Seismological Research Letters

Using deep learning to derive shear wave velocity models from surface wave dispersion data

--Manuscript Draft--

Manuscript Number:	SRL-D-19-00222
Full Title:	Using deep learning to derive shear wave velocity models from surface wave dispersion data
Article Type:	Article - Regular Section
Corresponding Author:	Haijiang Zhang, Ph.D. University of Science and Technology of China Hefei, Anhui CHINA
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Science and Technology of China
Corresponding Author's Secondary Institution:	
First Author:	Jing Hu
First Author Secondary Information:	
Order of Authors:	Jing Hu
	Hongrui Qiu
	Haijiang Zhang, Ph.D.
	Yehuda Ben-Zion
Order of Authors Secondary Information:	
Manuscript Region of Origin:	CHINA
Suggested Reviewers:	Michael Bianco mbianco@ucsd.edu
	Bertrand Rouet-Leduc bertrandrl@lanl.gov
	Greg Beroza beroza@stanford.edu
	Peter Gerstoft gerstoft@ucsd.edu
Opposed Reviewers:	

Using deep learning to derive shear wave velocity models from surface wave dispersion data

Jing Hu[1], Hongrui Qiu[2], Haijiang Zhang[1*], Yehuda Ben-Zion[2]

1. Laboratory of Seismology and Physics of Earth's Interior; School of Earth and Space Sciences, University of Science and Technology of China, Hefei, Anhui, China
2. Department of Earth Sciences, University of Southern California, Los Angeles, CA 90089-0740, USA

*Corresponding email: zhang11@ustc.edu.cn

Seis. Res. Lett.

In review, 2019

Keyword: shear wave velocity, dispersion curves, deep learning, convolutional neural network.

Abstract

We present a new algorithm for derivations of 1-D shear wave velocity models from surface wave dispersion data using convolutional neural networks (CNNs). The technique is applied for the continental China and the plate boundary region in Southern California. Different CNNs are designed for these two regions and are trained using theoretical Rayleigh wave phase and group velocity images computed from reference 1-D Vs models. The methodology is tested with 3260 phase-group images for continental China and 4160 phase-group images for Southern California. The conversions of these images to velocity profiles take ~23 s for continental China and ~30 s for Southern California on a personal laptop with the NVIDIA GeForce GTX 1060 core and a memory of 6 GB. The results obtained by the CNNs show high correlation with previous studies using conventional methods. The effectiveness of the CNN technique makes this fast method an important alternative for deriving shear wave velocity models from large data sets of surface wave dispersion data.

1. Introduction

Surface wave tomography has been widely used to image Earth structures at various scales in different tectonic regions. Surface wave dispersion curves are utilized mainly to determine shear wave speed (Vs) models, but are also sensitive to density and compressional velocity (Vp) models (Liu et al., 2018; Curtis et al., 1998; Lin et al., 2008; Zhou et al., 2006). In general, surface wave tomography adopts a two-step approach: group or phase velocity maps at different periods are first determined and then a series of 1-D Vs models beneath each grid cell are inverted using phase and/or group velocity values at that node. A linearized inversion approach requires selecting optimum regularization values and an appropriate initial velocity model to stabilize the inversion (e.g. Herrmann, 2013). In addition, the data volume of phase and group velocity measurements are becoming very large with more and denser deployments of local and regional seismic arrays (e.g. Lin et al. 2012, Ben-Zion et al., 2015) and surface waves extracted from natural earthquakes (e.g. Yang et al., 2008), ambient noise (e.g. Shapiro et al., 2005) and artificial sources (e.g. She et al., 2018). This makes classical 1-D Vs inversions very time-consuming. Nonlinear methods based on the random sampling (Mosegaard et al., 1995; Sambridge, 1999a, b) have been proposed to directly invert 1D Vs models from dispersion curves. This can avoid selecting regularization parameters but may yield biased solutions due to arbitrary sampling and could be time consuming for the optimization process. Another alternative is to use artificial neural network (Devilee et al., 1999; Meier et al., 2007). Compared to the conventional linear or nonlinear methods, once the network is trained, it can be used to map the 1D Vs model directly from surface wave dispersion measurements without inversion.

In recent years, deep learning techniques, and especially convolutional neural network (CNN) algorithms, have shown significant potential in various seismological applications including event detection (Perol et al., 2018; Yu et al., 2018), phase picking (Zhu et al., 2018; Ross et al., 2018; Wang et al., 2019), earthquake early-warning (Li, et al., 2018), first-motion polarity determination (Ross et al., 2018) and seismic phase association (McBrearty et al., 2018; Ross et al., 2019). Only a few studies applied neural networks to surface wave tomography (Devilee et al., 1999; Meier et al., 2007; Cheng et al. 2019). As noted by Meier et al. (2007), there are typically three major steps to solve the inverse problem with the neural network method (Fig. 1). (1) Assemble a large amount of 1-D Vs models (labels) and the corresponding phase and group velocity dispersion curves (inputs) for training the network. (2)

Design a neural network structure which takes phase and group velocities as inputs and outputs 1-D Vs models. (3) Train the designed neural network by minimizing the differences between its outputs and the labels. Once a neural network is trained, the best fitting 1-D Vs models are predicted based on the Rayleigh wave phase and group velocity dispersion curves.

In this study, we present two CNNs that are used to preform surface wave tomography for two different regions, China and Southern California. Our analyses are different from previous studies using neural networks (e.g. Meier et al., 2007 and Cheng et al., 2019) in three main aspects. (1) We use CNNs rather than the shallow neural network utilized in Meier et al. (2007), which can deal with more complicated nonlinear inverse problems. (2) We construct 1-D Vs models using finer layers (0.5 km layer thickness), whereas in Meier et al. (2007) and Cheng et al. (2019) the Vs models involve only five major layers (a sedimentary layer, three crustal layers, and an upper mantle layer). (3) We employ as inputs both phase and group velocities as Meier et al. (2007), whereas Cheng et al. (2019) utilized only phase velocities.

The reminder of the paper is organized as follows. In Section 2, we describe the methodology and demonstrate the process using an example training dataset from the central western USA. This includes data preprocessing steps, the CNN architecture, and a training process. In Section 3, we apply the method to datasets obtained from continental China (Section 3.1) and Southern California (Section 3.2). For the application in Southern California, we use a CNN with a slightly different architecture from the one illustrated in Section 2, which is trained using a dataset generated based on the regional model of Shaw et al. (2015). In section 4, we discuss and summarize the results.

2. Methodology

We utilize one of the most widely used deep learning algorithms, the CNN, to directly invert surface wave phase and group velocity dispersion curves for isotropic 1-D Vs models. First, we take a set of derived 1-D Vs models and calculate corresponding theoretical Rayleigh wave phase and group velocity dispersion curves and corresponding images as the training dataset (preprocessing step in Fig. 2). Then, a designed CNN takes pairs of phase-group dispersion images as inputs and provides outputs 1-D Vs profiles. The differences between the predicted Vs models and corresponding Vs models labeled in the training dataset are minimized to train the CNN (updating the weights of the CNN). The trained CNN can be used to quickly map large amounts of Rayleigh wave phase and group dispersion curves to 1-D Vs models. Since the output depth range needs to be changed with the period range of the input

Rayleigh wave velocity dispersions, we use two different CNNs for the applications in Section 3. The period ranges are 8-50s for continental China at depths of 0-150km (Section 3.1) and 2.5-16s for Southern California at depths of 0-49.5km (Section 3.2). For a brief introduction of the data preparation, neural network architecture, and training process, we use a training dataset generated from surface wave tomography results for the central western USA of Shen et al. (2013) and a test dataset of Rayleigh wave velocity dispersions for the continental China from Shen et al. (2016).

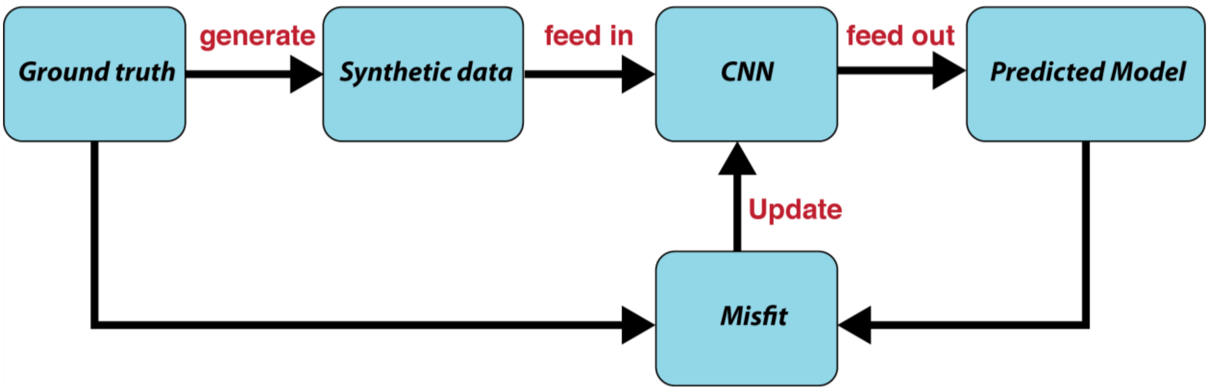

Figure 1. The workflow for data-driven inversion scheme. The ground truth is one of the 1-D Vs models used as a label in the training process. Synthetic data are the corresponding phase and group velocity dispersion curves. CNN stands for the convolutional neural network (the architecture shown in Fig. 2).

2.1 Data preparation and preprocessing

The data diversity is important for training a neural network (Deng et al., 2009). We first extract 6803 1-D Vs models for the central western USA from the surface to a depth of 150 km with each layer thickness of 0.5 km (Fig. S1) based on the surface wave tomography of Shen et al. (2013). Corresponding Vp models are computed above 120 km from Vs using the relationship of Brocher (2005) and a fixed Vp/Vs=1.79 from 120 km to 150 km (Kennett et al., 1995). Density is also computed from Vs following the empirical relation of Brocher (2005). We then generate corresponding theoretical Rayleigh wave phase and group velocity dispersion curves for periods in the range 8-50s (8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 35, 40, and 50 s) via the Computer Programs in Seismology (CPS) software package (Herrmann, 2013). Considering the good performance of CNNs on image processing, we transform the dispersion curves to energy images (preprocessing part in Fig. 2) by allowing for uncertainties via a Gaussian function $g_T(v_0)$,

$$g_T(v_0) = \mathrm{e}^{-(v-v_0)^2/r} \quad (1)$$

where v_0 is the value of phase velocity or group velocity at period T, v is a constant array of 60 elements that varies from 2 km/s to 5 km/s with a spacing of 0.05 km/s and r is the radius of the Gaussian function and representative of the estimated uncertainty. The velocity range and spacing can be changed according to the specific training dataset, and here we set r as 0.1 km/s for a rough estimate of the uncertainty in the dispersion curves. After converting dispersion curves into images, we obtain 6803 pairs of phase-group dispersion images with a dimension of 60x17 (height = 60; width = 17).

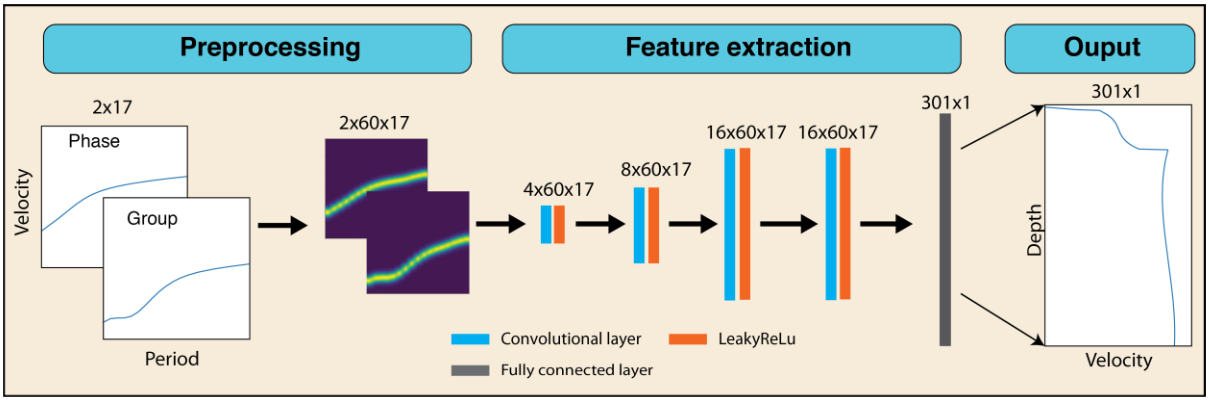

Figure 2. Data preprocessing and the architecture of the convolutional neural network (used in Section 3.1). In Section 3.2, the number of output elements of CNN is changed from 301 to 99 but other parameter setups are the same.

