

Article

A deep-learning-based approach for seismic surface-wave dispersion inversion (SfNet) with application to the Chinese mainland

Feiyi Wang^{1,2}, Xiaodong Song^{1,2,✉} and Mengkui Li³

¹ SinoProbe Laboratory, School of Earth and Space Sciences, Peking University, Beijing 100871, China

² Center of Artificial Intelligence Geosciences, Institute for Artificial Intelligence, Peking University, Beijing 100871, China

³ School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China

Key points:

- We proposed a novel deep-learning method to invert surface-wave dispersions and obtained an updated v_s model of the Chinese continent.
- Synthetic models and flexible sampling strategy were used to build training dataset, which enhances greatly the usability of the method.
- Synthetic tests and real-data application demonstrate that the method is robust and is superior to conventional methods in accuracy.

ABSTRACT

Surface-wave tomography is an important and widely used method for imaging the crust and upper mantle velocity structure of the Earth. In this study, we proposed a deep learning (DL) method based on convolutional neural network (CNN), named SfNet, to derive the v_s model from the Rayleigh wave phase and group velocity dispersion curves. Training a network model usually requires large amount of training datasets, which is labor-intensive and expensive to acquire. Here we relied on synthetics generated automatically from various spline-based v_s models instead of directly using the existing v_s models of an area to build the training dataset, which enhances the generalization of the DL method. In addition, we used a random sampling strategy of the dispersion periods in the training dataset, which alleviates the problem that the real data used must be sampled strictly according to the periods of training dataset. Tests using synthetic data demonstrate that the proposed method is much faster, and the results for the v_s model are more accurate and robust than those of conventional methods. We applied our method to a dataset for the Chinese mainland and obtained a new reference velocity model of the Chinese continent (ChinaVs-DL1.0), which has smaller dispersion misfits than those from the traditional method. The high accuracy and efficiency of our DL approach makes it an important method for v_s model inversions from large amounts of surface-wave dispersion data.

Keywords: deep learning; surface-wave inversion; shear-wave velocity; Chinese mainland



Production and Hosting by Elsevier on behalf of KeAi

✉ **Corresponding author.** Song XD, email: xiao.d.song@gmail.com

Article history:

Received 29 September 2022

Received in revised form 9 November 2022

Accepted 6 December 2022

Available online 15 December 2022

© 2023 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by/4.0>).

<https://doi.org/10.1016/j.eqs.2023.02.007>

Citation: Wang FY, Song XD and Li MK (2023). A deep-learning-based approach for seismic surface-wave dispersion inversion (SfNet) with application to the Chinese mainland. *Earthq Sci* **36**(2): 147–168, doi: 10.1016/j.eqs.2023.02.007.

1. Introduction

Seismic velocity inversion of surface-wave dispersions is a widely used method to image Earth structures (e.g., Shen WS and Ritzwoller 2016; Li MK et al., 2022). However, such an inversion requires a highly non-linear, multi-parameter, and multi-solution geophysical optimization (Nolet et al., 1986; Tarantola, 1987; Mora and Tarantola, 1989; Bassin et al., 2000; Meier et al., 2007). Traditional methods for such inversion can be divided into linearization and non-linearization approaches. Linearization approach is based on Taylor expansion, which linearizes the original non-linear problem and solves it iteratively (e.g., Hauksson and Haase, 1997; Herrmann, 2013). Despite their less computation time, these methods rely on the optimum regularization parameters and initial models. If the initial model differs significantly from the true model, local linearization methods may not be able to determine the global minimum solution of the objective function. In contrast, non-linearization methods are based on sufficient sampling, which removes the dependence on the initial models and avoids the selection of optimal regularization parameters; however, they have high computational costs (e.g., Mosegaard and Tarantola, 1995; Sambridge, 1999; Xu ZJ et al., 2013a; Shen WS et al., 2016).

In recent years, with the deployment of an increasing number of seismic stations and other related instruments, such as the United States Array (IRIS Consortium, 1999; Levander et al., 1999) and the China Array (Ding ZL and Wu ZF, 2013), available observational data has grown exponentially (Ben-Zion et al., 2015). Rapid advancement of deep learning (DL) has introduced new possibilities for solving large-scale geophysical problems (Bergen et al., 2019; Yu SW and Ma JW et al., 2021). DL methods have been successfully used in seismology for earthquake detection and to solve phase-picking problems (Lehuteur et al., 2018; Wang J et al., 2019; Mousavi et al., 2020; Zhou YJ et al., 2022; Jiang C et al., 2021; Zhang L et al., 2020; Li ZF, 2021). Some studies have used DL for surface-wave inversion. Cheng XQ et al. (2019) used deep neural networks to estimate the crustal thickness from Rayleigh waves. Hu J et al. (2020) used a convolutional neural network to derive the v_S model at depths of 0–150 km from Rayleigh-wave dispersions at periods of 8–50 s. Based on U-Net, Fu L et al. (2021) obtained a v_S model from Rayleigh wave phase velocity. Cai A et al. (2022)

inverted the 1D v_S model using the cycle-Generative Adversarial Networks algorithm, which improved the training stability and spatial continuity of the output v_S model. Luo YH et al. (2022) constructed a v_S model from dispersion curves using DL with a training dataset derived from CRUST1.0. Recently, a DL method was used to invert the receiver functions for crustal thickness and v_P/v_S ratio (Wang FY et al., 2022).

In this study, we proposed a convolutional neural network (CNN) method, named SfNet, to invert for the v_S lithosphere structure from surface-wave dispersions. The remainder of this paper is organized as follows. First, we present the methods and data, wherein we present the key details of the design of our DL network SfNet. We also compare the proposed method with search-based inversion, which is a conventional non-linear inversion method and briefly describe the neighborhood algorithm (NA). The quality of inversions was evaluated using input models or data. We used both synthetic and real data in this study to test the network. In the Section 3, we demonstrate the performance of DL method using synthetic data in two cases: (1) comparing the inversion result with that from the NA for a small area and (2) comparing the inversion result with an input model of the entire Chinese continent to further test the stability and applicability of the proposed method. The DL method was then applied to a real dataset, the dispersion curves of which were measured by Li MK et al. (2022), to invert the model for Chinese mainland. In Section 4, we discuss the rationality of the network design (i.e., different structures of the network) and the diversity of the training dataset.

2. Methods and data

2.1. Network architecture

CNN is an important branch of DL and has been widely used to automatically extract deep features embedded in data. A brief introduction to CNN can be found in a recent DL study of seismic receiver functions by Wang FY et al. (2022). Considering the excellent performance of U-Net and ResNet (Ronneberger et al., 2015; He KM et al., 2016), we designed a multilayer 1D CNN model, known as the SfNet (Figure 1a), to invert the v_S model from Rayleigh-wave phase and group velocity dispersions.

In our SfNet, a total of 70×2 nodes were used in the

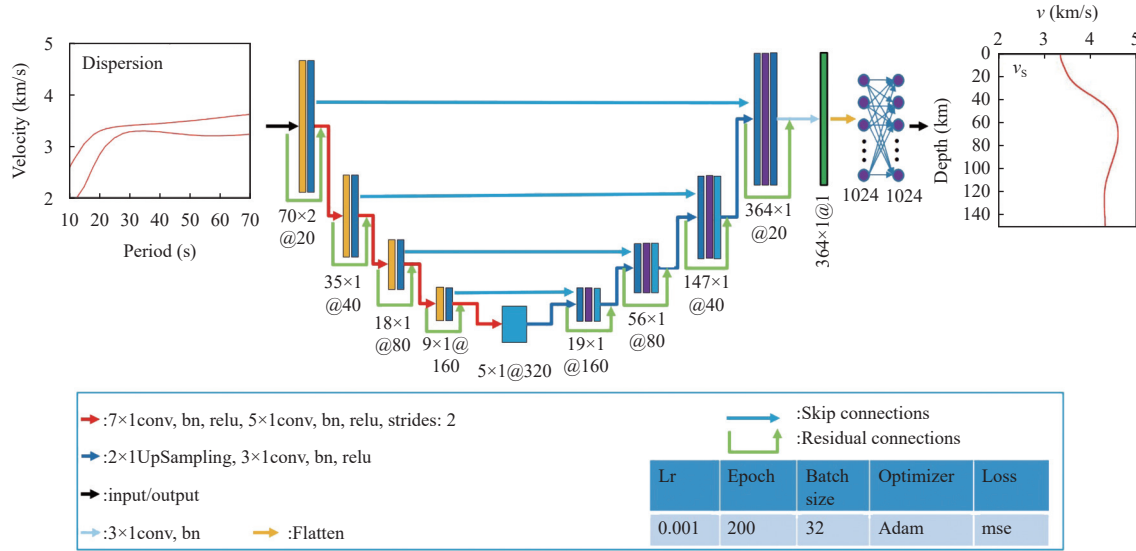


Figure 1. Architecture of SfNet. The symbols and acronyms are conventional to the convolutional neural network (CNN) (Ronneberger et al., 2015; Raschka and Mirjalili, 2017; Wang FY et al. 2022).

input layer, corresponding to the total number of sampling points of the phases and group velocity dispersions in 0–70 s at an evenly spaced interval of 1 s for either dispersion curve. First, we gradually extracted the deep information in the dispersions through a encoder part with convolution and short connections to merge the shallow and deep information. Both global and local features were considered while preventing the gradient from vanishing when the error propagated backward. A two-layer fully connected network was introduced at the end of the CNN to integrate the extracted features and enhance the learning and prediction capabilities. We discretized each 1D velocity model into 300 layers with a 0.5 km thick uniform layer and used the mean-square error to construct a loss function:

$$\text{Loss} = \frac{1}{N} \sum_{i=1}^N (V_{s_i} - V'_{s_i})^2, \quad (1)$$

where V_{s_i} is the true value of the i th layer and V'_{s_i} is the network output. $N = 300$ represents the total number of layers in the 1D velocity model.

2.2. Training and validation datasets

During network training, it is common to establish both the training and validation datasets. The training dataset is the knowledge for the DL network and forms the base for training. The validation dataset was excluded from training and used only for parameter tuning and avoiding overfitting.

A representative training dataset is crucial for DL methods and the trained DL model will have a good

generalization ability (the ability that the method can be used for different cases) only when the training dataset is sufficiently representative. In previous studies, the existing v_s models are usually directly used (Hu J et al., 2020; Cai A et al., 2022; Luo YH et al., 2022) or according to prior information of a specific study area to generated v_s models (Fu L et al., 2021), and then corresponding dispersion curves are calculated to build a training dataset. This way of building the training dataset will limit the generalization of a trained model, especially when applied to real data from new region. In this study, we relied on cubic splines in depth for model parameterization to generate 10,000 v_s models. The training dataset constructed this way is no longer limited to the features of a certain area, but can sample a wider model space, which enhances the generalization of the trained DL model.

The training dataset was generated as follows:

(1) We used two sets of cubic splines to represent the velocity models of the crust and the mantle, respectively. The Moho depth was randomly set within 25–75 km. In the crust, the number of nodes is set from 5 to 11, depending on the depth of Moho. We set uniform node spacing for the crust and the v_s value was set from 2.5 km/s to 4.0 km/s. In the mantle, we fixed the number of nodes to 7 and used a variable stretching parameter to control the node spacing to simulated the gradual or sharp Moho. The v_s value was set from 4.0 km/s to 5.0 km/s. Figure 2 shows the model space of the training dataset.

(2) In each model, the v_p/v_s ratio above the Moho was set to be 1.75, while the v_p/v_s ratio below Moho was set to be 1.8. The v_p/v_s ratio was obtained from the model PREM

(Dziewonski and Anderson, 1981). The effect of Q was not considered in the study of Li MK et al. (2022), and we will compare its results below, therefore, we also did not

$$\rho = 1.6612v_p - 0.4721v_p^2 + 0.0671v_p^3 - 0.0043v_p^4 + 0.000106v_p^5. \quad (2)$$

(3) Theoretical dispersion curves were calculated from the Robert Herrmann's program (Herrmann, 2013), and Gauss noise was then added (with σ set to 0.015 km/s for 5–20 s, 0.02 km/s for 20–35 s, 0.03 km/s for 35–45 s, and 0.05 km/s for 45–70 s). Noise amplitudes from Shen WS et al. (2016) were used to simulate real data.

(4) The period range of dispersion curve we use is 5–70 s. In previous studies of surface-wave inversion based on DL methods, the sampling points of the dispersion curves are fixed (Hu J et al., 2020; Fu L et al., 2021; Cai A et al., 2022; Luo YH et al., 2022). However, in different studies, although the ranges of the dispersion curves period used may be similar, the sampling points of the dispersion curves period may be different. A DL approach has strict requirements on the input format. If the dispersion periods of the data are different from those of the training dataset, the trained model will be unusable. This greatly limits the usability of DL methods.

