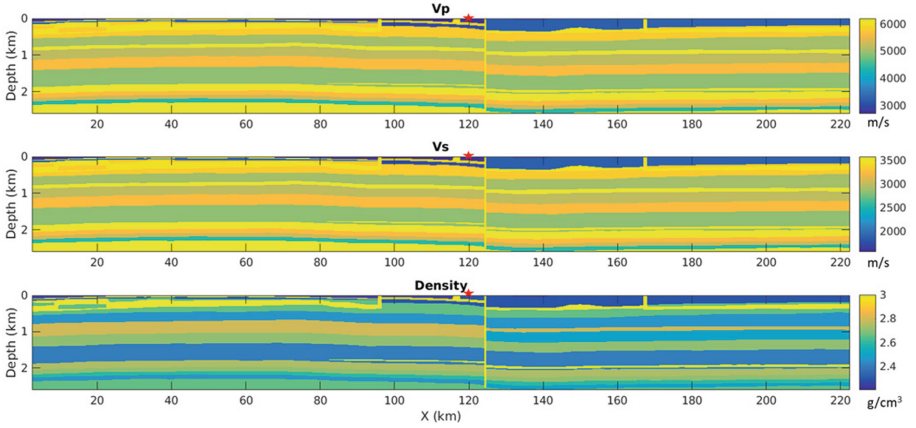


Fig. 1. Vanavar model

We constructed the distance matrix for the entire dataset, computed for the Vanavar model. The model is shown in Fig. 1. The size of the model was 220 km by 2.6 km. The acquisition included 1901 sources with a distance of 100 m. Similarly, we recorded the wavefield by 512 receivers for each shot with maximal source-receiver offsets of 6.4 km. The distance between the receivers was 25 m. The source wavelet was a Ricker pulse with a central frequency 30 Hz. Perfectly matched layers, including the top layer, were used at all boundaries.

We simulated wavefields using grids with steps 2.5 and 1.25 m. After that, for each set of seismograms, we computed the distance matrices, the one corresponding to the simulations with the step 2.5 m is presented in Fig. 2. The values of the distances are low within a narrow band near the main diagonal, which means that the distance between the seismograms is low if the sources are close enough, but it grows rapidly until reaching the value of about 100% of NRMS. To illustrate the relation between the NRMS of two seismograms with respect to the distance between the sources, we provide plots of several columns of the distance matrix in Fig. 3. Each line in this plot represents the NRMS-distance between the given seismogram and all the others. The distance is equal to zero if the seismogram is compared with itself. If the seismogram is compared with those corresponding to the nearby sources, the distance grows almost linearly, from some starting value to a limiting value. After that, the NRMS-distance is almost independent of the distance between the source position. Thus, for each source position \mathbf{x}_j^s exist two numbers k_j^+ and k_j^- and value e_j , so that for all $k < k_j^-$ and $k > k_j^+$ the following statement holds $d(\mathbf{u}(\mathbf{x}_j^s), \mathbf{u}(\mathbf{x}_k^s)) \approx e_j$. However, these values are individual for each source position. To analyze the boundaries k_j^\pm and error value e_j , we study the averaged values of the distance. For each seismogram, we compute the symmetric distance as:

$$d_j(\Delta j) = \frac{1}{2}(d(\mathbf{u}(\mathbf{x}_j^s), \mathbf{u}(\mathbf{x}_{j+\Delta j}^s)) + d(\mathbf{u}(\mathbf{x}_j^s), \mathbf{u}(\mathbf{x}_{j-\Delta j}^s)))$$