2.2 The CNN Architecture

The CNN used in this study comprises four convolutional layers and one fully connected layer (Fig. 2). The CNN has 2 input channels that take phase and group dispersion images and outputs a best fitting 1-D Vs model. For the continental China case, the input images have dimensions of 60x17 and the output 1-D Vs profile is discretized into 301 layers. The numbers of filters at each convolutional layer are, from shallow to deep, 4, 8, 16, and 16. For each convolutional layer, we set the kernel size equal to 3 and stride equal to 1, and apply a zero-padding operation in each convolutional layer. To further avoid the vanishing gradient problems, activation function LeakyReLu ($f(x) = x, if\ x > 0;\ otherwise\ 0.01x$) is applied at each activation layer (Maas et al., 2013) located right after the convolutional layer (red bars in Fig. 2). Since the inputs dimension is small, we do not employ a pooling layer that is often included in a conventional CNN architecture.

Mapping phase-group dispersion images to 1-D Vs profiles belongs to the category of regression problem. The weights of the neural network are optimized to minimize the mean squared error (MSE) using the training dataset. The loss function used to train the CNN is defined as:

$$J = \left\| \boldsymbol{m}_p - \boldsymbol{m}_t \right\|_2^2 \qquad (2)$$

where $\boldsymbol{m}_p$ and $\boldsymbol{m}_t$ are the predicted and ground-truth labels, respectively.

2.3 Training

We randomly split 80% of the whole dataset as the training dataset and 20% as the validation dataset. The validation dataset is excluded from the training process and only used to guide parameter tuning and avoid overfitting. The maximum number of epochs is set to 600 to ensure the training process converges. For each epoch, we randomly shuffle the whole input dataset to decrease the risk of creating batches that are not representative of the overall training dataset. We use the adaptive moment estimation (Adam) optimizer with a learning rate of 1e-5 and other parameters set by default to minimize the loss function (Kingma et al., 2014). We initialize weights with uniform distribution and use a batch size of 64 considering a tradeoff between efficiency and generalization performance (Keskar et al., 2016). The training parameter setups are the same for both Sections 3.1 and 3.2.

3.1 Application to continental China

In this section we conduct two experiments, named *Test1* and *Test2,* to study the impact of the training dataset on the performance of the CNN. For *Test1* and *Test2*, we employ two different training datasets, named *Dataset1* and *Dataset2*, with same control parameters (learning rate, batch size, etc.) to train the CNN. Then, we use two CNNs trained separately with those two training datasets to invert Vs models using the actual Rayleigh wave dispersion curves measured in continental China. The two CNNs based models are evaluated by comparing with the Vs model of Shen et al. (2016).

Dataset1 comprises 6803 1-D Vs models (left panel of Fig. S2) extracted from the central western USA tomographic model of Shen et al. (2013) and the corresponding theoretical Rayleigh wave phase and group dispersion images (section 2.1). For *Dataset2*, we augment *Dataset1* with additional 675 1-D Vs models (right panel of Fig. S2) extracted from the Tibet region (white box in the top panel of Fig. 3) results of Shen et al. (2016) and the corresponding theoretical dispersion images. Shen et al. (2016) measured the Raleigh wave group and phase velocity dispersion curves for a period range of 8-50s in continental China. Those dispersion measurements were used to determine a 3-D Vs model for the top 150 km of continental China via a Bayesian Monte Carlo inversion. Here we use the dispersion measurements of Shen et al. (2016) as the test dataset for *Test1* and *Test2*. For the test dataset, both phase and group velocity dispersion curves are required to be within the period range of 8-50s at each grid node. We linearly interpolate those phase and group velocity dispersion curves and generate phase-group dispersion images based on equation 1. A total of 3260 pairs of phase-group velocity dispersion images are produced for the test dataset of continental China, which covers most of continental China (the bottom panel in Fig.3).

The training losses as a function of epochs show for both tests a dramatic decrease (from ~4 km/s to ~0.15 km/s) after the first 15 epochs (Figs. S3 a&b). After 600 epochs, the final losses converge to ~0.05 km/s and we take the CNNs at epoch 600 as the final trained CNNs for both tests. These are then used to estimate 1-D Vs models from dispersion measurements for continental China. It takes ~1.5 hour to train the CNN and ~23 seconds to generate 3260 1-D Vs models with the entire test dataset for both *Test1 and Test2*. The CNN-based inversion is much more efficient computationally than the Bayesian Monte Carlo inversion used in Shen et al. (2016), which usually requires more than 200 computing hours for the same test dataset.

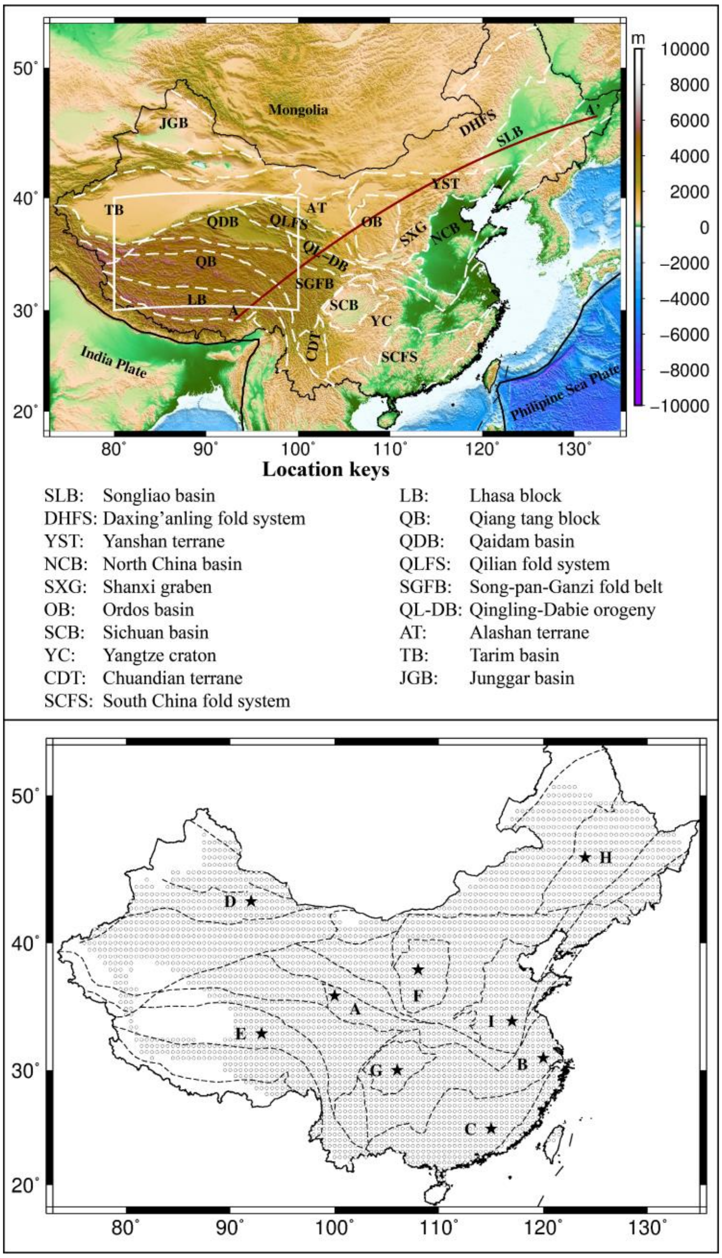

Figure 3. (Top panel) Map view of the surface topography and major tectonic features of China and surrounding area. White dashed lines outline the main tectonic units and basins, and black bold lines indicate the plate boundary. The dark red line denotes AA' profile (shown in Fig. 5). The white box outlines the region in which the 675 1-D Vs models of Shen et al. (2016) are extracted to build up the training dataset for *Test2*. (Bottom panel) Map view of the area covered by the test dataset. Black stars mark grid nodes that are used to demonstrate the comparison between the observed and predicted Rayleigh wave dispersion curves in Fig. 6.

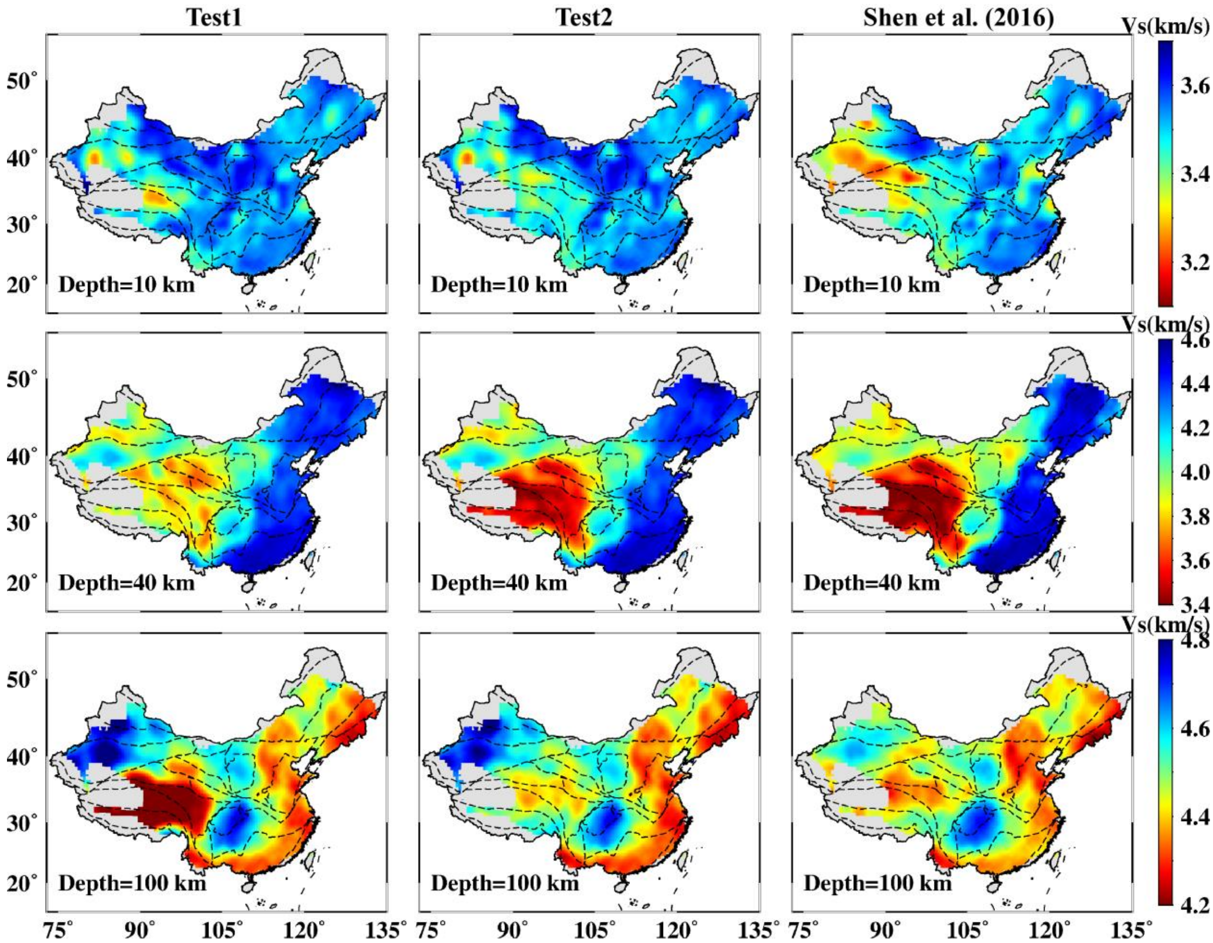

Figure 4. Comparison of Vs depth slices obtained from (left) *Test1*, (middle) *Test2*, and (right) Shen et al. (2016) at 10 km (top), 40 km (center), and 100 km (bottom), respectively. The thick dashed lines delineate the tectonic units.