In this study, we used a random sampling strategy of the dispersion period when calculating the dispersion curves in the training dataset, which alleviates the problem

consider the influence of the Q model in our inversion (i.e. the $1/Q$ values are set to zero). Density was calculated using the Brocher relationship (Eq. 2) (Brocher, 2005).

that the real data used must be sampled strictly according to the periods of training dataset. In the range of 5–40 s of the dispersion curve, the sampling interval is randomly set to 2 or 3 s, in the range of 40–50 s it is randomly set to 2, 3, or 5 s, and in the range of 50–70 s it is set to 5 s. The values of period points (at integer period) that are not sampled are set to zero. Interestingly, we also tried to interpolate the unsampled period points according to the sampled period points before the training, and the result is the virtually the same as setting zero. This is because the source of the information is the same, which is the information from the sampled period points. When the training dataset is large enough, the totality of the random samples contains enough information to uncover the underlying structure, regardless of what happens to the random unsampled points.

(5) We randomly split the entire dataset to 80% for training and 20% for validation.

We conducted 200 training epochs that were validated after each epoch to monitor the training effect. The loss functions during the training process are shown in Figure 3a. We used the open-source libraries TensorFlow and Keras to support all the above-mentioned DL concepts and optimization algorithms. All computations were performed on a desktop PC equipped with an Intel(R) Xeon(R) Gold 6244 CPU and an NVIDIA GeForce RTX 2080Ti GPU with 11 GB memory. The total training time was ~ 3000 s, and the total time required to predict the 1000 1D v_s models using the trained SfNet was ~ 0.8 s, indicating that predicting each model requires < 0.001 s.

2.3. Neighborhood algorithm (NA)

NA (Sambridge, 1999) was used in this study to verify the accuracy and efficiency of SfNet by comparing the results of the synthetic tests. Here, we briefly discuss some

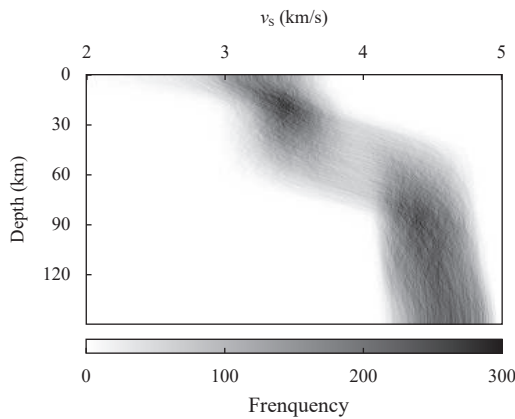


Figure 2. v_s space of the training dataset.

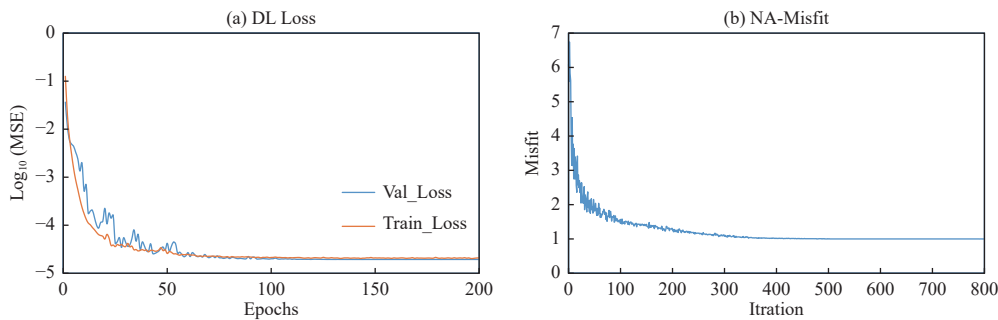


Figure 3. Misfit and loss curves in the two methods. (a) Misfit vs. iteration curve by neighborhood algorithm (NA). (b) Loss vs. epoch curve by SfNet.

of the details of the NA method, which has been widely used in geophysical inversions (e.g., [Snoke and Sambridge, 2002](#); [Yao HJ et al., 2008](#); [Xu ZJ et al., 2013b](#); [Li JT et al., 2017](#)). NA searches the model space to find the “good” data-fitting region using the Voronoi cell, while simultaneously exploring a relatively large variety of different models. Thus, with well-controlled selection of parameters, one can search the complete model space without being trapped in a local minimum ([Xu ZJ et al., 2013b](#)).

Our NA inversion was based on the implementation of [Xu ZJ et al. \(2013b\)](#), who designed an effective joint inversion scheme using NA to determine the v_s model. They used two sets of stretchable splines in the inversion to obtain smooth velocity profiles bounded by a specifically defined Moho (the connecting node of the crustal and mantle splines, the depth of which can be changed as a search parameter). In the first step, we set 15 nodes at 10 km intervals over a depth of 150 km for a sufficient resolution to generate the initial velocity models with a reasonable computational cost. The solution that we seek satisfies the following minimizer in the L1 norm:

$$\min(\|D^{\text{obs}} - D^{\text{pred}}\| + \varphi \|Lm\|), \quad (3)$$

where D^{obs} is the observed dispersion data and D^{pred} is the prediction from a random model generated by NA. The second term is the smoothing constraint. We experimented different φ and choose the one ($\varphi=0.8$) that has best performance. L is the Laplacian operator ([Lees and Crosson, 1989](#)), which requires the second derivative of the model to be zero.

We converted the spline model into 1-km thick homogeneous layers, resulting in a 150-layered model for calculating the theoretical Rayleigh wave group and phase velocities. The velocity of each layer was calculated from the spline coefficients. Similar to [Xu ZJ et al. \(2013b\)](#), we fixed the v_p/v_s ratio at 1.75. Initially, 200 random models were generated, dividing the model space into 200 corresponding Voronoi cells. The five top-ranked (smallest misfits defined in Eq. 2) Voronoi cells were resampled by 100 new models (20 models in each cell), and the process was reiterated 800 times to obtain the final model. The total number of models generated in each inversion was 80200, and the time required for each model was ~ 980 s parallelly using 24 cores. The misfit curve ([Figure 3b](#)) showed a decreasing trend, and the inversion converged after 800 iterations.

2.4. Synthetic data for model tests and real data for model constructions

To test the stability and performance of SfNet, we used v_s models from [Li MK et al. \(2022\)](#) to generate synthetic

data. The period range and interval of the dispersion data is the same as the [Li MK et al.'s](#) data, which will be described below. Li's model is an updated 3D v_s model (on $0.5^\circ \times 0.5^\circ$ grids) for the crust and uppermost mantle of Chinese mainland, obtained from a joint dataset of newly measured Rayleigh-wave dispersions and those previously obtained from ambient noise and earthquakes ([Sun XL et al., 2010](#); [Xu ZJ et al., 2013a](#); [Bao XW et al., 2015](#); [Wang Q et al., 2017](#)).

In addition to synthetic data tests, we also apply our method to real data. We directly applied our method to the real data of [Li MK et al. \(2022\)](#), which allows us to compare the results based on the same data set but different methods. The dispersion measurements of Rayleigh waves from [Li MK et al. \(2022\)](#) include 8–70 s Rayleigh-wave phase velocities and 10–70 s Rayleigh-wave group velocities. The total number of sampling points of phase velocity dispersions is 20 and the sampling points are 8, 10, 12, 15, 18, 20, 22, 25, 28, 30, 32, 35, 38, 40, 45, 50, 55, 60, 65, and 70 s. The sampling points of group velocity dispersions are 10, 12, 15, 18, 20, 22, 25, 28, 30, 32, 35, 38, 40, 42, 45, 50, 55, 60, 65, and 70 s. [Li MK et al. \(2022\)](#) used a linearized inversion method for the v_s model of Chinese mainland ([Bao XW et al., 2013](#); [Xu ZJ et al., 2013a](#)). Details regarding data processing and linearized inversion are provided in [Li MK et al. \(2022\)](#).

2.5. Performance evaluations

For synthetic data, we evaluated the inversion performance in two ways. In the first method, we compared the error of the inversion models with the input models using the root-mean-square (RMS) misfit and the lateral distribution characteristics of all depths using the correlation coefficient (CC). The input model is used to generate the input dispersion data during inversion, so it can be regarded as the correct answer for the inversion result. In the second method, we compared the dispersion curves of the inversion models with those of the input model. For real data, we evaluated the inversion models by comparing the results with those of [Li MK et al. \(2022\)](#) and analyzing the misfit in their dispersion curves.

The RMS misfit of inversion models with the input models is defined as:

$$E_{\text{RMS}} = \left[\frac{1}{N} \frac{1}{M} \sum_{i=1}^N \sum_{j=1}^M \left(v_{i,j}^{\text{inv}} - v_{i,j}^{\text{input}} \right)^2 \right]^{1/2}, \quad (4)$$

where $v_{i,j}^{\text{inv}}$ is the velocity of the M th layer in the N th inversion model and $v_{i,j}^{\text{input}}$ is the velocity of the M th layer in the N th input model. $M = 300$ denotes the number of layers in the model. N is the number of models.

The CC between the two models at a given depth is defined as:

$$CC = \frac{\sum [(x_i - \bar{x}) \times (y_i - \bar{y})]}{\sqrt{\sum (x_i - \bar{x})^2 \times \sum (y_i - \bar{y})^2}}, \quad (5)$$

where x and y are the velocities of the two models and i is the index for a grid point common to both models. CC was used to verify the similarity between the perturbation patterns of each model pair. It should be noted that CC only indicates the similarity in the patterns of velocity change at a certain depth and is not related the absolute amplitude of the change.

At each location, the dispersion misfit is defined as:

$$\text{Dispersion misfit} = \left[\frac{1}{N} \sum_{i=1}^N (d_i^{\text{obs}} - d_i^{\text{cal}})^2 \right]^{1/2}, \quad (6)$$

where d_i^{obs} is the input or real measurement dispersion and d_i^{cal} is the dispersion calculated from the inversion model. N is the number of periods sampling points of the dispersion curve.

3. Results

3.1. Synthetic tests

We conducted two synthetic tests for our proposed DL method on small and large regions (Figure 4) to examine the accuracy and stability of the method. The SfNet results were compared with those of the conventional search-based NA method. As the NA method is computationally time-consuming, we selected a relatively small region of Sichuan-Yunnan (SY) in southwestern China. The second synthetic test was conducted on the entire Chinese

mainland. For both tests, the v_s models of Li MK et al. (2022) were used as input models to generate synthetic Rayleigh wave dispersion curves. We also added noise to the synthetic data. We then used our DL method (and the NA method for the SY region) to invert the synthetic data to obtain the inversion models and compared them with the input models.

A common problem of DL methods is that the uncertainty of their results is difficult to estimate (Hu J et al., 2020). However, after adding noise to the dispersion data, the average error from several thousand inversion tests could be used to estimate the uncertainty of the DL method.

3.1.1 Comparison of DL and NA method: Test for the Sichuan-Yunnan region

The SY region is located on the southeastern margin of the Tibetan Plateau (Figure 4), and it is one of the most seismically active regions in the Chinese mainland. Its special tectonic location, frequent seismic activity, and complex relationships between earthquakes and geological structures make the SY region an important area for seismological studies (Gkarlaoui et al., 2008; Wen XZ et al., 2013; Wang F et al., 2013; Bao XW et al., 2015; Xu ZJ et al., 2013b; Chang LJ et al., 2015; She YY et al., 2018).

A total of 165 pairs of Rayleigh-wave phase and group velocity dispersion curves were calculated from the SY region and added with Gaussian noise (with different levels for different periods, Section 2.2) to generate synthetic data for inversion. Figure 5 shows a comparison of the results obtained using NA and SfNet with the input v_s model at a depth of 100 km. Comparisons at depth of 50 km is shown in Figure S1.

