

Adaptive Data-Based Optimization of the Training Dataset for the NDM-net

Kirill Gadylshin[1], Vadim Lisitsa[2(✉)], Kseniia Gadylshina[2], and Dmitry Vishnevsky[2]

[1] LLC RN-BashNIPIneft, 3/1 Behtereva st., Ufa 450103, Russia
[2] Institute of Petroleum Geology and Geophysics SB RAS, 3 Koptug ave., Novosibirsk 630090, Russia
lisitsavv@ipgg.sbras.ru

Abstract. This paper presents a new approach to construct the training dataset for the Numerical Dispersion Mitigation network (NDM-net). The network was designed to suppress numerical error in seismic modeling results. In this approach, a small number of seismograms generated using coarse and fine girds are used to train the network, mapping the inaccurate coarse-grid data to the high-quality fine-grid data. After that, the network is applied to the entire set of seismograms, precomputed using the coarse grid to reduce the numerical error. The most time-consuming part of the suggested approach is the training dataset generation. Thus, we need to minimize the number of seismograms in the training dataset without the loss of the training quality. We suggest constructing the training dataset preserving the Hausdorff distance between the training dataset and the entire dataset. However, the level of the limiting distance may vary depending on the seismogeological model used for simulation. We illustrate that the adaptive strategy is preferred over the fixed Hausdorff distance limiting level because it allows reducing the training dataset without loss of accuracy.

Keywords: Deep learning · seismic modelling · numerical dispersion

1 Introduction

The Numerical Dispersion Mitigation neural network (NDM-net) was suggested recently [3] to suppress numerical error in seismic modeling efficiently. The idea of the approach is to use the deep-learning to post-process the numerically generated seismic wavefields and reduce the numerical dispersion caused by the

Vadim Lisitsa developed the algorithm of optimal dataset construction and Kseniia Gadylshina performed numerical experiments on NDM-net training under the support of RSF grant no. 22-11-00004. Dmitry Vishnevsky performed seismic modeling using NKS-30T cluster of the Siberian Supercomputer Center under the support of basic research project FWZZ-2022-0022. Kirill Gadylshin optimized the NDM-net hyperparameters.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
O. Gervasi et al. (Eds.): ICCSA 2023 Workshops, LNCS 14106, pp. 76–90, 2023.
https://doi.org/10.1007/978-3-031-37111-0_6

wavefield under resolution. Seismic modeling requires the numerical solution of acoustic, elastic, anelastic, or poroelastic wave equation [22] with different types of extra complexities, such as free surface topography or rough sea level [10,11,18,20], anisotropy [12,17], fluid intrusions, and karsts. The numerical experiments' design typically compromises the numerical error due to the coarse grid and the extreme time and computational resources consumption due to fine grid usage. In particular, if the finite difference schemes on symmetric stencils are applied [7,21], the numerical error manifests itself as a numerical dispersion, which may destroy the signal. Typically the grid is constructed so that the time delay due to the dispersion in the recorded signal does not exceed one-quarter of the wave period. This condition should be satisfied for the signal traveling 100 to 200 wavelengths. So, the condition is strict, and for typical set-ups may lead to a grid size of as small as one meter. Taking into account the size of the standard domain of interest of 10 by 10 by 5 km, one would finish up with the estimate of $5 \cdot 10^{11}$ grid points to discretize the model, or $6 \cdot 10^{12}$ degrees of freedom to simulate the wavefield in isotropic elastic media. Indeed, modern supercomputers are capable of such simulations. However, this estimate is valid for the wavefield modeling corresponding to a single source, whereas typical acquisition systems contain up to 10^5 shots. Simulation of the entire dataset would require an unacceptably high amount of computational resources. Indeed, there are alternative approaches to seismic modeling, such as finite elements, spectral elements, discontinuous Galerkin [4,11], pseudo-spectral method [15], and others. However, they allow using coarser mesh but with a higher number of degrees of freedom per element, resulting in similar overall estimates of the degrees of freedom and computational intensity to approximate the entire problem [1,9].

There is also a set of approaches based on the post-processing of the simulated data to reduce the numerical error. It includes direct subtraction of the dispersion due to coarse spatial discretization [6,13], which is not capable of reducing dispersion due to discretization of time direction. The other direction of the research is the use of machine learning to resolve this problem, as suggested in [3,5,19]. In particular, in [3] the Numerical Dispersion Mitigation network or NDM-net was suggested. The approach is essentially based on the necessity to solve numerous similar problems with different right-hand sides (different source positions). As a result, the training dataset can be constructed as an accurate enough solution for a small number of problems, and then the net is applied to fix the entire dataset. However, the generation of the training dataset assumes simulation of elastic wavefield using a fine-enough grid which is the most time and resource-consuming part of the method. Thus, it is worth minimizing the training dataset to improve the algorithm's performance. Recently, we suggested an approach to construct the training dataset preserving the particular value of the Hausdorff distance between the training dataset and the entire dataset [2]. This approach to dataset construction lets us reduce the number of the seismograms in the training dataset by a factor of three compared to equidistantly distributed sources. However, this approach still has a drawback. If the model has strong lateral inhomogeneities, such as dykes or salt intrusions, the average

distance between the seismograms may vary. Thus, the use of a low Hausdorff distance limit leads to an unreasonable dense set of sources involved in the training dataset in some areas. On the contrary, adjusting the limit to the worst case causes an underrepresentation of the data from relatively smooth parts of the model. To overcome this drawback, we suggest the adaptive training dataset construction, where the limiting distance between the training dataset and the entire dataset is adjusted to the model.

