

Interactive comment on “Random Noise Attenuation of Sparker Seismic Oceanography Data with Machine Learning” by Hyunggu Jun et al.

Hyunggu Jun et al.

htjou@kiost.ac.kr

Received and published: 18 May 2020

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 Reviewer 1's comments.

Q1. Li 6-7. The opening sentence of the abstract is a bit confusing. SO exploits water column reflections to interpret the oceanic features (fronts, eddies, water mass boundaries) as well as ocean fine structure (internal waves etc.). Furthermore, “compensating for the drawbacks of conventional PO equipment” is a very strong (and erroneous) statement. Perhaps “supplements the conventional PO observations”. We should also

C1

be very careful when interpreting the seismic images to describe PO quantitatively.

A1. We modified the sentence to “... to supplement the conventional physical oceanographic observation methods”

Q2. Li 8. The low / high frequency band introduction is not very helpful unless you relate it to spatial resolution.

A2. We added the vertical resolution of each equipment as “Most SO studies obtain data using air guns, which have relatively low-frequency bands with vertical resolution around or larger than ten meters. For higher-frequency bands with vertical resolution ranging from several centimeters to several meters...”

Q3. Li 10. Reword “To solve the problem”? For example, “To extract reliable signal from the low S/N ...”

A3. We modified the sentence to “To attenuate the random noise and extract reliable signal from the low S/N ratio of sparker SO data, we applied machine learning.”

Q4. Li 23: “measurements [from cruises] are performed. . . . observation [stations].”

A4. We modified the sentence to “Conventional physical oceanography measurements from cruises are performed by dropping equipment at the observation stations.”

Q5. Li 27: mention how the sea water characteristics can be estimated (through the acoustic impedance contrasts and expand a bit more to inform the reader)

A5. We modified the sentence to “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. The difference of temperature and salinity between the Labrador Current and the North Atlantic Current generated the difference of acoustic impedance which reflected the seismic signals, and the reflected signals recorded at the receivers were processed to image the thermohaline find structure of the Atlantic Ocean.”

C2

Q6, Q7. L29-32: Fine, but please do not oversell. Perhaps mention “qualitative images” and then move to “quantitative information after careful analysis where temperature/salinity contrasts produce well-defined horizons of seismic reflections” or similar. Also SO is not “widely used” Li 32: reword “determine the behavior of turbulence and internal waves” to, for example, “quantify the internal wave spectral distribution and infer turbulence”

A6., A7. We modified the sentence to “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 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”

Q8. Li 33: clarify what central frequency is (since the source covers a range of frequencies)

A8. We modified the central frequency to “peak frequency” which is more widely used word.

Q9. Li 43: vertical resolution of 1.5 m is not much superior to the vertical resolution of “several meters” stated in line 34. Perhaps specify the latter as 5-10 m?

A9. We modified “several meters” to “around or larger than ten meters”

Q10. Li 67-68: If not using MLP and AE (and any other acronym), no need to introduce them. It is difficult to read the text.

A10. We removed the explanation of MLP and AE

Q11. Li 64-75: If there's a possibility to thin out various methods introduced (and refer to a few key references and citations therein), it can be easier for the reader to follow.

C3

A11. We thinned and removed several explanations.

Q12. Li 77: East Sea appears very abruptly here, out of context.

A12. We removed the sentence “Therefore, this study applies the DnCNN to attenuate random noise in East Sea sparker SO data”.

Q13. Li 135: delete “On the other hand,”?

A13. We removed “on the other hand”

Q14. Sec 2.1 and 2.2: can any of these descriptions refer to Fig 1? (I only see a reference in the end, at li 141, and it is not very instructive.)

A14. We relocated the sentence “Fig. 1 shows the DnCNN architecture used in this study, where Conv and BN indicate convolution and batch normalization, respectively.” at the early part of the paragraph and matched explanation of each block to Fig. 1.

Q15. Li 147: This is actually one line, but two repeats (in different travelling directions). Please mention the date of data collection, vessel speed during data collection. Transect duration etc.

A15. We added date of data collection with transect duration and vessel speed.

Q16. Li 163-164: there're CTD /XCTD profiles, but the authors shown only 2xtemperature profiles from XBTs. It would be nice to increase the oceanographic context in the paper.

