JointNet: A Multimodal Deep Learning-Based Approach for Joint Inversion of Rayleigh Wave Dispersion and Ellipticity

Xiang Huang[1,2], Ziye Yu[*1,2], Weitao Wang[1,2], and Fang Wang[1]

ABSTRACT

Joint inversion of multitype datasets is an effective approach for high-precision subsurface imaging. We present a new deep learning-based method to jointly invert Rayleigh wave phase velocity and ellipticity into shear-wave velocity of the crust and uppermost mantle. A multimodal deep neural network (termed JointNet) is designed to analyze these two independent physical parameters and generate outputs, including velocity and layer thicknesses. JointNet is trained using random 1D models and corresponding synthetic phase velocity and ellipticity, resulting in a low cost for the training dataset. Evaluation using synthetic and observed data shows that JointNet produces highly comparable results compared to those from a Markov chain Monte Carlo-based method and significantly improves inversion speed. Training using synthetic data ensures its generalized application in various regions with different velocity structures. Moreover, JointNet can be easily extended to include additional datatypes and act as a joint inversion framework to further improve imaging resolution.

KEY POINTS

- A novel multimodal network capable of jointly inverting Rayleigh wave dispersion and ellipticity into V_S.
- It requires less prior knowledge and is significantly faster compared to traditional nonlinear techniques.
- It can be easily extended to include more input types and serve as a framework for joint inversion.

INTRODUCTION

Surface-wave tomography is a classical method for studying the crustal and upper mantle structure of the Earth (Press, 1956; Forsyth *et al.*, 1998; Shapiro and Campillo, 2004; Huang *et al.*, 2009). The surface-wave velocity at different periods is sensitive to shear-wave velocity (V_S) at different depths, so the 1D V_S model of the crust and upper mantle can be inverted using observed surface-wave velocity dispersions. However, due to the limited period range of high-quality dispersion, especially at short periods, in regional imaging studies, constraints on the shallow crust are limited (Lin *et al.*, 2012).

To achieve more precise constraints on both shallow and deep crustal structures, the joint inversion of Rayleigh wave dispersion and ellipticity (also known as the ZH ratio) has been widely applied for fine crustal structure imaging (Lin *et al.*, 2012; Li *et al.*, 2016; Huang *et al.*, 2021). The ZH ratio at 10–24 s periods has a high sensitivity to V_S at shallow depths (<5 km), thus it can form a good complement to the dispersion to obtain more detailed information about the crustal structure by joint inversion. However, traditional joint inversion methods have their limitations. The linear methods are fast, but require optimal regularization and an appropriate initial model. The nonlinear methods based on global optimization algorithms can avoid the heavy dependence on starting models (Bodin *et al.*, 2012; Afonso *et al.*, 2013; Shen *et al.*, 2013; Guo *et al.*, 2015; Zhang *et al.*, 2020). However, they are usually time consuming and may be difficult to converge on due to incomplete sampling.

Deep neural network (DNN) has undergone rapid development recently and has been widely applied in geophysical research. DNN has been applied to solve different problems, including event detection (Perol *et al.*, 2018; Tous *et al.*, 2020), phase picking (Ross *et al.*, 2018; Yu *et al.*, 2018; Wang *et al.*, 2019), surface-wave dispersion picking (Dai *et al.*, 2021; Wang *et al.*, 2021; Yang *et al.*, 2022), and model inversion (Devilee *et al.*, 1999; Meier *et al.*, 2007; Hu *et al.*, 2020; Cai *et al.*, 2022; Luo *et al.*, 2022). To some extent, the DNNs can provide alternative approaches to traditional processing with human participation, such as manual picking of seismic phases,

1. Institute of Geophysics, China Earthquake Administration, Beijing, China, https://orcid.org/0000-0002-1720-3811 (ZY); 2. Key Laboratory of Earthquake Source Physics, China Earthquake Administration, Beijing, China

*Corresponding author: yuziye@cea-igp.ac.cn

Cite this article as Huang, X., Z. Yu, W. Wang, and F. Wang (2023). JointNet: A Multimodal Deep Learning-Based Approach for Joint Inversion of Rayleigh Wave Dispersion and Ellipticity, *Bull. Seismol. Soc. Am.* **114**, 627–641, doi: 10.1785/0120230199

TABLE 1
Model Parameterization and Perturbation Ranges of the Model Parameters

Model Layers	Model Parameters	Reference	Range	Number of Parameters	Number of Sublayers
Sediment	Sediment thickness (km)		0–10		
	V_S at the top of the sediment layer (km/s)	CRUST1.0	0–3.0	3	8
	V_S at the bottom of the sediment layer (km/s)	CRUST1.0	0–3.0		
Crust	Crystalline crustal thickness (km)		20–80	6	20
	5 B-spline coefficients for the crystalline crust (km/s)	CRUST1.0	3.1–4.4		
Upper mantle	5 B-spline coefficients for the upper mantle (km/s)	CRUST1.0	4.0–5.4	5	20

earthquake classification, setting initial models and inversion parameters, and so on, and are usually efficient for data-driven seismological studies.

The DNN-based methods have been used for surface-wave dispersion inversion (Hu *et al.*, 2020; Cai *et al.*, 2022; Luo *et al.*, 2022). However, most of them are based on surface-wave dispersion only. The joint inversion framework to fully utilize multicomponent surface-wave characters is still rare. The inclusion of the ZH ratio in the joint inversion can improve the imaging resolution of the shallow crust. Moreover, the flexibility and generalization of these DNNs need improvement. Many existing networks are based on convolutional neural networks (CNN) adopted from computer vision models. For instance, Hu *et al.* (2020) transformed the dispersion curve into an image format as input, which may result in data precision reduction when converting the data to images. In addition, most networks have strict requirements for the input data format. The input dispersions are required to have the same period range and sampling interval as the training data (Cai *et al.*, 2022; Luo *et al.*, 2022). However, in reality, the period ranges and sampling intervals of the dispersion data can vary between different studies, and retraining the network is required, limiting the generalized application of the models. To address this concern, Wang *et al.* (2023) proposed a random sampling technique for the dispersion period. Nevertheless, the period range of the input dispersion is still fixed. A DNN-based approach with fewer input requirements and generalized model sampling will help to apply the trained model to various regions.

In this study, we design a multimodal DNN model, named JointNet, to jointly invert Rayleigh wave phase velocity and ZH ratio into a 1D V_S model, and delineate the thicknesses of sedimentary and crustal layers. The DNN model is trained using a large quantity of random 1D models based on the global CRUST1.0 model (Laske *et al.*, 2013). The trained model can capsule the nonlinear relationship between the V_S model and the phase velocity and ZH ratio. The model performance is validated using synthetic data and observed phase velocity and ZH ratio obtained from North China. The comparison with 1D models inverted using the Markov chain Monte Carlo (McMC)-based method suggests that the trained model can produce competitive results with high efficiency. Furthermore, it can be extended to include more input types as a joint inversion framework.

DATA

Training data

The train data are purely synthetic, and the dispersion and ellipticity are calculated based on a large number of 1D velocity models in a reasonable model space. These 1D models are parameterized using the strategy of Huang *et al.* (2023). Specifically, 14 parameters are used to parameterize a 1D model (Table 1), including the sedimentary thickness, the V_S at the upper and lower boundaries of the sedimentary layer, the crustal thickness, five B-splines representing the V_S in the crust, and likewise five B-splines for the upper mantle. Based on the B-splines, the crust and upper mantle are interpolated into 20 layers, respectively. Using B-splines to represent V_S can help parameterize the 1D velocity model with fewer parameters, thereby reducing computational costs. Moreover, using B-splines helps to obtain a smoother model, which can prevent the occurrence of unreasonable models with large velocity jumps. On the flip side, it may smooth out velocity–discontinuity interfaces within the crust and upper mantle, resulting in the loss of some detailed information on these discontinuities. Nevertheless, because the sedimentary layer, crust, and upper mantle are interpolated separately, the major discontinuities at the bottom of the sedimentary layer and the Moho interface are not smoothed. As demonstrated by Lin *et al.* (2012), Rayleigh wave phase velocity and ellipticity can help to constrain *P*-wave velocity and density. However, they also discovered a strong trade-off between density and the V_P/V_S ratio, thus suggesting that it could be reduced by employing empirical relationships. Because of the relatively lower sensitivity of the Rayleigh wave phase velocity and ellipticity to *P*-wave velocity and density, the V_P and density in the sedimentary layer and the crystalline crust are scaled from the V_S using the relationship of Brocher (2005) in this study. Similarly, we employ the relationship of Birch (1961) to obtain the V_P and density in the upper mantle. This might lead to some deviation between the model and the real situation, but it should be acceptable considering the lower sensitivity of Rayleigh wave phase velocity and ellipticity to *P*-wave velocity and density.