Although the exact Vs distribution of continental China is unknown, we can take the 3-D Vs model of Shen et al. (2016) as the baseline model and compare results from *Test1* and *Test2*. Note that the CNN training dataset for *Test*2 includes Vs information from the Tibet region, which is significantly different compared to the Vs model of the central western USA in Shen et al. (2013). Figure 4 shows comparisons of Vs distributions given by *Test1*, *Test2*, and Shen et al. (2016) at depths of 10, 40, and 100 km. The results of *Test1* and *Test2* show high similarity to the baseline 3-D Vs model in eastern, southern, northern, and northeast China, especially for several main tectonic units including SCB, OB, NCB (see keys in Fig. 3). For JGB and TB, the Vs values from *Test1* and *Test2* are systematically larger than those of the Shen et al. (2016) for all three depths. In the Tibet region, however, Vs values from *Test1* are larger than those of the baseline model at the depths of 40 and 100 km, while Vs values from *Test2* are close to those of Shen et al. (2016) at all three depths. This is because the 1-D Vs profiles in Tibet region are significantly different from those of the training dataset used to train the CNN in *Test1*.

Furthermore, we compare Vs models from *Test1* and *Test2* to results of Shen et al. (2016) along a vertical cross-section AA' at a depth range of 5–120 km (Fig. 5), which crosses SLB, OB and Tibet region. We exclude the top 5 km Vs structures due to a lack of sensitivity at shallow depth as illustrated in Shen et al. (2016). Overall, the Vs distributions of all three cross-sections are similar except for the Tibet region. In general, the crustal thickness (the areas marked by red-yellow color in Fig.5) decreases from Tibet to SLB for all three cross-sections, which is consistent with previous imaging results (He et al., 2014; Xin et al., 2018).

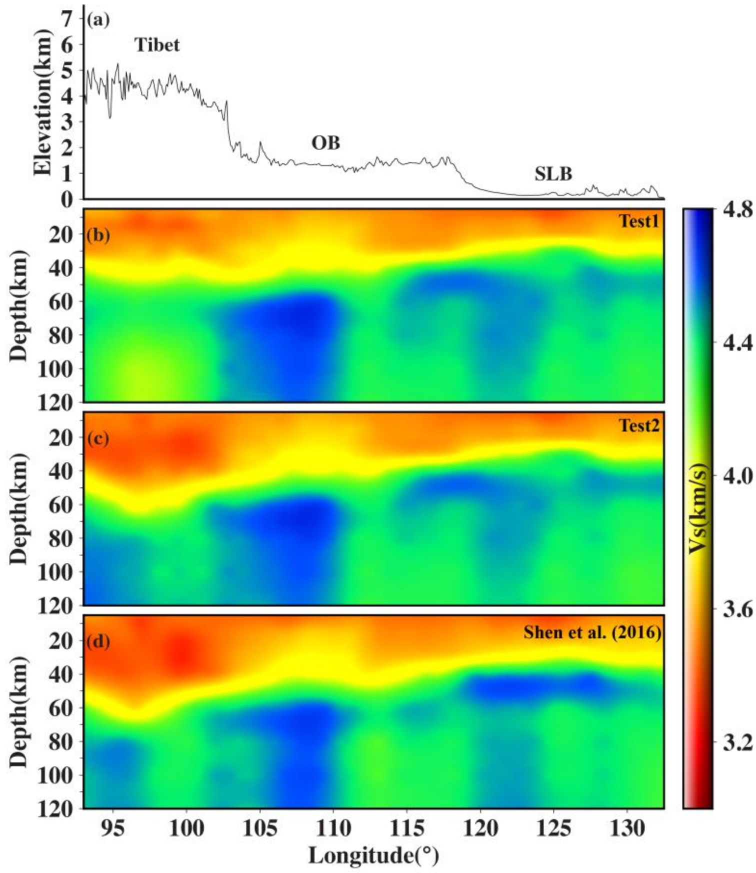

Figure 5. Comparison of Vs vertical cross sections along profile AA' for *Test1*, *Test2* and Shen et al. (2016). (a) Topography along AA' profile. OB and SLB represent Ordos basin and Songliao basin, respectively (shown in Fig. 3). (b) Vs model of *Test1*. (c) Vs model of *Test2*. (d) Vs model of Shen et al. (2016).

To quantify the quality of the model given by the CNN, we examine the data fitting between the input and model-predicted dispersion curves. For each location, the data misfit is defined as:

$$Misfit = \left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{d_i^{obs} - d_i^{cal}}{\sigma_i}\right)^2\right]^{\frac{1}{2}} \qquad (3)$$

where d_i^{obs} is the observed Rayleigh wave phase or group velocity, d_i^{cal} is the value predicted by the corresponding model, σ_i is the uncertainty of the phase or group velocity, i is the index of the discrete phase and group velocity measurements, and N is the number of dispersion measurements.

Figure 6 shows the resulting dispersion fitting at two example gird nodes located in Tibet region and OB for both tests. The 1-D Vs profiles (*Test1*, *Test2*, and baseline model of Shen et al. (2016)) and corresponding dispersion curves are also shown for comparison. The misfits at the Eth grid node (upper panels of Figs. 6a&b) for *Test1*, *Test2*, and the baseline model are 9.29, 1.06, and 1.05, respectively. The misfits at the Fth grid node (bottom panels of Figs. 6a&b) for *Test1*, *Test2*, and the baseline model are 0.87, 0.73, and 1.93, respectively. Both models from *Test1* and *Test2* at the Fth grid node in OB yield smaller misfit values than the model of Shen et al. (2016), while the model from *Test1* fails to fit the dispersion measurements at the Eth grid node in Tibet region. This indicates that the CNNs give inaccurate estimate 1-D Vs models when a biased Vs distribution is assumed in the training process. More examples of such dispersion fitting are shown in Figs. S4 & S5 in the supplementary material.

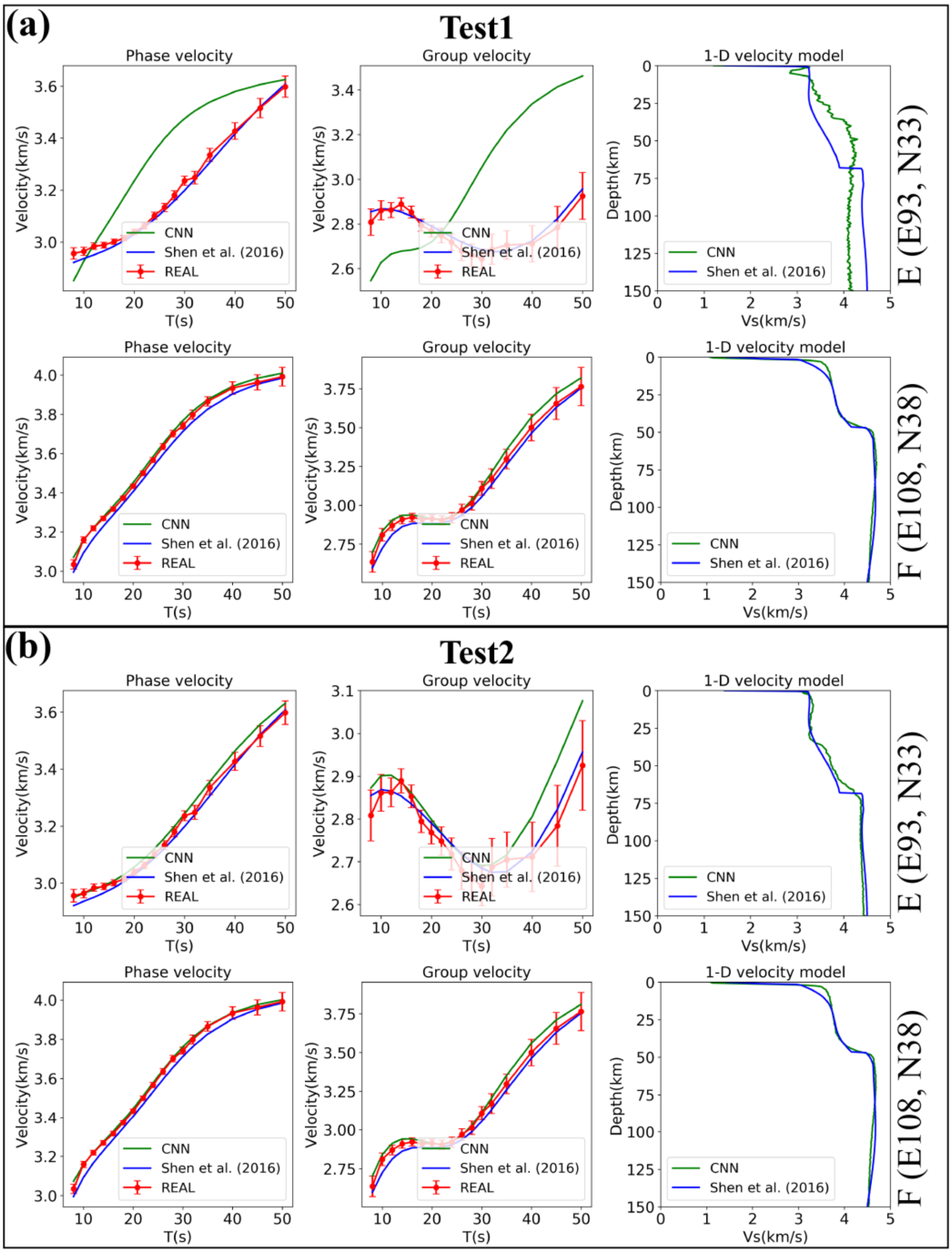

Figure 6. Comparison of observed and predicted dispersion curves for *Test1* (a) and *Test2* (b) at two selected nodes as shown in Fig. 4. Left and middle panels show the comparisons of Rayleigh wave group and phase dispersion curves, respectively. The red line in each panel represents the observed dispersion curve and error bars indicate a range of one standard deviation about each respective mean value. The right panels illustrate the comparison of 1-D Vs profiles obtained from the CNN based method and Shen et al. (2016). Green and blue lines depict the predicted dispersion curves from *Test1* and Shen et al. (2016), respectively. For the Eth grid node, the misfits are 9.29 for *Test1*, 1.06 for *Test2*, and 1.05 for Shen et al. (2016). For

the F^{th} grid node, the misfits are 0.87 for *Test1*, 0.73 for *Test2*, and 1.92 for the Shen et al. (2016) model.

The spatial distribution and histogram of dispersion misfits for *Test1*, *Test2*, and baseline model of Shen et al. (2016) are shown in Figs. 7a&b. The average misfit values are 3.03 for *Test1*, 1.97 for *Test2*, and 1.37 for the baseline model (Fig. 7b). The average misfit value of the baseline model is slightly larger than ~0.76 reported in Shen et al. (2016). This is because Shen et al. (2016) used a different scaling relationship to obtain Vp and density values from Vs, and the amount of data from continental China and surrounding area (Shen et al. 2016) is larger than that used in this study.

As shown in Fig. 7a, regions in the eastern part of continental China show generally small misfit values (< 2), suggesting the dispersion curves are well fitted for both *Test1* and *Test2*, whereas the dispersion data is poorly fitted in *Test1* for the Tibet region. In the CDT region the misfit of *Test2* is less than that of *Test1*, indicating that the CNN of *Test2* provides better estimations of the Vs structure there. Since the Vs distribution of the CDT region is not included in the training datasets of both tests, this likely suggests that the diversity of 1-D Vs profiles in the Tibet region is sufficient to represent the complexities in the CDT region.

