Review of esd-2024-1, "Estimating ocean heat content from the ocean thermal expansion parameters using satellite data", submitted by Kondeti and Palanisamy for possible publication in Earth System Dynamics.

This manuscript presents a methodology for making maps of ocean heat content (and thermosteric sea level) using an artificial neural network that incorporates Argo CTD data together with what the authors identify as satellite SST and SSS data, as well as the results of the application of that method. There are some very major issues that will need to be resolved prior to publication. Specific minor comments follow, first major, then minor, indexed by line number (L) where possible.

Thank you for your constructive comments that have improved the manuscript. Please find the authors' responses to each comment.

1. Since ocean heat content and thermometric anomalies are generally surface-intensified, so integrating from the surface to each depth of mapping does little to reveal how well the method works for mapping ocean heat content or thermometric anomalies at depths below the thermocline. In that respect it would be much more useful, and transparent, to map these quantities in distinct pressure or depth layers. Typical layers for using Argo might be 0-100 m, 100-300 m, 300-450 m, 450-700 m, 700-1000 m and 1000-2000 m, but others might be chosen if desiring more vertical resolution. The statistics for these would be much more revealing of the skill of the method below the surface layers than as it is currently formulated.

The primary objective of this work is to develop a model to estimate OHC accumulated between the sea surface to 700 m depth, which is the standard depth considered in reporting OHC (Domingues et al., 2008; Levitus et al., 2009, 2012; Balmaseda et al., 2013; Abraham et al., 2013; IPCC, 2014; Palmer et al., 2017; Cheng et al., 2017; Meyssignac et al., 2019; Irrgang et al., 2019). This depth covers most oceanic basins except shallower portions of bathymetry levels less than 700 m. As the OHC changes are more dynamic in shallower/upper oceanic layers, the current work has been extended to shallower depths lesser than 700 m and reported the models' performance at each depth extent in section 4.

On the other hand, OHC can be modeled based on the depth layers such as 0-20 m, 20 – 30 m, 30 – 40 m, …., and 650 – 700 m. It involves the layer-wise preparation of in-situ OHC databases, theoretical formulation derivation, collocation of satellite data of all the input parameters (based on the derived theoretical formulations), and development and validation of ANN models. Hence, we considered this layer-wise OHC modeling as a future scope of the current study.

2. The maps of ocean heat content (or thermometric sea level) are validated only against another machine learning estimate, but there are a number of objectively mapped and machine learning mapped products out there which are more widely used against which this work should be compared. See Johnson et al. (2023, https://doi.org/10.1175/BAMS-D-23-0076.2) and/or von Schuckman et al. (2023, https://doi.org/10.5194/essd-15-1675-2023) for an idea of the range of products available.

The main objective of the current study is to develop ANN models to provide daily OHC estimates from satellite data. Hence, direct measurements from the Argo systems are employed to compute in-situ OHC estimates rather than the interpolated/synthetic gridded data of monthly/quarterly/yearly scales to accurately capture physical oceanographic conditions during ANN training. To the best of our knowledge, OHC daily products are not available except the NRSC OHC products, and thus we compared our model results with the NNRSC OHC products (as presented in section 4.3). The mapped products derived from the ocean reanalysis systems and objectively analyzed techniques are of monthly to yearly temporal resolutions. Hence, we couldn't validate/compare our model results with other OHC products. However, the model-derived long-term mean distribution of OHC will be compared against monthly data of ARMOR3D and IAP products in section 5 (to be included in the revised manuscript).

3. Sea surface height is more directly related to both ocean heat content and of course thermosteric anomaly, so why aren't say CNEMS SSH maps, available since 1993, used as the neural network analysis?

Though SSH data have a profound relevance with OHC variations, critical issues associated with satellite altimetry-based SSH data in the precise implementation of the ocean thermal expansion method were identified and discussed in section 4.3. SSH (SLA) data must be corrected for haline and water mass changes (barotropic flows) before being used in OHC estimations (Sato et al., 2000; Jayne et al., 2003). Water mass change-related correction can be implemented in real-time from GRACE mission data on monthly scales, but not on daily scales. On the other hand, remote sensing-based approaches are still lacking in estimating the haline correction factors (halosteric sea level anomaly) in real-time. Moreover, the base period employed in preparing CMEMS SSH data and World Ocean Atlas-2018 climatological OHC differs from one to another (lines 57-73, 331-338).

       The current study aims to provide a novel approach to implement the ocean thermal expansion method by eliminating all the potential sources of uncertainties as identified and reported in this study. Hence, the prime criterion followed in choosing the parameters for ANN modeling is based on the theoretical relationship between the input and output parameters rather than the direct usage of all the relevant parameters. The one-to-one relationship between OHC and TSL is employed in the OHC model. To arrive at TSL, the theoretical dependency of TSL on temperature and salinity is considered in TSL modeling work. This approach, when implemented in the current study, eliminates all the uncertainties associated with haline and water mass change corrections, base period, and other issues, as explained in sections 1 and 4.3.

4. Why are only Argo data used in the training? There are many more temperature profiles from shipboard CTDs, XBTs, marine mammals, and other sources readily available from the WOD that expand the spatial and temporal coverage Argo data considerably. Furthermore, Argo data should be downloaded directly from an Argo global data assembly centre, not WOD at NCEI, for the most recent version, as the data are subject to delayed-mode quality control and revisited frequently. As a result NCEI archives may be out-of-date.

The criteria considered in choosing the in-situ data sources are uniformity (in measuring techniques, instrumentation, and bias corrections), quality control (to eliminate erroneous or spurious data), and data coverage (in three-dimensional ocean space and time scales). Argo data have dense coverage in three-dimensional ocean space and time scales compared to other sources of CTD profile data. Employing Argo data alone suffice in achieving spatial and temporal coverage with uniformity.

