Responses to the Comments and Suggestions

Reviewer 2:

The authors produced a heat flux dataset based on a statistical neural network trained over model reanalyses and / or buoy data (I am not sure, it is not so clear to me after reading their manuscript). They compare their product to other products, and mostly find that their product performs better.

Re: Thank you for your comments. We have carefully considered all your comments and suggestions and made corresponding point-by-point responses and revisions. Specifically, reviewer comments are shown in black, our responses in blue, and the corresponding revisions in the manuscript are highlighted in red. We hope that our responses and the revised manuscript would be satisfactory.

There are strong chances that conceptually, this whole work would be no use, since the reanalyses used to train their network already provide the surface fluxes. Therefore, I really don't see the point in producing what I would call a 'statistical shortcut' of an existing model.

Re: Thank you for your comment. We would like to clarify that the target variables in our study are sourced from in-situ buoy observations, rather than from outputs provided by reanalysis products. Although the input features for our model include variables from reanalysis, the neural network is trained to reproduce observed air-sea turbulent heat fluxes, rather than merely replicating outputs of reanalysis. Accordingly, our approach should not be considered a "statistical shortcut" of existing models, but rather a methodology aimed at improving air-sea turbulent heat fluxes estimation by integrating observations with machine learning techniques.

My interpretation of the context is that historically, global surface flux datasets were developed at a time when model reanalyses were not accurate enough. In this context, independent blended-analyses gathering various satellite sensor fields and sometimes model forecasts (for stability and / or near surface air temperature) could be helpful for documenting the heat budget and it is spatial variability. Nowadays, satellite sensor data as well as in situ observations are widely assimilated in models, which results -in my opinion- in an optimum mix between physics (equations in the models) and observations, in terms of surface heat fluxes. Therefore, I don’t see why independent flux products (which are not even an ounce independent from models, since they are trained on them) should be developed any longer, the reason for which I left this field.

Re: Thank you for your comment. We respectfully note that we do not fully agree with the reviewer’s perspective. Currently, multiple global reanalysis products exist, and these products are still under development and not fully mature, which contrasts with the implication that additional independent flux products are unnecessary and that reanalysis represents an optimal mix between physics and observations. While we acknowledge that assimilating satellite data and in-situ observations into process-based models can improve the accuracy of air–sea turbulent heat fluxes simulations, it should be recognized that the accuracy of flux estimates is significantly influenced by the physical representation of air–sea exchange processes, model parameterizations, and biases in inputs. Therefore, assimilation alone does not necessarily guarantee high-accuracy flux estimates, which partially explains the continued need for model development and optimization.

With the rapid growth in the availability of flux observations, integrating machine learning models while fully accounting for the key physical and environmental factors influencing air–sea turbulent heat exchange has become an important approach for improving the accuracy and reliability of air–sea turbulent heat fluxes estimations. Indeed, in recent years, global estimations of carbon, water, and energy fluxes, ocean currents, and temperature/salinity fields using machine learning trained on in situ observations have become increasingly common (Chen et al., 2019; Cummins et al., 2023; Cutolo et al., 2024; Ge et al.; Zhou et al., 2024). Moreover, AI-driven models such as AlphaFold have achieved breakthrough progress in protein structure prediction, illustrating the substantial potential of artificial intelligence (Jumper et al., 2021). Other notable examples include OpenAI's GPT series in natural language understanding and generation, DeepMind's AlphaZero in surpassing human performance in complex strategy games, and deep-learning–based climate model parameterization and Earth system prediction (Brown et al., 2020; Rasp et al., 2018; Silver et al., 2018). Collectively, these successes demonstrate that scientific research is increasingly embracing a "data-driven + AI-assisted". In our view, flux estimation should be continuously improved by integrating emerging technologies in order to provide more accurate and reliable results.

