Dear Reviewer,

Thank you for your effort on review of my submission. Your comments and suggestions are helpful for my current submission and future research. Now, We respond to your comments item-by-item. Your comments in blue and my response in black.

1- There is an overall lack of general context and discussion on the added value of the Transformer_CNN method: why do the authors directly use the ML based reconstruction without trying a more classical (physics-based) method of flux computation? This paper is submitted to GMD as a ‘development and technical paper’; and as such, it should clearly assess the performance of the model presented with respect to existing methods (not limited to ML-based gap filling). If the aim of the study is rather, as stated l. 93-94, to ‘complete the imputation of turbulent heat flux for this site spanning from 2007 to 2016 and make this dataset publicly accessible’, this study will be more usefully published as a ‘data paper’ in some dedicated journal. Can the authors explain why a basic flux computation algorithm is not usable here? If so, is it due to the different regimes of atmospheric conditions and soil covers encountered in the site throughout the year? Also, can you add information about the significance of the differences between the methods, based on the statistical indicators used in part 4? The SVM method already provides very good results in my view. Is the difference between the SVM and Transformer_CNN MAE significant? Or the distance between the SVM and Transformer_CNN positions in the Taylor diagram in Fig. 5a? The SVM method is simple and ready-to-use, what is the added value of developing a new, ML-based method for gap filling time series of EC fluxes? This added value could be efficiently demonstrated by a comparison of the resulting time series (by adding the SVM reconstruction to Fig. 8 for instance, or to a close up of it). There would also be an interest in identifying the time periods where ML algorithms yield different results from physical parameterization, to demonstrate the contribution of ML, and possibly in discussing the reasons of this discrepancy.

The variation in turbulent heat flux represents a profoundly complex process. This implies that simple linear models might not accurately capture the intricate relationships between meteorological elements and turbulent heat flux. As a result, traditional statistical methods may fail to provide accurate predictions in certain circumstances. Moreover, these conventional techniques typically depend on historical data, meaning that if specific weather conditions or combinations of meteorological elements have not been previously recorded, the model might be incapable of effectively predicting such situations. This issue is particularly pertinent in the context of climate change and the increasing frequency of extreme weather events, as these phenomena could lead to the emergence of new and unknown meteorological conditions. Machine learning models offer an advantage in this scenario because their combination of environmental driving variables can address the complex non-linear relationships among predictive variables, regardless of their interdependencies or correlations, and the intended outcomes.

The current techniques for interpolating missing data on turbulent heat flux include linear interpolation, variable relationships, neighboring substitutions, daily mean variation, non-linear regression, and lookup tables. Linear interpolation is only suitable for small gaps (1-3 consecutive missing data points), but at the QOMS site on the Tibetan Plateau, turbulent heat flux often exhibits long-term missing observations, rendering linear interpolation unreliable. Variable relationships, which use the linear relationships between meteorological variables for mutual interpolation, share similarities with ML methods. However, due to the strong non-linear relationship between turbulent heat flux and environmental driving factors, mere variable relationship methods struggle to accurately capture changes in turbulent heat flux.

Neighboring observations involve interpolation using nearby observation stations, which have standard meteorological measurements and share similar terrain and environmental conditions. However, even if geographically close to flux observation towers, the significant spatial variability in microclimates, especially in complex terrains like mountainous regions, may lead to changes in meteorological elements. The extensive area and sparse station distribution on the Tibetan Plateau make this method unsuitable.

The daily mean variation method involves establishing a time window, typically 4-15 days, with 7 and 14 days being the most common choices. Observations at the same time within this window are averaged to obtain a set of daily mean variations. Missing data within these means are interpolated using linear interpolation, with the corresponding time's daily mean variation data used to fill in the gaps. However, if the data used for calculating daily variations are based on specific conditions, then the interpolated data may not accurately reflect real-world situations.

Non-linear regression is based on an understanding of the primary factors controlling the flux, thus effectively capturing the impact of major environmental element changes on the flux, allowing for more accurate data interpolation. However, pre-determined equations may not always be effective, as, under conditions such as drought, moisture availability becomes a crucial controlling factor, suitable for short-term data gaps.