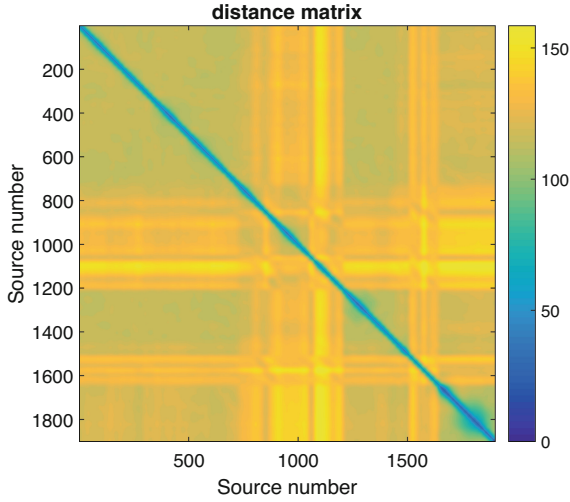


Fig. 2. Vanavar model. The distance matrix

If $j + \Delta j > J_s$ or $j < \Delta j < 1$ we assume that the corresponding distance is equal to symmetric one. After that we compute the mean and standard deviation of the with respect to j obtaining the functions of Δj :

$$M_d(\Delta j) = \frac{1}{J_s} \sum_{j=1}^{J_s} d_j(\Delta j), \tag{4}$$

$$\Sigma_d(\Delta j) = \frac{1}{J_s} \sum_{j=1}^J s(d_j(\Delta j) - M_d(\Delta j))^2. \tag{5}$$

The plots of $M_d(\Delta j)$, $M_d(\Delta j) \pm \Sigma_d(\Delta j)$, and $M_d(\Delta j) \pm 3\Sigma_d(\Delta j)$ are presented in Fig. 4. It illustrates that the discrepancy increases if $j < 30$, after that the error stabilizes in average at the value of approximately 120%. The standard deviation starts from 10% for nearby sources to 20% for long-distance sources. It follows from the plot, that if we use equidistantly distributed sources to construct the training dataset we can not use fewer than each thirties, otherwise, we will lose the representativity of the dataset. This analysis allows restricting the lowest number of sources that are reasonable to use if equidistantly distributed (in this particular case it should be more than 3%). On the other hand, it allows us to choose the seismograms to keep the prescribed discrepancy level within the training dataset and this level. In particular, it is worth considering discrepancies between 60 and 90 %.

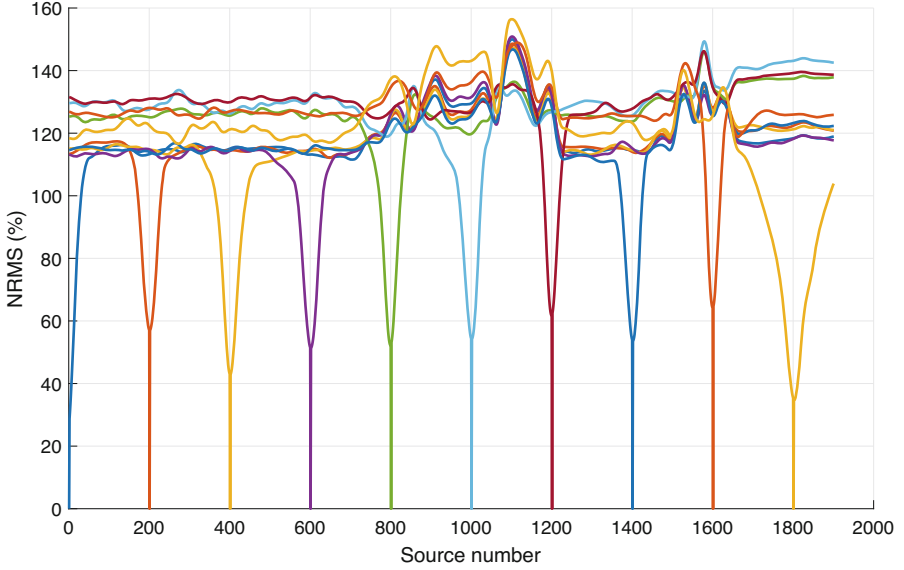


Fig. 3. Vanavar model. The distances for 19 seismogramms.