At best, the authors' product will perform the same as model fields, which is obvious according to Figure 3 and 4 (compare ERA in panels d, to 'BrTHF' in panel i). Worse, there is one risk when aiming at getting the highest accuracy with artificial neural networks: overtraining. This could have been discussed in the manuscript. Please note that the proposed BrTHF product does not account for negative LHF values (Figure 4i).

Re: Thank you for your comment. With regard to the concern that our product may not outperform reanalysis, Figures 3 and 4 show that BrTHF achieves substantial improvements over ERA5, with RMSE reductions of ~1 W/m^2 (14%) for SHF and ~5 W/m^2 (16%) for LHF.

To address the reviewer's concern about potential overfitting, we implemented two measures to ensure the robustness and generalizability of our model. First, we employed a spatial 10-fold cross-validation, which provides a rigorous evaluation of model performance. Second, following the suggestions of another reviewer, we conducted targeted cross-validation by withholding two high-latitude buoy sites in the Southern Hemisphere, largely independent from the training dataset, as a validation set. As shown in Tables S4 and S5, BrTHF maintained higher accuracy than the other products and models, demonstrating its reliable generalization ability.

Regarding negative LHF values, we note that small negative values remain in Figure 4 (i), but their magnitudes are close to zero. This is mainly due to the uneven distribution of observations and the constraints applied to the BrTHF model, which prioritizes simultaneous high-accuracy estimation of SHF, LHF, and β. Consequently, the predicted range is compressed. We acknowledge this limitation and have discussed it in Section 3.5 of the original manuscript.

The authors focus on the Bowen ratio, which is supposed to give more ‘consistency in physics’, I don’t even know how to define/name it as they do… I am not convinced at all. Technically, I think it is just a matter of optimizing their neural network configuration.

Re: Thank you for your comment. Regarding “consistency in physics,” we would like to clarify that our main goal is to ensure that model outputs satisfy the physical relationship SHF/LHF = β. While this relationship can indeed be maintained in the reanalysis products highlighted by the reviewer, as shown in Figures 3–6, these reanalysis models cannot simultaneously provide SHF, LHF, and β with high accuracy. Conversely, using machine learning to model SHF, LHF, and β separately can achieve high accuracy for each variable individually, but such predictions do not necessarily preserve the physical relationship SHF/LHF = β. Therefore, our work emphasizes achieving both physical consistency (SHF/LHF = β) and high-accuracy estimation of all three variables, which demonstrates that our approach is not merely an optimization of the neural network configuration.

To me, it seems that the authors have downloaded a lot of data and model fields, and that they desperately look for a way to add some value using these datasets. If so, I would rather encourage the authors to analyze what is inside and produce case analyses, statistical analyses.

Re: Thank you for your comment. We would like to clarify that our study is not a simple aggregation of existing data. Instead, it aims to improve simultaneous estimation of SHF, LHF, and β through a physically-informed neural network—a novel approach beyond existing case studies or statistical analyses. The results demonstrate that BrTHF reduces both RMSE and bias of SHF and LHF compared to the existing state-of-the-art products. We believe this constitutes a meaningful contribution to the ongoing efforts in improving air–sea turbulent heat fluxes estimation.

In this manuscript, the principal publications are not even cited, which I consider to be a lack of respect to authors that did a pioneering work more than twenty years before them!

- Bourras, D., L. Eymard, & Liu, W. T. (2002). A neural network to estimate the latent heat flux over oceans from satellite observations, International Journal of Remote Sensing, 23(12), 2405-2423. doi: http://doi.org/10.1080/01431160110070825
- Bourras, D., Liu, W. T., Eymard, L., & Tang, W. (2003). Evaluation of Latent Heat Flux Fields from Satellites and Models during SEMAPHORE, Journal of Applied Meteorology, 42(2), 227-239. doi: https://doi.org/10.1175/1520-0450(2003)0422.0.CO;2
- Bourras, D. (2006). Comparison of Five Satellite-Derived Latent Heat Flux Products to Moored Buoy Data, Journal of Climate, 19(24), 6291-6313. doi: https://doi.org/10.1175/JCLI3977.1
- Bourras, D., Reverdin, G., Caniaux, G., & Belamari, S. (2007). A Nonlinear Statistical Model of Turbulent Fluxes, Monthly Weather Review, 135(3), 1077-1089. doi: https://doi.org/10.1175/MWR3335.1