For model uncertainty estimates, we take *Test1* as an example to illustrate the results. In order to perform a statistical analysis, we split the dataset in the same way to that of Section 2.3. The randomly splitting dataset process is performed 15 times to produce different training datasets. Then, those different datasets are subsequently used to train 15 CNNs, respectively. The trained CNNs are used to estimate 1-D Vs models with the entire test dataset of *Test1*, independently. The standard deviation of Vs at each layer is calculated (Figs. S6a&b). The mean standard deviation of Vs values from those CNNs is ~0.06 km/s (Fig. S6c), suggesting that the Vs model given by CNN is insensitive to the selection of training dataset when 1-D Vs profiles are sufficiently accurate. We cannot estimate the actual Vs model uncertainty, which is related to the uncertainty of weights updated in the CNN during the training as well as errors in dispersion curves, so the reported uncertainty is underestimated.

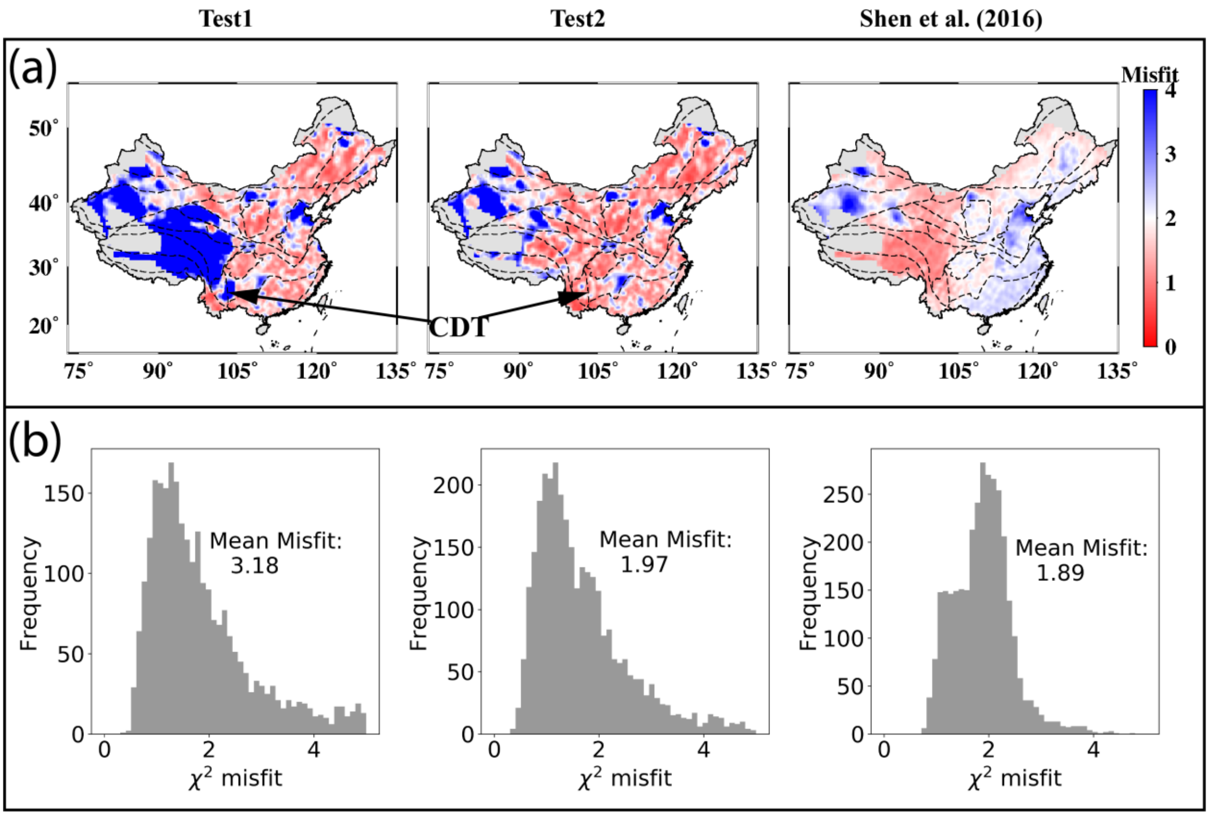

Figure 7. Dispersion misfit maps (a) and the corresponding histograms (b) for *Test1* (left), *Test2* (middle), and Shen et al. (2016) (right). For *Test1*, *Test2*, and Shen et al. (2016) models, the mean misfits are 3.03, 1.95, and 1.89, respectively. We do not compare the misfit maps of Shen et al. (2016) with those of *Test1* and *Test2* due to a difference in Vp and density setups between our study and Shen et al. (2016).

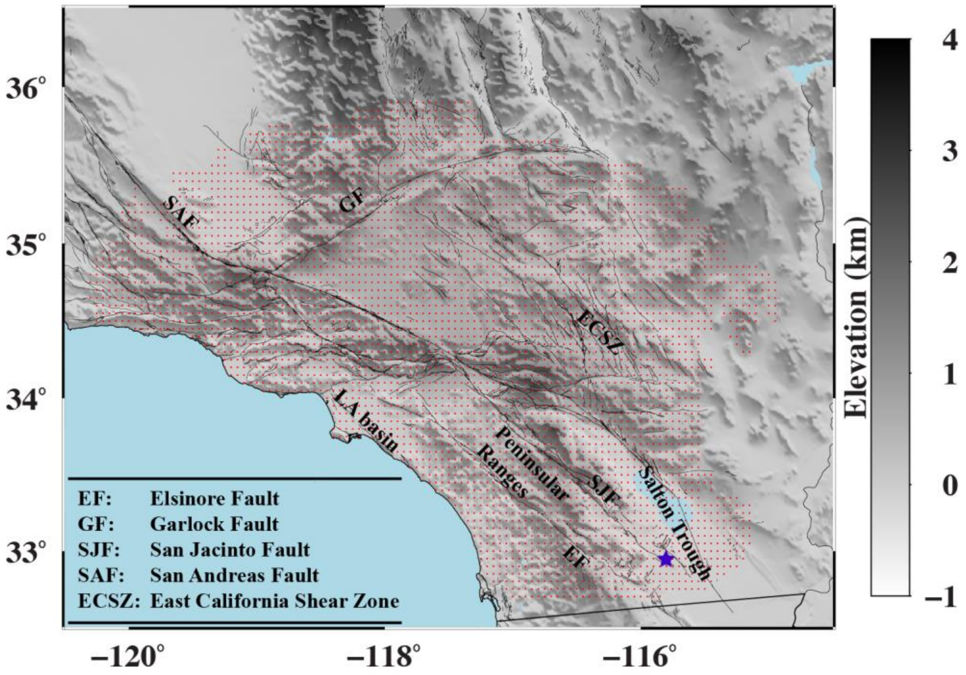

Figure 8. Map view of the Southern California plate boundary region. Red dots represent the spatial distribution of the observed Rayleigh wave phase and group dispersion curves used in this study. Thin black lines denote the main fault surface traces in Southern California. The gray-shaded background depicts the topography. The blue star marks the selected node that is used to show the comparison between the observed and predicted Rayleigh wave dispersion curves in Fig. 11.

3.2 Application to Southern California

Assuming the same initial 3-D Vs model, can the CNN based tomography perform as well as the traditional methods (e.g. Herrmann, 2013; Shen et al., 2013)? The southern California region provides a good opportunity to answer this question as several dense seismic networks and seismic velocity models (Lee et al., 2014; Shaw et al., 2015) are available in this plate boundary region. Recently, Qiu et al. (2019) performed Eikonal tomography for Southern California using more than 300 seismic stations and provided isotropic Rayleigh wave phase and group velocity maps with a grid size of 0.05°×0.05° over a period range of 2.5-16s (Fig. 8). These Rayleigh wave velocity maps were jointly inverted at each grid node using the velocity model of Shaw et al. (2015; hereinafter referred to as CVMH) as an initial model and the CPS (Herrmann, 2013) to obtain a set of 1-D Vs profiles for the top 50 km. The 1-D Vs profiles were assembled to construct a 3-D Vs model.

To perform a direct comparison between the CNN based model and the results of Qiu et al. (2019), we use the same Rayleigh wave phase and group velocity maps of Qiu et al. (2019) and depth discretization from the surface to a depth of 49.5 km with 0.5 km layer thickness. The CNN architecture is the same as Section 3.1 except for the output dimension size (99 – the number of layers). For the training dataset, we extract 24554 1-D Vs profiles from the CVMH and generate corresponding theoretical phase and group velocity dispersion curves of Rayleigh waves over a period range of 2.5-16s. The Vp and density are also given by the CVMH. We then generate corresponding phase and group dispersion images following the processing steps described in Section 2.1 (equation 1). We use a velocity range of 1.0km/s-4.5km/s for constructing the dispersion images. For the test dataset, phase and group dispersion images are generated using the Rayleigh wave velocity maps of Qiu et al. (2019). Finally, we have 4160 pairs of Rayleigh wave phase and group dispersion images as the input test dataset.

The training parameter setups (maximum number of epochs, batch size, learning rate, etc.) are the same as those used in Section 3.1. The final loss converges to ~0.06 km/s (Fig. S3c). Convergent results are achieved after 600 epochs without overfitting and the trained network at epoch 600 is used to estimate the 1-D Vs models from the input dispersion images. It takes ~4.5 hours to train the new CNN and ~30 seconds to generate all the 1-D Vs models, while more than 30 hours are required to invert the same dataset in Qiu et al. (2019).

Due to a lack of Rayleigh wave velocity data at periods shorter than 3s or longer than 16s, we cannot resolve the structures in the top 3 km and below 20 km (Qiu et al., 2019). For this reason, Fig. 9 only shows the comparison of Vs depth slices between the CNN based model, results of Qiu et al. (2019) and the initial model CVMH at depths of 3-15 km. The derived CNN-based model overall shows consistent features compared to Qiu et al. (2019). However, the velocities in the Peninsular Ranges and Salton Trough region from the CNN model are, respectively, higher and considerably lower than those of Qiu et al. (2019). The large-scale geological features observed in the CNN model are consistent with those of the CVMH model, while the variations in velocity values are much larger in the CNN-based model. Figure 10 shows the spatial distributions of Rayleigh wave velocity dispersion misfit (equation 3) for the CVMH, CNN based model, and Qiu et al. (2019). The average misfits are 4.53, 1.49, and 1.72 for the CVMH, CNN based model and Qiu et al. (2019), respectively. In Qiu et al. (2019) the Vp/Vs ratio and Moho depth are fixed in the inversion and the average misfit is ~0.6, which is smaller than the value obtained in this study. The larger misfit from the CNN based model is likely caused by different scaling relations between Vp and Vs used to compute theoretical Rayleigh wave dispersion curves. In general, large misfits are seen in the Salton Trough region relative to other regions in the results of Qiu et al. (2019). However, the misfit values in the

Salton Trough region is similar to the other regions in the CNN based model, suggesting the dispersion data are likely better fitted by the CNN model than Qiu et al. (2019) in the Salton Trough region (Fig. 11). The average misfit of the initial model CVMH is a factor of ~3 larger than that of the CNN based result, suggesting the CNN based model significantly improves the fitting of the input dispersion data. Overall, the CNN based Vs model fits the input Rayleigh wave dispersion data better than CVMH and similar to the results of Qiu et al. (2019).

