Interactive comment on "Using a data-driven statistical model to better evaluate surface turbulent heat fluxes in weather and climate numerical models: a demonstration study (https://doi.org/10.5194/egusphere-2024-568)"

Dear reviewer 2,

We thank the reviewer for his/her helpful suggestions, which led to significant improvements of our paper. Below we detailed how his/her comments are addressed in the revised version of the paper. The corrections are cited here in *italic*. *We refer to specific pages by "P" and lines by "L". For example, "**P1, L1**" stand for page 1, line 1 in the revised version.*

The results section could use some additional context and discussion, as well as revision for clarity. The connection between the background/motivation/introduction and the results gets lost at times. One small change that could assist this is a more specific naming of the cases. In section 5 in particular, I found myself frequently confusing the two MLP based fluxes and struggled at times to immediately understand the comparison being made. Perhaps assigning abbreviated case names (one for each of the four (or five if you count the different grid cells): estimated fluxes, observed fluxes, fluxes in the same environment, simulated fluxes) as well as text reminding what exactly they represent within the section could improve this clarity. The authors could use the naming already present in the figures in the text, for example, to have strong consistency. Language throughout sections 4 and 5 connecting back to the goals and motivation in the beginning of the paper would also help promote cohesion and make it easier to interpret.

Abbreviations are now used to clearly distinguish between observed fluxes (OBS), simulated fluxes (SIM), and MLP-based fluxes in the observed (MLP_{OBS}) and simulated environments (MLP_{Grd}). All these abbreviations were added in Figure 1. The comparison figures have been significantly updated to correspond with the schematic illustration in Figure 1. Accordingly, the discussion in section 4 and 5 have been revised for clarity.

Finally, there are a few differences between the simulations and the tower observations (and the MLPs based on them) that hinder comparison. While the authors do not necessarily avoid talking about them, the discussion on them is scattered throughout and could be enhanced with a more detailed and focused discussion. In particular answering:

- How does the mismatch of temporal resolution (30 min vs 3 hour) affect the results, particularly since we would not expect fluxes to be stationary over 3 hours (especially during the mornings/evenings)?

The mismatch in temporal resolution between observational and model data (30 minutes vs. 3 hours) can obviously introduce some challenges, especially since fluxes are not typically stationary over a 3-hour period, particularly during transition between stable and unstable regimes. In the revised version, a section was added (Section 5.1) to clearly discuss the impact of the temporal resolution in both the comparison and the performance of the

statistical model. We initially applied this approach to an already existing climate simulation, and in future steps, it will be used for simulations performed in the framework of the MOSAI project.

- How effectively can we compare between 20 km grid cells with different (and heterogeneous) land cover and a tower with ~4m agl flux readings which is likely only reading a small area from a grassland?

This study aimed to develop an evaluation method that firstly addresses the disparity in environmental conditions between simulation and observation. Future work in MOSAI project will address other challenges that hinder reliable evaluation of land surface scheme. These challenges include, non-closure of SEB in observed fluxes and the representativeness of in-situ measurement at the coarser horizontal resolution of numerical model. An enhanced observation period of one-year has been conducted for this purpose, by adding additional flux measurements over different types of land use around the main site. A general paper presenting the MOSAI project will be submitted soon. The conclusion has been modified for clarity (***P33, L613-620***).

- What is lost by neglecting the soil/surface temperature? Those should have a strong correlation with the sensible heat flux in particular.

Surface temperature and sensible heat flux are indeed closely related. However, due to a significant amount of missing data (~40%), we chose not to include surface temperature when deriving the input variables for the statistical model.
Table below shows matrix of correlation coefficient between sensible heat flux (H), net radiative flux (Rnet) and potential temperature θ_{surf} (calculated with surface pressure and surface brightness temperature) for all available half-hourly observational data (between July 2015 to December 2022) .

	H	R_{net}	θ_{surf}
H	-	0.912880	0.721587
R_{net}	0.912880	-	0.721798
θ_{surf}	0.721587	0.721798	-

It can be observed that H is more strongly correlated with R_{net}, while both R_{net} and H show similar correlations with θ_{surf}. As R_{net} is one of the input variables, there is no significant loss of critical information regarding the variability in H and even LE. A new sentence was added in the revised version (***P06, L133***) :

Sensitivity analysis indicates no significant loss of key information concerning the variability in H and LE.

Technical Corrections have been done