Inferring shear wave velocity structure from surface wave dispersion data using Neural network

Group members: Chuanming Liu and Mengyu Wu
Team Name: SWT- ML

1. Problem space

Seismology is a data-driven field from its origin. With the development of the modern digital seismic station network, the archive of seismic waveforms publicly available from the Incorporated Research Institutions for Seismology (IRIS) has increased dramatically. Such an exponentially growing volume of data is hard to be processed manually. Based on the intrinsic property of machine learning (ML), ML is designed to deal with large databases and extract the wanted information. ML has proven its potential in seismology in recent years. Mainly, ML significantly helps seismological applications such as small earthquake event detection, earthquake phase picking, and earthquake early warning, but few studies use the ML to resolve the structure inversion problems.

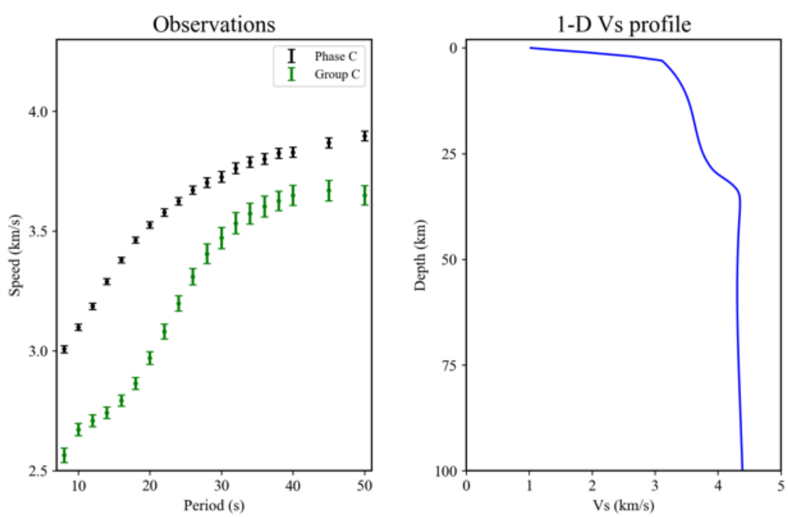

Figure 1. Example of input (left) and output (right) of the study at (151°W, 64.5°E, Alaska)

In this project, we try to investigate the application of ML in the classical surface wave inversion problem. When an earthquake (or another similar thing) happens, several different kinds of waves are generated and propagate through the earth. One of them is called the surface wave (including the Rayleigh wave and Love wave, we only use the Rayleigh wave in this study), and its speed varies when the wave oscillates in different periods. This relation (speed vs. period curve) is called the dispersion curve, which is observable and is the input of our project. This dispersion curve is controlled by the properties of earth rocks and how those properties vary in space. One possible simplification is to relate this curve only to one of the properties, shear velocity and its variation with depth. This shear velocity structure is our output of the inversion. Our problem space is to make an inversion, feed the input of observable dispersion curves to ML procedure to get the output, the estimation of shear velocity structure.

Fig 1. shows the input and output of our study. The input data includes the phase and group dispersion curves, and the output is the 1-D shear wave velocity (Vs) profile at the location (151° W, 64.5° E). This surface wave inversion problem can be solved by the nonlinear inversion approach, such as the Monte Carlo method, which can sample a broader range of model space randomly. But these methods could be very time-consuming with the increasing dimensionality of model space. In this study, we try to apply the convolutional neural network (CNN) method to two datasets we used in our research and compare the results using CNN with those from the Monte Carlo method (baseline model) to check the performance of the CNN method.

2. Approach

We use CNN to solve the inversion problem. That is to invert phase speed dispersion curves for the 1-D Vs model. The idea of this project is like Hu et al., 2020, which also use CNN to solve the surface wave inversion problem in China and Southern California.