4 Training Dataset Construction Algorithms

Before describing particular algorithms of dataset construction let us introduce a measure, that characterizes the representativity of a dataset. Assume we have chosen a dataset

$$D_t = \bigcup_{j \in J_t} \mathbf{u}(\mathbf{x}_j^s),$$

where J_t is a set of training dataset sources indices, so that $J_t \subset \{1, \dots, J_s\}$. Thus the training dataset is also a subset of the entire dataset $D_t \subset U$. Let us define the distance from a single seismogram to the dataset as

$$b(\mathbf{x}_k^s, D_t) = \min_{j \in J_t} d(\mathbf{u}(\mathbf{x}_j^s), \mathbf{u}(\mathbf{x}_k^s)).$$

This function indicates the distance from a given seismogram to the closest one from the training dataset. It is clear that for $k \in J_t$ the distance will be zero. After that, the distance between the datasets can be introduced as

$$B_{\text{inf}} = \max_{k \in \{1, \dots, J_s\}} b(\mathbf{x}_k^s, D_t) = \max_{k \in \{1, \dots, J_s\}} \min_{j \in J_t} d(\mathbf{u}(\mathbf{x}_j^s), \mathbf{u}(\mathbf{x}_k^s)).$$

In our further considerations we will use both function $b(\mathbf{x}_k^s, D_t)$ and the distance between the datasets B_{inf} .

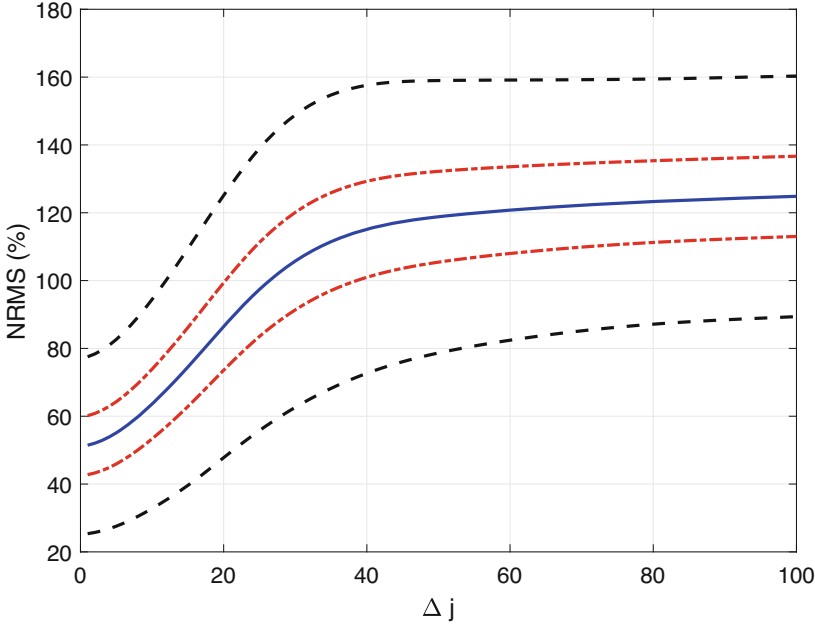


Fig. 4. Vanavar model. Mean NRMS-distance with respect to distance between the sources (solid line), dash-dotted lines correspond to $M_d \pm \Sigma_d$, and dashed lines correspond to $M_d \pm 3\Sigma_d$.

4.1 Equidistantly Distributed Dataset

The first and the simplest algorithm to construct the training dataset is to take the seismograms corresponding to the equidistantly distributed sources. In particular, we considered the datasets composed of 5, 10, and 20% of the entire seismograms denoting them as $D_{5\%}$, $D_{10\%}$, and $D_{20\%}$, respectively. We constructed the functions $b(\mathbf{x}_k^s, D_t)$ for the three datasets, as presented in Fig. 5. It is clear, that if a source position belongs to the training dataset the distance $b(\mathbf{x}_k^s, D_t)$ is equal to zero, so we excluded these points from the plots. Note, that functions $b(\mathbf{x}_k^s, D_{10\%})$ and $b(\mathbf{x}_k^s, D_{20\%})$ are almost indistinguishable, whereas variation of $b(\mathbf{x}_k^s, D_{5\%})$ is much stronger. This means that dataset $D_{5\%}$ may be under-representative providing poor data for NDM-net. On the contrary, increasing the number of sources in the dataset above 10% does not provide new valuable information.