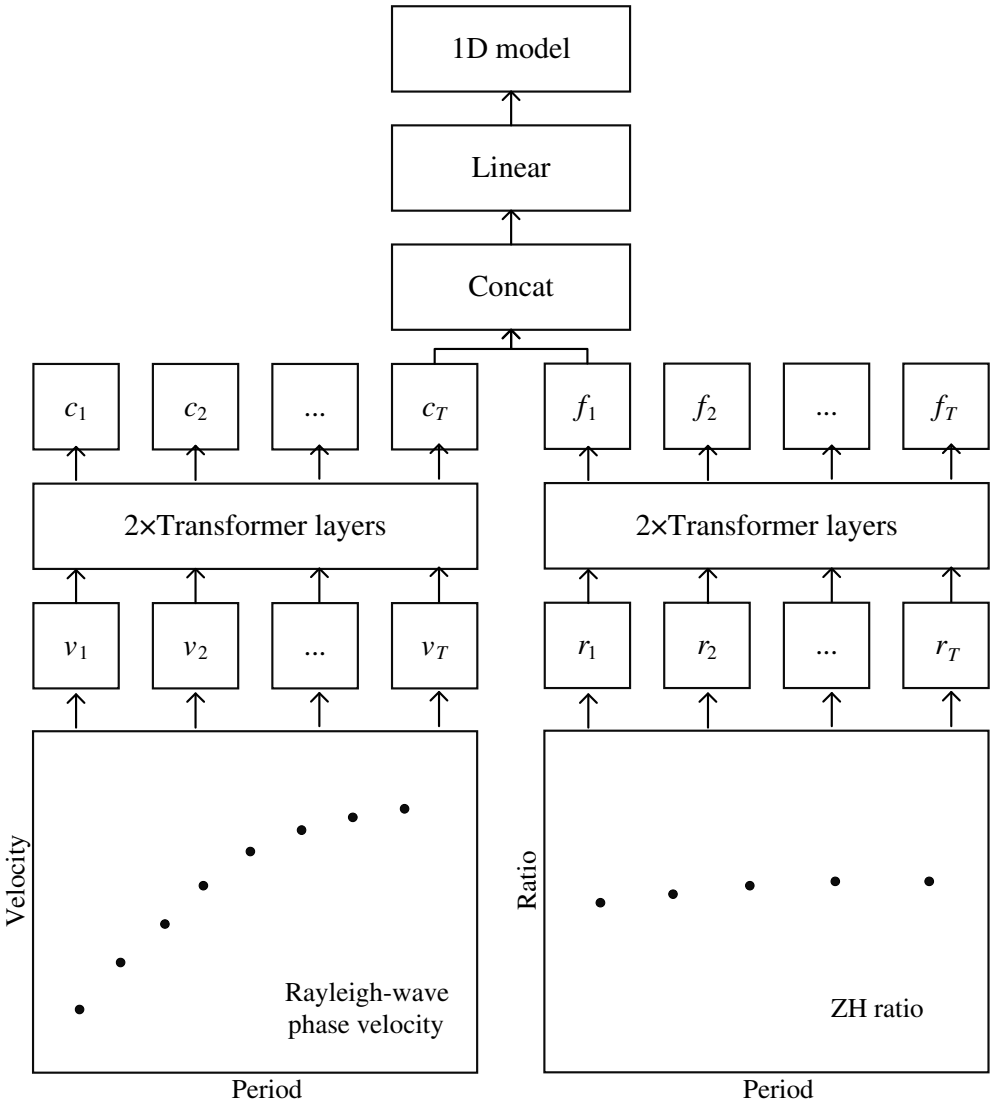


Figure 1. Structure of JointNet. The bottom two panels represent the input phase velocity curve and ZH ratio curve. $v = [v_1, v_2, \ldots, v_{T1}]$ stands for the phase velocities at different periods, and $r = [r_1, \ldots, r_{T2}]$ stands for different ZH ratios at different periods. They are fed to two transformer layers for extracting corresponding features c and f, respectively. Then, the concatenation of c and f are converted to 1D model using a linear layer.

To cover as many potential 1D models as possible for generalization, we randomly generate models in quite large model space. To represent a realistic Earth structure, the models are perturbated based on the CRUST1.0 model (Laske *et al.*, 2013). Here, we did not directly use the 1D models from CRUST1.0, mainly because the number of models from CRUST1.0 is limited, which may cause overfitting during training. During the training process, random samples are generated at each step, which helps to prevent overfitting. The perturbation ranges of the model parameters are shown in Table 1. To add more physical constraints, we additionally apply several prior conditions to the randomly generated models: (1) V_S increases linearly with depth in the sedimentary layer; (2) V_S at the top of the upper mantle should be larger than that at the bottom of the crust. Once each 1D model is randomly generated within the defined model space, we calculate the synthetic Rayleigh wave phase velocity dispersion of 6–40 s period and ellipticity (ZH ratio) of 10–24 s period using the Computer Programs in Seismology software package (Herrmann, 2013). Finally, these 1D models and the corresponding synthetic Rayleigh wave phase velocity dispersions and ellipticities are used as training datasets.

Observed data

We collect the observed Rayleigh wave phase velocity dispersions and ZH ratios for 306 stations of the ChinArray Phase III deployed in North China (Huang *et al.*, 2021) to evaluate the performance of our trained DNN model. For each station, the 6–40 s period Rayleigh wave phase velocity dispersion is derived from ambient noise tomography, and the 10–24 s period ZH ratio is obtained from ambient noise cross-correlations and teleseismic earthquake records. Then, the dispersions and ZH ratios are used to invert the crustal shear-wave velocity models with a McMC-based inversion method. The detailed procedures can be found in Huang *et al.* (2021). In this study, these observed dispersions and ZH ratios serve as inputs for JointNet. The model outputs are jointly inverted 1D models for all stations. The model reliability is evaluated by comparing the model outputs with those jointly inverted using a McMC-based method (Huang *et al.*, 2021) using the same observed dispersions and ZH ratios.

METHODOLOGY

DNN architecture

Inversion methods based on Bayesian theory typically maximize the posterior probability of an inversion problem:

$$P = p(m|d) = \frac{p(m)p(d|m)}{Z}, \quad (1)$$

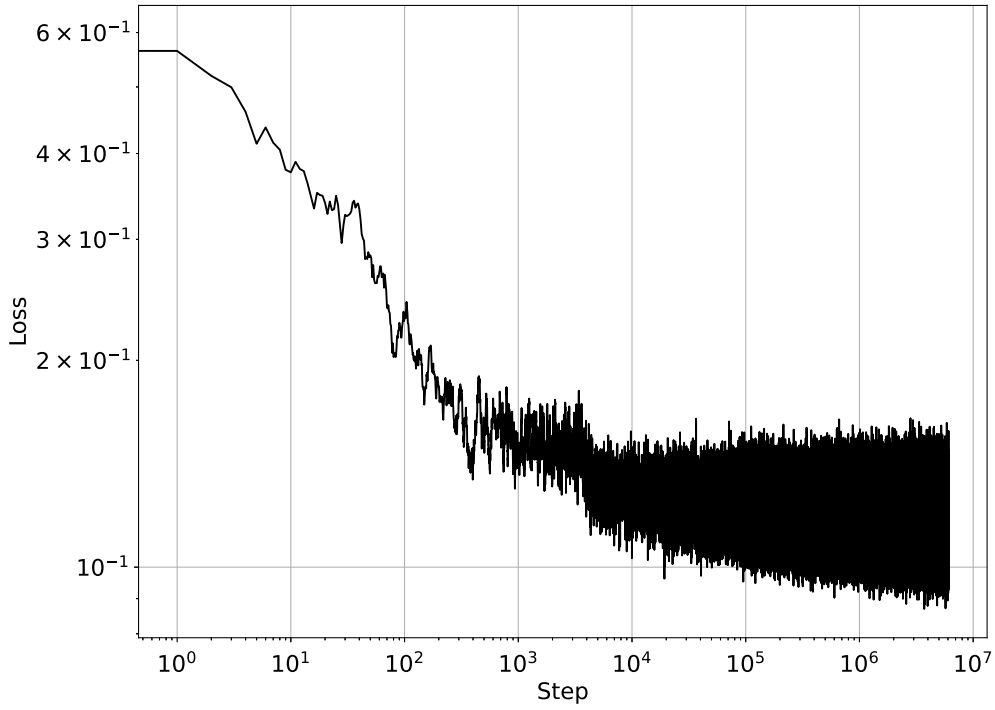


Figure 2. Loss curve in training.

in which m is the model, d is the input data, $p(m)$ is the prior probability of m, $p(d|m)$ is the likelihood or forward calculation, and Z represents the distribution of the input data ($Z = p(d)$). The McMC-based methods iterate to adjust the parameter m and find the posterior probability distribution. However, the prior distribution $p(m)$, or regularization of m, needs to be manually determined. In contrast, the inversion process based on DNN has the advantage of directly fitting the maximum value of the posterior probability. This approach eliminates the need to manually set regularization and avoids the problem of getting trapped in local minima by removing the iterative process.

In this work, we develop a novel inversion approach based on the transformer model (Vaswani *et al.*, 2017), termed JointNet, to directly invert Rayleigh wave phase velocity and ZH ratio for an isotropic 1D shear-wave velocity model. The architecture of JointNet is shown in Figure 1.

As shown in Figure 1, the phase velocity and ZH ratio are fed to the network separately. Here, $v = [v_1, v_2, \ldots, v_{T1}]$ stands for the phase velocities at different periods, and $r = [r_1, \ldots, r_{T2}]$ stands for different ZH ratios at different periods. v and r are fed to different transformer layers and get $c_1, \ldots, c_{T1}$ and $f_1, \ldots, f_{T2}$ as features. As c and f contain the global information of phase velocity and ZH ratio, we can use $g = \mathrm{concat}(c, f)$, the concatenation of c and f, to perform inversion. A linear layer is then constructed to obtain a 1D velocity model from the concatenated input g.

We use the weighted mean absolute error function as the loss function of the DNN:

$$\mathrm{Loss} = \frac{1}{N}\sum_{i}^{N} w_i |v_i^p - v_i^d| + \sum_{j \in \{\mathrm{moho,sed}\}}^{2} w_j |h_j^p - h_j^d|, \quad (2)$$

in which N represents the number of layers. v_i^p and v_i^d represent the predicted and input velocity of the ith layer, respectively. h_k^p and h_k^d represent the predicted and input thickness of the sedimentary and crustal layers. w represents the corresponding weight coefficient.

Training

The JointNet is trained using randomly generated 1D models, along with forward-calculated dispersion and ZH ratio data. There are virtually an infinite number of samples available for both training and testing purposes. This extensive dataset covers all potential solutions so that the trained model can be applied to various regions. Furthermore, the training samples can be considered as the prior information for the Bayes function, and the inversion results will fall within the range of the training samples. The output of JointNet will not exceed the scope of the potential solutions during training. Thus, the previously mentioned parametrizations ensure that we can obtain reasonable structures.

During the training of JointNet, the program randomly generated 32 samples for each batch of data. The learning rate is set to 1×10^{-4}. Because the training samples are completely randomly generated, there is no risk of overfitting, and the weight decay (l2 regularization) is set to 0. We trained the model for 100,000 steps, which is equivalent to approximately 3.2 million training samples until the loss function converged (Fig. 2). Here, we constructed a testing dataset to test the accuracy of inversion and stopped iteration when the accuracy of the testing dataset became stable. The instability of the loss function comes from two aspects. First, during the training process, we add random noise with a normal distribution to the dispersion and ZH ratio. It can not only improve the generalization ability of the network, but it can also cause oscillations in the loss function. Second, due to the multisolution problem of the inversion problem, small errors in the input data can lead to changes in the output model, which can cause

