

Object: Referee report for Earth System Science Data

Paper: Reconstructing ocean subsurface salinity at high resolution using a machine learning approach

Authors: Tian Tian & al.

.....

In this work the authors describe a new product of global subsurface salinity at a high resolution (0.25°x0.25°) covering 41 vertical levels (in the range 1-2000m), named IAP0.25°.

This product is obtained using a Feed Forward Neural Network model designed by the authors, which is trained from several input variables taken from satellite sets and reanalysis to reconstruct the subsurface salinity product. Interpolated *in situ* observations of salinity profiles has been considered as ground truth to compare reconstruction findings. Finally IAP0.25° is evaluated by comparison with three independent ocean products of salinity.

The overall presentation of the new product, results and evaluation metrics is of high quality, hence in my opinion the manuscript should be considered positively for publication in ESSD. I would suggest a minor revision, mainly because I find that improvements in presentation and some additional comments have to be considered. I give also line by line suggestions to ease revision.

- The paper is quite long: I find it is possible to shorten some parts by avoiding repetitive sentences or being more concise and direct in some subsections. This is particularly true, for example, for the Introduction. I believe that shortening the manuscript would aid the overall readability, hence facilitating the choice of using the authors' dataset.

- The link of the IAP0.25° DOI should be given for the English version of the website page. Indeed you end up in the Chinese one and when you switch to English you are brought back to the home page, which is confusing. In the website the dataset is told to cover 0-2000m instead of 1-2000m (which is correctly reported in the manuscript and seen from the netcdf downloaded)

- The FFNN model description needs some additional details. In general the authors have fully described data and reconstruction and evaluation, but less attention has been paid in motivating the NN choice and structure. I believe that adding this kind of comments would aid in understanding the background ratio.

- Understanding which input mostly influences the salinity reconstruction and causes greater propagation error would be very interesting in this work, and in my opinion would enhance completeness of the presentation (this is stated as a future step in the Summary section, hence to be considered only as a suggestion).

- In several points some statements sound very vague or too general (see line by line comments hereafter). I suggest to be more precise. For example the authors could revise how they refer to machine learning in a vague way: it might be more interesting to focus on neural networks only, since this is the model choice for this study.

.....

1. Introduction

- (54) typo *below the ocean surface

- (62) what do you mean with sufficient data quality? This is vague

- (75-76) state better the limitations

- (77-78) in this study you inspect NN for reconstruction of salinity field only, not for other ocean subsurface fields. state better

- (86) I would not talk about "major deficiencies", which sounds too strong for the subsequent comments. Maybe *some differences, *some limitations or similar

- (94-96) a bit repetitive with (77-79). revise by shortening

- (97) the second objective introduced does not sound as an objective
- (103) currently your section 6 is of Data availability (this could become your last section 7, probably more appropriate)

2. Data & Methods

- (107-109) this should be said in the introduction, not here. I would not say "a machine learning model", which sounds very general; instead point out it is your model- this happens also in other points of the paper.
 - (110, 115, 121) sentences "data were from.." could be improved with expressions as: we use, we downloaded, data was extracted from...
 - (112) sounds like you have done the optimal interpolation
 - (116) take off "and extrapolating"
 - (119-120) take off last sentence, you say this at the end of the 2.1.1 ssec, repetitive
 - Table 1. Should be better organised, I suggest you should have: Data type, Variable, Dataset, Data Source, Horizontal resolution, Vertical coverage and resolution, Time period, Reference, DOI; hence correct information given (e.g. for SST you would have Variable:SST, Dataset: OISST, Data Source: NOAA, etc.). Pay attention to classifying Salinity observations as Input, since indeed you use it as ground truth to which you compare the model output.
 - Eventually insert DOIs in text for each product for completeness
 - (131) say something more of interpolation technique adopted
 - (139-140) state better, e.g. anomaly profiles are derived by subtracting monthly climatologies from salinity profiles.
 - (140-141) not clear to me, to state better
 - why do you say that SSA were averaged in 0.25 resolution? Did you mean interpolated?
 - (146) "certain tests" is too vague. Say if you tested your method on raw profiles and found no differences, or say something more about the tests (without needing to show figures or results about)
 - (150) remove "including", since you are stating all sets
 - (156-157) remove "which have spatial resolution of 0.25", repetitive
 - (169) You should start this subsec by describing the FFNN, not the data. These first sentences could be moved in data subsec or later on.
 - (186-187) complexity of a NN depends not only on how many neurons or layers are chosen, but also on type of layers and activation functions.
 - (190-192) make it shorter and less repetitive
 - (195) you should not refer to your model generically as "a FFNN" but "the FFNN"
 - (196) take off "based on training set" uninformative
 - (199) I would not talk about "reconstruction effect"
 - Figure 1. when showing the output information you give several maps of subsurface salinity anomalies, since your method is reconstructing each time step separately from the others I suggest you leave only one map also in the figure. Caption: I suggest you change in "Schematics of the FFNN architecture for subsurface salinity anomalies reconstruction (IAP0.25)"
 - which is your cost function minimised for training? did you use a GD algorithm? say more
 - (209) I would mention that the independent dataset are those introduced in ssec 2.2
- ### MANCA 2.3.3
- (243-244) is variance the standard deviation? usually $\text{variance} = \text{std}^2$
 - (256) I would add "estimate machine learning methods uncertainty"
 - (263) missing ending dot
 - Figure 2. (b&d) it seems to me that there might have been higher values than 0.2, flattened to be exactly 0.2. if this is the case please adapt colorbar with pointed end for out-of-range values. Eventually it would be interesting to have the four panels shown with same colorbar limits to be able to compare effectively two different depths' results (if graphically informative)