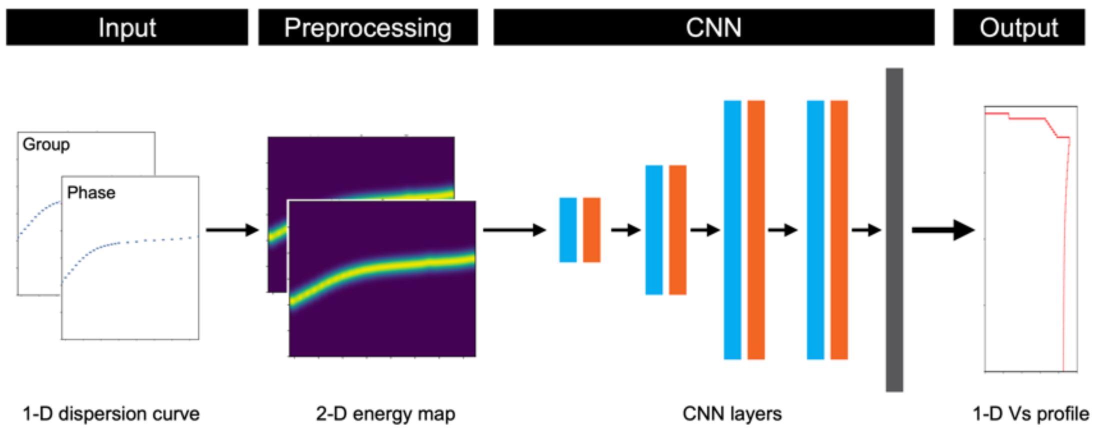

Figure 2. Workflow

To make full use of the ability of CNN on image processing, rather than using raw input, we transfer the 1-D dispersion curve to a 2-D energy map via a Gaussian function (Hu et al., 2021):

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

where v_0 is the value of phase speed at period T, v is the constant array from 2 to 5 km/s with an interval of 0.05 km/s, and r controls the radius of the Gaussian function. In this way, the height of the 2-D energy map is 60, and the width depends on the number of periods, N_T. The example of the transferred phase speed energy map is shown in Fig. 2 and Fig. 3b.

First, we generate the 1-D input dispersion curves for both phase velocity and group velocity or just phase velocity based on the reference Vs model for training. Then we transfer the 1-D input data into 2-D energy maps in preprocessing. The designed CNN then takes the 2-D energy maps as the input to perform training processing to predict the 1-D Vs model profile. The L-2 norm difference between the predicted Vs model and ground true Vs model is treated as the loss (objective) function. Once the training is done, the well-trained CNN can be used to predict Vs model efficiently given the input dataset at the study regions. The whole prediction processing is shown in Fig. 2.

As for the architecture of CNN, we use four convolutional layers, including 4, 8, 16, and 16 filters in each layer from shallow to deep and one fully connected layer in the end. ReLU activation function is used in each convolutional layer.

3. Dataset

We perform two applications in this project using the datasets in Alaska (section 4.2) and Cascadia (section 4.3). The example of the input data phase speed curve is shown in Fig. 3a. In the example, the surface wave phase velocity dispersion curve contains 24 periods in the range of 8-85 s (8, 10, 12, 14, 16, 28, 20, 22, 24, 26, 28, 30, 32, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, and 85 s). The input of CNN is the 2-D energy map (speed image) shown in Fig. 3b. The output of CNN is the 1-D Vs profile (Fig. 3c)

For the application in Alaska, we have both Rayleigh wave phase and group velocity observations from 8 to 50 s period, including 17 discrete periods at 547 geological locations as the input. At each location, we have 2×17×60 features.

For the application in Cascadia, we have only Rayleigh wave phase velocity observations from 8 to 80 s period, including 23 discrete periods at around 2500 geological locations as the input. In this way, we have 1×23×60 features for each location.

We perform two training with two CNN models for the applications in Alaska and Cascadia. The period ranges of the training data depend on the corresponding prediction dataset in either Alaska or Cascadia. To obtain the dataset used in training, we use the 1-D Vs models in the western US from Shen et al. 2016 as the reference model to generate the theoretical dispersion curves. In this way, from the view of the input energy map, we have 6803 1-D Vs models/entries in the training dataset. For each entry, we have 2×17×60 or 1×23×60 features for the applications in Alaska and Cascadia, respectively