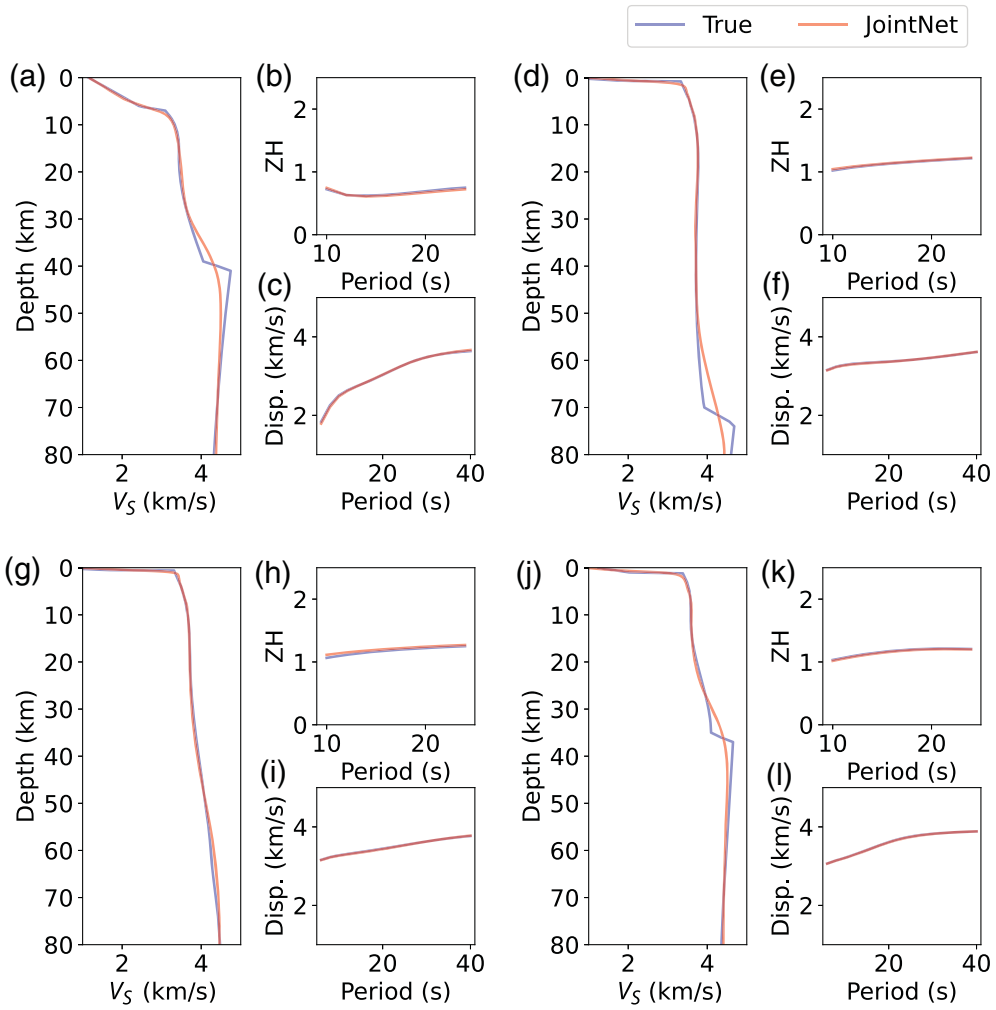


Figure 3. (a,d,g,j) Examples of the inverted 1D V_S models from JointNet compared with the true models. The blue and red solid lines represent the true and inverted models using JointNet, respectively. (b,e,h,k) The blue and red solid lines represent the input and calculated ZH ratio from the inverted models, respectively. (c,f,i,l) The same as panels (b, e, h, and k) but for phase velocity dispersion.

the loss function to oscillate as well. In this study, the training process was performed using a single NVIDIA A100 GPU. The time for a single iteration was 23.6 ms, and the time for 100,000 iterations was 2364 s. The well-trained DNN model is then applied to synthetic and observed dispersions and ZH ratios to obtain V_S profiles for performance evaluation.

PERFORMANCE EVALUATION OF JOINTNET

We evaluate the inversion performance of JointNet in view of both 1D V_S profiles and horizontal slices of the 3D V_S models. This evaluation is conducted using both the synthetic data and the observed data obtained from North China.

Comparisons for 1D V_S profiles

First, we use the synthetic Rayleigh wave phase velocity and ZH ratio calculated from a subset of the randomly generated 1D model to feed JointNet and obtain the inversion results. The misfit between the generated models and the inverted models is calculated to evaluate the differences.

Figure 3 shows a comparison between the inverted 1D models obtained using JointNet and the input models, as well as the data fitting. In most cases, the inverted models closely resemble the input models, and the data fitting is good, demonstrating the robustness of the DNN-based method. It is worth observing that there may be slight deviations observed at the depths in which velocity varies sharply. However, during the generation of the training models, the bottom interface of the sedimentary layer and the Moho interface were not smoothed. As shown in Figure 4, the phase velocity and ZH ratio for a specific period are sensitive to the V_S within a certain depth range, thereby reducing the resolution of velocity–discontinuity interfaces. Thus, we attribute these slight deviations to the lower sensitivity of phase velocity and the ZH ratio to the velocity discontinuities of the models. However, the velocity estimates between the interfaces generally exhibit high accuracy, proving that the DNN model can effectively approximate the nonlinear mapping between the surface-wave datasets and the V_S model.

The histograms of misfits for V_S at different depth ranges and the layer thicknesses are presented in Figure 5. Because the thicknesses of the sedimentary and crustal layers are independent inversion parameters added to the loss function (equation 2), the neural network output includes the thicknesses of the sedimentary and crustal layers. The misfit values exhibit a normal distribution with a mean centered around zero, which indicates the effectiveness and reliability of JointNet. Apart from the relatively larger errors in crustal thickness, the sediment layer thickness and shear-wave velocities at different depths have been accurately recovered. The large errors in the crustal thickness are mainly caused by some models with insignificant velocity contrasts across the Moho. The overall error in the crustal thickness obtained from JointNet is

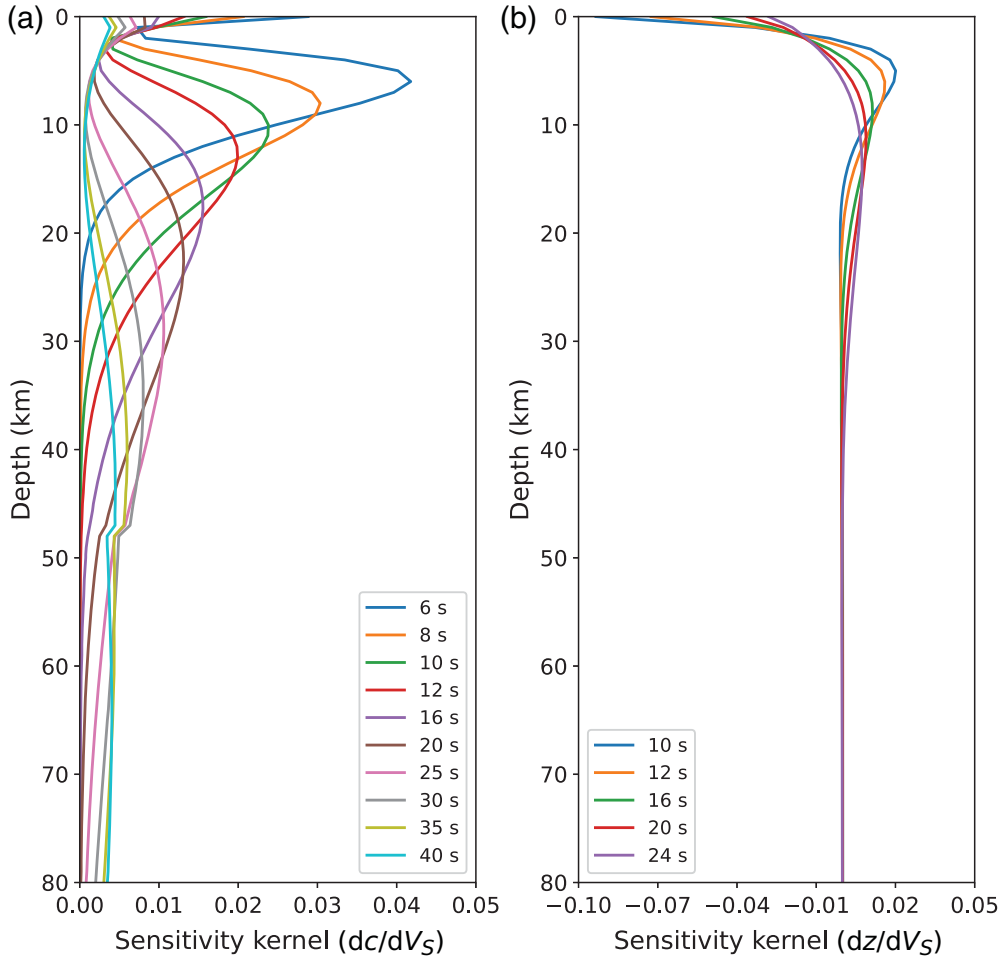


Figure 4. (a) Rayleigh wave phase velocity depth sensitivity kernels at 6–40 s related to V_S. (b) Rayleigh wave ZH ratio depth sensitivity kernels at 10–24 s related to V_S.

relatively large. To improve its accuracy, additional data, such as the receiver function, could be added to future research.

To further evaluate the effectiveness of this approach, we compared it with a linear inversion method and an McMC approach. The linear inversion method is computationally efficient but heavily relies on the accuracy of the initial model. Inversion methods based on the McMC algorithm are known for their robustness against initial model selection and their ability to avoid local minima. Therefore, we compare JointNet's results with those from a linear method (Zhang and Yao, 2017) and a McMC-based method (Li *et al.*, 2016; Huang *et al.*, 2021). Through testing with synthetic data, we found that the accuracy of the inverted results obtained by the linear method is lower than that of JointNet (Fig. 6). This may be related to the selection of the initial model and/or the inversion parameters. Because the accuracy of the results obtained by the McMC-based method is comparable to that of JointNet, we only conduct a systematic comparative analysis between JointNet and the McMC-based method.

Figure 7 shows the 1D V_S models inverted using JointNet and the McMC-based method, and the corresponding calculated phase velocity and ZH ratio, respectively. The inputs are both synthetic dispersion and ellipticity. These models exhibit striking similarity and are in close agreement with the input V_S models. The true models and the JointNet inverted models are generally included in the high probability range of the results from the McMC-based method. Because the forward calculations fit very well with the input values for phase velocity and ZH ratio, the relatively larger V_S difference around the Moho interface between the true models and the inverted models can be attributed to the inherent nonuniqueness of the inversion problem. This applies to both JointNet and the McMC-based method. The joint inversion of dispersion and ZH ratio effectively helps to recover the velocity structures in both shallow and deep layers for both the methods. For a detailed comparison, we calculate the misfits between the V_S models inverted by these two methods and the true models, as well as the misfits between the forward calculations and the input values for phase velocity and ZH ratio. These results are presented in Figure 8.

In both cases, the misfits all follow a normal distribution centered around 0 km/s. Larger misfits are observed at depths of 0–5 km for both the methods, which is likely related to the velocity jump between the sediments and crystalline crust. These results indicate that JointNet can achieve inversion accuracy comparable to the conventional McMC-based method to a significant extent. The misfits of phase velocity and ZH ratio for JointNet are slightly larger than the McMC-based method. This is because the McMC-based method iteratively inverts the V_S model by fitting the input data, whereas JointNet directly fits the V_S model. Because the inverted models have similar accuracy, we consider this small difference to be acceptable.