Re: Thank you for your comment. We fully acknowledge and respect the contributions of the pioneering studies, and in the revised manuscript, we have now carefully revised the manuscript to include appropriate citations to these important references. We thank the reviewer for pointing this out.

Some comments for the introduction:

-L46 ‘the evaporative latent heat flux’: the term ‘evaporative’ is not appropriate in this sentence

Re: Thank you for your comment. We have removed the redundant term “evaporative” and now simply use “latent heat flux” for clarity.

-L47 ‘the conductive sensible heat flux’: wrong, it is convection, not conduction, except in the first microns above the water surface

Re: Thank you for your comment. We sincerely apologize for the typo and have corrected the relevant description accordingly.

-L51 ‘the Bowen ratio…revealing the partitioning of water and energy over the ocean and atmosphere’: this sentence does not make any sense, and it is not helpful, in addition to what the definition of the Bowen ratio is common knowledge in this field

Re: Thank you for your comment. In the revised manuscript, we have removed the related description and now provide the definition of the β upon its first appearance.

-L52-L54: ‘Accurate estimation of these three parameters is an essential prerequisite for advancing our understanding of atmosphere-sea interaction’… I don’t see why the Bowen ratio would be key, and the fluxes as well as the Bowen ratio are not ‘parameters’ but ‘variables’, in this context

Re: Thank you for your comment. We agree that the use of the term “parameters” in this context could be misleading, and we have revised it to “SHF, LHF and their ratio—

the Bowen ratio (β = SHF/LHF)". We also acknowledge the reviewer's concern regarding the role of β. We would like to clarify that while SHF and LHF individually describe the components of turbulent heat fluxes, β provides additional insight into their relative partitioning at the air–sea interface. This ratio not only captures differences in climate regimes (e.g., large β in cold and dry regions such as the subpolar North Atlantic, and small β in tropical and subtropical oceans), but also reflects the synergistic variations between SHF and LHF (e.g., both SHF and LHF may increase while β remains unchanged), which cannot be inferred from either flux alone. Therefore, we consider β to be an essential variable for advancing the understanding of atmosphere–ocean interactions.

-L57-L61: 'To map global air-sea... as developed and widely adopted as a primary approach'. This sentence is nonsense. The Monin-Obukhov (1954) similarity theory was not developed for that, and I am not aware of any 'primary approach'

Re: Thank you for your comment. We agree that the Monin–Obukhov similarity theory was not originally developed for mapping global air–sea fluxes, and it is not accurate to describe it as a 'primary approach' in this context. In the revised manuscript, we have revised the sentence to more appropriately reflect its role as a theoretical foundation widely used in flux parameterization schemes in the second paragraph of Section 1 as follows:

"To estimate global air–sea turbulent heat fluxes, the semi-empirical bulk aerodynamic method was developed based on the Monin–Obukhov similarity theory (Monin and Obukhov, 1954). It establishes scaling relationships between fluxes and near-surface meteorological variables such as wind speed, humidity, and temperature (Yu, 2019)."

-L58: 'easily': I don't see why it would be 'easy' to measure mean meteorological quantities, it is rather complicated, just try to get a reliable information with two thermometers mounted close to each other on a ship or on a buoy, it is a real challenge.

In addition, this includes SST, which is not a meteorological variable, strictly speaking

Re: Thank you for your comment. We apologize for the inappropriate wording and have made the corresponding corrections in the manuscript. We also acknowledge that including sea surface temperature (SST) in this context was misleading, and we have now corrected this accordingly.