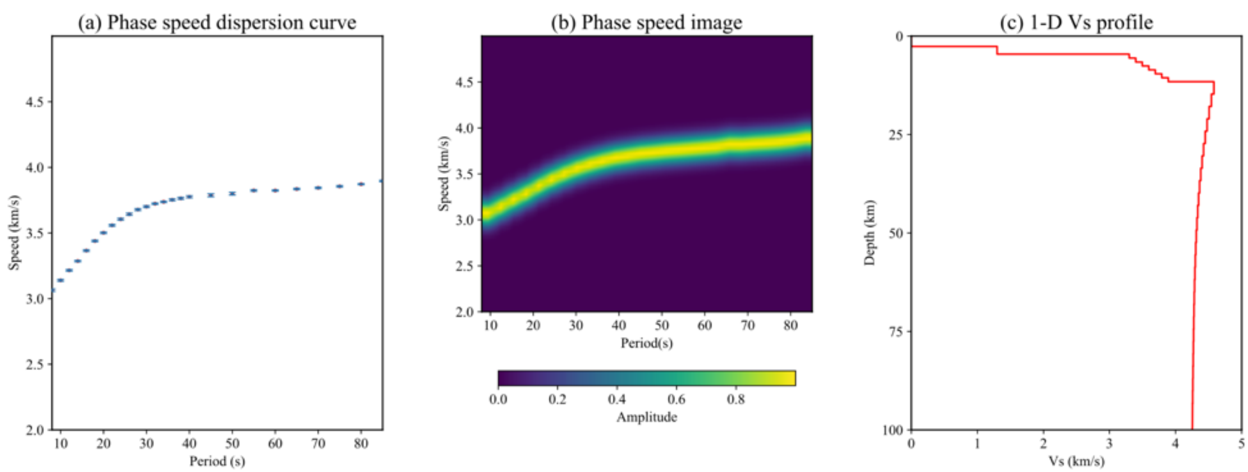

Figure 3. Examples for input dispersion curve (a), 2-D energy map (b), and 1-D output Vs profile (c).

4. Results

4.1 Training process

In the CNN training process, the model gets observable improvements in the first twenty epochs. Fig. 4 shows how the predicted Vs profile move close to the ground truth. The initial prediction is a homogeneous prediction with some random noise, which indicates the parameters are also initialized in this method. In several subsequent epochs, prediction gradually moves close to the truth and stops when it reaches the truth. Starting from about epoch 7, the prediction basically fits our ground truth and continue to make minor changes to further lower the loss function. Most points reach relative stability at epoch from 15 to 20 and after that, the improvements become non-observable.

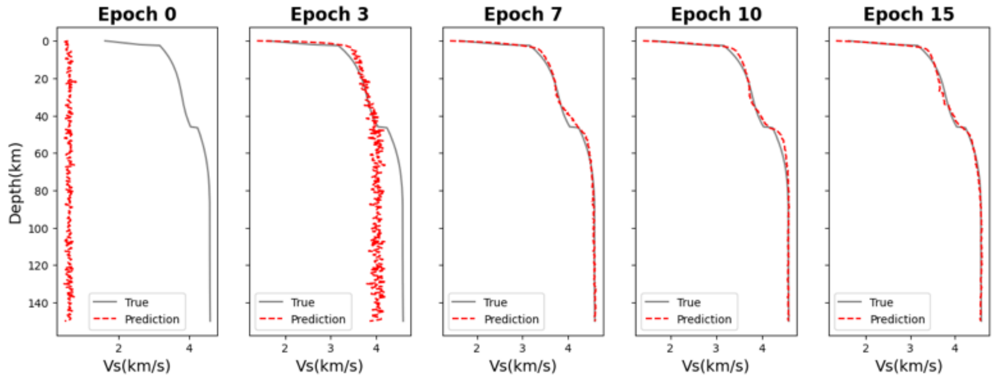

Figure 4. CNN training process

4.2 Application in Alaska

In application to Alaska, we imply the well-trained CNN to invert Vs models using the real Rayleigh wave dispersion curves of the input data. The input dataset from Feng & Ritzwoller (2019) comprises 547 grid points. At each grid location, the input dataset contains the observations, including both Rayleigh wave phase velocity and group velocity dispersion curves with corresponding uncertainty estimates in the period range from 8 to 50 s periods. The Bayesian Monte Carlo inversion model (MC) (Feng & Ritzwoller, 2019) is used as the baseline model to be compared with the CNN model both qualitatively and quantitatively. The whole prediction process takes less than 1 minute in contrast with the around 100 computing hours used by MC inversion on the same 20 core local server.