The NA and SfNet results showed similar CC and RMS misfit with the input v_s at all the depth. However, the CC and RMS of SfNet are more stable than those of the NA model in the deep part. Figure 6 shows the CC and RMS misfit of the two inversion models compared to the input v_s for the depth of 0–150 km, indicating that SfNet performed better than the NA model for both the average CC and RMS misfit. Overall, SfNet has a comparable CC relative to the NA model but reduced the RMS misfit by 30%. Notably, in case of poor constraints of the NA inversion on the deeper velocity structure, we can introduce prior constraints on the velocity to improve the inversion results. However, the phase and group velocities of 10–70 s are sufficient to constrain the model at 0–150 km depth (Shen WS et al., 2016; Hu J et al., 2020) and for a fair comparison with SfNet, we did not add any prior constraint in the NA inversion. See the Supplementary information for examples of the individual nodes marked

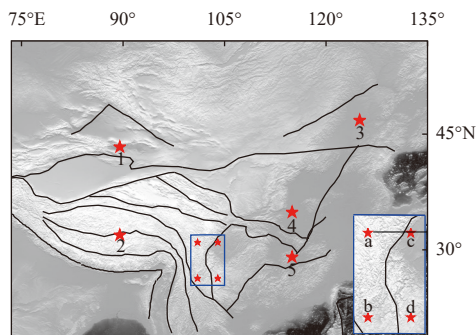


Figure 4. Map of the locations for the study. We tested the entire Chinese mainland (inside the border) and the smaller Sichuan-Yunnan (SY) region in China (blue box). Red stars mark the selected points that are used to show the inversion results in Supplementary information (Figures S2–S10). The black lines represent the boundaries of geological units.

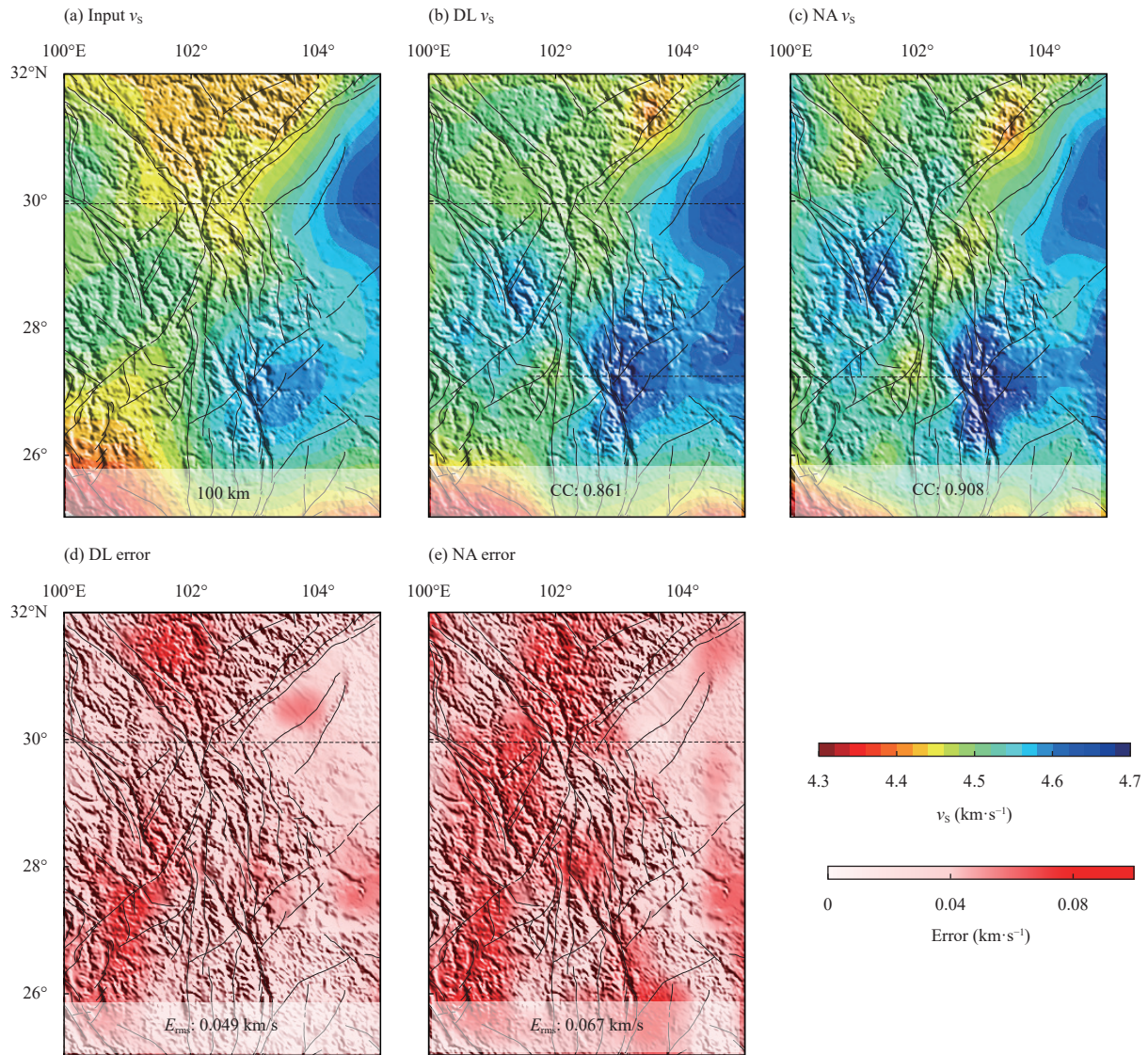


Figure 5. Comparison of result of our DL method and NA for the SY region at depth of 100 km. (a) Input v_s model. (b) Inverted v_s model from DL method. (c) Inverted v_s model from NA. (d) Absolute difference between the inverted v_s model from DL method and input models. (e) Absolute difference between the inverted v_s model from NA and input models. The CC values between the input and inverted models and the RMS errors are also labeled, respectively. The black lines represent the faults.

by red stars in Figure 4 (Figures S2–S5).

It is interesting to note that despite a much larger number of models generated in each NA inversion (80, 200) than that in the training dataset (10, 000), the DL method achieves a better result. The essence of DL is to find the complex mapping between dispersion curves and velocity model. When the given training dataset can meet the network's learning requirements, the network can invert those unseen dispersion curves and obtain a credible velocity model. For the NA method, its essence is to find the current optimal solution through simple trial and error. In other words, the NA method cannot make “prediction”. Even if the result is very close to the true value, it has to

“try” to get a more accurate result, so a large number of generative models are required. But deep learning can be done in one step after mastering the mapping relationship.

3.1.2 Synthetic test for Chinese mainland

To further test the applicability and stability of the proposed DL method, we applied it to the entire Chinese mainland. Random noise was added to the synthetic dispersion curves at different levels for different periods (Section 2.2). Figure 7 shows the results of v_s models at 50 km and the CC and RMS misfit compared to the input models in the range of 0–150 km.

Results showed that the v_s model obtained by the DL method at various depths corresponded well to the input

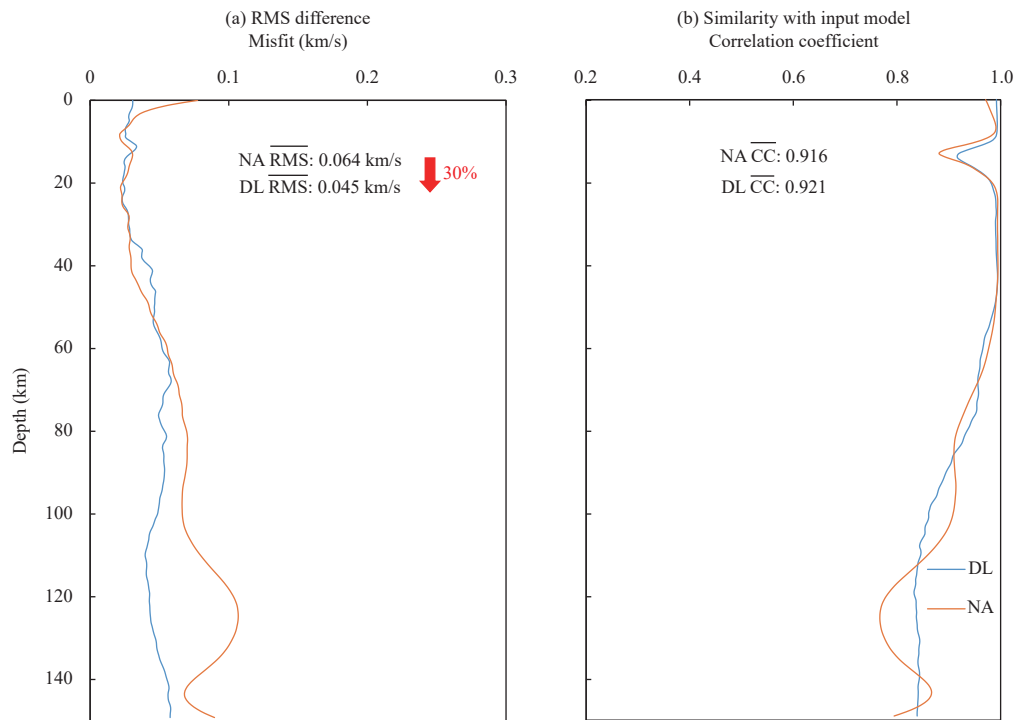


Figure 6. Comparison of the SfNet and NA results. (a) CC with the input v_s model and (b) RMS misfit for the depth of 0–150 km. The average CC and the RMS misfit are also provided. Red arrow indicates that reduces RMS misfit by 30% over the NA method.

model, with an average CC of 0.968 and RMS misfit of 0.042 km/s. Similar to the results for the SY region (CC: 0.921, RMS misfit: 0.045 km/s), our method is also robust throughout Chinese mainland. See the Supplementary information for examples of individual nodes marked by red stars in Figure 4 (Figures S6–S10). At 60–150 km, the inversion results for CC were poor. However, the RMS misfit was relative stable, which suggests that although the method works well, the complexity of the geological structure and the lateral changes were relatively drastic in the Chinese mainland, resulting in the instability of CC.

Overall, the tests for the Chinese mainland showed that our proposed method has good stability and adaptability. The convincing v_s results indicate that SfNet deduced a complex mapping relationship between the dispersion curves and the velocity structure, rather than simple data fitting.

3.2. Application to real data

Systematic synthetic tests (previous section) demonstrated the effectiveness of the proposed method. In this section, we describe the application of the proposed DL method to real data. We applied the method to the Rayleigh-wave phase velocities in 8–70 s and group velocities in 10–70 s from Li MK et al. (2022) and obtained a 3D v_s model (named as ChinaVs-DL1.0).

Figures 8 and 9 show some examples of the model and the model misfits compare the newly inverted model with the model by Li MK et al. (2022) using a conventional method.

Figure 8 shows that both models corresponded well to the geological structures at depth of 50 km. For example, the Tibetan Plateau exhibits an obvious low-velocity anomaly, whereas the stable North China Craton shows a high-velocity feature (Xu PF and Zhao DP et al., 2009; Bao XW et al., 2013). In comparison, the differences between the two models in eastern China were small. However, the difference is larger in the west, where the data coverage is poorer (Li MK et al., 2022) and the deep structure is more complex from the collision between the Indian and Eurasian plates. The similarity between the two models gradually decreased with depth, and the RMS difference increased. In the shallow and intermediate depths (< 60 km), the difference between the two models was small, reflecting the reliability of the DL model; however, in the deep layer (> 60 km), the RMS difference was large and the CC was small. The reason for this difference may be that the surface-wave sensitivity to the structure is smaller in the deeper part, which causes a multi-solution problem. However, through previous tests on synthetic data, it is proved that our method is relatively reliable in the deep (Figure 6 and Figure 7). Refer to the

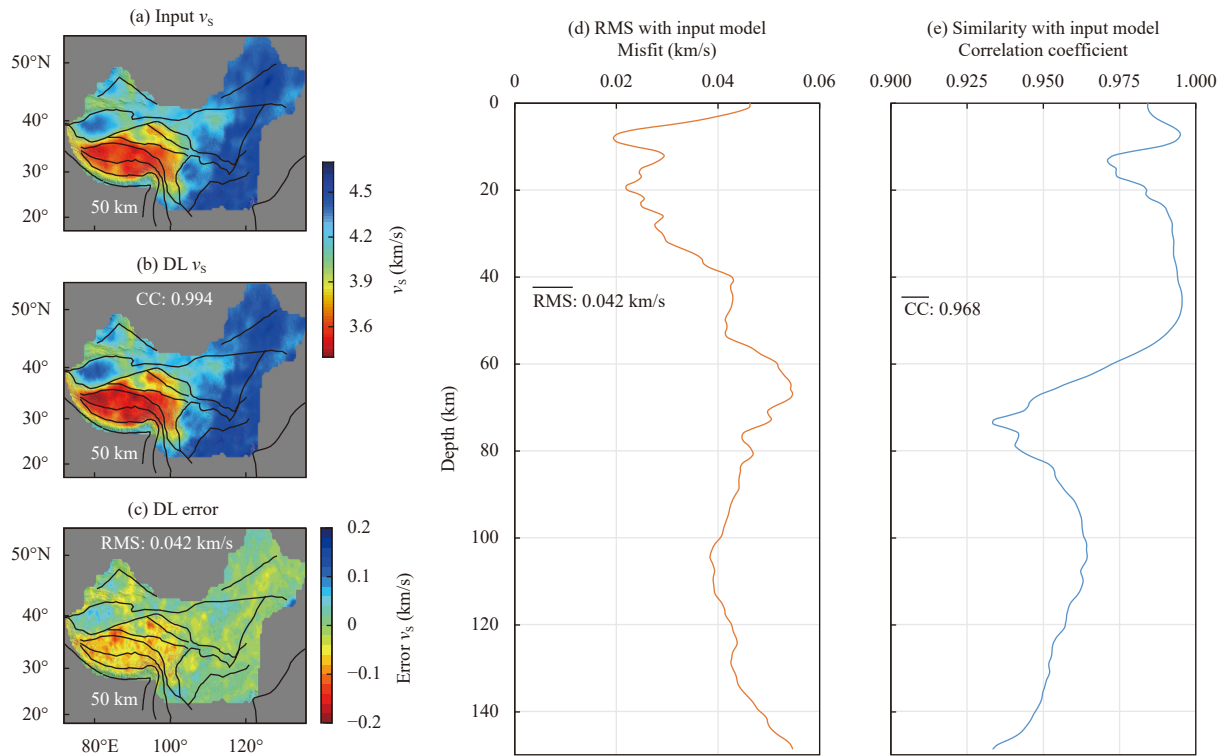


Figure 7. Comparison of the v_s models obtained from the proposed DL method with the input models. (a) Input v_s model at 50 km, (b) inverted v_s model at 50 km, (c) difference between the inverted v_s model and the input model at 50 km, CC and RMS misfit are also provided. (d) RMS misfit at 0–150 km and (e) CC. Average CC and RMS misfit are also provided.