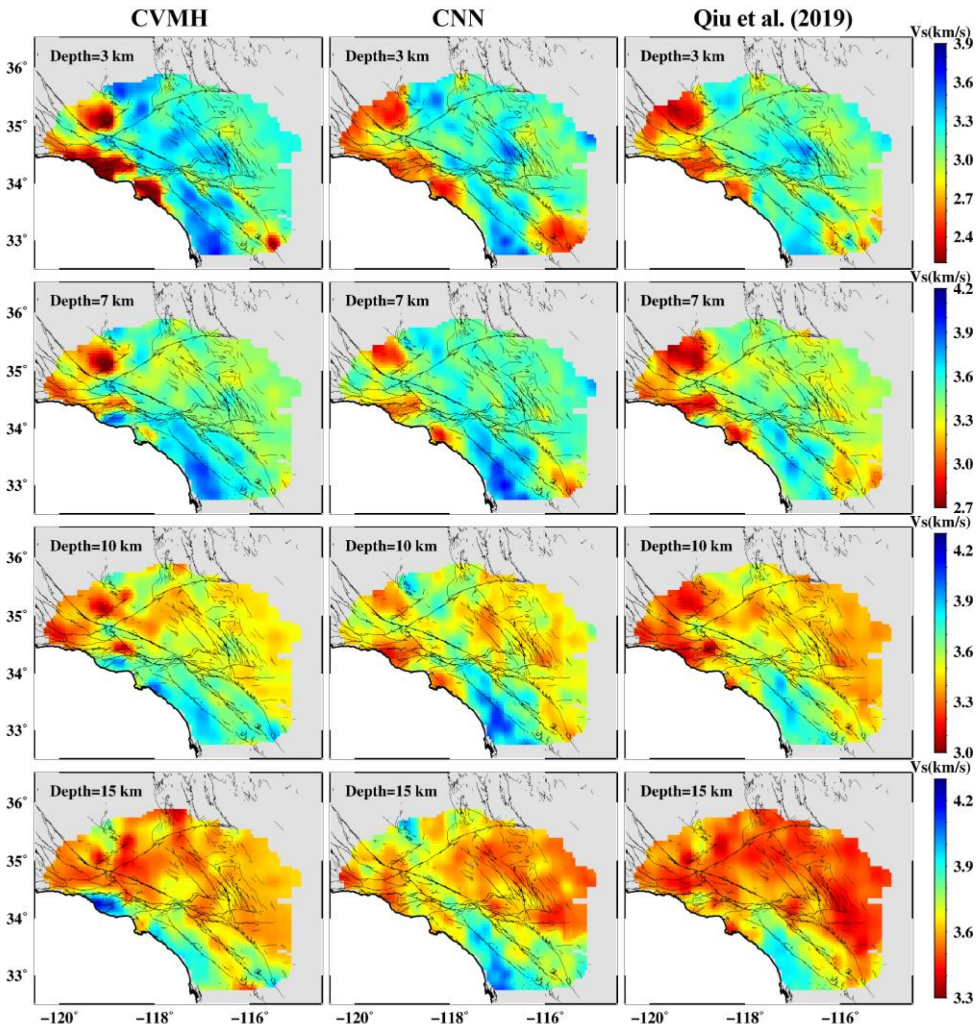

Figure 9. Comparison of the Vs depth slices extracted from the CVMH (left), CNN (middle), and results of Qiu et al. (2019) (right) at 3 km, 7 km, 10 km, and 15km. The thin black lines delineate surface traces of main faults in Southern California. The thick line represents the coastal line.

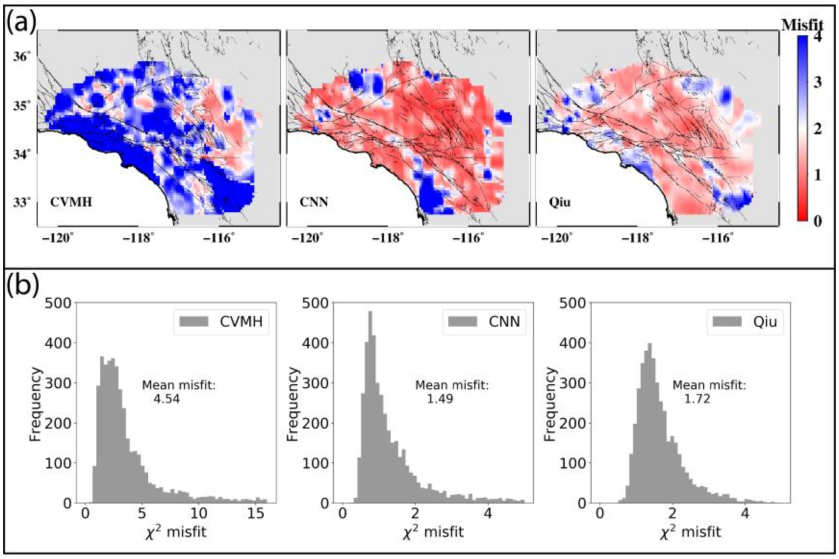

Figure 10. Misfit maps (a) and histograms (b) of the CVMH, CNN, and Qiu et al. (2019). The average misfits are 4.53 for CVMH, 1.49 for CNN, and 1.72 for Qiu et al. (2019).

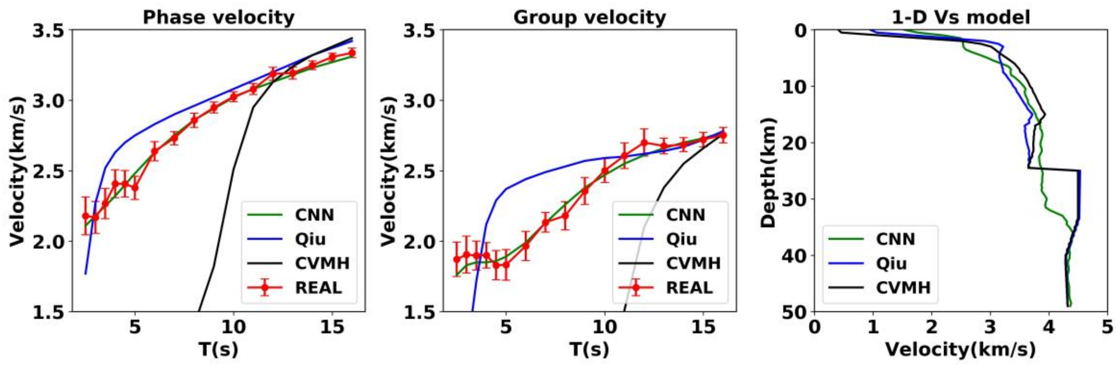

Figure 11. Comparison of the observed and predicted Rayleigh wave dispersion curves for the CNN based model, Qiu et al. (2019), and CVMH at the selected node marked by blue star in Fig. 8. Left and middle panels show the comparisons of the observed and predicted Rayleigh wave group and phase dispersion curves, respectively. The red curve in each panel represents the observed dispersion curves and error bars indicate a range of one standard deviation about each respective mean value. 1-D Vs models used to generate predicted dispersion curves are illustrated in the right panel. Green, black, and blue lines depict the dispersion curves predicted by the CNN model, Qiu et al. (2019), and CVMH, respectively. The misfit values are 0.58 for the CNN based model, 3.3 for the Qiu et al. (2019), and 16.2 for the CVMH model.

4. Discussion and conclusions

The ongoing significant increase in the number of seismic stations produces big datasets that require fast processing methods. In this study we demonstrate that properly trained CNNs provide highly effective tools for converting surface wave dispersion measurements directly to shear wave velocity models. The CNNs bypass the need to select carefully an appropriate initial model and inversion parameters, and they replace time-consuming nonlinear inversions with rapid direct mapping of phase and group velocity dispersion curves to model velocity results.

The effectiveness of the proposed CNN-based technique is tested on two different datasets associated with continental China and Southern California. Compared with the earlier method of Meier et al. (2007), we use deep neural network that can better represent highly complex velocity models. Different from Cheng et al. (2019) that used deep neural network for surface wave tomography from phase velocities, we utilize both phase and group velocities as inputs. In addition, in contrast to Meier et al. (2007) and Cheng et al. (2019) that generated dataset based on a reference model with a small number of layers, we use training datasets associated with Vs models derived in previous studies with many layers that are 0.5 km thick each. The application of the CNN to continental China shows great potential for deriving Vs models using a relatively small training dataset (*Test1* in Section 3.1). The analysis also demonstrates that increasing diversity in the training dataset enhances the performance of the CNNs (*Test2* in Section 3.1). Both the CVMH model and results of Qiu et al. (2019) have large dispersion misfits in the Salton Trough region. The significant improvement in fitting dispersion measurements in that region (Figs. 10&11) using CNN suggests that the CNN-based results are less affected by the initial model than those of classical inversion schemes (Qiu et al. 2019).

Our applications of CNNs to surface wave tomography employ dispersion curves data over different period ranges for continental China and Southern California, targeting a depth range of 0-150 km for continental China and 0-50 km for Southern California. The results are consistent with previous studies using conventional methods but our method is computationally far more efficient. This advantage will become increasingly important as training datasets accumulate and vast new datasets are recorded. Future applications can include monitoring structures (e.g., fault zones, volcanos, reservoirs) in real time with CNN-based time-dependent tomography using all results at the monitored locations as training datasets.

The model uncertainty associated with the CNN results are difficult to estimate. This is a common problem in deep learning, which impacts the ability to interpret the outputs. We obtain low bound uncertainty estimates of the results by using an ensemble of perturbed training datasets. An improve procedure may implement validation and iterative improvements, where in each iteration forward calculation results based on the derived CNN models are compared to data. Additional future improvements include using deep mixture density network to estimate the uncertainty of the outputs, and including receiver function results (Bodin et al., 2012; Shen et al., 2013) and lateral discontinuities between geological units as constrained in the output models.

Data and code availability

The CNN is implemented using the deep-learning frame of Pytorch-0.4 library. The training and prediction processes are performed on a laptop with a NVIDIA GeForce GTX 1060 core and a memory of 6 GB. For Section 3, scripts and training and test data sets are available at https://github.com/jhsa26/SurfTomoCNN.

Acknowledgements

This study was supported by National Key Research and Development Program of China (grant 2018YFC1504102), National Natural Science Foundation of China (U1839205), China Scholarship Council (CSC), and the Department of Energy (award DE-SC0016520).

References

Abdel-Hamid, O., Mohamed, A. R., Jiang, H., Deng, L., Penn, G., & Yu, D. (2014). Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on audio, speech, and language processing*, *22*(10), 1533-1545.

Ben-Zion, Y., F. L. Vernon, Y. Ozakin, D. Zigone, Z. E. Ross, H. Meng, M. White, J. Reyes, D. Hollis and M. Barklage (2015). Basic data features and results from a spatially dense seismic array on the San Jacinto fault zone. *Geophysical Journal International*, *202*(1), 370-380.

Bergen, K. J., Johnson, P. A., Maarten, V., & Beroza, G. C. (2019). Machine learning for data-driven discovery in solid Earth geoscience. *Science*, *363*(6433), eaau0323.

Bodin, T., Sambridge, M., Tkalčić, H., Arroucau, P., Gallagher, K., & Rawlinson, N. (2012). Transdimensional inversion of receiver functions and surface wave dispersion. *Journal of Geophysical Research: Solid Earth*, *117*(B2).

Brocher, T. M. (2005). Empirical relations between elastic wavespeeds and density in the Earth's crust. *Bulletin of the seismological Society of America*, *95*(6), 2081-2092.

Cheng, X., Liu, Q., Li, P., & Liu, Y. (2019). Inverting Rayleigh surface wave velocities for crustal thickness in eastern Tibet and the western Yangtze craton based on deep learning neural networks. *Nonlinear Processes in Geophysics*, *26*(2), 61-71.

Collobert, R., & Weston, J. (2008, July). A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th international conference on Machine learning* (pp. 160-167). ACM.

Curtis, A., Trampert, J., Snieder, R., & Dost, B. (1998). Eurasian fundamental mode surface wave phase velocities and their relationship with tectonic structures. *Journal of Geophysical Research: Solid Earth*, *103*(B11), 26919-26947.

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). IEEE.

Devilee, R. J. R., Curtis, A., & Roy-Chowdhury, K. (1999). An efficient, probabilistic neural network approach to solving inverse problems: Inverting surface wave velocities for Eurasian crustal thickness. *Journal of Geophysical Research: Solid Earth*, *104*(B12), 28841-28857.

He, R., Shang, X., Yu, C., Zhang, H., & Van der Hilst, R. D. (2014). A unified map of Moho depth and Vp/Vs ratio of continental China by receiver function analysis. *Geophysical Journal International*, *199*(3), 1910-1918.