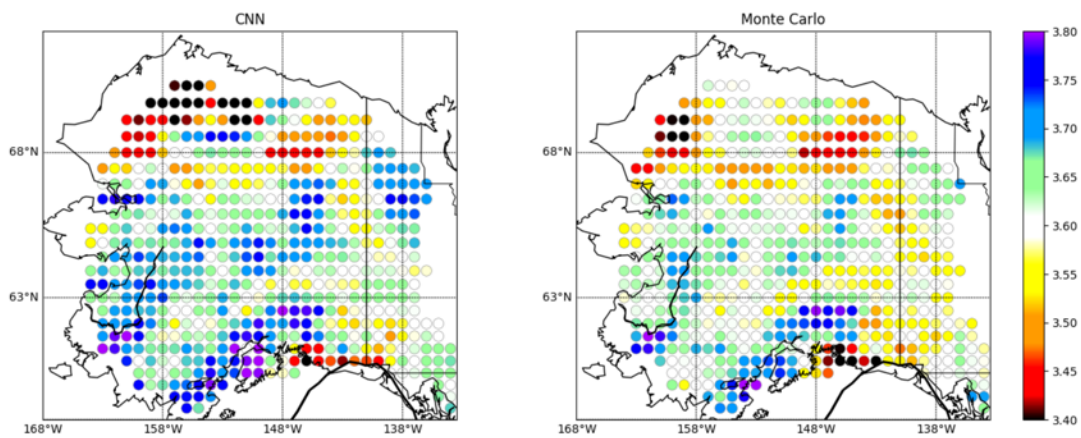

Figure 5. Comparison of Vs depth slice 20 km from CNN (left) and Monte Carlo (right).

Figs. 5 and 6 compare resulting Vs depth slices at 20 and 40 km. Because we use the Rayleigh wave dispersion curves from 8 to 50 s, the primary sensitive depth range is at the crust and uppermost mantle. Both methods generate similar Vs patterns at 20 and 40 km, but the spatial discrepancies still exist. At 20 km, the models primarily resolve the Vs in the crust. The CNN model shows higher velocities in southern Alaska and slower velocities in the Alaska North Slope than the MC model. However, at 40 km, the CNN model is systematically faster than the MC model in central Alaska

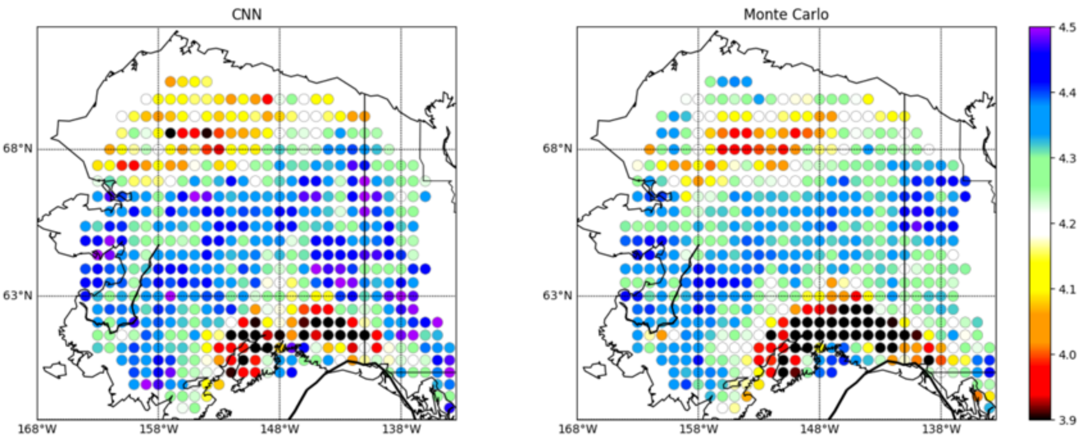

Figure 6. Same as Fig. 5 but for Vs depth slice at 40 km.

