Thank you for your careful review and constructive comments. We have studied all of your comments carefully and revised our manuscript. Followings are the response to the Editor, Referee 1 and Referee 2's comments. "Q" is a comment from the reviewer, "A" is a response to the comment and "Change" the lines and details of the modification. The comments are black in color and responses / changes are blue in color.

Editor's comments.

Q1. Line 67. "McCormack" (spelling)

A1. We modified the typo.

Change: Line 73.

Q2. Lines 95-98. The text and figure 1 do not match well here despite the figure being cited: no "residual learning" in figure 1; no ReLU in text; no "shortcut" in figure 1.

A2-1. We added "shortcut" in the Figure 1.

Change: Figure 1.

A2-2. Figure 1 is not needed to be cited here because Figure 1 is the actual network architecture used in this study and this part explains the characteristics of the DnCNN, thus we remove the sentence "The architecture of the DnCNN is shown in Fig. 1 and will be explained in more detail below". Figure 1 is now firstly cited in "Network architecture" section and we added ReLU in text.

Change: Line 140-142.

- We modified the sentence from "Fig. 1 shows the DnCNN architecture used in this study, where Conv and BN indicate convolution and batch normalization, respectively." to "Fig. 1 shows the DnCNN architecture used in this study, where Conv, BN, and ReLU indicate convolution, batch normalization, and rectified linear units (Krizhevsky et al., 2012), respectively."

A2-3. The network architecture using shortcut is called as residual learning and we explained the residual learning as "Residual learning was first suggested by He et al. (2016) and it applied residual block which consisted of several convolution processes and a shortcut connection to the neural network to overcome the problem of machine learning when networks delve deeper." Therefore, "residual learning" is not needed to be shown in Figure 1.

Q3. Line 103. "by adding noise-free seismic data (y) and noise (n)" -> "as a sum of "true" seismic signal (y) and noise (n)"

A3. We modified the sentence

Change: Line 109-110.

- The seismic data including noise (y) can be expressed as a sum of true seismic data (x) and noise (n) as follows:

Q4. Lines 108-116. This section is difficult for oceanographers owing to the use of several unfamiliar and unexplained terms: residual learning, residual network, network depth (and why should it increase?), residual unit, trainable nonlinear reaction diffusion, batch, batch normalization (there is a gap to its explanation), mini-batch.

A4. We tried to explain many of the difficult terms and removed several unnecessary terms. However, some terms are so basic that even beginners who are not very interested in artificial intelligence already know them. Currently, many researchers in many fields are very interested in artificial intelligence, and so is the oceanographers. Therefore, some terms were not explained because they were unnecessary to explain in the paper.

Changes: Line 116-117, Line 119-120, Line 104-105, Line 124-125.

- We modified the sentence from "it is different from the conventional residual network" to "it is different from the conventional neural network using residual learning (residual network)" to explain what is residual network.
- We removed trainable nonlinear reaction diffusion which is unnecessary.
- We modified "residual unit"" to "residual block". We modified the sentence from "Residual learning was first suggested by He et al. (2016) and it added the shortcut connection to the neural network to overcome the problem of machine learning when networks delve deeper." to "Residual learning was first suggested by He et al. (2016) and it applied residual block which consisted of several convolution processes and a shortcut connection to the neural network to overcome the problem of machine learning when networks delve deeper." to supplement the residual learning and explain what is residual block.
- We added the sentence "Instead of using entire training data at the same time, the machine learning sequentially uses mini-batch which is a small part of the entire data as input for the efficient training." to explain the mini-batch. Since the explanation of mini-batch has been supplemented, the batch normalization can be better understood by readers.

Q5. Lines 133-140. More difficulty for oceanographers owing to unfamiliar and unexplained terms: rectified linear units (previously mentioned but not explained), activation function, why add nonlinearity? Is "3x3xc" and reference to colour (c=3) necessary since (line 140) c = 1? Why 64 feature maps?

A5-1. If there is not a nonlinear activation function, the network with many layers can be expressed single layer network because the result of $\mathrm{Matrix1} \times \mathrm{Matrix2} \times \ldots \times \mathrm{MatrixN}$ is the same as a single matrix $\mathrm{Matrix0}$ (matrix multiplication is linear operation). However, if we add nonlinearity between the matrix multiplication, the multiplication cannot be expressed with a single matrix. This is why the nonlinear activation function is needed and more detail explanation can be found in Huang and Babri (1998).

Change: Line 146.

- We added the reference of Huang and Babri (1998).

Huang, G. B., and Babri, H. A.: Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions, IEEE transactions on neural networks, 9(1), 224–229, doi: 10.1109/72.655045, 1998.

A5-2. We modified and added the explanations of unfamiliar terms.

Changes: Line 146-150.

- We removed the sentence "The conventional DnCNN performs denoising from the image file (.jpg, .png, etc.), thus the size of the convolution filter is $3 \times 3 \times 3$ in the color image and $3 \times 3 \times 1$ in the gray image." because we did not use image file but binary file.
- We added "which is the same number of feature maps in Zhang et al. (2017)"
- We added the explanation of ReLU as "ReLU which returns input value for the positive input and 0 for the negative input is used for the activation function in this study"

Q6. Line 159. "multiplying by the square root of time"

A6. We modified "multiplying the square root of time" to "multiplying by the square root of time".

Change: Line 170.

Q7. Line 166. "Internal waves in the research area propagate"

A7. We modified "internal wave of the research area propagates" to "internal waves in the research area propagate"

Change: Line 179.

Q8. Line 172. "(assuming a constant density 1 g/cm3)" is only necessary for the XBT data. If you applied it to the XCTD data, did you ignore conductivity?

A8. We assumed a constant density for XBT and used measured density for XCTD.

Change: Line 185-186.

- We modified the sentence from "calculated with the XBT and XCTD data (assuming a constant density $1\ \mathrm{g/cm^3}$)" to "calculated with the XBT (assuming a constant density $1\ \mathrm{g\ cm^{-3}}$) and XCTD data"

Q8. Line 189. "that" -> "those".

A8. We modified "that" to "those"

Change: Line 206.

Q9. Lines 225, 227-228. What do you mean by "epoch", "iteration"? "is a process" does not explain them.

A9. We rewrote the explanation of "epoch" and "iteration."

Change: Line 242-249.

- The modified sentence is "In this study, training data were newly generated at every epoch with the fit_generator function in Keras (Keras Documentation, 2020). Therefore, the number of training data used in the training is determined by the size of mini-batch and the number of iterations per epoch. The number of training using a mini-batch is called as iteration and the number of training using entire training data is called as epoch. If one mini-batch pass through the training, one iteration ends. If all mini-batches pass through the training and entire training data has been used for training, one epoch

ends. The mini-batch size was 128 and the number of iteration of an epoch was 220, thus the fit_generator function generated 28,160 training data patches at every epoch."

Q10. Line 233. "or" -> "of"

A10. We modified "or" to "of"

Change: Line 256.

Q11. Line 234. What is the purpose of using the "Adam optimizer"?

A11. Adam optimizer minimizes the differences between the true data and estimated data during training. It gives better convergence than many other optimizers. Because the Zhang et al (2017) used adam optimizer in DnCNN, we also used adam optimizer. The adam optimizer is one of the most popular optimizers in the machine learning field and we gave the reference of the adam optimizer, thus we did not modify the sentence.

Q12. Line 332. "point" -> "patch".

A12. We modified "point" to "patch"

Change: Line 356.

Q13. Line 340. "data slope" and "horizontal slope" make no sense. You have not given the definition asked for by Referee 1.

A13. We modified the sentence by following Referee 1's suggestion

Change: Line 366-369.

- The sentence was modified to "The data slope spectrum is a slope spectrum obtained directly from seismic amplitude instead of tracked seismic reflections. The obtained horizontal wavenumber (k_x) spectrum of the seismic reflection amplitude is multiplied by $(2\pi k_x)^2$ to produce a data slope spectrum, which is useful in identifying noise contamination of seismic data to reveal the cutoffs from an internal wave to turbulence subrange"

Q14. Lines 395-396. ". . scaled the seismic sections again, multiplying the signal by the square-root of return time at each time step to make the amplitude correction."

A14. We modified the sentence.

Change: Line 423-425.

- The sentence was modified to "we scaled the seismic sections again, multiplying the signal by the square root of time at each time step (consequently multiplying the time at each time step) for the spherical divergence correction."

Q15. Pages 40-42, 45 are blank

A15. We removed blank pages.

Referee 1 comments

The authors have addressed my comments satisfactorily. The presentation is also improved with the revised figures. I think the paper can be published in OS.

We thank Referee 1 for detailed comments that improved the manuscript. Followings are the responses to each comment.

A few minor technical issues:

Q1. There are several blank pages.

A1. We removed blank pages.

Q2. li340. I think the description of the data slope spectrum is still confusing for the reader. Perhaps the narrative could be: obtain the horizontal wavenumber spectrum of the seismic reflection amplitude; multiply this by (2\pi k_x)^2 to produce a so-called "slope spectrum".

A1. We modified the sentence.

Change: Line 367-369.

- We modified the sentence as "The data slope spectrum is a slope spectrum obtained directly from seismic amplitude instead of tracked seismic reflections. The obtained horizontal wavenumber (k_x) spectrum of the seismic reflection amplitude is multiplied by $(2\pi k_x)^2$ to produce a data slope spectrum, which is useful in identifying noise contamination of seismic data to reveal the cutoffs from an internal wave to turbulence subrange (Holbrook et al., 2013; Fontin et al., 2017)."

Q3. li 421, which generate data

A3. We modified "generates" to "generate".

Change: Line 449.

Q4. Conclusions: perhaps rename as "Summary".

A4. We modified "Conclusions" to "Summary"

Change: Line 447.

Q5. li 437-438: the attenuation of the random noise is repeated in two sentences.

A5. We modified the sentence and rewrote the paragraph.

Change: Line 465-473.

- We removed "The observed random noise is successfully attenuated in the seismic section" and modified the sentence as "First, the calculated data slope spectrum indicates that a noise with a slope of k_x^2 is not removed completely at wavenumbers above 0.02 cpm." The rest of the paragraph is rewroted as "Therefore, future studies should include a detailed analysis of the slope spectra of the SO data and

establish an improved noise attenuation algorithm suitable for higher wavenumbers. Moreover, the data were collected and processed using 2D seismic exploration technology, which cannot efficiently deal with out-of-plane contamination. We expect that 3D seismic exploration can improve the resolution of SO data."

Q6. references: The authors do not give DOIs for the references.

A6. We added DOIs for the references where available.

Q7. conference papers: In physical oceanography papers, citation of conference papers, expanded abstracts are not very meaningful. We even avoid citing unpublished papers, theses etc.

A7. We also notice that citing the conference papers may not be very meaningful in many of research area. However, application of machine learning to the seismic data processing has become popular recently. Therefore, there are not many published full papers yet but many of studies are introduced in conference. Because of this reason, some of the references are citing the conference papers. When the expanded abstract was published in full paper, we modified the citation such as Liu et al. (2020).

Liu, D., Wang, W., Wang, X., Wang, C., Pei, J., and Chen, W.: Poststack Seismic Data Denoising Based on 3-D Convolutional Neural Network, IEEE Transactions on Geoscience and Remote Sensing, 58(3), 1598-1629, doi: 10.1109/TGRS.2019.2947149, 2020.

Q8. I was hoping the paper was going to be shorter but it came back with 20 figures. Font size in Figs 18, 19 and 20 can be larger. Sub-panel labels (a, b, c, d) can be moved to side or into the panels.

A8-1. This is the first study of applying the machine learning to attenuate the noise from the SO data. Therefore, the manuscript needed to include many detailed information of the machine learning and SO. In addition, we had to add a synthetic data test during the first revise, which required additional sub-section and several figures. For these reasons, the manuscript has been extended, but we think it is essential.

A8-2. We modified the fond size of Figs 18, 19, 20. The labels of the sub-panels can be relocated during the publication stage to make more suitable for publication.

Referee 2 comments

General comments– The paper has an improved balance of content, though I still think much of the detail of the filtering method could be moved to supplementary material as it will be of low interest to the principal readership of this journal.

Editor to advise if American spelling is acceptable and some phasing is still awkward in places but it does not detract from the understanding.

We thank Referee 2 for detailed comments that improved the manuscript.

Our study, we think, is the first attempt to attenuate noise in SO data using machine learning. Therefore, the readers will benefit much from our study if detailed procedures are illustrated in the manuscript. Followings are the responses to the comments and corrections.

Comments and corrections.

Q1. line 8 punctuation

A1. We modified "." to ","

Change: Line 8.

Q2. line 23: "dropping equipment" ? Rephrase as this is ambiguous

A2. We rephrased this sentence.

Change: Line 23-25.

- The sentence was modified from "Conventional physical oceanography measurements from cruises are performed by dropping equipment at observation stations" to "Conventional physical oceanography measurements from cruises are performed by deploying instrument such as a conductivity/temperature/depth (CTD), an expendable conductivity/temperature/depth (XCTD) or an expendable bathythermograph (XBT) at observation stations."

Q3. line 25: first published use of technique was by Gonella, J., Michon, D., 1988. Deep internal waves measured by seismic-reflection within the eastern Atlantic water mass. Comptes Rendus de l Academie Des Sciences Serie II 306, 781–787 (in French with English abstract). However it was Holbrook who first used the term seismic oceanography.

A3. We added Gonella and Michon (1988).

Gonella, J., and Michon, D.: Ondes internes profondes rèvèlèes par sismique rèflexion au sein des masses d'eua en atlantique-est, C. R. Acad. Sci., Ser. II, 306, 781–787, 1988.

Change: Line 28-31.

- The added sentence is "Seismic oceanography (SO) is a method that obtains the water column reflections via seismic exploration and analyzes seismic sections to estimate the oceanographic characteristics of sea water. It was firstly attempted by Gonella and Michon (1988) to measure deep internal waves in the

eastern Atlantic and later became widely known after the work of Holbrook et al. (2003).”

Q4. Line 27/28: “The differences in temperature and salinity between water column generate ...” poor English

A4. We modified the sentence.

Change: Line 31-34.

- The sentence was modified from “The differences in temperature and salinity between water column generate the difference in acoustic impedance, which reflect the seismic signals, and the reflected signals recorded at the receivers are processed to image the thermohaline fine structure of the ocean” to “The difference in temperature and salinity between water mass generates an acoustic impedance contrast, resulting in the reflection of seismic signals. The reflected seismic signals are processed to image the thermohaline fine structure of the ocean (Ruddick et al., 2009)”

Q5. line 29: need reference here suggest Ruddick, B., Song, H., Dong, C., Pinheiro, L., 2009. Water column seismic images as maps of temperature gradient. Oceanography 22 (1), 192–205.

A5. We added a reference.

Change: Line 34.

Q6. line 31: need reference for “conventional oceanographic methods” and link to the equipment that was dropped in line 23.

A6. We modified the sentence and added a reference.

Change: Line 36.

- We modified the sentence from “it has the advantage of generating data with a high horizontal resolution compared to conventional oceanographic methods” to “it has the advantage of generating data with improved horizontal resolution over conventional probe-based oceanographic methods (Dagnino et al., 2016).”

Q7. 34: redundant word “careful”

A7. We removed “careful”

Change: Line 40.

Q8. line 39: “...higher exploration expenses...” ambiguous as higher than what

A8. We modified the “higher exploration expenses” and added explanation.

Change: Line 45-46.

- We modified sentence from “SO also has the disadvantage of higher exploration expenses when using air guns and streamers that are several kilometers long” to “SO also has the disadvantage of high exploration expenses when using air guns and streamers of several kilometers long, which require large vessel and many operators.”

Q9. line 45: the vertical resolution of 1.5 m is based on the theoretical Rayleigh limit however in a 3D environment it is unlikely that you can achieve this due to interference from out-of-plane scattering.

A9. Piété et al. (2013) indicated that their data obtained by using SIG sparker with a 250Hz central frequency has the vertical resolution of 1.5 m, thus we mentioned "a high vertical resolution of 1.5m". They also seem to ignore the out-of-plane scattering issue, thus we modified the sentence

Change: Line 51.

- The sentence was modified from "a high vertical resolution of 1.5m" to "a high vertical resolution of approximately 1.5m"

Q10. Line 49/50: the maximum impedance contrasts are smaller

A10. We modified "impedance contrasts" to "maximum impedance contrasts."

Change: Line 55-56.