       Regarding the data used in this study, the Argo global data assembly centers distribute CTD data in separate files for each float at their observed depth levels. These depth levels vary from one float to another. WOD at NCEI provides the Argo-based objectively analyzed CTD profile data at standard depth levels in a single NetCDF file. Moreover, WOD provides quality flags for each measurement (Cheng and Zhu, 2016). All these considerations made the WOD data highly suited for reading, eliminating erroneous profiles, and preparing in-situ OHC estimates from Argo data (Levitus et al., 2012; Ali et al., 2012; Jagadeesh et al., 2015; Cheng et al., 2017).

5. The data section covers only the in situ data, and not the fields used for prediction. Please consider moving the description of the SST and SSS "data" into the data section. Also, hopefully SSH data too, as that should certainly be useful in constructing the network. Another point, the ORAS5 "SSS" data are not from satellite measurements! Those maps are output of an ocean reanalysis, that is to say, an assimilation of in situ data (including all the Argo profile data) into a numerical ocean model. So ORAS5 SSS fields are certainly not independent from Argo data, in fact, they directly incorporate those data. This fact, and the model used and datasets ingested, are all well documented on the ORAS5 websites. Please study the documentation and revise the analysis, manuscript, and claims within accordingly.

The description of satellite-derived data will be moved to "Data" section in the revised manuscript.

The source of ORAS5 SSS will be specified as Ocean reanalysis system in the revised manuscript when it refers to the ORAS5 SSS dataset.

Owing to the dependency of ORAS5 SSS data on Argo profiles, we have considered the different time periods during model training and validation. The SSS data used during model training is from the Argo measurements over the period 2005-2016 and satellite-based validation is from ocean reanalysis data over the period 2017-2020.

6. Section 4.1. Given the relatively large (100s to 1000s of km) and long (30-120 days) decorrelation scales of ocean temperature anomalies, it is misleading to call or treat these points as "independent". Almost all of them will be well correlated with some of the data used to construct the ANN. This probably means the true uncertainties are much higher than estimated here.

In the current study, the database has been divided into training (2005-2016) and testing/validation (2017-2020) based on the time period. This eliminates the correlation between the data used during model training and testing phases. However, the data used during late 2016 and early 2017 possess some correlation, which is in the order of a few weeks, and its effect on the reported validation metrics can be neglected. However, by considering this part of SST data and ORAS5 SSS data, the manuscript will be revised by referring testing/validation datasets as *unseen* datasets instead of independent datasets.

7. What is the point of section 5? It gives neither new oceanographic information nor insight into the accuracy or usefulness of the maps generated. It might better to compare and contrast trends with trends from other products. If a section on the topic is retained, it would be beneficial to recruit a co-author familiar with oceanography, as the section as written is haphazard.

Thank you for your critical points. We will revise section 5 by providing spatial distribution of long-term mean OHC estimates for the period 1993-2020. OHC estimates across all the standard depth levels considered from 20 m to 700 m can be merged based on the GEBCO-2020 grid to arrive at depth-varying OHC estimates. Further, the observed model-derived OHC estimates and annual time series will be compared with annual OHC estimates obtained from the ARMOR3D and IAP products. Among the available mapped CTD fields and OHC products, ARMOR3D and IAP products comprise temperature and salinity data at all standard depth levels that are considered in the current work except 150 m and 650 m depth levels.

Minor comments

8. L24. 93% is an old estimate at this point. See von Schuckman et al. (2023) for an update.
The percentage OHC absorbed by oceans will be updated as 89% in the revised manuscript.

9. L27. The statement that this understanding is "inevitable" seems an odd word choice. It is important, or societally relevant, but that does not make it inevitable.
The word 'inevitable' will be replaced with the phrase 'is of great importance to the scientific community' in the revised manuscript

10. L32-33. Actually, just temperature profiles with depth information are also useful (e.g. XBT data). Measurements of conductivity on all profiles are not strictly necessary.
Sea water density and conservative temperature (Meyssignac et al., 2019) employed in OHC computation are functions of in-situ temperature, salinity, and pressure (IOC et al., 2010). Hence, simultaneous measurements of in-situ-based temperature, salinity, and pressure (depth) profile data are employed in this study to estimate OHC changes from in-situ data.

11. L42-87. This paragraph is very long and rather stream-of-consciousness in style. Consider breaking it up into a few paragraphs covering the different topics to better organize the exposition.
This paragraph will be divided into smaller paragraphs.

12. L48. There is a recent Zhao manuscript on the subject that should be referenced.
Zhao (2017) manuscript titled "Propagation of the Semidiurnal Internal Tide: Phase Velocity Versus Group Velocity" will be referenced in the revised manuscript.

13. L177-178. Here and elsewhere in the manuscript, sentences like "blah blah blah are shown in Figure 3" generally duplicate figure captions and interrupt the flow of ideas. It's much better to write a sentence that communicates the main thing the authors want the readers to get out of the figure and refer to it parenthetically.
We will implement this modification throughout the revised manuscript.

14. L241-249. The more energetic regions of the oceans will of course be more difficult to model. It might be appropriate to reference studies that have documented the spatial variability of eddy energy in the oceans, highlighting the western boundary current extensions and Antarctic Circumpolar Current, among other regions.
We will provide references to Beech et al., 2022 and Ni et al., 2023 in the revised manuscript.

15. L466. The phrase "Future releases of WOA will certainly resolve the complex patterns..." again seems odd, arguably incorrect, and out of place. Please consider deleting or rephrasing this sentence.
This sentence will be removed in the revised manuscript as it is conveyed well enough in the previous statement "More number of in-situ CTD profiles need to be collected and analyzed in such oceanic regions to address the associated complex patterns".