The remainder of the paper has the following structure. We describe the main concepts of seismic modeling and the NDM-net in Sect. 2. Different strategies for the training dataset construction are presented in Sect. 3. Numerical experiments illustrate the advantages of the suggested approach in Sect. 4.

2 Preliminaries

Seismic modeling aims to solve the elastic wave equation with multiple right-hand sides, providing the solution at the given surface (usually the free surface). The most commonly used approach to perform seismic modeling nowadays is the finite differences because they combine high computational efficiency with simplicity of implementation and complex models treatment [8,22,23]. The standard staggered grid schemes are commonly used to simulate wavefields in isotropic elastic media [7,21], which has the following representation in 2D:

$$\begin{gathered}
\hat{\rho}_{i+1/2,j} D_t[u_x]^{n+1/2}_{i+1/2,j} = D_x[\sigma_{xx}]^{n+1/2}_{i+1/2,j} + D_z[\sigma_{xz}]^{n+1/2}_{i+1/2,j}, \\
\hat{\rho}_{i,j+1/2} D_t[u_z]^{n+1/2}_{i,j+1/2} = D_x[\sigma_{xz}]^{n+1/2}_{i,j+1/2} + D_z[\sigma_{zz}]^{n+1/2}_{i,j+1/2}, \\
D_t[\sigma_{xx}]^{n}_{i,j} = (\lambda + 2\mu)_{i,j} D_x[u_x]^{n}_{i,j} + \lambda_{i,j} D_z[u_z]^{n}_{i,j} + (f_{xx})^{n}_{i,j}, \\
D_t[\sigma_{zz}]^{n}_{i,j} = \lambda_{i,j} D_x[u_x]^{n}_{i,j} + (\lambda + 2\mu)_{i,j} D_z[u_z]^{n}_{i,j} + (f_{zz})^{n}_{i,j}, \\
D_t[\sigma_{xz}]^{n}_{i+1/2,j+1/2} = \hat{\mu}_{i+1/2,j+1/2} \left(D_x[u_z]^{n}_{i+1/2,j+1/2} + D_z[u_x]^{n}_{i+1/2,j+1/2} \right) + \\
+ (f_{xz})^{n}_{i+1/2,j+1/2},
\end{gathered} \tag{1}$$

where $\boldsymbol{u} = (u_x, u_z)^T$ is the velocity vector, σ_{xx}, σ_{zz}, and σ_{xz} are the components of the stress tensor. The corresponding grid functions are defined at the staggered grid, so that $g^N_{I,J} = g(N\tau, Ih_x, Jh_z)$, where τ is the time step, h_x and h_z are the spatial steps, indices I, J, N can be either integer or half-integer and g is a smooth enough function. Functions f_{xx}, f_{zz}, and f_{xz} represent the right-hand side. Parameter ρ is the mass density, λ and μ are the Lame parameters. Typically all the model parameters are defined at the integer grid pints. In contrast, the arithmetic average for density and the harmonic average for μ are used to obtain them in the fractional points [14,23]. Operators D_t, D_x, and D_z are:

$$\begin{gathered}
D_t[g]^N_{I,J} = \frac{g^{N+1/2}_{I,J} - g^{N-1/2}_{I,J}}{\tau} = \frac{\partial g}{\partial t} + O(\tau^2), \\
D_x[g]^N_{I,J} = \frac{1}{h_x} \sum_{m=0}^{M} \left(g^N_{I+m+1/2,J} - g^N_{I-m-1/2,J} \right) = \frac{\partial g}{\partial x} + O(h_x^{2(m+1)}), \\
D_z[g]^N_{I,J} = \frac{1}{h_z} \sum_{m=0}^{M} \left(g^N_{I,J+m+1/2} - g^N_{I,J-m-1/2} \right) = \frac{\partial g}{\partial z} + O(h_z^{2(m+1)}).
\end{gathered} \tag{2}$$

Note that the scheme (1) approximates linear hyperbolic system, thus due to the Courant stability criterion $\tau = Ch$, where h can be considered as the $\min\{h_x, h_z\}$. Thus, we can consider the scheme depending only on parameter h.