Supplementary information for examples of individual nodes marked by red stars in Figure 4 (Figures S11–S15).

Figure 9 shows the distributions of the misfits between the dispersions predicted by the two models. Misfits of the proposed DL model were smaller in most regions in the Chinese mainland, and significantly lower than those of Li MK et al. (2022) (Figure 9c and 9f). The phase velocity misfit and group velocity reduced by 24% and 23%, respectively. The dispersions curves misfit reduction showed that our proposed method is more reliable.

4. Discussion

We applied our method to the synthetic data and real data and obtained reliable results, which demonstrate that the SfNet has the potential for broad applications without retraining the network, because the Chinese continent is known to have highly complex crustal structure (e.g., Li MK et al., 2022; Wang FY et al., 2022). Two important factors affect the performance of the network. First, the structure of the network determines how well the network can learn the complex mapping between the dispersion curve and the v_s model. Second, the diversity of training dataset determines the generalization of the method. Two additional experiments were performed to further discuss

the proposed method.

4.1. Design of the SfNet architecture

The design of a DL architecture is often based on experience, as it often involves several choices, and the differences are unclear unless each choice is tested out, which is neither possible nor necessary. A structure that is too simple will make it difficult for the network to learn the intrinsic relationship of the data, and a structure that is too complex will make the network overfit. To evaluate the choices of the SfNet structure, we changed some of its components and compared the results using synthetic data. The tested parameters included the size of the receptive field of the convolution kernels in the decoder (ks), the number of convolution kernels set at the first layer (nk) and number of neurons in the first fully connected layer (n1), and that in the second fully connected layer (n2). The training dataset was the same as that described in Section 2.2. Table 1 lists the performances of the different structures.

It can be seen from the table that for different network structures, v_s misfits are small (<0.1 km/s) and CC is close to 1 (>0.94), which shows that the robustness of our method. However, the performance can still differ somewhat and we tried to fine tune the structure to find more optimal solution. The structure used in this study (#5

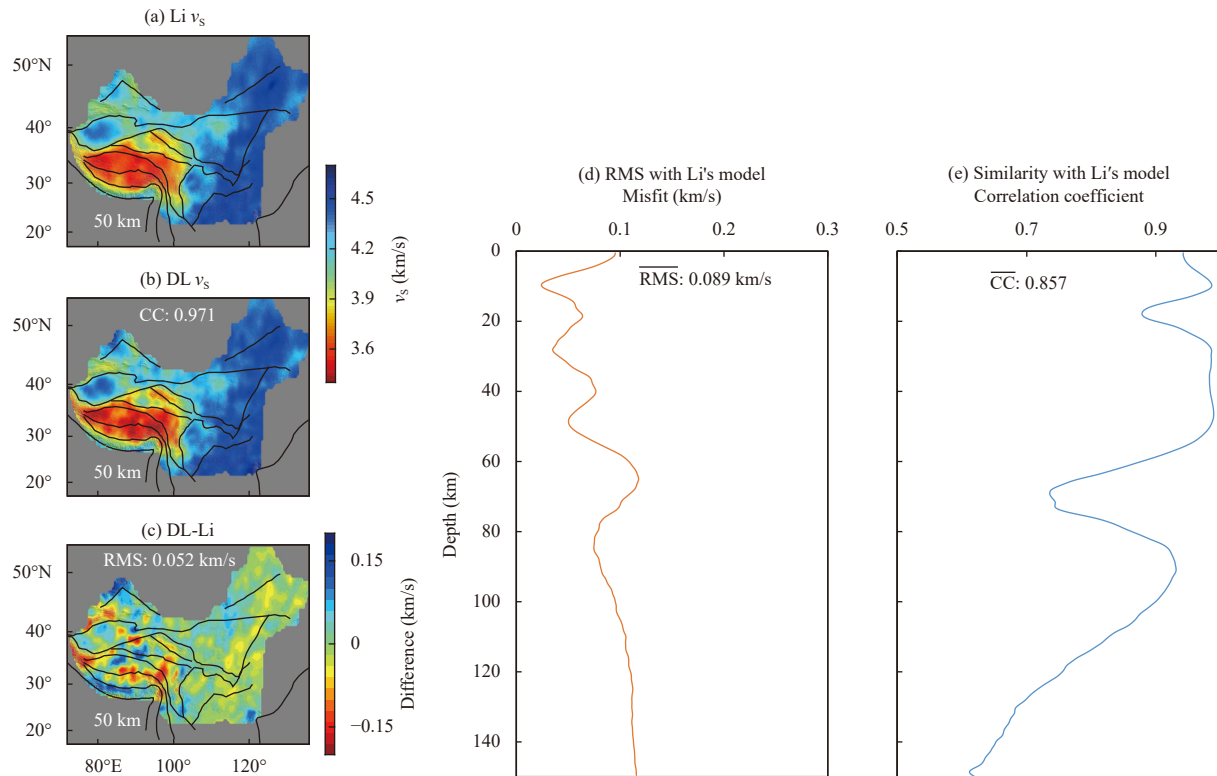


Figure 8. Comparison of the v_s model obtained from DL (ChinaVs-DL1.0) with the model of Li MK et al. (2022). (a) v_s model of Li MK et al. (2022) at 50 km; (b) DL v_s model at 50 km, and (c) difference between the v_s models of the DL method and Li MK et al. (2022) at 50 km. CC and RMS difference are also provided. (d) CC and (e) RMS difference at 0–150 km. Average CC and RMS misfit are also provided.

in Table 1) performed relatively better on the all metrics. Tests #1 and #2 show that the size of the receptive field does not have a significant impact on the performance of the network. Tests #3, #4 and #5 show that the number of neurons in the fully connected layer significantly effects the performance of network while the training time is longer as the number increases. However, the extra time is completely negligible compared to the time taken by traditional inversion methods and the huge amount of data processing. Tests #5 and #6 show that as the complexity of the network increases, the performance begins to decline, which means the network appears over-fitting. The network structure in Test #5 may be not the optimal, but considering that the results of both synthetic data and real data can already meet the requirements, no further tests are performed.

4.2. The diversity of the training dataset

The diversity of the training dataset is critical to the performance of DL methods. If the training dataset does not contain potential structures for the real dataset, the DL methods may perform worse than the conventional inversion methods (Hu J et al., 2020). In contrast to previous studies, because we relied on spline-based

method to build training dataset, the number of training dataset represents diverse sample space of v_s models. Small training dataset will make it difficult for the network to learn the mapping relationship between dispersion curves and v_s models. However, if the training dataset is too large, it will unnecessarily cost computing resources. We tested the network by training with 5000, 10000, and 20000 training datasets, respectively, and evaluated its performance like the Section 4.1. Table 2 lists the performances of the different training datasets.

Table 2 shows that compared with using 10000 or 20000 training datasets for training the network, using only 5000 training dataset yielded much worse results, which means the 5000 training dataset is significantly less diverse. Training with 10000 and 20000 dataset has no effects on the results, indicating that the 10000 training dataset is enough for diversity. Furthermore, owing to certain typical velocity structures shared by different regions (Stankiewicz et al., 2010; Lekic et al., 2012; Braeuer and Bauer, 2015; Eymold and Jordan, 2019) and the geological structures beneath China are very complex (Huang JL and Zhao DP, 2006; Teng JW et al., 2013; Chen B et al., 2013; Jiang WL et al., 2020; Li MK et al., 2022), we believe that our trained DL model, which obtained reliable results for

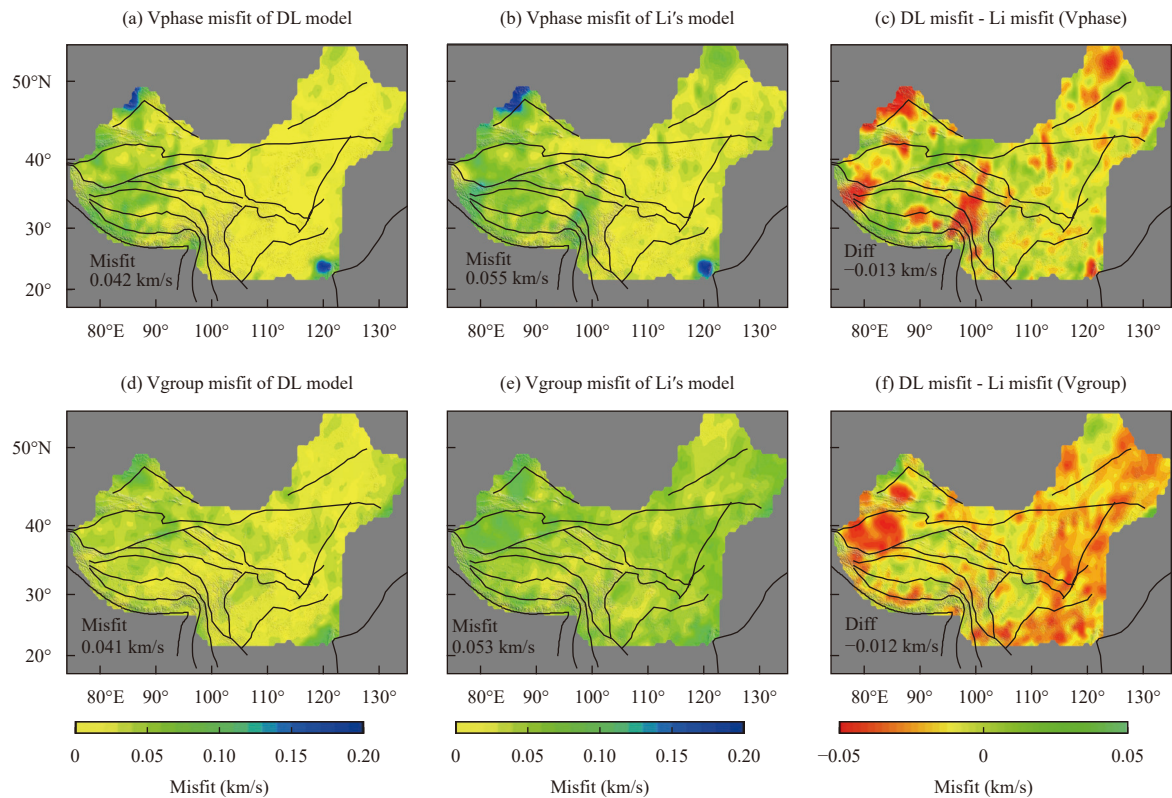


Figure 9. Misfit maps between the dispersions predicted by the inversion and real data. (a) Misfit maps of Rayleigh-wave phase velocity model from the ChinaVs-DL1.0 model and (b) Li MK et al. (2022); (c) difference of two misfits; (d–f) same as (a–c) but for group velocity. Average misfits for the two models and the difference between them are also provided.

Table 1. Performances of different structures.