To compare the data fitting between input and predicted dispersion curves quantitatively at each grid location, we define data misfit as:

$$Misfit = \left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{d_i^{obs}-d_i^{pred}}{\sigma_i}\right)^2\right]^{1/2},$$

where d_i^{obs} is the observations of Rayleigh wave phase or group velocity, d_i^{pred} is the corresponding predictions by forward calculation using either CNN or Monte Carlo model, σ_i is the uncertainty estimates for the observation, *i* is the index of measurements, and *N* is the total number of measurements.

Fig. 7 shows the comparison of misfit maps from CNN and Monte Carlo inversion. Monte Carlo inversion achieves better data fitting in most of the inversion regions with an average misfit o 0.83. In contrast, CNN obtains an average misfit of 1.16, the data fitting is good in the central study region but gets worse in the margin region, especially in southern Alaska.

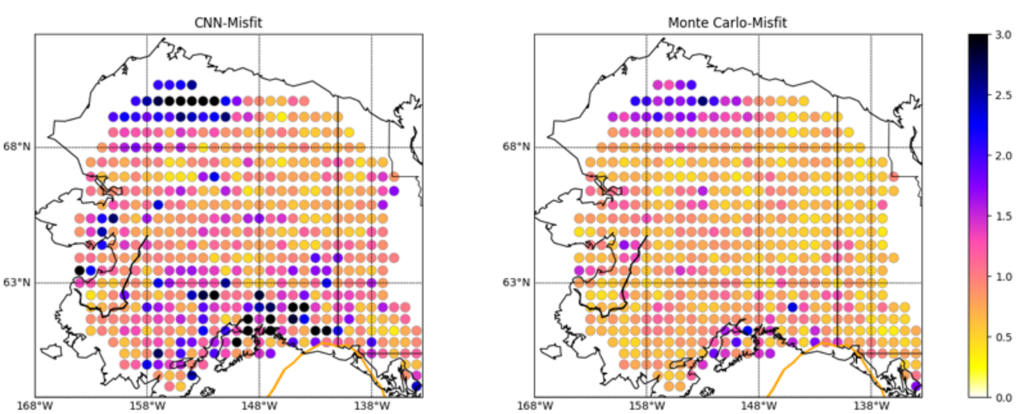

Figure 7. Maps of a misfit.

4.3 Application in Cascadia

In the application in Cascadia, our target is to resolve the Vs structures both onshore and offshore. Due to the low quality of group velocity measurements offshore, our Cascadia dataset only includes the Rayleigh wave phase velocity from 8 to 80 s (Zhang et al., 2021). That is different from the Alaska dataset, which contains both phase and group velocity measurements from 8 to 50 s. The more extended periods band also involves deeper structure information than just shorter periods. Our Cascadia amphibious dataset covers most of Washington and Oregon, part of northern California, and extends eastwards to the mid-ocean ridge. Besides, we are still using the same US structure model to generate a training dataset. Other settings are the same as the application in Alaska.

The inversion result is not as good as in Alaska. Two shear velocity slices are given below. Like in Alaska, CNN and Monte Carlo offer similar shear velocity patterns in continental part in both shallow and deep slice. However, at 25km depth, CNN shows much more considerable variation than the Monte Carlo method, and this significant variation is nearly impossible for earth material. When moving towards the deeper part, the different shallow structure between ocean and continent plays a less critical role in our input-output relation. And this results in the variation decrease to a similar value as Monte Carlo's result at 100km depth, though the absolute difference is still remarkable.

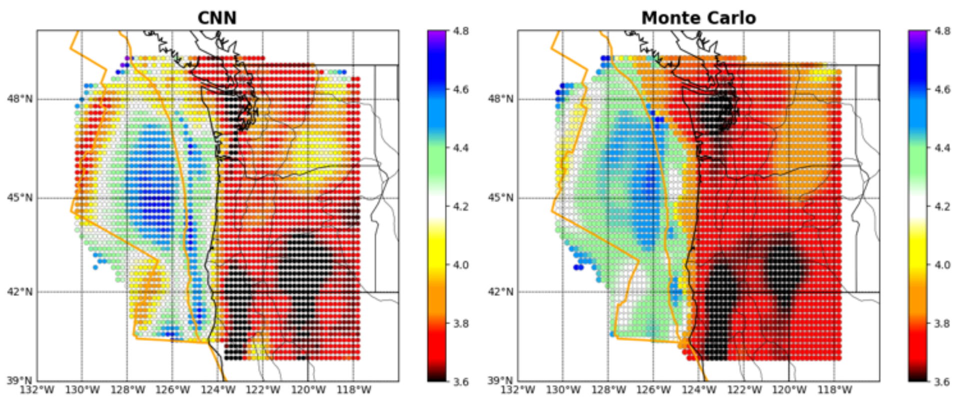

Figure 8. Shear velocity slice at 25km.