In short notations, the finite-difference problem can be rewritten as:

$$R_h[\boldsymbol{v}_h] = \boldsymbol{\phi}_h(t)\delta(x - x_s^l)\delta(z - z_s^l),$$

where $\boldsymbol{\phi}(t)$ is the time impulse of the source, δ is the delta-function, (x_s^l, z_s^l) are the coordinates of the $l-$th source. Vector $\boldsymbol{v}_h$ is the vector of the solution combined with velocity vector components and the stress tensor components.

According to the theory of finite-difference schemes, if the scheme is stable and it approximates the original differential operator, then the numerical solution converges to the true solution, and the following estimation takes place:

$$\|\boldsymbol{v} - \boldsymbol{v}_h\| \leq Ch^r, \tag{3}$$

where the convergence rate r coincides with the order of approximation, which is 2 for the considered case. Constant C is independent of the grid step.

Consider now the scheme (1) but with two different grid steps $h_1 \leq h_2$. According to the estimation (3) one gets:

$$\|\boldsymbol{v} - \boldsymbol{v}_{h_1}\| \leq \|\boldsymbol{v} - \boldsymbol{v}_{h_2}\|.$$

However, using the smaller grid steps $h_1 \leq h_2$ leads to the problem size increase. The idea of the NDM-net is to construct a map:

$$\mathcal{M}[\boldsymbol{v}_{h_2}] = \tilde{\boldsymbol{v}}_{h_2},$$

so that

$$\|\tilde{\boldsymbol{v}}_{h_2} - \boldsymbol{v}_{h_1}\| \leq \varepsilon_{21} << Ch_2^r.$$

If the error ε_{12} is small one gets:

$$\|\tilde{\boldsymbol{v}}_{h_2} - \boldsymbol{v}\| \leq \|\tilde{\boldsymbol{v}}_{h_2} - \boldsymbol{v}_{h_1}\| + \|\boldsymbol{v}_{h_1} - \boldsymbol{v}\| \leq \varepsilon_{21} + Ch_1^r < Ch_2.$$

In other words, we suggest using NDM-net to map the data computed on a coarse mesh to the data computed using a fine enough grid.

In our research, we apply the Convolutional Neural Network (CNN), and in particular, a U-Net [3,16]. As it was mentioned above, the most time-consuming part of the NDM-net implementation is the generation of the training dataset, which is the simulation of the wavefield for a small number of sources using a fine enough grid. We suggest optimizing the choice of sources to minimize the training dataset and, thus, improve the NDM-net performance.

3 Training Dataset Construction Approaches

In this paper, we discuss three ways to construct training datasets. Two of them were presented in [3] and [2]; the third one is new and will be compared against the other two.

3.1 Equidistantly Distributed Seismograms

The first approach was suggested together with NDM-net in [3]. Introduce a seismogram corresponding to a fixed source number l as:

$$\boldsymbol{u}^l = \boldsymbol{u}(t, x, z_0, x_s^l, z_s^l),$$

where $\boldsymbol{u}$ is the velocity vector. We need to point out that it is convenient to introduce a change of variables $x_o = x - x_s^l$, variable x_o is called the offset, and it is fixed for all source positions. As a result, the seismic data can be viewed as a set of 2D functions of time and offset or a single 3D function of time, offset and source position. We prefer the first option, thus, the entire set of seismograms can be represented as:

$$U = \bigcup_{l=1,\dots,L_s} \boldsymbol{u}(x_s^l, x_o, t) = \bigcup_{l=1,\dots,L_s} \boldsymbol{u}(\boldsymbol{x}_s^l),$$

and our goal is to construct the subset

$$U^{tr} = \bigcup_{l \in M_t} \boldsymbol{u}(\boldsymbol{x}_s^l) \subseteq U,$$

where M_t is the set of indices of the seismograms from the subset; i.e., $M_t \subseteq \{1, ..., J_L\}$.

The first approach to construct the dataset U^{tr} was implemented to preserve the Housdorff distance between the training dataset and the entire dataset, where the distance is measured as the physical distance between the source positions. The Hausdorff distance between two sets D_1 and D_2 is defined as

$$\begin{aligned} H(D_1, D_2) &= \max\{\max_{i \in L_1} b(\boldsymbol{u}_i, D_2), \max_{j \in L_2} b(\boldsymbol{u}_j, D_1)\} = \\ &= \max\{\max_{i \in L_1} \min_{j \in L_2} d(\boldsymbol{u}_i, \boldsymbol{u}_j), \max_{j \in L_2} \min_{i \in L_1} d(\boldsymbol{u}_j, \boldsymbol{u}_i)\}. \end{aligned}$$

where

$$b(\boldsymbol{u}_i, D_2) = \min_{j \in L_2} d(\boldsymbol{u}_i, \boldsymbol{u}_j),$$

is the distance between an element $\boldsymbol{u}_i$ and the set D_2, and $d(\boldsymbol{u}_i, \boldsymbol{u}_j)$ is the distance between two elements (seismorgamms).