-L59: ‘metrological’: Wrong, I think the authors mean ‘meteorological’

Re: Thank you for your comment. We have revised “metrological” to “meteorological”.

After reading this one and half paragraph I have noted so many inaccuracies and / or wrong statements, that I don’t feel compelled to review in detail the rest of the manuscript. This manuscript looks like a science paper, but from far. To me, it is way too weak to be published.

Re: We sincerely appreciate the reviewer’s feedback. We fully acknowledge the concerns raised regarding inaccurate statements in the manuscript, and have carefully considered all comments, undertaking substantial revisions to address these issues. At the same time, we have incorporated the constructive suggestions and comments provided by another reviewer, which have further enhanced the clarity, rigor, and overall quality of the manuscript. We believe that, following these revisions, the manuscript now presents meaningful and valuable scientific contributions.

Other comments, maybe not in order of line numbering:

-The manuscript is unnecessarily long, difficult to read. It contains unnecessary acronyms such as THF, and it contains unnecessary equations, such as the equation 1 that relates the relative humidity to the dew point temperature, which is common knowledge

Re: Thank you for your comment. In the revised manuscript, we have removed the unnecessary acronyms and equations for conciseness.

-Figure 1 is unclear

Re: Thank you for your comment. We have reorganized the flowchart to improve its readability, as shown below:

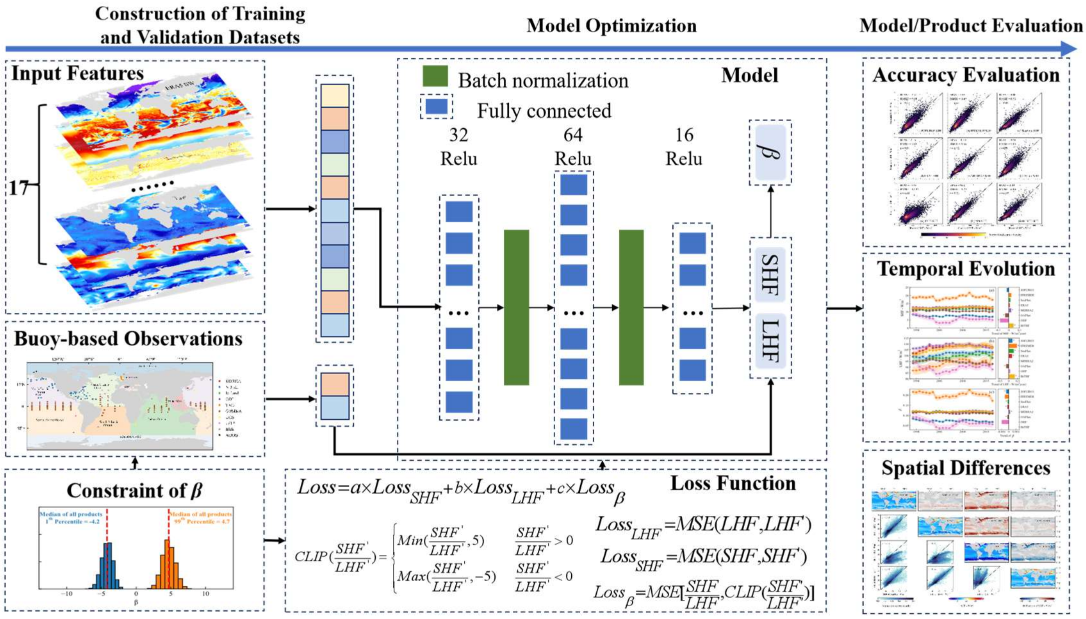

Figure 1. flowchart of the generation of a global product of air-sea SHF, LHF and β by the BrTHF model