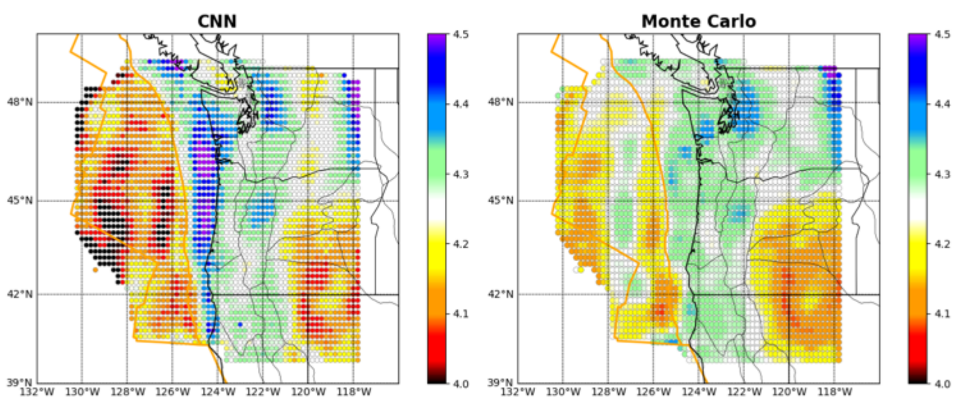

Figure 9. Shear velocity slice at 100km.

In the view of misfit, we could find our CNN model performs much worse than the Monte Carlo method in the oceanic part. The prediction from the CNN model gives misfits more than 5 in nearly all points, which is considered as not fitting at all in traditional evaluation criteria. The result of the Monte Carlo method looks much better, though its misfits also go above 5 near the coast. In continental region, CNN's performance is not stable. In some areas, misfits are higher than 3 or even 5. However, in the interval, the misfit could also decrease to around 1, which is usually considered as a threshold of the good fitting. The Monte Carlo method gives a much smoother image of misfit but might be higher than CNN in some specific points.

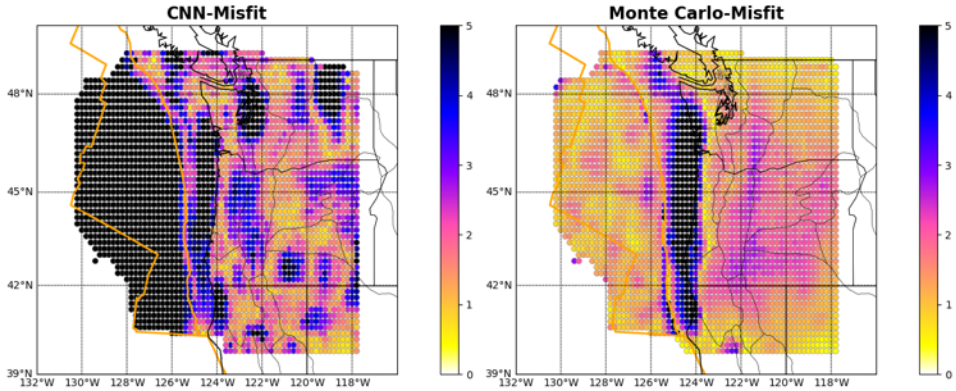

Figure 10. Maps of the misfit. Misfit values above 5 are all shown as black.

Like application in Alaska, though not as good as the traditional Monte-Carlo method, the CNN approach still eliminates meaningful geological features. The more significant and uneven misfit makes it not solid evidence for solving geoscience problems but might be good enough to be used as input in other techniques. However, it performs much worse in the ocean. Though we could still see some features, the very distinct shallow structure of the ocean plate results in a significant bias.

