Geophysical Prospecting
EAGE EUROPEAN ASSOCIATION OF GEOSCIENTISTS & ENGINEERS
Geophysical Prospecting, 2021, **69**, 1405–1432 doi: 10.1111/1365-2478.13107

An artificial neural network approach for the inversion of surface wave dispersion curves

Alexandr V. Yablokov[1,2,3]*, Alexandr S. Serdyukov[1,2,3], Georgy N. Loginov[1] and Valery D. Baranov[4]

[1]*Trofimuk Institute of Petroleum Geology and Geophysics of SB RAS, Novosibirsk, 630090, Russia,* [2]*Chinakal Institute of Mining of SB RAS, Novosibirsk, 630091, Russia,* [3]*Novosibirsk State University, Novosibirsk, 630090, Russia, and* [4]*Laboratory of Technologies of Inverse Problems, Moscow, 115035, Russia*

Received November 2020, revision accepted April 2021

ABSTRACT

We describe a new algorithm for the inversion of one-dimensional shear-wave velocity profiles from dispersion curves of the fundamental mode of Rayleigh surface waves. The novelties of our approach are that the layer velocities and thicknesses are set as unknowns, and an artificial neural network is proposed to solve the inverse problem. We suggest that training data should be calculated for a set of random synthetic velocity layered models, while layer thicknesses and velocities should be set to fixed intervals, with ranges estimated based on the systematic application of empirical relations between Rayleigh and S-wave velocities to the dispersion data. Our main challenge is a total overhaul of the artificial neural network, which includes selecting the optimal artificial neural network architecture and parameters by performing a large number of numerical experiments. Our synthetic results show that the accuracy of the proposed approach outperforms that of the Monte Carlo approach. We illustrate our proposed method with West Siberia data processing obtained from an area of approximately 800 km^2. From a user perspective, the main strength of our method is the computationally efficient processing of large amounts of dispersion data, which make it well suited for four-dimensional near-surface monitoring.

Key words: Inversion, Surface wave.

INTRODUCTION

We consider the inversion of surface wave dispersion curves to obtain shear-wave velocity (V_S) profiles, which is a key step in the multichannel analysis of surface waves (MASW) (Park *et al.*, 1999; Socco *et al.*, 2010b). MASW is widely employed for noninvasive geotechnical site investigations (Strobbia *et al.*, 2010; Mahvelati and Coe, 2017; Serdyukov *et al.*, 2017; Rahman *et al.*, 2018) and can be applied to the estimation of statics corrections, which can increase the accuracy of deep reservoir exploration by seismic reflection data (Socco *et al.*, 2010a; Roohollah, 2013). Surface wave analysis is the preferred method for shallow subsurface time-lapse four-dimensional (4D) monitoring (Ikeda *et al.*, 2018), which is particularly important because of the growing demand for CO_2 monitoring during the injection and storage of CO_2 in carbon capture projects (White, 2009).

The standard method for inverting surface wave data is to minimize the misfit between the observed dispersion curve and the calculated dispersion curve (Xia *et al.*, 1999). With the assumption of a vertical one-dimensional (1D) layered elastic medium, the Rayleigh wave phase velocity depends on the shear-wave velocities (V_S), compressional wave velocities (V_P), density (ρ) and layer thicknesses (h). V_S has a dominant influence on surface wave dispersion data; moreover, inverting the Rayleigh wave phase velocity as a function of h seems to be

* E-mail: yablokov.alexandr93@gmail.com

more feasible than inverting it as a function of either V_P or ρ (Xia *et al.*, 1999).

In the inversion approach adopted by many previous MASW studies (Park *et al.*, 1999; Lai, 2005), the subsurface is subdivided into a reasonable number of layers, and each layer has a constant V_S. The number of layers and their thicknesses are defined by the user and kept fixed. The values of V_P and ρ are either assumed to be known or related to the values of V_S by a rule of thumb; for example, a constant Poisson's ratio is employed. A variety of local optimization methods have been developed to solve the problem of inverting surface wave phase velocity dispersion curves in the above-mentioned fixed-layer formulation. In the near-surface region, seismic velocities typically exhibit abrupt changes with depth (i.e. by a factor of 2–3 with a decrease in depth or an increase in depth). In this case, the standard gradient descent minimization method often provides V_S models with strong nonphysical fluctuations (Lai, 2005). This problem can be solved by using Occam's inversion (Lai, 2005; Rubaiyn *et al.*, 2018), which was originally proposed by Constable *et al.* (1987) to generate smooth models from electromagnetic sounding data. Generally, the use of a large number of layers, in addition to some rules of thumb to build the initial V_S profile, model constraints (Xia *et al.*, 1999; Cercato, 2009) and 'smooth' solution techniques, for example, Occam's inversion (Lai, 2005), can provide reliable results on the relative errors in V_S. These outcomes are sufficiently accurate for estimating the seismic site response.

Exact near-surface depths are essential for numerous applications, such as bedrock mapping (Miller *et al.*, 1999) and landslide analysis (Prayitna *et al.*, 2019). Although a fine-layered approximation can be utilized to determine the layer depth by the inverted V_s, a more promising approach is to directly determine these depths; that is, the unknown parameters should include both the V_S of each layer and the h of each layer. However, an increase in the number of unknown parameters increases the ill-posedness of the considered inverse problem. The most appropriate regularization strategy is to reduce the entire space of possible layered earth models. Cox and Teague (2016) proposed a method for conducting multiple inversions by the global search utilizing systematically varied layering parameterization of the model. Parameterization is defined by a unique 'layering ratio', and the ratio values are based on the user's experience. The 'layering ratio' represents a multiplier that is used to estimate the number of layers and corresponding depth ranges and systemically increases the potential thickness of each layer in the inversion parameterization based on the potential thickness of the layer directly above it. Global optimization is the preferred method for addressing both unknown velocities and thicknesses. A variety of global search methods, including the Monte Carlo method (Socco and Boiero, 2008), genetic algorithm (Yamanaka and Ishida, 1996; Dal Moro *et al.*, 2007), simulated annealing (Beaty *et al.*, 2002; Pei *et al.*, 2007), neighbourhood algorithm (Cox and Teague, 2016) and Grey Wolf optimizer (Song *et al.*, 2015), can be utilized to evaluate the minimum misfit solution in the entire space of possible parameters.

In our study, we utilize a fully connected artificial neural network (ANN) to reconstruct the velocity and thickness values for each layer of the model. Çaylak and Kaftan (2014) proposed a multilayer perceptron neural network for MASW processing of Rayleigh wave data. However, Çaylak and Kaftan (2014) presented only preliminary results and did not describe how the possible ranges of unknown parameters to calculate the training dataset were selected. Cao *et al.* (2020) proposed an interesting alternative to classical neural networks by using mixed density networks (MDN) to separately obtain the Bayesian posterior distribution over the S-wave velocity for each layer of the model. The drawbacks of this approach are that the number of layers and velocity ranges for each layer must be known, and it only yields S-wave velocity values. Moreover, Cao *et al.* (2020) do not explain how to obtain a model a priori to train the MDN or how the architecture of the MDN is designed.

In several studies, seismological applications of neural networks were considered, in which surface velocities were inverted for the layer thickness (Moho depth) (Devilee *et al.*, 1999; Meier *et al.*, 2007; Cheng *et al.*, 2019) or smooth velocity models were used (Hu *et al.*, 2020). In recent research (Hu *et al.*, 2020), a new algorithm for deriving 1D shear-wave velocity models from surface-wave dispersion data with convolutional neural networks (CNNs) is proposed. Hu *et al.* (2020) use phase and group velocity images that are generated from dispersion curves computed from reference 1D V_S models for CNN training. Constructing the velocity images from the dispersion curves seems redundant; however, this approach provides the possibility of processing the surface wave dispersion data via full wave inversion. Hu *et al.* (2020) stated that CNNs can address more complicated nonlinear inverse problems than fully connected neural networks. However, since CNNs depend on a larger number of parameters, it is more difficult to tune them and generalize an inversion algorithm. With the approach proposed by Hu *et al.* (2020), a neural network architecture must be created and configured when processing new data. Hu *et al.* (2020) designed different CNNs for processing data from two regions (continental China and southern California).

In summary, almost all existing methods for the application of neural networks to seismic surface wave data inversion for geological structures, except the method of Çaylak and Kaftan (2014) (which is preliminary research), are aimed at the inversion of layer thicknesses or velocities but not both simultaneously. Previous papers provide few details on neural network architectures, parameter adjustments, inversion accuracy tests and comparisons with other inversion methods (e.g. Monte Carlo inversion). However, these issues are important for the efficient and accurate inversion of dispersion data. In addition, most previous studies have focused on seismological data processing, which differs from MASW data processing. Our study seeks to address these gaps.

The remainder of this paper is organized into three main sections, covering methodology, synthetic examples and an example using field data. In the methodology section, we propose an algorithm to accurately define the space of possible layered models for calculating the training set for the ANN. In the second section, 'Synthetic data processing', we show that the ANN inversion has reasonable accuracy that outperforms that of the Monte Carlo approach. The comparison of the ANN method with the Grey Wolf optimizer , which is a recently proposed advanced global search method, reveals similar accuracies. In the last section, 'Western Siberia field data processing', we illustrate the prospects of applying the proposed ANN method for the efficient processing of large data via the inversion of 35,010 surface wave dispersion curves from seismic exploration data gathered in Western Siberia.

METHOD

In our research, we consider the inversion of the surface wave dispersion curve for a one-dimensional (1D) stratified V_S model that is composed of a number of layers. The proposed inversion method consists of three steps. In the first step, we follow Cox and Teague (2016) and set the variation ranges of V_S and h to define the space of possible layered models. In the second step, we create a training dataset; that is, we calculate the dispersion curves for numerous random models from the previously defined space. In the last step, a fully connected ANN is trained (on the previously generated training dataset) and implemented using the dispersion curve data to obtain 1D V_S profiles. We consider each step here.

Estimation of the layered model ranges

The frequency-dependent surface wave velocity is considered a function of its wavelength ($V_R(\lambda)$). To estimate the depth range of the ith layer ($d^i_{\min}$ and $d^i_{\max}$), Cox and Teague (2016) introduced an empirical parameter that is referred to as the 'layering ratio' (Ξ), which represents a multiplier that systematically increases the potential thickness of each layer based on the potential thickness of the layer directly above it. The value of Ξ determines the number of layers. Cox and Teague (2016) considered the set of possible layering ratios $\Xi \in \{1.2, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 5.0\}$ and showed that the incorrect choice of Ξ leads to maximum errors of 150% for V_S and 130% for depth in the resulting model when using the neighbourhood algorithm for the inversion of dispersion curve data.

Consequently, Cox and Teague (2016) proposed that the minimum misfit between the experimental dispersion curve and the theoretical dispersion curve should be monitored to select the optimal value of Ξ. The algorithm for the estimation of the ranges of the layered model is given in the appendix of the paper. We propose a similar algorithm for the adjustment of the layering ratio but use the Monte Carlo method with a small-scale kernel (1000 velocity models and dispersion curves) for the benefit of computational efficiency. A more complete inversion is performed after adjustment of the optimal value of the layering ratio. The proposed algorithm consists of the following steps:

1. Choose the trial value of Ξ.

2. Estimate the possible model parameter ranges.

3. Set the Monte Carlo kernel: 1000 velocity models are selected according to a uniform distribution within the estimated model parameter ranges, and dispersion curves of phase velocities are calculated using parallelization on CPUs by OpenMPI.

4. Perform Monte Carlo inversion: the solution is the velocity model from the kernel for which the dispersion curve has the minimum root-mean-square misfit with the observed dispersion curve.

5. Compute the mean squared error: $V_R^{\mathrm{MSE}} = \frac{1}{M}\sum_{i=1}^{M}(V_R^{\mathrm{true},i} - V_R^{\mathrm{restored},i})^2$, where M is the number of frequencies, N is the number of layers and V_R^{true} and V_R^{restored} are the true and restored phase velocities, respectively.

6. Repeat stages 2–5 100 times.

7. Repeat stages 1–6 for all trial values $\Xi \in \{1.2, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 5.0\}$

8. Obtain the optimal value Ξ at which the median average of all errors V_R^{MSE} reaches a minimum.

An example of the use of the algorithm for the adjustment of the layering ratio is given in the section 'Synthetic data processing experiments'.

Creating a training dataset

Recall that for a horizontal homogeneous medium, the Rayleigh wave phase velocity is a solution of the dispersion equation (Lai, 2005):

$$\mathcal{F}_R[V_S(z), V_P(z), \rho(z), \omega] = 0, \tag{1}$$

where $\mathcal{F}_R(\omega)$ is the unknown Rayleigh wave velocity; $V_P(z)$, $V_S(z)$ and $\rho(z)$ are the seismic compressional wave velocity, shear-wave velocity and density, respectively, which are functions of the depth z; and ω is the frequency. Equation (1) is nonlinear and yields multiple solutions for the Rayleigh wave modes. We consider the fundamental mode, namely, the minimum solution of equation (1).

The particular case of a horizontal homogeneous medium, that is a finite-layer stratified velocity model that can be described by four vectors of parameters, is considered: $\mathbf{V_S} = (V_S^1, \ldots, V_S^N)$, $\mathbf{V_P} = (V_S^1, \ldots, V_S^N)$, $\boldsymbol{\rho} = (\rho^1, \ldots, \rho^N)$, $\mathbf{h} = (h^1, \ldots, h^{N-1})$, where the last Nth layer is the half-space. Since the Rayleigh wave phase velocity is much less sensitive to V_P and ρ than to V_S and h (Pan *et al.*, 2018), we follow a common approach and assume the constant velocity ratio $V_P/V_S = \text{Const} = \sqrt{2(1-\nu)/(1-2\nu)}$, where ν is Poisson's ratio. In our research, which also includes synthetic and field data processing examples (presented below), we set $\nu = 0.35$ and $\rho = \text{Const} = 1900$ kg/m^3. These parameters are typical for soils. Let us denote the corresponding vector of Rayleigh wave velocity values $V_R(w_i)$, $i = 1, \ldots, M$ for the observed series of frequencies as $\mathbf{V}_R = (V_R^1, \ldots, V_R^M)$.

Thus, the considered forward problem operator is written as follows:

$$\mathbf{V}_R = \mathbf{F}\left[V_S^1 \ldots V_S^N, h^1 \ldots h^{N-1}\right]. \tag{2}$$

To calculate the artificial neural network (ANN) training dataset, the forward problem (2) should be solved many times. We suggest using multicore computing when solving equation (2) with the transfer matrix method (Thomson, 1950; Haskell, 1953), which is also referred to as the Thomson–Haskell method.

Artificial neural network training and dispersion data inversion

The training process consists of obtaining the weight values, that is, the parameters of the ANN data model, that provide the best fit to the V_S and h values of the training dataset. A trained ANN is an operator that maps the observation space into the model parameter space:

$$\mathbf{V_S}, \mathbf{h} = \hat{\mathbf{F}}[\mathbf{V_R}], \tag{3}$$

where $\hat{\mathbf{F}}$ is the ANN approximation of the inverse problem operator. The output parameters of the inversion problem (3) are the vectors of V_S and h.