Similarly, we also evaluated the performance of JointNet in practical applications using observed data. Huang *et al.* (2021) employed a DRAM-McMC inversion method to obtain 1D V_S models beneath 306 stations deployed in North China by

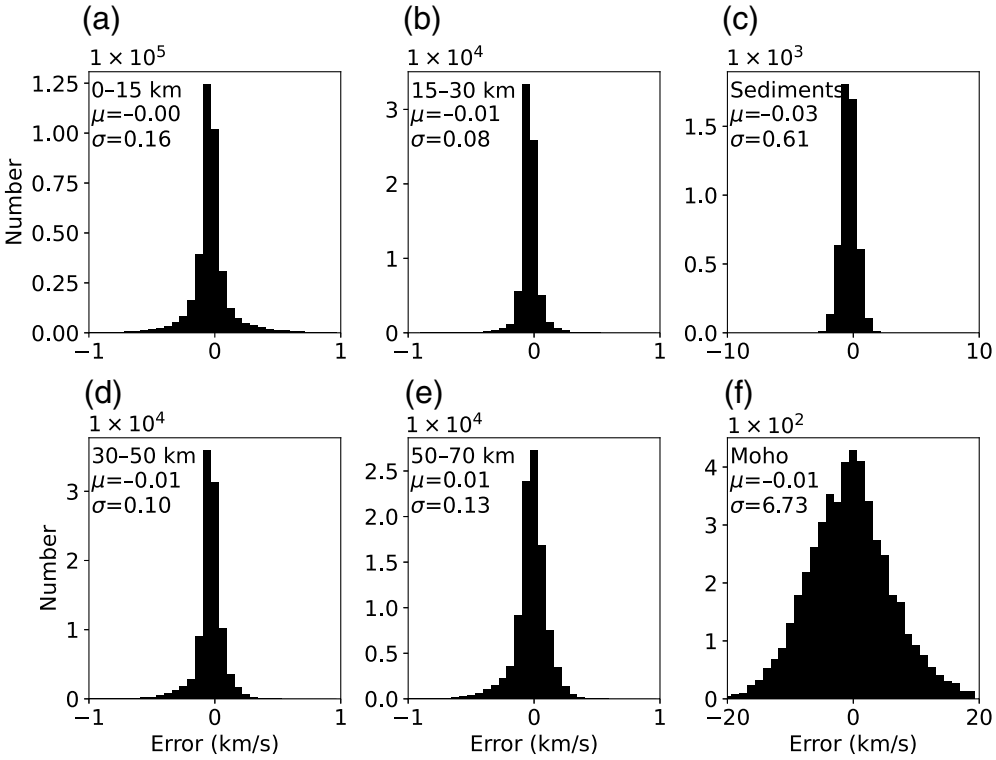


Figure 5. Distribution of the misfits between the inverted 1D models and the corresponding input models. (a,b,d,e) Distribution of the misfits of V_S at different depth ranges. (c,f) Distribution of the misfits of sedimentary and crustal thicknesses. The depth range or thickness type as well as the corresponding mean and standard deviation are annotated in the top-left corner of each panel.

jointly inverting the Rayleigh wave phase velocity and ZH ratio data. We feed JointNet with these observed datasets and compare the output with those from the McMC-based method. The comparison of the inverted 1D models and the data fitting is shown in Figure 9. The results exhibit a high level of consistency between JointNet and the McMC-based method, and the data fitting of both the methods is quite good. This indicates that JointNet is capable of obtaining reliable inversion results for actual observed data as well. Here, clearer Moho interfaces can be observed in the models obtained by the McMC-based method, because the crustal thickness from the previous study is input as an initial value during the inversion (Huang *et al.*, 2021).

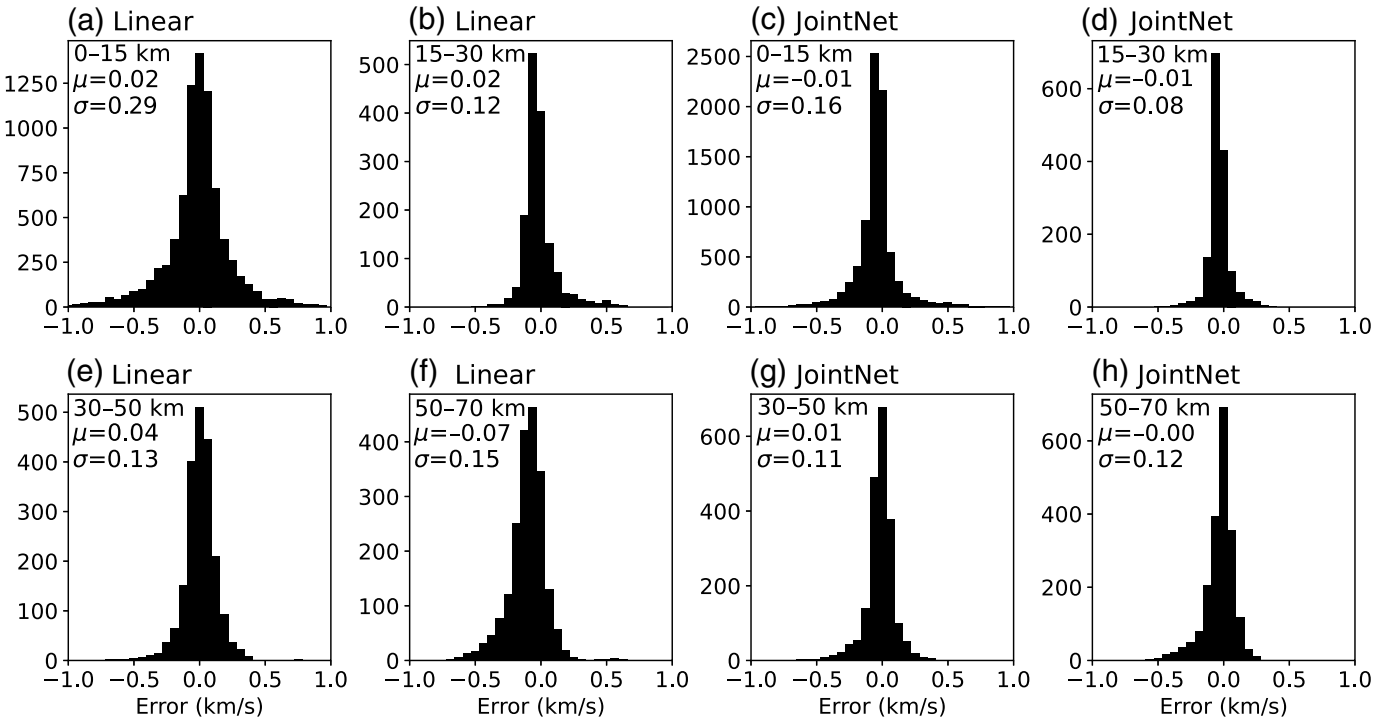
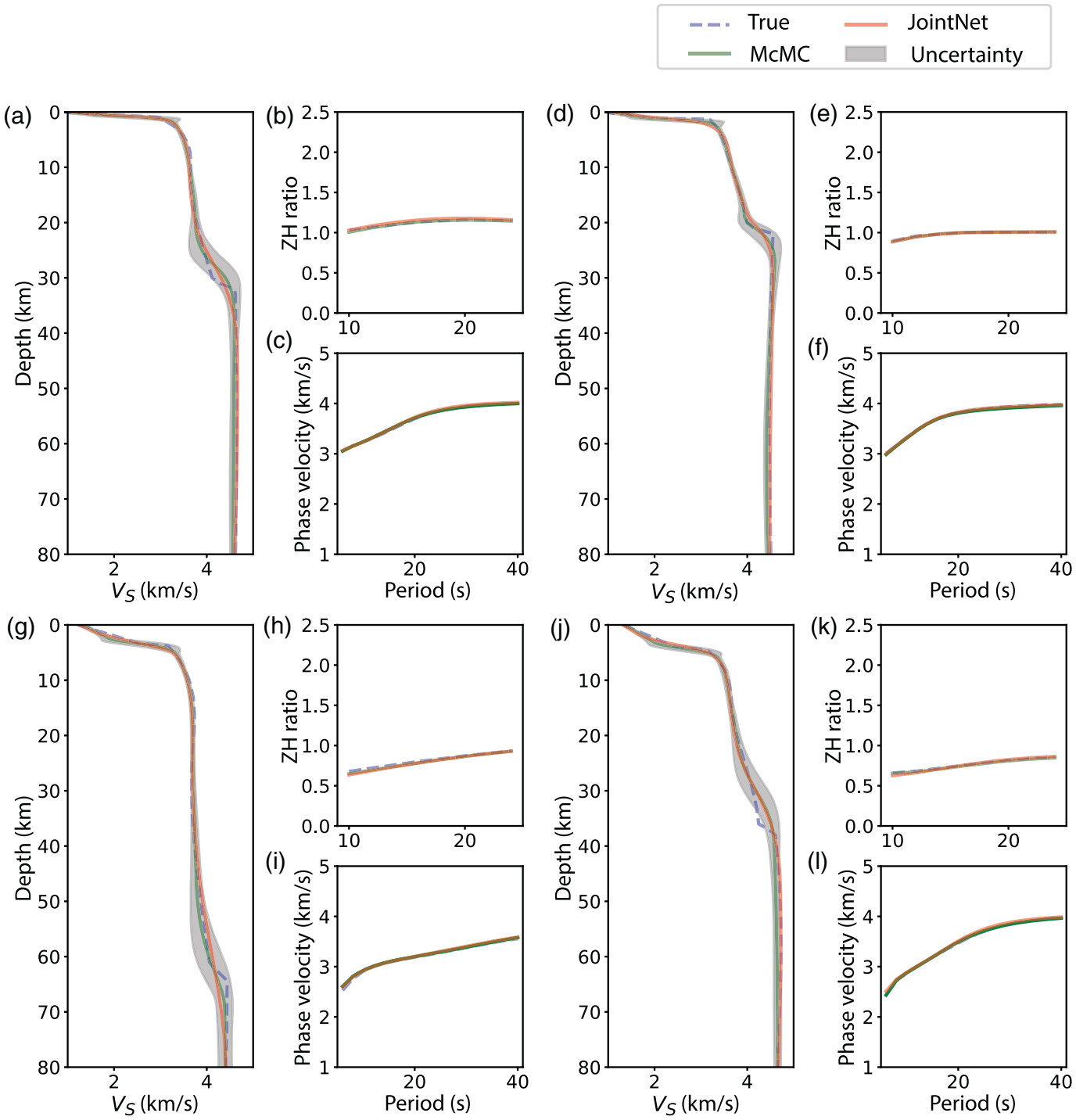


Figure 6. (a,b,e,f) Distribution of the misfits between the inverted models and the true models for the linear method. (c,d,g,h) The same as panels (a–d) but for JointNet.