A16. We added 2 XCTD data.

Q17. Li 167: please describe what a reflection coefficient is.

A17. We added “Reflection coefficient is a value that defines the ratio between the reflected and incident wave.”

Q18. Li 184-185: what do you mean by “thus, the subsurface seismic data have a better S/N ratio than the SO data.”? Is subsurface seismic data not SO data? I suspect

C4

you mean beneath seabed by subsurface. Please clarify.

A18. We modified the “subsurface” to “below the sea floor and beneath seabed” We also modified “sparker subsurface seismic data” to “SEZ seismic data”

Q19. Li 187: It is confusing: “We used the interval from 0.2 to 0.6 s of the original data where the noise level is relatively low”. Earlier you mentioned that part was just noise!

A19. There might be misunderstanding. The interval from 0.2 to 0.6 s of the SEZ data contains seismic data below the sea floor because the SEZ data is obtained shallow part of the East Sea where the water depth is approximately shallower than 200 m. However, the East Sea SO data which is the target data of this study is obtained from the deep part of the East Sea and the water depth is approximately deeper than 1000 m. Therefore, the East Sea SO data contains random noise below 0.28 s and SEZ data contains high S/N signal between 0.2 to 0.6 s.

Q20. Li 190: Reword “the data are field data recorded with the same equipment.” as “the data are collected by the same equipment”

A20. We modified the sentence.

Q21. Li 204: what is g/cc? Please use SI units.

A21. We modified g/cc to g/cm³

Q22. Li 249: bottom right (instead of right bottom)

A22. We modified right bottom to bottom right

Q23. Li 249-250: The sentence is confusing: “. . .using training dataset 1 has one problem. The ground truth of test data 5 contains noise in the right bottom part, and training dataset 1 also contains noise in some parts of the ground truth”. Dataset 1 has 6 test data. With the last reference to dataset 1 do you mean test data 1 or the entire dataset 1? Perhaps cut out the entire last part after the comma. Overall, I would appreciate a more distinct wording for test data. For example subset 1 to 6, or patch

C5

(you use it in line 280)?

A23. We modified “test data 5” to “5th patch of the test data. We also modified the “test data” to “patch”.

Q24. Li 279-280: 20th and 30th traces from the last patch: which epoch is this? Are the traces from the 50x50 patch? Can you please mention for the reader: “. . .traces out of the 50x50 size patch 6 of the test data”.

A24. We modified the sentence to “We extracted the 20th (Fig. 11 (a)) and 30th (Fig. 11 (b)) vertical traces from the last (6th) patch of the test data, which is 50x50 size, as shown in Fig. 9. For the denoised trace, we extracted trace from the denoised patch of the 40th epoch”.

Q25. Li 310-311: can be cut out; simply cross reference Fig 13 after 25 epochs. Overall there are repetitions throughout the authors could try to simplify.

A25. We simplified the paragraph by removing some repetitions.

Q26. Li 325: perhaps specify, “is the number of test data patch (3072)”

A26. We modified the sentence.

Q27. Li 327-328: too many significant digits at RMS errors? (perhaps enough with 6.37 and 6.34). For which epoch are these values? (Also the normalized values in line 331 could be 0.27 and 0.15)

A27. We modified 6.374 to 6.37 and 6.339 to 6.34. They are the RMS error of the test data before applying DnCNN. To clarify, we modified “RMS errors of test dataset 1 and 2 before noise attenuation. . .” to “initial RMS errors of test dataset 1 and 2 before noise attenuation. . .”

Q28. Li 332: delete “than that of the D1 model”

A28. We removed “than that of the D1 model”.

C6

Q29. Eq 6, is a division by nmode missing?

A29. We wanted to calculate the average RMS error of each test data patch (not each node in a patch). Therefore, we divided the errors by ntest only.

Q30. Li 364: "The data slope spectrum is the slope spectrum. . ." this is all very confusing. The data slope spectrum is first referred to in line 276-277 (again without explanation). Please introduce what the data slope is. For example, "the slope spectrum is the horizontal wavenumber, k_x , spectrum of the horizontal gradient of the vertical displacement of a digitized horizon. The data slope spectrum is . . .?" (or a similar explanation. Note my interpretation of the slope spectrum can be in error.)