Numerical studies

A selection of the optimal ANN architecture is crucial for accurate and efficient inversion of dispersion curves. For these purposes, we conduct a large number of synthetic data processing experiments. After the selection of the optimal ANN architecture and parameters, we study the accuracy of the ANN inversion and compare it with the Monte Carlo inversion by processing another large synthetic dataset.

Artificial neural network architecture selection

The following key factors have the greatest influence on the convergence of the ANN learning process:

1. Input and output dataset scaling,
2. The volume of the training and validation datasets,
3. The numbers of hidden layers and neurons,
4. Loss function type,
5. Activation function type and
6. Optimization algorithm and learning rate decreasing model.

We separately study the influence of every factor by changing each factor while locking the remaining factors. We estimate the ANN learning efficiency by analysing the learning curves, loss function values for the training and validation datasets at the end of the training and the inversion error distribution and its statistical moments (expected value, variance, skewness and kurtosis) for a large set of synthetic models. We examine the mean absolute error (MAE), which is computed based on a comparison of the synthetic 'true' and 'restored' (by the ANN) model parameters and dispersion curves:

$$\begin{aligned} V_S^{\text{MAE}} &= \frac{\sum_{i=1}^{N} \left| V^i_{S,\text{true}} - V^i_{S,\text{restored}} \right|}{N}, \\ h^{\text{MAE}} &= \frac{\sum_{i=1}^{N-1} \left| h^i_{\text{true}} - h^i_{\text{restored}} \right|}{N-1}, \\ V_R^{\text{MAE}} &= \frac{\sum_{i=1}^{M} \left| V^i_{R,\text{true}} - V^i_{R,\text{restored}} \right|}{M}, \end{aligned} \tag{4}$$

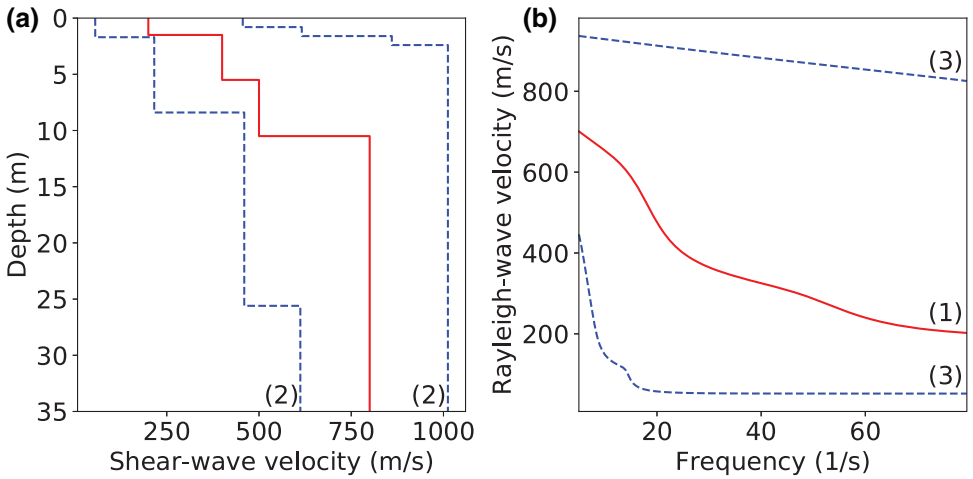

Figure 1 (a) Velocity model used for estimation of the possible model parameter ranges (red line) and the limits of the estimated possible model ranges (blue dashed lines); (b) synthetic dispersion curve (red line) observed from the model denoted by the red line in (a) and the limits of the possible dispersion curve ranges (dashed blue lines). The computed order of the graphs is denoted by the numbers in brackets.

where $V_{S,\text{true}}$ and h_{true} are the true values of the S-wave velocities and thicknesses, respectively; $V_{S,\text{restored}}$ and h_{restored} are the restored values of the S-wave velocities and thicknesses, respectively; N is the number of layers; $V_{R,\text{true}}$ and $V_{R,\text{restored}}$ are the true dispersion curves and restored dispersion curves, respectively, computed by $V_{S,\text{restored}}$ and h_{restored}, respectively; and M is the number of frequency samples.

The learning curve represents the loss function values of every training epoch. During a successful learning process, the loss function should decrease for every subsequent epoch. The loss function for the validation dataset or testing dataset, which is not applied for learning, represents the success of the restoration of the model parameters by the ANN.

We consider a large number of various synthetic models that are composed of 2–10 layers to solve the ANN architecture selection problem. Our results indicate that the optimal ANN architecture and hyperparameters are similar for models with different numbers of layers. We present the ANN architecture selection problem with numerical results for a reference four-layer model. The model is demonstrated in Fig. 1(a) (red line).

The values of the S-wave velocity and thickness of the model are also given in column 2 and column 3, respectively, of Table 1. The P-wave velocity values are computed using a fixed ratio $V_P/V_S = \text{Const} = \sqrt{2(1-\nu)/(1-2\nu)}$ with $\nu = 0.35$. The density is set to Const = 1900 kg/m^3. The synthetic dispersion curve for this model is plotted in Fig. 1(b) (red line). We consider the dispersion phase velocity curves in the frequency range of [0.5; 80] Hz with the discretization step of 0.5 Hz. We estimate the ranges of the possible values of V_S and h using the synthetic dispersion curve and the above-mentioned algorithm ($\Xi = 2.5$). The model parameter ranges are presented in Table 1 (columns 4 and 5) and plotted in Fig. 1(a) by blue dashed lines. We also show the synthetic dispersion curves, which define the limits of the possible phase velocities, in Fig. 1(b) (blue dashed lines).

Table 1 Synthetic model parameters for the ANN architecture selection experiments

Layer Number	True V_S (m/s)	True h (m)	Range Limits ($\Xi = 2.5$) V_S (m/s)	h (m)
1	200	1.5	56–456	0.8–1.7
2	400	4	216–616	0.8–6.7
3	500	5	460–860	0.8–17.2
4	800	∞	612–1012	∞

Dataset scaling

Data scaling is a necessary pre-processing step of ANN learning. We perform data transformation of the original range so that all values are within the range [0;1] according to the following formula:

$$x^{\text{scaled}} = \frac{x - \min(x)}{\max(x) - \min(x)}. \quad (5)$$

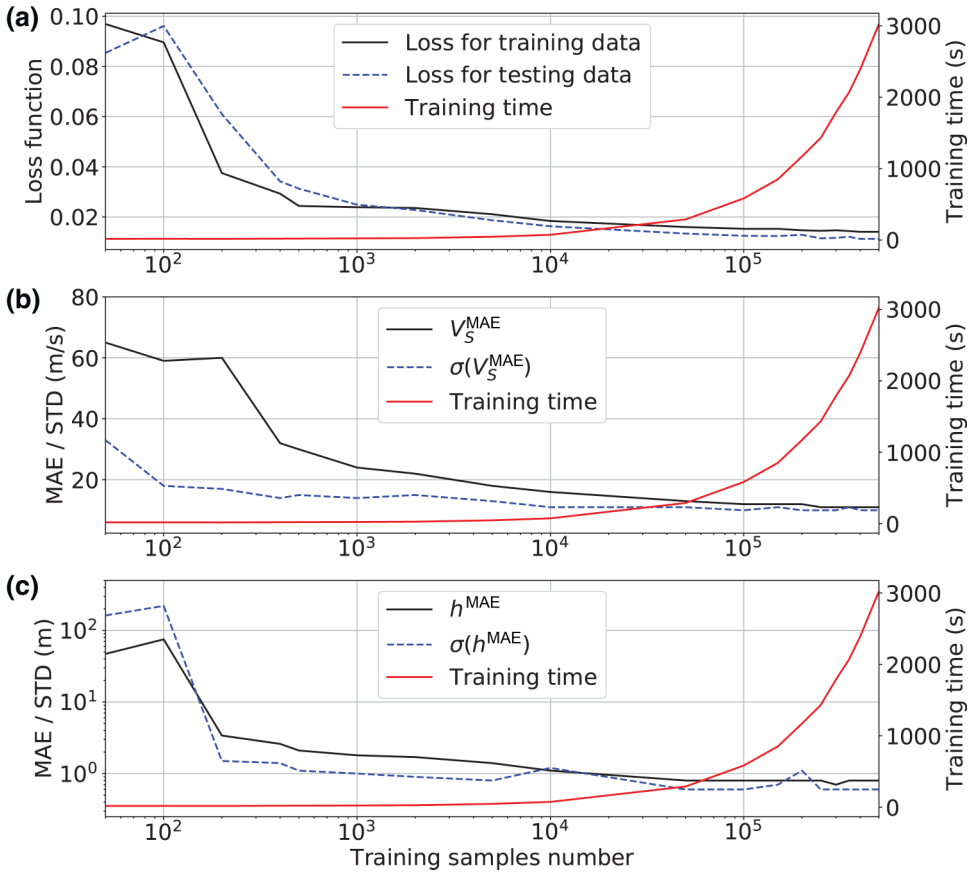

Figure 2 (a) Loss function dependencies on dataset volume for the training and testing datasets and training time; (b) dependence of V_S^{MAE} and its STD on the dataset volume; and (c) dependence of h^{MAE} and its standard deviation (STD) on the dataset volume.

Training data volume

The volume of the training dataset is a key parameter of ANN training. We consider training datasets with different volumes, from 50 to 5×10^5 samples. The dependence of the loss function values on the volumes of training data after 100 epochs of training is presented in Fig. 2(a).

The MAE (4), which is computed between the true model parameters and the parameters restored by ANN inversion of the test data, and the dependence on the volume of training data is presented in Fig. 2(b, c). Thus, every point in panels 2(a–c) corresponds to a separate synthetic data processing experiment.

Both the loss function (Fig. 2a) and MAE values (Fig. 2b,c) show that the ANN learning efficiency improves with an increase in the dataset volume. However, the training time (red curve in Fig. 2) increases excessively at the interval of 10^4–10^5 samples. Thus, we are forced to obtain a trade-off between an acceptable error level and the computation time.

Layer thickness (h) has less of an influence on the phase velocity than the S-wave velocity (V_S), so this parameter is harder to recover. Since we aim for accurate inversion of dispersion curves for both velocities and thicknesses, we pay more attention to the reconstruction of the layer thicknesses during our experiments. Note that h^{MAE} decreases until approximately 10^5 samples, whereas V_S^{MAE} continues to decrease after 10^5 samples. Taking into account the need for a suitable ANN training time, we choose a dataset volume of 2.5×10^5 samples as the optimal dataset. The testing dataset volume is selected in proportion to the volume of the training dataset according to a fixed ratio. From the experimental results, we conclude that the size of the validation dataset should be approximately 30% of the size of the training dataset.

Hidden layers and neuron numbers

General rules have not been provided for selecting the number of hidden layers and neurons. Introducing more than one

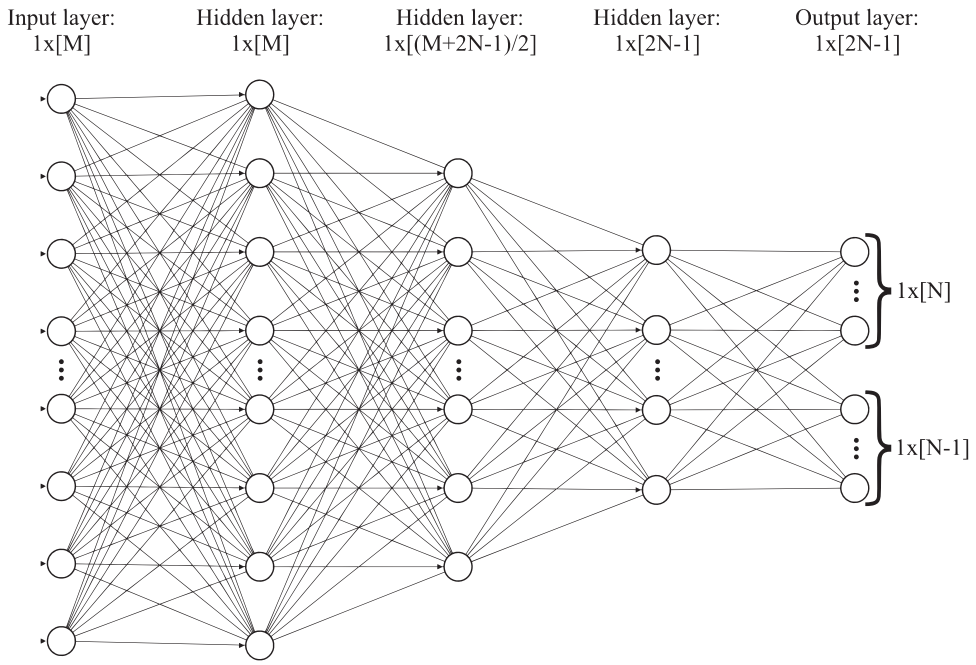

Figure 3 ANN architecture.

hidden layer allows the ANN to approximate the nonlinear dependencies between the input data and output data. Since dispersion curve inversion is a highly nonlinear problem, the number of hidden layers should exceed one. On the other hand, too many layers can lead to overtraining. Overtrained ANNs can provide almost 100% restoration accuracy of the training data, but the new data inversion results will not be reliable.

Our investigation of the optimal number of ANN hidden layers reveals that, regardless of the other ANN hyperparameters (e.g. the number of neurons or the type of activation function), the minimum number of hidden ANN layers to achieve inversion results is three (not including the input and output layers). Using a larger number of hidden layers does not significantly improve the accuracy but increases the computational complexity.

Based on the results of our study, we suggest an empirical rule: the number of input layer neurons equals the number of dispersion curve frequency samples $l^{\text{input}} = M$; the number of output layer neurons equals the total number of components of the unknown V_S and h vectors, that is $l^{\text{output}} = N + (N-1)$, where N is the number of layers; the number of neurons in the first hidden layer l_1^{hidden} equals the number of input layer neurons l^{input}; the number of neurons in the third hidden layer l_3^{hidden} equals the number of output neurons l^{output}; and the number of neurons in the second hidden layer is equal to an intermediate value, that is, $l_2^{\text{hidden}} = (M + 2N - 1)/2$. The structure of the considered three-hidden-layer ANN is shown in Fig. 3.

Loss function selection

A large number of ANN loss functions are employed to solve regression problems; the most popular ANN loss functions are the MAE, mean squared error (MSE) and mean squared logarithmic error (MSLE). The MSE is more sensitive to outliers than the MAE, and the MSLE is a kind of MSE. We also consider two additional advanced loss functions based on Canberra and Bray–Curtis metrics. We conduct experiments for training the ANN and synthetic data inversion for the abovementioned loss functions.

We assume error values during synthetic model parameter prediction as a major criterion of the selection of the loss function type. We present the distribution envelopes of the MAE (4) between the true velocity and the restored velocity V_s in Fig. 4(a) (V_S^{MAE}) and thickness h in Fig. 4(b) (h^{MAE}). In addition, we show the general statistical characteristics of the V_S^{MAE} and h^{MAE} distributions in Table 2. The optimal loss function should provide lower expected values and variance but large skewness and kurtosis. The MAE and Bray–Curtis provide lower expected values of h^{MAE} and have equal variance.