#	Description	Training time (s)	v_s misfit (km/s)	CC
1	ks: 3; nk: 16; n1: 512; n2: 512	3 702	0.090	0.943
2	ks: 5; nk: 16; n1: 512; n2: 512	3 815	0.090	0.960
3	ks: 3; nk: 20; n1: 512; n2: 512	4 023	0.090	0.944
4	ks: 3; nk: 20; n1: 512; n2: 1 024	4 352	0.055	0.966
5	ks: 3; nk: 20; n1: 1 024; n2: 1 024	4 418	0.042	0.968
6	ks: 3; nk: 20; n1: 2 048; n2: 1 024	7 402	0.072	0.948

Note: The optimal value of each indicator is in bold font.

the entire Chinese mainland, has a greater generalization capability.

4.3. Applicability of the SfNet

Note the method is not applicable if the periods of the data are outside the range of 5–70 s or the assumptions (e.g., v_p/v_s ratio, density, Q model) are different (Kaypak,

2008; Jo E and Hong TK, 2013). However, the network can be easily re-trained, for which practical strategies were proposed. There are two basic cases in which SfNet is used for surface-wave tomographic inversion. First, use Rayleigh-wave phase and group velocity dispersion data at 5–70 s (as in this study) and input directly into SfNet to obtain velocity models for 0–150 km. The trained SfNet

Table 2. Performances of different training datasets.

#	The number of training dataset	Training time (s)	v_s misfit (km/s)	CC
1	5 000	2 097	0.111	0.873
2	10 000	4 418	0.042	0.968
3	20 000	8 605	0.043	0.963

Note: The optimal value of each indicator is in bold font.

and the training dataset have been made publicly available (at Zenodo, see below). Second, if the measured dispersion data are in other period ranges, one needs to generate the training dataset (Section 2.2) and train it first before applying it to the real data, e.g., for 5–35 s and 10–140 s, respectively. We used the Robert Herrmann's program (Herrmann, 2013) to calculate the Rayleigh-wave phase and group velocity dispersion curves at period of 5–35 and 10–140 s to generate training datasets, respectively. We further added Gaussian noise to the calculated theoretical dispersion curves (with standard errors σ set to 0.015 km/s for 5–20 s, 0.02 km/s for 20–35 s, 0.03 km/s for 35–45 s, 0.05 km/s for 45–70 s, 0.08 km/s for 70–100 s, 0.10 km/s for 100–120 s, and 0.12 km/s for 120–140 s).

As surface-wave dispersions for 5–35-s periods were sensitive to shallower depths, we only used them to invert the velocity model at 0–80 km. For dispersion data of 10–140 s, we inverted for models at 0–300 km. We also used the v_s models of the Chinese mainland from Li MK et al. (2022) for testing. Results of the inversion for 5–35 and 10–140 s dispersion data are shown in Supplementary information Figure S16. The corresponding network structure (input and output nodes) was fine-tuned according to the changes in the dispersion data and depth range of the velocity structure, and the datasets and corresponding SfNet were also uploaded to Zenodo.

Results for the period range of 5–35 s and 10–140 s (Figure S16) showed that the proposed DL method performed well. The results of v_s were relatively stable, and the error increased slightly with depth. The RMS misfits of v_s for two periods range (0.03 km/s and 0.05 km/s, respectively) were comparable to those of the previous tests (0.04 km/s) in the period range (Figure 7).

The two examples suggest that SfNet can be easily retrained and adapted for other surface wave dispersion inversions. Similarly, if we only have group or phase velocity dispersion curves, we can retrain the SfNet before application. Once trained, it can be used for other

applications with dispersion curves within the same period range.

5. Conclusions

In this study, we proposed a new DL-based method to invert the v_s models from Rayleigh wave phase and group velocity dispersion curves. Utilizing the powerful data-mining ability of CNN, SfNet could quickly and efficiently extract the complex mapping relationship between surface-wave dispersions and velocity structures. Finally, we also provided an updated Chinese mainland regional v_s model (ChinaVs-DL1.0).

Compared with traditional methods, the proposed method was faster and more accurate, avoiding the selection of initial models and empirically determined inversion parameters. With the increasing availability of seismic records, the processing of large amounts of data will increase the efficiency of the proposed method. We used spline-based method and random sampling strategy to build training dataset, which enhance the generalization and usability of the trained DL model. The network can be applied directly to new data at the same period range (5–70 s). Although not applicable if period range is different, the network can be easily re-trained. Application to real data showed that the method was powerful and feasible; the real surface-wave dispersion data were never used in the construction of the SfNet, but smaller misfits were obtained for the dispersion data than those from the traditional inversion.

Open research

The data and newly obtained v_s model for the Chinese mainland and the new software can be obtained upon request from the authors or directly from Zenodo: <https://doi.org/10.5281/zenodo.7056170>.

The Rayleigh wave dispersion data used in this study were obtained from Li MK et al. (2022), <https://doi.org/10.1016/j.eqs.2022.05.004>. Some figures were plotted using GMT 6.1.1 software (<https://www.generic-mapping-tools.org/download/> [Wessel and Smith, 1998]).

Acknowledgments

We thank two anonymous reviewers for helpful comments. Most of the figures were generated using Generic Mapping Tools (GMT, Wessel et al., 2019). This research was supported by the Open Fund from SinoProbe Laboratory (Grant No. Sinoprobe Lab 202201) and the

National Natural Science Foundation of China (No. U1939204).

References

- Bao XW, Song XD, Xu MJ, Wang LS, Sun XX, Mi N, Yu DY and Li H (2013). Crust and upper mantle structure of the North China Craton and the NE Tibetan plateau and its tectonic implications. *Earth Planet Sci Lett* **369–370**: 129–137 <https://doi.org/10.1016/j.epsl.2013.03.015>.
- Bao XW, Song XD and Li JT (2015). High-resolution lithospheric structure beneath Mainland China from ambient noise and earthquake surface-wave tomography. *Earth Planet Sci Lett* **417**: 132–141 <https://doi.org/10.1016/j.epsl.2015.02.024>.
- Bassin C, Laske G and Masters GG (2000). The current limits of resolution for surface wave tomography in North America. *Eos Trans AGU* **81**(48): F897.
- Ben-Zion Y, Vernon FL, Ozakin Y, Zigone D, Ross ZE, Meng HR, White M, Reyes J, Hollis D and Barklage M (2015). Basic data features and results from a spatially dense seismic array on the San Jacinto fault zone. *Geophys J Int* **202**(1): 370–380 <https://doi.org/10.1093/gji/ggv142>.
- Bergen KJ, Johnson PA, De Hoop MV and Beroza GC (2019). Machine learning for data-driven discovery in solid earth geoscience. *Science* **363**(6433): eaau0323 <https://doi.org/10.1126/science.aau0323>.
- Brauer B and Bauer K (2015). A new interpretation of seismic tomography in the southern Dead Sea basin using neural network clustering techniques. *Geophys Res Lett* **42**(22): 9772–9780 <https://doi.org/10.1002/2015gl066559>.
- Brocher TM (2005). Empirical relations between elastic wavespeeds and density in the earth's crust. *Bull Seismol Soc Am* **95**(6): 2081–2092 <https://doi.org/10.1785/0120050077>.
- Cai A, Qiu HR and Niu FL (2022). Semi-supervised surface wave tomography with Wasserstein cycle-consistent GAN: Method and application to southern California plate boundary region. *J Geophys Res:Solid Earth* **127**(3): e2021JB023598 <https://doi.org/10.1029/2021jb023598>.
- Chang LJ, Ding ZF and Wang CY (2015). Upper mantle anisotropy beneath the southern segment of North-South tectonic belt, China. *Chin J Geophys* **58**(11): 4052–4067 <https://doi.org/10.6038/cjg20151114> (in Chinese with English abstract).
- Chen B, Chen C, Kaban MK, Du JS, Liang Q and Thomas M (2013). Variations of the effective elastic thickness over China and surroundings and their relation to the lithosphere dynamics. *Earth Planet Sci Lett* **363**: 61–72 <https://doi.org/10.1016/j.epsl.2012.12.022>.
- Cheng XQ, Liu QH, Li PP 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. *Nonlin Processes Geophys* **26**(2): 61–71 <https://doi.org/10.5194/npg-26-61-2019>.
- Ding ZF and Wu ZL (2013). Advances of ChinArray program. In: AGU Fall Meeting Abstracts. AGU, San Francisco, CA.
- Dziewonski AM and Anderson DL (1981). Preliminary reference earth model. *Phys Earth Planet Inter* **25**(4): 297–356.
- Eymold WK and Jordan TH (2019). Tectonic regionalization of the southern California crust from tomographic cluster analysis. *J Geophys Res:Solid Earth* **124**(11): 11840–11865 <https://doi.org/10.1029/2019jb018423>.
- Fu L, Pan L, Ma QB, Dong S and Chen XF (2021). Retrieving S-wave velocity from surface wave multimode dispersion curves with DispINet. *J Appl Geophys* **193**: 104430 <https://doi.org/10.1016/j.jappgeo.2021.104430>.
- Gkarlaoui C, Papadimitriou EE, Karakostas VG, Wen XZ, Jin XS, Kiliyas A, Pan H and Yang JL (2008). Implication of fault interaction to seismic hazard assessment in Sichuan-Yunnan provinces of southeastern China. *Earthq Sci* **21**(2): 181–201 <https://doi.org/10.1007/s11589-008-0009-7>.
- Hauksson E and Haase JS (1997). Three-dimensional v_p and v_p/v_s velocity models of the Los Angeles basin and central Transverse Ranges, California. *J Geophys Res: Solid Earth* **102**(B3): 5423–5453 <https://doi.org/10.1029/96jb03219>.
- He KM, Zhang XY, Ren SQ and Sun J (2016). Identity mappings in deep residual networks. In: 14th European Conference on Computer Vision. Springer, Amsterdam, the Netherlands, pp 630–645. https://doi.org/10.1007/978-3-319-46493-0_38.
- Herrmann RB (2013). Computer programs in seismology: an evolving tool for instruction and research. *Seismol Res Lett* **84**(6): 1081–1088 <https://doi.org/10.1785/0220110096>.
- Hu J, Qiu HR, Zhang HJ and Ben-Zion Y (2020). Using deep learning to derive shear-wave velocity models from surface-wave dispersion data. *Seismol Res Lett* **91**(3): 1738–1751 <https://doi.org/10.1785/0220190222>.
- Huang JL and Zhao DP (2006). High-resolution mantle tomography of China and surrounding regions. *J Geophys Res:Solid Earth* **111**(B9): B09305 <https://doi.org/10.1029/2005jb004066>.
- IRIS Consortium (1999) USArray—A component of Earthscope, report, Earth Sci Div, Nat Sci Found.