3. Reconstruction Results

- (275) typo: take of "as"
- (276) you should quantify the consistency between your reconstruction and IAP1 & ARMOR3D with some metric

- (280) take off "indicating greater resolution", uninformative. Next sentence take off "of change", the patterns you are showing are of one particular month.
- (282) How did you perform subtraction of IAP0.25 and IAP1? Did you first have IAP1 interpolated?
- I find it curious that you first show results of 5m and 100m depth globally, but then you turn to 100m & 300m for specific regions. Why is this? if there is no reason I would try to be consistent between global and detail analysis
- Figure 3 & 4. Refer in caption to each product by giving subfigures letters. You don't mention comparison between IAP025 and IAP1 in the caption (panel e that could be given close to the two products images)
- (298) You talk about reliability of the reconstruction by giving very qualitative comments, this should be quantified for robustness via some metric
- (327) typo: Figure 2b and *d
- (334) It is not clear to me when you state that ADT change should correspond better with subsurface salinity. I would suggest to state the whole ADT paragraph better
- (345) take off "certain", it sounds vague
- Figure 7a. Legend should be put all together possibly internal to the image. I suggest to put longitude coordinates on the top of the figure to aid readability of comments.
- (359) "seems that" does not sound as a statement. Substitute "in the global area" with "At a global scale"
- (363-365) very heavy to read, try to improve
- Figure 9. The best way for comparing graphically rmse over these regions is to have same figure limits for the upper panels (same for the lower panels).
- (395) take off parenthesis. why did you include lower quality data?
- (397) I would say even something about lowest reduction
- Figure 11. The y-axis limits should be the same for all subfigures for a more informative graphical comparison of results. In the caption I would mention that time series are averaged over the different areas
- (404) *reveals, instead of suggests, more precise
- (406) *with the other products considered

4. Five-fold cross validation and uncertainty estimate

- (412) I would recall these independent observations, referring also to subsec 2.2.
- (421) typo: * are obviously not biased
- (424-435) it is confusing how you comment the findings. First you talk about Figure 12e, then 12a-12d and then 12e again. Revise the whole paragraph, possibly shortening Figure 12. in the caption the sentence on correlation coefficient should be found when talking about panels a-d

5. Evaluation of the major climatic patterns

- (454) saying a "very significant linear trend" is too vague. Have you computed it with a p-value? what is the annual rate of change? state better
- Figure 13. the title should be Salinity Contrast index at Surface/0-2000m. I believe no y label is needed, and anyways should not be salinity since SC is shown
- (461) *for both the salinity at surface and averaged over the 0-2000m volume
- (475) To me the sentence regarding increased SC is not clear
- (479) you're talking about anomalies, why should average values impact the change? state better
- (486) typo: *systematic twice
- Figure 14 & 15. if possible all subfigures should have same y axis limits to ease comparison. Are these salinity anomalies?

6. Data availability

- (494) take off *mainly
- (498) *at a monthly resolution

7. Summary and Discussion

- (502) *were given as inputs to the FFNN algorithm for reconstruction
- (508) it is vague to say that IAP025 performs best compared with many available gridded products. state better
- (510) you should mention that for point (2) you are referring to subsurface salinity field. Indeed I don't see why separating this point from previous one. Instead I would say something of your findings of rmse or climatic patterns, which are not cited in this summary
- I suggest that this section should be revised trying to make it more appealing, and pointing more clearly to your findings.