We used these three datasets to train the NDM-net to map solution computed on a grid with the step of 2.5 m to that simulated using grid with the step of 1.25 m. Consequently, we estimate the accuracy of the NDM-net by introducing the measure

$$q(\mathbf{x}_j^s, D_t) = d(\mathbf{u}_{h_1}(\mathbf{x}_j^s), \tilde{\mathbf{u}}_{h_2}(\mathbf{x}_j^s)) = d(\mathbf{u}_{h_1}(\mathbf{x}_j^s), \mathcal{N}(\mathbf{u}_{h_2}(\mathbf{x}_j^s))), \quad (6)$$

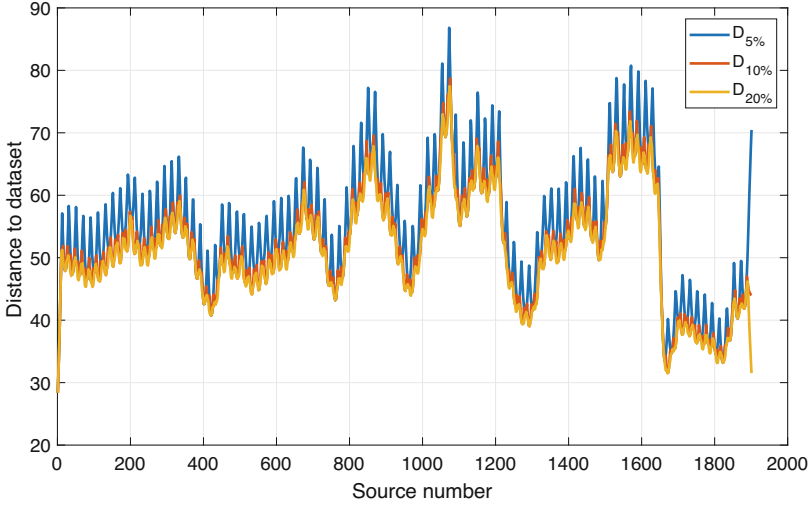


Fig. 5. Distances between the seismograms and the training datasets $b(\mathbf{x}_k^s)$ for different dataset of equidistantly distributed source $D_{5\%}$, $D_{10\%}$, and $D_{20\%}$.

where $h_1 < h_2$. That is a source-by-source seismograms comparison (in this study we simulated the wavefield using the fine mesh for all source positions to be able to validate the NDM-net action). We computed the mean value of

$$M_q = \frac{1}{J_s} \sum_{j=1}^{J_s} q(\mathbf{x}_j^s, D_t)$$

overall source positions. The results mean NRMS between the fine grid solution and NDM-net action for three considered training datasets are presented in Table 1. Note, that the error is relatively high for the case of $D_{5\%}$, however, the cases of $D_{10\%}$ and $D_{20\%}$ provide approximately the same accuracy of the NDM-net prediction. Thus, the dataset $D_{10\%}$ can be considered as an optimal one among the datasets of equidistantly distributed sources.

Table 1. Datasets of equidistantly distributed sources.

Dataset	Number of sources	Average NRMS
$D_{5\%}$	283	44.28%
$D_{10\%}$	191	31.91%
$D_{20\%}$	86	29.41%

4.2 Distance-Preserving Datasets

We assume that the entire dataset computed on a coarse mesh is available. Thus, we can compute the distances between all seismograms and chose the source positions to compute the training dataset. We suggest constructing the training dataset by solving the max-min problem:

$$\max_{k \in \{1, \dots, J_s\}} b(\mathbf{x}_k^s, D_t) = \max_{k \in \{1, \dots, J_s\}} \min_{j \in J_t} d(\mathbf{u}(\mathbf{x}_j^s), \mathbf{u}(\mathbf{x}_k^s)) \leq Q,$$

where Q is the desired error level. We considered several values of Q starting from 60% to 100% (according to the mean NRMS distances presented in Fig. 11). We considered five datasets $D_{60\%}^{NRMS}$, $D_{70\%}^{NRMS}$, ..., $D_{100\%}^{NRMS}$, so that $D_{m\%}^{NRMS}$ correspond to the case of

$$\max_{k \in \{1, \dots, J_s\}} b(\mathbf{x}_k^s, D_t) \leq m,$$

In Figs. 8, 9 and 10 we provide the functions $b(\mathbf{x}_k^s, D_{m\%}^{NRMS})$. We kept the sources that belong to the training dataset to visualize the number of sources in the dataset. In particular, the distance is equal to zero if the source belongs to the dataset. For example, the dataset $D_{60\%}^{NRMS}$ contains all sources with numbers from 1500 to 1650. This means that even for the two neighboring sources in this range the NRMS between the seismograms exceeds 60%. Increasing the level of the acceptable NRMS one may reduce the number of sources in the training dataset accelerating the NDM-net. However, it may lead to significant accuracy degradation.