5. Discussion

In our two applications, CNN achieves a better result in Alaska than in Cascadia. The primary problem lies in the different geological structures between continental and oceanic environments.

In the application to Alaska, both the training and testing use the continental dataset, but for the application to Cascadia, using only the continental dataset for training is not a good choice. The shallow structure in the ocean area is very different from the continent. First, the oceanic plate is covered by water that shear waves could not travel through. Then the thickness of the oceanic crust is only 6-7 km, much thinner than the continental's average crust thickness of 30 km. Besides these major factors, many other relatively minor factors like composition differences could further alter the input-output relation. The big structure difference between oceanic and continent is also displayed using two example points in Fig. 11.

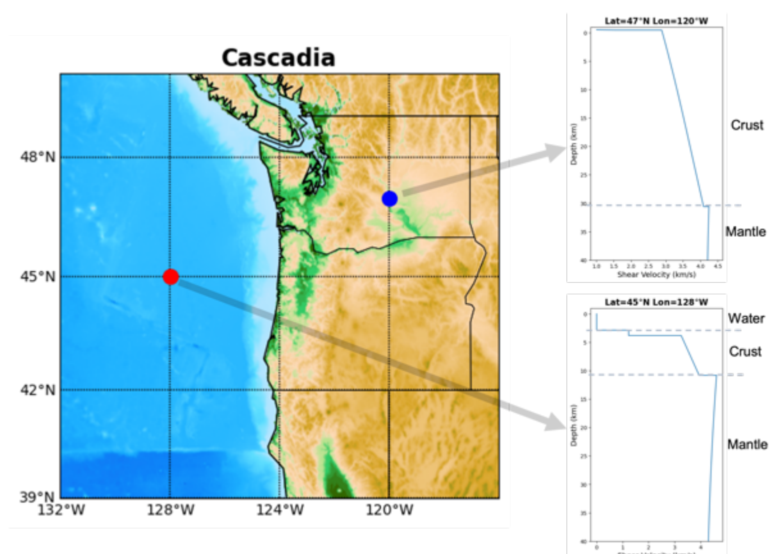

Figure 11 Left: Topography map of the research area. Right: Shear velocity profile examples.

So, to study structure in the ocean area, we need to introduce the dataset of the oceanic model in training. But limited by observation, there is no available oceanic model like the US model that is both solid enough to be treated as ground truth and large enough to contain enough variability in structure.

Overall, CNN works well for the Alaska dataset and continental region of the Cascadia dataset. The poor data fitting in the offshore Cascadia shows that oceanic structure must be included in training data to get a reasonable prediction. One approach to adding this structure is to use some theoretical method like the half-space cooling model. Further study is required to solve this problem.

In the future, more applications of CNN in the field of surface wave inversion are worth investigating, such as to resolve the inversion for the azimuthal and radial anisotropic model, and the joint inversion with ellipticity and receiver functions. In addition, the construction of a global seismic velocity model database including model at various geological environments will be beneficial to the training of applications at different regions.

Reference:

Kong, Q., Trugman, D. T., Ross, Z. E., Bianco, M. J., Meade, B. J., & Gerstoft, P. (2019). Machine Learning in Seismology: Turning Data into Insights. *Seismological Research Letters*, *90*(1), 3–14. https://doi.org/10.1785/0220180259

Hu, J., Qiu, H., Zhang, H., & Ben-Zion, Y. (2020). Using Deep Learning to Derive Shear-Wave Velocity Models from Surface-Wave Dispersion Data. *Seismological Research Letters*, *91*(3), 1738–1751. https://doi.org/10.1785/0220190222

Shen, W., & Ritzwoller, M. H. (2016). Crustal and uppermost mantle structure beneath the United States. Journal of Geophysical Research: Solid Earth, 121(6), 4306–4342. https://doi.org/10.1002/2016jb012887

Zhang, S., Wang, H., Wu, M., & Ritzwoller, M. H. (2021). Isotropic and azimuthally anisotropic Rayleigh wave dispersion across the Juan de Fuca and Gorda plates and US Cascadia from earthquake data and ambient noise two- and three-station interferometry. *Geophysical Journal International*, *226*(2), 862–883. https://doi.org/10.1093/gji/ggab142