Response to Reviewers – Manuscript gmd-2024-185 “Correction of Air-Sea Heat Fluxes in the NEMO Ocean General Circulation Model Using Neural Networks” by Storto et al., 2024

We thank the reviewers and the Editor for the careful reading and the suggestions to improve the quality and readability of the manuscript. Below, we provide a detailed reply to each of their comments (Reviewers’ comment/question in bold, our reply in italic font, and new text in the manuscript in blue).

Response to Reviewer 1

The use of artificial neural networks (ANNs) to model nonlinear relationships between atmospheric and oceanic state predictors supported by stationary predictors and heat flux errors appears remarkably effective.

Thank you for the encouraging comments.

It would be of interest to include in the manuscript information on the performance of the neural network in predicting the nudging increments using standard metrics like R2 or normalized mean square error or error maps. It appears that the network is able to learn a substantial part of the corrections. Figure 4 provides a hint.

Thanks for the suggestion, we now include the normalized (and not) error map applied to the validation (independent) data. See below. The figure is indeed interesting, and shows the low residual error (<10% everywhere, on average equal to 4% = 1.36 W/m2), with the dominant effect of variability error compared to the negligible systematic error of the ANN reconstruction. We have added a paragraph in the text to comment on the new figure:

In Figure 2 we show the error maps of the inferred heat flux correction from validation (i.e., independent) data. The Normalized RMSE (panel a) shows errors smaller than 10%, and on average equal to 4% (corresponding to 1.36 W m-2); while errors peak in areas of large mesoscale activity (western boundary currents and ACC), there exist other non-obvious local peaks. The systematic error of the ANN reconstruction is very low (panel c), generally not exceeding 0.7 W m-2, indicating that the RMSE is explained, to a great extent, by random errors (panel d shows the standard deviation of the differences).

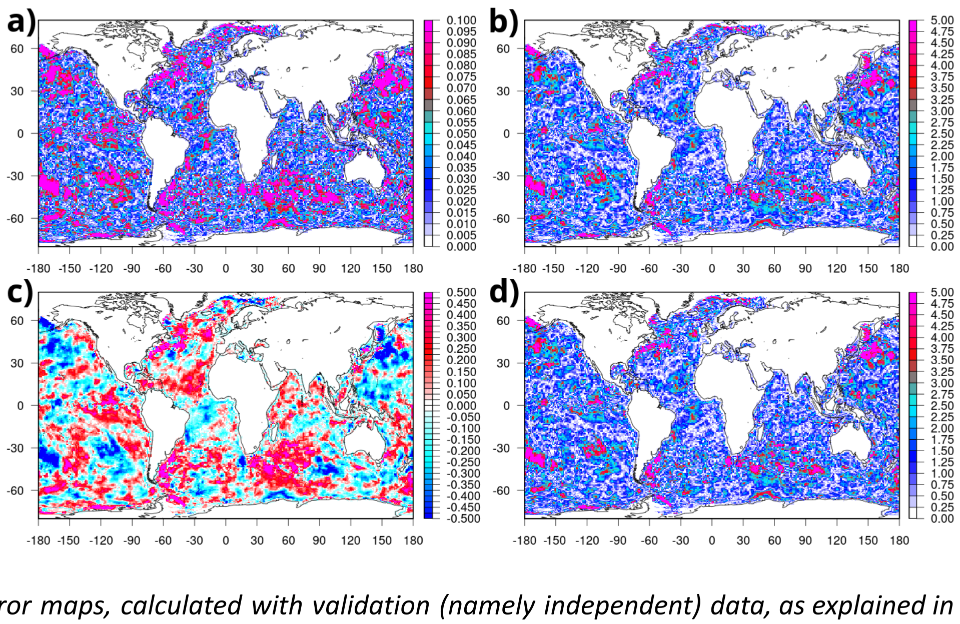

Figure R1. Error maps, calculated with validation (namely independent) data, as explained in the main text. a) normalized RMSE; b) RMSE in units of heat flux (W m-2); c) bias (W m-2); d) standard deviation of differences (W m-2).

