Comment on essd-2022-236
Anonymous Referee #1

Referee comment on "Reconstructing ocean subsurface salinity at high resolution using a machine learning approach" by Tian Tian et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2022-236-RC1, 2022

It is a well-written paper on reconstructing subsurface salinity profiles using satellite data and some reanalysis data with the machine learning approach. A very important new contribution to the problems is the development of a better product to examine the vertical structure of the salinity field than ARM and EN4. I have the SST background, so I only have a few minor comments (which I leave to the authors to decide on how to address).

1. Define the abbreviation the first time it is used in the text; also, there is no need to define it twice. The author should check it thoroughly. For example, The SST was defined in line 108 but still used in lines 66,84.

2. The authors put the DOI link in the abstract. Whether the link can point to another English version which is http://english.casodc.com/data/metadata-special-detail?id=1546377368443076609
Then more people can use this data.

3. This study used the FFNN approach to reconstruct the salinity dataset. Why the authors use this specific method? Have they did the comparison between FFNN-VAR with other models, such as XGBoost, GANs, random fores, for reconstructing the ocean subsurface salinity dataset?

4. Figure 11 shows the seasonal fluctuation of RMSE of IAP 0.25 AND IAP 1. It's better to add the seasonal fluctuations of other insitu and reanalysis products to compare their performance.
5. The authors should analyze the errors with respect to the input (SLA, SSTA, SSSA, UWSA, VWSA): Which input contributes the most errors? Which input could be the primary surface independent parameter for estimating salinity?