- Jiang C, Fang LH, Fan LP and Li BR (2021). Comparison of the earthquake detection abilities of PhaseNet and EQTransformer with the Yangbi and Maduo earthquakes. *Earthq Sci* **34**(5): 425–435 <https://doi.org/10.29382/eqs-2021-0038>.
- Jiang WL, Tian T, Chen Y, Wang X and Zhang JF (2020). Detailed crustal structures and seismotectonic environment surrounding the Su-Lu segment of the Tan-Lu fault zone in the eastern China mainland. *Geosci J* **24**(5): 557–574 <https://doi.org/10.1007/s12303-019-0041-8>.
- Jo E and Hong TK (2013). v_p/v_s ratios in the upper crust of the southern Korean Peninsula and their correlations with seismic and geophysical properties. *J Asian Earth Sci* **66**: 204–214 <https://doi.org/10.1016/j.jseaes.2013.01.008>.
- Kaypak B (2008). Three-dimensional v_p and v_p/v_s structure of the upper crust in the Erzincan basin (eastern Turkey). *J Geophys Res:Solid Earth* **113**(B7): B07307 <https://doi.org/10.1029/2006jb004905>.
- Lees JM and Crosson RS (1989). Tomographic inversion for three-dimensional velocity structure at Mount St. Helens using earthquake data. *J Geophys Res: Solid Earth* **94**(B5): 5716–5728 <https://doi.org/10.1029/jb094ib05p05716>.
- Lehujeur M, Vergne J, Schmittbuhl J, Zigone D, Le Chenadec A and EstOF Team (2018). Reservoir imaging using ambient noise correlation from a dense seismic network. *J Geophys Res:Solid Earth* **123**(8): 6671–6686 <https://doi.org/10.1029/2018jb015440>.
- Lekic V, Cottaar S, Dziewonski A and Romanowicz B (2012). Cluster analysis of global lower mantle tomography: a new class of structure and implications for chemical heterogeneity. *Earth Planet Sci Lett* **357–358**: 68–77 <https://doi.org/10.1016/j.epsl.2012.09.014>.
- Levander A, Humphreys ED, Ekstrom G, Meltzer AS and Shearer PM (1999). Proposed project would give unprecedented look under North America. *Eos Trans AGU* **80**(22): 245–251 <https://doi.org/10.1029/99eo00181>.
- Li JT, Song XD, Zhu LP and Deng YF (2017). Joint inversion of surface wave dispersions and receiver functions with P velocity constraints: application to southeastern Tibet. *J Geophys Res: Solid Earth* **122**(9): 7291–7310 <https://doi.org/10.1002/2017jb014135>.
- Li MK, Song XD, Li JT and Bao XW (2022). Crust and upper mantle structure of East Asia from ambient noise and earthquake surface wave tomography. *Earthq Sci* **35**(2): 71–92 <https://doi.org/10.1016/j.eqs.2022.05.004>.
- Li ZF (2021). Recent advances in earthquake monitoring II: Emergence of next-generation intelligent systems. *Earthq Sci* **34**(6): 531–540 <https://doi.org/10.29382/eqs-2021-0054>.
- Luo YH, Huang Y, Yang YJ, Zhao KF, Yang XZ and Xu HR (2022). Constructing shear velocity models from surface wave dispersion curves using deep learning. *J Appl Geophys* **196**: 104524 <https://doi.org/10.1016/j.jappgeo.2021.104524>.
- Meier U, Curtis A and Trampert J (2007). Fully nonlinear inversion of fundamental mode surface waves for a global crustal model. *Geophys Res Lett* **34**(16): L16304 <https://doi.org/10.1029/2007gl030989>.
- Mora P and Tarantola A (1989). Seismic imaging. In: *Geophysics. Encyclopedia of Earth Science*. Springer, Boston, MA, pp 1024–1033. https://doi.org/10.1007/0-387-30752-4_124.
- Mosegaard K and Tarantola A (1995). Monte Carlo sampling of solutions to inverse problems. *J Geophys Res:Solid Earth* **100**(B7): 12431–12447 <https://doi.org/10.1029/94jb03097>.
- Mousavi SM, Ellsworth WL, Zhu WQ, Chuang LY and Beroza GC (2020). Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nat Commun* **11**(1): 3952 <https://doi.org/10.1038/s41467-020-17591-w>.
- Nolet G, van Trier J and Huisman R (1986). A formalism for nonlinear inversion of seismic surface waves. *Geophys Res Lett* **13**(1): 26–29 <https://doi.org/10.1029/gl013i001p00026>.
- Raschka S and Mirjalili V (2017). *Python machine learning: Machine learning and deep learning with Python, Scikit-Learn, and TensorFlow*. 2nd ed. Packt Publishing, Birmingham.
- Ronneberger O, Fischer P and Brox T (2015). U-Net: Convolutional networks for biomedical image segmentation. In: *18th International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Munich, Germany, pp 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
- Sambridge M (1999). Geophysical inversion with a neighbourhood algorithm—I. Searching a parameter space. *Geophys J Int* **138**(2): 479–494 <https://doi.org/10.1046/j.1365-246x.1999.00876.x>.
- She YY, Yao HJ, Zhai QS, Wang FY and Tian XF (2018). Shallow crustal structure of the middle-lower Yangtze river region in eastern China from surface-wave tomography of a large volume airgun-shot experiment. *Seismol Res Lett* **89**(3): 1003–1013 <https://doi.org/10.1785/0220170232>.
- Shen WS and Ritzwoller MH (2016). Crustal and uppermost mantle structure beneath the United States. *J Geophys*

- Res:Solid Earth **121**(6): 4306–4342 <https://doi.org/10.1002/2016jb012887>.
- Shen WS, Ritzwoller MH, Kang D, Kim Y, Lin FC, Ning JY, Wang WT, Zheng Y and Zhou LQ (2016). A seismic reference model for the crust and uppermost mantle beneath China from surface wave dispersion. *Geophys J Int* **206**(2): 954–979 <https://doi.org/10.1093/gji/ggw175>.
- Snoke JA and Sambridge M (2002). Constraints on the S wave velocity structure in a continental shield from surface wave data: comparing linearized least squares inversion and the direct search neighbourhood algorithm. *J Geophys Res: Solid Earth* **107**(B5): ESE4–1–ESE4-8 <https://doi.org/10.1029/2001jb000498>.
- Stankiewicz J, Bauer K and Ryberg T (2010). Lithology classification from seismic tomography: additional constraints from surface waves. *J Afr Earth Sci* **58**(3): 547–552 <https://doi.org/10.1016/j.jafrearsci.2010.05.012>.
- Sun XL, Song XD, Zheng SH, Yang YJ and Ritzwoller MH (2010). Three dimensional shear wave velocity structure of the crust and upper mantle beneath China from ambient noise surface wave tomography. *Earthq Sci* **23**(5): 449–463 <https://doi.org/10.1007/s11589-010-0744-4>.
- Tarantola A (1987). Inversion of travel times and seismic waveforms. In: Nolet G (ed). *Seismic Tomography*. Springer, Netherlands, pp 135–157. https://doi.org/10.1007/978-94-009-3899-1_6.
- Teng JW, Zhang ZJ, Zhang XK, Wang CY, Gao R, Yang BJ, Qiao YH and Deng YF (2013). Investigation of the Moho discontinuity beneath the Chinese mainland using deep seismic sounding profiles. *Tectonophysics* **609**: 202–216 <https://doi.org/10.1016/j.tecto.2012.11.024>.
- Wang F, Shen ZK, Wang M, Wang YZ and Tao W (2013). Seismic hazard analysis of the Sichuan-Yunnan region, western China. *Seismol Geol* **35**(1): 101–112 <https://doi.org/10.3969/j.issn.0253-4967.2013.01.009> (in Chinese with English abstract).
- Wang FY, Song XD and Li JT (2022). Deep learning-based H - κ method (HkNet) for estimating crustal thickness and v_p/v_s ratio from receiver functions. *J Geophys Res: Solid Earth* **127**(6): e2022JB023944 <https://doi.org/10.1029/2022jb023944>.
- Wang J, Xiao ZW, Liu C, Zhao DP and Yao ZX (2019). Deep learning for picking seismic arrival times. *J Geophys Res: Solid Earth* **124**(7): 6612–6624 <https://doi.org/10.1029/2019JB017536>.
- Wang Q, Song XD and Ren JY (2017). Ambient noise surface wave tomography of marginal seas in East Asia. *Earth Planet Phys* **1**(1): 13–25 <https://doi.org/10.26464/epp2017003>.
- Wen XZ, Fang D, Yi GX, Long F, Fan J, Yang PX, Xiong RW, Liu XX and Liu Q (2013). Earthquake potential of the Zhao-tong and Lianfeng fault zones of the eastern Sichuan-Yunnan border region. *Chin J Geophys* **56**(10): 3361–3372 <https://doi.org/10.6038/cjg20131012> (in Chinese with English abstract).
- Wessel P, Luis JF, Uieda L, Scharroo R, Wobbe F, Smith WHF and Tian D (2019). The generic mapping tools version 6. *Geochemistry, Geophysics, Geosystems* **20**(11): 5556–5564 <https://doi.org/10.1029/2019GC008515>.
- Xu PF and Zhao DP (2009). Upper-mantle velocity structure beneath the North China craton: implications for lithospheric thinning. *Geophys J Int* **177**(3): 1279–1283 <https://doi.org/10.1111/j.1365-246x.2009.04120.x>.
- Xu ZJ, Song XD and Zheng SH (2013a). Shear velocity structure of crust and uppermost mantle in China from surface wave tomography using ambient noise and earthquake data. *Earthq Sci* **26**(5): 267–281 <https://doi.org/10.1007/s11589-013-0010-7>.
- Xu ZJ, Song XD and Zhu LP (2013b). Crustal and uppermost mantle S velocity structure under Hi-CLIMB seismic array in central Tibetan Plateau from joint inversion of surface wave dispersion and receiver function data. *Tectonophysics* **584**: 209–220 <https://doi.org/10.1016/j.tecto.2012.08.024>.
- Yao HJ, Beghein C and van der Hilst RD (2008). Surface wave array tomography in SE Tibet from ambient seismic noise and two-station analysis – II. Crustal and upper-mantle structure. *Geophys J Int* **173**(1): 205–219 <https://doi.org/10.1111/j.1365-246x.2007.03696.x>.
- Yu SW and Ma JW (2021). Deep learning for geophysics: Current and future trends. *Rev Geophys* **59**(3): e2021RG000742 <https://doi.org/10.1029/2021rg000742>.
- Zhang L, Fang LH, Wang WL and Lv ZY (2020). Seismic phase picking in China Seismic Array using a deep convolutional neuron network. *Earthq Sci* **33**(2): 72–81 <https://doi.org/10.29382/eqs-2020-0072-03>.
- Zhou YJ, Yue H, Fang LH, Zhou SY, Zhao L and Ghosh A (2022). An earthquake detection and location architecture for continuous seismograms: Phase picking, association, location, and matched filter (PALM). *Seismol Res Lett* **93**(1): 413–425 <https://doi.org/10.1785/0220210111>.

Supplementary information

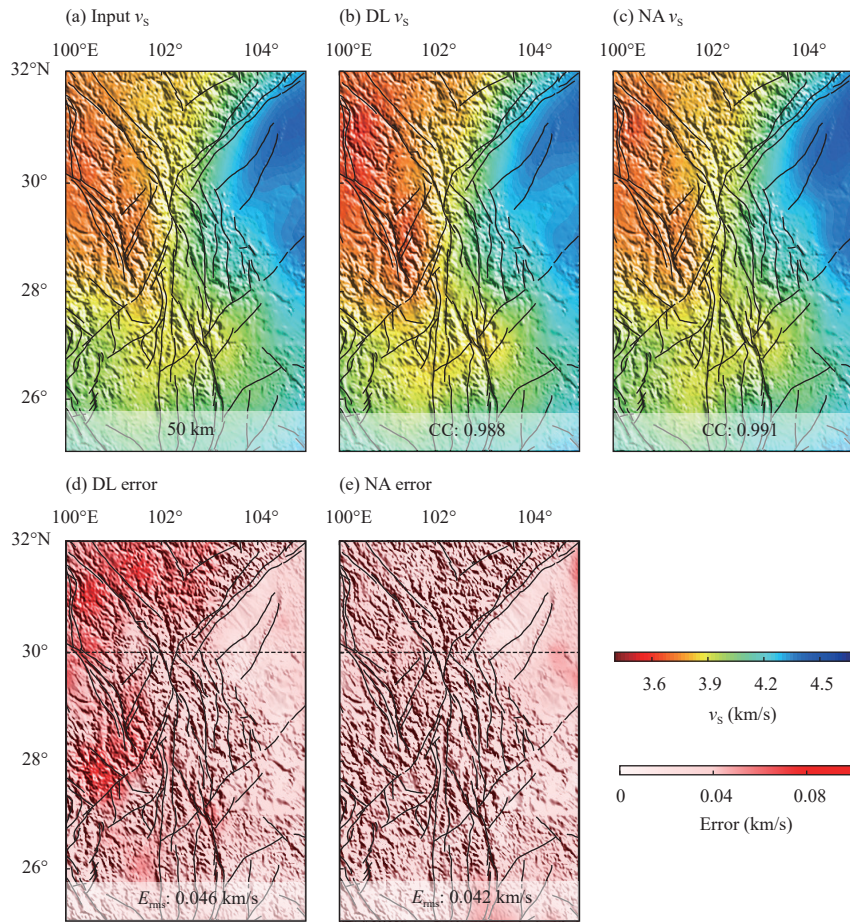
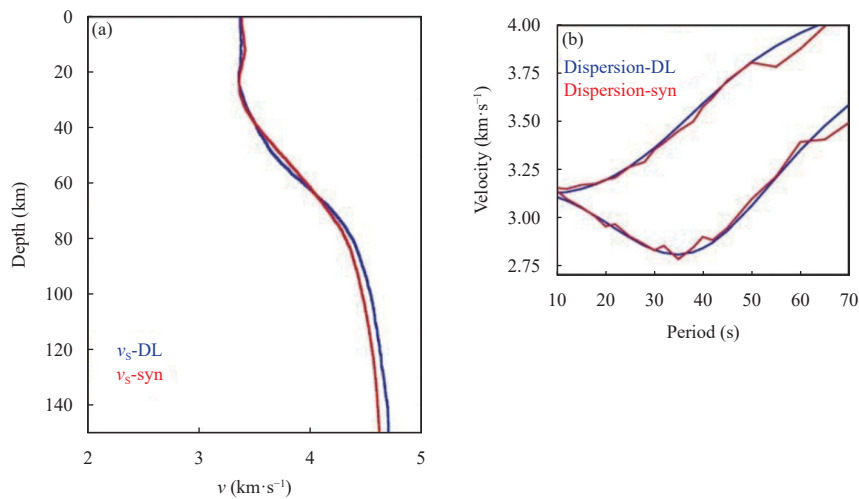


Figure S1. Comparison of the results of SfNet and NA for the SY region at 50 km. (a) Input v_s model, (b) inverted v_s model SfNet, (c) inverted v_s model from NA, (d) absolute difference between the inverted v_s model from DL method and input models, and (e) absolute difference between the inverted v_s model from NA and input models. CC values between the input and inverted models and the RMS of the difference are also labeled.



