Reply to Reviewer 2

Estimating ocean currents from the joint reconstruction of absolute dynamic topography and sea surface temperature through deep learning algorithms

By Daniele Ciani , Claudia Fanelli , and Bruno Buongiorno Nardelli

The authors are following their previous work on super-resolving and inpainting sea-surface height, this time focusing on a network learned through a new OSSE experiment. Their new study focuses on the Mediterranean Sea and is used to evaluate sea currents.
There are significant strong points in their approach. The article is well structured and scientifically sound, and can further a very active research field tied to the OceanChallenges data challenge. I especially appreciate the care taken to evaluate the physical fields obtained.

We thank the Reviewer for the constructive comments provided . Please find below our replies to the specific points.

In general, the authors are keenly aware of the bibliography in the field. However, given the many similitudes (there are differences too) I would like to see their positioning in regards to Archambault, Théo, et al. "Learning sea surface height interpolation from multi-variate simulated satellite observations." Journal of Advances in Modeling Earth Systems 16.6 (2024): e2023MS004047.

We thank the reviewer for suggesting this reference, which was added in the introduction and discussion section (please double check lines 68, 443-446 of the revised manuscript). There are indeed similarities between these two approaches, as both exploit the "transfer learning problem" from numerically simulated data to real-world data for the mapping of sea surface height (SSH), exploiting the combined use of SSH and Sea Surface Temperature (SST) data. Additionally, both methodologies result from fine-tuning/comparing different types of Loss function to train the neural network. Reading the manuscript we noticed the major following advantages (+) and disadvantages (-) of their approach, compared to ours.

Archambault et al. 2024 (A24 hereinafter)

+ Unlike our study, A24 rely on a significantly much longer time series to train the neural network, as they use a 20 years time series of simulated SSH/SST data, also including along-track altimeter observations, and they rely on one full year for the test. Such an approach certainly reduces the possibility of residual overfitting issues, as pointed out in our study.

+ The Availability of one full year for the final test is indeed extremely interesting, enabling one to assess the reconstruction performances as a function of the season.
+ The A24 set-up allows both supervised and unsupervised training, making the reconstruction methodology more applicable directly to satellite-derived observations
- According to the OSSE results, the A24 methodology is preferably applicable to predict observations in a temporal window centered at +/- 10 days, indicating that the methodology is more fit for delayed-time mode application. In an operational context, conversely, the methodology presented in our manuscript (e.g. using NRT SSH/SST L4 maps or L4 SSH and L3 SST maps) would result immediately applicable for near-real-time (NRT) processing.
- To the best of our understanding, the generation of the simulated satellite observations adopted in Archambault et al., although providing a realistic satellite-equivalent L3 dataset for SSH and SST, does not reproduce the optimal interpolation processing used within actual operational services (as done in our study for both SST and SSH data). We believe that our approach (in which we mimic both the SSH and SST 2D mapping artifacts introduced by the current algorithms within the Copernicus operational production) could be more directly implemented to improve present-day operational SSH and SST. We already identified the possibility of extending our OSSE in time and to account for training relying on "real-world observations". We thus believe it is worth mentioning this by making a link with the A24 publication. We inserted a comment on this at lines 443-446 of the revised manuscript.

I have some reservations in regards to the validation procedure since it has some contradictory information in the paragraph that starts at line 183. There seems to be attention paid to avoid data leakage but at the same time, early in the paragraph, the 40 days seem to be selected randomly. A clarification of which is the actual approach in this paper, and how it guarantees a significant enough time lag between data used in train, validation, and test is important.

In this work, the 40 dates kept aside for the validation are not purely random selected samples. As stated at line 192 of the revised manuscript, those dates are selected on purpose in order to perform the final test over different dynamical regimes, i.e. accounting for the full seasonality. Figure 8 indeed shows that the Julian Days over which we performed the final test are fairly equally distributed throughout the four seasons. Such 40 dates only represent ~11% of the entire time series. The rest of the time series is only used for training/validation purposes. The novelty of this work, which is summarized by figure 3 and explained at lines 200-204, 402-406 of the revised manuscript, lies in a different training/validation strategy. In particular, training is performed systematically excluding samples at the end of our time series i.e. the ones characterized by late autumn/winter dynamics. In other words, during training, the neural networks never "sees" features of the typical late autumn/winter period, where a more frequent presence of small scale features is expected (in full agreement with the findings of Callies et al. 2015). In this way, the neural network is asked to infer beyond what explicitly

learned in previous studies (e.g. BBN22). In BBN22, the validation procedure was based on randomly selected samples (from the ~89% percent of the timeseries used for training/validation purposes). In this way, it is likely that the neural network was trained/validated using samples that are too close to each other in time, potentially generating overfitting issues. In the newly proposed approach, the occurrence of overfitting can be reasonably excluded, as discussed at lines 320-327 of the revised manuscript. Moreover, the new architecture demonstrated to outperform the reconstruction illustrated by BBN22, confirming the validity of our upgrades.
However, we agree with the Reviewer that our approach still presents some limits. Relying on a single year for training/validation and test is suboptimal. Future applications plan to build an OSSE based at least on one full year for each of the following operations: training, validation and test, as claimed in the manuscript at lines 327-329, 443-446 of the revised manuscript.