Herrmann, R. B. (2013). Computer programs in seismology: An evolving tool for instruction and research. *Seismological Research Letters*, *84*(6), 1081-1088.

Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A convolutional neural network for modelling sentences. *arXiv preprint arXiv:1404.2188*.

Kennett, B. L., Engdahl, E. R., & Buland, R. (1995). Constraints on seismic velocities in the Earth from traveltimes. *Geophysical Journal International*, *122*(1), 108-124.

Keskar, N. S., Mudigere, D., Nocedal, J., Smelyanskiy, M., & Tang, P. T. P. (2016). On large-batch training for deep learning: Generalization gap and sharp minima. *arXiv preprint arXiv:1609.04836*.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kong, Q., Inbal, A., Allen, R. M., Lv, Q., & Puder, A. (2018). Machine learning aspects of the MyShake global smartphone seismic network. *Seismological Research Letters*, *90*(2A), 546-552.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436.

Lee, E. J., Chen, P., Jordan, T. H., Maechling, P. B., Denolle, M. A., & Beroza, G. C. (2014). Full-3-D tomography for crustal structure in southern California based on the scattering-integral and the adjoint-wavefield methods. *Journal of Geophysical Research: Solid Earth*, *119*(8), 6421-6451.

Li, Z., Meier, M. A., Hauksson, E., Zhan, Z., & Andrews, J. (2018). Machine learning seismic wave discrimination: Application to earthquake early warning. *Geophysical Research Letters*, *45*(10), 4773-4779.

Lin, F. C., Ritzwoller, M. H., & Snieder, R. (2009). Eikonal tomography: surface wave tomography by phase front tracking across a regional broad-band seismic array. *Geophysical Journal International*, *177*(3), 1091-1110.

Lin, F. C., Schmandt, B., & Tsai, V. C. (2012). Joint inversion of Rayleigh wave phase velocity and ellipticity using USArray: Constraining velocity and density structure in the upper crust. *Geophysical Research Letters*, *39*(12).

Liu, Y., Zhang, H., Fang, H., Yao, H., & Gao, J. (2018). Ambient noise tomography of three-dimensional near-surface shear-wave velocity structure around the hydraulic fracturing site using surface microseismic monitoring array. *Journal of Applied Geophysics*, *159*, 209-217.

Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013, June). Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml* (Vol. 30, No. 1, p. 3).

McBrearty, I. W., Delorey, A. A., & Johnson, P. A. (2019). Pairwise association of seismic arrivals with convolutional neural networks. *Seismological Research Letters*, *90*(2A), 503-509.

Meier, U., Curtis, A., & Trampert, J. (2007). Global crustal thickness from neural network inversion of surface wave data. *Geophysical Journal International*, *169*(2), 706-722.

Mosegaard, K., & Tarantola, A. (1995). Monte Carlo sampling of solutions to inverse problems. *Journal of Geophysical Research: Solid Earth*, *100*(B7), 12431-12447.

Paszke, A., Gross, S., Chintala, S., Chanan, G., et al. (2017). Automatic differentiation in pytorch.

Perol, T., Gharbi, M., & Denolle, M. (2018). Convolutional neural network for earthquake detection and location. *Science Advances*, *4*(2), e1700578.

Qiu, H., F.-C. Lin and Y. Ben-Zion, 2019. Eikonal tomography of the Southern California plate boundary region, *J. Geophys.*, *Res*., in review.

Ross, Z. E., Meier, M. A., & Hauksson, E. (2018). P wave arrival picking and first-motion polarity determination with deep learning. *Journal of Geophysical Research: Solid Earth*, *123*(6), 5120-5129.

Ross, Z. E., Yue, Y., Meier, M. A., Hauksson, E., & Heaton, T. H. (2019). PhaseLink: A deep learning approach to seismic phase association. *Journal of Geophysical Research: Solid Earth*, *124*(1), 856-869.

Sambridge, M. (1999a). Geophysical inversion with a neighbourhood algorithm—I. Searching a parameter space. *Geophysical journal international*, *138*(2), 479-494.

Sambridge, M. (1999b). Geophysical inversion with a neighbourhood algorithm—II. Appraising the ensemble. *Geophysical Journal International*, *138*(3), 727-746.

Shapiro, N. M., Campillo, M., Stehly, L., & Ritzwoller, M. H. (2005). High-resolution surface-wave tomography from ambient seismic noise. *Science*, *307*(5715), 1615-1618.

Shaw, J.H., Plesch, A., Tape, C., Suess, M.P., Jordan, T.H., Ely, G., Hauksson, et al. (2015). Unified structural representation of the southern California crust and upper mantle. *Earth and Planetary Science Letters, 415*, pp.1-15.

She, Y., Yao, H., Zhai, Q., Wang, F., & Tian, X. (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. *Seismological Research Letters*, *89*(3), 1003-1013.

Shen, W., Ritzwoller, M. H., Kang, D., et al. (2016). A seismic reference model for the crust and uppermost mantle beneath China from surface wave dispersion. *Geophysical Journal International*, *206*(2), 954-979.

Shen, W., Ritzwoller, M. H., & Schulte-Pelkum, V. (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. *Journal of Geophysical Research: Solid Earth*, *118*(1), 262-276.

Wang, J., Xiao, Z., Liu, C., Zhao, D., & Yao, Z. (2019). Deep-Learning for Picking Seismic Arrival Times. *Journal of Geophysical Research: Solid Earth*.

Xin, H., Zhang, H., Kang, M., He, R., Gao, L., & Gao, J. (2018). High-Resolution

Lithospheric Velocity Structure of Continental China by Double-Difference Seismic Travel-Time Tomography. *Seismological Research Letters, 90*(1), 229-241.

Yang, Y., & Ritzwoller, M. H. (2008). Teleseismic surface wave tomography in the western US using the Transportable Array component of USArray. *Geophysical Research Letters*, *35*(4).

Yao, H., van Der Hilst, R. D., & De Hoop, M. V. (2006). Surface-wave array tomography in SE Tibet from ambient seismic noise and two-station analysis—I. Phase velocity maps. *Geophysical Journal International*, *166*(2), 732-744.

Yu, Y., Lin, J., Zhang, L., Liu, G., Hu, J., Tan, Y., & Zhang, H. (2018, August). Identification of Seismic Wave First Arrivals from Earthquake Records via Deep Learning. In *International Conference on Knowledge Science, Engineering and Management* (pp. 274-282). Springer, Cham.

Zhou, Y., Nolet, G., Dahlen, F. A., & Laske, G. (2006). Global upper-mantle structure from finite-frequency surface-wave tomography. *Journal of Geophysical Research: Solid Earth*, *111*(B4).

Zhu, W., & Beroza, G. C. (2018). PhaseNet: a deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, *216*(1), 261-273.

Using deep learning to derive shear wave velocity models from surface wave dispersion data

Jing Hu[1], Hongrui Qiu[2], Haijiang Zhang[1*], Yehuda Ben-Zion[2]

1. Laboratory of Seismology and Physics of Earth's Interior; School of Earth and Space Sciences, University of Science and Technology of China, Hefei, Anhui, China
2. Department of Earth Sciences, University of Southern California, Los Angeles, CA 90089-0740, USA

*Corresponding email: zhang11@ustc.edu.cn

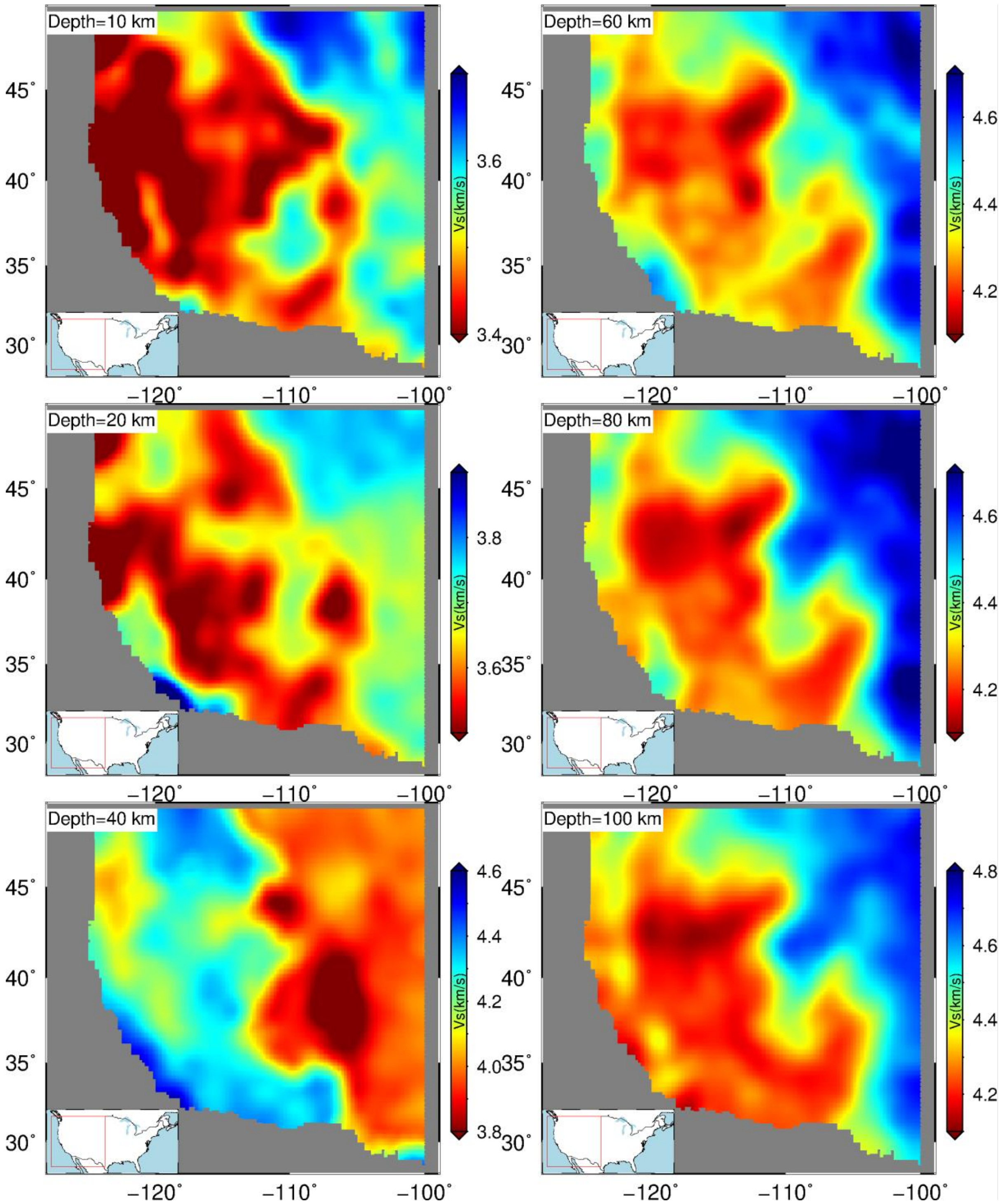

Figure S1. Vs labels of the training dataset in *Test1*. Six different depths are shown by nearest-neighbor interpolation with 6803 1-D Vs profiles from the surface to a depth of 150 km. The left bottom inset shows the map of USA and the red rectangular outlines the region of the Vs labels.