The lookup table method is based on creating a data retrieval table from valid data, searching for valid data under similar environmental conditions by primary environmental factors, averaging the found data to interpolate missing data. However, the environmental elements used for the search must be complete, and the data may be dispersed, as even under similar environmental conditions, differences in wind direction and terrain could lead to significant variation in the found data, thereby increasing the uncertainty in the interpolation.

Given the harsh geographical conditions of the Tibetan Plateau, we employ the daily mean variation method, non-linear regression, and lookup tables to interpolate data at the QOMS site, exploring the gap between these methods and machine learning approaches.

Table 1 Interpolation effect of traditional method and Transformer_CNN method on test set

	H			LE		
Sets	RMSE	MAE	R^2	RMSE	MAE	R^2
Mean Diurnal Variation	40.26	5.91	0.68	95.17	17.19	0.49
Nonlinear Regressions	38.44	6.90	0.72	107.54	18.76	0.53
Look Up Tables	52.64	9.86	0.54	142.18	21.43	0.40
Transformer_CNN	**18.56**	**3.598**	**0.95**	**15.45**	**2.909**	**0.89**

The above is a comparison between classic physical methods and the Transformer_CNN method. From the table, it is apparent that the deep learning method Transformer_CNN has a clear advantage in predicting turbulent heat flux.

The differences between SVM and Transformer_CNN can be distinctly observed through the use of monthly average daily variation graphs. Figure 1 presents the monthly average daily variation graph for simulating sensible heat flux using the SVM model, while Figure 2 displays the corresponding graphs for both Transformer_CNN and Transformer as discussed in the article. It is evident from these figures that due to the prevalence of low-value data points within the turbulent heat flux dataset, the training outcomes of the SVM model are significantly influenced by these low-value instances, leading to challenges in accurately simulating daily variations in turbulent heat flux. While the Transformer model introduces improvements over the SVM baseline, the incorporation of CNN within the Transformer_CNN architecture markedly enhances this aspect. The integration of various convolutional kernels allows the Transformer_CNN model to accurately capture the periodic changes in turbulent heat flux, showcasing its superior performance in this domain.

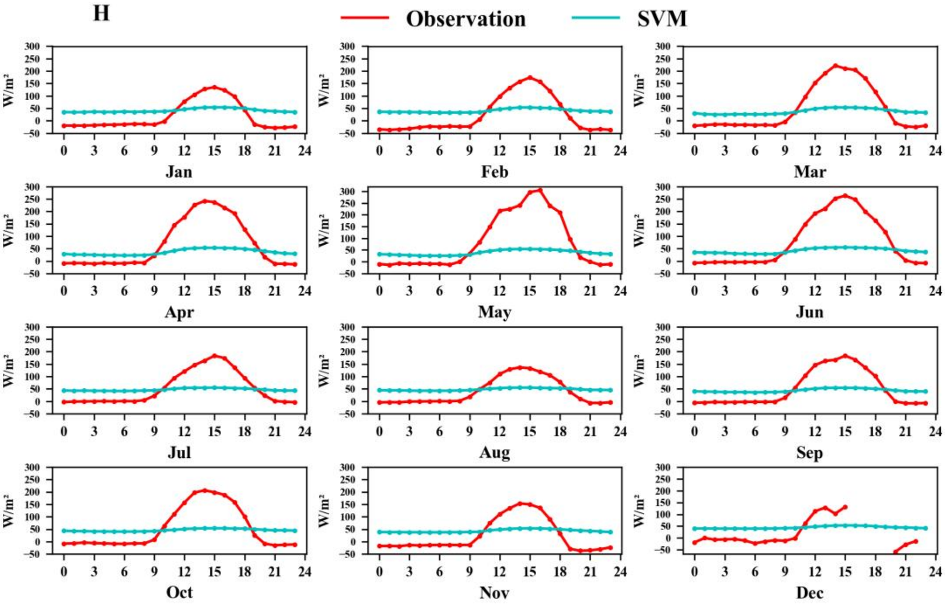


Figure 1 Observed values and SVM estimated values for the monthly average diurnal variation curves in the test set (2012) are shown.