A30. To explain the data slope spectrum and avoid confusion, we removed "data slope spectrum" in line 276-277 which is unnecessary. Instead, we modified the explanation of data slope spectrum in line 364 as "The data slope spectrum is a horizontal slope spectrum obtained directly from seismic data by calculating the horizontal wavenumber (k_x) spectrum of the seismic reflection amplitude and it is useful to identify noise contamination of seismic data and the cutoffs from an internal wave to turbulence sub-range clearly (Holbrook et al., 2013; Fontin et al., 2017)."

Q31. Li 367: replace "we calculated the data slope spectrum . . . and compared the data slope spectra" with "we calculated and compared the data slope spectra using the outcome of the D1 and D2 models. . ."

A31. We modified the sentence.

Q32. Li 376: "slope" is missing before "at wavenumbers"

A32. We added "slope".

Q33. Li 377-378: I cannot quite follow the subranges and the mentioned slopes in this panel. Perhaps mark on the figure?

A33. We marked guide lines of each subrange in figure.

C7

Q34. Li 390: Here again mention why sparker SO data may be desirable

A34. We added "Despite the low S/N problem, the sparker source has advantage of generating relatively high frequency band signal, which can provide information with higher vertical resolution."

Q35. Fig 1. Please offer some more explanation in the caption. If not possible, defer reader to the main text

A35. We added explanation of figure in caption.

Q36. Fig 2. Elevation is grayed out for >0m, so the colorbar can stop at 0. It would be useful to add a few isobaths. I would call Line 1 and Line 2, Repeat 1 and Repeat 2.

A36. We added isobaths in the figure.

Q37. Figs 4 and 5 can be combined into 1 figure. I suggest two panels, T profiles in one panel with different color. Reflection coeff in the second panel with different colors and one offset by 0.0001 unit. Does the coefficient have a unit?

A37. We merged and modified figures. The reflection coefficient does not have a unit.

Q38. Fig 6 can be removed. It is already shown in Fig 3 and with the statement time > 0.28 s. You can mark the region by a rectangle in Fig 3.

A38. We modified the Figure.

Q39. Fig 10 (and Fig 13)- this is the average PSNR and SSIM for the 6 subsets of dataset 1? Would it not be better to show all 6 lines, or the average with one standard deviation? Actually the number of test data is 3072 (6 is an arbitrary pick), why not show the mean and std over all 3072? And also show a histogram?

A39. It is the average PSNR and SSIM of dataset 1 (Figure 10) and dataset 2 (Figure 13) We calculated standard deviation over all 3072 test data.

Q40. Fig 11 caption: in the end the cross reference should be to Fig 9.

C8

A40. We modified the cross reference.

Q41. Fig 14 caption: in the end the cross reference should be to Fig 12. In the text the model is referred to as D2 (but here D1).

A41. We modified the cross reference.

Q42. Fig 16. This figure is not needed either. It is simply the upper 0.28s of Fig 3. However, I appreciate that it is zoomed in and compared to the cleaned sections. See suggestion below Fig 17 comment. Fig 17. Please consider removing xaxis labels from panels c to d, and placing panel labels in the upper left corner of panels (it's grayed out anyway), so that you can have a more condensed 4-panel, 1 page figure with minimum white space vertically between panels. I think a reorganized version of Figs 16- 18 will be much better for the reader. I suggest 2 figures, each with 5 panels (with identical x-axis limits and width, and minimum white space between them). New Fig 16: Line 1 results. New Fig 17 Line 2 results with corresponding 5 panels for each figure: a) data (as in Fig 16a) b) clean after D1 c) clean after D2 d) noise after D1 e) noise after D2.

A42. We re-ordered the figures and made new Figure 16, 17. The layout of the figures will be re-arrayed at the publication stage. We will request the Journal to place the Figure 16 and 17 at the same page because it would be better for reader to read. If it is difficult, we will modify the figure to make it possible.

* We are still modifying the manuscript and will try to improve the manuscript more.

* There may be some changes in the sentences which are not grammatically correct in the future because we will get the English editing service after finishing revise.

Interactive comment on Ocean Sci. Discuss., <https://doi.org/10.5194/os-2020-13>, 2020.