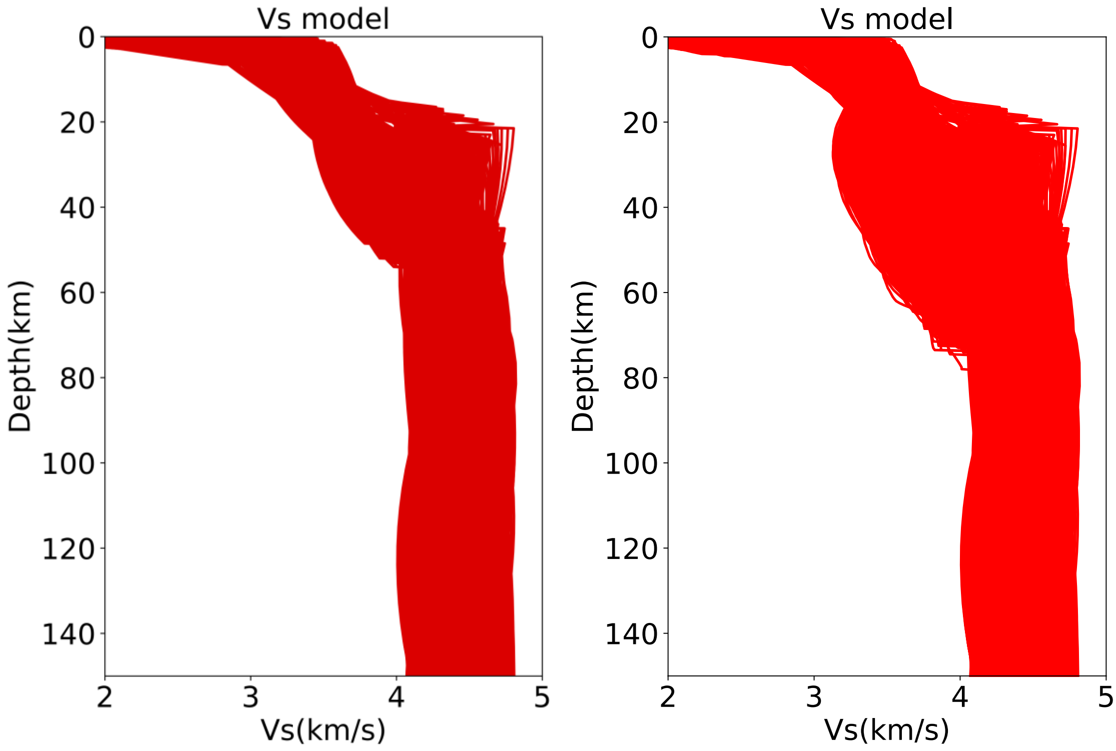

Figure S2. 1-D Vs profiles from the training datasets of *Test1* (left) and *Test2* (right).

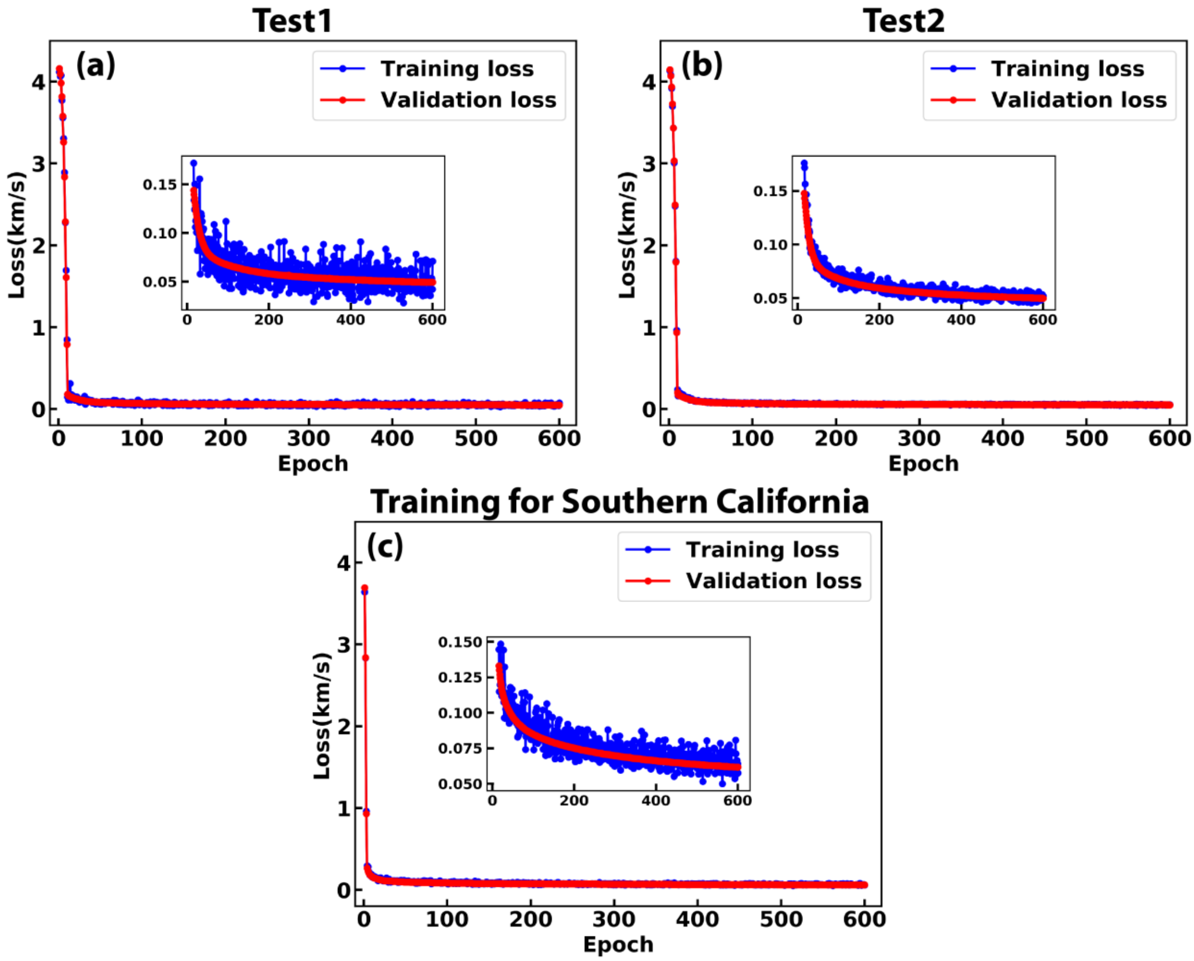

Figure S3. The training loss as a function of epoch for the training dataset (blue dotted curve) and validation dataset (red dotted curve). (a) CNN trained in *Test1*. (b) CNN trained in *Test2*. (c) CNN trained in the Southern California case. The inset is a zoom-in from epoch 15 to epoch 600.

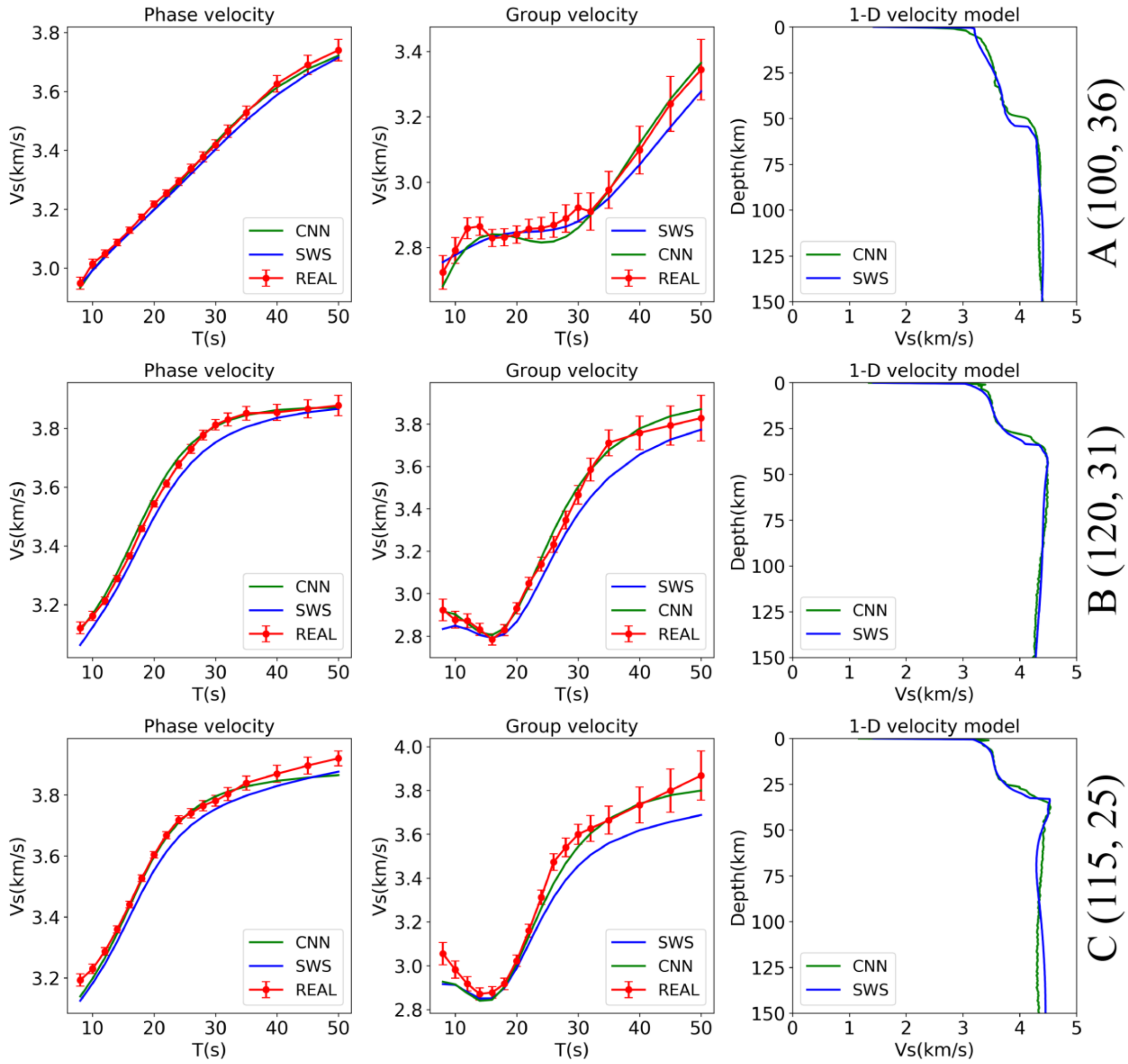

Figure S4. Dispersion fitting for *Test1* at nodes of A, B, C, G, H, and I. All symbols are the same as those shown in Fig. 7. SWS is short for Shen et al. (2016).