Q11. line 53 & 64: most noise sources are not random so the authors need to define which noise sources that consider as problematical, especially for SO data where a dominant noise is caused by pressure variations at the receiver due to waves on the surface. This peaks at frequencies below ~5 Hz so with sparker SO you should do better than airguns as you can set your low-cut filter at a higher frequency without damaging your primary signal bandwidth. I think what the authors are looking at here is noise generated from the ship as it has a distinct time characteristic and I would not be surprised if this was during periods of tidal flow against the direction of travel. So perhaps the ships propellers were run at a faster rate to maintain speed over the ground which has resulted in more cavitation. However, that period does not match with the statement that the ship speed was 5.5 knots. Suggest authors include auto-correlation functions both in space and time of the red boxes (Fig 3) indicated to justify that the synthetic random noise used in training datasets is appropriate.

A11. In seismic exploration, the term "noise" refers to the unwanted responses which are not meet the purpose of the exploration. There are many sources of noise such as feature of seismic source, ship, wave, weather, and even marine organisms (plankton, fish, etc.). The noise which has coherent characteristic is called as "coherent noise" (multiple, bubble effect, backscattering, etc.) and this can be removed during the seismic processing stage. The noise excluding coherent noise is generally referred to as "random noise". Directly removing random noise is difficult and random noise usually attenuated during the stacking procedure which can increase S/N ratio. However, when the random noise is strong and does not present white characteristics in the frequency band, removal becomes very difficult.

For the conventional noise attenuation methods, it was very important to estimate and analyze the source of noise and apply the appropriate algorithm. On the other hands, the machine learning proposed in this study uses random noise already measured in the survey area, thus the characteristics of noise for various sources can be automatically considered. Even the noise generated by the source that we have not yet figured out can also be taken into account. Of course, estimating and analyzing the source of noise is a very important issue, but we think detailed analysis

of the actual source of the noise is a little off from the machine learning perspective which is the main content of the paper. We added the explain of the possible source of the random noise in the seismic data and advantage of using suggested machine learning in the "East Sea SO data" section.

Change: Line 190-193.

- We added the sentence "Random noise in the seismic data can be created by rough weather conditions, ocean swells, a tail buoy jerk, the engine and the propeller of the vessel, etc. For the conventional noise attenuation methods, it is important to estimate the noise sources and their properties. In contrast, for machine learning-based noise attenuation suggested in this study, estimation of those propertied is not needed because noise itself is used as training data."

Q12. Figure 2 – does this show the whole line or just the portions after the shelf sections were omitted?

A12. The solid line in Figure 2 was the whole line of the exploration. We modified the Figure 2 to show the whole survey line and the study area simultaneously.

Change: Figure 2.

- The caption of Figure 2 was changed from "The black solid line is the survey line, and the black dashed lines with arrow indicate the exploration directions of lines 1 and 2, and red dots are the locations of XBTs and XCTDs." to "The solid line with gray and black color is the survey line. Gray line is shelf and slope parts which were removed from the seismic section during data processing and black line is the target area of this study. The black dashed lines with arrow indicate the exploration directions of lines 1 and 2, and red dots are the locations of XBTs and XCTDs."

Q13. Line 163: no comma in velocity value and possibly need to correct style to m s-1 and other units as necessary. Also this is ambiguous as 'water velocity' in the context of oceanography describes the mass movement of a body of water. An acceptable term is 'sound-speed'. As the temperature and salinity of the water is known (XCTD) then authors could do better than just use a generic sound speed by using the equations of state.

A13. We modified "constant velocity of 1,500 m/s" to "constant sound speed of 1500 $m\ s^{-1}$". We also modified "g/cm^3" to "$g\ cm^{-3}$". It would be better using true temperature and salinity of the water to calculate the sound speed of water instead of using 1500m/s. However, the difference between true sound speed and 1500 m/s is small and does not give meaningful differences in the seismic imaging. Because of this reason, several studies such as Tsuji et al. (2005), Holbrook et al. (2013), piete et al. (2013), Moon et al. (2016) also used constant sound speed for seismic data processing.

Changes: Line 174-175.

Q14. Line 173; define what you mean by very small as it looks like the calculated reflection coefficients are at the level of the minimum resolution of the XBT and given that you have two XCTD why did you use a constant density for those?

A14. We added reflection coefficient value in the sentence. We calculated reflection coefficients of the XCTD

data by using measured sound speed and density and rewrote the sentence.

Changes: Line 185-186.

- The sentence was modified from "The reflection coefficients are very small" to "The reflection coefficients are very small (~0.00005)". Figure 2 was re-plotted.
- The sentence "calculated with the XBT and XCTD data (assuming a constant density $1\ g\ cm^{-3}$)." was modified to "calculated with the XBT (assuming a constant density $1\ g\ cm^{-3}$) and XCTD data."

Q15. Line 209: justify why you used 1000 kg m^-3 for the density as true density is provided as part of the model – why does this change make the synthetic more SO-like if you don't apply a scaling an shift to the compressional wave velocities too. It sounds like a fix to overcome inefficiencies in your modelling process eg calling depth time and followed by a convolution. Given the level of adjustment and modification (see line 224), I see no purpose for using these sophisticated models as a set of time-shifted spikes (eg an enlarged version of Fig 16) would have probably done equally as well.

A15. To generate the noise-free synthetic seismic section, appropriate synthetic impedance model is necessary. There are several ways to obtain the synthetic impedance model: generating the synthetic model manually, modifying the existing synthetic model, directly using the existing synthetic model, etc. If we want to generate the synthetic model manually, it needs additional programing and computing cost. The easiest way is modifying existing synthetic model. The adjustment and modification of the synthetic models performed in this study is much easier and requires less time and effort compared to manually generating the synthetic model (similar to the model in Figure 16). Therefore, we used Marmousi and Sigsbee2A synthetic model to generate the synthetic seismic section.

There is no advantage of using synthetic true density instead of constant density in the generation of synthetic seismic section. Therefore, we just assumed constant density.

Q16. Line 210: were the simulations acoustic or elastic?

A16. Since we used acoustic impedance model, it is acoustic.

Q17. Line 247-260: you don't actually know the 'ground truth' as the section without added noise is, as you correctly point out, still contaminated with noise and will contain inter-layer peg-legs and out-of-plane events which you algorithm will also treat as primary signal.

A17. The SEZ field data still contains the noise because it is almost impossible to remove all kinds of noise in the seismic data. However, the "ground truth" term of the machine learning is the "true value" in the misfit function of the machine learning and the SEZ field data is used as the "true value" in the training data 1, thus we used "ground truth".

Q18. Line 275: what is x and y in equation 5

A18. We explained what is x and y.

Change: Line 299.

- The sentence is modified from "σ_{xy} is the covariance of x and y" to "σ_{xy} is the covariance of reference image (x) and test image (y)."

Q19. line 339 suggest a sub heading here to indicate you have changed to dealing with the SO data

A19. We added subheading "3.4 Calculation of the data slope spectrum from the synthetic seismic section"

Change: Line 364.

Q20. line 342 "and" → "to reveal"

A20. We modified "and" to "to reveal"

Change: Line 369.

Q21. lines 342-344 suggest: Holbrook et al. (2013) suggested analysing the slope of the spectrum for the complete data before calculating the slope of the spectrum from the water reflections because the noise that needs to be suppressed is more evident in the spectrum of the complete data.

A21. We modified the sentence

Change: Line 370-372.

- The modified sentence is "Holbrook et al. (2013) suggested analyzing the slope of the spectrum for the complete data before calculating the slope of the spectrum from the water reflections because the noise that needs to be suppressed is more evident in the spectrum of the complete data."

Q22. line 369 "clearly imaged" → "improved"

A22. We modified "clearly imaged" to "improved".

Change: Line 397.

Q23. Line 375-378, 385-387 & 389-393: I think you are making too much of this as the differences between the approaches is marginal as shown by your fig 20. Essentially both methods worked on the data to a sufficient degree to enable meaningful estimates of the slope of the internal wave to be discerned. If you feel you want to emphasise the benefit of D2 vs D1 then show a difference plot of the two.

A24. We removed comment of several minor differences between D1 model result and D2 model result.

Change: Line 404-406, Line 414

Q24. Line 397: what sound speed model did you use for conversion?

A24. We used constant sound speed model of 1500 m/s.

Change: Line 425.

- We added "using a constant sound speed of 1500 $\mathrm{m\ s^{-1}}$"

Q25. Line 406: would be useful to quote horizontal limit of resolution due to the Fresnel zone which I estimate, by eye-balling your data, to be ~0.02 cpm so above this value you would expect the amplitude to the perturbations to be strongly filtered so the fact the signal becomes noise limited is to be expected and further noise reduction will not provide any useful information. I note that your data appear to have a lower frequency content than I would expect – how do you explain that? What were the acquisition parameters? As these directly effect the data signal bandwidth.

A25. The horizontal resolution due to the first Fresnel zone radius (r) is approximately 17.3 m when the dominant frequency is 250 Hz, depth is 100 m (middle depth of the target zone), and sound speed is 1500 m/s based on following equation (Yilmaz, 2001)

$$r = \sqrt{\frac{z_0 \lambda}{2}}$$

z_0 is depth of reflector and λ is wavelength. The horizontal resolution of the depths ranging from 30 m to 200 m is approximately 9.5 m to 24.5 m.

The receiver interval is 6.25 m and we made super gather by using 4 neighboring CMPs (resulting 12.5m (0.08 cpm) interval of super gather CMP). Therefore, the horizontal resolution of the seismic data is larger than CMP interval at the section deeper than 52.1 m (r=12.5 m when depth is 52.1 m).

The horizontal wavenumbers of 30 m depth and 200 m depth is approximately 0.1 and 0.04 cpm, respectively. It means that if we can remove the noise at wavenumbers above 0.02 cpm, it would be possible to obtain more information. We mentioned that we made super gather during the processing stage and added the vertical and horizontal resolution of the East Sea data.

Change: Line 177-179.

- We added "The calculated vertical and horizontal resolution (Yilmaz, 2001) of the processed seismic section are approximately 1.5 m and 17.3 m when the central frequency is 250 Hz, sound speed is 1500 m s^{-1}, and depth is 100 m." to give the information of vertical and horizontal resolution in the "2.3 East Sea SO data" section.

Yilmaz, O.: Seismic data analysis: Processing, inversion, and interpretation of seismic data, second edition, Society of Exploration Geophysicists, Tulsa, Oklahoma, 2001.

Random Noise Attenuation of Sparker Seismic Oceanography Data with Machine Learning

Hyunggu Jun [1], Hyeong-Tae Jou[1], Chung-Ho Kim[1], Sang Hoon Lee[1], Han-Joon Kim[1]

[1]Korea Institute of Ocean Science & Technology, Busan, 49111, Republic of Korea

Correspondence to: Hyeong-Tae Jou (htjou@kiost.ac.kr)

Abstract. Seismic oceanography (SO) acquires water column reflections using controlled source seismology and provides high lateral resolution that enables the tracking of the thermohaline structure of the oceans. Most SO studies obtain data using air guns, which can produce acoustic energy below 100 Hz bandwidth~~.~~, with ~~a~~ vertical resolution of approximately ten meters or more. For higher-frequency bands, with vertical resolution ranging from several centimeters to several meters, a smaller, low-cost seismic exploration system may be used, such as a sparker source with central frequencies of 250 Hz or higher. However, the sparker source has a relatively low energy compared to air guns and consequently produces data with a lower signal-to-noise (S/N) ratio. To attenuate the random noise and extract reliable signal from the low S/N ratio of sparker SO data without distorting the true shape and amplitude of water column reflections, we applied machine learning. Specifically, we used a denoising convolutional neural network (DnCNN) that efficiently~~successfully~~ suppresses random noise in a natural image. One of the most important factors of machine learning is the generation of an appropriate training dataset. We generated two different training datasets using synthetic and field data. Models trained with the different training datasets were applied to the test data, and the denoised results were quantitatively compared. To demonstrate the technique, the trained models were applied to an SO sparker seismic dataset acquired in the East Sea and the denoised seismic sections were evaluated. The results show that machine learning can successfully attenuate the random noise in sparker water column seismic reflection data.

1 Introduction

Conventional physical oceanography measurements from cruises are performed by ~~dropping~~ deploying ~~equipment~~ instrument such as a conductivity-temperature-depth (CTD), an expendable conductivity-temperature-depth (XCTD) or an expendable bathythermograph (XBT) at observation stations. In general, due to time and cost limitations, the distance between observation points is large, from hundreds of meters to tens of kilometers; thus, the acquired water column information has a low horizontal resolution. ~~Holbrook et al. (2003) suggested a seismic oceanography (SO) method that obtained water column reflections via seismic exploration and analyzed seismic sections to estimate the oceanographic characteristics of sea water.~~ Seismic oceanography (SO) is a method that obtains the water column reflections via seismic exploration and analyzes seismic sections to estimate the oceanographic characteristics of sea water. It was firstly attempted by Gonella and Michon (1988) to measure deep internal waves in the eastern Atlantic and later became widely known after the work of Holbrook et al. (2003). The difference~~s~~ in temperature and salinity between water ~~column~~ mass generates an acoustic impedance contrast, resulting in the reflection of seismic signals.~~the difference in acoustic impedance, which reflect the seismic signals, and t~~ The reflected seismic signals ~~recorded at the receivers~~ are processed to image the thermohaline fine structure of the ocean (Ruddick et al., 2009). Seismic exploration acquires data continuously in the horizontal direction; thus, it has the advantage of generating data with ~~a high~~improved horizontal resolution over~~compared to~~ conventional probe-based oceanographic methods (Dagnino et al., 2016). Therefore, SO is used to image the structure of water layers (Tsuji et al., 2005; Sheen et al., 2012; Piété et al., 2013; Moon et al., 2017) and provide quantitative information such as physical properties (i.e, temperature, salinity) (Papenberg et al., 2010; Blacic et al., 2016; Dagnino et al. 2016; Jun et al., 2019) or the spectral distribution of the internal wave and turbulence (Sheen et al., 2009; Holbrook et al., 2013; Fortin et al. 2016) after ~~careful~~ analysis where temperature or salinity contrasts produce clear seismic reflections.

SO has been conducted mainly using air guns, a high-energy source, and the central frequency, the geometric center of the frequency band (Wang, 2015), of air guns is usually below 100 Hz. Therefore, the vertical resolution of the acquired seismic data using air guns is approximately ten meters or more, which is lower than that of conventional physical oceanography observation equipment. SO also has the disadvantage of high~~er~~ exploration expenses when using air guns and streamers ~~that are~~of several kilometers long, which require large vessel and many operators. Ruddick (2018) highlighted the limitations of current SO studies using multichannel seismic (MCS) exploration and argued that using a small-scale source instead of a large-scale air gun and a relatively shorter streamer with a length shorter than 500 m can make SO more widely available.

Piété et al. (2013) implemented a sparker source with a central frequency of 250 Hz and a short 450-m streamer (72 channels at 6.25 m intervals) to examine the oceanographic structure. Since high-frequency band sources were implemented, data with a high vertical resolution of approximately 1.5 m were acquired, and the short source signature enabled the thermocline structure to be imaged even in very shallow areas between 10 and 40 m. However, the signal-to-noise (S/N) ratio of the seismic section was lower than that of the air gun source, and the amplitude of the thermocline feature was small; thus, it was difficult to interpret. Generally, using a low-energy source and a short streamer in seismic exploration causes the low-S/N ratio problem.

This problem is more accentuated in SO because the maximum impedance contrasts between the water layers are smaller than the maximum impedance contrasts between the layers beneath the seabed. If a low-energy source is used, the water column reflections recorded by the receiver become too weak, and the influence of the background noise becomes larger than when using a high-energy source. The improvement in vertical resolution is evident when using higher-frequency band sources such as a sparker source; therefore, if appropriate methods can effectively suppress the random noise in the seismic section, more useful information can be derived compared to SO data using an air gun source.

There are various types of noise recorded by the receiver in seismic exploration, and several data processing steps are usually applied to the seismic data to attenuate noise. However, the noise attenuation method not only removes noise but also potentially alters important seismic signals (Jun et al., 2014). Especially for SO data, careful processing is essential to recover the actual shape of the water column reflections (Fortin et al., 2016), which contain internal wave and turbulence information. It is difficult to apply various noise attenuation methods to SO data because analyzing the internal wave and turbulent subranges of the water column requires the horizontal wavenumber spectrum (Klymak and Moum, 2007) of the seismic data, which is liable to be damaged by data processing. Therefore, minimized noise attenuation processes have been applied to SO data, and for this reason, studies calculating the wavenumber spectrum by using SO data such as those by Holbrook et al. (2013) and Fortin et al. (2016, 2017) have only applied bandpass and notch filters to remove random and harmonic noise. However, when the sparker is used as a seismic source, the bandpass filter alone is not sufficient to attenuate random noise, resulting in great difficulties in analyzing the wavenumber spectrum. Therefore, it is necessary to apply additional data processing to properly attenuate noise without damaging the wavenumber characteristics of SO data.