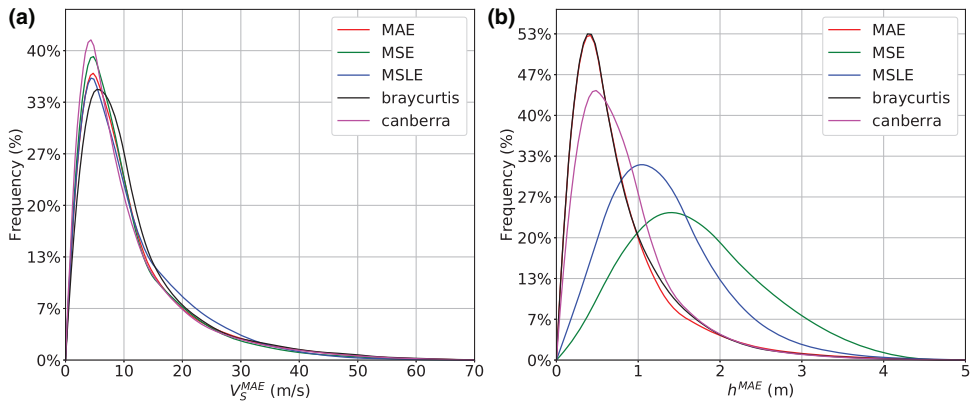

Figure 4 Distribution envelopes of the MAE between true and ANN-restored model parameters for different loss functions: (a) V_S^{MAE} envelopes and (b) h^{MAE} envelopes.

Table 2 General statistical characteristics of the distributions of V_S^{MAE} and h^{MAE} for different kinds of loss functions

Model Parameter	Loss Function	Expected Value	Variance	Skewness	Kurtosis
V_S^{MAE}	Canberra	9.16	48.33	1.25	3.95
	MSE	9.19	41.9	1.18	3.72
	MAE	9.74	49.56	1.24	3.92
	MSLE	9.92	48.96	1.01	3.17
	Bray–Curtis	10.24	50.12	1.21	3.97
h^{MAE}	MAE	0.65	0.19	1.17	3.84
	Bray–Curtis	0.65	0.19	1.04	3.39
	Canberra	0.73	0.18	0.85	3.14
	MSLE	1.22	0.32	0.36	2.49
	MSE	1.62	0.51	0.23	2.3

However, the MAE yields large skewness and kurtosis values. Of these two loss functions, the MAE yields better V_S^{MAE} statistical characteristics. Thus, we use the MAE as the loss function for ANN learning.

Activation function selection

To choose the type of activation function, we conduct a series of synthetic data processing experiments considering different types of activation functions and locking the remaining hyperparameters. We examine and compare the inversion results for the nine most popular types of ANN activation functions. A list of the examined activation functions is given in Table 3 (column 2). The loss function for the training and testing datasets, MAE (4) distributions between the true and restored V_S, and h for a large set of synthetic models (2.5×10^5 samples) are considered.

The MAE distributions of the model parameters are presented in Fig. 5. The general statistical characteristics of these distributions (expected values, variance, skewness and kurtosis) are presented in Table 3. The rows of Table 3 are separately sorted with the expected values in increasing order for V_S^{MAE} and h^{MAE}. Our focus is on restoring the accuracies of the thicknesses of the layers.

Note that the 'sigmoid' activation function provides the most reasonable statistical characteristics of h^{MAE} and is one of the three best activation functions according to the V_S^{MAE} statistical characteristics. Thus, we suggest using the 'sigmoid' activation function.

Batch normalization is employed after applying the activation function, which is a standard approach. The batch normalization procedure scales the data in hidden layers to the range [0, 1], which accelerates ANN training.

Optimizer selection

ANN learning is an optimization process, so a suitable optimization technique should be chosen. We test the

Table 3 General statistical characteristics of the distributions of V_S^{MAE} and h^{MAE} obtained by learning and inversion ANN for different kinds of activation functions

Model Parameter	Activation	Expected Value	Variance	Skewness	Kurtosis
V_S^{MAE}	Softplus	8.94	48.06	1.33	4.15
	Elu	9.63	50.92	1.19	3.75
	Sigmoid	9.97	52.23	1.22	3.87
	SELU	10.09	50.85	1.22	3.91
	Hard sigmoid	11.48	63.36	1.11	3.58
	Softsign	12.62	56.71	1.02	3.58
	ReLU	12.64	69.02	0.97	3.21
	Tanh	12.90	70.77	1.02	3.36
	Linear	29.64	166.48	0.29	2.37
h^{MAE}	Sigmoid	0.67	0.2	1.07	3.53
	Softplus	0.68	0.18	1.01	3.46
	Elu	0.72	0.2	1.0	3.4
	SELU	0.74	0.2	0.93	3.23
	Softsign	0.83	0.22	0.8	3.01
	Tanh	0.89	0.26	0.8	2.97
	Linear	1.44	0.49	0.28	2.17
	ReLU	4.49	2.47	−0.07	1.97
	Hard sigmoid	4.49	2.47	−0.08	2.01

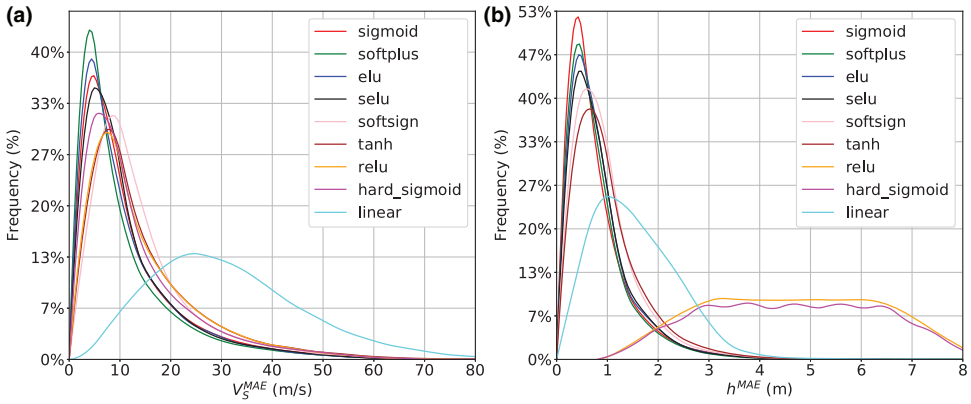

Figure 5 Distribution envelopes of the MAE between true and restored synthetic model parameters obtained for different activation functions: (a) V_S^{MAE} envelopes and (b) h^{MAE} envelopes.

following popular optimization algorithms that are widely employed for ANN learning: 'Nadam', 'Adam', 'SGD', 'RMSprop', 'Adagrad', 'Adadelta' and 'Adamax'. First, we implement the above-mentioned algorithms with default parameters from the Keras open-source software library (Keras, 2020). Figure 6(a,b) depicts the distributions of V_S^{MAE} and h^{MAE} (4), which are calculated based on a comparison of the true and restored synthetic model parameters after ANN inversion of the dispersion curves. Note that several optimizers (RMSprop, Adam, Adamax, Nadam) are more effective than others because of the adaptive decrease in the learning rate with the number of epochs.

The learning rate is a manageable option for the optimizer algorithms in the Keras library. We examine several models in which the learning rate decreases during the learning process. The examined common models with decreasing learning rates are shown in Fig. 7.

After testing the considered learning rate models for every optimizer algorithm, we conclude that the linear decreasing model of the learning rate (green line in Fig. 7) is the most effective. The optimal value of the initial learning rate is 10^{-2}.

We show the experimental results for ANN training and synthetic data inversion at various optimizer algorithms with the linear decreasing model of the learning rate in Table 4.

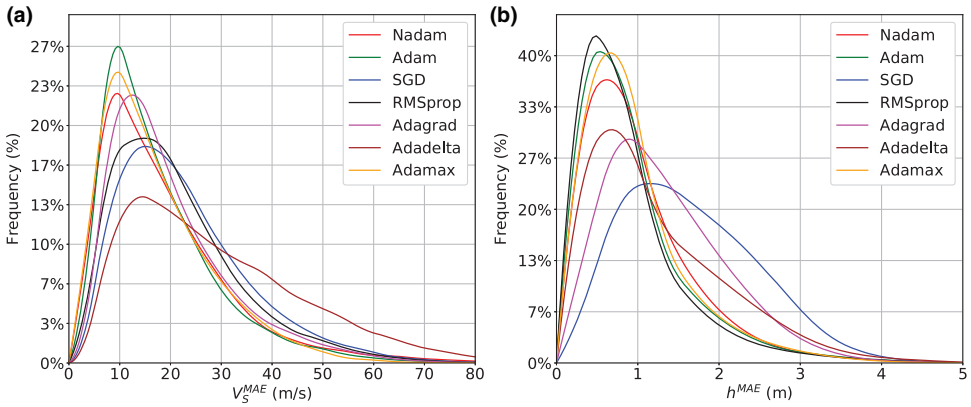

Figure 6 Distribution envelopes of the MAE between true and restored synthetic model parameters obtained for various optimizers from the Keras library with default parameters: (a) V_S^{MAE} envelopes and (b) h^{MAE} envelopes.

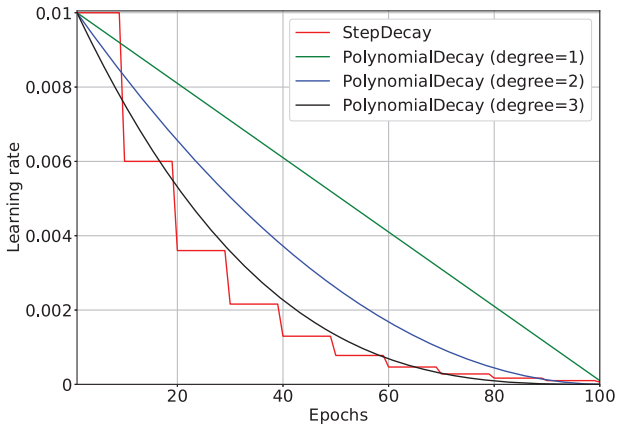

Figure 7 Learning rate-decreasing models for optimizer algorithms.

Distributions of the MAE are presented in Fig 8. Figure 9(a,b) reveal a change in the loss function with the number of epochs for the various optimizer algorithms for training the datasets (learning curve) for standard optimizers and optimizers of the linear decreasing model of the learning rate, respectively.

We select the Nadam (Dozat, 2016) optimization algorithm based on the synthetic data processing results.

Proposed ANN architecture

We summarize the results for the selection of the optimal ANN architecture and parameters. The proposed feedforward ANN architecture scheme is given in Table 5. During the preprocessing step, the input and output datasets are scaled from the original range so that all values fall within the range [0,1]

Table 4 General statistical characteristics of the distributions of V_S^{MAE} and h^{MAE} obtained by the learning and inversion of the ANN for different kinds of optimizers

Model Parameter	Optimizer	Expected Value	Variance	Skewness	Kurtosis
V_S^{MAE}	Adamax	8.71	47.26	1.31	4.05
	Nadam	9.79	50.12	1.25	4.01
	RNSprop	10.7	53.71	1.13	3.67
	Adam	10.16	51.34	1.21	3.88
	Adagrad	16.04	102.07	0.85	2.96
	SGD	20.08	117.54	0.67	2.72
	Adadelta	21.76	123.08	0.61	2.64
h^{MAE}	Nadam	0.67	0.17	1.03	3.56
	Adamax	0.7	0.18	0.91	3.19
	Adam	0.72	0.19	0.93	3.25
	RNSprop	0.74	0.2	0.91	3.29
	Adagrad	1.32	0.43	0.39	2.35
	Adadelta	1.56	0.56	0.3	2.24
	SGD	1.62	0.59	0.27	2.22

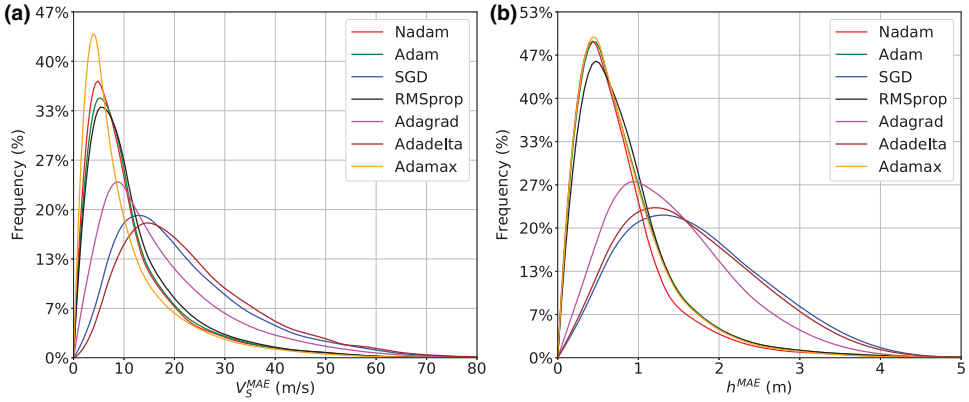

Figure 8 Distribution envelopes of the MAE between true and restored synthetic model parameters for various optimizer algorithms for the linear decreasing model of the learning rate: (a) V_S^{MAE} envelopes and (b) h^{MAE} envelopes.

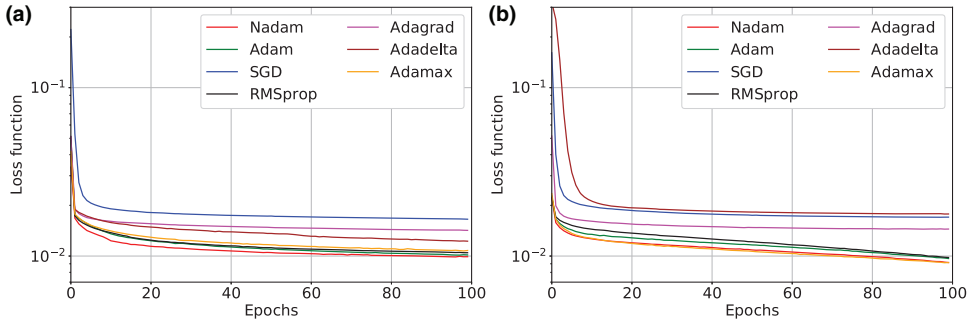

Figure 9 Change in the learning curve for the various optimizer algorithms: (a) standard optimizers and (b) optimizers for the linear decreasing model of the learning rate.

by formula (5). The loss function is the MAE. The optimizer algorithm is 'Nadam' with the linear decreasing model.

ANN inversion accuracy test

To study the accuracy of the proposed algorithm, we process synthetic dispersion curves for numerous velocity models. We consider velocity models composed of two to six layers. A frequency range of [0.5, 80] Hz with a discretization step of 0.5 Hz is considered. We summarize the whole testing workflow, which includes the generation of training and testing datasets and the performance of data inversion and error analysis.

1. We select five synthetic models with different numbers of layers (from two to six).The model parameters are presented in columns 3 and 4 of Table 6.

2. The Rayleigh wave fundamental mode phase velocity curve is calculated for each model. We estimate the ranges of the possible model parameters using the calculated synthetic phase velocity curve. The ranges of the model parameters are presented in Table 6 (columns 5 and 6).

3. We calculate 2.5×10^5 samples of the training $[V_S^{\mathrm{training}}; h^{\mathrm{training}}, V_R^{\mathrm{training}}]$ and 2.5×10^5 samples of the testing $[V_S^{\mathrm{true}}; h^{\mathrm{true}}, V_R^{\mathrm{true}}]$ datasets for each model (total 1.25×10^6 training and total 1.25×10^6 testing datasets).

