Detail response for reviews “Learning-based prediction of the particles catchment area of deep ocean sediment traps”

Dear editor and reviewers,

We are truly grateful for the time and effort you dedicated to carefully assessing our manuscript. Please find attached our revised manuscript, incorporating the revisions based on your feedback. We sincerely hope that you find this updated version to be reflective of your insightful suggestions. Please note that the line numbers mentioned here refer to the track changes file.

general comments

This paper presents a novel method for estimating the near surface origin of sinking particles that reach a sediment trap at 1000m. A neural network based on the U-Net architecture is trained to predict backward trajectories of sinking particles from environmental variables. The authors present the method and demonstrate its skill in a twin experiment setting. They also provide compelling analysis of the behavior of their estimation method depending on dynamical situation.

I feel this is a strong paper and recommend it to be accepted for publication after minor revision. The paper is generally very well written and clear. The method being introduced opens up interesting perspectives to help better interpret sediment trap observations, but the authors also acknowledge difficulties that lie ahead if such method was to be applied to real data. One of the strengths of this paper in my view is the analysis of results in relation to physical processes. A weakness is that we don't know how/if the method's skill could be quantified in the real world.

specific comments

- L66 : please provide a reference to justify the choice of "50m/day" as realistic on average in the real world, assuming that this indeed is the case. Is 50m/day representative of a large fraction of sinking biomass that sediment traps capture? Wang et al 2022, cited later, appears to be a modeling study by the same authors. Please state in the paper whether the choice of 50m/day is backed by real world observations, or otherwise.

As mentioned in the discussion, representing only one sinking speed is a significant limitation that will be evaluated in a subsequent study. However, as a preliminary step, we have chosen 50 m/day, which represents the slow range of sinking particles observed in the region at depth (Villa-Alfageme et al., 2014). Slow particles are more susceptible to being affected by ocean circulation, which in turn makes the prediction of catchment areas more challenging. We revised the manuscript to better express this choice lines 69-72 :

“*We focus on particles with a vertical sinking velocity of 50 m/day, which represents the slow range of sinking particles observed in the region (Villa-Alfageme et al. 2014). Slow particles*

are more susceptible to being influenced by ocean circulation, which in turn makes the prediction of catchment areas more challenging."

- L107 : "based on the statistical results of Wang et al 2022" is vague. What statistical results?

We have revised the sentence lines 116-118:

"*Based on Wang et al 2022, who computed the source region at 200 m for particles collected by the moored sediment traps over the seven years (2002–2008), an horizontal domain of 800 km x 800 km centered on the ST was chosen to encompass all the source particles position*."

- L138 : please explain and/or provide a reference for "Combining these steps with skip connections enables the detection of hydrodynamic structures at different spatial scales".

- Fig 4 : more explanation about "concatenation" and "channels" would also be useful in section 5.2 for the non specialist reader.

We have revised the paragraph. We have added Ronneberger et al. 2015 reference to justify that "skip connections enable the detection of hydrodynamic structures at different spatial scales" and we have detailed the meaning of the terms "channel" and "concatenation" :

" *A 3-step pooling process is used to downsample the channel, i.e. the dimension representing the number of dynamical images for one experiment. At each step, this number of features is doubled. This successive resolution downscaling from 8 km to 32 km, combined with skip connections and concatenation (addition over the channel dimension), could facilitate the detection of structures at different spatial scales (Ronneberger et al. 2015)*."

- L178 : can it be excluded that Unet4layers simlply benefits from a larger number of data constraints (Nx) that in UnetSurf rather than their subsurface location? I wonder if providing Nx time as many surface data to train UnetSurf would make it match the performance of Unet4layers.

The number of images taken as an input (N_{inputs}) shouldn't affect the score performance.

We tested this by increasing the number of channels in the first layer (128 instead of 64 channels), but this had no impact on the final score. In other words, we believe that the predictions are not limited by the number of parameters in the Unet but by the lack of information required in the inputs.

- L220-223 : state how much of the training sets was "chaotic situations" versus nonchaotic ones. I wonder : if you trained a NN separately on just "chaotic situations", would you get improved performance? maybe match the performance obtained for nonchaotic ones.

According to our criterion based on vorticity and KE, there are 1818 chaotic cases (27% of the experiences in the testing dataset). We have added this information in the manuscript line 269 :

“This "chaotic situation" represents 27% of our dataset.”

Since the number of chaotic cases is significant, it would indeed be interesting to train specific NN depending on the dynamical situation and this will be investigated in a future study.

- L294 : it seems to me that a confidence index would need to correctly account for several sources of model error (uncertain sinking velocity distributions, possible biases in POLGYR statistics, transport rates, etc) to avoid being misleading. Please discuss this in a bit more detail.

Indeed, the uncertainties associated with the numerical simulation must also be taken into account when calculating an index of confidence. We have mentioned this in the discussion lines 296-299 :

“*However, it is worth noting that other sources of uncertainty, such as sinking velocity distributions, simplification of numerical simulations, transport rates, etc., would also be important to evaluate.* “

- in section 5 please describe the kind of real world observational experiment and data sets that would be needed to demonstrate / quantify the method's skill outside of a twin experiment configuration. I feel that many would be concerned with using real world results of such a method if its skill can only be assessed within a model world.

A possible strategy to validate the model predictions with real observations would be to evaluate the cross-correlation between satellite chlorophyll in the prediction area and carbon fluxes measured at PAP (Frigstad et al. 2015). The cross-correlation coefficient can be compared with the cross-correlation obtained by considering a simplified catchment area, such as a 100 km or 200 km box around the PAP location, which is classically still used today (Lampitt et al. 2023). If a better correlation is found, this would indicate a better connection between deep fluxes at PAP and surface images from satellites and validate the relevance of the model to real observations.

We have added this details lines 301-306:

“*A possible strategy to validate the model predictions with real observations would be to evaluate the cross-correlation between satellite chlorophyll in the prediction area and carbon fluxes measured at PAP (Frigstad et al. 2015). The cross-correlation coefficient can be compared with the cross-correlation obtained by considering a simplified catchment area, such as a 100 km or 200 km box around the PAP location, which is a classical method still used today (Lampitt et al. 2023). If a better correlation is found, this would indicate a better connection between deep fluxes at PAP and surface images from satellites and confirm the relevance of the model application to real observations.*”

technical corrections

L57 : "used to design"

The correction has been applied.

L94 : space missing in "1000 m.Biological"

The correction has been applied.

L96 : spell out what 3D+t means. also, consider saying 4D instead of 3D+t or 3D+T

We gave a more explicit definition of “3D+T” line 103 :

“*The PDFs saved at vertical levels lower than 200 m provide information on the 3 spatial dimensions and temporal dimension (3D+T)*”

L100 : "To avoid common particle between two experiments" is unclear and seems grammatically incorrect

We removed the sentence and we added in the “Training, validation and test data” section lines 138-143 :

“*With the Lagrangian experiments presented above, we create a training and a validation dataset using the first simulation setup POLGYR1 (2002 to 2008). During this period, Lagrangian experiments are realised every 10 days time step and at 36 fictive ST positions around the PAP region resulting in a total of 10260 samples (Figure 3).*”

L101 : rephrase as e.g. "to set up the patch centers 36km apart"

The correction has been applied.

L104 : nx,ny would seem a better notation than deltax, deltay; no?

The correction has been applied.

L109 : "for storage constrain" is vague and seems grammatically incorrect.

We have removed the term.

Fig 3 caption : "superimposed" (one "s" only)

The correction has been applied.