We appreciate the community's comment from Dr. Etienne Pauthenet. Please find below a point-by-point response to comment. The comments are in black, our response is in blue, text changes are in orange.

Manuscript number: essd-2022-236
MS type: Data description paper
Title: Reconstructing ocean subsurface salinity at high resolution using a machine learning approach
Author(s): Tian Tian, Lijing Cheng, Gongjie Wang, John Abraham, Shihe Ren, Jiang Zhu, Junqiang Song, and Hongze Leng
Submission date for revisions: 4th Oct 2022

CC #1:

I would like to thank the authors for the very interesting work and the thorough validation of the results. However I believe there is overstatements in the presentation of the paper compared to what the authors actually did, that would require some clarification before publication. The paper is presented as a reconstruction of gridded salinity fields from observation, but it is in fact an upscaling of the IAP1 product from 1 degree to 0.25 degree resolution. I have three main comments.

Re: We'd like also thank Dr. Etienne Pauthenet for the insightful comments and introduction of their newly published paper (a regional reconstruction in Gulf Stream with machine learning approach). In response, we have clarified several issues and cite/discuss the suggested paper from Dr. Etienne Pauthenet in the revised manuscript. These comments help to improve our manuscript. However, we don't think there is "overstatement" in the presentation of our paper by doing gridded averages, as discussed below (it is a common practice to do the gridded average on observations before mapping/spatial interpolation).

1 - This paper uses data already gridded by Global Circulation Models (GCM) as inputs (IAP1 and SSW). The IAP1 data is also "interpolated into unified monthly and 0.25°× 0.25°spatial resolution fields" (L131-L132). Therefore the following sentences are overstatements because they are not mentioning the presence of GCM in the inputs already gridded at 0.25x0.25 resolution :

—> L94-L96 : "This paper explores the possibility of using a machine learning approach in which in situ observations and remote sensing data are merged to reconstruct the salinity changes at a 0.25°×0.25°horizontal resolution and a monthly temporal resolution from the surface down to a depth of 2000 m."

Re: Several sentences have been added to expand our introduction to IAP1° product, which is an already well-established dataset. "This product combines *in situ* salinity profiles with coupled model simulations (from phase 5 of the Coupled Model Intercomparison Project) to derive an objective analysis with the Ensemble Optimal Interpolation approach (Cheng et al., 2020)".

—> L500-L503 : "This study used an FFNN approach to reconstruct a high-resolution (0.25°×0.25°) ocean subsurface salinity dataset (1–2000 m) for the period 1993–2020, in which the spatial and temporal information (time, longitude, latitude, depth), and satellite remote sensing data (ADT, SST, SSW) were input into the reconstruction. By training the functional relationship between input variables and truth values (observed salinity), the reconstruction model was established."

To correct this would require a change of title, it is not a "reconstruction" but an "upscaling". And also correct the abstract, introduction, and conclusion, by stating first that the network is taking IAP1 as an input. I believe IAP1 is the main source of information the network is learning from. An analysis of the relative importance of each input for each output could answer this question (see figure 10 and 11 of Pauthenet et al 2022).

Re: First, the above-mentioned sentence has been revised to "This study used an FFNN approach to reconstruct a high-resolution (0.25°×0.25°) ocean subsurface salinity dataset (1–2000 m) for the period 1993–2018, in which the spatial and temporal information (time, longitude, latitude, depth), **previously available the 1°×1° resolution salinity dataset,** and satellite remote sensing data (ADT, SST, SSW) were given as inputs to the FFNN algorithm for reconstruction. By training the functional relationship between input variables and truth values (observed gridded averaged salinity), the reconstruction model was established". With one more sentence (in bold) added.

Second, we also include an analysis to discuss the relative importance of different inputs using Shap-method in the revised manuscript. We find that IAP1° is the most important input, which is physically meaningful because it sets the large-scale pattern of salinity, and the other inputs are indirect information and should play a complementary role (i.e. mainly to add small scale signals).