We are interested in the distance between the training dataset D_t and the entire dataset U with $D_t \subset U$. Thus,

$$H(D_t, U) = \max_{i \in [1,\dots,J_s]} b(\boldsymbol{u}_i, D_t) = \max_{i \in [1,\dots,J_s]} \min_{j \in J_t} d(\boldsymbol{u}_i, \boldsymbol{u}_j).$$

Let us consider the distance $d(\boldsymbol{u}_i, \boldsymbol{u}_j) = |x_s^i - x_s^j|$, and require the corresponding Hausdorff distance to be fixed at value $M \cdot ds$, where M is an integer number, and ds is the distance between two neighboring sources (typically fixed for a seismic acquisition system). We get the training datasets U_s^N.

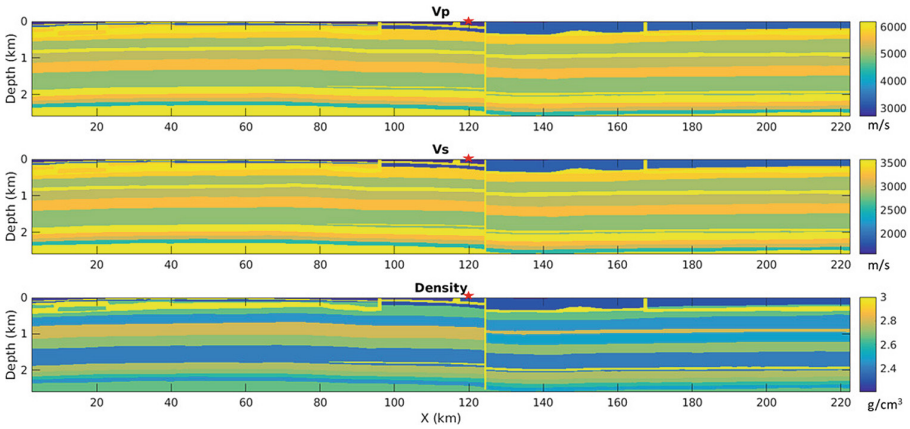

Fig. 1. Vanavar model

3.2 NRMS-Based Distance Between the Seismogramms

The second approach to construct the training dataset is based on preserving the Hausdorff distance but using the similarity measure between two seismograms instead of the indirect measure based on the distance between the sources. The similarity of the seismograms is often measured by the Normalized Root Mean Square of NRMS:

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

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 $d(\boldsymbol{u}^j, \boldsymbol{u}^k)$ as an average NRMS between $\boldsymbol{u}^j$ and $\boldsymbol{u}^k$ seismograms.

Having introduced this measure, one may define the Hausdorff distance and construct the training datasets to preserve this distance. We will denote such datasets U^Q_{NRMS}, where Q represents the particular value of NRMS in percentages.

To illustrate the construction of the dataset we consider a model presented in Fig. 1.

We simulated two sets of seismograms for this model, using grids with the step 2.5 m and 5 m. The entire datasets contain a 1901 seismogram with the sources placed 100 m apart from each other. After that, we constructed the training datasets and then computed the NRMS-based distances $b(\boldsymbol{u}_i, D_t)$ between each seismogram and the training datasets as a function of the source position. Figure 2 represents the distances to datasets U_s, and Fig. 3 illustrates the

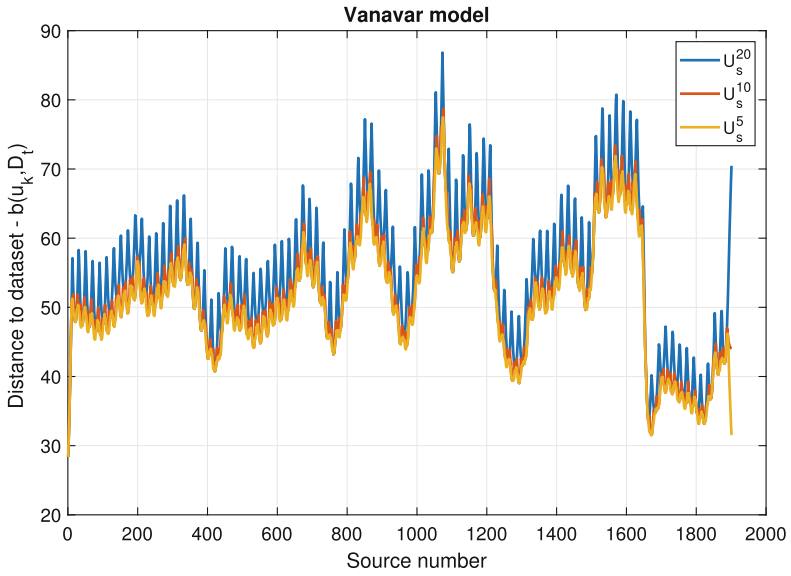

Fig. 2. Distances from the seismograms and the training datasets $b(\boldsymbol{u}^k, U_s^N)$ for different dataset of equidistantly distributed source U_s^5, U_s^{10}, and U_s^{20}.

distances to the dataset U_{NRMS}^{70}. Note that in the second case, we manage to preserve the maximal distance to the training dataset. However, there are domains where almost all seismograms were used, which strongly affects the algorithm's performance.