4. We train the ANN using the training dataset (this step is performed five times for each of the five ANNs, corresponding to models 1–5 from Table 6).

5. We perform the inversion for each sample model dispersion curve from the testing dataset $[V_R^{\mathrm{true}}]$ and obtain the corresponding resulting set of velocities and thickness values $[V_S^{\mathrm{restored}}; h^{\mathrm{restored}}]$ (this step is performed five times for each of the five ANNs, corresponding to models 1–5 from Table 6).

6. We calculate the MAE (4) between the 'true' and restored velocities $[V_S^{\mathrm{true}}; V_S^{\mathrm{restored}}]$ and thicknesses $[h^{\mathrm{true}}; h^{\mathrm{restored}}]$ (for all models 1-5 from Table 6).

We compare the proposed ANN inversion with Monte Carlo inversion for the considered testing datasets. We employ the same training datasets $[V_S^{\mathrm{training}}, h^{\mathrm{training}}, V_R^{\mathrm{training}}]$ as the Monte Carlo inversion core and the same testing dataset $[V_S^{\mathrm{true}}]$ for the inversion. For each dispersion curve $V_R^{\mathrm{true},i}$

Table 5 Selected ANN architecture. *M* indicates the frequency number in the dispersion curve, and *N* indicates the layer numbers in the velocity model

ANN Layer	Process	Size
1. Input layer	1. Input 2. Output	$1 \times [M]$
2. Hidden layer	1. Input 2. Neurons computing 3. Activation: sigmoid 4. Batch normalization 5. Output	$1 \times [M]$
3. Hidden layer	1. Input 2. Neurons computing 3. Activation: sigmoid 4. Batch normalization 5. Output	$1 \times [(M + 2N - 1)/2]$
4. Hidden layer	1. Input 2. Neurons computing 3. Activation: sigmoid 4. Batch normalization 5. Output	$1 \times [2N - 1]$
5. Output layer	1. Input 2. Activation: sigmoid 3. Output	$1 \times [2N - 1]$

from the testing dataset $[V_S^{\text{true}}]$, we perform the following processing:

1. Root mean square misfits (**S**) between $V_R^{\text{true},i}$ are computed for each curve from $[V_R^{\text{training}}]$ (we obtain vector **S** as the result).
2. A minimum element of vector **S**: $\min(\mathbf{S}) = S^{\min}$ and a corresponding dispersion curve $V_R^{\text{training},j}$ (that best fits the $V_R^{\text{true},i}$) are selected.
3. Velocity model parameters from the training dataset that correspond to $V_R^{\text{training},j}$ are the restored values $V_S^{\text{restored}}; h^{\text{restored}}$.
4. Steps 1–3 are performed for all curves, that is, for each sample dispersion curve from the testing dataset $[V_R^{\text{true}}]$.
5. The MAEs (4) between the true and restored velocity model parameters are calculated (separately between velocities $[V_S^{\text{true}}; V_S^{\text{restored}}]$ and thicknesses $[h^{\text{true}}; h^{\text{restored}}]$).

To compare the ANN and Monte Carlo inversion results, we consider the MAE distributions for shear-wave velocity V_S^{MAE} and thickness h^{MAE} (refer to Fig. 10). We also compute the moments of the V_S^{MAE} and h^{MAE} distributions, which are given in Table 7.

We believe that the expected values of $V_S^{\text{MAE}} = 8.1$ m/s and $h^{\text{MAE}} = 0.8$ m are acceptable in practice. The practical constraint of the multichannel analysis of surface waves (MASW) vertical resolution is determined by the minimal measurable surface wave wavelength. Typically, the minimal layer thickness, which can be defined from field seismic data (using 5–10 Hz geophones and other standard seismic exploration equipment) is approximately (or greater than) 1 m. Thus, the expected value of h^{MAE} for ANN (0.8 m) is near the MASW vertical resolution limit. The MAE is acceptable as an accuracy measure for the comparison of synthetic data results for different inversion algorithms using the same data. However, the MAE provides an absolute residual with physical dimensions, which may depend on the data that are being processed. To obtain the dimensionless value of the inversion accuracy, we use the weighted mean absolute percentage error (WAPE):

$$V_S^{\text{WAPE}} = \frac{\sum_{i=1}^{N} \left| V_S^{\text{true},i} - V_S^{\text{restored},i} \right|}{\sum_{i=1}^{N} V_S^{\text{true},i}} \times 100,$$

$$h^{\text{WAPE}} = \frac{\sum_{i=1}^{N-1} |h^{\text{true},i} - h^{\text{restored},i}|}{\sum_{i=1}^{N-1} h^{\text{true},i}} \times 100, \quad (6)$$

where N is the number of layers. WAPE (Baeldung, 2020) is a popular measure of the prediction accuracy and indicates the general effectiveness of the proposed algorithm. The expected values of WAPE are $V_S^{\text{WAPE}} = 1.7\%$ and $h^{\text{WAPE}} = 17.8\%$ for the ANN inversion. The expected values of WAPE for the Monte Carlo inversion are $V_S^{\text{WAPE}} = 3.4\%$ and $h^{\text{WAPE}} = 28\%$.

SYNTHETIC DATA PROCESSING

We consider the artificial neural network (ANN) inversion results for several synthetic models. We also compare the accuracy and computational performance of the proposed ANN method with those of Monte Carlo inversion (Socco and Boiero, 2008) and the Gray Wolf optimizer (GWO) algorithm (Song *et al.*, 2015) (which is an advanced global search optimization method) by processing both pure data and noisy synthetic data. Three typical near-surface synthetic models are considered. We study the dependence of the inversion results on the layering parameter Ξ, which is also employed when computing the training dataset and possible velocity models within Monte Carlo inversion and the GWO algorithm.

The synthetic one-dimensional (1D) velocity model parameters are presented in Table 8: the model with a positive velocity gradient is denoted model A; the velocity model with a low-velocity layer is denoted model B; and the model with a high-velocity layer is denoted model C. The model with the positive velocity gradient (model A) is a typical model for near-surface investigations and represents sequential sedimentary deposits. The model with a low-velocity layer (model B) simulates the pavement conditions in urban areas. The

Table 6 Synthetic model parameters used for the algorithm accuracy test

Model	V_S (m/s)	h (m)	Range Limits ($\Xi = 2.5$)	
			V_S (m/s)	h (m)
Model 1 (two layers)	200	1.5	100–220	0.9–1.9
	400	∞	368–500	∞
Model 2 (three layers)	200	1.5	100–344	0.8–1.7
	300	4	150–447	0.8–8.7
	400	∞	200–516	∞
Model 3 (four layers)	200	1.5	100–363	0.8–1.7
	400	4	240–520	0.85–6.6
	500	5	489–769	0.85–17.2
	800	∞	635–915	∞
Model 4 (five layers)	200	1.5	100–363	0.8–1.7
	400	4	225–505	0.85–5.1
	500	4	379–659	0.85–10.2
	600	5	533–833	0.85–20.4
	800	∞	628–908	∞
Model 5 (six layers)	200	1.5	100–363	0.8–1.7
	400	4	219–499	0.85–4.5
	500	4	344–624	0.85–8.3
	600	3	513–793	0.85–14.9
	700	5	599–879	0.85–26.8
	800	∞	627–907	∞

presence of a high-velocity layer in model C simulates a permafrost structure. The V_S and h ranges for the layering ratio $\Xi = 2.5$ estimated by the previously mentioned algorithm are given in columns 5 and 6 of Table 8.

Model A

The model A inversion results for various Ξ values are presented in Figure 11. The synthetic model parameters (layer depths and velocities) and ranges (minimum and maximum layer velocities and depths) of the reconstructed model are plotted as solid red lines and dashed-dotted blue lines, respectively, in Figure 11. Note that the total number of layers in the reconstructed model depends on Ξ. The estimated number of layers is larger than (Fig. 11a,b), equal to (Fig. 11c–f) or less than (Fig. 11g–i) that in the 'true' synthetic model. We invert the data for all nine considered values of Ξ. The results are illustrated in Fig. 11, in which the Monte Carlo results are plotted as dashed magenta lines; the GWO results are plotted

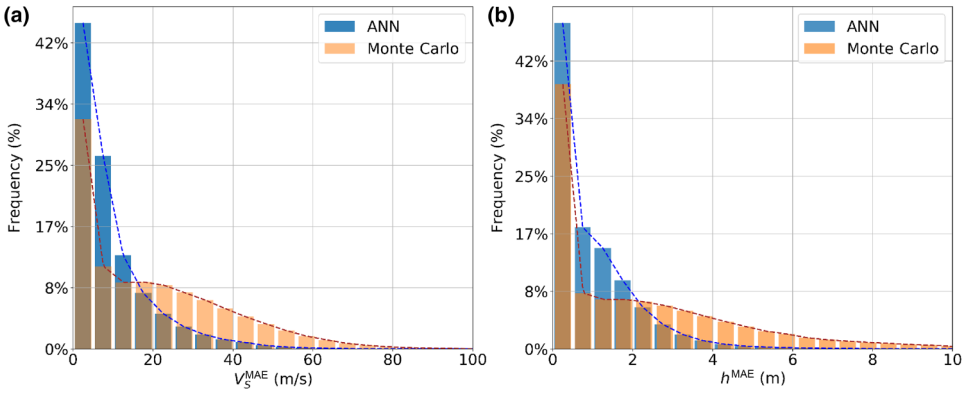

Figure 10 MAE distributions between the true and restored synthetic model parameters for the shear-wave velocity V_S^{MAE} (a) and thickness h^{MAE} (b) for the ANN and Monte Carlo inversion algorithms.

Table 7 General statistical characteristics of the distributions between the true and restored synthetic model parameters for shear-wave velocity V_S^{MAE} (a) and thickness h^{MAE} (b) for the ANN and Monte Carlo inversion algorithms

Type of Model Parameter	Inversion Algorithm	Expected Value	Variance	Skewness	Kurtosis
V_S^{MAE}	ANN	8.10	44.58	1.30	3.98
	Monte Carlo	17.23	235.12	0.62	2.25
h^{MAE}	ANN	0.84	0.54	0.99	3.04
	Monte Carlo	1.83	3.59	0.90	2.77

Table 8 Parameters of the velocity models used for training and for estimating ranges with the layering ratio $\Xi = 2.5$

				Range Limits ($\Xi = 2.5$)	
Model	Layer	True V_S (m/s)	True h (m)	V_S (m/s)	h (m)
Model A	1	200	1.5	100–345	0.83–1.7
	2	300	4	150–451	0.83–6.2
	3	500	8	278–628	0.83–15.7
	4	600	∞	364–714	∞
Model B	1	300	1.5	100–357	0.89–1.8
	2	200	4	150–374	0.89–6.7
	3	500	8	246–599	0.89–16.7
	4	600	∞	353–703	∞
Model C	1	200	1.5	100–363	0.85–1.7
	2	400	4	150–454	0.85–6.3
	3	300	8	200–509	0.85–16.0
	4	600	∞	340–690	∞

as dashed green lines; and the ANN results are plotted as dashed black lines. The ANN and Monte Carlo results are less sensitive to the layering ratio Ξ than the GWO results. For nearly all considered Ξ values, the ANN inversion results are closer to the true model than the Monte Carlo results. However, for appropriate values of Ξ, particularly $\Xi = 3.0$ (refer to Fig. 11d), the GWO inversion is more accurate than the ANN inversion.

We apply the proposed algorithm of the adjustment of the layering ratio (which was previously given) to obtain an optimal layering ratio parameter for model A synthetic data. The velocity of model A is plotted as a red line in Fig. 13. The model parameter values are also listed in Table 8.

The dependence of V_R^{MSE} between the true dispersion curve and restored dispersion curve on the layering ratio Ξ is depicted by the blue points in Fig. 12, and the blue line merges median values of V_R^{MSE}. The minimum of the median values of V_R^{MSE} indicates the optimal value of Ξ, which is 2.5 for the considered example.

We consider this case in more detail (refer to Fig. 13). In addition, synthetic noise is generated with a uniform distribution over the range of ±5% of the dispersion curve velocity at each frequency. The results of the noisy data processing are presented in Fig. 13(c, d).

The MAEs between the true and restored velocity model parameters (V_S^{MAE} and h^{MAE}) and dispersion curves (V_R^{MAE}) for model A are presented in Table 9. Note that the GWO inversion provides a better match between the dispersion curves; however, the ANN velocity is closer to the true velocity than is the GWO velocity. Based on the results of noisy data processing, the ANN method is evidently the most robust approach because it most closely matches the model parameters.

The randomness of the selection of velocity models for the ANN training dataset affects the ANN learning process and, therefore, the data inversion results. This kind of uncertainty associated with the choice of random parameters in the inversion algorithm is typical of global search optimization methods. In particular, the Monte Carlo inversion results are influenced by the random selection of the velocity models for its core. The GWO algorithm is also stochastic (e.g. it depends on the initial positions of the agents; refer to details in Song *et al.* (2015)). We compare these uncertainties for the ANN, Monte Carlo and GWO methods. Note that the considered uncertainties are not the only uncertainties, and accurate uncertainty appraisal within multichannel analysis of surface

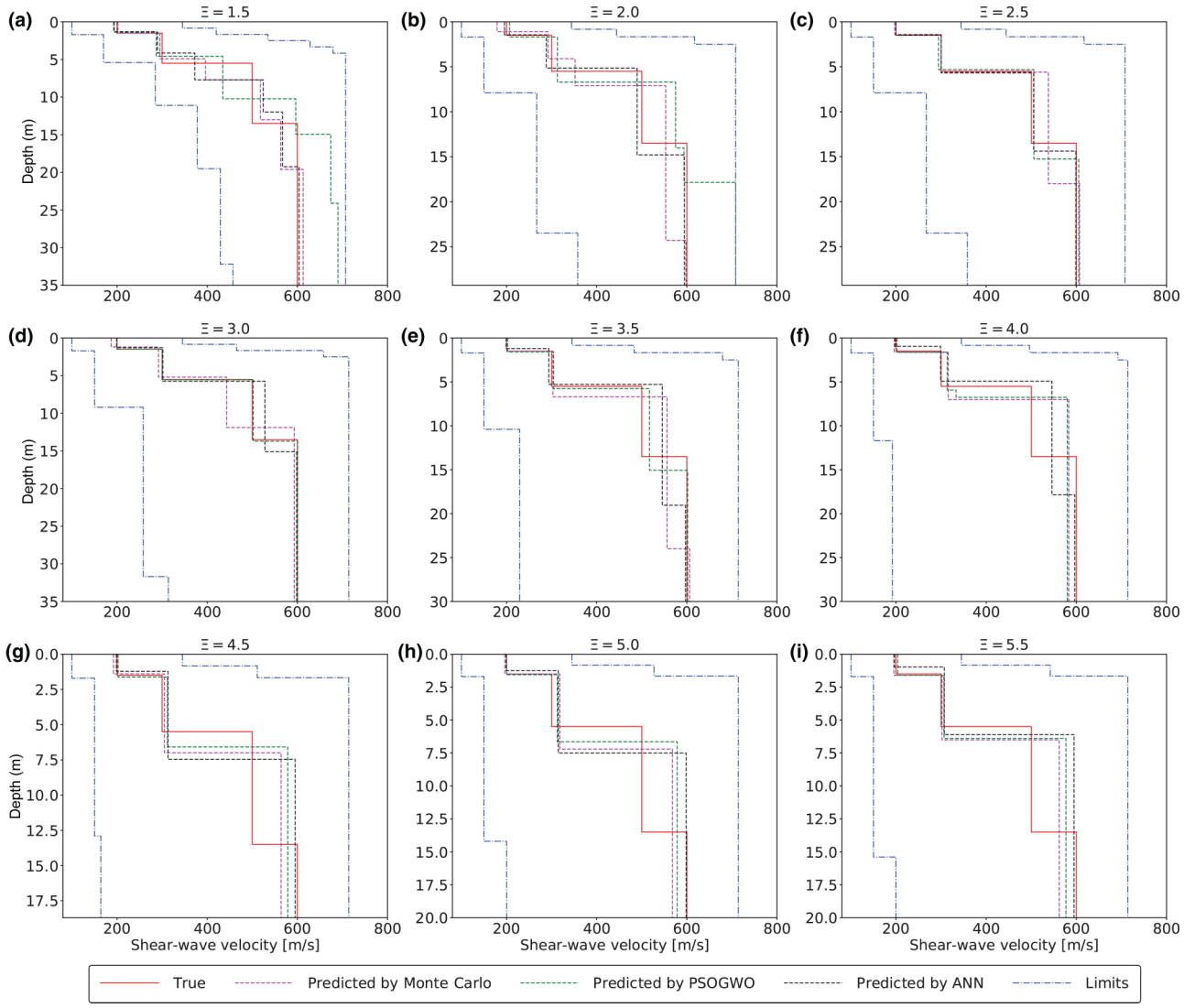

Figure 11 Comparison of the inversion results of model A among the ANN, GWO and Monte Carlo algorithms for different values of the layering ratio (Ξ).

waves (MASW) inversion is a more complex problem (see the Discussion section). In other words, we only study the stability of the results of the inversions according to the choice of the considered random parameters.