Finally, we decided to keep "reconstruction" instead of "upscaling", because "reconstruction" is a standard word for derivation of gridded field from observations. In previous temperature/salinity reconstruction efforts, *in situ* observations are firstly gridded into grid averages and then do the mapping (spatial interpolation), for example (Cheng et al. 2017; Cheng et al. 2020; Gaillard et al. 2016; Good et al. 2013; Ishii et al. 2003; Ishii et al. 2017; Levitus et al. 2000; Levitus et al. 2012; Lyman and Johnson 2014; Lyman et al. 2010; Roemmich and Gilson 2009), and also in a recent machine-learning effort (Bagnell and DeVries 2021). You can find more references in our recent review paper in Nature Reviews Earth & Environment (Cheng et al. 2022). Therefore, this is process is a common standard. Doing gridded averages aims to partly homogenize the data and reduce the sub-grid variability.

2 - The "truth" data is made of in situ profiles gridded by simple arithmetic averaging (L143). It is referred to as the "in situ observations" several time in the paper, which I find confusing. Here are three examples :

Re: Thanks, we have clarified this issue at several locations.

—> The statement L86 to L89 leads to believe that the present study does use in situ profiles directly as an input as opposed to previous studies : "First, some studies (Lu et al. 2019; Su et al. 2020; Wang et al. 2021) used Argo gridded data rather than in situ salinity observations data as the "truth" to train the machine learning model, and thus the reconstruction error in the Argo gridded data is embedded in the final reconstruction."

Re: This is a state of fact because it does not refer to our study, thus we prefer no change made. Because we have clarified this in the following texts, we believe it is sufficient.

—> L94 : "in situ observations" refer to the gridded 0.25x0.25 field here.

Re: Revised to "in situ profile observations (processed to a gridded 0.25° × 0.25° arithmetic mean field)"

—> L503 : "truth values (observed salinity)"

Re: "truth values (observed salinity)" revised to "truth values (observed gridded averaged salinity)"

Could you please differentiate clearly between the in situ profiles and the "truth" gridded fields of salinity? A gridded field is not "observation".

Re: We disagree that "gridded average" is not "observation", it is "observation". Actually, many of the gridded datasets after reconstruction have been commonly regarded as "observation", for example IPCC reports (please check chapter-2 of IPCC-AR6, time series for climate monitoring in WMO State of Climate Report, and BAMS State of Climate report) (Gulev et al. 2021; Johnson et al. 2021).

In addition, we want to clarify an important concept: the term "observation" is not limited to "*in situ* observation", the former has broader context: gridded averages and data products after interpolation are all regarded as "observations" although they are not "in situ observation".

3 - Figure 9 shows the RMSE between IAP0.25, IAP1, Armor3D and EN4 against "observation data" (L355). But is this observation data the "0.25°×0.25°, 1-month, and 41-level grid averages" or the in situ profiles? if it is the grid average, this presentation is biased towards IAP0.25 as it is the only product that was trained with the grid average, so the other products are expected to have higher RMSE.

Re: We agree that the comparison is not perfect, because the *in situ* data and gridded averaged fields are all processed by the authors' group, the data used for this validation is not independent. But we don't think the presentation is "biased" because all these products are using essentially the same *in situ* profile database (for example, >99% data in EN4 archive is from WOD, Good et al. 2013). We did use more evaluation approaches to compensate this shortcoming.

And we oppose to compare the final gridded field with *in situ* profiles, because the results will be biased to the regions with more data, especially where there are high-frequency observations (sometimes, there are even >1000 of profiles in one grid box by several moorings). Besides, as the RMSE (between gridded field versus in-situ-profiles) = Reconstruction error + Observational error + Sub-grid variability (<0.25°× 0.25° grid), the results will have loose physical meaning.

If it is the in situ profiles compared to colocated profiles from EN4, Armor3D, IAP0.25 and IAP1, then it would be interesting to see the RMSE between the in situ profiles and the grid average used a "truth" data. Could you add this RMSE profile to the figure?

Re: The RMSE between *in situ* profiles and the gridded averages represents the the sub-grid (<0.25°× 0.25° grid) variability, reconstruction error and the observational error. If the purpose is to reconstruct a complete 0.25°× 0.25° field, the sub-grid variability is also "noise" or "error" in the reconstruction (we don't want these variability), thus this RMSE represents the level of noise in different grids. In our reconstruction, we did properly account for these errors (sub-grid variability and observational errors) in reconstruction (see Section 2, uncertainty estimate).