3.3 Adaptive Dataset Construction

As it follows from Fig. 2, NRMS between the seismograms corresponding to the neighboring sources may vary depending on the source position, and just a formal dissimilarity of the seismograms may not be the reliable criterion for dataset construction. In this paper, we suggest combining the two approaches so that we will choose the limiting NRMS distance to the training dataset, depending on the physical position of the sources. We introduced several levels of the Hausdorff distance limits in the NRMS metric. To do so, we considered the mean NRMS between the seismograms and its standard deviation as the function of the distance between the sources, as presented in Fig. 4. We constructed the limits recurrently as follows:

- $L_1 =< NRMS(1) > +\alpha\sigma(1)$, where $< NRMS(1) >$ is the mean NRMS at a distance $1 \cdot ds$, and $\sigma(1)$ is the standard deviation of the NRMS for the sources at a distance $1 \cdot ds$, and α is a parameter to be changed.
- Start the recursion:

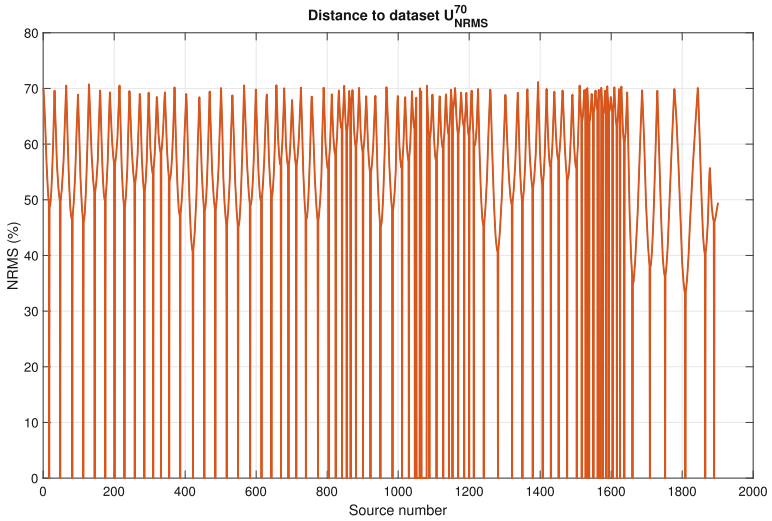

Fig. 3. Distance from the seismograms and the training dataset $b(\boldsymbol{u}^k, D^{70}_{NRMS})$.

- Solve the problem $\Delta_j = argmin|L_{j-1} - < NRMS(\Delta) > |$, where Δ_j is the physical distance between the sources.
- $L_j =< NRMS(\Delta_j) > +\alpha\sigma(\Delta_j)$

The algorithm stops if the solution of the *argmin* problem is the maximum of Δ. There is a parameter in this algorithm that is subject to change.

Constructing the set of limiting NRMS values, we may apply the procedure of the dataset generation according to the rule:

- start from the seismogram number 1, subset 1;
- accumulate the seismograms (in ascending order) to the current cluster while

$$\min_{i \in C_k} \max_{j \neq i} d(\boldsymbol{u}_i, \boldsymbol{u}_j) \leq L_m$$

  where i and j are indices of the seismograms from the current cluster C_k, and m is a current NRMS limit (which starts from one for a new cluster).
- if the set contains enough seismograms, which is

$$\min_{i \in U_k} \max_{j \neq i} |x^i_s - x^j_s| \geq R,$$

  we state that the cluster is complete, and the new element of the training dataset is the solution to the problem:

$$I_k = argmin_{i \in C_k} \max_{j \neq i} d(\boldsymbol{u}_i, \boldsymbol{u}_j).$$

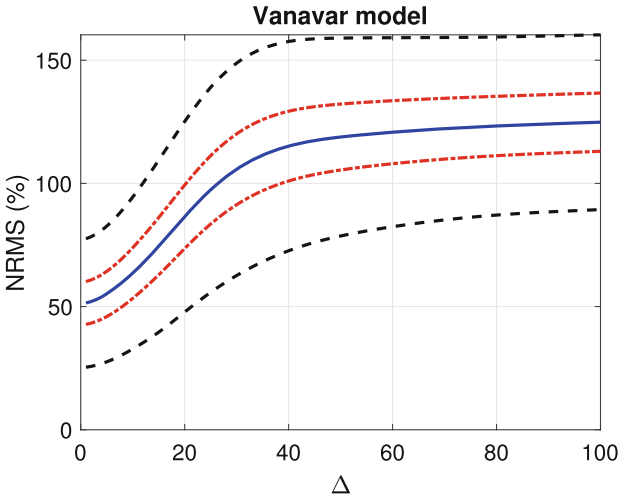

Fig. 4. Mean (solid line), STD (dash-dotted), and 3 STD (dashed) of NRMS-distance as a function of the distance between the sources.