Comparisons for horizontal V_S slices

In addition, we compare the horizontal V_S slices obtained from the observed data to validate whether JointNet can benefit from the joint usage of the two types of data and produce results spatially consistent with the McMC-based method. The horizontal V_S slices at 5, 10, 20, and 50 km depths obtained from JointNet and the McMC-based method (Huang *et al.*, 2021) are shown in Figure 10. The inclusion of ZH ratio data enhances the resolution of shallow structures, which is evident in the results of both methods. At 5 km depth, the

Figure 7. (a,d,g,j) Comparison of the 1D models inverted from synthetic dispersion and ZH ratio using JointNet and Markov chain Monte Carlo (McMC)-based method. The blue dashed lines represent the true models. The green and red solid lines represent the inverted models using the McMC-based method and JointNet, respectively. The width of the gray shading represents one standard deviation of uncertainty from the McMC-based method. (b,e,h,k) The blue dashed lines represent the calculated ZH ratio from the true models. The green and red solid lines represent the calculated ZH ratio from the inverted models using the McMC-based method and JointNet, respectively. (c,f,i,l) The same as panels (b, e, h, and k) but for phase velocity.

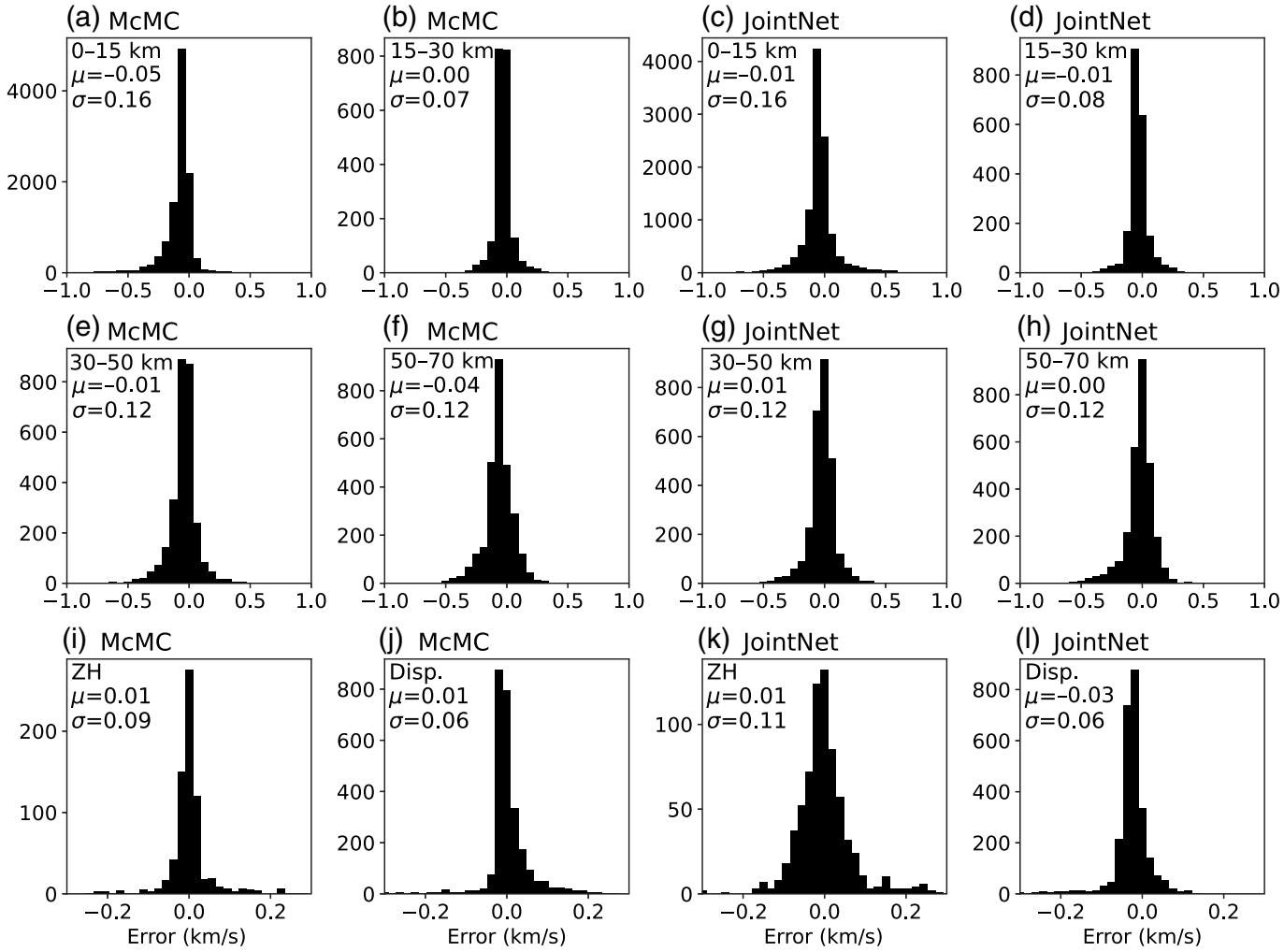


Figure 8. (a,b,e,f) Distribution of the misfits between the inverted models and the true models for the McMC-based method. (c,d,g,h) The same as panels (a,b,e,f) but for JointNet. Distribution of the misfits between the calculated (i) ZH ratio and (j) phase velocity dispersion from the inverted models using the McMC-based method and the input values. (k,l) The same as panels (i and j) but for JointNet.

results from both the methods reveal that the Huabei basin, Ordos basin, and Weihe-Shanxi Rift System generally exhibit low velocity, likely due to the presence of thick sediments at shallow depth. At 10 and 20 km depths, the velocity patterns gradually change, which can be observed in the results from both methods. The Ordos basin and Huabei basin gradually transition into high velocity, whereas the Yinshan Mountains and Taihang Mountains turn into low velocity. At 50 km depth, the low-velocity zone is centered around the Datong Volcano, and extends along the Yinshan and Taihang Mountains, indicating the extent of magmatism in the uppermost mantle. The results from JointNet and the McMC-based method demonstrate high similarity, proving the ability of JointNet to recover a high-resolution V_S model from phase velocity and ZH ratio data. The results at 10 km depth are slightly different due to the influence of the velocity interface at the sedimentary basement.

The vertical V_S profiles crossing the Datong volcano in both the east–west and north–south directions, obtained through these two methods, are shown in Figure 11. These vertical profiles also exhibit remarkable similarity and consistency with the geological settings. In summary, the comparisons of the 1D profiles and 2D slices both suggest that JointNet is a reliable method for crustal structure inversion and can be considered a viable alternative to conventional inversion methods.

Efficiency of the DNN method

Compared to the McMC-based methods, one notable advantage of JointNet is its significantly reduced inversion time (as demonstrated in Table 2). Specifically, for one sample and using the same computer equipment, the McMC-based method takes ∼1000 s for inversion (100,000 iterations), whereas JointNet only requires ∼0.005 s. That means JointNet achieves a speed improvement of ∼200,000 times. Even when accounting for training time, this method remains significantly more efficient than the McMC approach.

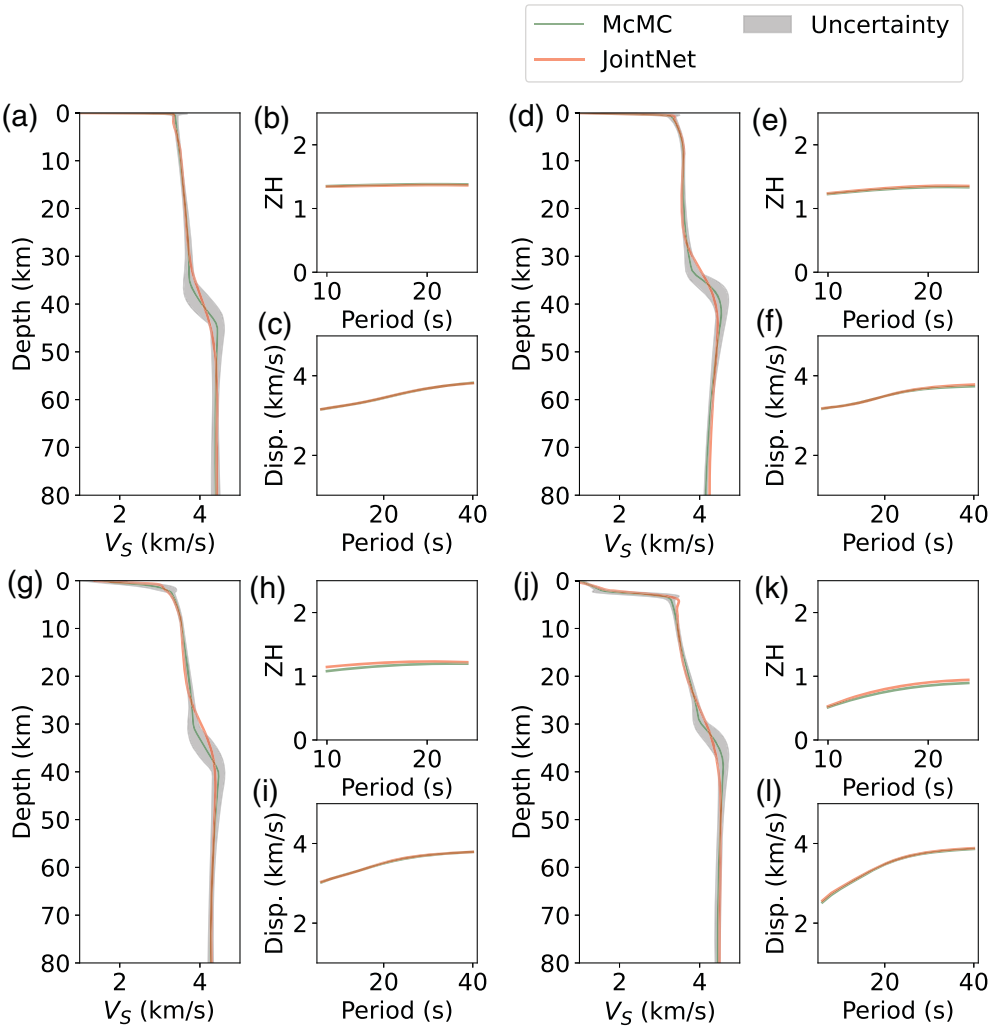


Figure 9. (a,d,g,j) Comparison of the 1D V_S models inverted from observed Rayleigh swave phase velocity and ZH ratio using JointNet (red lines) and the McMC-based method (green lines). The width of the gray shading represents one standard deviation of uncertainty from the McMC-based method. (b,e,h,k) The red and green lines represent the calculated ZH ratio from the inverted models obtained using JointNet and the McMC-based method. (c,f,i,l) The same as (b, e, h, and k) but for phase velocity dispersion.