A series of independent experiments is performed to invert the synthetic dispersion curves of models A, B and C. The independence of the experiments means that for each inversion by the ANN or Monte Carlo method, a new set of training data is calculated using the new set of the randomly selected velocity models; that is, the ANN is retrained each time based on the new training dataset, and the core of the Monte Carlo method is recomputed. The GWO inversion involves the application of a global optimization search 20 times (as proposed in Song *et al.* (2015)) and selection of a result (from these 20 trials) that provides a minimum misfit between the dispersion curves for one run of the inversion. Thus, we consider 100 results of independent inversion experiments using different inversion algorithms. The uncertainty assessments are presented in Fig. 14. The results show that the ANN inversion uncertainties are much smaller for this considered model than the Monte Carlo and GWO methods.

Model B

The results of synthetic data inversion for the model with a low-velocity layer (model B) are presented in Fig. 15.

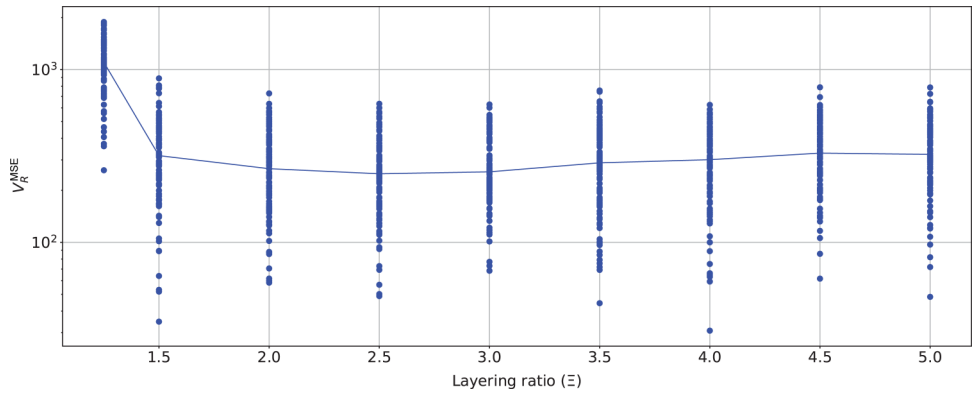

Figure 12 Values of the mean squared error V_R^{MSE} for different values of the layering ratio.

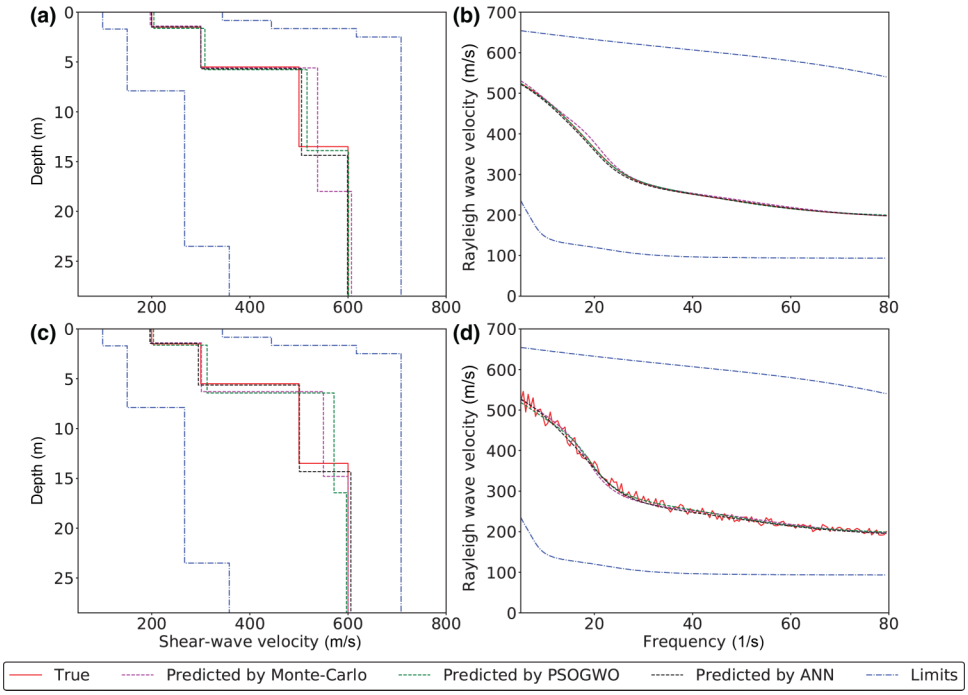

Figure 13 Comparison among the inversion results of model A by the ANN, GWO and Monte Carlo algorithms.

Table 9 Mean absolute errors between the true and restored model parameters for the various inversion algorithms of model A

Inversion Algorithm	V_S^{MAE}(m/s)	h^{MAE}(m)	V_R^{MAE}(m/s)
Noise free data			
ANN	2	0.5	2
Monte Carlo	12	2.3	3.5
GWO	8	0.3	0.8
Noise data			
ANN	4	0.45	7.5
Monte Carlo	14	0.9	7.8
GWO	23	1.73	7.37

The MAEs between the true and restored velocity model parameters (V_S^{MAE} and h^{MAE}) and dispersion curves (V_R^{MAE}) for model B are presented in Table 10. The ANN inversion yields a slightly better match with the estimated true model parameters than the global search method; however, the ANN technique also produces a larger dispersion curve misfit than either Monte Carlo inversion or the GWO algorithm (refer to Fig. 15b). The noisy data inversion results are presented in Fig. 15(c,d). In the case with noisy data, the ANN provides the best match between the estimated parameters and the model

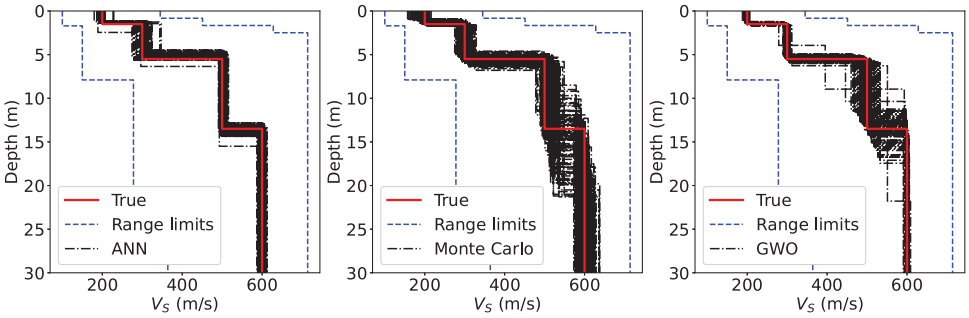

Figure 14 Comparison of the model A synthetic data inversion uncertainties associated with the choice of random inversion parameters for the ANN, GWO and Monte Carlo methods.

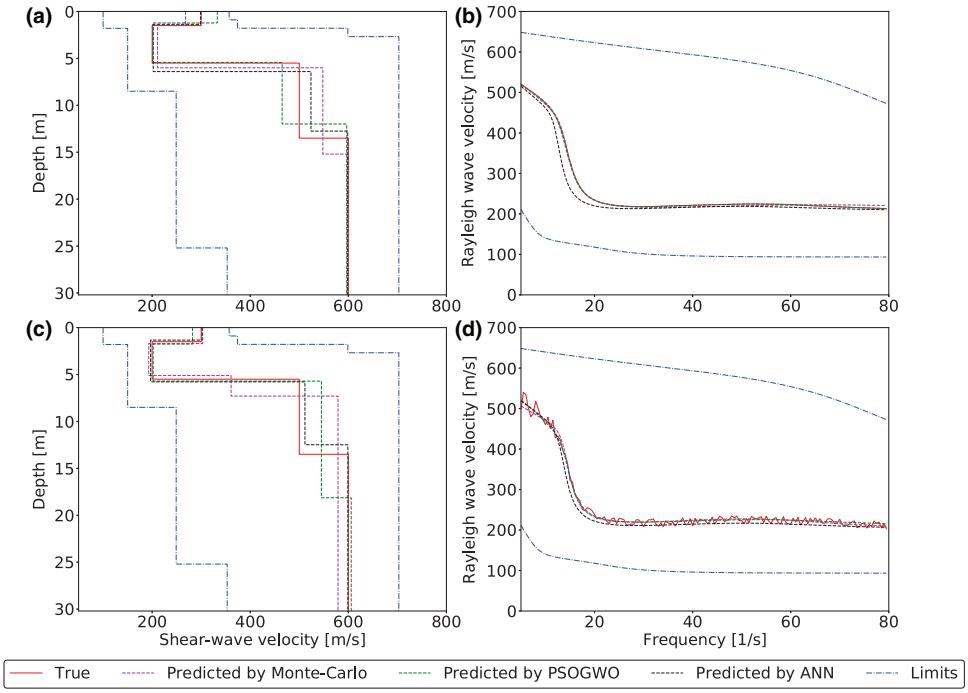

Figure 15 Comparison of the inversion results of model B among the ANN, GWO and Monte Carlo algorithms.

Table 10 Mean absolute errors between the true and restored model parameters for various inversion algorithms of model B

Inversion Algorithm	V_S^{MAE}(m/s)	h^{MAE}(m)	V_R^{MAE}(m/s)
Noise free data			
ANN	7	0.6	10
Monte Carlo	23	0.9	3
GWO	18	0.8	0.54
Noise data			
ANN	4	0.6	11
Monte Carlo	42	3	7
GWO	18	2.4	7

parameters in the presence of noisy data, despite the large misfit in the dispersion curves.

The uncertainty assessments of model B synthetic data inversions, which are similar to those previously given for model A, are presented in Fig. 16. The ANN inversion uncertainties are the smallest.

Model C

The results of synthetic data inversion for the model with a high-velocity layer (model C) are presented in Fig. 17.

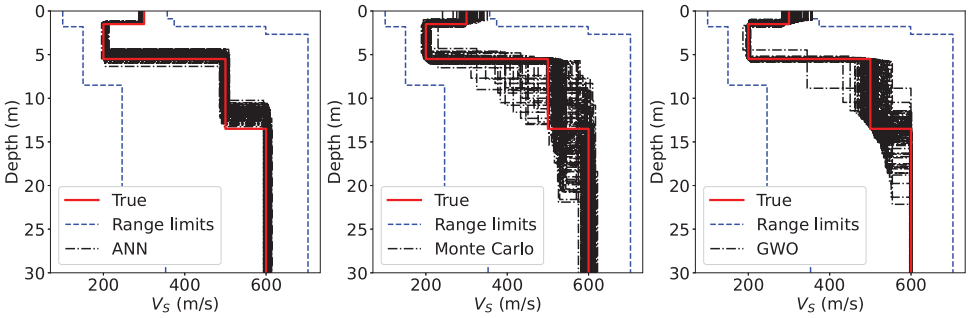

Figure 16 Comparison of the model B synthetic data inversion uncertainties associated with the choice of random parameters among the ANN, GWO and Monte Carlo methods.

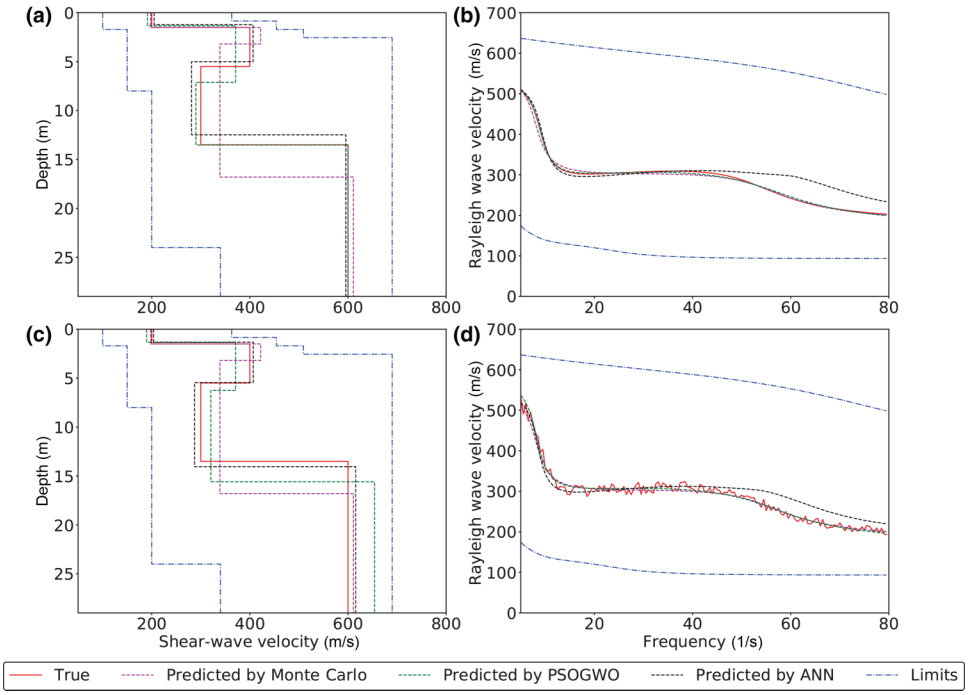

Figure 17 Comparison of the inversion results of model C among the ANN, GWO and Monte Carlo algorithms.

The MAEs between the true and restored velocity model parameters (V_S^{MAE} and h^{MAE}) and dispersion curves (V_R^{MAE}) for model C are presented in Table 11. The ANN provides a better match with the true model than the global search methods despite the relatively large misfit in the dispersion curves. The noisy data-processing results are presented in Fig. 17(c, d). With these noisy data, the ANN inversion results are the most robust.