– if the current set contains a small number of seismograms which is

$$\min_{i \in U_k} \max_{j \neq i} |x_s^i - x_s^j| < R,$$

then we update NRMS limit to L_{m+1} and continue collecting seismogramms to the cluster.

This algorithm allows for overcoming the situation where the seismograms in the training dataset correspond to unreasonably close sources.

4 Numerical Experiments

4.1 Effect of the Parameters

First, we studied the effect of the parameters used in the algorithm on the number of seismograms in the training dataset. There are two empirically adjusted parameters: α, the coefficient defining the increase in the limiting NRMS level, and R, which governs the minimal number of seismograms in a cluster. Change in the minimal distance between the sources R would directly affect the number of seismograms in the training dataset. In particular, the higher R is, the fewer seismograms will be used. Moreover, the following estimate takes place $N_{cl} \leq N_{sg} \cdot ds/R$, where N_{cl} is the number of clusters or the number of seismograms in the training dataset, N_{sg} the total number of seismograms. The effect of α is less evident because α directly affects the segmentation of the limiting NRMS level, not the number of seismograms. However, the use of too small α would lead to the very fine discretization of the limiting NRMS. Thus, only the

physical proximity of the sources will govern the training dataset construction. We consider these two parameters varied as follows $R \in \{2, 5, 10, 20, 50\}$, and $\alpha \in \{0.2, 0.5, 0.7, 0.9, 1, 1.2, 1.5, 2, 3\}$. Table 1 presents the relation between the parameter α and the number of the segmentation levels of the limiting NRMS. As expected, the higher α, the smaller the number of segmentation levels. Similarly, when α increases, the number of clusters gets smaller (Table 2). It is due to the rapid increase of the limiting NRMS level that wider clusters of seismograms are allowed to form. The number of clusters increases with the decrease of the minimal distance between the sources, as expected. In general, the number of clusters varies slightly with the change of limiting distance (at least for intermediate values of α). However, for $R = 50 \cdot ds$, the number of clusters decreased significantly, making the training dataset underrepresentative.

Table 1. Number of segmentation levels for different α.

α	0.2	0.5	0.7	0.9	1	1.2	1.5	2	3
N_L	30	12	9	7	6	5	5	4	3

Table 2. Number of seismograms in the training dataset for different α and R.

$R \setminus \alpha$	0.2	0.5	0.7	0.9	1	1.2	1.5	2	3
2	197	135	117	110	104	96	84	75	58
5	165	129	111	108	101	94	83	74	58
10	131	110	100	92	92	88	80	71	56
20	85	78	75	73	73	70	66	62	52
50	37	35	34	32	32	32	17	10	8

4.2 Implementation of the NDM-Net

In our further study, we focused on two scenarios with $R = 10 \cdot ds$ in both cases and $\alpha = 0.9$ and $\alpha = 1.2$. We trained the NDM-net using two adaptive datasets and compared the results against our previous results, obtained with the training datasets with equidistantly distributed sources and with the NRMS-preserving technique. The NDM-net setup was the same for all numerical experiments as described in [3]. Having trained the NDM-net, we applied it to the entire dataset and compared the results with the accurate solution computed on the fine mesh. Figures 5 and 6 provide the seismogramm-by-seismogramm NRMS distances between the accurate and predicted solutions for the datasets constructed preserving NRMS and using the adaptive technique, respectively. Both approaches illustrate similar accuracy of the seismogram prediction. To make a qualitative comparison, it is worth computing an average NRMS over all seismograms and considering the number of seismograms in the training dataset.

In addition, we add the data corresponding to equidistantly distributed sources to complete the comparison. All data is provided in Table 3 and Fig. 7. Note that the proposed adaptive dataset construction illustrates both superiority in accuracy and the number of seismograms used in the training dataset.