The method has been demonstrated to be effective in the ocean only model. Non-linear feedbacks in the coupled models could affect the effectiveness of the ANN-based corrections. Do you foresee the method to work well in the coupled models? Perhaps applicability in the context of coupled modelling/coupled reanalysis could be discussed.

Thanks for this suggestion. We have now included a discussion on the potential for air-sea coupled modelling, extending the sentence already mentioning the issue in the original version of the manuscript.

The neural network-based heat flux correction method has proven effective in the ocean-only model by correcting systematic air-sea heat flux biases and improving surface and subsurface ocean temperature predictions. However, applying this method within a coupled ocean-atmosphere model may benefit climate drift correction (e.g., Gupta et al., 2013) but introduces additional complexities due to nonlinear coupled feedback. In a coupled model, the atmosphere could respond to modified fluxes in a nonlinear and potentially unpredictable manner. Heat flux corrections that work well in an uncoupled system may introduce unintended biases when the atmosphere reacts dynamically, potentially leading to unrealistic SST adjustments. Atmospheric variability (e.g., cloud cover, wind stress, and humidity) will alter in response to changes in SST, which could impact the efficacy of the NN-based correction. Corrections applied at short timescales may also have long-term impacts on coupled modes of variability (e.g., ENSO, MJO).

To make the NN approach more suitable for coupled applications, it could be retrained using data from coupled model reanalyses (e.g., CMIP simulations or CERA reanalysis datasets, for instance, Chapman and Berner, 2024). This would allow the NN to learn heat flux corrections in a system where atmospheric responses are accounted for, in analogy with flux correction or flux adjustment techniques (e.g., Sausen et al., 1988). The NN-based correction could be implemented to maintain the overall coupled energy balance while addressing systematic errors.

Sausen, R., Barthel, K. & Hasselmann, K. Coupled ocean-atmosphere models with flux correction. Climate Dynamics 2, 145–163 (1988). https://doi.org/10.1007/BF01053472

Gupta, A. S., N. C. Jourdain, J. N. Brown, and D. Monselesan, 2013: Climate Drift in the CMIP5 Models. J. Climate, 26, 8597–8615, https://doi.org/10.1175/JCLI-D-12-00521.1.

Chapman, W. E., and Berner, J.. A State-Dependent Model-Error Representation for Online Climate Model Bias Correction. ESS Open Archive . November 23, 2024. doi:10.22541/essoar.172526800.05354621/v2

Have you tested other ANN architectures like CNN? While CNN may be more difficult to apply online, it would be interesting to assess if it would be superior to the simple column model?

Unfortunately, it is difficult to test architectures embedding convolutional layers within our framework; this is mostly because of technical challenges linked with online inference in NEMO. Indeed, convolutional layers need to know exactly the MPI decomposition of NEMO, because each iteration requires data exchange across neighboring domains/cores. Inference software that is available at the moment (our ANNIF module, ECMWF/Infero, neural-fortran, or Cambridge-ICCS/fortran-tf-lib) does not allow specifying an MPI decomposition of the calling module. Another solution is being developed, based on leveraging Python API for the OASIS coupler (see Barge and Le Sommer, 2024), but it will be released in the future for the latest NEMO version, so for the time being we were forced to choose a model architecture not relying on convolutional layers. We have added a short discussion on that, in the final section:

In this study, we demonstrate the significant impact of online inference, which allows high-frequency (3-hourly) updates of the correcting fields. For this reason, testing different model architectures, e.g., those relying on convolutional layers, which require MPI communication across NEMO domains inside the convolutional filters, was technically complex and demanding. It is not obvious whether convolutional layers are beneficial compared to grid-point-wise corrections (see, e.g., different conclusions in Chen et al., 2022; Chapman and Berner, 2024), as the potential advantage of retaining horizontal patterns is balanced by the computational needs of coarsening the spatial resolution. In the future, more sophisticated inference libraries and tools for online prediction are expected to be available, paving the way for testing different neural network architectures.

References

Barge, A. and Le Sommer, J.: Online deployment of pre-trained machine learning components within Earth System models via OASIS, EGU General Assembly 2024, Vienna, Austria, 14–19 Apr 2024, EGU24-16148, https://doi.org/10.5194/egusphere-egu24-16148, 2024.