McMC-based methods typically rely on trial-and-error approaches to obtain reliable inversion results by maximizing the posterior probability $p(m|x)$, in which m represents the velocity model, and x represents the observed data. In trial-and-error methods, the prior probability $p(m)$ is manually defined, and forward calculations are performed to determine the likelihood $p(x|m)$. However, calculating the $p(x|m)$ often demands substantial computing resources. Because of the requirement of discarding samples based on $p(m|x)$, the McMC-based methods utilize fewer samples for inversion. In contrast, the DNN-based method utilizes all available samples to construct the inversion function and directly fits the maximum posterior probability $p(m|x)$:

$$m = \mathrm{DNN}(x) \approx \arg\max_m p(m|x). \quad (3)$$

This allows for more efficient and comprehensive use of the dataset, reducing the overall inversion time. The DNN-based method only requires forward calculations during training, thus making it faster than traditional McMC-based inversion methods. The reduction in computational time is crucial, because the large computational cost is a common issue for nonlinear inversion methods.

Moreover, because the randomly generated models are all within a reasonable model space, the prior probability is already included in the training models. Therefore, JointNet eliminates the step of inputting prior probability. In addition, it minimizes the concern about getting trapped in local minima. It is capable of achieving accurate inversion results without the need for an initial model or prior information. This allows for a more flexible and unbiased inversion procedure to obtain structural information directly from the data.

TABLE 2
Comparison of Error and Speed between JointNet and the Markov Chain Monte Carlo (McMC)-Based Method

Methods	JointNet	McMC
Stand variation of error (km/s)	0.239	0.223
Inversion speed per 1000 inputs (s)	~5	~1,000,000

DISCUSSIONS

Our results demonstrate that JointNet is capable of producing reliable joint inversion results. In addition to its fast inversion speed, which is typically advantageous for DNN-based methods, JointNet is also a multimodal learning model (Srivastava and Salakhutdinov, 2014). Most of the existing DNN-based methods take individual input physical parameter, typically dispersion curve, for surface-wave tomography (e.g., Hu *et al.*, 2020; Cai *et al.*, 2022; Luo *et al.*, 2022). In contrast, JointNet

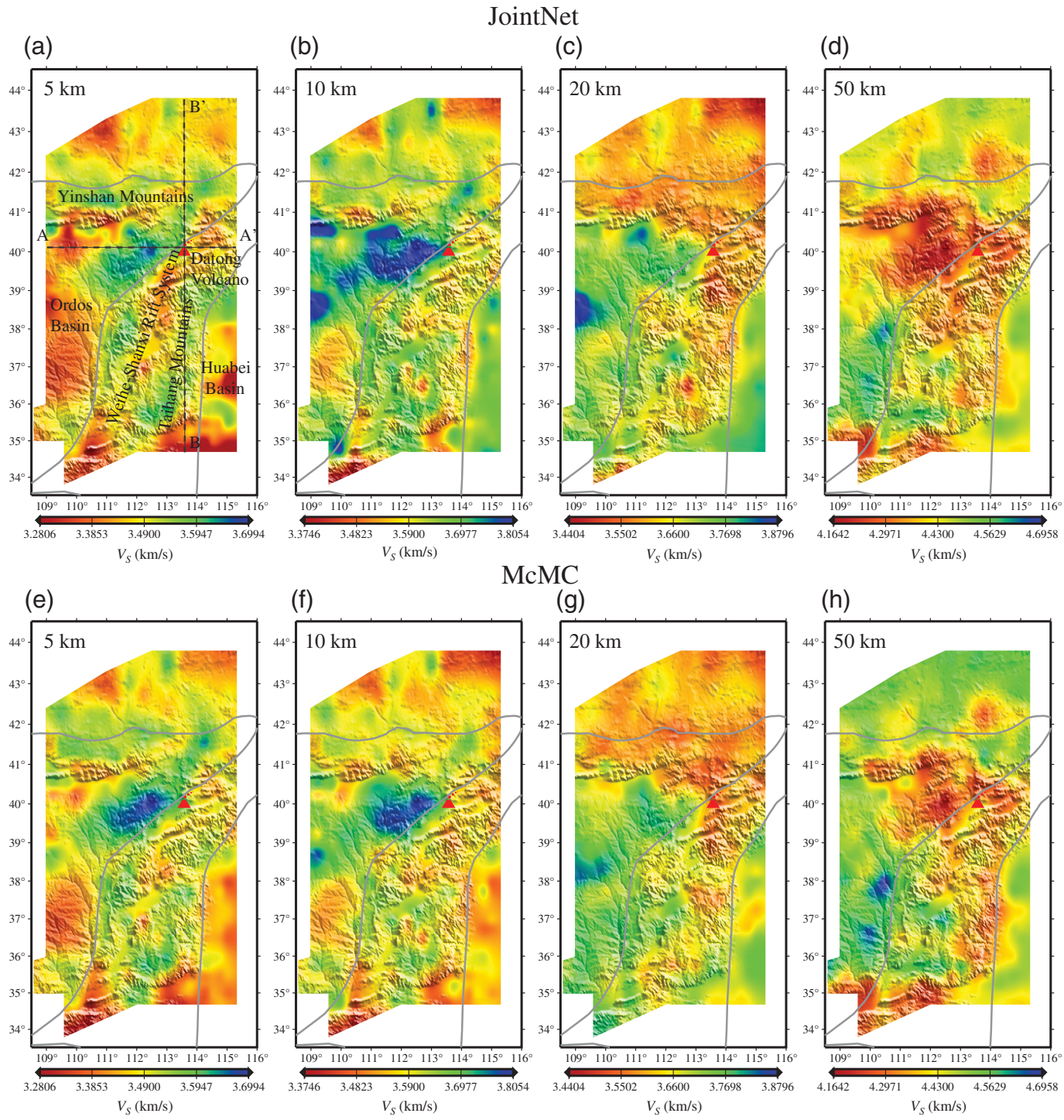


Figure 10. Horizontal V_S slices at 5–50 km depths from (a–d) JointNet and (e–h) the McMC-based method. The depth is labeled in the top-left corner of each panel. The tectonic units and the locations of the vertical profiles in Figure 11 are marked in panel (a).

accepts dispersion curves and ZH ratios as input, which can capture the information of two independent physical parameters. Furthermore, it produces multimodal outputs, including the 1D velocity model and the thicknesses of the sedimentary and crustal layers. This improves the capability of shallow structure imaging, which is an advantage that most existing networks do not have.

JointNet employs two separate transformer layers to represent and summarize the dispersion and ZH ratio data, and employs two fully connected layers to output the velocity and thickness information. This makes it highly adaptable to adding more physical quantities, such as raw waveform data. Therefore, it can act as a framework for duplicated joint inversion processing. This is also an innovation of JointNet compared to the existing networks.

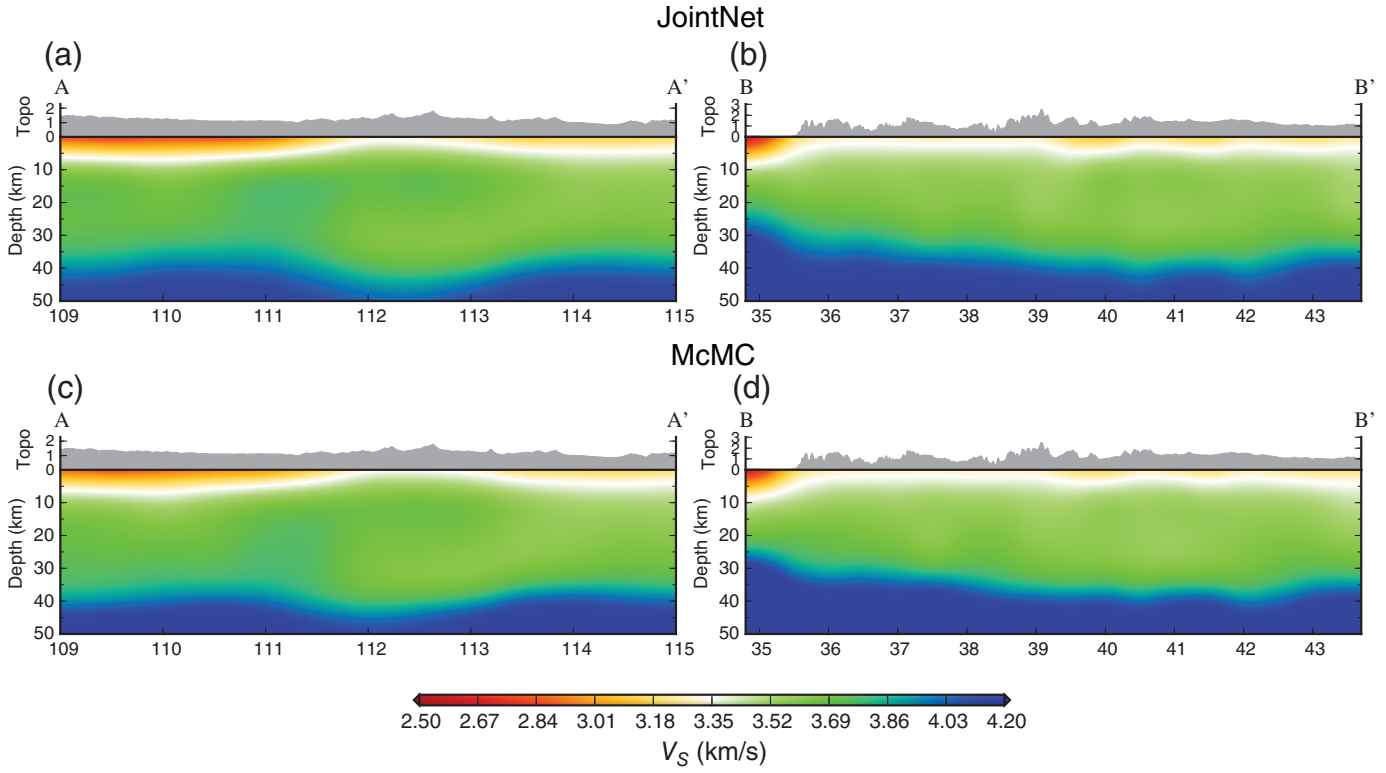


Figure 11. Vertical V_S profiles at 0–50 km depths from (a,b) JointNet and (c,d) the McMC-based method. The locations of these profiles are delineated with gray dashed lines in Figure 10a.

JointNet provides flexibility for joint inversion. Hu *et al.* (2020) proposed that deep learning-based inversion methods may underperform compared to conventional inversion methods if the training dataset does not adequately represent the real structures. JointNet is trained using a large number of synthetic datasets with the necessary physical constraints based on a reference model with reasonable perturbations. This feature, combined with the parameterization strategy, can ensure a broader model space. Therefore, it is capable of recovering the majority of 1D crustal models worldwide, eliminating the need for further training when applied to different regions. In addition, a sufficient number of random models helps to mitigate the issue of overfitting during neural network training.

