EGUsphere, referee comment RC2
https://doi.org/10.5194/egusphere-2022-25-RC2, 2022

Comment on egusphere-2022-25

Michel CREPON (Referee)

Referee comment on "Four-dimensional temperature, salinity and mixed-layer depth in the Gulf Stream, reconstructed from remote-sensing and in situ observations with neural networks" by Etienne Pauthenet et al., EGUsphere, https://doi.org/10.5194/egusphere-2022-25-RC2, 2022

Review of the paper submitted to egusphere 2022

Four dimensional temperature, salinity and mixed layer depth in the Gulf Stream, reconstructed from remote sensing and in-situ observations with neural networks

By E. Pauthenet et al

The paper aims at providing four-dimensional temperature, salinity, and mixed layer depth in the Gulf Stream, from sea surface satellite observations (SST and altimetry). Interpolations of surface data at depth are done with a NN trained on 67767 vertical profiles. In the operational phase, satellite data are associated with vertical profiles (Temperature, Salinity, Density and MLD) through the NN. The authors also present a procedure based on density stability to improve the MLD estimation. The subject is of scientific interest due to the lack of vertical profiles in the ocean with respect to satellite surface data. The procedure presented (OSnet) seems efficient to associate sea surface satellite data with their vertical profiles. But I found the paper difficult to read and poorly structured. It can be published after the following corrections and the rewriting of some sections.

Major comments

The paper is quite long and can shorten by 30%. I suspect it presents the results of Ph.d. work of an enthusiastic student who would like to present all the details of his work and has some difficulties extracting the major conclusions.

The readers of Ocean Sciences are physicists and most of them are not familiar with neural networks. Section 3.1 must be rewritten with care.

I recommend specifying that the use of a NN can be decomposed into two phases well separated:

- a learning phase in which the weights of the neurons are estimated from a learning data base.
- an operational phase consisting in retrieving the profiles from the satellite data (input data base)

The learning data must be described with care: mention the origin of the profiles, which is unclear in the present form. The input data must be justified. It appears that there is some redundancy among them: are MDTs and SLAs independent data? I do not think that geostrophic currents content added information with respect to SLA. How do you compute geostrophic current anomalies? Are they seasonal anomalies or anomalies with respect to whole observation period? Information included in SLA are also included in the geostrophic currents. These remarks are comforted by section 4.4 which shows that some variables do not play an important role and can be neglected. Section 4.4 could be suppressed if the input variables are chosen adequately in section 3 by a simple physical reasoning or by doing an EOF on the input data.

Can you comment?

The procedure for improving the MLD developed in section 3.3 is an important feature of this work, but it is hard to understand. Can you reformulate it in a simpler manner? How do you estimate the parameter lambda in the K estimation? A simpler procedure would be to apply a median filter onto the density profiles for removing the hydrostatic instability.

Can you discuss this?

The significance of the sentence printed in lines 199-200 is difficult to understand.

I have appreciated the scientific content of appendix A which aims at removing the density inversion with a physical constrained loss function, which is an original contribution of OSnet

The OSnet procedure has the characteristic of a multi-entry data base. It interpolates the profiles but does not model the physical laws connecting satellite observations and the associated vertical profiles. An original procedure using hidden Markov chain, which models these physical laws has been recently developed for retrieving vertical Chl-a vertical profiles from ocean color satellite observations (Charantonis et al, 2015, Puissant et al, 2021). Can you say some words about the philosophy of these two methods, their advantages, and disadvantages?

Minor comments

Most of the figures are very small. It is difficult to extract information from them. As an

example, in Figure 4, it is difficult to identify the different profiles from each other. Besides, the significance of the two horizontal dotted lines must be mentioned in the figure legend

In figure 2, what are the units for the T rmse, S rmse, sigma0 rmse?

In figure 5, why the density distributions of SST and SSS are so different.

Section 4.2 : horizontal maps of T and S (Figures 6, 7) are not very useful since the authors focus their interest on the four dimensional representation of these two variables. Besides the figures are very small. I suggest replacing them by vertical sections.

Section 4.5 is interesting. OSnet is able to reproduce the SST due to global change. It could be used to process ocean data in climate study contexts. But I do not understand the sentence (lines 413-414) "The long term…. Based on loess" What do you mean by loess?

Section 5.1 justifies the use of OSnet for providing T S profiles at any location. Do be too modest! I would change line 353 as "One **major** feature of OSnet is the possibility…". The detection by OSnet of the big warm eddy crossing the mooring is impressive. Some problem, the position of the mooring W3 presented in the map figuring in little cartoon at the left top of figure 14 does not correspond to the coordinates mentioned in the figure legend!

English must be corrected by a native English-speaking person:

There are many English mistakes

Examples: line 91 "is shown **on** figure 1a"; line 2007 "is shown **on** figure 4"; line 299 "**The** figures 10 and 11……", instead of "**F**igures 10 and 11……..",

line 23 "the ocean's surface is observed ……" Instead of "the ocean surface has been observed……"

Too many uses of the possessive case: line 23 "the ocean surface has been observed ….. ". In modern English, possessive case is mainly dedicated to persons.

Data is a plural noun (singular: datum)

Conclusion

This paper is useful contribution to ocean data sampling. It can be published after the above corrections are done. I also suggest 30% concatenation of the text which is too long with unnecessary presentations.