The use of artificial intelligence (AI) has been studied in geophysics for decades (McCormack, 1991; McCormack et al., 1993; Van der Baan and Jutten, 2000), but recent advances in computer resources and algorithms have spurred AI research, and several studies have been conducted to apply machine learning in the field of seismic data processing (Araya-Polo et al., 2019; Yang and Ma, 2019; Zhao et al., 2019). Among them, one of the most actively studied areas is prestack and poststack data noise attenuation. After convolutional neural networks (CNNs) were introduced, various noise attenuation methods based on the CNN architecture have been proposed (Jian and Seung, 2009; Gordonara, 2016; Lefkimmiatis, 2017), and the denoising convolutional neural network (DnCNN) suggested by Zhang et al. (2017) attained good results in random noise suppression in natural images. Recently, the DnCNN was applied to attenuate various types of noise in seismic data (Li et al., 2018; Si and Yuan, 2018; Liu et al., ~~2018~~2020). The DnCNN uses residual learning (He et al., 2016) and has the advantage of minimizing damage to the seismic signal by estimating the noise from seismic data rather than directly analyzing the signal. The original shape of the water column reflector in SO data remains unchanged during data processing, so the DnCNN, which learns noise characteristics, is a suitable SO data denoising algorithm.

As important as the proper neural network architecture when conducting training through machine learning is the use of an appropriate training dataset. When using the DnCNN to attenuate noise, the training data require noise-free and noise-only (or noise containing) data. In this study, we use both field and synthetic data as training data and compare which training data are more suitable for the DnCNN in attenuating random noise in SO data.

First, we introduce the DnCNN architecture used in this study and explain the construction method for the training and test datasets using field and synthetic data, respectively. Then, we perform training using the constructed training datasets and verify the trained models using test datasets. Finally, the trained models are applied to the East Sea sparker SO data, and the results are compared and evaluated.

2 Data and Methodology

2.1 Review of the DnCNN

The purpose of this study is to attenuate the random noise in sparker SO data, and the machine learning architecture used in this study is the DnCNN, which was suggested by Zhang et al. (2017). DnCNN is a neural network architecture based on the CNN for the purpose of removing the random noise in natural images. DnCNN reads the noisy image in the input layer and extracts the noise from the noisy image in the hidden layer. A layer is a module containing several computing processes (e.g. convolution, pooling or activation). At the output layer, the extracted noise is subtracted from the noisy image and generates the denoised result. ~~The architecture of the DnCNN is shown in Fig. 1 and will be explained in more detail below.~~ The DnCNN has three distinctive characteristics: 1) residual learning, 2) batch normalization, and 3) the same input and output data size for each layer.

Residual learning was first suggested by He et al. (2016) and it applied residual block which consisted of several convolution processes and ~~added the~~ a shortcut connection to the neural network to overcome the problem of machine learning when networks delve deeper. The DnCNN adopted residual learning and a single shortcut to estimate the noise from natural images. The estimated noise was subtracted from the noisy natural image, and the noise-attenuated image remained. If the DnCNN is applied to seismic data denoising, the target noise is estimated from the noisy prestack or poststack seismic data, and the estimated noise is subtracted from the noisy seismic data. The seismic data including noise (y) can be expressed as a sum of true seismic data~~by adding noise-free seismic data~~ (x) and noise (n) as follows:

$$y = x + n. \tag{1}$$

When the deep learning architecture that estimates noise from the noisy seismic data is $\boldsymbol{D}(y; n)$, the cost function of the DnCNN (C) can be expressed as follows:

$$\mathrm{C} = \frac{1}{2N}\sum_{i=1}^{N} \|\boldsymbol{D}(y_i; n_i) - (y_i - x_i)\|^2, \tag{2}$$

where n is the estimated noise from the original noisy seismic data (y), N is the number of the training data and $\|\ \ \|^2$ is the sum of squared errors (SSE). Although the DnCNN uses residual learning, it is different from the conventional neural network using residual learning (residual network). The conventional residual network utilizes residual learning to solve the performance degradation problem when the network depth increases; thus, it includes many residual ~~units~~blocks. On the other hand, the DnCNN uses residual learning to predict noise from noisy images~~, which is related to trainable nonlinear reaction diffusion (TNRD) (Chen and Pock, 2016)~~ and includes a single residual ~~unit~~block. For example, ResNet (He et al., 2016), which is a well-known image recognition network using residual learning, has more than tens or hundreds of network depth layers with many residual blocks, but the DnCNN has fewer than 20 network depth layers with a single residual block.

Moreover, the DnCNN applies batch normalization (Ioffe and Szegedy, 2015) after each convolution layer ~~to transform the mini-batch data distribution~~. Instead of using entire training data at the same time, the machine learning sequentially uses mini-batch which is a small part of the entire data as input for the efficient training. The distribution of input data varies during training and the neural network has a risk of updating the weights in the wrong direction. Batch normalization is a method to normalize the distribution of each mini-batch by making the mean and variance of the mini-batch equal to 0 and 1, respectively. The normalized mini-batch is transformed through scaling and shifting. Batch normalization is widely used in many deep learning neural networks because it can stabilize learning and increase the learning speed (Ioffe and Szegedy, 2015). The authors of the DnCNN empirically found that residual learning and batch normalization create a synergistic effect. In addition, unlike the encoder-decoder type denoising architecture, the size of the input data of the DnCNN is the same as the size of the output data in each layer. The DnCNN directly pads zeros at the boundaries during convolution and does not contain any pooling layer; thus, the data size remains unchanged during training. This procedure has the advantage of minimizing the data loss occurring during the encoding and decoding process. As mentioned above, the amplitude and shape of the seismic reflections are important for spectrum analysis using SO data. To minimize possible deformation of the seismic signals during the denoising procedure, the DnCNN, which predicts noise using residual learning and avoids information loss due to the absence of an encoding-decoding model, could be an appropriate algorithm.

2.2 Network architecture

The DnCNN uses three different kinds of layers, and we use the same layers as suggested by Zhang et al. (2017). Fig. 1 shows the DnCNN architecture used in this study, where Conv ~~and~~, BN, and ReLU indicate convolution ~~and~~, batch normalization, and rectified linear units (Krizhevsky et al., 2012), respectively.

The first layer type consists of ~~"~~convolution + rectified linear units ~~(ReLUs; Krizhevsky et al. (2012))"~~ and is used only at the first layer of the network architecture. This layer is shown as "Conv+ReLU" in Fig. 1. In the convolution process, 2-dimensional convolution between a certain size of kernel and data is performed. The outputs of the convolution process are passed through the activation function to add the nonlinearity in the network (Huang and Babri, 1998). ReLU which returns input value for the positive input and 0 for the negative input is used for the activation function in this study. The size of the convolution filter is $3 \times 3 \times c$ and generates 64 feature maps which is the same number of feature maps in Zhang et al. (2017), where c is the number of channels of the input data. ~~The conventional DnCNN performs denoising from the image file (.jpg, .png, etc.), thus the size of the convolution filter is $3 \times 3 \times 3$ in the color image and $3 \times 3 \times 1$ in the gray image.~~ This study extracts noise from binary files, ~~and~~ thus a $3 \times 3 \times 1$ convolution filter is adopted. The second layer type consists of convolution + batch normalization + ~~ReLUs~~ rectified linear units and is applied from layers 2 to L-1, where L is the total number of network layers. This layer is shown as "Conv+BN+ReLU" in Fig. 1. Sixty-four $3 \times 3 \times 64$ convolution filters are used because the number of feature maps of the hidden layer is 64, which is the same for all hidden layers. After convolution, batch normalization and the ReLU activation function are applied. The third layer type is convolution and uses only the last

layer to generate output noise data, and one $3 \times 3 \times 64$ convolution filter is used. This layer is shown as "Conv" in Fig. 1. After training is completed, the predicted noise is subtracted from the input data to produce denoised data.

2.3 East Sea SO data

The purpose of this study is to attenuate the random noise in the East Sea sparker SO data. The East Sea sparker SO data were obtained with a 5,000-J SIG PULSE L5 sparker source to investigate the propagation of the internal tide and characteristics of turbulent mixing. Two seismic lines were explored: line 1 traveled from southwest to northeast, and line 2 traveled from northeast to southwest (Fig. 2). The survey was performed from October 7th to 11th in 2018 (approximately 38 hours for one line) and the vessel speed was 5.5 knots. The seismic data include the shallow continental shelf and slope with a water depth of ~ 200 m, but we removed the continental shelf and slope area and used 280.4 km of line 1 and 280.9 km of line 2 because the data from these sections did not target the layers below the sea floor but the water layer. The shot interval was approximately 15 m, and 24 receivers were used at intervals of 6.25 m.

The acquired seismic data were processed through conventional time processing consisting of instrument delay and amplitude corrections, bandpass filtering, common-midpoint (CMP) sorting and stacking. Amplitude correction was performed by empirically multiplying by the square root of time at each time step. The corner frequencies of the trapezoidal bandpass filter (Dickinson et al. 2017) were 60-80-250-300 Hz, which were higher than those in air gun seismic data processing. Sparker source data have a lower S/N ratio due to the weak energy source compared to air gun source data and generally rely on a shorter streamer length; thus, it is common to generate supergathers (Piété et al., 2013) to enhance the S/N ratio. We combined 4 neighboring CMP gathers (Tang et al., 2016) to construct one supergather. A constant ~~velocity~~ sound speed of 1~~,~~500 m s^{-1} ~~m/s~~ was adopted for normal move-out. After CMP stacking, data recorded before 0.03 s were eliminated from the stack section because only direct waves and noise were present, and water layer reflections were rarely recorded. The processed seismic sections are shown in Fig. 3. The calculated vertical and horizontal resolution (Yilmaz, 2001) of the processed seismic section are approximately 1.5 m and 17.3 m when the central frequency is 250 Hz, sound speed is 1500 m s^{-1}, and reflector depth is 100 m. The internal waves in~~of~~ the research area propagate~~s~~ above a depth of 200 m, which is approximately 0.26 s in the seismic section. In addition, the physical properties of the research area were measured with oceanographic equipment, such as ~~expendable conductivity/temperature/depth (~~XCTD~~)~~ and ~~expendable bathythermograph (~~XBT~~) instruments~~, during exploration. Fig. 4 (a) shows the temperature profiles from two XBTs and two XCTDs. From the measurement data, the mixed layer ranged from the sea surface to a depth of 30 m, the depth of the thermocline ranged approximately from 30 to 200 m and deep water occurred below approximately 200 m depth. Fig. 4 (b) shows the reflection coefficients, defining the ratio between the reflected and incident wave, calculated with the XBT (assuming a constant density 1 g~~/~~ cm^{-3}) and XCTD data ~~(assuming a constant density 1 g/cm^3)~~. The reflection coefficients are very small (~0.00005) at depths shallower than 30 m, which seems to be the mixed layer, and deeper than approximately 200 m, which seems to be the deep water layer. Deep water exhibits a very slight water temperature/salinity variation with the depth, which makes it difficult to generate reflections, as indicated by

the seismic sections and reflection coefficients. Therefore, data after 0.28 s are considered random noise, and we used this part as noise data for the DnCNN. Random noise in the seismic data can be created by rough weather conditions, ocean swells, a tail buoy jerk, the engine and the propeller of the vessel, etc. For the conventional noise attenuation methods, it is important to estimate the noise sources and their properties. In contrast, for machine learning-based noise attenuation suggested in this study, estimation of those propertied is not needed because noise itself is used as training data.

2.4 Training data

The most important noise attenuation aspect of machine learning is generating an appropriate training dataset. Noise-free seismic sections (the ground truth) and sections with noise are required to generate the training dataset, and the training dataset can be constructed by combining these two datasets. As previously explained, the purpose of this study is to effectively attenuate noise in the water column seismic section acquired in the East Sea. Thus, the noisy section can be easily obtained by extracting the deep water zone of the water column seismic section without reflections. At this point, we assume that the random noise of the top and bottom parts of the water column seismic section exhibits similar features. The noise parts of the East Sea SO data are shown as red boxes in Fig 3. There are no notable reflections in the noise parts. However, it is almost impossible to obtain noise-free seismic sections from field data. Therefore, we constructed training datasets using two different methods and compared these datasets.

Training dataset 1 obtains the ground truth based on the field sparker seismic section below the sea floor. The reflection coefficients of the major reflectors below the sea floor are tens to hundreds of times larger than ~~that~~ those of the water column; thus, the seismic data below the sea floor have a better S/N ratio than the SO data. In addition, after the proper data processing steps, the S/N ratio of the seismic data beneath the sea bed can be further enhanced. We used 14 lines of field sparker seismic data targeting below the sea floor (SEZ data) acquired with the same equipment used to record the East Sea SO data. We used the interval from 0.2 to 0.6 s of the original data where the noise level is lower than in other parts of the data. A bandpass filter, FX-deconvolution, a Gaussian filter and noise muting above the sea floor were applied. Fig. 5 (a) shows an example of the SEZ data used to generate training dataset 1. This method has the advantage of using data with similar characteristics to those of the target data (the East Sea SO data) as the ground truth because the data are collected by the same equipment. Even if the S/N ratio of the sparker seismic data beneath the sea bed is relatively higher than that of the sparker seismic data of the water column and noise is suppressed during processing, it is difficult to completely eliminate noise from seismic data. Therefore, this method has the disadvantage that there is a possibility that the remaining noise would have a detrimental effect on training. Training dataset 2 uses synthetic data as the ground truth. The method for generating a synthetic seismic section from the velocity model is to perform time or depth domain processing using prestack synthetic data or to convolve the reflection coefficient and source wavelet. The former method has the advantage of generating synthetic seismic sections with features more similar to those of the actual field seismic section, but the generation and processing of prestack data are time consuming, and artificial noise is often generated during processing. The latter method has the advantage of generating noise-free seismic

sections with a very simple procedure. However, the generated synthetic seismic section has much different features from the target seismic section, which is, in this study, the East Sea water column sparker seismic section. Therefore, when the trained model is applied to the target seismic section, there is a risk that the trained model will regard the reflection signal as noise. In this study, we used the latter method to generate the ground truth because we needed to avoid artificial noise. Marmousi2 and Sigsbee2a synthetic velocity models with a constant density (1 g~~/~~ cm^{-3}) were employed to calculate the reflection coefficient, and the first derivative Gaussian wavelet was the synthetic source wavelet. The original Marmousi-2 and Sigsbee 2A synthetic velocity models are depth domain velocity models, but we assumed that these velocity models were time domain models to generate time domain seismic sections via 1-dimensional convolutional modeling. Fig. 5 (b) and (c) show the generated seismic sections of Marmousi-2 and Sigsbee 2A, respectively.

Each ground truth was first divided into with 300×300 sections. Then, amplitude values higher than the top 1% and lower than the bottom 1% were replaced by the top 1% and bottom 1% values, respectively, to prevent outliers from significantly affecting training. In addition, the outlier-removed ground truth and noise section were normalized to the maximum value of each section. This procedure balances the amplitudes of the ground truth and noise before generating the training dataset. Finally, training data with field seismic noise were generated by combining the ground truth and noise at a random ratio. Eq. 3 is the method to construct the training data, and Fig. 6 shows an example training data compilation.

$$T = r_1 \times G + r_2 \times N, \tag{3}$$

where T is the noise-added seismic patch (training data), G is the ground truth patch, N is the noise patch extracted from the noisy part of the East Sea seismic section (noisy data), and r_1 and r_2 are random values ranging from 0.2~0.8 ($r_1 + r_2 = 1$). The dimensions of T, G and N are 50×50; G and N were extracted at a random location of the ground truth and noisy section. To increase the number of ground truth data, data augmentation was applied by zooming in/out and randomly rotating or flipping the data. In this study, training data were newly generated at every epoch with the fit_generator function in Keras (Keras Documentation, 2020). Therefore, the number of training data used in the training is determined by the size of mini-batch and the number of iterations per epoch. The number of training using a mini-batch is called as iteration and the number of training using entire training data is called as epoch. If one mini-batch passes through the training, one iteration ends. If all mini-batches pass through the training and entire training data has been used for training, one epoch ends. ~~Training data were newly generated at every epoch with the fit_generator function in Keras (Keras Documentation, 2020).~~ The mini-batch size was 128 and the number of iteration of an epoch was 220, thus the fit_generator function generated 28,160 training data patches at every epoch. ~~The fit_generator function generated 28,160 data patches at every epoch because the mini-batch size was 128 and the iteration of each epoch was 220. The epoch is a process using all training data, and iteration is a process using a mini-batch; thus, an epoch usually consists of several iterations.~~

3 Training

3.1 Experimental setting

The experiment was conducted using 28,160 training data patches per epoch, and the size of each patch was 50×50. The mini-batch size was 128, the network depth which is the total number ~~or~~ of layers in the network architecture was 17, the number of feature maps of each layer was 64 and the Adam optimizer (Kingma and Ba, 2015) was implemented by following Zhang et al.'s (2017) DnCNN experiments. The network architecture used in this study is shown in Fig. 1. We performed training by using the two different training datasets generated from the field data (training dataset 1) and synthetic data (training dataset 2). The DnCNN model was trained for 40 epochs, and the total training time was approximately 1 hour using a single NVIDIA Quadro P4000 GPU.