Next, we consider the one-to-one NRMS between the coarse- and fine-mesh solutions $q(\mathbf{x}_j^s, D_t)$ as defined by formula (6) for the considered datasets in Fig. 11. We also considered the averaged values of $q(\mathbf{x}_j^s, D_t)$ over the source position. The results are provided in Table 2. According to the plots in Fig. 11, all datasets provide similar accuracy in the leftmost part of the model (source numbers up to 800), where the model was relatively simple but original NRMS between fine- and coarse-mesh solutions was about 70%. The main difference between the datasets is associated with source numbers 800 to 1500, where the results of NDM-net applications are different for different datasets. For the source numbers 1500 to 1650, where $D_{60\%}^{NRMS}$ includes all the sources, its accuracy is higher than that of any other dataset. However, in the simplest part of the model (source numbers 1650–1900), all adaptive datasets include very sparse sources distribution,

Table 2. Distance-preserving datasets

Dataset	Number of sources	Average NRMS
$D_{60\%}^{NRMS}$	414	30.28%
$D_{70\%}^{NRMS}$	109	34.69%
$D_{80\%}^{NRMS}$	56	35.11%
$D_{90\%}^{NRMS}$	43	35.68%
$D_{100\%}^{NRMS}$	34	36.26%
$D_{10\%}$	191	31.91%

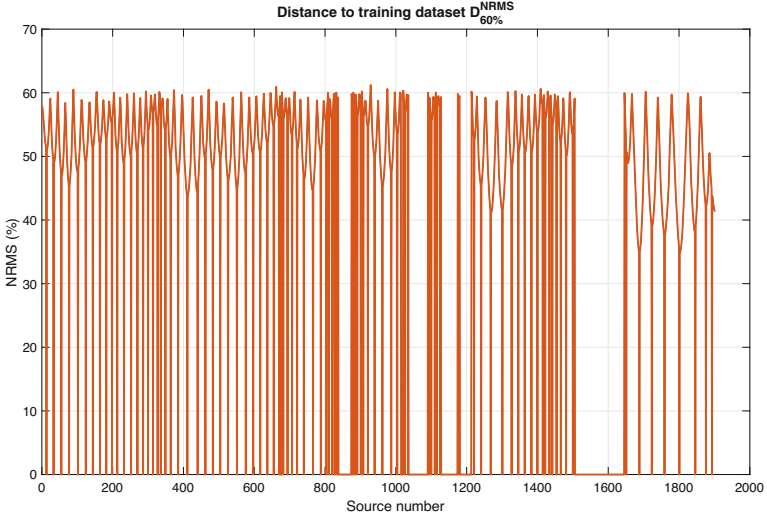


Fig. 6. Distances between the seismograms and the training dataset $b(\mathbf{x}_k^s, D_{60\%}^{NRMS})$.