JointNet also reduces the strict input format requirements. Most published inversion networks are based on CNN or DNN (Hu *et al.*, 2020; Luo *et al.*, 2022; Wang *et al.*, 2023), which have strict requirements for the input data format. An important difference between JointNet and the existing neural networks is that JointNet is a transformer-based model, which is suitable for processing sequence data. By leveraging the capabilities of the transformer model, it can learn comprehensive information about the dispersion curve and ZH ratio curve. Therefore, during the inversion process, the input dispersion and ZH ratio of JointNet can have different period ranges and intervals compared to the training data. We conducted experiments on this and found that even if we remove the dispersion points of two random periods for inversion, similar inversion results can still be obtained, as shown in Figure 12. However, when more than six dispersion points are removed, the inversion results may have large deviations. In general, the phase velocity within 6–40 s periods and the ZH ratio within 10–24 s periods are easy to obtain using ambient noise tomography based on regional seismic arrays. However, in certain situations, the period ranges of reliable dispersion and/or ZH ratio may be different. In such cases, this loose input requirement of JointNet provides additional flexibility, making it applicable to most regional-scale crustal imaging studies.

By including both the dispersion and ZH ratio, the imaging resolution of shallow structures is significantly improved. This is particularly beneficial in basins with thick sediments. The presence of shallow low-velocity layers can have a substantial impact on the inversion of crustal velocity structure (Yang *et al.*, 2019). Our results show that JointNet can reliably delineate the structures of the sedimentary layer. However, due to the nature of the surface wave, it cannot provide ideal constraints for velocity discontinuity. The velocity values around the bottom of the sedimentary layer and Moho might have larger errors. The same is true for traditional methods such as the McMC-based method. To address this issue, one possible solution is to incorporate the receiver function or teleseismic *P* waveform data into the joint inversion (Li *et al.*, 2019; Berg *et al.*, 2020). Because JointNet is a multimodal learning model, it should be easily adapted to include a new input dataset, such as the receiver function, to

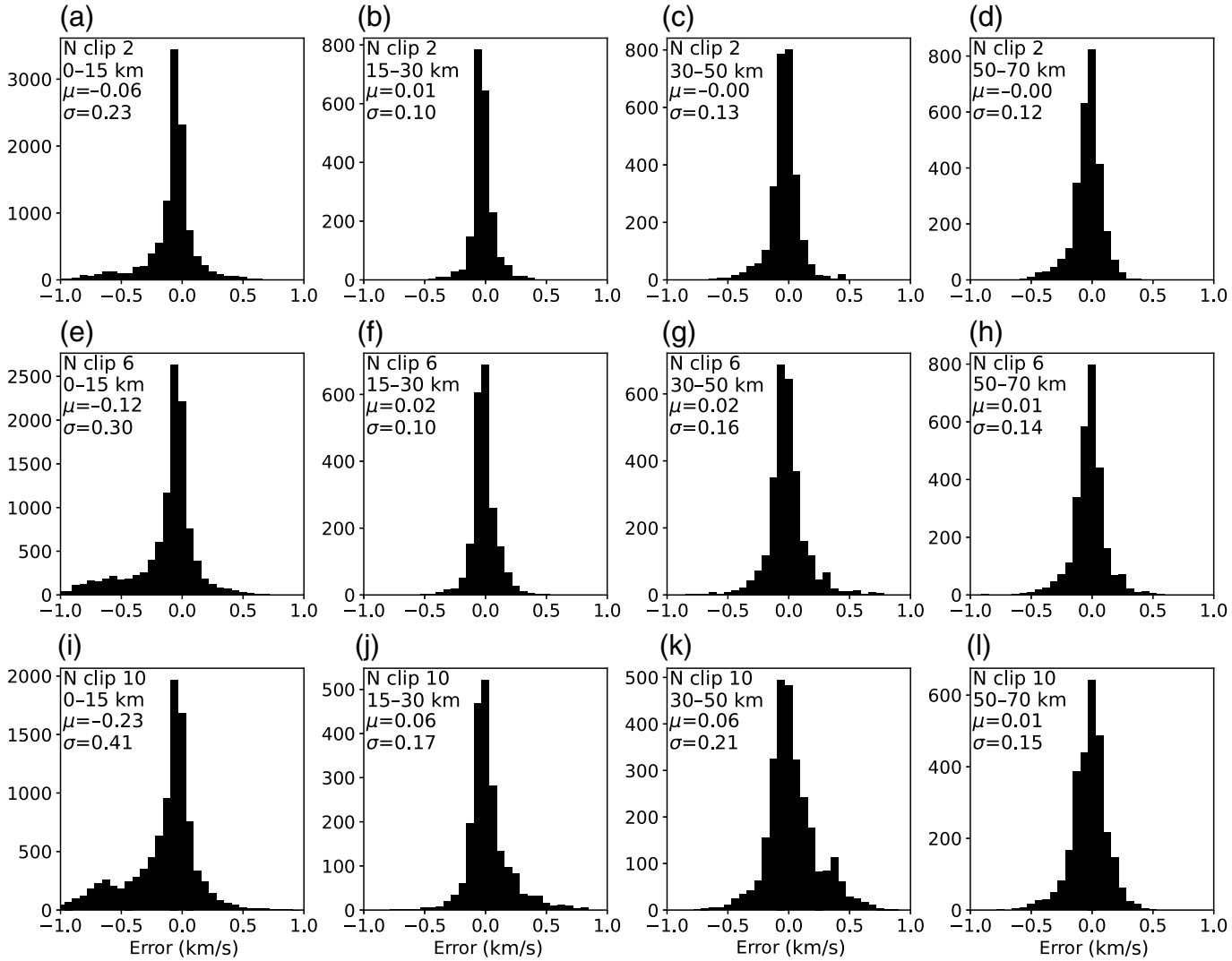


Figure 12. (a–d) Distribution of the misfits between the inverted models from JointNet and the true models. We randomly removed two dispersion points for inversion. The number of removed dispersion points, depth range, mean, and standard deviation are annotated in the top-left corner of each panel. (e–h) The same as panels (a–d) but for six removed dispersion points. (i–l) The same as panels (a–d) but for 10 removed dispersion points.

improve constraints on velocity discontinuities. This would be further systematic research.

In this study, we use JointNet to invert the structures of the crust and uppermost mantle. For studies with significantly different lateral and/or vertical scales, such as deep mantel imaging using surface waves with a longer period, direct application of JointNet may not be feasible. In such cases, it would be necessary to modify the parameters and retrain the DNN model. This process is relatively straightforward to implement.

Because DNN directly fits the maximum value of the posterior probability, it outputs a 1D model without the posterior probability distribution (Hu *et al.*, 2020; Cai *et al.*, 2022; Luo *et al.*, 2022). Thanks to the efficiency of JointNet, we assessed the uncertainty of the inversion results by performing inversions on 100 datasets that were augmented with random Gaussian noise. The mean and standard deviation of the V_S models obtained from these 100 inverted datasets are considered as the final inversion results and uncertainty, respectively. Figure 13 demonstrates that the uncertainty in the inverted models increases as the noise level is elevated. In practical applications, datasets with larger errors will result in greater uncertainty in the inverted models. It can help to evaluate the reliability of the inverted models, similar to the McMC-based method.

In addition, JointNet still has certain limitations in terms of the training models and inverting results. For instance, the 1D models are parameterized with B-splines to present V_S in the crust and upper mantle, respectively. It can produce relatively smooth models, but, at the same time, it may lose some interface information inside the crust or upper mantle. In addition, our use of empirical relationships to calculate *P*-wave velocity and density may lead to deviations between the synthetic and real models. Because our model parameterization strategy is based on the previous research, we did not evaluate the appropriateness of the parameterization in this study. Although our inversion tests show that these issues have a very small impact on the

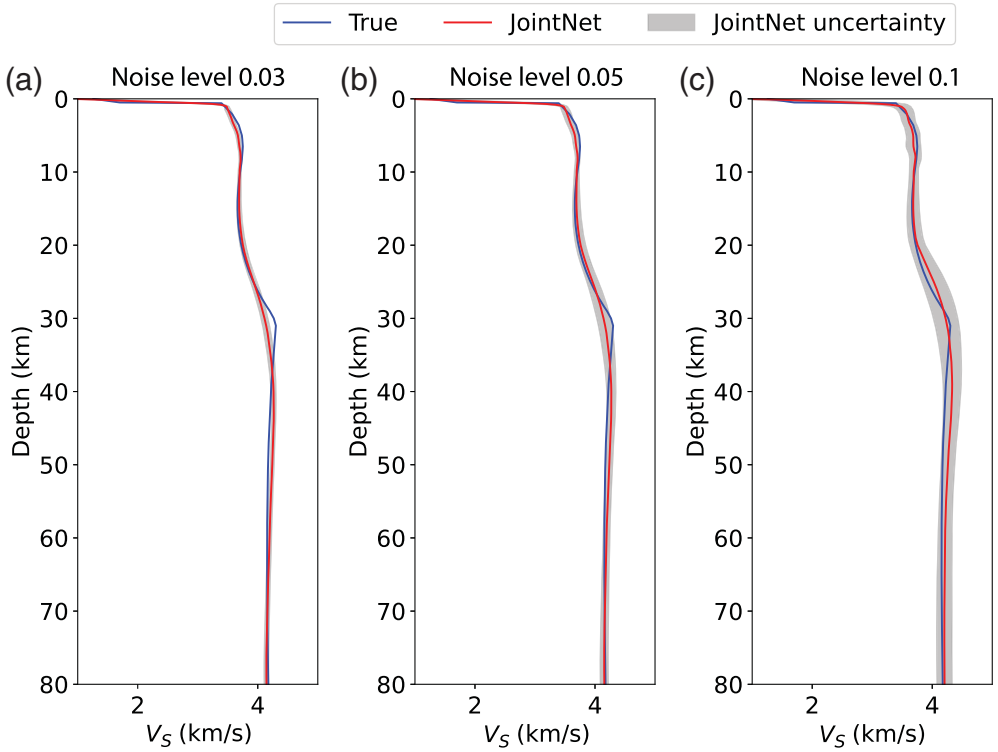


Figure 13. Comparison of the inversions results from input data with different levels of noise. The blue and red solid lines represent the true models and inverted models from JointNet, respectively. The width of the gray shading represents one standard deviation of uncertainty. (a) The noise levels are 0.03 km/s for phase velocity and 0.03 for ZH ratio. (b) The same as panel (a) but for a noise level of 0.05 km/s and 0.05. (c) The same as panel (a) but for a noise level of 0.1 km/s and 0.1.

results, it may still be worth further improvement. Another limitation of JointNet is the relatively larger errors in the inverted crustal thickness, which could potentially be addressed by incorporating receiver function data in the future research.