3.2 Experiment using training dataset 1

Training dataset 1 was generated with the SEZ field data and noise obtained from the East Sea seismic section. After training the DnCNN model (D1 model) using training dataset 1, we evaluated the trained model against the test data. The test data were generated with the same procedure as that for the training data, and we used the other lines of SEZ data that were not used to generate the training data. Eighty-six 300×300 size test data were available, and we divided the test data into 50×50 size patches, which is the same size as the training data patch. Then, we discarded the remaining data divided by 128 (mini-batch size) for computational efficiency; thus, the number of test data points was 3,072. Fig. 7 shows 6 randomly selected test data subset patches, ground truths and denoised results after applying the D1 model at the 5^{th}, 10^{th} , 20^{th} and 40^{th} epochs. The depicted test data patches (1 to 6) include noise, but most of the noise has been successfully removed after training for 40 epochs. In the 3^{rd} and 6^{th} patches of test data subset in particular, the reflections are hardly recognized because of the severe noise, but the D1 model successfully attenuated the noise and generated a denoised section almost identical to the ground truth. In addition, there is a water layer without any signal at the top of the 4^{th} test data patch, and the trained model properly attenuated the noise at the water layer. This means that the trained model can determine those parts where no signal occurs.

However, the trained model using training dataset 1 has one problem. The ground truth of the 5^{th} test data patch contains noise in the bottom right part, and training dataset 1 might also contain noise in some parts of the ground truth. Even though the ground truth of training dataset 1 was generated from a processed sparker seismic section below the sea floor, noise still remained because it is almost impossible to perfectly remove noise from field data. The ground truth of training dataset 1 (SEZ data in Fig. 5 (a)) is obtained using the same equipment as was used for the East Sea SO data, which is the target of this study. Therefore, the ground truth signal has similar characteristics to the signal of the East Sea SO data, but its noise feature could also be similar to the noise of the East Sea SO data. This means that noise with similar characteristics would be trained to be eliminated in some cases and not in other cases during training. Training inconsistency can degrade the performance of the trained model.

To evaluate the test result quantitatively, we calculated the peak S/N ratio (PSNR) and structural similarity index measure (SSIM) by using entire test data. The PSNR reflects the amount of noise contained in the data and can be calculated as follows (Hore and Ziou, 2010):

$$PSNR = 20\,log_{10}(MAX_I) - 10\,log_{10}(MSE) \tag{4}$$

where MAX_I is the maximum value of the image and MSE is the mean squared error between the data with and without noise. The PSNR is high when noise is successfully removed, while the PSNR is low when noise is not sufficiently removed. Fig. 8 (a) shows the average PSNR and standard deviation of the test results. At the early stage of training, the average PSNR is low, which indicates that noise has not been sufficiently removed, but it increases as training progresses and converges at approximately 36 dB after 25 epochs. Even though the denoising algorithm attenuates noise successfully, the reflection shape, which is important information of the SO data, can be altered. Therefore, it is necessary to measure the structural distortion to verify the effectiveness of the proposed method. The SSIM is a quality metric that calculates the structural similarity between two datasets and can be calculated as follows (Hore and Ziou, 2010):

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_{\sout{u}y}^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{5}$$

where μ is the average, σ^2 is the variance, σ_{xy} is the covariance of reference image (x) and test image (y), and c is a stabilizing parameter. The value of SSIM ranges from 0 to 1, and if the structure is distorted during the denoising process, the SSIM will be low. On the other hand, the SSIM will be close to 1 if the denoised data are similar to the ground truth. Fig. 8 (b) shows the average SSIM of the test results. Similar to the PSNR result, the SSIM is also low at the early stage of training but increases as training progresses and converges at approximately 0.88 after 21 epochs. We also plotted the PSNR and SSIM histogram of the test data before and after applying the D1 model (40th epoch) in Fig. 9. Both the PSNR and SSIM are clearly improved after applying D1 model.

For seismic data, it is important to determine how well the actual amplitude and shape of the true reflection are recovered through the denoising process. Therefore, we extracted seismic traces from the denoised section and ground truth and compared the extracted traces, as shown in Fig. 10, to ensure that the trained model recovers the actual amplitude of the signal. We extracted the 20th (Fig. 10 (a)) and 30th (Fig. 10 (b)) vertical traces from the last (6th) patch of the test data, which had a size of 50x50. For the denoised trace, we extracted trace from the denoised patch of the 40th epoch. The amplitude and shape of the trace from the noisy data are different from those of the ground truth because the data are severely contaminated with noise. Even though there is much noise, the denoised traces have similar amplitudes and shapes to those of the ground truth. These results indicate that the DnCNN can recover important information of the true reflections and can be useful for random noise attenuation of sparker SO data.

3.3 Experiment using training dataset 2

Training dataset 2 was generated by using the modified Marmousi2 and Sigsbee2a synthetic seismic sections and noise obtained from the East Sea seismic section. After training the DnCNN model (D2 model) with training dataset 2, we evaluated the trained model against test data. The test data were generated with the same procedure used to generate the training data, and we selected part of the 1994 Amoco static test dataset (SEG Wiki, 2020), which is a different model from that used for the training data. The size of the test data patch was the same as that of the training data patch (50×50), and the number of test data points was 3,072. Fig. 11 shows 6 randomly selected test data subset patches, ground truths and denoised results after applying the D2 model at the 5th, 10th, 20th and 40th epochs. Even though the test data patches contain noise at different levels, the trained model at the 40th epoch attenuated most of the noise successfully and generated almost identical seismic sections to the ground truth. The second test data patch contained relatively little noise compared to other test data patches, and most of the noise was removed after approximately 10 training epochs. Test data patches 1 and 3 contained simple reflections with much noise, and the noise was sufficiently removed after approximately 20 training epochs. The noise in test data patches 4 and 6 was more severe than the noise in the other test data patches. After 40 training epochs, most of the noise was attenuated but not perfectly removed. The noise was dominant in test data patch 5, and only a weak signal existed in the bottom part of ground truth 5. If we evaluate the denoised result of the 5th test data patch, noise had been successfully removed, and only a weak signal remained in the bottom part of the patch after 40 training epochs. This indicates that the trained DnCNN model can accurately discriminate between signal and noise.

Unlike training dataset 1, training dataset 2 was generated with synthetic data. Therefore, it has the advantage of using noise-free seismic sections as the ground truth. In addition, generating many different kinds of synthetic seismic sections does not require much time or effort; thus, it is easy to increase the amount of training data compared to using field data as training data. However, the features of synthetic seismic sections can be different from those of the target data requiring noise attenuation because the synthetic seismic sections were generated by simply convolving the reflection coefficient with the source wavelet. Several studies have applied machine learning to field seismic data by training the model using synthetic data, such as automated fault detection with synthetic training data (Wu et al., 2018), but machine-learning-based noise attenuation of SO data using synthetic training data has not yet been studied.

Similar to the first experiment, the average PSNR (Fig. 12 (a)) and SSIM (Fig. 12 (b)) converged after approximately 25 epochs. The histograms of PSNR and SSIM of the test data before and after applying the D2 model (40th epoch) are also plotted in Fig. 13. As shown, the PSNR and SSIM are improved after DnCNN is applied. The average PSNR and SSIM in the second experiment are higher than those in the first experiment. These results could be caused by the use of a noise-free synthetic seismic section as the ground truth of training dataset 2 and might indicate that training dataset 2 is more appropriate for random noise attenuation of SO data. Fig. 14 shows the extracted traces before and after applying the D2 model. We extracted

the 20th (Fig. 14 (a)) and 30th (Fig. 14 (b)) vertical traces from the 1st patch of the test data. The denoised traces successfully recovered the true amplitude and shape, although the input data were severely contaminated by random noise.

In the second experiment, the noise-attenuated traces are closer to the ground truth traces than those in the first experiment. However, the comparison of the several extracted traces does not indicate which training data are more suitable for suppressing noise of sparker SO data. Therefore, we calculated the root-mean-square (RMS) error between the denoised test data and ground truth of the test data and evaluated which training data produced a lower RMS error. The RMS error was calculated as follows:

$$RMS\ error = \sqrt{\frac{1}{ntest}\sum_{i=1}^{ntest}\sum_{j=1}^{nnode}\left(g_{ij} - d_{ij}\right)^2} \quad (6)$$

where g is the ground truth of the test data, d is the denoised test data, $ntest$ is the number of test data patches (3,072) and $nnode$ is the size of each data ~~point~~patch (50×50). Even though test datasets 1 and 2 were generated using the same noisy data (the part containing noise of the East Sea SO section), the initial RMS errors of test datasets 1 and 2 before noise attenuation were different, 6.37 and 6.34, respectively, because noise was randomly extracted from the noise data. Therefore, we normalized the RMS error by that of the test data before noise attenuation. Fig. 15 illustrates the normalized RMS error of the first and second experiments at every epoch, and the normalized RMS errors were properly decreased in both results. The normalized errors converged at 0.27 in the first experiment and at 0.15 in the second experiment. The normalized RMS error of the second experiment is lower than that of the first experiment, indicating that the performance of the D2 model is better.

3.4 Calculation of the data slope spectrum from the synthetic seismic section

Water column reflection data can be used to obtain the physical oceanographic information by calculating the slope spectrum. The data slope spectrum is a ~~horizontal~~ slope spectrum obtained directly from seismic ~~data~~ amplitude instead of tracked seismic reflections. ~~by calculating the~~The obtained horizontal wavenumber (k_x) spectrum of the seismic reflection amplitude is multiplied by $(2\pi k_x)^2$ to produce a data slope spectrum, ~~and it~~which is useful ~~to identify~~in identifying noise contamination of seismic data ~~and~~ to reveal the cutoffs from an internal wave to turbulence subrange (Holbrook et al., 2013; Fontin et al., 2017). Holbrook et al. (2013) suggested analyzing ~~calculating~~ the data slope spectrum for the complete data before calculating the ~~reflector~~ slope spectrum from the water reflections because the ~~random~~ noise that ~~should be removed~~ needs to be suppressed is more~~before analyzing the seismic data becomes~~ evident in the spectrum of the complete data.~~data slope spectrum.~~ Therefore, we calculated and compared the data slope spectrum of noise-free, noise-added and noise-attenuated seismic data by using synthetic seismic section to verify that the proposed denoising method can recover the true data slope spectrum. The synthetic seismic section was generated by convolving the source wavelet with a randomly generated reflection coefficient section. Then, the noise extracted from the East sea SO data was added. Fig. 16 (a) shows the generated synthetic water column reflection section, and Fig. 16 (b) shows the noise added section. We applied the trained D1 model and D2 model to attenuate the noise, and the results are in Fig. 16 (c) (D1 model) and (d) (D2 model). Most of the noise was successfully attenuated, but the noise was not perfectly removed in the D1 model result at a distance from 20 to 25 km and depth from 140 to 180 m. Fig. 17 shows the calculated data slope spectra. The data slope spectrum of the noise-added section follows a k_x^2 slope, which is the slope of the random noise. After the noise attenuation, the data slope spectrum of the D2 model result (red line) follows the data slope spectrum of the noise-free section (greed line) almost identically. The data slope spectrum of the D1 model result (blue line) does not follow the noise slope, but the data slope spectrum is distorted compared to the noise-free data. The comparison of data slope spectra using synthetic data shows that the D2 model can recover the true data slope spectrum better than the D1 model.

4 Application to the East Sea SO data

The DnCNN models trained with training datasets 1 and 2 (the D1 and D2 models, respectively) were applied to the East Sea SO data. We applied the trained DnCNN models to the seismic sections from 0.03 to 0.28 s (approximately 22.5 to 210 m) where the reflections exist

Fig. 18 shows the results of applying the DnCNN to line 1. Fig 18 (a) is the line 1 seismic section from 0 to 0.28 s before the noise attenuation. The seismic section shallower than 0.03 s is dominated by noise from direct waves, which is muted at the data processing stage, and the section deeper than 0.28 s mainly contains random noise. Fig. 18 (b) and (c) are the denoised seismic section after applying the D1 and D2 model, respectively. In both results, most of the random noise was successfully removed, and the reflections became clearer. The strong random noise that occurred in the shallow part of the processed seismic sections was substantially attenuated, and the noise located between 150 and 200 km were also properly removed. Since noise was successfully attenuated, reflections that were difficult to distinguish due to a low S/N ratio were ~~clearly imaged~~improved. In particular, the weak signals between 0 and 50 km and between approximately 0.1 and 0.18 s became clearer after noise attenuation. Fig. 18 (d) and (e) are the estimated noise using the D1 and D2 model, respectively. As shown, both models successfully discriminated the noise component from the reflections; thus, the estimated noise sections are almost identical to the noise component of the processed seismic section. Even though both models successfully attenuated the noise in the seismic section of line 1, there are several differences. Reflections are not observed from 150 to 200 km and at approximately 0.2 s in the line 1 seismic section. The result from the D1 model still contains noise in that part, while the result from the D2 model contains lower noise levels compared to that from the D1 model. ~~In addition, for the weak reflections between 70 and 150 km and between 0.1 and 0.2 s, the reflections in the result from the D2 model are clearer and more continuous than those in the result from the D1 model.~~

Fig. 19 shows the results of applying the DnCNN to line 2. Fig .19 (a) shows the line 2 seismic section from 0 to 0.28 s before the noise attenuation. Fig. 19 (b) and (c) show the denoised seismic section after applying the D1 and D2 model, respectively. The seismic section of line 2 was contaminated by severe noise, but the D1 and D2 model properly removed the noise. In particular, the strong random noise located between 0 to 50 km was removed; thus, it became possible to recognize the reflections that were illegible. In addition, the reflections with steep slopes between 240 and 260 km and between 0.12 and 0.2 s were obscured by severe noise, but the D1 and D2 models successfully attenuated the noise and clearly recovered the reflections. However, similar to the line 1 result, the D2 model attenuated the noise better than the D1 model in some parts of the section. From 20 to 50 km ~~and 250 to 280 km~~, noise can still be observed when the D1 model is applied, but most of the noise has been sufficiently suppressed when the D2 model is applied.

Despite the successful noise attenuation of the D1 and D2 models, we found some differences. We presume that these differences are caused by the characteristics of the SEZ data which are the ground truth used to train the D1 model. The SEZ data are field data and contain noise to a certain degree because it is almost impossible to perfectly remove the noise from the field data. In other words, the D1 model is likely to regard the noise in the seismic section with similar characteristics to those

contained in the ground truth as a signal rather than noise. On the other hand, the D2 model does not suffer from this kind of problem because its ground truth is noise-free synthetic data.

To validate the noise attenuation results, we also calculated and compared the data slope spectra by using the outcome of the D1 and D2 models. Before calculating the data slope spectrum, we scaled the seismic sections again, ~~by~~ multiplying the signal by the square root of time ~~to~~ at each time step (consequently multiplying the time ~~to~~ at each time step) for the spherical divergence correction. Then, we converted the seismic section from the time axis to the depth axis using a constant sound speed of 1500 m s^{-1} and extracted the part from 150 to 175 km and at a depth from 75 to 150 m. Fig. 20 (a), (b) and (c) show the seismic sections extracted from the section before and after noise attenuation using models D1 and D2, respectively. The seismic section before noise attenuation was severely contaminated with random noise, but most of the noise was removed in the sections after noise attenuation. Fig. 20 (d) shows the calculated data slope spectra. From the KM07 model (Klymak and Moum, 2007), noise has a k_x^2 slope in the slope spectrum, and we plotted the k_x^2 slope with the green dashed line in Fig. 20 (d) for comparison. The data slope spectrum of the section before noise attenuation has a k_x^2 slope at wavenumbers above 0.002 cpm, which indicates that noise dominates these wavenumbers. Because of the severe noise, it is impossible to analyze the seismic data before noise attenuation. On the other hand, the data slope spectra after noise attenuation seem to contain internal waves subrange from 0.0015 to 0.006 cpm and turbulence subrange from 0.009 to 0.015 cpm that approximately follow the $k_x^{-1/2}$ (yellow dashed line) and $k_x^{1/3}$ (purple dashed line) slopes (Klymak and Moum, 2007), respectively. This result indicates that noise was properly attenuated and the seismic data could be analyzed, even though noise with a slope of k_x^2 still occurred at wavenumbers above 0.02 cpm. There is a shift in the data slope spectrum after noise attenuation at wavenumbers smaller than 0.001 cpm. This shift is also observed in the synthetic data slope spectrum experiments. In Fig. 17, there is a difference between the spectrum of the noise-added section and that of the noise-attenuated sections at wavenumbers smaller than 0.001 cpm. However, the difference is also observed between the spectrum of the noise-free section and that of the noise-added section. Therefore, this shift seems to be caused by the characteristic of the noise extracted from the East Sea SO data.