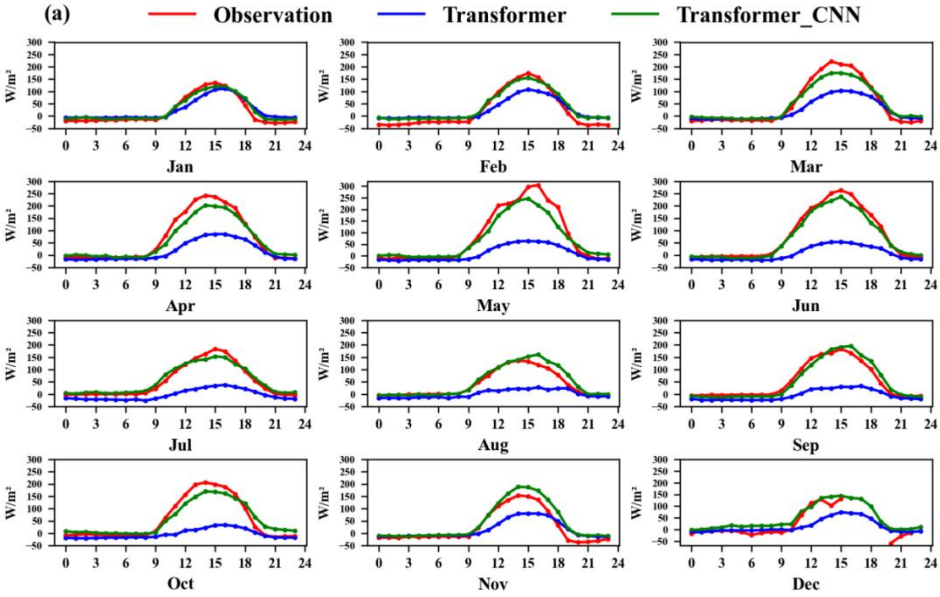


Figure 2 Observed values, Transformer estimated values, and Transformer_CNN estimated values for the heat flux monthly average diurnal variation curves in the test set (2012) are shown.

2- The physical meaning of the results is probably worth a discussion. In 2.3, what is the physical meaning of the variables ranking first for H/LE? Please comment why the numbers of variables, and variables themselves are different for H and LE. Some correlations between subgroups of variables are probably rather high (e.g. Ta_2m and Ta_1.5m, RH_1.5m and RH_2m): could the same results be obtained with less variables? Not all sites provide measurements of the soil temperature between the surface and 4 m, or air temperature between 1.5 and 10 m. Could the same (very good) fit be obtained with measurements at first levels (Ts 0m and Ta 1.5m) only?

At the QOMS site on the Tibetan Plateau, the surface cover characteristics are defined by barren, relatively flat, and open terrains with sparse and low-lying vegetation, composed primarily of sand and gravel from the surface to deeper soil layers. Initially, the barren surface and sparse vegetation imply that less solar radiation is absorbed by plants for photosynthesis, allowing more shortwave radiation energy to reach the ground directly, thereby increasing surface heating. Secondly, the flat and open terrain, coupled with a sandy and gravelly texture, facilitates these areas to efficiently absorb and re-radiate solar energy, significantly influencing the formation of sensible heat flux. Consequently, downward shortwave radiation, as the primary energy input source, plays a particularly crucial role under such environmental conditions, elucidating its top-ranking importance in predicting sensible heat flux.

In the context of these specific surface conditions, when employing the Random Forest model to rank the importance of variables for latent heat flux, soil moisture content emerges as the most significant. This predominance is attributable to the crucial role of soil moisture in regulating the surface energy balance within such arid and barren environments. Soil moisture content not only affects surface evaporation and vegetation transpiration, thereby controlling the magnitude of latent heat flux, but even minimal variations can have a substantial impact on latent heat flux under conditions of water scarcity. Therefore, in arid regions like the Tibetan Plateau, soil moisture content becomes a pivotal variable for predicting latent heat flux, significantly outweighing other factors.

As illustrated in Figures 3-c and 3-d, employing a reduced set of environmental driving variables results in a slight decrease in the accuracy of turbulent heat flux interpolation. However, overall, the performance remains commendably high.