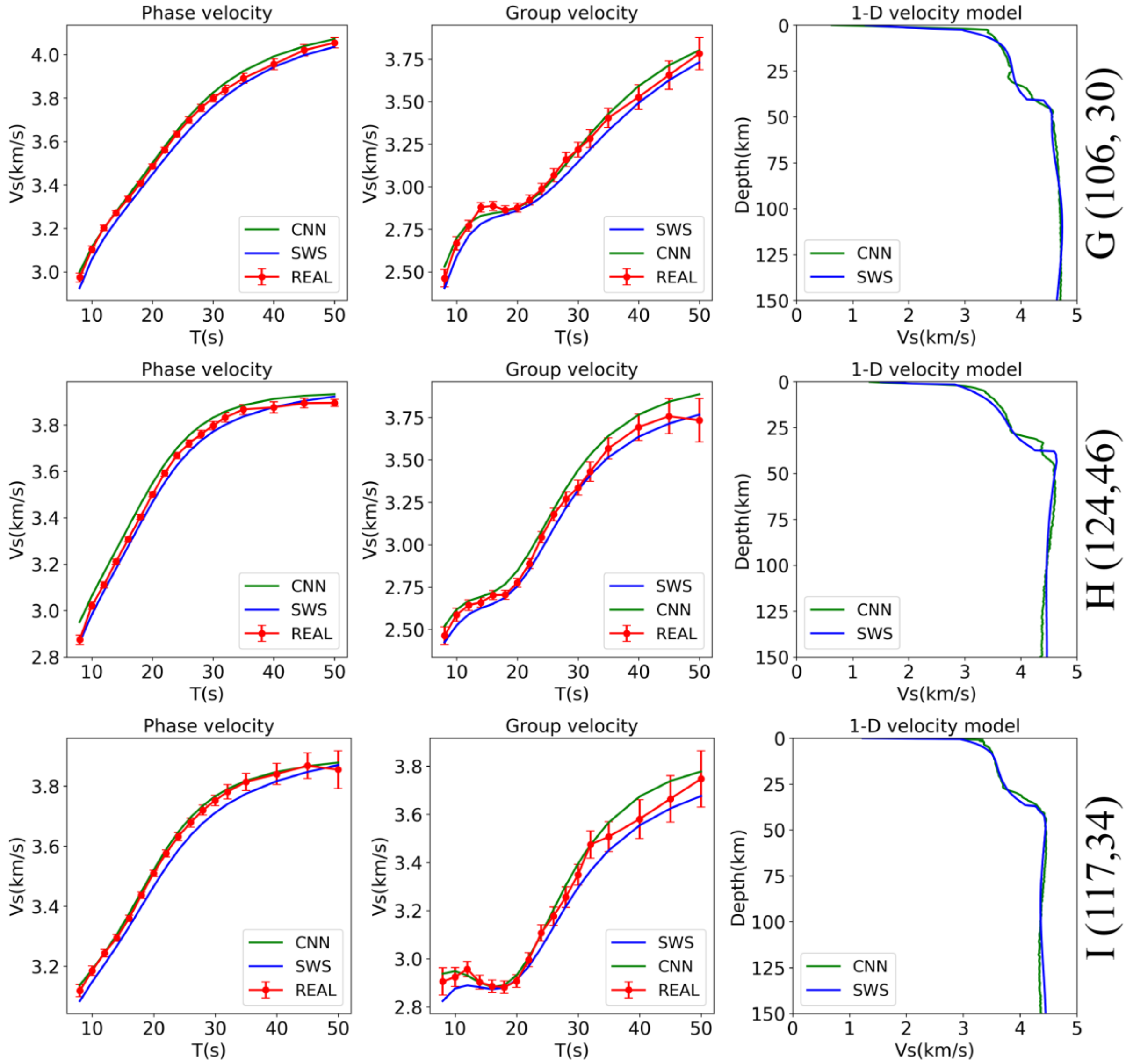

Figure S4. (Continued).

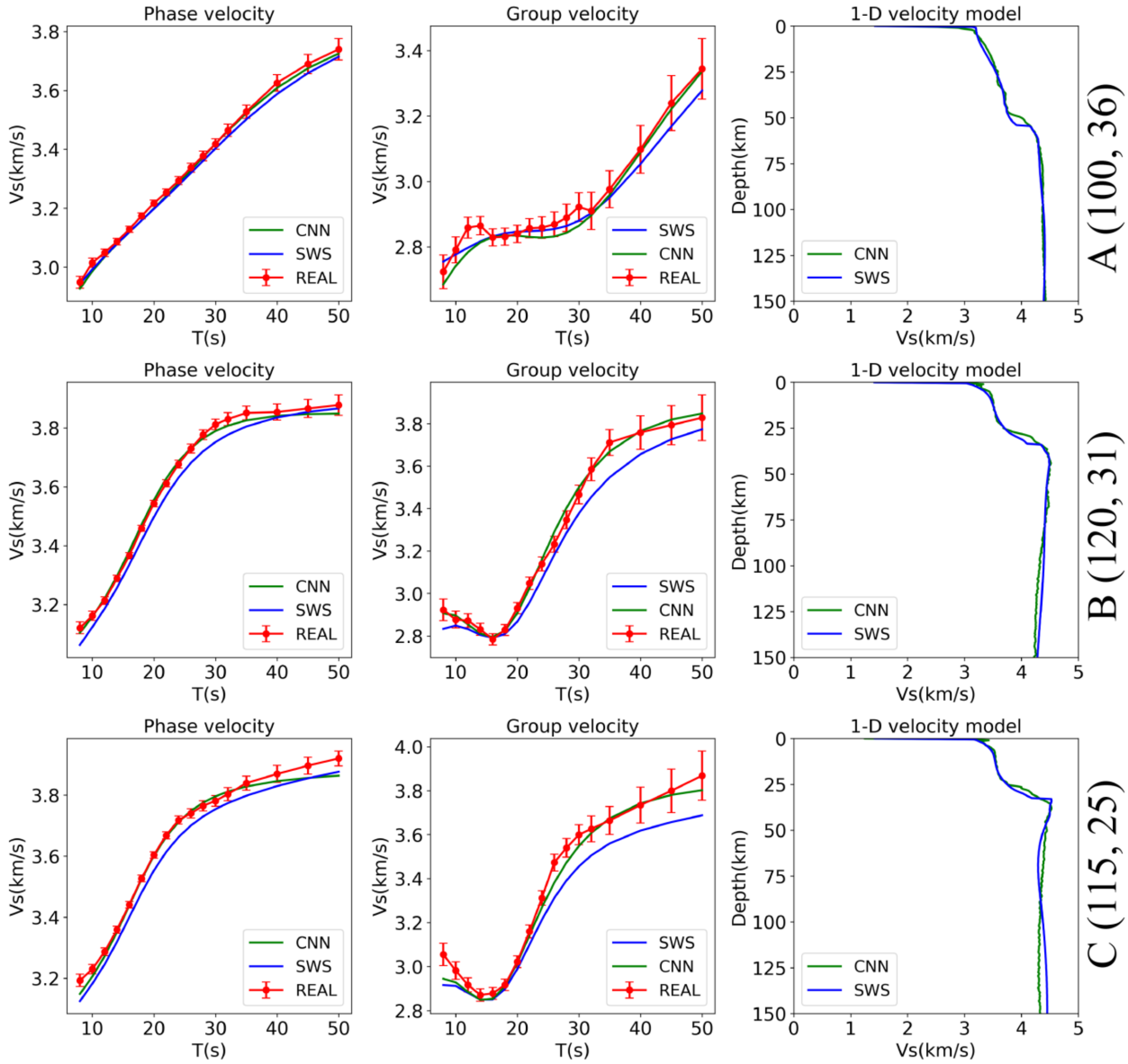

Figure S5. Same as Fig. S4 but for *Test2*.

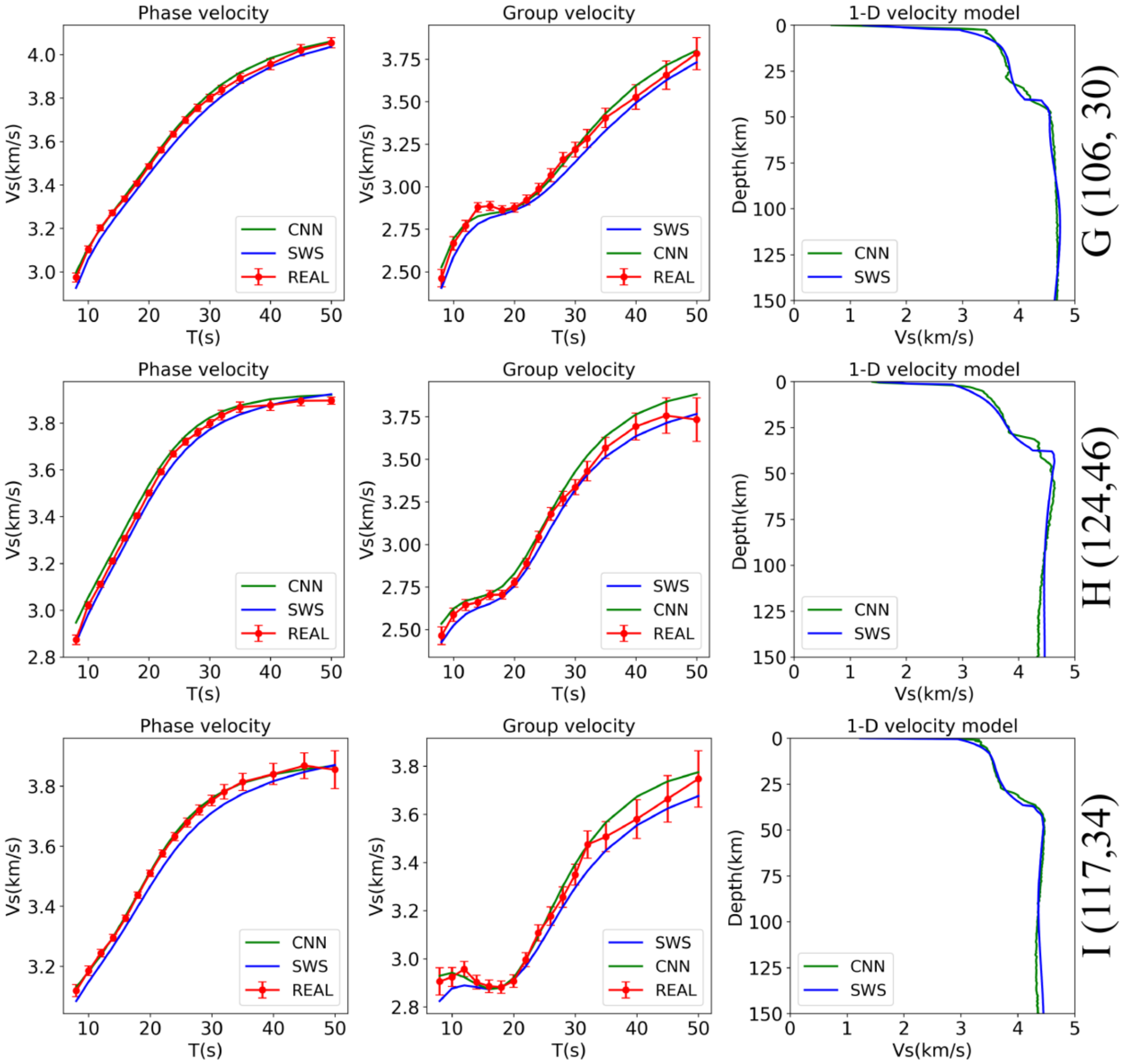

Figure S5. (Continued).

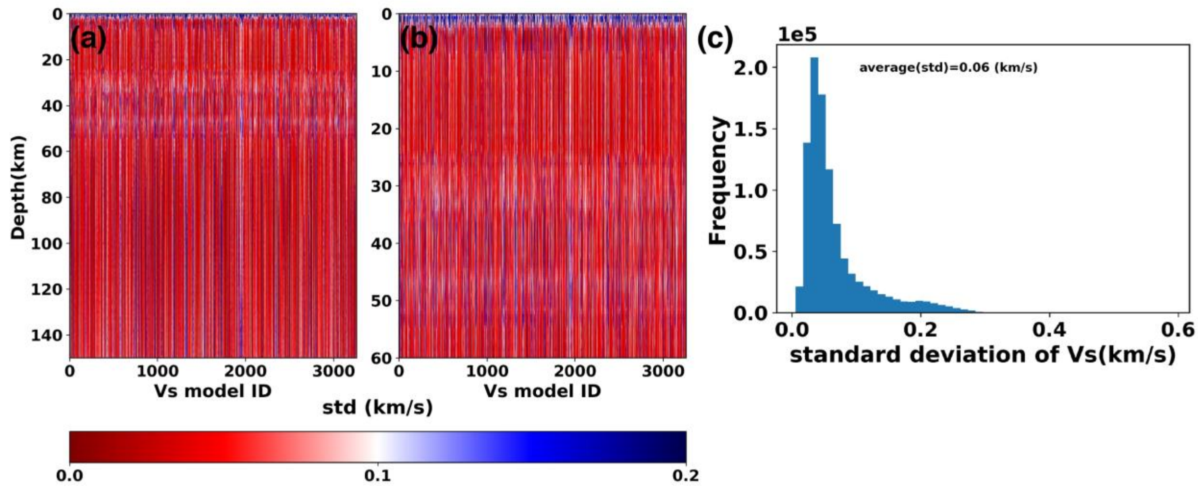

Figure S6. (a) The map of Vs standard deviations at each layer for the depth range of 0-150km based on the tested 3260 1-D Vs models. (b) A zoom-in map of (a) at a depth range of 0-60km. (c) The histogram of Vs standard deviations at each layer of all tested 1-D Vs models. The average standard deviation of Vs is 0.06 km/s for all layers of tested models.