Review of 'Bowen ratio-constrained global dataset of air-sea turbulent heat fluxes from 1993 to 2017'

By Yizhe Wang, Ronglin Tang, Meng Liu, Lingxiao Huang, and Zhao-Liang Li

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.

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.

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. 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)

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.

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.

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)042<0227:EOLHFF>2.0.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 Air–Sea Fluxes, Monthly Weather Review, 135(3), 1077-1089. doi: https://doi.org/10.1175/MWR3335.1

Some comments for the introduction:

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

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

-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

-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

-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’

-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

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

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

-Figure 1 is unclear

-Figure 5 is statistically pointless

-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’

-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

-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

-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