CONCLUSIONS

We develop a new multimodal network, JointNet, to jointly invert Rayleigh wave phase velocity and ZH ratio for the 1D V_S model. It can produce consistent and stable results compared to the conventional McMC-based method but with a much higher inversion speed. Its network structure alleviates the requirement for fixed-period inputs, and can be easily trained and extended to include more input types. Therefore, JointNet can serve as a joint inversion framework for regional crustal structure investigation, with the characteristics of simple training, fewer prior assumptions, and faster inversion speed compared to conventional nonlinear methods.

DATA AND RESOURCES

The CRUST1.0 model can be obtained from the IRIS Data Management System (http://ds.iris.edu/ds/products/emc-crust10, last accessed August 2023). The Rayleigh wave phase velocity dispersion and ZH ratio data used in this study are derived from Huang *et al.* (2021) and accessible at DOI: 10.17632/9tcj862hrh.2.

DECLARATION OF COMPETING INTERESTS

The authors acknowledge that there are no conflicts of interest recorded.

ACKNOWLEDGMENTS

The authors thank the two anonymous reviewers and Associate Editor Ricardo Taborda for their insightful comments that helped to improve this article. The JointNet is implemented using the deep learning framework of PyTorch. This work was supported by the National Key R&D Program of China (Grant 2021YFC3000702), the Special Fund of the Institute of Geophysics, the China Earthquake Administration (Grant DQJB22R34), and the National Natural Science Foundation of China (U2239206, 42004046).

REFERENCES

Afonso, J. C., J. Fullea, W. L. Griffin, Y. Yang, A. G. Jones, D. J. A. Connolly, and S. Y. O'Reilly (2013). 3-D multiobservable probabilistic inversion for the compositional and thermal structure of the lithosphere and upper mantle I: A priori petrological information and geophysical observables, *J. Geophys. Res.* **118,** no. 5, 2586–2617, doi: 10.1002/jgrb.50124.

Berg, E. M., F.-C. Lin, A. Allam, V. Schulte-Pelkum, K. M. Ward, and W. Shen (2020). Shear velocity model of Alaska via joint inversion of Rayleigh wave ellipticity, phase velocities, and receiver functions across the Alaska transportable array, *J. Geophys. Res.* **125,** e2019JB018582, doi: 10.1029/2019JB018582.

Birch, F. (1961). The velocity of compressional waves in rocks to 10 kilobars (Part II), *J. Geophys. Res.* **66,** no. 7, 2199–2224, doi: 10.1029/JZ066i007p02199.

Bodin, T., M. Sambridge, c. H. Tkalcˇ, P. Arroucau, K. Gallagher, and N. Rawlinson (2012). Transdimensional inversion of receiver functions and surface wave dispersion, *J. Geophys. Res.* **117,** no. B02301, doi: 10.1029/2011JB008560.

Brocher, T. M. (2005). Empirical relations between elastic wavespeeds and density in the Earth's crust, *Bull. Seismol. Soc. Am.* **95,** no. 6, 2081–2092, doi: 10.1785/0120050077.

Cai, A., H. Qiu, and F. Niu (2022). Semi-supervised surface wave tomography with Wasserstein cycle-consistent GAN: Method and application to southern California plate boundary region, *J. Geophys. Res.* **127,** e2021JB023598, doi: 10.1029/2021JB023598.

Dai, T., J. Xia, L. Ning, C. Xi, Y. Liu, and H. Xing (2021). Deep learning for extracting dispersion curves, *Surv. Geophys.* **42,** 69–95.

Devilee, R. J. R., A. Curtis, and K. Roy-Chowdhury (1999). An efficient, probabilistic neural network approach to solving inverse

problems: Inverting surface wave velocities for Eurasian crustal thickness, *J. Geophys. Res.* **104,** no. B12, 28,841–28,857.

Forsyth, D. W., S. C. Webb, L. R. M. Dorman, and Y. Shen (1998). Phase velocities of Rayleigh waves in the MELT experiment on the East Pacific Rise, *Science* **280,** no. 5367, 1235–1238.

Guo, Z., Y. J. Chen, J. Ning, Y. Feng, S. P. Grand, F. Niu, H. Kawakatsu, S. Tanaka, M. Obayashi, and J. Ni (2015). High resolution 3-D crustal structure beneath NE China from joint inversion of ambient noise and receiver functions using NECESSArray data, *Earth Planet. Sci. Lett.* **416,** 1–11.

Herrmann, R. B. (2013). Computer programs in seismology: An evolving tool for instruction and research, *Seismol. Res. Lett.* **84,** no. 6, 1081–1088, doi: 10.1785/0220110096.

Hu, J., H. R. Qiu, H. J. Zhang, and Y. Ben-Zion (2020). Using deep learning to derive shear wave velocity models from surface-wave dispersion data, *Seismol. Res. Lett.* **91,** no. 3, 1738–1751, doi: 10.1785/0220190222.

Huang, X., Z. Ding, J. Ning, F. Niu, G. Li, X. Wang, and X. Xu (2021). Sedimentary and crustal velocity structure of Trans-North China Orogen from joint inversion of Rayleigh wave phase velocity and ellipticity and some implication for Syn-rift volcanism, *Tectonophysics* **819,** 229104.

Huang, X., W. Wang, and Z. Ding (2023). Sedimentary and crustal structure of eastern North China from joint inversion of Rayleigh wave phase velocity and ellipticity and its implication for magmatism, *Tectonophysics* **846,** 229680.

Huang, Z. X., H. Y. Li, Y. J. Zheng, and Y. J. Peng (2009). The lithosphere of North China Craton from surface wave tomography, *Earth Planet. Sci. Lett.* **288,** 164–173.

Laske, G., G. Masters, Z. Ma, and M. Pasyanos (2013). Update on CRUST1.0 - A 1-degree global model of earth's crust, *Geophys. Res. Abstr.* **15,** Abstract EGU2013-2658, doi: 10.6092/1970-9870/128.

Li, G., H. Chen, F. Niu, Z. Guo, Y. Yang, and J. Xie (2016). Measurement of Rayleigh wave ellipticity and its application to the joint inversion of high-resolution S wave velocity structure beneath northeast China, *J. Geophys. Res.* **121,** no. 2, 864–880, doi: 10.1002/2015JB012459.

Li, G., F. Niu, Y. Yang, and K. Tao (2019). Joint inversion of Rayleigh Wave phase velocity, particle motion, and teleseismic body wave data for sedimentary structures, *Geophys. Res. Lett.* **46,** no. 12, 6469–6478, doi: 10.1029/2019GL082746.

Lin, F. C., B. Schmandt, and V. C. Tsai (2012). Joint inversion of Rayleigh wave phase velocity and ellipticity using USArray: Constraining velocity and density structure in the upper crust, *Geophys. Res. Lett.* **39,** no. 12, doi: 10.1029/2012GL052196.

Luo, Y., Y. Huang, Y. Yang, K. Zhao, X. Yang, and H. Xu (2022). Constructing shear velocity models from surface wave dispersion curves using deep learning, *J. Appl. Geophys.* **196,** 104524.

Meier, U., A. Curtis, and J. Trampert (2007). Global crustal thickness from neural network inversion of surface wave data, *Geophys. J. Int.* **169,** no. 2, 706–722.

Perol, T., M. Gharbi, and M. Denolle (2018). Convolutional neural network for earthquake detection and location, *Sci. Adv.* **4,** no. 2, e1700578.

Press, F. (1956). Determination of crustal structure from phase velocity of Rayleigh waves part I: southern California, *Geol. Soc. Am. Bull.* **67,** no. 12, 1647–1658.

Ross, Z. E., M. A. Meier, and E. Hauksson (2018). P wave arrival picking and first-motion polarity determination with deep learning, *J. Geophys. Res.* **123,** no. 6, 5120–5129.

Shapiro, N. M., and M. Campillo (2004). Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise, *Geophys. Res. Lett.* **31,** L07614, doi: 10.1029/2004GL019491.

Shen, W., M. H. Ritzwoller, and V. Pelkum Schulte (2013). A 3-D model of the crust and uppermost mantle beneath the central and western US by joint inversion of receiver functions and surface wave dispersion, *J. Geophys. Res.* **118,** no. 1, 262–276, doi: 10.1029/2012JB009602.

Srivastava, N., and R. Salakhutdinov (2014). Multimodal learning with deep Boltzmann machines, *J. Machine Learn. Res.* **15,** no. 1 (January 2014), 2949–2980.

Tous, R., L. Alvarado, B. Otero, L. Cruz, and O. Rojas (2020). Deep neural networks for earthquake detection and source region estimation in north-central Venezuela, *Bull. Seismol. Soc. Am.* **110,** no. 5, 2519–2529.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, *et al.* (2017). Attention is all you need, in *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Editors), Curran Associates Inc., Red Hook, New York, 30.

Wang, F. Y., X. D. Song, and M. K. Li (2023). A deep-learning-based approach for seismic surface-wave dispersion inversion (SfNet) with application to the Chinese mainland, *Earthq. Sci.* **36,** no. 2, 147–168, doi: 10.1016/j.eqs.2023.02.007.

Wang, J., Z. Xiao, C. Liu, D. Zhao, and Z. Yao (2019). Deep-learning for picking seismic arrival times, *J. Geophys. Res.* **124,** no. 7, 6612–6624.

Wang, Z., C. Sun, and D. Wu (2021). Automatic picking of multimode surface-wave dispersion curves based on machine learning clustering methods, *Comput. Geosci.* **153,**104809.

Yang, C., G. Li, F. Niu, and Y. Ben-Zion (2019). Significant effects of shallow seismic and stress properties on phase velocities of Rayleigh waves up to 20 s, *Pure Appl. Geophys.* **176,** no. 3, 1255–1267.

Yang, S., H. Zhang, N. Gu, J. Gao, J. Xu, J. Jin, J. L. Li, and H. J. Yao (2022). Automatically extracting surface-wave group and phase velocity dispersion curves from dispersion spectrograms using a convolutional neural network, *Bull. Seismol. Soc. Am.* **93,** no. 3, 1549–1563.

Yu, Y., J. Lin, L. Zhang, G. Liu, J. Hu, Y. Tan, and H. Zhang (2018). Identification of seismic wave first arrivals from earthquake records via deep learning, *Int. Conf. on Knowledge Science, Engineering and Management*, Springer, Cham, Switzerland, 274–282.

Zhang, P., and H. Yao (2017). Stepwise joint inversion of surface wave dispersion, Rayleigh wave ZH ratio, and receiver function data for 1D crustal shear wave velocity structure, *Earthq. Sci.* **30,** 229–238.

Zhang, X., F. Hansteen, A. Curtis, and S. deRidder (2020). 1D, 2D and 3D Monte Carlo ambient noise tomography using a dense passive seismic array installed on the North Sea seabed, *J. Geophys. Res.* **125,** e2019JB018552, doi: 10.1029/2019JB018552.

Manuscript received 10 August 2023
Published online 19 December 2023