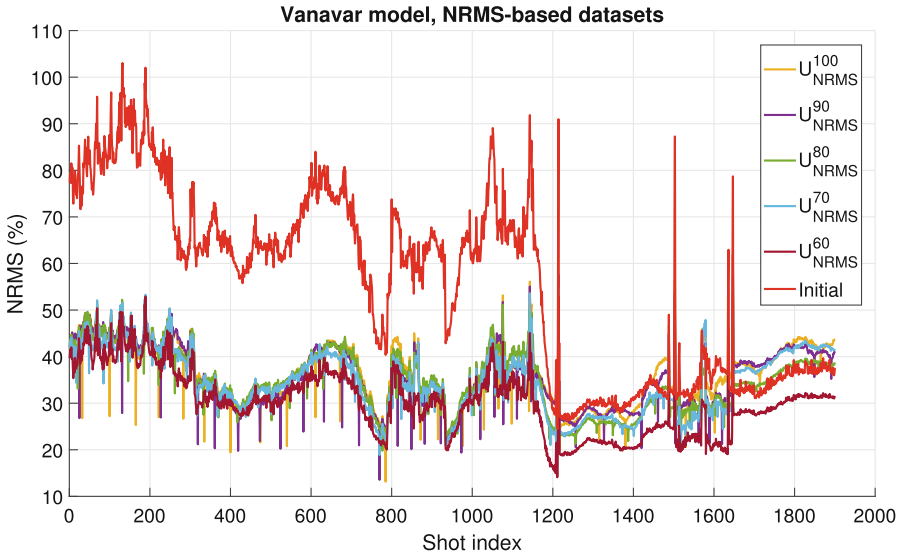

Fig. 5. seismogramm-by-seismogramm NRMS distances between the fine-grid solution and NDM-net corrected solutions for different training datasets based on the NRMS preserving.

5 Conclusions

This study presented a new strategy to construct the training dataset for the numerical dispersion mitigation network or NDM-net. In opposite to the two previous approaches, where the Hausdorff distance from the training dataset to the entire dataset was preserved globally, the new approach is adaptive which allows avoiding the use of unreasonably dense training datasets. In particular, in the previous works, the datasets were constructed to preserve the Hausdorff distance based on the Euclidean distance between the source positions. The resulting datasets consisted of the seismograms corresponding to the equidistantly distributed sources, with no information about the data proximity or dissimilarity accounted for. The other approach was based on the preservation of the distance measured in the data space. In particular, the NRMS measure of seismogram similarity was used. This method reduced the number of seismograms in the training dataset by a factor of three compared with equidistantly distributed sources. However, this approach may lead to the unreasonably dense spatial distribution of the sources corresponding to the seismograms from the

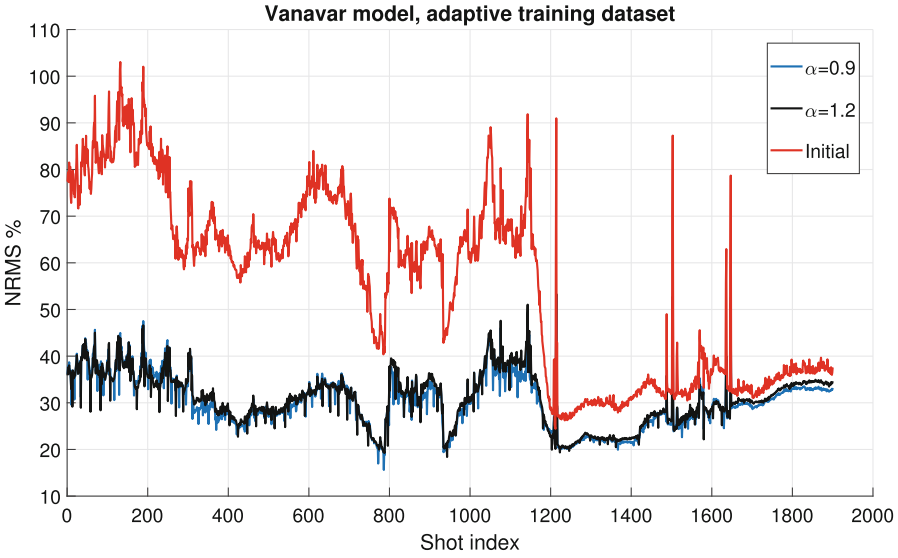

Fig. 6. seismogramm-by-seismogramm NRMS distances between the fine-grid solution and NDM-net corrected solutions for different training datasets constructed using the adaptive technique.

Table 3. Number of seismograms and resulting NRMS for different strategies of training datasets constructions

NRMS-preserving datasets		
Dataset	Number of sources	average NRMS
U^{60}_{NRMS}	414	28.16%
U^{70}_{NRMS}	109	30.28%
U^{80}_{NRMS}	56	34.69%
U^{90}_{NRMS}	43	35.11%
U^{100}_{NRMS}	34	35.68%
equidistantly distributed sources		
Dataset	Number of sources	average NRMS
U^{5}_{s}	86	44.28%
U^{10}_{s}	191	31.91%
D^{20}_{s}	283	29.41%
adaptive datasets		
Dataset	Number of sources	average NRMS
$U^{\alpha=0.9}$	92	30.32%
$U^{\alpha=1.9}$	88	30.83%