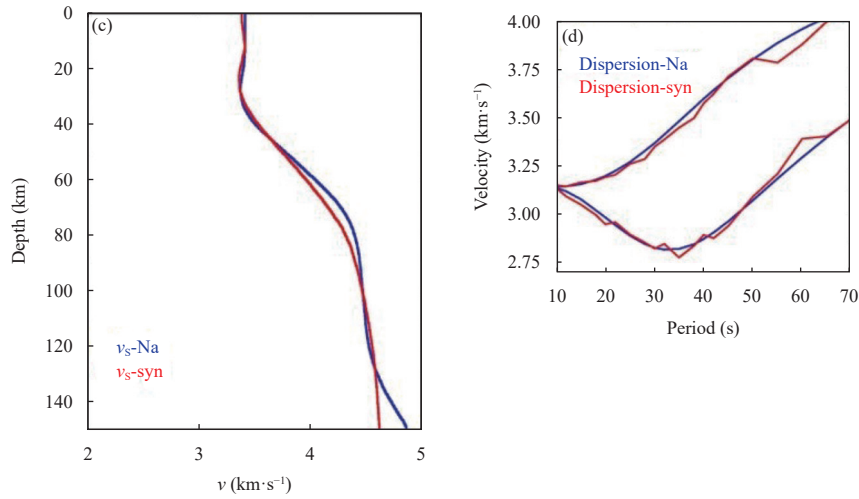


Figure S2. Continued. Inversion results at nodes a. (a) The input velocity models and DL inversion models; (b) The simulate real dispersions and dispersions of DL models; (c) The input velocity modes and NA inversion models; (d) The simulate real dispersions and dispersions of NA models.

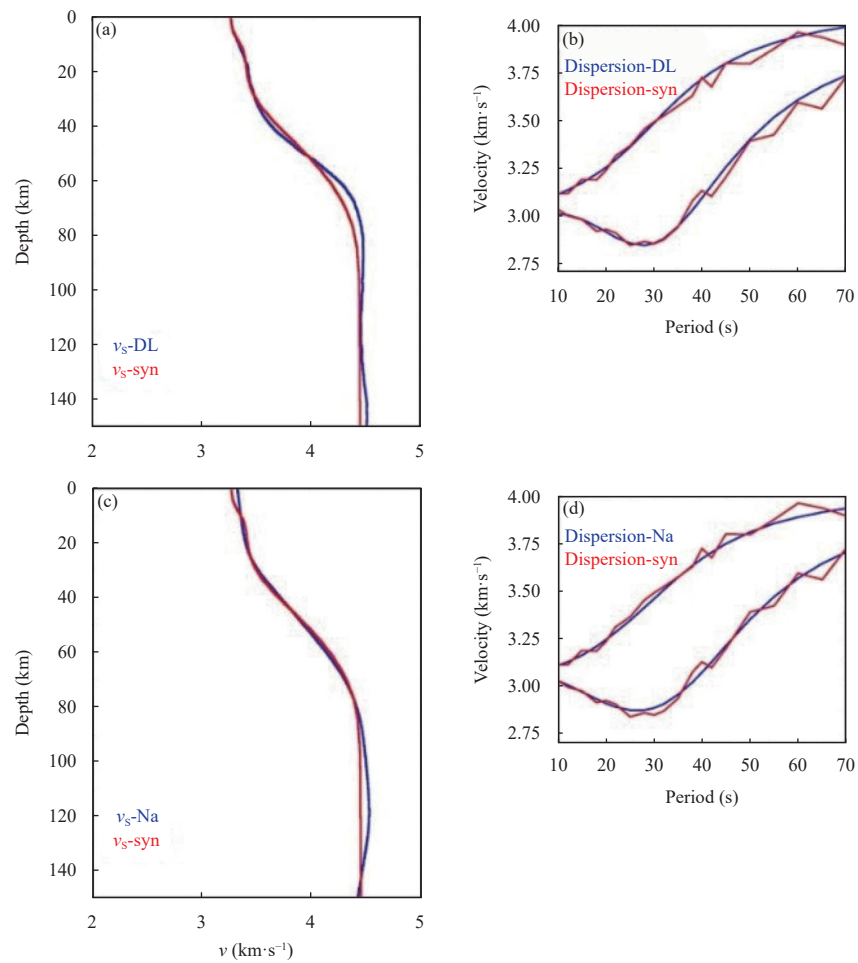


Figure S3. Inversion results at nodes b. (a) The input velocity models and DL inversion models; (b) The simulate real dispersions and dispersions of DL models; (c) The input velocity modes and NA inversion models; (d) The simulate real dispersions and dispersions of NA models.

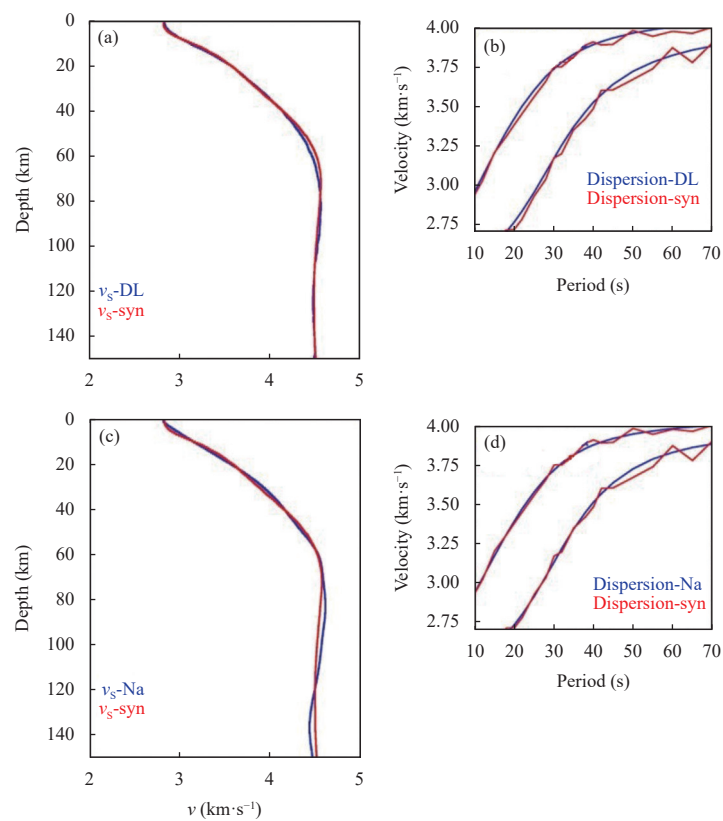


Figure S4. Same as Figure S2 but at nodes c.

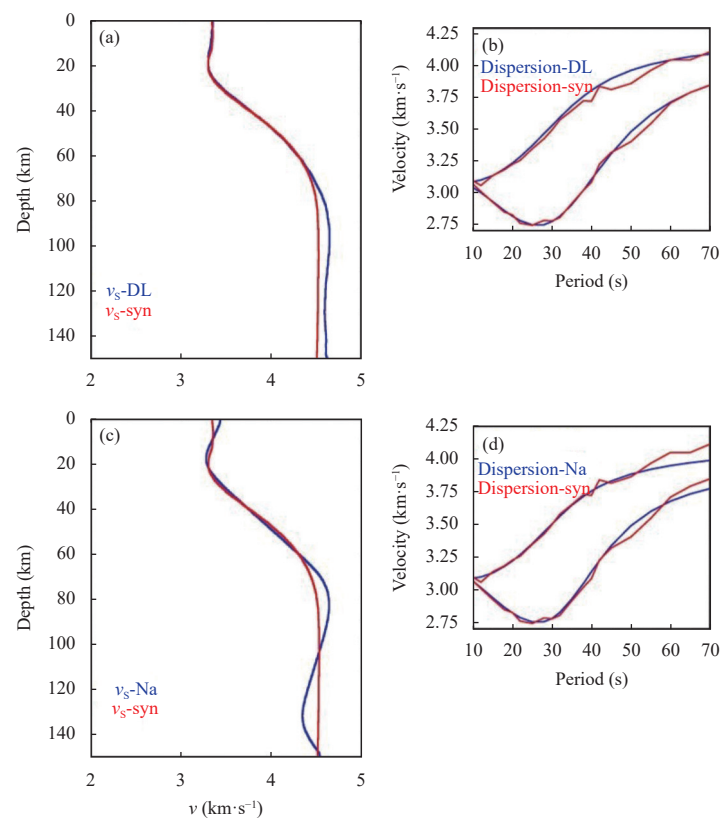


Figure S5. Same as Figure S2 but at nodes d.

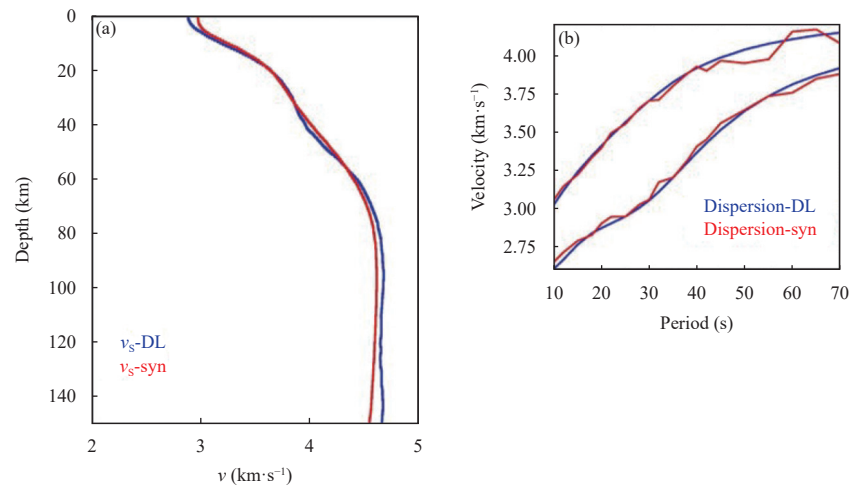


Figure S6. Inversion results at nodes 1. (a) The input velocity models and DL inversion models; (b) The simulate real dispersions and dispersions of DL models.

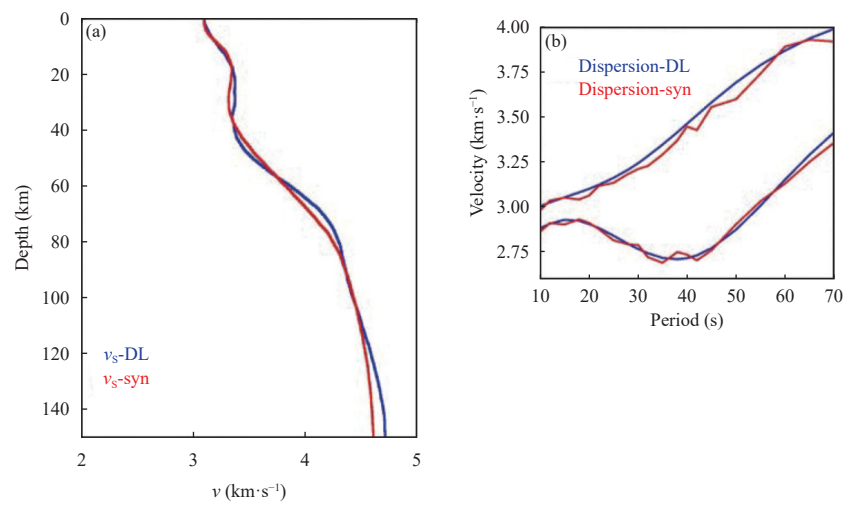


Figure S7. Same as Figure S6 but at nodes 2.