From the noise attenuation results obtained by applying the trained models to the East Sea sparker SO data, we showed that the DnCNN architecture used in this study can successfully suppress random noise. The comparison of the D1 and D2 model results showed that the training data generated using noise-free synthetic data are more suitable for random noise attenuation of sparker SO data than those generated using field data with a relatively high S/N ratio.

5 ~~Conclusions~~Summary

Random noise is one of the major obstacles in analyzing SO data. Conventionally, the noise in SO data has been attenuated through simple data processing methods because most of the SO data are obtained with air guns, which generate~~s~~ data with a high S/N ratio. However, the simple noise attenuation method is not sufficient for data with a low S/N ratio, such as sparker SO data. Despite the low S/N problem, the sparker source has advantage of generating a higher-frequency band signal than an air gun source, which can provide information with higher vertical resolution. Therefore, we applied machine learning to attenuate the random noise in East Sea sparker SO data, which contains ~~much~~ significant random noise. The DnCNN architecture was used to construct ~~the~~ a neural network, and training data were generated by combining the ground truth and noise extracted from the target seismic data at random amplitude ratios. Two different training datasets were generated, and they used either field or synthetic data as ~~the~~ a ground truth. The trained DnCNN models were applied to the test datasets that were generated with the same procedure of generating the training datasets. The test results were verified based on the PSNR, SSIM, trace extraction and normalized RMS error. The data slope spectrum test using synthetic seismic section was also performed. The test results revealed that both trained DnCNN models were able to successfully attenuate random noise and the training data generated using noise-free synthetic data showed better results than the training data generated using high-S/N ratio field data. We applied the trained DnCNN models to the East Sea sparker SO data, which is the target of this study, and the models successfully attenuated random noise. The comparison of the denoised seismic sections after applying the two different trained models also showed that the training dataset generated from the noise-free synthetic data was more suitable for sparker SO data noise attenuation than that generated from the high-S/N ratio field data.

Even though ~~the~~ random noise is almost completely attenuated in the seismic section, the proposed method still needs several improvements. ~~The observed random noise is successfully attenuated in the seismic section, but t~~First, the calculated data slope spectrum ~~still~~ indicates that ~~the section contains~~a noise with a slope of k_x^2 is not removed completely at wavenumbers above 0.02 cpm. Therefore, future studies should include a detailed analysis of the slope spectra of the ~~East Sea~~ SO data and establish an improved noise attenuation algorithm suitable for higher wavenumbers. Moreover, the data were collected and processed using 2D seismic exploration technology, which cannot efficiently deal with out-of-plane contamination. ~~which can degrade the seismic resolution during the data processing stage because of the limitations of 2D seismic exploration such as out-of-plane contamination.~~ We expect that 3D seismic exploration can improve the resolution of SO data.~~Therefore, to improve the resolution of SO data, it is necessary to acquire data by using 3D seismic exploration.~~

The network architecture used in this study is straightforward and efficient. In addition, the proposed method of generating the training dataset is very simple and easy because it only requires synthetic data, which are readily generated, and noise data, which can be extracted from the target seismic data. Moreover, only approximately one hour is required to train the DnCNN model with a single GPU. Therefore, the noise attenuation method suggested in this study ~~has the advantage that it~~ can be widely and easily applied in noise attenuation of the various kinds of SO data.

Data availability

The synthetic training data can be downloaded from https://github.com/hgjun1026/so_dncnn.git. The field data and program can be made available upon request to authors.

Author contribution

Hyunggu Jun and Hyeong-Tae Jou constructed the machine learning program and performed experiments. Chung-Ho Kim, Sang Hoon Lee and Han-Joon Kim acquired seismic data and performed data processing.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements.

This research is supported by the Korea Institute of Ocean Science and Technology (Grant number PE99841, PE99851) and the Korea Institute of Marine Science and Technology promotion (Grant number 20160247).

References

Araya-Polo, M., Farris, S., and Florez, M.: Deep learning-driven velocity model building workflow, The Leading Edge, 38(11), 872a1-872a9, doi:10.1190/tle38110872a1.1, 2019.

Blacic, T. M., Jun, H., Rosado, H., and Shin, C.: Smooth 2-D ocean sound speed from Laplace and Laplace-Fourier domain inversion of seismic oceanography data, Geophysical Research Letters, 43(3), 1211-1218, doi:10.1002/2015GL067421, 2016.

Chen, Y., and Pock, T.: Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration, IEEE transactions on pattern analysis and machine intelligence, 39(6), 1256-1272, doi: 10.1109/TPAMI.2016.2596743, 2016.

Dagnino, D., Sallares, V., Biescas, B., and Ranero, C. R.: Fine-scale thermohaline ocean structure retrieved with 2-D prestack full-waveform inversion of multichannel seismic data: Application to the Gulf of Cadiz (SW Iberia), Journal of Geophysical Research: Oceans, 121(8), 5452-5469, doi:10.1002/2016JC011844, 2016.

Dickinson, A., White, N. J., and Caulfield, C. P.: Spatial Variation of Diapycnal Diffusivity Estimated From Seismic Imaging of Internal Wave Field, Gulf of Mexico, Journal of Geophysical Research: Oceans, 122(12), 9827-9854, doi:10.1002/2017JC013352, 2017.

Fortin, W. F. J., Holbrook, W. S., and Schmitt, R. W.: Mapping turbulent diffusivity associated with oceanic internal lee waves offshore Costa Rica, Ocean Science, 12, 601–612, doi: 10.5194/os-12-601-2016, 2016.

Fortin, W. F., Holbrook, W. S., and Schmitt, R. W.: Seismic estimates of turbulent diffusivity and evidence of nonlinear internal wave forcing by geometric resonance in the South China Sea, Journal of Geophysical Research: Oceans, 122(10), 8063-8078, doi:10.1002/2017JC012690, 2017.

Gondara, L.: Medical image denoising using convolutional denoising autoencoders. In: 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), doi:10.1109/ICDMW.2016.0041, 241-246, 2016.

He, K., Zhang, X., Ren, S., and Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778, doi: 10.1109/CVPR.2016.90, 2016.

Holbrook, W. S., Páramo, P., Pearse, S., and Schmitt, R. W.: Thermohaline fine structure in an oceanographic front from seismic reflection profiling, Science, 301(5634), 821-824, doi:10.1126/science.1085116, 2003.

Holbrook, W. S., Fer, I., Schmitt, R. W., Lizarralde, D., Klymak, J. M., Helfrich, L. C., and Kubichek, R.: Estimating oceanic turbulence dissipation from seismic images, Journal of Atmospheric and Oceanic Technology, 30(8), 1767-1788, doi:10.1175/JTECH-D-12-00140.1, 2013.

Hore, A., and Ziou, D.: Image quality metrics: PSNR vs. SSIM, In: 2010 20th International Conference on Pattern Recognition, 2366-2369, doi:10.1109/ICPR.2010.579, 2010.

Huang, G. B., and Babri, H. A.: Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions, IEEE transactions on neural networks, 9(1), 224–229, doi: 10.1109/72.655045, 1998.

Ioffe, S., and Szegedy, C.: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, In: International Conference on Machine Learning, 448-456, 2015.

Jain, V., and Seung, S.: Natural image denoising with convolutional networks, In: Advances in Neural Information Processing Systems, 769-776, 2009.

Jun, H., Kim, Y., Shin, J., Shin, C., and Min, D. J.: Laplace-Fourier-domain elastic full-waveform inversion using time-domain modelling, Geophysics, 79(5), R195-R208, doi:10.1190/geo2013-0283.1, 2014.

Keras Documentation: https://keras.io/models/sequential/, last access: 20 March 2020.

Klymak, J. M., and Moum, J. N.: Oceanic isopycnal slope spectra. Part I: Internal waves, Journal of physical oceanography, 37(5), 1215-1231, doi:10.1175/JPO3073.1, 2007.

Krizhevsky, A., Sutskever, I., and Hinton, G. E.: Imagenet classification with deep convolutional neural networks, In: Advances in neural information processing systems, 1097-1105, doi:10.1145/3065386, 2012.

Lefkimmiatis, S.: Non-local color image denoising with convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3587-3596, doi:10.1109/CVPR.2017.623, 2017.

Li, H., Yang, W., and Yong, X.: Deep learning for ground-roll noise attenuation, In: SEG Technical Program Expanded Abstracts 2018, 1981-1985, doi:10.1190/segam2018-2981295.1, 2018.

Liu, D., Wang, W., Wang, X., Wang, C., Pei, J., and Chen, W.: Poststack Seismic Data Denoising Based on 3-D Convolutional Neural Network, IEEE Transactions on Geoscience and Remote Sensing, 58(3), 1598-1629, doi: 10.1109/TGRS.2019.2947149, 2020.

~~Liu, D., Wang, W., Chen, W., Wang, X., Zhou, Y., and Shi, Z.: Random noise suppression in seismic data: What can deep learning do?, In: SEG Technical Program Expanded Abstracts 2018, 2016-2020, 2018.~~

McCormack, M. D.: Neural computing in geophysics, The Leading Edge, 10(1), 11-15, doi:10.1190/1.1436771, 1991.

McCormack, M. D., Zaucha, D. E., and Dushek, D. W.: First-break refraction event picking and seismic data trace editing using neural networks, Geophysics, 58(1), 67-78, doi:10.1190/1.1443352, 1993.

Papenberg, C., Klaeschen, D., Krahmann, G., and Hobbs, R. W.: Ocean temperature and salinity inverted from combined hydrographic and seismic data, Geophysical Research Letters, 37(4), L04601, doi:10.1029/2009GL042115, 2010.

Piété, H., Marié, L., Marsset, B., Thomas, Y., and Gutscher, M. A.: Seismic reflection imaging of shallow oceanographic structures, Journal of Geophysical Research: Oceans, 118(5), 2329-2344, doi:10.1002/jgrc.20156, 2013.

Ruddick, B. R., Song, H., Dong, C., and Pinheiro, L.: Water column seismic images as maps of temperature gradient. Oceanography, 22(1), 192–205, doi:10.5670/oceanog.2009.19, 2009.

Ruddick, B. R.: Seismic Oceanography's Failure to Flourish: A Possible Solution, Journal of Geophysical Research: Oceans, 123(1), 4-7, doi:10.1002/2017JC013736, 2018.

SEG Wiki: https://wiki.seg.org/wiki/1994_BP_statics_benchmark_model, last access: 22 June 2020.

Sheen, K. L., White, N. J., and Hobbs, R. W.: Estimating mixing rates from seismic images of oceanic structure, Geophysical Research Letters, 36(24), doi:10.1029/2009GL040106, 2009.

Sheen, K. L., White, N. J., Caulfield, C. P., and Hobbs, R. W.: Seismic imaging of a large horizontal vortex at abyssal depths beneath the Sub-Antarctic Front, Nature Geoscience, 5(8), 542, doi:10.1038/ngeo1502, 2012.

Si, X., and Yuan, Y.: Random noise attenuation based on residual learning of deep convolutional neural network, In: SEG Technical Program Expanded Abstracts 2018, 1986-1990, doi:10.1190/segam2018-2985176.1, 2018.

Tang, Q., Hobbs, R., Zheng, C., Biescas, B., and Caiado, C.: Markov Chain Monte Carlo inversion of temperature and salinity structure of an internal solitary wave packet from marine seismic data, Journal of Geophysical Research: Oceans, 121(6), 3692-3709, doi:10.1002/2016JC011810, 2016.

Tsuji, T., Noguchi, T., Niino, H., Matsuoka, T., Nakamura, Y., Tokuyama, H., Kuramoto, S. and Bangs, N.: Two-dimensional mapping of fine structures in the Kuroshio Current using seismic reflection data, Geophysical Research Letters, 32(14), doi:10.1029/2005GL023095, 2005.

Van der Baan, M., and Jutten, C.: Neural networks in geophysical applications, Geophysics, 65(4), 1032-1047, doi:10.1190/1.1444797, 2000.

Wang, Y.: Frequencies of the Ricker wavelet, Geophysics, 80(2), A31-A37, doi:10.1190/geo2014-0441.1, 2015.

Wu, X., Liang, L., Shi, Y., and Fomel, S.: FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation, Geophysics, 84(3), IM35-IM45, doi:10.1190/geo2018-0646.1, 2019.

Yang, F., and Ma, J.: Deep-learning inversion: A next-generation seismic velocity model building method, Geophysics, 84(4), R583-R599, doi:10.1190/geo2018-0249.1, 2019.

Yilmaz, O.: Seismic data analysis: Processing, inversion, and interpretation of seismic data, second edition, Society of Exploration Geophysicists, Tulsa, Oklahoma, 2001.

Zhang, K., Zuo, W., Chen, Y., Meng, D., and Zhang, L.: Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising, IEEE Transactions on Image Processing, 26(7), 3142-3155, doi:10.1109/TIP.2017.2662206, 2017.

Zhao, X., Lu, P., Zhang, Y., Chen, J., and Li, X.: Swell-noise attenuation: A deep learning approach, The Leading Edge, 38(12), 934-942, doi:10.1190/tle38120934.1, 2019.

Figures

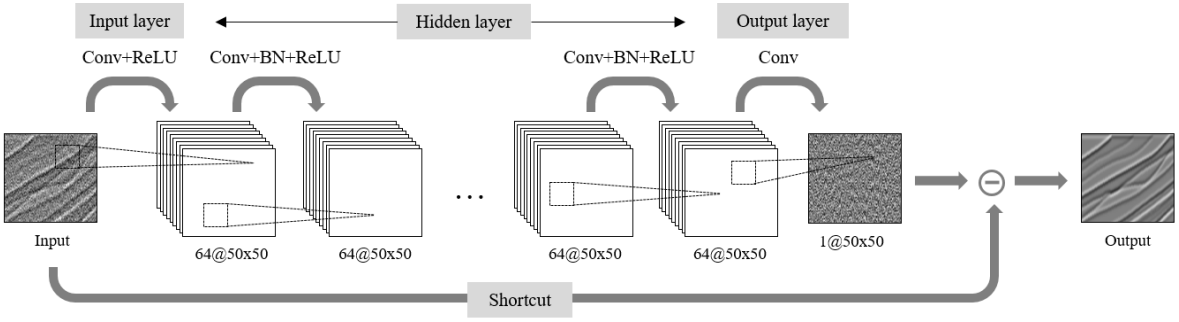


Figure 1: DnCNN architecture. 64@50x50 indicates 64 feature maps with 50x50 size, Conv is two dimensional 3x3 convolution kernel, BN is batch normalization, and ReLU is rectified linear unit activation function.