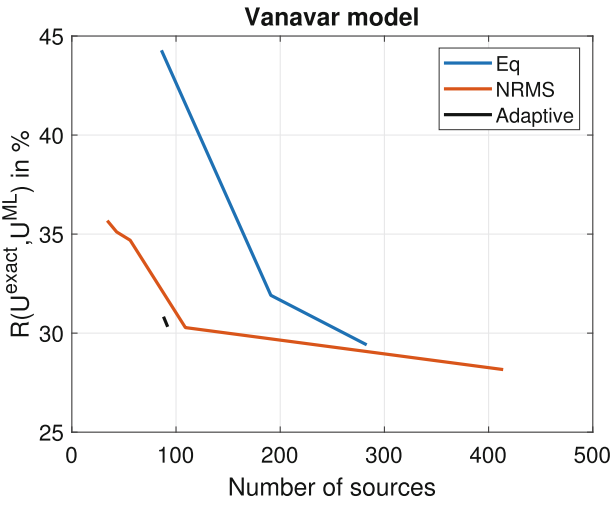

Fig. 7. The error depending on the number of seismograms used in the training dataset for different strategies of training dataset construction.

training dataset because of the local complexities of the model. Moreover, the two presented approaches require empirically chosen input parameters (the limiting Hausdorff distance) which strongly affect the result.

To overcome the drawbacks of the two approaches, we suggested a new strategy that adjusts the distance to the training dataset according to the current data. In particular, we introduce the segmented levels of limiting NRMS based on the mean of NRMS for different distances between the sources over the entire dataset. After that, we applied the procedure of constructing the training dataset with preserved NRMS. If the sources corresponding to the training dataset appears to be close to each other, we locally increase to limiting NRMS level, according to the segmented levels. We illustrated that this approach provides the tiniest training datasets among the three considered methods. Moreover, using the adaptive training dataset produces the lowest error of the NDM-net implementation. Additionally, the adaptive approach does not require input parameters calibration because it produces stable results for the appropriate set of parameters.

References

1. Ainsworth, M., Wajid, H.A.: Dispersive and dissipative behavior of the spectral element method. SIAM J. Numer. Anal. **47**(5), 3910–3937 (2009)
2. Gadylshin, K., Lisitsa, V., Gadylshina, K., Vishnevsky, D.: Optimization of the training dataset for numerical dispersion mitigation neural network. In: Gervasi, O., Murgante, B., Misra, S., Rocha, A.M.A.C., Garau, C. (eds.) Computational Science and Its Applications - ICCSA 2022 Workshops. ICCSA 2022. Lecture Notes

in Computer Science, vol. 13378, pp. 295–309. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-10562-3_22
3. Gadylshin, K., Vishnevsky, D., Gadylshina, K., Lisitsa, V.: Numerical dispersion mitigation neural network for seismic modeling. Geophysics **87**(3), T237–T249 (2022)
4. Kaser, M., Dumbser, M., Puente, J.d.L., 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., Broggini, F., Andersson, F.: Eliminating time dispersion from seismic wave modeling. Geophys. J. Int. **213**(1), 169–180 (2017)
7. Levander, A.R.: Fourth-order finite-difference P-SV seismograms. Geophysics **53**(11), 1425–1436 (1988)
8. Lisitsa, V., Podgornova, O., Tcheverda, V.: On the interface error analysis for finite difference wave simulation. Comput. Geosci. **14**(4), 769–778 (2010)
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. Lisitsa, V., Tcheverda, V., Vishnevsky, D.: Numerical simulation of seismic waves in models with anisotropic formations: coupling Virieux and Lebedev finite-difference schemes. Comput. Geosci. **16**(4), 1135–1152 (2012)
13. Mittet, R.: Second-order time integration of the wave equation with dispersion correction procedures. Geophysics **84**(4), T221–T235 (2019)
14. Moczo, P., Kristek, J., Vavrycuk, V., Archuleta, R.J., Halada, L.: 3D heterogeneous staggered-grid finite-differece modeling of seismic motion with volume harmonic and arithmetic averagigng of elastic moduli and densities. Bull. Seismol. Soc. Am. **92**(8), 3042–3066 (2002)
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. Shragge, J., Konuk, T.: Tensorial elastodynamics for isotropic media. Geophysics **85**(6), T359–T373 (2020)
19. Siahkoohi, 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

20. 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
21. Virieux, J.: P-SV wave propagation in heterogeneous media: velocity-stress finite-difference method. Geophysics **51**(4), 889–901 (1986)
22. Virieux, J., Calandra, H., Plessix, R.E.: A review of the spectral, pseudo-spectral, finite-difference and finite-element modelling techniques for geophysical imaging. Geophys. Prospect. **59**(5), 794–813 (2011). https://doi.org/10.1111/j.1365-2478.2011.00967.x
23. Vishnevsky, D., Lisitsa, V., Tcheverda, V., Reshetova, G.: Numerical study of the interface errors of finite-difference simulations of seismic waves. Geophysics **79**(4), T219–T232 (2014)