This observation is based on the outcomes utilizing fundamental meteorological elements (These elements include single-layer air temperature, pressure, single-layer air humidity, single-layer wind speed, single-layer wind direction, site hourly average precipitation, ground net radiation, single-layer soil temperature, and single-layer soil moisture content) on the test set (year 2012), demonstrating that the interpolation effectiveness does not significantly diminish with fewer variables（Table 2）.

Table 2 When the environmental driving variables are fundamental meteorological elements, the performance of Transformer and Transformer_CNN in simulating sensible and latent heat fluxes on the test set (year 2012) is evaluated.

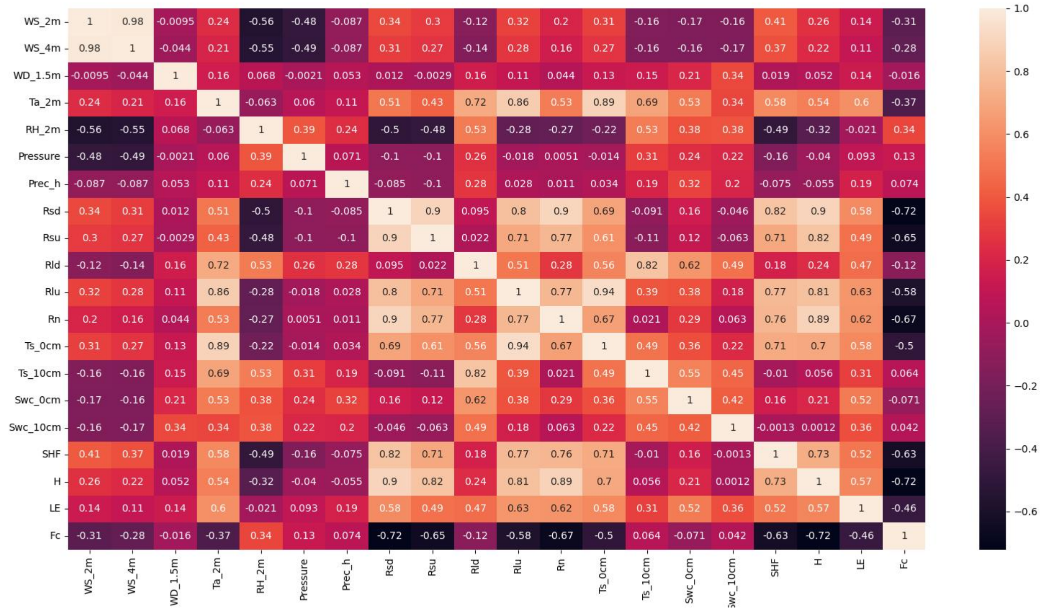
Sets	QOMS						SETORS					
	H			LE			H			LE		
	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2
Transformer	34.76	4.36	0.74	37.58	4.77	0.69	39.96	5.28	0.70	42.48	4.79	0.67
Transformer_CNN	29.34	3.44	0.83	30.25	3.93	0.78	31.66	4.83	0.80	34.22	4.61	0.79

3- Can the method presented here be used directly or with adaptation at different sites? It would be really interesting to add sensitivity tests of the importance of the different variables selected in the H (18) and LE (16) subgroups. Would the fit be significantly lower when excluding RH_2m and/or RH_4m from the H subgroup?

The model proposed in the article can be directly applied to other sites, as evidenced by Figure 3 in the article, which demonstrates a gradual improvement in the model's fitting performance. The removal of RH_2m and RH_4m does not significantly impact the effectiveness of the model. Additionally, the data presented in Table 2 further corroborate that when utilizing basic meteorological elements for fitting, the results are notably commendable.

4- The preprocessing part is insufficiently explained. Why is it relevant to use random forest

(RF) rather than a principal component analysis (PCA) to select relevant variables? PCA are based on linear correlation between variables. Are you sure that the linear model is not appropriate here? Did you checked that a simple correlation / covariance analysis does not lead to the same results as in Fig 3? Also, can you please explain briefly the meaning of the ‘importance ranking’ (Fig. 3 a and b) and, more importantly, of the ‘OOB score’ for the readers not familiar with the RF terminology?