The choice of architecture is interesting, and I expect that a lot of other approaches were tested. It would be interesting to include them in the annexes, as negative results are often ill-represented in literature.

Thanks for this comment. As pointed out at lines 78-80 of the revised manuscript, several architectures have been tested and thoroughly discussed in a former study, ranging from the baseline CNN configuration proposed by Dong et al. 2015 to the dADR-SR developed by Buongiorno Nardelli et al. 2022 (BBN22 hereinafter). Here, as mentioned at lines 81-85 of the revised manuscript, we investigated the effect of introducing simple, yet substantial improvements to the dADR-SR, which outperformed the other configurations tested by BBN22. The dADR-SR , in the former implementation, considered a non-realistic input SST field, limiting the potential of accounting for more sophisticated loss functions. In our study we thus introduced realistic SST / SST error fields which allowed us to modify:

1) The set of input/output data for the CNN training (as described at lines 210-212 of the revised manuscript)
2) The Loss function, with the introduction of a Physics-Informed approach (lines 288-313 of the revised manuscript)

Nevertheless, our achievements on the dADR-SR with the physics informed approach was not straightforward and required a couple of tests in order to understand how to shape the new, customized loss function given by eq (2) of the revised manuscript.

In a first attempt, the Physics Informed term of the Loss Function (LF) was given by (1.R2) and its Loss_phy term was expressed by (2.R2)

$$LF = \alpha[\overline{(SST_{pred} - SST_{ref})^2}] + \beta[\overline{(ADT_{pred} - ADT_{ref})^2}] + \gamma\{\overline{[(\partial_t SST)_{pred} - (\partial_t SST)_{ref}]^2}\} + \delta Loss_{phy} \quad (1.R2)$$

$$Loss_{Phy} = \left(\frac{\partial SST}{\partial t}\right)_{pred} - \frac{g}{f}\frac{\partial ADT_{pred}}{\partial y}\frac{\partial SST_{pred}}{\partial x} + \frac{g}{f}\frac{\partial ADT_{pred}}{\partial x}\frac{\partial SST_{pred}}{\partial y} \quad (2.R2)$$

Minimizing the Loss_phy term as in (2.R2) equals asking the predicted SST and ADT to obey to the horizontal geostrophic SST advection. When performing the initial test to find the weights of the Loss_phy term "delta" (as explained at lines 460-468 of the revised manuscript) we obtained delta=1.38. After training the neural network and reconstructing the ocean currents with satellite derived data, the comparison to in-situ measured currents yielded the results detailed in table t1.R2. Interestingly, in order to get a further RMS error reduction, we had to reduce the "delta" factor manually, which made the approach rather empirical.

We however noticed that the Loss_phy term given by (2.R2) was linear, unlike the ADT, SST and ∂tSST terms appearing in 1.R2, for which the minimization is given in a least squared sense. We thus decided to homogenize all the terms contributing to the Loss function , modifying Loss_phy as in (3.R2)

$$Loss_{Phy} = \overline{\left(\left(\frac{\partial SST}{\partial t}\right)_{pred} + \frac{g}{f}\frac{\partial ADT_{pred}}{\partial y}\frac{\partial SST_{pred}}{\partial x} + \frac{g}{f}\frac{\partial ADT_{pred}}{\partial x}\frac{\partial SST_{pred}}{\partial y}\right)^2} \quad (3.R2)$$

In this way, the weighting factor "delta" automatically adjusted to 0.025 and the effect on the reconstructed surface currents was an improvement for both components of the surface motion (i.e. a lower Root Mean Square error compared to in-situ measured currents), as expressed by table t1.R2. We thus decided to adopt the quadratic formulation for the Loss_phy term (shown in (3.R2)).

	RMS zonal (cm/s)	RMS meridional (cm/s)
Loss_phy "linear" (eq. 2.R2)	12.18	12.09

Loss_phy “quadratic” (eq. 3.R2)	11.90	11.85

Table t1.R2. Root mean square error computed by means of in-situ measured currents for both components of the surface currents and for the two formulations of the Loss_phy term.

We added such comments in Appendix A1, please see lines 470-485 of the revised manuscript

There is also a technical question of how the authors reconstruct the whole Mediterranean basin, and whether there are discontinuities in the full reconstruction.