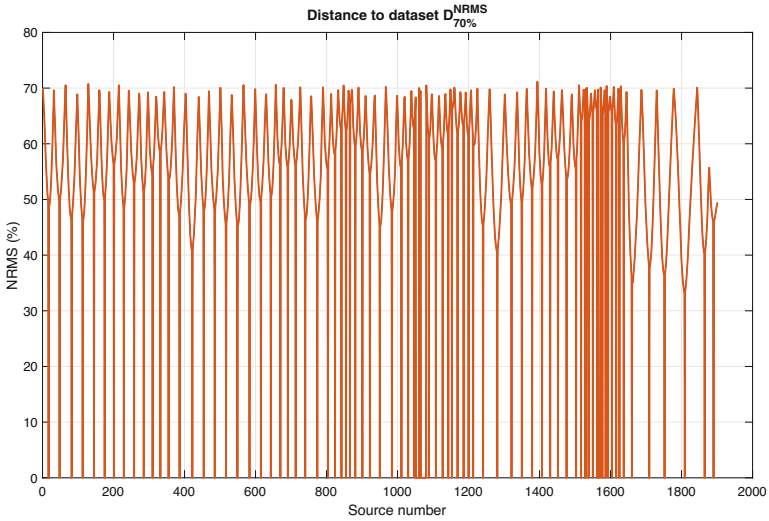


Fig. 7. Distances between the seismograms and the training dataset $b(\mathbf{x}_k^s, D_{70\%}^{NRMS})$.

which leads to an increase of the NRMS after the NDM-net application. On average, dataset $D_{60\%}^{NRMS}$ provides the highest accuracy of the NDM-net. However, the number of sources in $D_{60\%}^{NRMS}$ is 414 which is twice as many as in $D_{10\%}$. Dataset $D_{70\%}^{NRMS}$ includes half of the sources of $D_{10\%}$ but it caused a significant increase of the NRMS. According to the presented results, it is reasonable to construct the datasets with adaptive NRMS levels (Figs. 6 and 7).

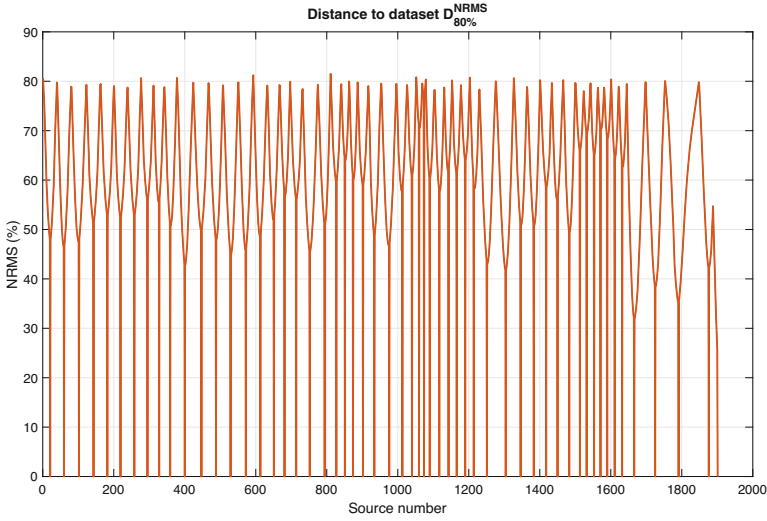


Fig. 8. Distances between the seismograms and the training dataset $b(\mathbf{x}_k^s, D_{80\%}^{NRMS})$.

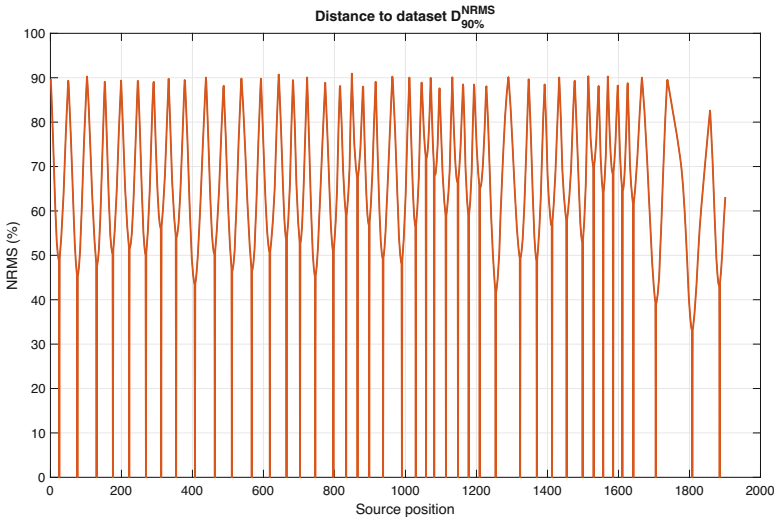


Fig. 9. Distances between the seismograms and the training dataset $b(\mathbf{x}_k^s, D_{90\%}^{NRMS})$.