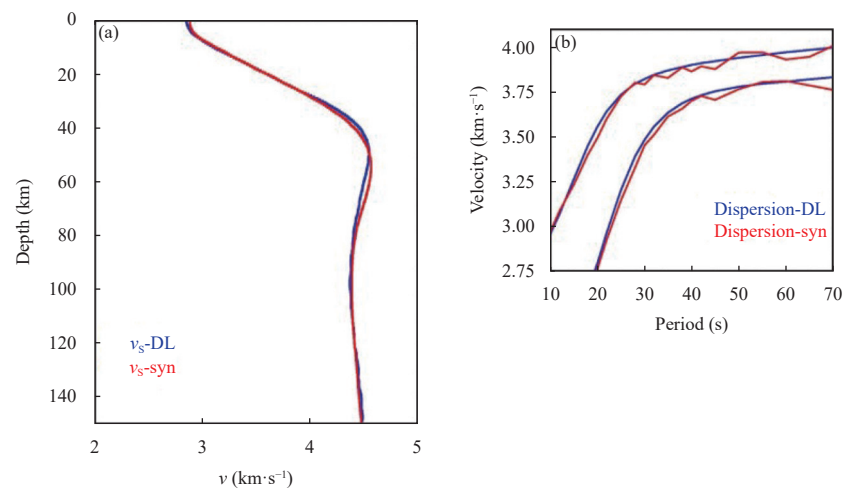


Figure S8. Same as Figure S6 but at nodes 3.

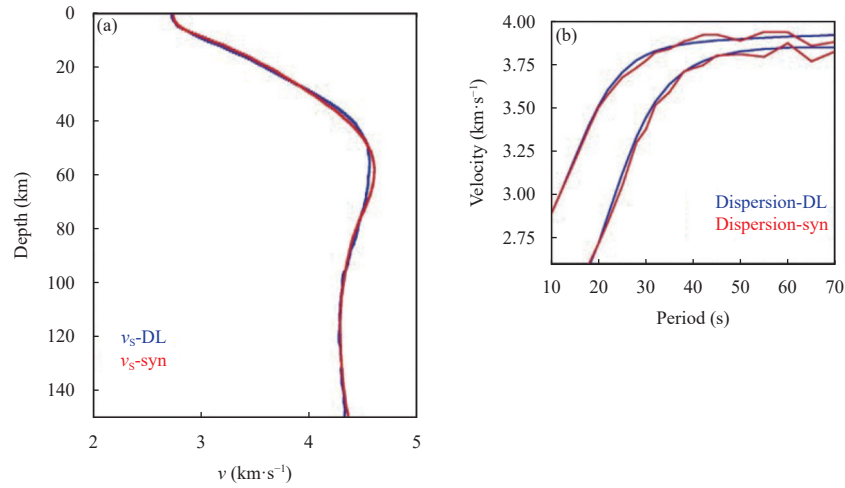


Figure S9. Same as Figure S6 but at nodes 4.

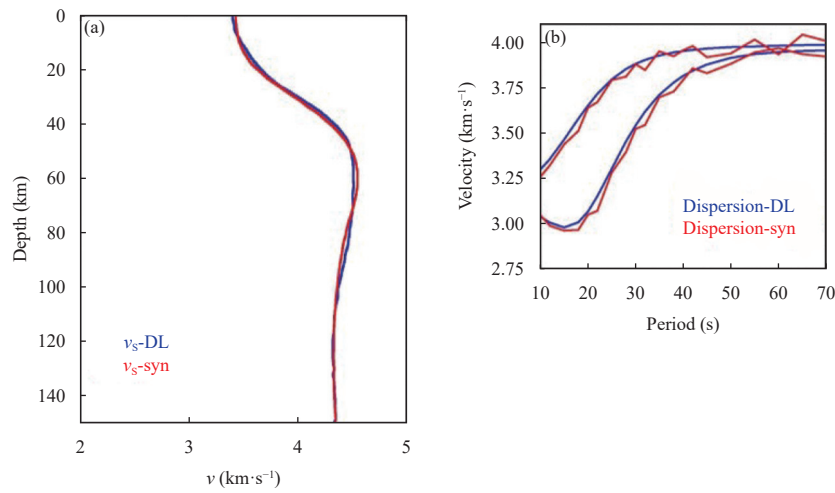


Figure S10. Same as Figure S6 but at nodes 5.

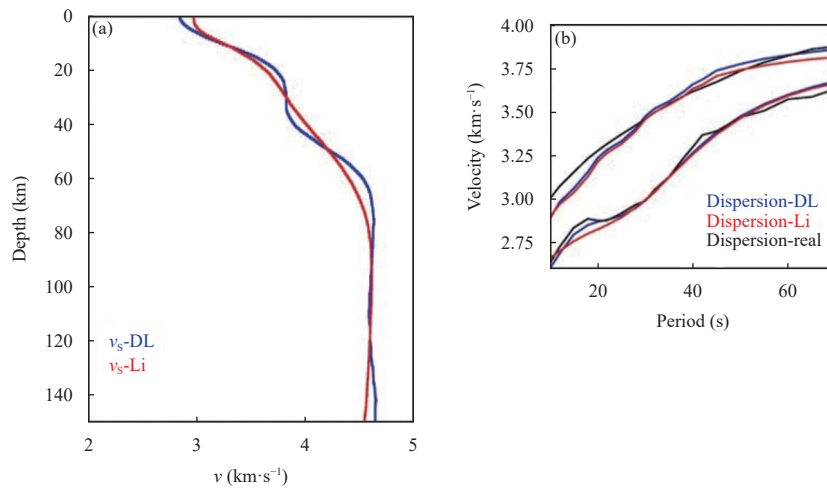


Figure S11. Comparison of DL inversion results with Li's at nodes 1 (Li et al., 2022). (a) The Li's inversion model and DL inversion model; (b) The dispersions of observation, DL model and Li's model.

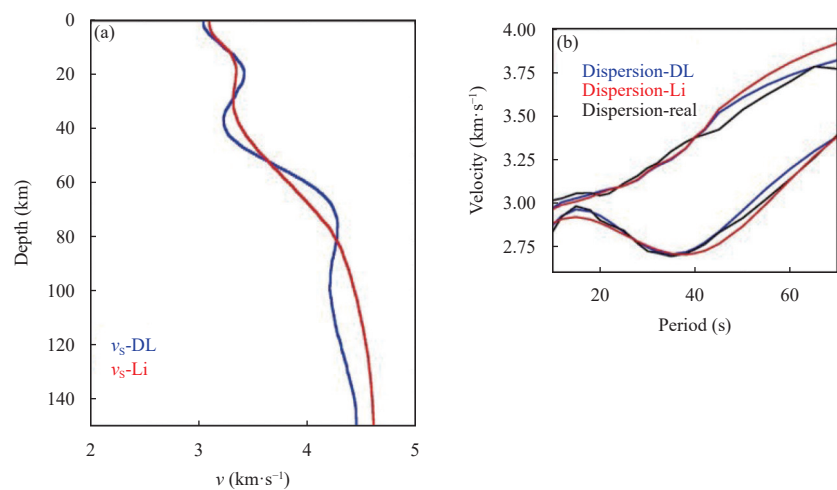


Figure S12. Same as Figure S11 but at nodes 2.

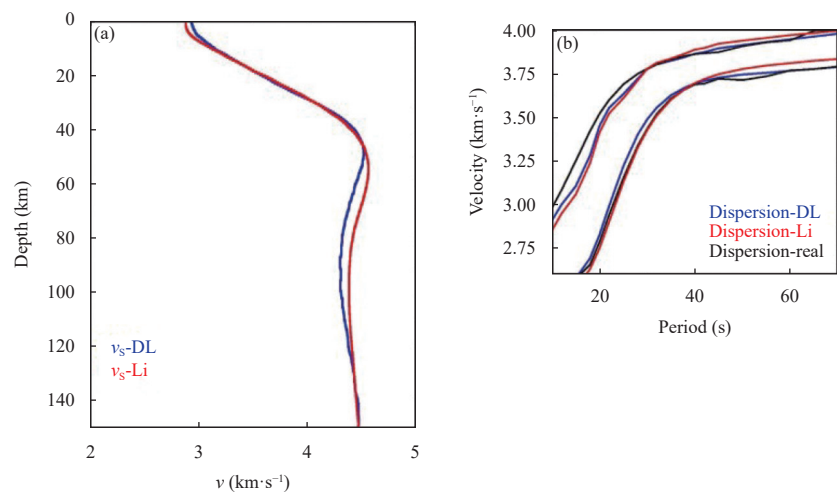


Figure S13. Same as Figure S11 but at nodes 3.

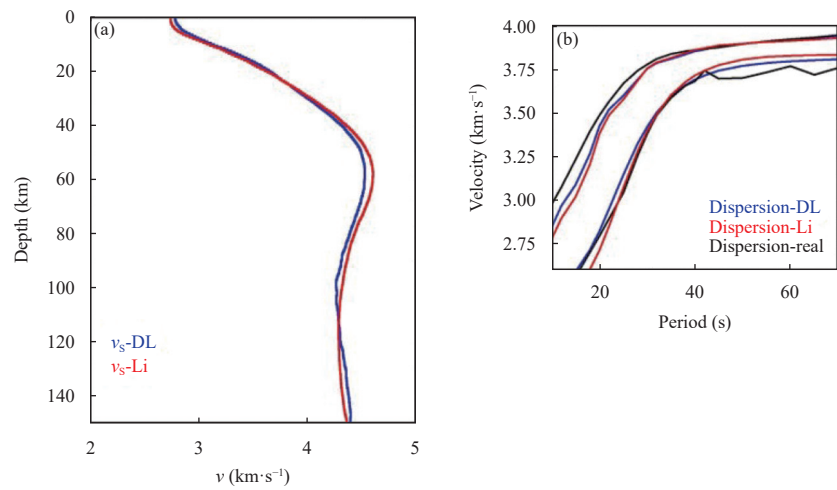


Figure S14. Same as Figure S11 but at nodes 4.

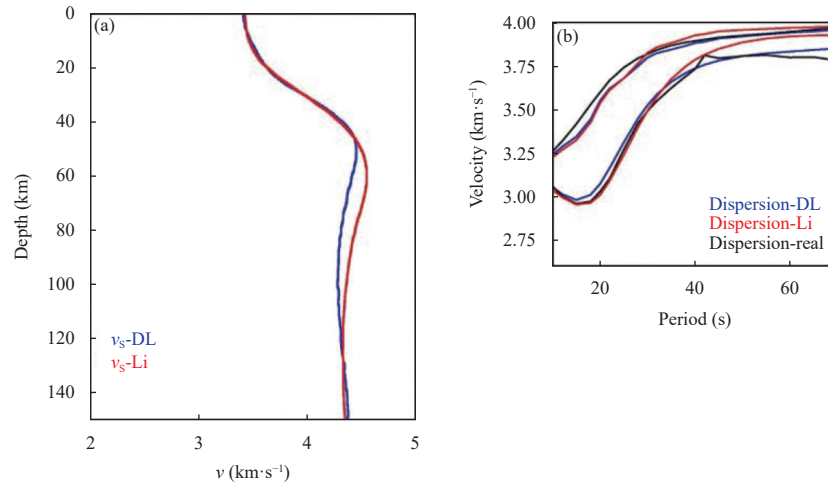


Figure S15. Same as Figure S11 but at nodes 5.

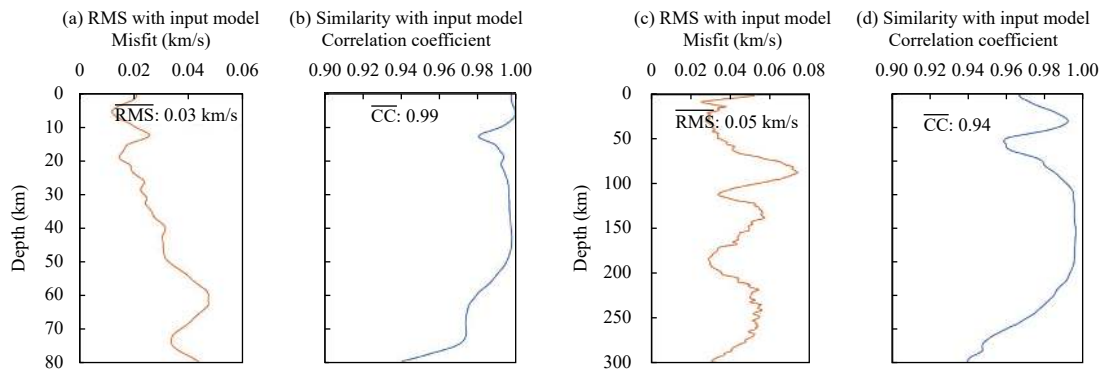


Figure S16. Comparison of the v_s models obtained from the proposed DL method with the input models: (a) RMS misfit at 0–80 km of inversion using 5–35 s dispersion data. (b) CC of inversion using 5–35 s dispersion data. (c) (d) same as (a) and (b) but for 10–140 s dispersion data. Average CC and RMS misfit are also provided.