	WS_2m	WS_4m	WD_1.5m	Ta_2m	RH_2m	Pressure	Prec_h	Rsd	Rsu	Rld	Rlu	Rn	Ts_0cm	Ts_10cm	Swc_0cm	Swc_10cm	SHF	H	LE	Fc
WS_2m	1	0.98	-0.0095	0.24	-0.56	-0.48	-0.087	0.34	0.3	-0.12	0.32	0.2	0.31	-0.16	-0.17	-0.16	0.41	0.26	0.14	-0.31
WS_4m	0.98	1	-0.044	0.21	-0.55	-0.49	-0.087	0.31	0.27	-0.14	0.28	0.16	0.27	-0.16	-0.16	-0.17	0.37	0.22	0.11	-0.28
WD_1.5m	-0.0095	-0.044	1	0.16	0.068	-0.0021	0.053	0.012	-0.0029	0.16	0.11	0.044	0.13	0.15	0.21	0.34	0.019	0.052	0.14	-0.016
Ta_2m	0.24	0.21	0.16	1	-0.063	0.06	0.11	0.51	0.43	0.72	0.86	0.53	0.89	0.69	0.53	0.34	0.58	0.54	0.6	-0.37
RH_2m	-0.56	-0.55	0.068	-0.063	1	0.39	0.24	-0.5	-0.48	0.53	-0.28	-0.27	-0.22	0.53	0.38	0.38	-0.49	-0.32	-0.021	0.34
Pressure	-0.48	-0.49	-0.0021	0.06	0.39	1	0.071	-0.1	-0.1	0.26	-0.018	0.0051	-0.014	0.31	0.24	0.22	-0.16	-0.04	0.093	0.13
Prec_h	-0.087	-0.087	0.053	0.11	0.24	0.071	1	-0.085	-0.1	0.28	0.028	0.011	0.034	0.19	0.32	0.2	-0.075	-0.055	0.19	0.074
Rsd	0.34	0.31	0.012	0.51	-0.5	-0.1	-0.085	1	0.9	0.095	0.8	0.9	0.69	-0.091	0.16	-0.046	0.82	0.9	0.58	-0.72
Rsu	0.3	0.27	-0.0029	0.43	-0.48	-0.1	-0.1	0.9	1	0.022	0.71	0.77	0.61	-0.11	0.12	-0.063	0.71	0.82	0.49	-0.65
Rld	-0.12	-0.14	0.16	0.72	0.53	0.26	0.28	0.095	0.022	1	0.51	0.28	0.56	0.82	0.62	0.49	0.18	0.24	0.47	-0.12
Rlu	0.32	0.28	0.11	0.86	-0.28	-0.018	0.028	0.8	0.71	0.51	1	0.77	0.94	0.39	0.38	0.18	0.77	0.81	0.63	-0.58
Rn	0.2	0.16	0.044	0.53	-0.27	0.0051	0.011	0.9	0.77	0.28	0.77	1	0.67	0.021	0.29	0.063	0.76	0.89	0.62	-0.67
Ts_0cm	0.31	0.27	0.13	0.89	-0.22	-0.014	0.034	0.69	0.61	0.56	0.94	0.67	1	0.49	0.36	0.22	0.71	0.7	0.58	-0.5
Ts_10cm	-0.16	-0.16	0.15	0.69	0.53	0.31	0.19	-0.091	-0.11	0.82	0.39	0.021	0.49	1	0.55	0.45	-0.01	0.056	0.31	0.064
Swc_0cm	-0.17	-0.16	0.21	0.53	0.38	0.24	0.32	0.16	0.12	0.62	0.38	0.29	0.36	0.55	1	0.42	0.16	0.21	0.52	-0.071
Swc_10cm	-0.16	-0.17	0.34	0.34	0.38	0.22	0.2	-0.046	-0.063	0.49	0.18	0.063	0.22	0.45	0.42	1	-0.0013	0.0012	0.36	0.042
SHF	0.41	0.37	0.019	0.58	-0.49	-0.16	-0.075	0.82	0.71	0.18	0.77	0.76	0.71	-0.01	0.16	-0.0013	1	0.73	0.52	-0.63
H	0.26	0.22	0.052	0.54	-0.32	-0.04	-0.055	0.9	0.82	0.24	0.81	0.89	0.7	0.056	0.21	0.0012	0.73	1	0.57	-0.72
LE	0.14	0.11	0.14	0.6	-0.021	0.093	0.19	0.58	0.49	0.47	0.63	0.62	0.58	0.31	0.52	0.36	0.52	0.57	1	-0.46
Fc	-0.31	-0.28	-0.016	-0.37	0.34	0.13	0.074	-0.72	-0.65	-0.12	-0.58	-0.67	-0.5	0.064	-0.071	0.042	-0.63	-0.72	-0.46	1