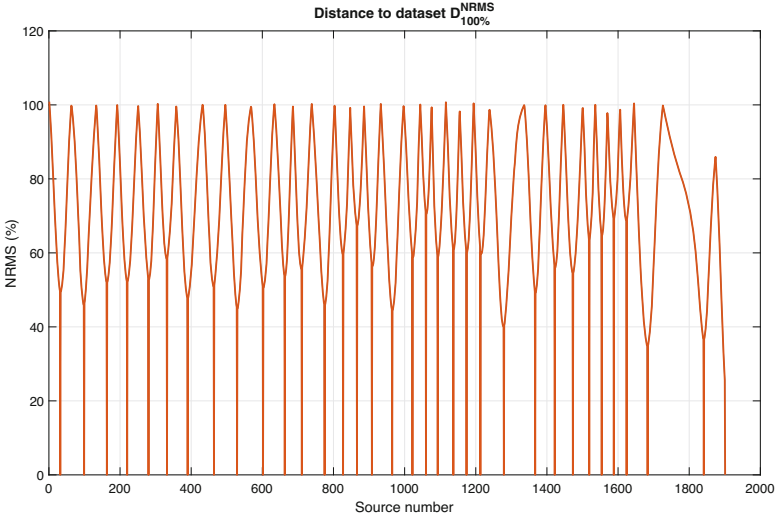


Fig. 10. Distances between the seismograms and the training dataset $b(\mathbf{x}_k^s, D_{100\%}^{NRMS})$.

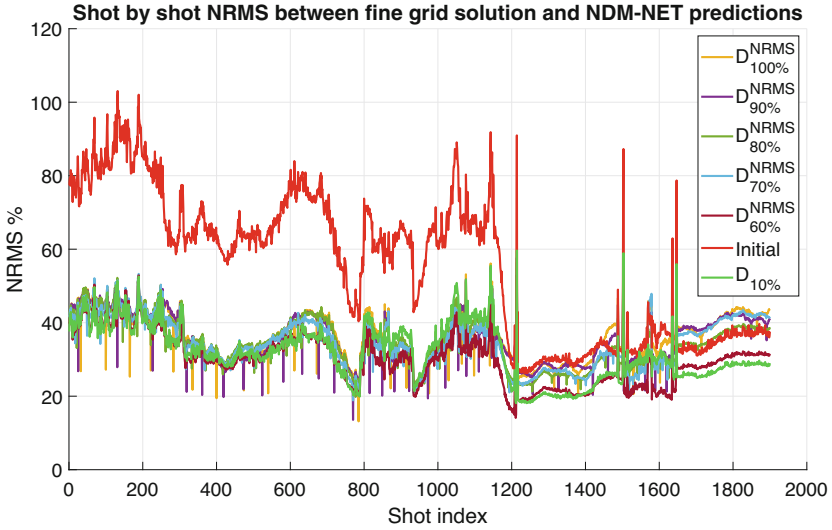


Fig. 11. Shot-by-shot distances between the fine-grid solution and NDM-net corrected solutions for different training datasets.

5 Conclusions

In this study, we consider two possible ways to construct the training datasets for the Numerical Dispersion Mitigation network or NDM-net. The network was designed to suppress numerical error in the seismic modeling results. It is con-

structed to map a noisy solution computed using a coarse mesh to that computed by a fine mesh. The training dataset is composed of the wavefields corresponding to a small number of sources from the considered acquisition system. Thus, the smaller the number of seismograms in the training dataset the more efficient the approach is. We considered two ways to construct the training datasets. The first one is based on the equidistantly distributed sources. In this case, the optimal set of sources to generate the training dataset is 10% of the entire number of the sources. Reduction of the number of sources leads to rapid error increase. The use of a denser system of sources requires higher computational time to generate the training dataset without significant accuracy improvement. The second way to construct the training dataset is the requirement, that the NRMS-based distance from the entire dataset and the training dataset does not exceed a prescribed level. In this case, the error can be reduced, however, an extremely dense source system may be needed, which may significantly increase the number of source positions in the training dataset. Moreover, due to the variation of the NRMS distance between the seismograms depending on the source position, it seems reasonable to consider an adaptive choice of the NRMS-based distance level to construct the training dataset.