The reconstruction of the Mediterranean Basin ADT (and derived geostrophic currents) is operated tile by tile. However, such tiles are not simply arranged adjacent to each other, as they consider a 50% overlap along the zonal and meridional directions (longitude-latitude). When reconstructing the tiles, we expect enhanced performances in the central area of the tile and potential spurious features towards the tile periphery, due to edge effects related to convolutional kernels. To overcome this, the reconstruction of the basin-scale field is obtained accounting for the longitude-latitude overlap but also applying a pixel wise weighting function that assigns a progressively decreasing weight to pixels lying at larger distances from the tiles centers. Such a methodology enables a seamless basin-scale reconstruction starting from the 76x100 tiles. We inserted a comment on that at lines 230-236 of the revised manuscript.

More details would be useful as to the process of selecting the hyperparameters of the new loss function, and some ablations on its usefulness would not be remiss.

We provide as an example the way we tuned the hyperparameters (alpha, beta) for the first two terms of the loss function, that we report below for convenience (equation 4.R2)

$$LF = \alpha[\overline{(SST_{pred} - SST_{ref})^2}] + \beta[\overline{(ADT_{pred} - ADT_{ref})^2}] \quad (4.R2)$$

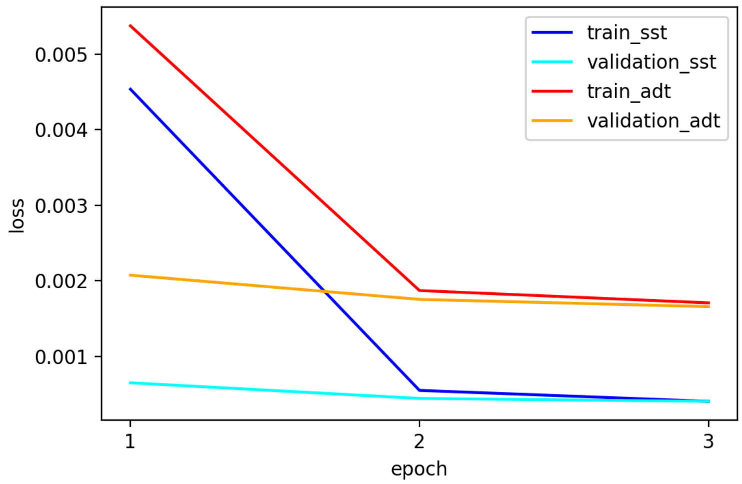

Figure 1.R2. Train validation losses employing a loss function built from ADT (red, orange, respectively) and SST (blue,cyan, respectively) separately.

The figure shows the behavior of the train/validation loss curves for the CNN employing the Loss function given by Eq. 4.R2, considering only the ADT or SST term, separately. This is shown for the first three epochs, as those curves converge quite quickly afterwards. The ratio between the ADT and SST train/validation loss curves can be quantified by comparing the curves and, from epoch 3 onward is around 4. As such, considering we impose alpha=1, beta has to be 0.25 in order to equally weight ADT and SST contributions. Such an exercise has been repeated for all the other terms of the loss function reported in the manuscript and has led to the final loss functions with the hyperparameters described at lines 460-468 of the revised manuscript. This was inserted in Appendix A1, together with the tuning of the physics - informed term. Please go to lines 470-485 of the revised manuscript.

While the evaluation of the currents is extremely important, the method relies on geostrophic approximation, which holds in the Mediterranean, but should be discussed in the case of applying this approach to other basins closer to the equator where this approximation does not hold.

We thank the Reviewer for this comment. The dADR-SR, in our approach, was built to improve the satellite derived ADT (as also stressed by Reviewer 1) and derived geostrophic circulation in the Mediterranean Sea. We believe this approach could be easily adaptable to other mid-latitude areas of the ocean. Equatorial areas could be studied via simple updates of the present-day architecture in which the training is not based on ADT but directly on surface currents. For example, the training of the CNN from numerical model outputs and based on the Observing System Simulation Experiment approach would account for the following datasets:

1) The satellite equivalent SST, ∂tSST and their error fields - (INPUT);

2) The satellite-equivalent-derived geostrophic currents (including as well the error field) - (INPUT)
3) The model-derived SST and ∂tSST - (TARGET)
4) The model-derived total surface currents - (TARGET)

In this way the neural network would be optimized to reconstruct directly the two components of the surface currents. This approach is also part of future studies on this topic, as already mentioned in the discussion section of the manuscript (see lines 433-436 of the revised manuscript)

In general, should the authors address these small points I would be very pleased to see their work published.

References

Buongiorno Nardelli, B., Cavaliere, D., Charles, E., and Ciani, D.: Super-resolving ocean dynamics from space with computer vision algorithms, Remote Sensing, 14, 1159, 2022

Callies, J., Ferrari, R., Klymak, J. M., and Gula, J.: Seasonality in submesoscale turbulence, Nature Communications, 6, 6862, 2015.

Dong, C., Loy, C. C., He, K., and Tang, X.: Image super-resolution using deep convolutional networks, IEEE transactions on pattern analysis and machine intelligence, 38, 295–307, 2015