The uncertainty assessments of the model C synthetic data inversions, which are similar to those previously given for models A and B, are presented in Fig. 18. The ANN inversion uncertainties are the smallest.

Table 11 Mean absolute errors between the true and restored model parameters for various inversion algorithms of model C

Inversion Algorithm	V_S^{MAE}(m/s)	h^{MAE}(m)	V_R^{MAE}(m/s)
Noise free data			
ANN	9	0.7	20
Monte Carlo	18	2.2	4
GWO	12	0.5	2
Noise data			
ANN	10	0.3	19
Monte Carlo	18	2.2	8
GWO	28	1.3	8

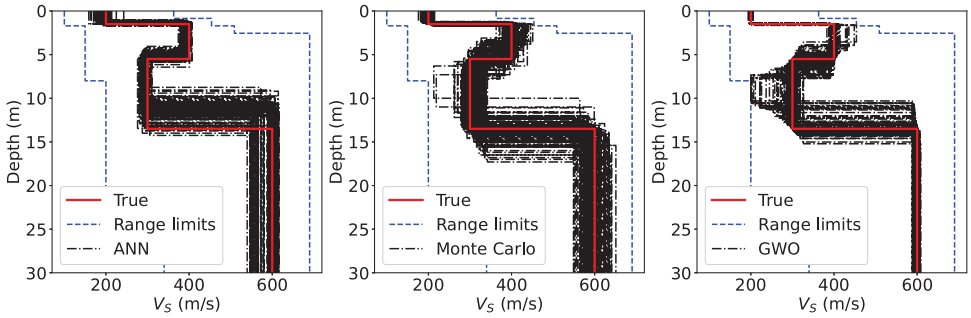

Figure 18 Comparison of the model C synthetic data inversion uncertainties associated with the choice of random inversion parameters among the ANN, GWO and Monte Carlo methods.

Table 12 Computation times of the various inversion algorithms (in minutes) for various numbers of dispersion curves

Amount of the Dispersion Curves	PSOGWO	Monte Carlo	ANN Inversion			
			Total	Calculation	Training	Application
1	2	19	41	19	22	3×10^{-5}
$1 \times 10^{+2}$	147	19	41	19	22	7×10^{-5}
3.2×10^{4}	47,153	58	41	19	22	2×10^{-2}
1.25×10^{6}	1,841,914	2,263	201	90	110	4×10^{-1}

Comparison of the computation time

We compare the computation times of the considered surface wave dispersion data inversion algorithms for various amounts of data: 1, 100 (typical amount of near-surface seismic data), 3.2×10^4 (typical amount of surface seismic exploration data) and 1.25×10^6 (typical amount of four-dimensional time-lapse monitoring data; a similar number of curves are applied in the example from the section 'ANN inversion accuracy test'). We assume that these dispersion data are extracted from 1D seismic profiles placed in the neighbourhood of each other or from one 1D profile at different times (e.g. during time-lapse monitoring). Thus, we assume that the inverted dispersion curves are close to each other and can be inverted by the same ANN, which is trained on only one training dataset.

In the first three examples (Table 12, rows 1–3), the training data consist of 2.5×10^5 dispersion curves, which are computed for the randomly generated four-layer velocity models. In the fourth example (Table 12, row 4), the training data consist of 1.25×10^6 dispersion curves, which are computed for the randomly generated velocity models from two to six layers (for more details, refer to subsection 'ANN inversion accuracy test'). The uniform distributions in the ranges, which are defined from typical (averaged) dispersion curves by the method considered here, are applied to select the layer velocities and thicknesses for the randomly sampled models that compose the training dataset. These dispersion data are also employed for the Monte Carlo inversion method.

The computation times of the GWO, Monte Carlo and ANN inversion algorithms are presented in Table 12. The total computation time of the ANN inversion presented in the fourth column of Table 12 is the sum of three components. The first component, which is denoted 'Calculation' (fifth column of Table 12), is the time spent on calculating the dispersion data for the training dataset. The second component, which is denoted 'Training' (sixth column of Table 12), is the time spent on training the ANN. The third component (last column of Table 12) is the time spent on applying the trained ANN to the inverted dispersion data. Note that 'Application' is the only component that depends on the number of dispersion curves within the inverted data. The 'Application' time is negligible compared to the 'Calculation' and 'Training' times; thus, the total ANN inversion time is almost the same for all considered amounts of data. For a large number of dispersion curves, for example, 1.25×10^6, the ANN is the fastest algorithm.

The Monte Carlo inversion times, which are presented in the third column of Table 12, comprise the sum of the

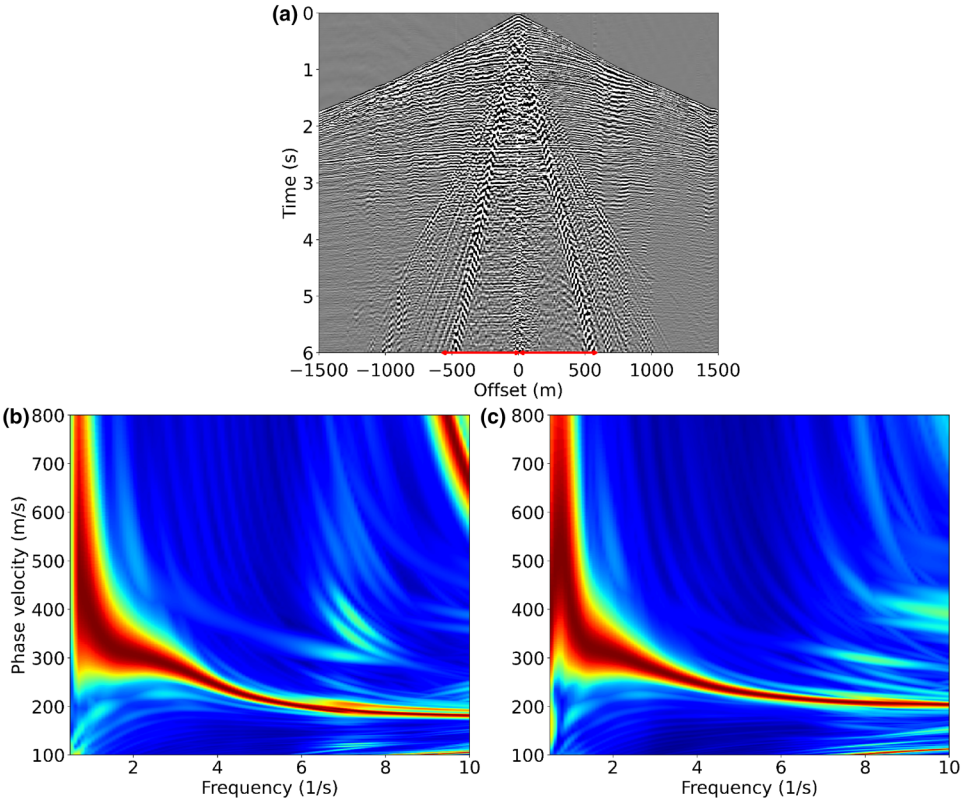

Figure 19 Examples from the dataset: (a) example of the gather; (b) v–f spectrum of the left-side receiver data, offsets (−500, 0) m; and (c) v–f spectrum of the right-side receiver data, offsets (0, 500) m. The MASW line length is marked by a red arrow in the shot gather.

ANN 'Calculation' time (fifth column) and the time needed to compare the given dispersion data with the training dataset. This comparison takes approximately 0.15 s for one inverted dispersion curve. For a relatively small number of dispersion curves (1–100), this comparison does not have a strong effect on the computation time, whereas for 3.2×10^4 inverted dispersion curves this comparison requires more than 70% of the Monte Carlo inversion time.

The GWO computation time, which is smaller than that of the other two methods for the inversion of one dispersion curve, linearly increases with the number of inverted dispersion curves and becomes more than 1000 times larger than the ANN computation time for 3.2×10^4 dispersion curves (refer to Table 12).

WESTERN SIBERIA FIELD DATA PROCESSING

We now present the results from processing data gathered in Khanty-Mansiysk Autonomous Okrug, Western Siberia, Russia. The two-dimensional (2D) common-depth-point seismic-reflection survey is considered. The exploration area is 995 km^2 along the outline of the profiles. The data contain the records of approximately 900 km of linear receiver profiles and are composed of 17,902 shot gathers.

The basic technical indicators and acquisition parameters of the considered seismic data are presented in Table 13.

The study area is in the Western Siberian Plain in a swampy and watered plain. The near-surface section is composed of sandy, sandy-loam and peat-bog subsoils. The elevation map is presented in Fig. 21(a).

The River Konda (Irtysh River tributary) is located in a low valley in the middle of the study area. Along the river, there is sandy and loamy soil. Seismic profiles are marked in Fig. 21(a) by black and red lines.

To process these data by the multichannel analysis of surface waves (MASW) method, each of the common depth point (CDP) lines is transferred to single-ended spread shot gathers. We separately process the records of 24 receivers (or 600 m of the survey line length) for every shot gather. We choose this length as a trade-off between the accuracy of spectral processing and the resulting lateral resolution of the MASW method

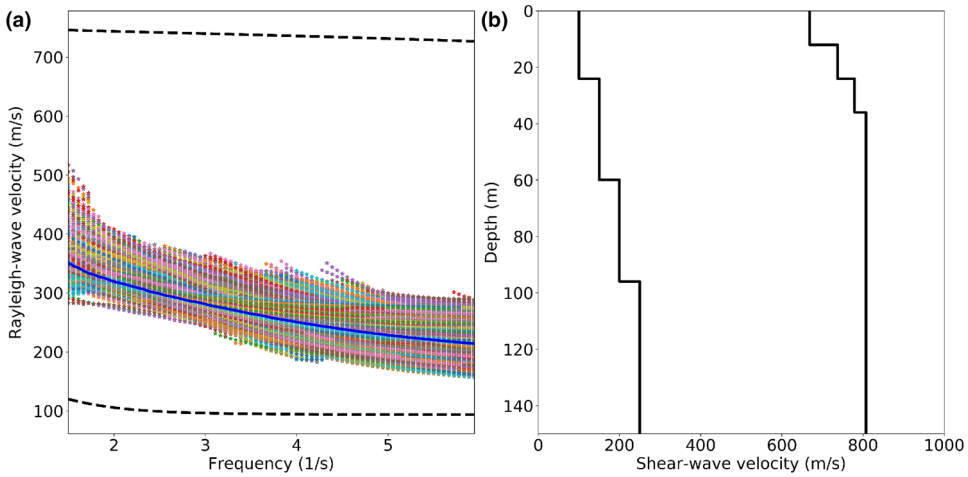

Figure 20 (a) Experimental dispersion curves (diamonds with different colours), average dispersion curve (blue line), the ranges of possible dispersion curves used for ANN training (dotted black lines); and (b) estimated ranges of the possible model parameters.

(Park, 2005). We process both parts of the gathers, namely, the left and right sides, from the source separately.

A necessary step of the MASW is the selection of surface wave phase velocity dispersion curves from the experimental data. The fundamental modes of surface waves can be distinguished from noise and other waves in the frequency–wavenumber (f–k) or frequency–phase velocity (f–v) domain. Note that additional time-domain filtering based on time-frequency representations can also be applied to separate surface waves from noise (Holschneider *et al.*, 2005; Askari and Ferguson, 2012; Ikeda and Tsuji, 2019; Serdyukov *et al.*, 2019). In particular, Serdyukov *et al.* (2019) proposed enhancing the f–k transform based on the S-transform, whereas Serdyukov *et al.* (2019) showed that the slant f–k (SFK) transform is much more robust than the standard f–k transform.

Table 13 Technical methodical indicators and acquisition parameters

Type of Survey Design	2D Seismic Acquisition, Central, Symmetrical
Receiver step, m	25
Stacking fold	60
Record length, s	6
Time step, ms	2
Minimal offset, m	12.5
Maximal offset, m	2987.5
Explosion conditions	Single wells
Source depth, m	16.5
Source step, m	50
charge weight, kg	0.5
Receivers	8-Hz vertical geophones

In this paper, we use the SFK transform to process the field data. An example of the data is presented in Fig. 19(a). The resulting f–v images for both the left side and right side are presented in Fig. 19(b,c).

We automatically select 17,505 phase velocity curves for the left-side receiver data and the same number of curves for the right-side receiver data (the total number of dispersion curves is 35,010). These dispersion curves are plotted in Fig. 20(a). The fundamental mode of the surface wave is observed to range from 1.5 to 6 Hz. An assessment of the signal-to-noise ratio for the considered frequency range is given in the Appendix. The average dispersion curve is plotted in blue in Fig. 20(a). We use the average dispersion curve when estimating the ranges of the possible model parameters by the previously mentioned algorithm. These ranges are shown in Fig. 20(b) by black lines and are utilized when selecting the training dataset for ANN learning. Note that all experimental dispersion curves fall within the defined ranges with a large margin. This result indicates that the considered method of estimation of the model ranges is applicable to both large-area studies and time-lapse studies.

The reconstructed velocity model layer depth interfaces and shear-wave velocity common depth maps are presented in Fig. 21. 2D shear-wave velocity models for inline and crossline profiles, which are marked by red lines in Fig. 21(a–h), are presented in Fig. 21(i,j). In addition, we calculate the synthetic phase velocities using the restored parameters V_s, h and assume a Poisson ratio of 0.35 and a density of 1900 kg/m^3. We compare these synthetic phase curves with the observed phase velocities and calculate the MAE (V_R^{MAE}). The distribution of V_R^{MAE} is presented in Fig. 22. The average mean of V_R^{MAE}