References

1. Ainsworth, M.: Dispersive and dissipative behaviour of high order discontinuous Galerkin finite element methods. *J. Comput. Phys.* **198**(1), 106–130 (2004)
2. Blanch, J., Robertsson, J., Symes, W.: Modeling of a constant Q: methodology and algorithm for an efficient and optimally inexpensive viscoelastic technique. *Geophysiscs* **60**(1), 176–184 (1995)
3. Gadylshin, K., Lisitsa, V., Gadylshina, K., Vishnevsky, D., Novikov, M.: Machine learning-based numerical dispersion mitigation in seismic modelling. In: Gervasi, O., et al. (eds.) ICCSA 2021. LNCS, vol. 12949, pp. 34–47. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-86653-2_3
4. Kaser, M., Dumbser, M., Puente, J.D.I., Igel, H.: An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes III. Viscoelastic attenuation. *Geophys. J. Int.* **168**(1), 224–242 (2007). <https://doi.org/10.1111/j.1365-246X.2006.03193.x>
5. Kaur, H., Fomel, S., Pham, N.: Overcoming numerical dispersion of finite-difference wave extrapolation using deep learning. In: SEG Technical Program Expanded Abstracts, pp. 2318–2322 (2019). <https://doi.org/10.1190/segam2019-3207486.1>
6. Koene, E.F.M., Robertsson, J.O.A., Brogini, F., Andersson, F.: Eliminating time dispersion from seismic wave modeling. *Geophys. J. Int.* **213**(1), 169–180 (2017)
7. Kostin, V., Lisitsa, V., Reshetova, G., Tcheverda, V.: Local time-space mesh refinement for simulation of elastic wave propagation in multi-scale media. *J. Comput. Phys.* **281**, 669–689 (2015)
8. Levander, A.R.: Fourth-order finite-difference P-SV seismograms. *Geophysics* **53**(11), 1425–1436 (1988)
9. Lisitsa, V.: Dispersion analysis of discontinuous Galerkin method on triangular mesh for elastic wave equation. *Appl. Math. Model.* **40**, 5077–5095 (2016). <https://doi.org/10.1016/j.apm.2015.12.039>

10. Lisitsa, V., Kolyukhin, D., Tcheverda, V.: Statistical analysis of free-surface variability's impact on seismic wavefield. *Soil Dyn. Earthq. Eng.* **116**, 86–95 (2019)
11. Lisitsa, V., Tcheverda, V., Botter, C.: Combination of the discontinuous Galerkin method with finite differences for simulation of seismic wave propagation. *J. Comput. Phys.* **311**, 142–157 (2016)
12. Liu, Y.: Optimal staggered-grid finite-difference schemes based on least-squares for wave equation modelling. *Geophys. J. Int.* **197**(2), 1033–1047 (2014)
13. Masson, Y.J., Pride, S.R.: Finite-difference modeling of Biot's poroelastic equations across all frequencies. *Geophysics* **75**(2), N33–N41 (2010)
14. Mittet, R.: Second-order time integration of the wave equation with dispersion correction procedures. *Geophysics* **84**(4), T221–T235 (2019)
15. Pleshkevich, A., Vishnevskiy, D., Lisitsa, V.: Sixth-order accurate pseudo-spectral method for solving one-way wave equation. *Appl. Math. Comput.* **359**, 34–51 (2019)
16. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015*. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
17. Saenger, E.H., Gold, N., Shapiro, S.A.: Modeling the propagation of the elastic waves using a modified finite-difference grid. *Wave Motion* **31**, 77–92 (2000)
18. Siahkoobi, A., Louboutin, M., Herrmann, F.J.: The importance of transfer learning in seismic modeling and imaging. *Geophysics* **84**, A47–A52 (2019). <https://doi.org/10.1190/geo2019-0056.1>
19. Tarrass, I., Giraud, L., Thore, P.: New curvilinear scheme for elastic wave propagation in presence of curved topography. *Geophys. Prospect.* **59**(5), 889–906 (2011). <https://doi.org/10.1111/j.1365-2478.2011.00972.x>