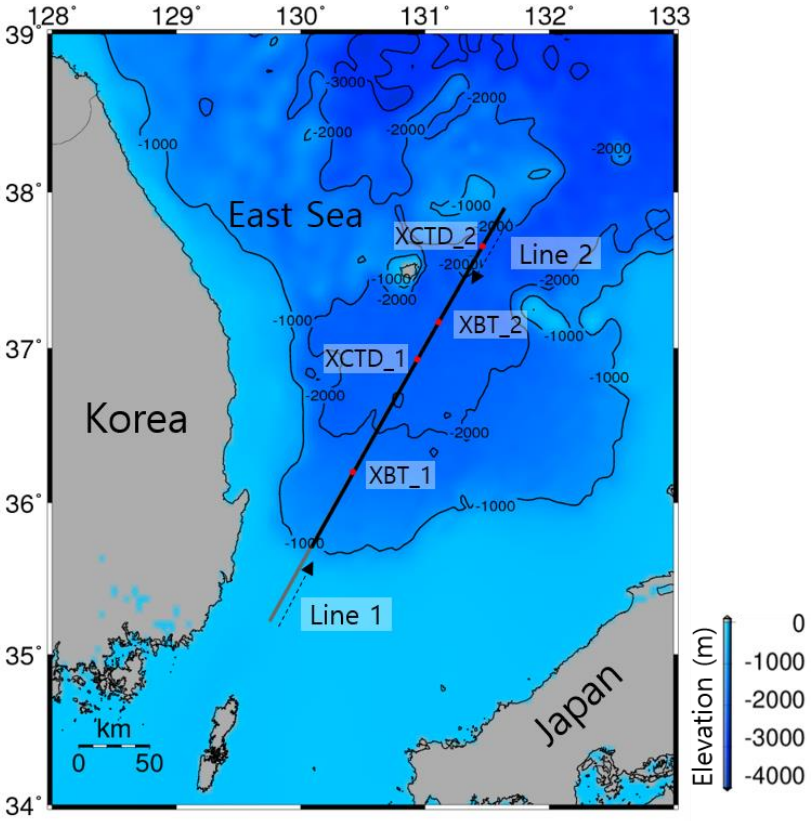
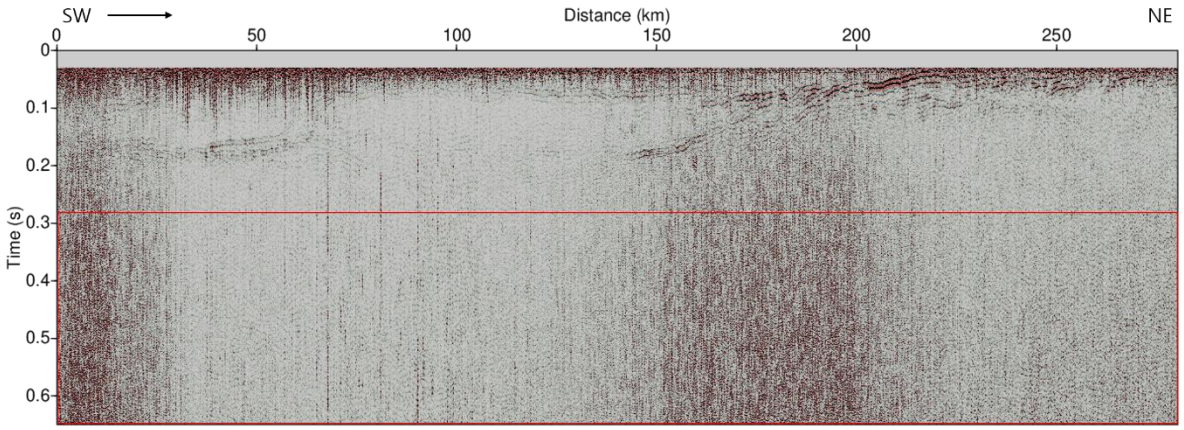
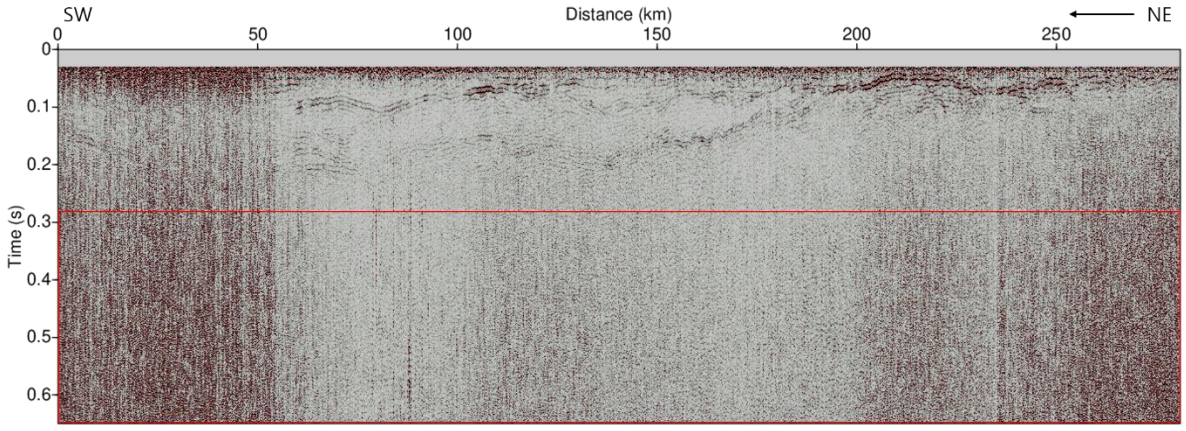


Figure 2: Location of seismic exploration. The solid line with gray and black ~~solid line is~~ color is the survey line. Gray line is shelf and slope parts which were removed from the seismic section during data processing and black line is the target area of this study. ~~, and~~ T~~t~~he black dashed lines with arrow indicate the exploration directions of lines 1 and 2, and red dots are the locations of XBTs and XCTDs.

635

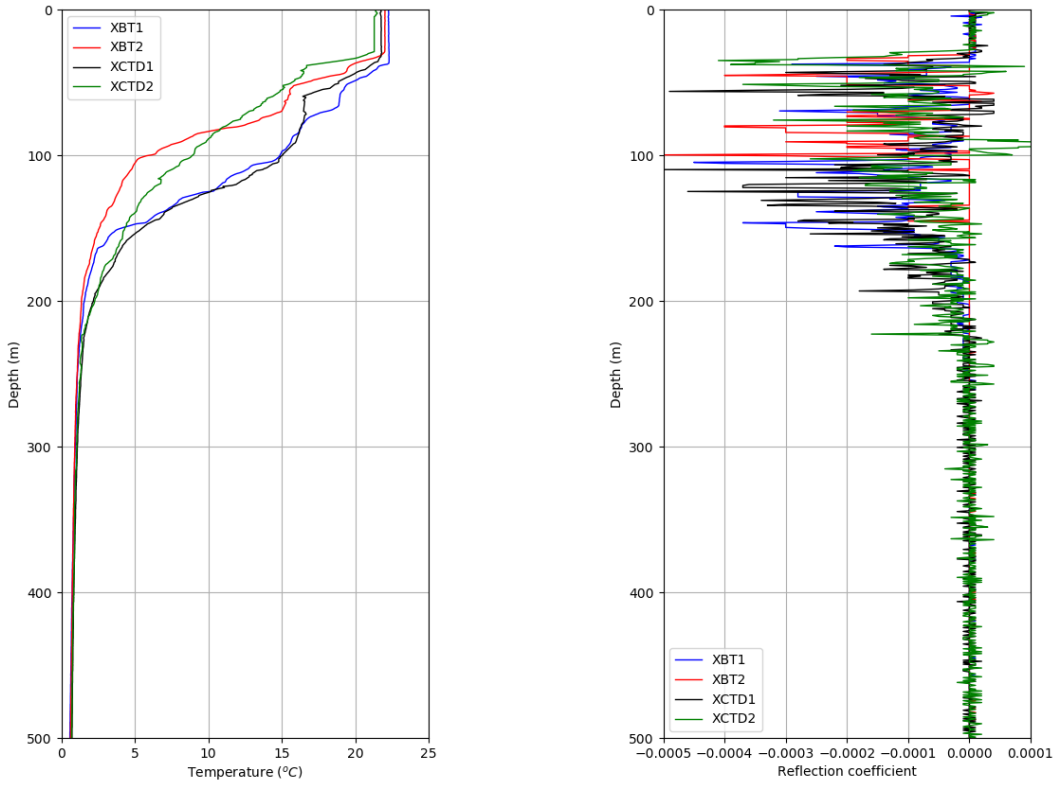


(a)



(b)

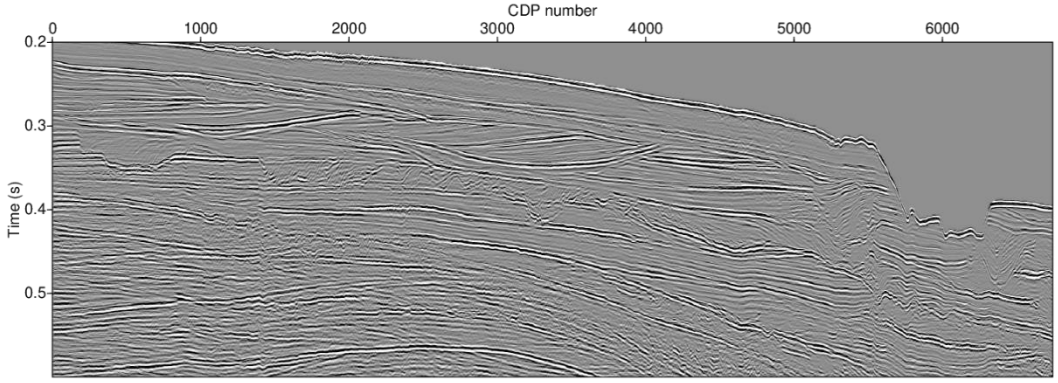
Figure 3: Processed seismic section of the East Sea: (a) line 1 and (b) line 2. The seismic section in the red rectangle is the noise part used to generate the training data. SW is south west, NE is north east, and black arrow indicates the data acquiring direction.



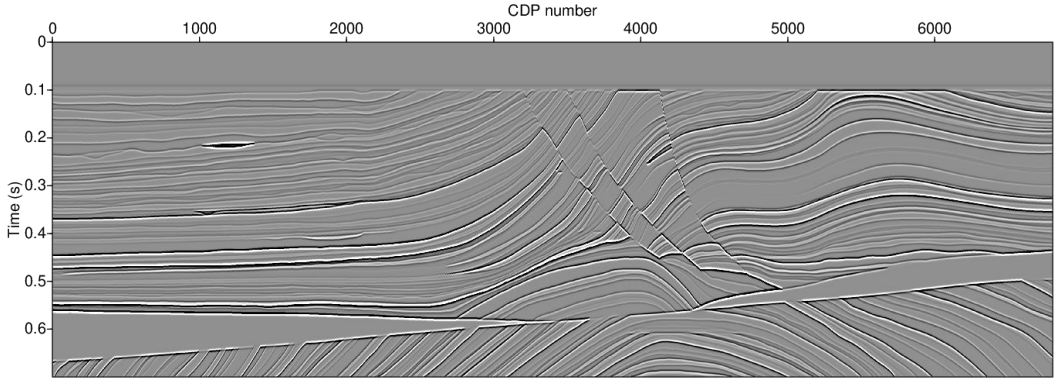
(a) (b)

Figure 4: (a) Temperature and (b) reflection coefficient profiles obtained using 2 XBTs and 2 XCTDs.

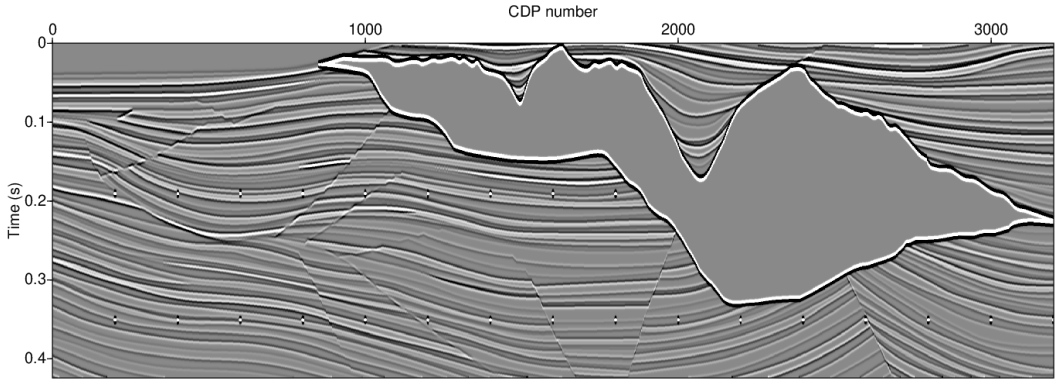
640



(a)



(b)



(c)

Figure 5: (a) Processed SEZ field seismic section, (b) Marmousi-2 synthetic seismic section and (c) Sigsbee 2A seismic section used to generate the training data.

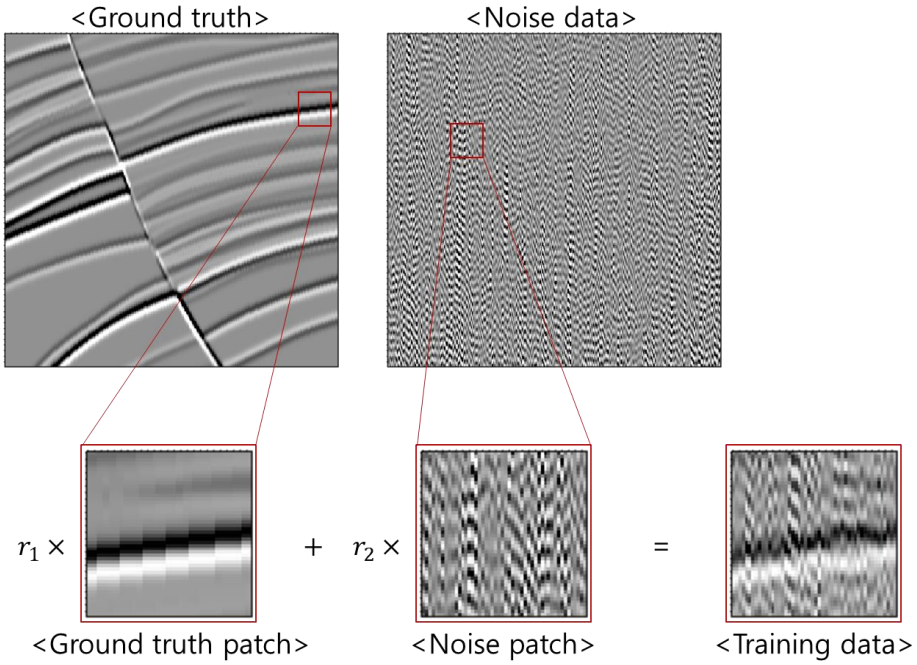


Figure 6: Example of constructing the training data.

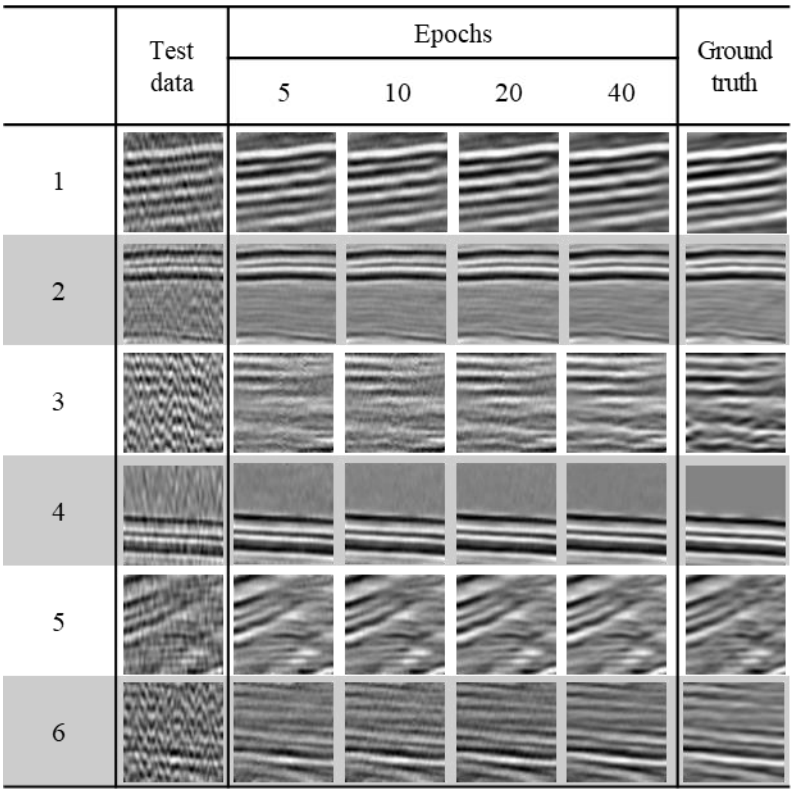
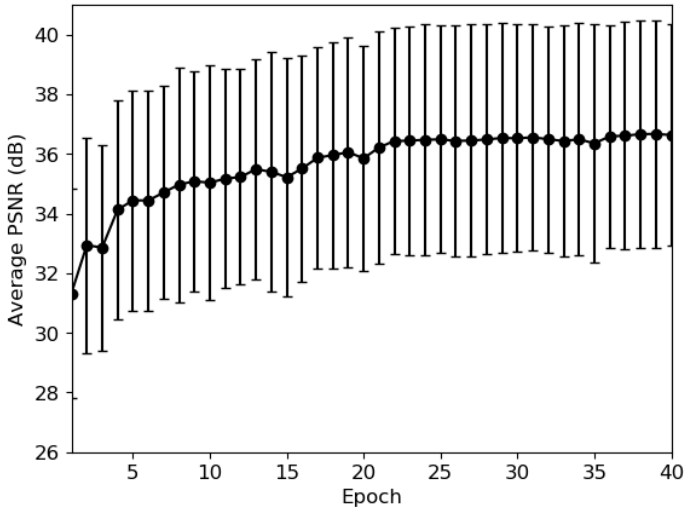
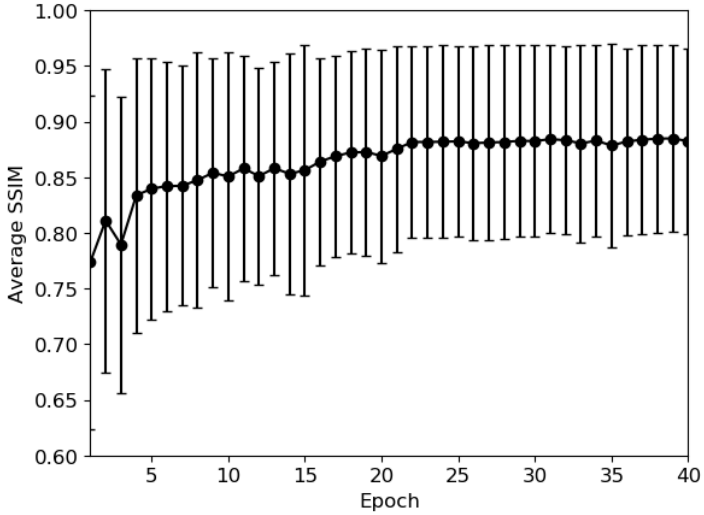


Figure 7: Test data, ground truth, and denoised results after applying the DnCNN models trained using training dataset 1.