One important concept in data reconstruction field is: "truth" depends on purpose of the reconstruction: it the purpose is to reconstruct a 0.25°× 0.25° resolution dataset, sub-grid (<0.25°) variability is error/noise, thus in situ profiles can't be regarded as "truth" as they have strong sub-grid variability.

Nevertheless, although our gridded average is not perfect because it can't completely remove sub-grid variability, but we do find our gridded averages works better than using *in situ* observations in our global reconstruction.

The IAP1 data is also "interpolated into unified monthly and 0.25°× 0.25°spatial resolution fields" (L131-L132). Could you plot the RMSE between this regridding of IAP1 and in situ profiles on figure 9 too? That would be a good justification for using the neural network for this upscaling.

Re: This is a good point and we already had the paper (Fig.2a, c, but for standard deviation). This plot shows the presence of the sub-grid (<1°× 1° grid) variability plus the observational errors.

Best regards,

Etienne Pauthenet

Reference :

Pauthenet, E., Bachelot, L., Balem, K., Maze, G., Tréguier, A. M., Roquet, F., ... & Tandeo, P. (2022). Four-dimensional temperature, salinity and mixed-layer depth in the Gulf Stream,

reconstructed from remote-sensing and in situ observations with neural networks. Ocean Science, 18(4), 1221-1244. **Citation**: https://doi.org/10.5194/essd-2022-236-CC1

Re: Thanks for introducing this very interesting paper. We have added this reference in the revised manuscript.

References:

Bagnell, A., and T. DeVries, 2021: 20(th) century cooling of the deep ocean contributed to delayed acceleration of Earth's energy imbalance. *Nat. Comm.*, **12,** 4604.

Cheng, L., K. E. Trenberth, J. Fasullo, T. Boyer, J. Abraham, and J. Zhu, 2017: Improved estimates of ocean heat content from 1960 to 2015. *Sci. Adv.*, **3,** e1601545.

Cheng, L., and Coauthors, 2020: Improved Estimates of Changes in Upper Ocean Salinity and the Hydrological Cycle. *J. Clim.*, **33,** 10357-10381.

Cheng, L., and Coauthors, 2022: Past and future ocean warming. *Nature Reviews Earth & Environment*.

Gaillard, F., T. Reynaud, V. Thierry, N. Kolodziejczyk, and K. von Schuckmann, 2016: In Situ–Based Reanalysis of the Global Ocean Temperature and Salinity with ISAS: Variability of the Heat Content and Steric Height. *J. Clim.*, **29,** 1305-1323.

Good, S. A., M. J. Martin, and N. A. Rayner, 2013: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *J. Geophys. Res. Oceans*, **118,** 6704-6716.

Gulev, S., and Coauthors, 2021: Changing state of the climate system. In *climate change 2021: The physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change* (eds Masson-Delmotte, V. et al.).

Ishii, M., M. Kimoto, and M. Kachi, 2003: Historical Ocean Subsurface Temperature Analysis with Error Estimates. *Mon. Weather Rev.*, **131,** 51-73.

Ishii, M., Y. Fukuda, S. Hirahara, S. Yasui, T. Suzuki, and K. Sato, 2017: Accuracy of Global Upper Ocean Heat Content Estimation Expected from Present Observational Data Sets. *Sola*, **13,** 163-167.

Johnson, G. C., and Coauthors, 2021: Ocean heat content [in State of the Climate in 2020]. *Bull. Am. Meteorol. Soc.*, **102,** S156-S159.

Levitus, S., J. I. Antonov, T. P. Boyer, and C. Stephens, 2000: Warming of the World Ocean. *Science*, **287,** 2225-2229.

Levitus, S., and Coauthors, 2012: World ocean heat content and thermosteric sea level change (0-2000 m), 1955-2010. *Geophys. Res. Lett.*, **39,** L10603.

Lyman, J. M., and G. C. Johnson, 2014: Estimating Global Ocean Heat Content Changes in the Upper 1800 m since 1950 and the Influence of Climatology Choice. *J. Clim.*, **27,** 1945-1957.

Lyman, J. M., and Coauthors, 2010: Robust warming of the global upper ocean. *Nature*, **465,** 334-337.

Roemmich, D., and J. Gilson, 2009: The 2004–2008 mean and annual cycle of temperature, salinity, and steric height in the global ocean from the Argo Program. *Prog. Oceanogr.*, **82,** 81-100.