-Figure 5 is statistically pointless

Re: Thank you for your comment. We would like to clarify that the main purpose of Figure 5 is to present and compare the distribution of β estimates from our model and other products against observations. The highlighting of outliers in the Figure 5 is intended to demonstrate that our model effectively avoids the outliers found in other models and products. Additionally, since each panel shows two modes (with and without outliers), to maintain the figure's clarity and avoid redundancy, detailed statistical information can be found in Table S2 and Figure 6. To prevent any misunderstanding, we have added an explanation in the caption of Figure 5 in the revised manuscript:

“Figure 5. Same as Figure 3 but for β. The samples out of the ranges of observed β ($-5 \leq \beta \leq 5$) were colored in blue, orange, green, red, purple, brown, pink and gray for JOFURO3, IFREMER, SeaFlux, ERA5, MERRA2, OAFlux, OHF products and the physics-free NN models, respectively. The corresponding statistical metrics can be found in Table S3 and Figure 6.”

-At several locations in the manuscript, the terminology used may be considered as misleading, such as L122 where they mention ‘the superiority of the model’. In this sentence, ‘model’ is ambiguous because it does not refer to a meteorological model or a physical model of any kind, but rather to a statistical model. At L136, there is also a reference to the ‘BrTHF model’

Re: Thank you for your comment. We agree that the term "model" may be ambiguous without clarification. In our original manuscript, we referred to the BrTHF model as "a Bowen ratio-constrained model using the machine learning technique," which implicitly indicates that it is a statistical model. However, to avoid potential ambiguity for readers, we have revised the first appearance of the BrTHF model to clearly state that it is a *“Bowen ratio-constrained statistical model using the machine learning technique”*. We continue to use the term “BrTHF model” throughout the manuscript for readability. Additionally, we have revised the sentence to specify that we are referring to the statistical model developed in this study.

-In the same fashion, section 2.2 is entitled ‘forcing datasets’, which I think also adds to the confusion, because forcing is usually used by ocean modelers. Here, it should be ‘learning’, which term is widely used in the field of multilayer perceptrons

Re: Thank you for your comment. We have revised the title of Section 2.2 to “Learning datasets for training the neural network” and updated related terminology throughout the manuscript.

-In section 2.2, I could not easily understand whether only model analyses were used for the learning (which I think), or if it is a mix with buoy data.

Re: Thank you for your comment. We would like to clarify that in our neural network framework, model analyses were used as input features, while buoy-based SHF and LHF observations served as the target variables for training. Accordingly, we have revised the relevant descriptions in the second paragraph of Section 2.2.1 to improve clarity as follows:

"Datasets of these learning variables used as input features for training the neural network were collected from multiple publicly available sources, as summarized in Table 2 and were used as the input features for training the neural network."

-L112-L113 and L104: the authors mention several times the EC fluxes are high quality compared to bulk estimates, which denotes a complete lack of knowledge in this field. EC fluxes are very difficult to obtain at sea because of platform motion and airflow distortion, even at turbulent scales. To get more insights, the authours should consider reading the following references, for example:

- Bourras, D., Weill, A., Caniaux, G., Eymard, L., Bourlès, B., Letourneur, S., Legain, D., Key, E., Baudin, F., Piguet, Traullé, O., Bouhours, G., Sinardet, G., Barrié, J., Vinson, J.-P., Boutet, F., Berthod, C., & Clémençon, A. (2009). Turbulent air-sea fluxes in the Gulf of Guinea during the AMMA Experiment, J. Geophys. Res., 114, C04014. doi: https://doi.org/10.1029/2008JC004951
- Bourras, D., Cambra, R., Marié, L., Bouin, M.-N., Baggio, L., Branger, Beghoura, H., Reverdin, G., Dewitte, B., Paulmier, A., Maes, C., Ardhuin, F., Pairaud, I., Fraunié, P., Luneau, C., & Hauser, D. (2019). Air-sea turbulent fluxes from a wave-following platform during six experiments at sea, J. Geophys. Res., 124, 4290–4321. doi: https://doi.org/10.1029/2018JC014803

