Review of **egusphere-2025-2782** “Estimating the AMOC from Argo Profiles with Machine Learning Trained on Ocean Simulations” by Yannick Wölker , Willi Rath , Matthias Renz , and Arne Biastoch

My expertise as a reviewer is mainly focused on the data science part, the parameter estimation, machine learning, statistics, error handling, and significance considerations, in the frame of oceanographic research questions.

General:

- The manuscript represents an important and interesting study on the potential of Argo floats to be used to estimate the AMOC, specifically the geostrophic part.
- I recommend publication after some minor clarifications, improvements and corrections.
- You mention the challenges of handling the irregular Argo data, which is reasonable. Then you overcome these difficulties with the embedding and graph-based NN approach, which is technically a very smart solution. However, you did not try an interpolation approach, bringing the Argo data on a regular grid and using those data as input for the feedforward NN. Thus, we don’t know if your graph-based approach is superior. In the end your results are good, which probably justifies your approach, however, for me it’s always the question if these results could have been achieved with simpler methods?
- Your approach, based on model data, shows that there is potential to reconstruct the AMOC utilizing Argo floats. However, for an application with real data, not enough data is available for a NN approach. So, what is not fully clear to me from the manuscript is, when we can reach “enough data”? Or, regarding to your discussions, is the only solution transfer learning, and enough data will not be available in a reasonable near future? Can you please clarify that?
- It is important to thoroughly always differentiate between the real elements in this study (AMOC, Argo, etc.) and the simulated. Please check all text.

Specific:

- Line 9-10: Add “… AMOC can be **potentially** data-drivenly …”
- Line 21: Are you referring to the North Atlantic Deep Water? But that is colder and **saltier** not fresher, or?
- Line 27: You say that ocean and climate models often fail to simulate the AMOC, but nevertheless you go for a full model analysis to draw inference on the real world.
- Line 53: Again make clear that you are not using real Argo.
- Line 93-95: I understand that the authors are going for a NN approach and real data is too limited in this case. However, why do the authors think that simpler approaches like linear regression may not be sufficient?
- Line 110: This sentence is confusing, what do you mean by “widely available observations”? If I understood correctly you are not using real observations.
- Line 113: You are not using Argo data!
- Line 110-124: Start the whole paragraph with explaining that you use simulated data.
- Page 5, Figure 1: c,d,e: Shouldn’t the middle x-label be the other way round 2024/1958?
- Line 182: First Argo floats have been deployed since 1997.

- Line 284: Difficult to understand. What is the “trained reconstruction”? And what means “the trained reconstruction is able to reconstruct…”?
- Line 303-327: You are saying that if you neglect the spatial info on Argo data, you can utilize a suitable neural network architecture. In the following you say you keep the spatial component using SUSTeR. I do not understand what in the end you do. In addition, understanding SUSTeR and explanations about traffic are not helpful. I suggest to remove this explanation and refer to the publication. Instead please make clearer what you have actually done in the end.
- Line 329-- Sect. 3.2: Regarding the training procedure of a NN, it would be interesting to see a “loss curve”. Often these loss curves are given for the performance of the model on the training set, during the training, as well as on the validation set (unknown).
- Line 374, Eqn. 4: I think the denominator is not Var(y), but the total sum of squares \sum{ (y_i-\overline{y})² }.
- Page 16, Fig. 4: I suggest to plot the reconstruction curve (blueish) on top of the ground truth (green) to better see it.
- Line 595: I guess you again randomized the Argo input data, not leaving it really out? Please mention in the text.
- Line 619: If I understand correctly by reading the full paragraph, the reason for no added value of deep Argo is probably just caused by not having enough training data. Thus the influence of deep Argo stays rather unknown. If that is true, please mention already here.
-

<u>Corrections:</u>

- Line 41: Rewrite this part, which sounds strange “… cables that measurement …”
- Line 48: Space missing “… balance(Mc …”
- Line 155: Figure ??
- Line 259: “… an high …” → “ … a high …” and “… an dedicated …” → “… a dedicated ….”
- Line 309: (?
- Line 396: “… due **to** larger …”
- Line 399: “The **the** …”
- Line 400: “… limits the compare …” → sounds strange, please rewrite.
- Line 427: “data(Jiang” → space
- Line 478: “… and the also the …”
- Line 536: “brach” ? → branch
- Line 542: “… due **to** the ...”
- Line 543: “… transport(Mc” → space
- Line 630: Change “We test if the test data lays within the training data and its ...” to “We investigate if the test data lie within the training data and if its ...”
- Line 706: “… amounts **of** diverse …”, delete: “… set of …”
- Line 736: “… mentioned,the …” → space