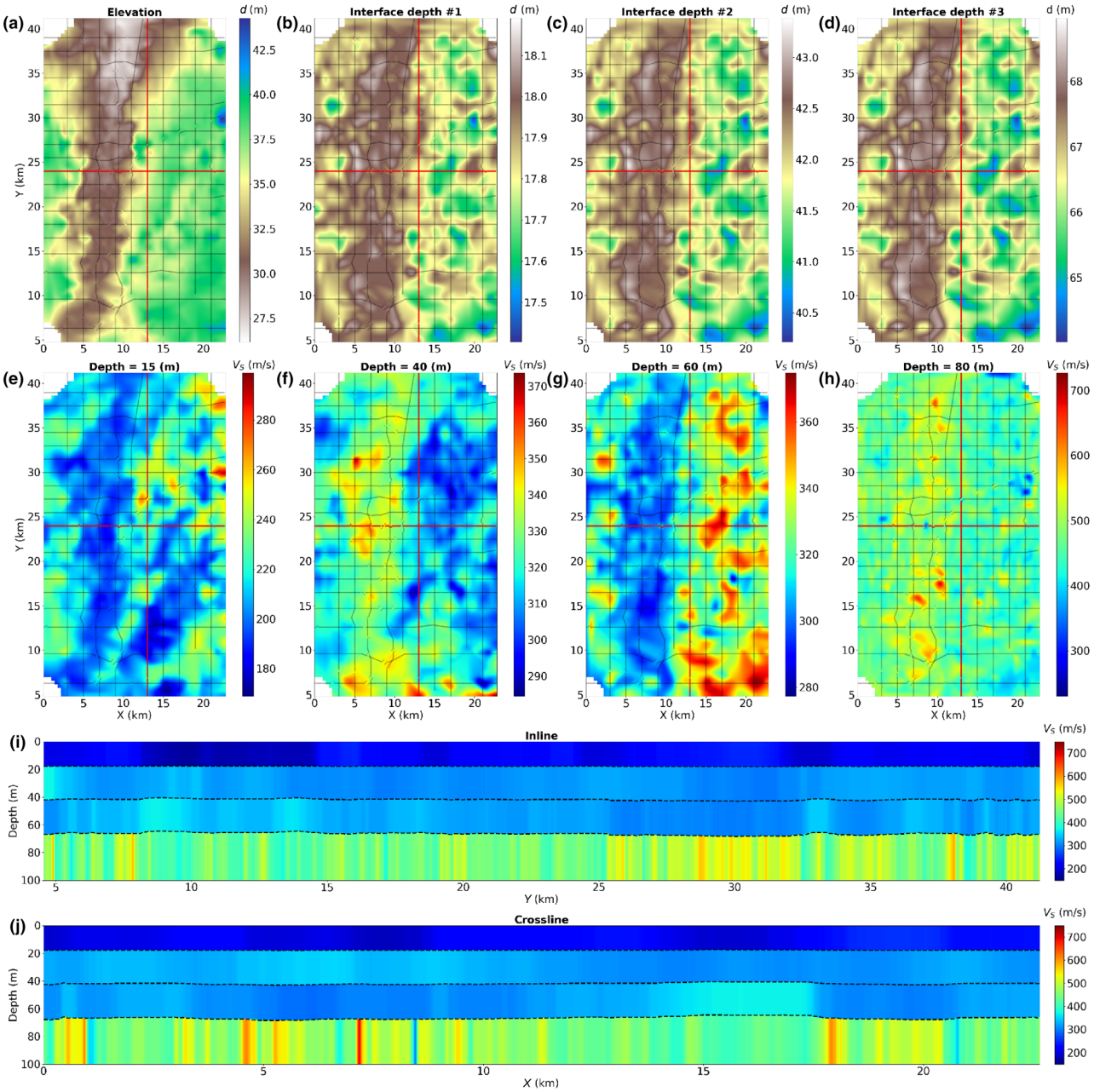

Figure 21 (a) Elevation map; (b)–(d) depth interface map distributions; (e)–(h) shear-wave velocity map distributions; (i) 2D shear-wave velocity model for the inline profile (marked by red line); and (j) 2D shear-wave velocity model for the cross-line profile (marked by red line).

is 8.6 m/s. This misfit among the dispersion data is similar to that of the synthetic data processing results. Therefore, similar accuracies should be expected for the field dispersion data inversion.

The restored shear-wave velocity model is characterized by a flat horizontal layer structure that is composed of four layers. The average depths are 18, 44 and 69 m. The range of variation in the shear-wave velocities is approximately 160–730 m/s to a depth of approximately 70 m for the full domain.

We do not take into account the elevation values during data processing. Nevertheless, the restored model parameters strongly correlate with elevation in terms of both layer

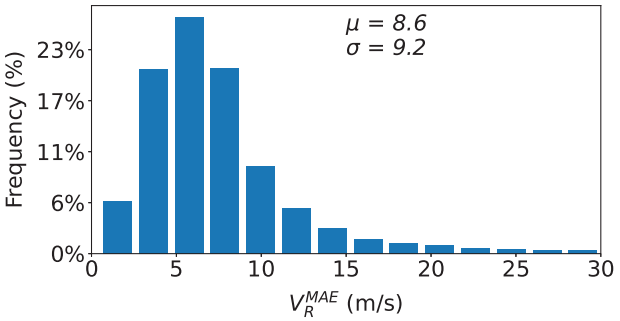

Figure 22 Distribution of the mean absolute errors between the observed and synthetic curves calculated using the restored model parameters.

thickness and velocity. We believe that this finding indicates the feasibility of the obtained results. The study area is located in a sporadic permafrost zone. Based on our research results, we conclude that permafrost has a piecewise-discontinuous characteristic. The presence of permafrost could explain why the velocity values decrease with depth. For example, consider the inverse velocity depth profiles in Fig. 21(j) at a depth of 20–40 m and distance of 5–10 km.

DISCUSSION

Multichannel analysis of surface waves (MASW) is an ill-posed inverse problem, and for this reason the estimation of the uncertainties affecting the recovered parameters is important. An accurate uncertainty appraisal within MASW artificial neural network (ANN) inversion is a complex problem. The ANN inversion seems to be well-suited for uncertainty estimation via Markov Chain Monte Carlo algorithms (see, e.g. Hansen and Cordua, 2017; Aleardi and Salusti, 2021) due to the computational efficiency of ANN. For an accurate uncertainty appraisal using ANN inversion, both data noise and modelling errors introduced by the network approximation must be projected onto the model space. Note that the dispersion data error is not Gaussian and is not purely random in nature. For instance, dispersion image energy flow (due to a spectral leakage) between different modes of the surface waves introduces additional biases to the observed dispersion curves. The presence of lateral heterogeneities not only blurs the image but also can lead to a splitting of the observed dispersion curve into two curves (see, e.g. Hayashi and Suzuki, 2004). These sources of data uncertainties should be properly taken into account. Since the aim of the paper is not to provide accurate uncertainty evaluations, we leave this problem for future research.

Next, we discuss a fine-layer approximation of the velocity model instead of inverting for layer thicknesses. As illustrated by the results presented in the paper, the resulting depths are less accurate than the velocity values, which is due to the different sensitivities of the surface wave phase velocities to the layered velocity model parameters. One can use fine layers to approximate thick layers, and the inverted V_s model would indicate the depths of different layers. From this point of view, the fine-layer approach seems more reasonable. In our previous research, we tried to apply the ANN method to invert dispersion data for a fine-layered V_s model composed of 50 layers with a fixed depth step (Yablokov *et al.*, 2019). We computed synthetic dispersion data for a set of models composed of two to five layers and tried to process it by ANN. We observed that for successful inversion, the training data should include the dispersion curves for the same model type (composed of the same number of layers between two and five); that is the ANN is unable to successfully predict the presence of more layers than are represented in the training data. We tried to increase the number of reconstructed layers but failed to find an architecture that would allow the ANN to be successfully trained on examples containing more than five layers simultaneously. Moreover, our numerical experiments showed that the ANN method presented above, which considers thicknesses as unknown parameters, almost always provides much more accurate results for few-layer velocity models than the ANN fine-layer approach. Actually, for a few-layer velocity model, searching for both layer thicknesses and velocity values leads to a smaller number of unknowns than fine-layer parameterization (e.g. nine unknowns for a five-layered model vs. 50 unknowns for the fine-layered model). The success of ANN training is significantly increased when the number of unknowns is reduced. However, the problem of implementing ANN MASW inversion into the study of near-surface models with a smooth velocity change with depth remains open. Perhaps the fine-layer approach could be stabilized by adding a kind of total variation regularization.

CONCLUSIONS

We propose a new method for the inversion of surface wave dispersion curves based on the application of an artificial neural network (ANN) to the data. The important issue in the proposed inversion method is the selection of a dataset to train the ANN. We suggest an observed dispersion data-driven approach for selecting the synthetic training dataset. First, we follow the strategy of Cox and Teague (2016) and define the possible near-surface parameter ranges for the layered model,

where the ranges of the model parameters depend on the inverted dispersion data (referring to the proposed method for generating the training dataset as a data-driven approach). We propose our rapid algorithm for the adjustment of the layering ratio to determine the optimal parameterization of the velocity model.

Second, we consider uniform distributions over the determined parameter intervals and create a training dataset by generating numerous random layered models. The numerical experiments reveal that a reasonable training dataset consists of approximately 2×10^5 dispersion curves. We accelerate the calculation of the training dataset by parallel computing.

Note that the introduced ranges of the layered model parameters further determine the range of dispersion curve variance. Thus, an ANN that was initially trained for the inversion of a single dispersion curve can be applied to the inversion of numerous dispersion curves that are similar to the initial curve. As a result of this feature, the proposed ANN method is computationally effective at processing large amounts of dispersion data. The processing results presented for the seismic exploration data reveal that the parameter ranges, which are defined for an average dispersion curve, are suitable for ANN training in the inversion of all data obtained from a large area of approximately 800 km^2.

An important aspect of our research is the rigorous adjustment of the ANN architecture and parameters. We conducted a large number of numerical tests, investigated the statistical characteristics of the inversion errors and defined the optimal ANN parameters. We tested the proposed ANN method for inverting surface wave dispersion data on synthetic data and compared the results with Monte Carlo and Grey Wolf optimizer (GWO) dispersion curve inversion methods. The accuracy of the ANN is similar to that of the GWO algorithm and outperforms that of Monte Carlo inversion. Moreover, the ANN inversion is more robust in the presence of noisy data and is dramatically faster than the two global search methods when a large number (10^4 or more) of similar dispersion curves must be processed.

The proposed method is especially promising for MASW time-lapse four-dimensional near-surface monitoring. We believe that the considered ANN inversion method should be adapted to include higher order modes of surface waves to improve the accuracy of inversion (Xia *et al.*, 2003). Another promising approach is to apply the ANN method to the skeletonized inversion of surface-wave dispersion curves, which can invert for lateral velocity variations Zhang *et al.* (2016).

ACKNOWLEDGEMENTS

The algorithmic and methodical parts of the reported study were funded by the Russian Science Foundation (RSF), project number 20-77-10023. The practical part of the reported study (field data processing) was funded by the Russian Foundation for Basic Research (RFBR), project number 19-35-90055. We would like to thank the associate editor of this paper, Clement Kostov, the reviewer Zheng-Dong Zhang, and another anonymous reviewer for their comments and corrections.

DATA AVAILABILITY STATEMENT

The synthetic data are available on request from the corresponding author. The field data are not publicly available due to privacy.

ORCID

Alexandr V. Yablokov
https://orcid.org/0000-0002-3251-0289
Alexandr S. Serdyukov
https://orcid.org/0000-0002-2726-6904
Georgy N. Loginov
https://orcid.org/0000-0002-4906-5986

REFERENCES

Aleardi, M. and Salusti, A. (2021) Elastic pre-stack inversion through discrete cosine transform reparameterization and convolutional neural networks. *Geophysics*, 86, R129–R146.

Askari, R. and Ferguson, R.J. (2012) Dispersion and the dissipative characteristics of surface waves in the generalized S-transform domain. *Geophysics*, 77, V11–V20.

Baeldung. (2020) Understanding forecast accuracy: MAPE, WAPE, WMAPE. Accessed April 28, 2021.https://www.baeldung.com/cs/mape-vs-wape-vs-wmape.

Beaty, K., Schmitt, D. and Sacchi, M. (2002) Simulated annealing inversion of multimode Rayleigh wave dispersion curves for geological structure. *Geophysical Journal International*, 151, 622–631.

Cao, R., Earp, S., de Ridder, S.A., Curtis, A. and Galetti, E. (2020) Near-real-time near-surface 3D seismic velocity and uncertainty models by wavefield gradiometry and neural network inversion of ambient seismic noise. *Geophysics*, 85, KS13–KS27.

Çaylak, Ç. and Kaftan, İ. (2014) Determination of near-surface structures from multi-channel surface wave data using multi-layer perceptron neural network (MLPNN) algorithm. *Acta Geophysica*, 62, 1310–1327.

Cercato, M. (2009) Addressing non-uniqueness in linearized multichannel surface wave inversion. *Geophysical Prospecting*, 57, 27–47.

Cheng, X., Liu, Q., Li, P. and Liu, Y. (2019) Inverting Rayleigh surface wave velocities for crustal thickness in eastern Tibet and the western

Yangtze craton based on deep learning neural networks. *Nonlinear Processes in Geophysics*, 26, 61–71.

Constable, S.C., Parker, R.L. and Constable, C.G. (1987) Occam's inversion: a practical algorithm for generating smooth models from electromagnetic sounding data. *Geophysics*, 52, 289–300.

Cox, B.R. and Teague, D.P. (2016) Layering ratios: a systematic approach to the inversion of surface wave data in the absence of a priori information. *Geophysical Journal International*, 207, 422–438.

Dal Moro, G., Pipan, M. and Gabrielli, P. (2007) Rayleigh wave dispersion curve inversion via genetic algorithms and marginal posterior probability density estimation. *Journal of Applied Geophysics*, 61, 39–55.

Devilee, R., Curtis, A. and Roy-Chowdhury, K. (1999) An efficient, probabilistic neural network approach to solving inverse problems: inverting surface wave velocities for Eurasian crustal thickness. *Journal of Geophysical Research: Solid Earth*, 104, 28841–28857.

Dozat, T. (2016) Incorporating Nesterov momentum into Adam. Workshop track - ICLR 2016. pp. 1–4.

Hansen, T.M. and Cordua, K.S. (2017) Efficient Monte Carlo sampling of inverse problems using a neural network-based forward—applied to GPR crosshole traveltime inversion. *Geophysical Journal International*, 211, 1524–1533.

Haskell, N.A. (1953) The dispersion of surface waves on multilayered media. *Bulletin of the Seismological Society of America*, 43, 17–34.

Hayashi, K. and Suzuki, H. (2004) CMP cross-correlation analysis of multi-channel surface-wave data. *Exploration Geophysics*, 35, 7–13.

Holschneider, M., Diallo, M.S., Kulesh, M., Ohrnberger, M., Lück, E. and Scherbaum, F. (2005) Characterization of dispersive surface waves using continuous wavelet transforms. *Geophysical Journal International*, 163, 463–478.

Hu, J., Qiu, H., Zhang, H. and Ben-Zion, Y. (2020) Using deep learning to derive shear-wave velocity models from surface-wave dispersion data. *Seismological Research Letters*, 91, 1738–1751.

Ikeda, T. and Tsuji, T. (2019) Two-station continuous wavelet transform cross-coherence analysis for surface-wave tomography using active-source seismic data. *Geophysics*, 85, 1–52.

Ikeda, T., Tsuji, T., Nakatsukasa, M., Ban, H., Kato, A., Worth, K., White, D. and Roberts, B. (2018) Imaging and monitoring of the shallow subsurface using spatially windowed surface-wave analysis with a single permanent seismic source. *Geophysics*, 83, EN23–EN38.

Keras. (2020) Keras: the python deep learning API. http://keras.io.

Lai, C.G. (2005) Surface waves in dissipative media: forward and inverse modelling. *Surface Waves in Geomechanics: Direct and Inverse Modelling for Soils and Rocks*. Springer, pp. 73–163.

Mahvelati, S. and Coe, J.T. (2017) The use of two dimensional (2D) multichannel analysis of surface waves (MASW) testing to evaluate the geometry of an unknown bridge foundation. *Geotechnical Frontiers 2017*, 12–15 March 2017, Orlando, FL, USA.

Meier, U., Curtis, A. and Trampert, J. (2007) Global crustal thickness from neural network inversion of surface wave data. *Geophysical Journal International*, 169, 706–722.

Miller, R.D., Xia, J., Park, C.B., Ivanov, J. and Williams, E. (1999) Using MASW to map bedrock in Olathe, Kansas. *SEG Technical Program Expanded Abstracts 1999*. Society of Exploration Geophysicists, pp. 433–436.

Pan, L., Chen, X., Wang, J., Yang, Z. and Zhang, D. (2018) Sensitivity analysis of dispersion curves of Rayleigh waves with fundamental and higher modes. *Geophysical Journal International*, 216, 1276–1303.