645

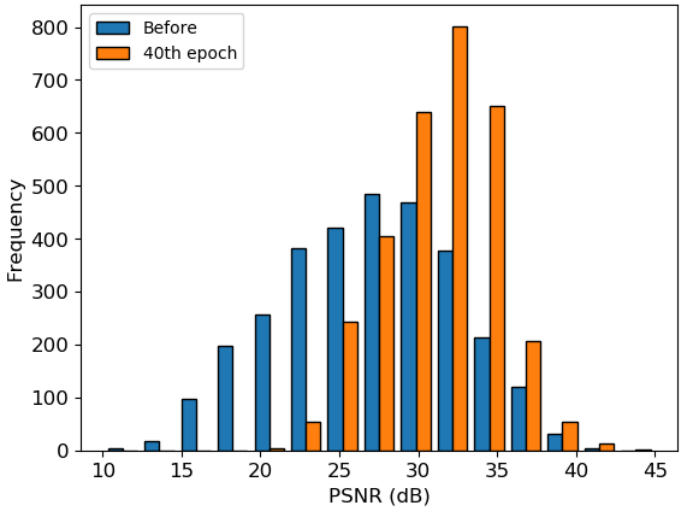


(a)

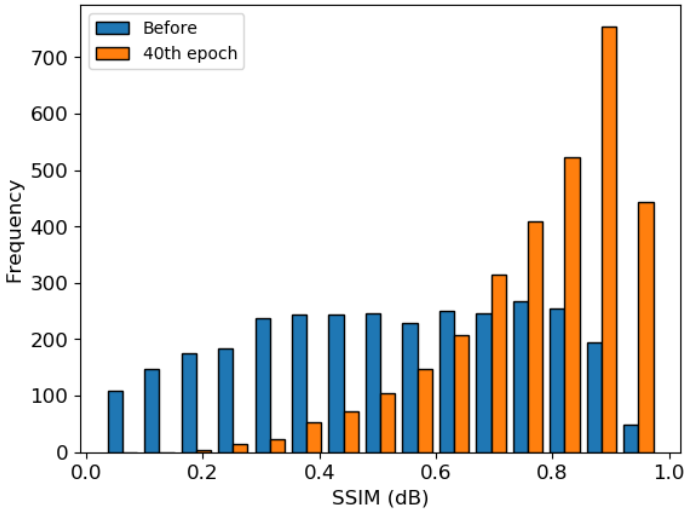


(b)

Figure 8: Average (a) PSNR and (b) SSIM with standard deviation of the test result of the first experiment.

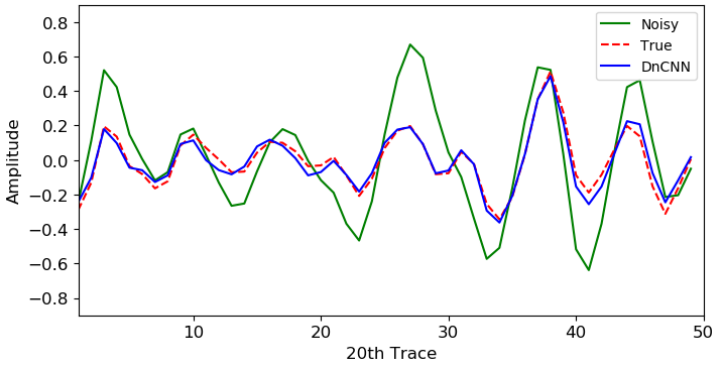


(a)

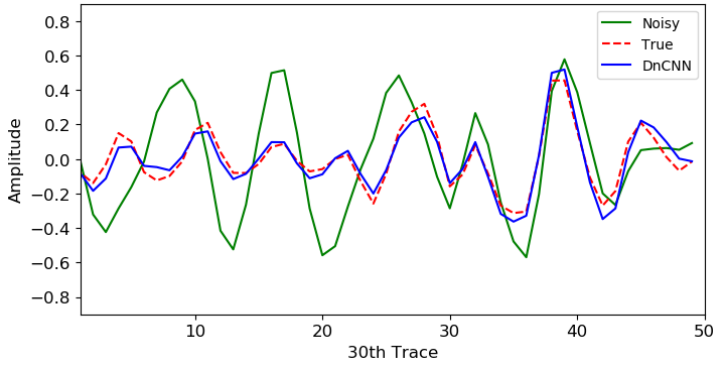


(b)

Figure 9: (a) PSNR and (b) SSIM histogram of the test data before and after applying the 40th epoch of the D1 model.



(a)



(b)

Figure 10: Comparison of the extracted traces before and after applying the 40th epoch of the D1 model. The green solid line is the trace from the noisy data, the red dashed line is the trace from the ground truth, and the blue solid line is the trace from the denoised data after applying the D1 model. (a) is the 20th and (b) is the 30th vertical trace of the last test patch in Figure 7.

650

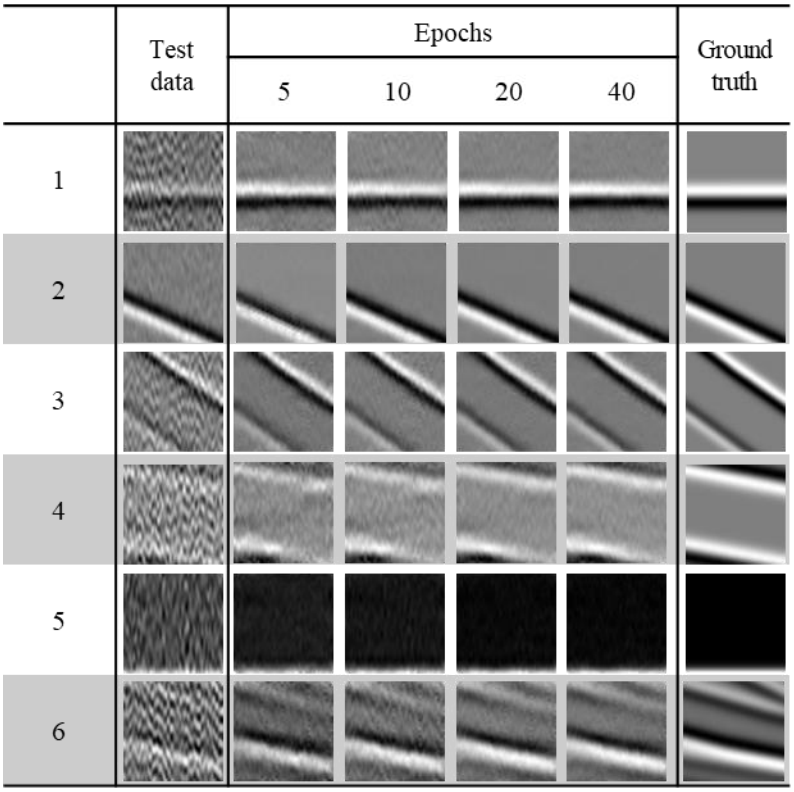
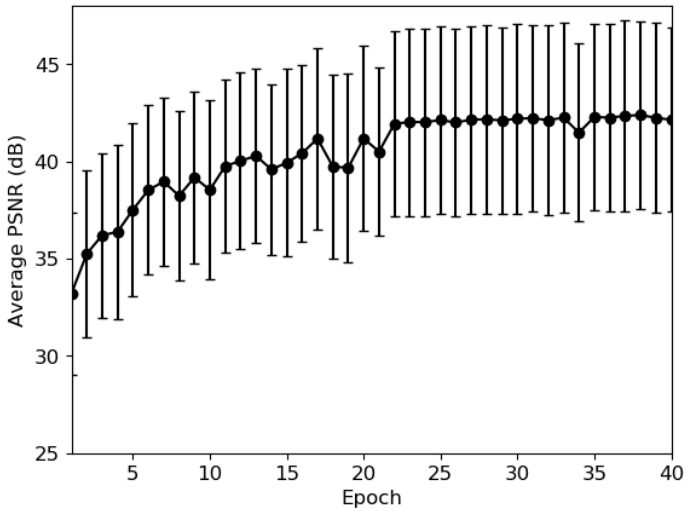
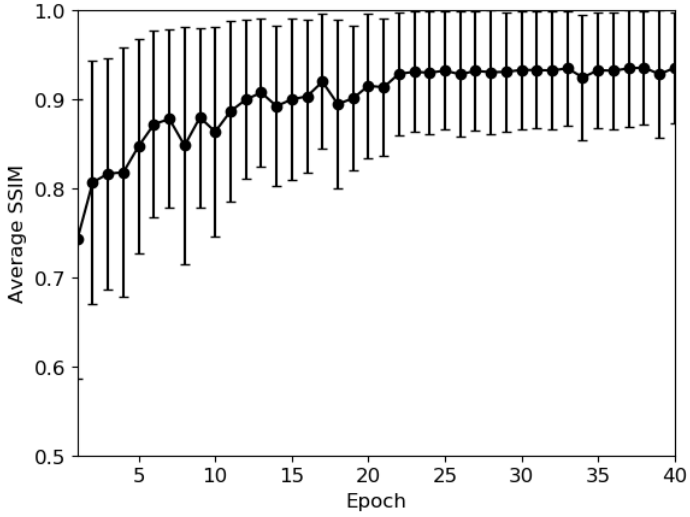


Figure 11: Test data, ground truth, and denoised results after applying the DnCNN models trained using training dataset 2.



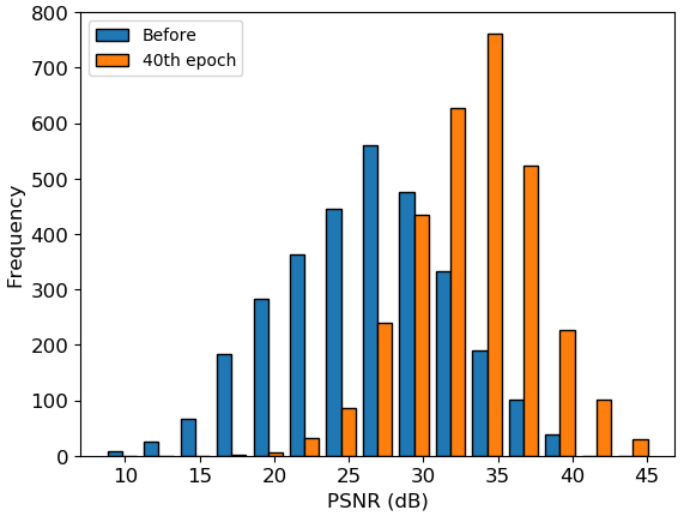
(a)



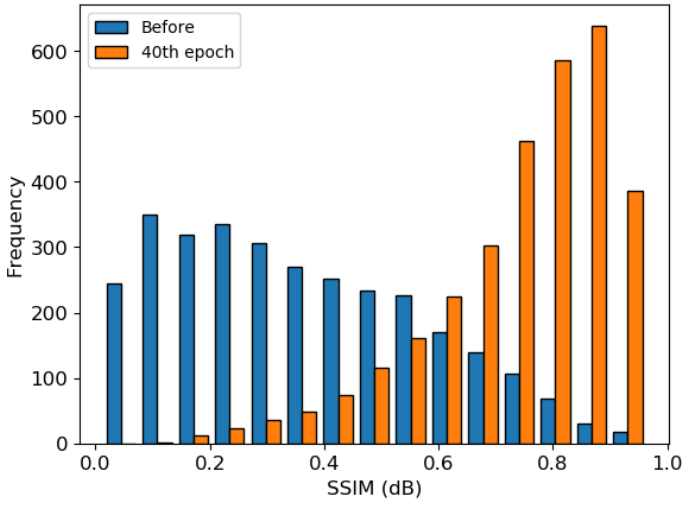
(b)

Figure 12: Average (a) PSNR and (b) SSIM with standard deviation of the test result of the second experiment.

655

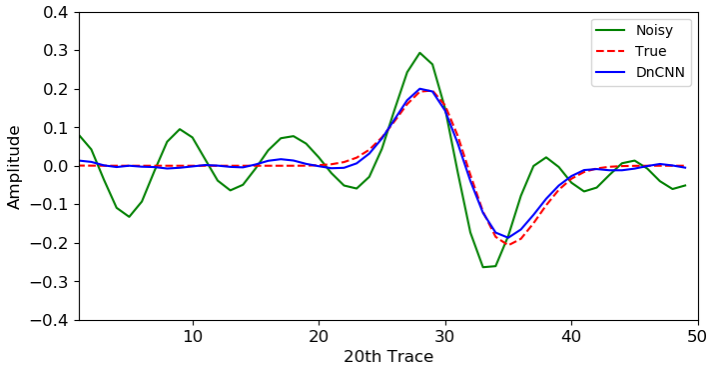


(a)

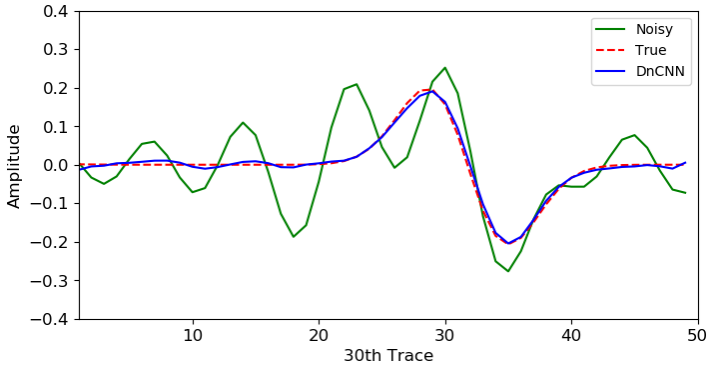


(b)

Figure 13: (a) PSNR and (b) SSIM histogram of the test data before and after applying the 40th epoch of the D2 model.



(a)



(b)

Figure 14: Comparison of the extracted traces before and after applying the 40th epoch of the D2 model. The green solid line is the trace from the noisy data, the red dashed line is the trace from the ground truth and the blue solid line is the trace from the denoised data after applying the D1 model. (a) is the 20th and (b) is the 30th vertical trace of the first test patch in Figure 11.

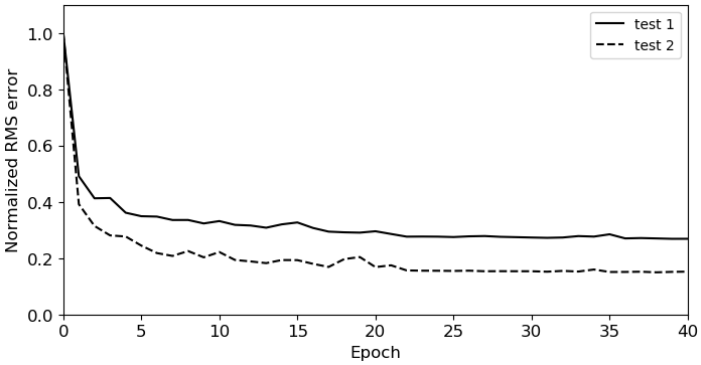


Figure 15: Normalized RMS error between the ground truth and denoised result of the first (solid) and second (dashed) experiments.