Re: Thank you for your comment. We would like to clarify that our reference to EC fluxes as "high quality" was intended to emphasize their value as direct measurements of turbulent heat fluxes, rather than to suggest that they are easy to obtain. We fully acknowledge that EC measurements at sea are challenging due to platform motion and airflow distortion, even at turbulent scales. In the revised manuscript, to avoid possible misinterpretation, we have removed the wording describing EC fluxes as “high quality” and have revised similar statements elsewhere in the manuscript. Furthermore, we have carefully reviewed the literature recommended by the reviewer and added these references in the fifth paragraph of Section 1 to highlight the challenges of obtaining EC measurements over the ocean as follows:

“However, since EC observations are difficult to obtain at sea due to platform motion and airflow distortion (Bourras et al., 2019; Bourras et al., 2009)—their limited spatio-temporal coverage constrains the application of the model for global mapping.”

Reference：

Bourras, D., Cambra, R., Marié, L., Bouin, M.N., Baggio, L., Branger, H., Beghoura, H., Reverdin, G., Dewitte, B., Paulmier, A., Maes, C., Ardhuin, F., Pairaud, I., Fraunié, P., Luneau, C. and Hauser, D., 2019. Air‐Sea Turbulent Fluxes From a Wave‐Following Platform During Six Experiments at Sea. Journal of Geophysical Research: Oceans, 124(6): 4290-4321.

Bourras, D., Weill, A., Caniaux, G., Eymard, L., Bourlès, B., Letourneur, S., Legain, D., Key, E., Baudin, F., Piguet, B., Traullé, O., Bouhours, G., Sinardet, B., Barrié, J., Vinson, J.P., Boutet, F., Berthod, C. and Clémençon, A., 2009. Turbulent air‐sea fluxes in the Gulf of Guinea during the AMMA experiment. Journal of Geophysical Research: Oceans, 114(C4).

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G. and Askell, A., 2020. Language models are few-shot learners. Advances in neural information processing systems, 33: 1877-1901.

Chen, S., Hu, C., Barnes, B.B., Wanninkhof, R., Cai, W.-J., Barbero, L. and Pierrot, D., 2019. A machine learning approach to estimate surface ocean pCO2 from satellite measurements. Remote Sensing of Environment, 228: 203-226.

Cummins, D.P., Guemas, V., Cox, C.J., Gallagher, M.R. and Shupe, M.D., 2023. Surface Turbulent Fluxes From the MOSAiC Campaign Predicted by Machine Learning. Geophysical Research Letters, 50(23).

Cutolo, E., Pascual, A., Ruiz, S., Zarokanellos, N.D. and Fablet, R., 2024. CLOINet: ocean state reconstructions through remote-sensing, in-situ sparse observations and deep learning. Frontiers in Marine Science, 11.

Ge, L., Wang, G., Huang, B., Cao, C., Chen, X. and Chen, G.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P. and Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873): 583-589.

Monin, A.S. and Obukhov, A.M., 1954. Basic laws of turbulent mixing in the surface layer of the atmosphere. Contrib. Geophys. Inst. Acad. Sci. USSR, 151(163): e187.

Rasp, S., Pritchard, M.S. and Gentine, P., 2018. Deep learning to represent subgrid processes in climate models. Proc Natl Acad Sci U S A, 115(39): 9684-9689.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D. and Graepel, T., 2018. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. Science, 362(6419): 1140-1144.

Yu, L., 2019. Global Air–Sea Fluxes of Heat, Fresh Water, and Momentum: Energy Budget Closure and Unanswered Questions. Annual Review of Marine Science, 11(1): 227-248.

Zhou, S., Shi, R., Yu, H., Zhang, X., Dai, J., Huang, X. and Xu, F., 2024. A Physical‐Informed Neural Network for Improving Air‐Sea Turbulent Heat Flux Parameterization. Journal of Geophysical Research: Atmospheres, 129(17).