Park, C.B. (2005) MASW horizontal resolution in 2D shear-velocity (v_s) mapping. *Open-File Report*, Lawrence: Kansas Geologic Survey.

Park, C.B., Miller, R.D. and Xia, J. (1999) Multichannel analysis of surface waves. *Geophysics*, 64, 800–808.

Pei, D., Louie, J.N. and Pullammanappallil, S.K. (2007) Application of simulated annealing inversion on high-frequency fundamental-mode Rayleigh wave dispersion curves. *Geophysics*, 72, R77–R85.

Prayitna, Y., Utama, W. and Warnana, D.D. (2019) Landslide analysis using seismic refraction tomography and MASW: a case study in Ponorogo, East Java, Indonesia. *IPTEK The Journal of Engineering*, 5, 43–47.

Rahman, M.Z., Kamal, A.M. and Siddiqua, S. (2018) Near-surface shear wave velocity estimation and v_s^{30} mapping for Dhaka City, Bangladesh. *Natural Hazards*, 92, 1687–1715.

Richart, F.E., Hall, J.R. and Woods, R.D. (1970) *Vibrations of Soils and Foundations*. Prentice Hall.

Roohollah, A. (2013) *Surface wave analysis and its application to the calculation of converted wave static corrections*. Ph.D. Thesis, University of Calgary.

Rubaiyn, A., Safani, J. and Priyono, A. (2018) Joint inversion of Rayleigh-wave dispersion curve for near-surface S-Wave velocity estimation. *EAGE-HAGI 1st Asia Pacific Meeting on Near Surface Geoscience and Engineering*.

Serdyukov, A., Yablokov, A., Chernyshov, G. and Azarov, A. (2017) The surface waves-based seismic exploration of soil and ground water. *IOP Conference Series: Earth and Environmental Science*, 53, 012010.

Serdyukov, A.S., Yablokov, A.V., Duchkov, A.A., Azarov, A.A. and Baranov, V.D. (2019) Slant f-k transform of multichannel seismic surface wave data. *Geophysics*, 84, A19–A24.

Socco, L.V. and Boiero, D. (2008) Improved Monte Carlo inversion of surface wave data. *Geophysical Prospecting*, 56, 357–371.

Socco, L.V., Boiero, D., Foti, S., Maraschini, M., Piatti, C., Bergamo, P., Garofalo, F., Pastori, M. and Del Molino, G. (2010a) Surface wave analysis for S-wave static correction computation. *SEG Technical Program Expanded Abstracts 2010*, 1929–1933. Society of Exploration Geophysicists.

Socco, L.V., Foti, S. and Boiero, D. (2010b) Surface-wave analysis for building near-surface velocity models — established approaches and new perspectives. *Geophysics*, 75, 75A83–75A102.

Song, X., Tang, L., Zhao, S., Zhang, X., Li, L., Huang, J. and Cai, W. (2015) Grey wolf optimizer for parameter estimation in surface waves. *Soil Dynamics and Earthquake Engineering*, 75, 147–157.

Strobbia, C., Vermeer, P., Laake, A., Glushchenko, A. and Re, S. (2010) Surface waves: processing, inversion and removal. *First Break*, 28, 85–91.

Thomson, W.T. (1950) Transmission of elastic waves through a stratified solid medium. *Journal of Applied Physics*, 21, 89–93.

Viktrov, I. (1967) Rayleigh and Lamb waves: physical theory and applications. Springer, Ch. II.
White, D. (2009) Monitoring CO_2 storage during EOR at the Weyburn-Midale Field. *The Leading Edge*, 28, 838–842.
Xia, J., Miller, R.D. and Park, C.B. (1999) Estimation of near-surface shear-wave velocity by inversion of Rayleigh waves. *Geophysics*, 64, 691–700.
Xia, J., Miller, R.D., Park, C.B. and Tian, G. (2003) Inversion of high frequency surface waves with fundamental and higher modes. *Journal of Applied Geophysics*, 52, 45–57.
Yablokov, A.V., Loginov, G.N., Serdyukov, A.S. and Duchkov, A.A. (2019) Solving of the inversion problem of the method multichannel analysis of surface waves based on the artificial neural network (in russian). *Interexpo GEO-Siberia*, 2, 191–200. http://geosib.sgugit.ru/wp-content/uploads/2019/sborniki/T2-3.pdf.
Yamanaka, H. and Ishida, H. (1996) Application of genetic algorithms to an inversion of surface-wave dispersion data. *Bulletin of the Seismological Society of America*, 86, 436–444.
Zhang, Z., Schuster, G., Liu, Y., Hanafy, S.M. and Li, J. (2016) Wave equation dispersion inversion using a difference approximation to the dispersion-curve misfit gradient. *Journal of Applied Geophysics*, 133, 9–15.

APPENDIX

ESTIMATION OF THE LAYERED MODEL RANGES

Following Cox and Teague (2016), we estimate the depth range of each layer using the following equations:

$$d^i_{\min} \approx \begin{cases} \lambda_{\min}/3 & \text{for } i = 1 \\ d^{i-1}_{\max} & \text{for } i > 1 \end{cases} \tag{A1}$$

$$d^i_{\max} \approx \begin{cases} \lambda_{\min} & \text{for } i = 1 \\ d^i_{\min} + \Xi \cdot \lambda_{\min}/2 & \text{for } i = 2\,. \\ d^i_{\min} + \Xi \cdot \left(d^{i-1}_{\max} - d^{i-1}_{\min}\right) & \text{for } i > 2 \end{cases} \tag{A2}$$

The maximum depth ($d^i_{\max}$) of the bottom layer should not exceed the vertical resolution (d_{res}), which is determined as one half of the maximum observed wavelength ($\lambda_{\max}/2$). Thus, the iterative search for $d^i_{\min}$ (A1) and $d^i_{\max}$ (A2) is stopped when $i = N$ such that ($d^N_{\max} > d_{res}$); that is, we assume that the last Nth layer is the underlying half-space. After the depth ranges are estimated, we calculate the thickness range of each layer ($h^i, i = 1, \ldots, (N-1)$):

$$\begin{cases} h^i_{\min} = \lambda_{\min}/3 \\ h^i_{\max} = d^i_{\max} - d^i_{\min} \end{cases}. \tag{A3}$$

Note that Cox and Teague (2016) did not provide an algorithm for estimating the range of V^i_S. We use the following approximate dependence of $V_S(\lambda)$ on the Rayleigh fundamental-mode velocity $V_R(\lambda)$ for the same wavelength λ (Viktrov, 1967):

$$V_S(\lambda) \approx 1.16 V_R(\lambda). \tag{A4}$$

Next, we employ the following correspondence between the depth d and the observed wavelength λ, as this relation is often applied in practice during the empirical inversion of Raleigh wave fundamental-mode dispersion curves (Richart *et al.*, 1970):

$$d \approx \lambda/2. \tag{A5}$$

In accordance with (A4) and (A5), V_S can be considered a function of depth ($V_S(d)$). For each ith layer $i < N$, we calculate the mean velocity value:

$$V^{i,\text{mean}}_S = V_S\left(d^i_{\text{mean}}\right), \tag{A6}$$

where the mean depth d^i_{mean} is the half sum of the depth range, which is defined as (A1) and (A2):

$$d^i_{\text{mean}} = \frac{d^i_{\max} + d^i_{\min}}{2}. \tag{A7}$$

The half-space velocity is defined as $V^{N,\text{mean}}_S = 1.16 V_R(\omega_{\min})$, where $\omega_{\min}$ is the minimum frequency in the dispersion data and $V^{N,\text{mean}}_S$ is the last element of the vector defined in equation (A6).

To determine the range of V_S for each layer, we add/subtract the constant velocity of ($100 \cdot \Xi$ m/s) to/from the mean value ($V^{i,\text{mean}}_S$), that is, $V^{i,\min}_S = V^{i,\text{mean}}_S - 100 \cdot \Xi$ m/s and $V^{i,\max}_S = V^{i,\text{mean}}_S + 100 \cdot \Xi$ m/s, respectively. During this process, the V_S of the first layer should not be less than 50 m/s, that is, $V^{1,\text{mean}}_S = \max[V_S(d^1_{\text{mean}}), 50]$. Thus, we obtain the range of the S-wave velocities and thicknesses (A3) of each layer.

ESTIMATION OF THE SIGNAL-TO-NOISE RATIO OF THE FIELD DATA

To evaluate the signal-to-noise ratio (SNR) of the field data, we perform a bandpass filter of the frequency range of 1–6 Hz, which contains a major portion of the utilized surface wave energy. Next, we study the distribution of the ratios of the root mean square (RMS) amplitude between the 'signal' and 'noise'.

The shot gathers are filtered using a fourth-order Butterworth filter, which is shown in Fig. A1(b). An example of a raw common shot gather from the study area is shown in Fig. A1(a), and the result of applying the bandpass filter is shown in Fig. A1(c).

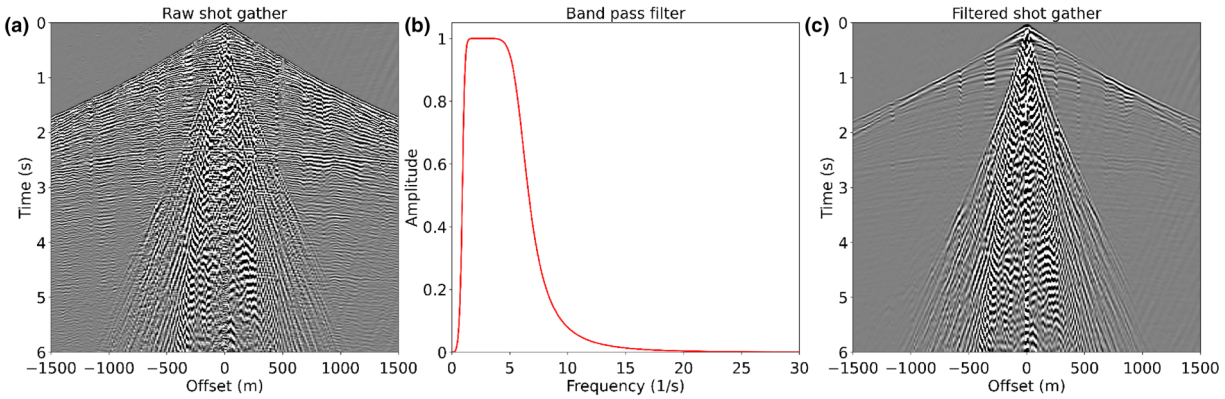

Figure A1 (a) Raw data of the field experiment (common shot gather); (b) fourth-order Butterworth bandpass filter; (c) the common shot gather after filtration.

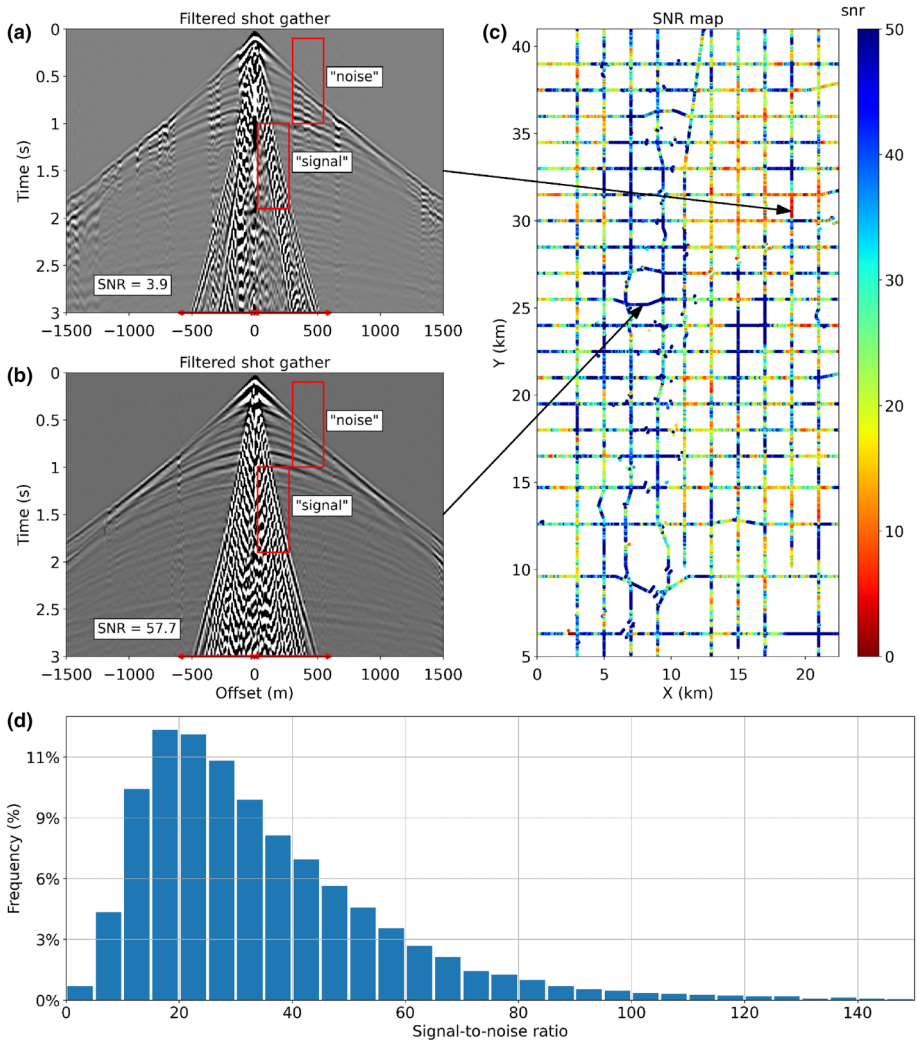

Figure A2 (a) The shot gathers with largest SNR = 57.7; (b) the shot gathers with smallest SNR = 3.9 after applying the fourth-order Butterworth bandpass filter; (c) the SNR distribution map for the study area; (d) the histogram of the distribution of all SNR values.

The analysis windows that correspond to the 'signal' and 'noise' are selected after filtering of shot gathers for the study area. In our study, the 'signal' window in the common shot gather primarily contains the surface-wave energy. The 'noise'' window contains the other types of waves: direct, reflected, refracted and microseisms (refer to red boxes in Fig. A2a,b). To calculate the SNR, the root-mean-square amplitudes in the 'signal' and 'noise' windows are calculated using the following formula:

$$A_{\mathrm{RMS}} = \sqrt{\frac{1}{n \cdot m} \sum_{i=1}^{n} \sum_{j=1}^{m} a_{i,j}^2}, \tag{A8}$$

where n is the number of traces in the analysis window, m is the number of time samples within the window for each trace and $a_{i,j}$ is the value of the amplitude of the jth sample on the ith trace. The SNR value is calculated as $\mathrm{SNR} = A_{\mathrm{RMS}}^{\mathrm{signal}} / A_{\mathrm{RMS}}^{\mathrm{noise}}$. Figure A2(c) shows the SNR distribution map for the entire study area. The SNR distribution histogram presented in Fig. A2(d) indicates that the majority of the data have a high SNR, greater than 10, and only approximately 5% of the gathers have an SNR less than 10. The examples of shot gathers (after applying the fourth-order Butterworth bandpass filter) with the largest SNR = 57.7 and the smallest SNR = 3.9 are given in Fig. A2(a,b).