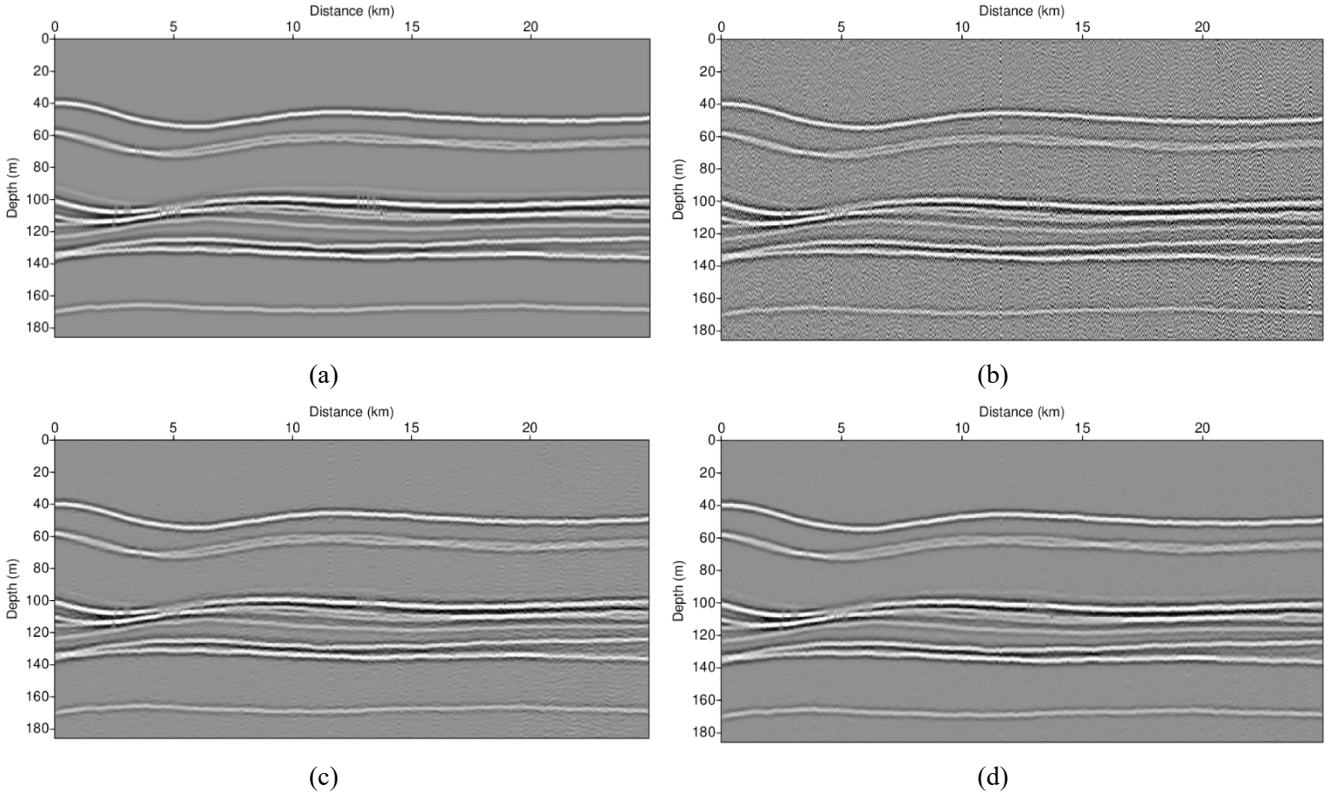


Figure 16: (a) Noise-free and (b) noise-added synthetic water column reflection section and noise-attenuated results using (c) the D1 model and (d) the D2 model.

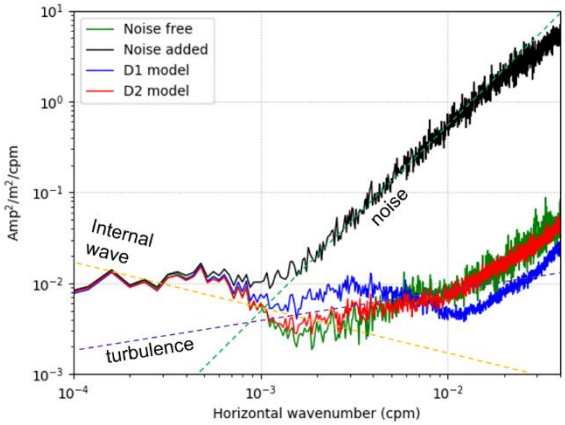


Figure 17: Data slope spectra of noise-free (green) and noise-added (black) synthetic seismic sections and noise-attenuated synthetic seismic section using the D1 model (blue) and D2 model (red).

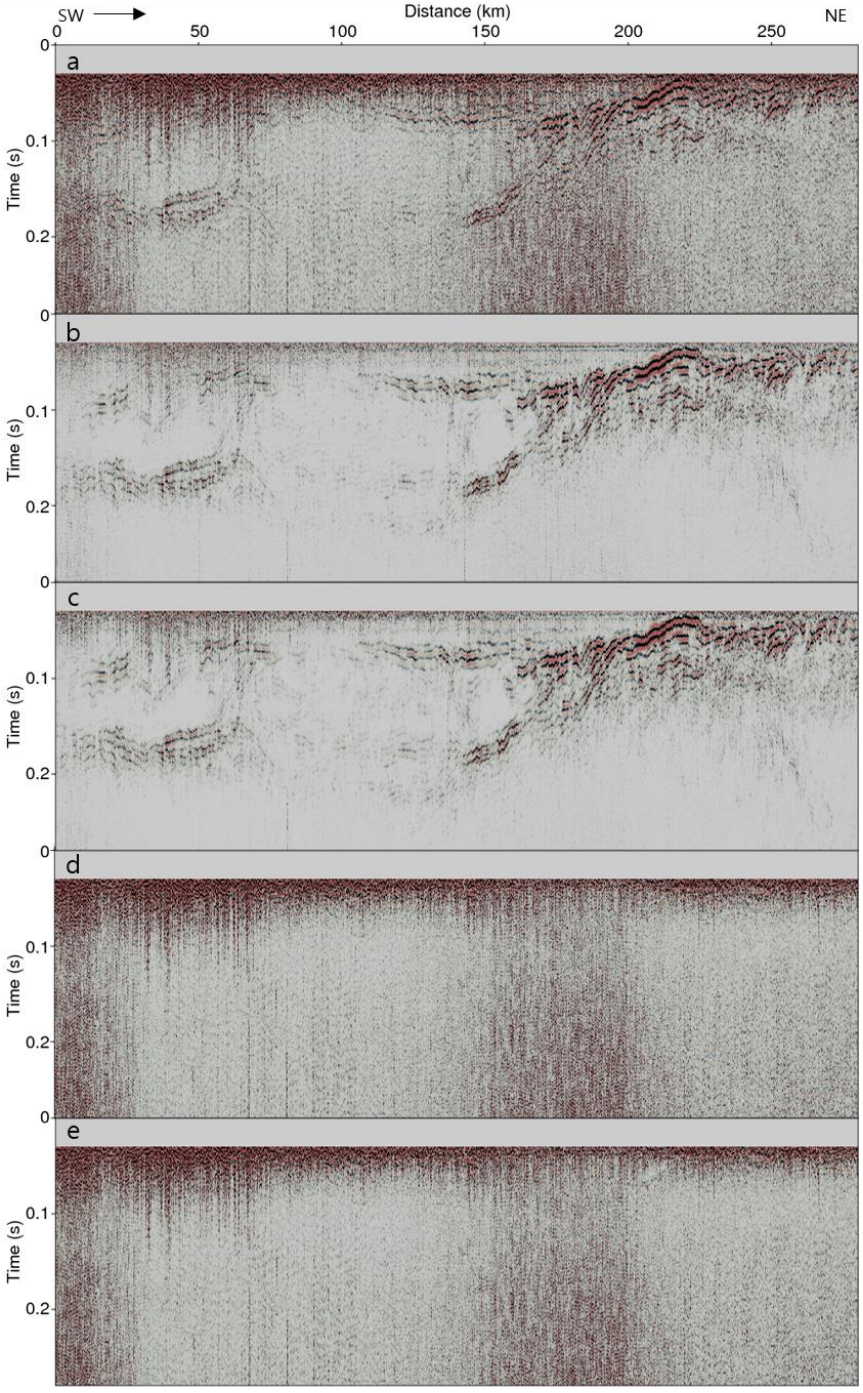


Figure 18: (a) Line 1 seismic section before applying DnCNN, noise-attenuated result using (b) the D1 model and (c) the D2 model, and estimated noise using (d) the D1 model and (e) the D2 model.

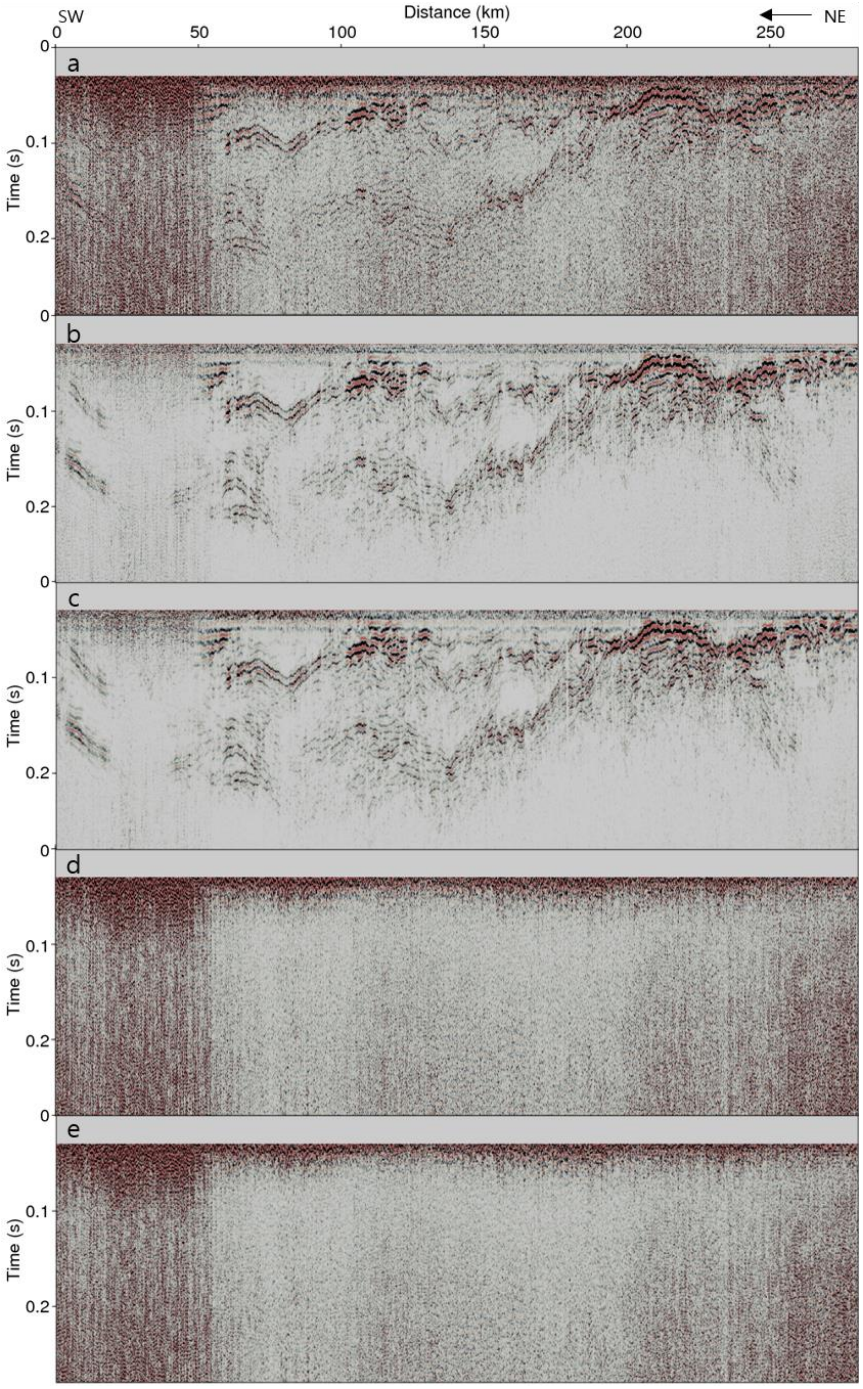
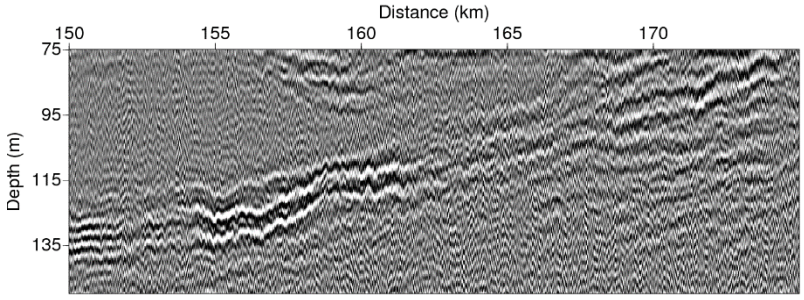
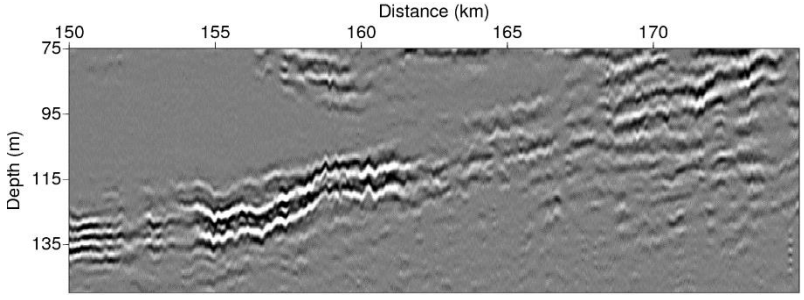


Figure 19: (a) Line 2 seismic section before applying DnCNN, noise-attenuated result using (b) the D1 model and (c) the D2 model, and estimated noise using (d) the D1 model and (e) the D2 model.

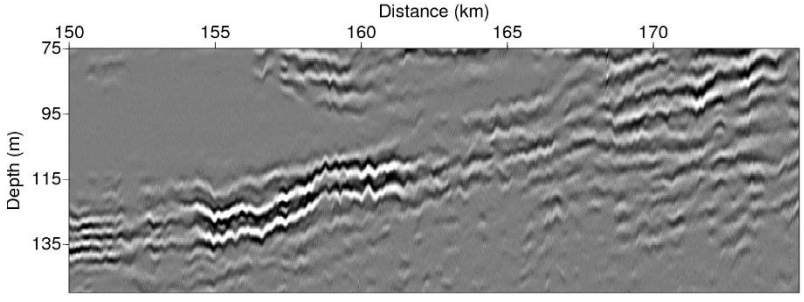
665


Distance (km)
150
155
160
165
170
Depth (m)
75
95
115
135

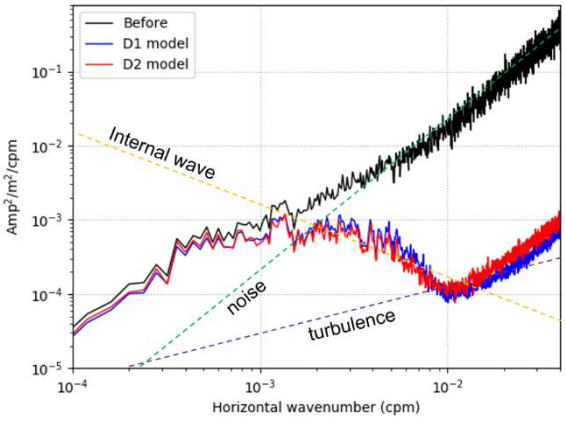
(a)


Distance (km)
150
155
160
165
170
Depth (m)
75
95
115
135

(b)


Distance (km)
150
155
160
165
170
Depth (m)
75
95
115
135

(c)


Before
D1 model
D2 model
Internal wave
noise
turbulence
Amp²/m²/cpm
Horizontal wavenumber (cpm)

(d)

Figure 20: Extracted seismic sections ((a) is the section before noise attenuation, (b) is the section after applying the D1 model and (c) is the section after applying the D2 model). (d) shows the calculated data slope spectra of (a), (b) and (c).