Figure 3 QOMS site variable heat map

Simple correlation/covariance analysis can yield results similar to Figure 3 in the article, but the outcomes are not pronounced (as shown in Figure 3). I believe that the importance ranking of random forests can achieve results more intuitively and accurately. My choice of random forests is primarily due to considerations of the nonlinear relationships in turbulent heat flux. Random forests are capable of handling nonlinear relationships between variables, whereas PCA is primarily based on the linear correlations among variables. Although PCA is a powerful dimensional reduction technique capable of simplifying datasets by extracting the most significant components, it is mainly suitable for uncovering linear structures within data. Given that the simulation of turbulent heat flux often involves complex nonlinear dynamics, directly applying PCA may not adequately capture these dynamics. This makes random forests a more suitable choice, as they can better capture and explain these nonlinear relationships.

Variable Importance Evaluation: Random forests offer an intuitive way to assess the contribution of each variable to the model's predictive performance through importance ranking. Variables that contribute more significantly to the model's predictive performance are ranked higher. This method not only helps identify the most important predictive variables but also reveals potential interactions between variables, which is difficult to achieve with PCA.

OOB Error Analysis: The Out-of-Bag (OOB) error in random forests is a method for evaluating model performance based on data not selected as part of the training set (i.e., OOB data) during the model training process. The OOB score provides an unbiased estimate of the model's generalization ability without the need for additional cross-validation processes. In terms of predictive performance, the OOB error can effectively assess the model's ability to predict unknown data, making the model's generalization capability a key consideration.

In summary, through capturing nonlinear relationships between variables and providing an unbiased estimate of generalization ability via OOB scores, random forests offer a more precise and reliable method for simulating turbulent heat flux. Although PCA can be useful in certain contexts, it may not be the optimal choice when selecting environmental variables in the presence of complex nonlinear relationships.

5- The building of the Transformer_CNN method itself is insufficiently detailed for the readers not familiar with ML in general. Is it something new, built on purpose for the present study, or has it been used previously? In the first case, can you please elaborate on the reasons leading to the choice of the different steps and modules used? In the latter case, please provide some references. It would also be interesting to study the attention weights determined by the transformer to analyze the causal link between certain variables and the reconstructed fluxes. This provides a perspective on the physical interpretability of the transformer performances.

The Transformer_CNN model is a novel deep learning framework specifically designed to address the complex physical phenomena of turbulent heat flux or similar challenges. It integrates the features of Convolutional Neural Networks (CNN) and Transformer, aiming to capture the intricate relationships temporal dimensions. Below are the reasons for choosing different steps and modules within the model:

Layer Normalization and Feed Forward Network:
Layer Normalization: Normalization layers stabilize the training process and accelerate model convergence. Commonly seen in Transformers, they help mitigate the vanishing or exploding gradients issue in deep networks.
Feed Forward Network: A standard component of the Transformer architecture, it processes each position in the sequence. Here, it's used for the initial processing of the input, mapping each input point to a higher-dimensional space for more complex processing.

Parallel Convolutional Layers:
Convolutional Kernels of Various Sizes (7, 5, and 3): This is a key improvement in the model. The design of parallel convolutional layers aims to capture spatial features at different scales. By employing convolutional kernels of varying sizes, the model learns features over different temporal cycles of turbulent heat flux, enhancing its performance in simulating changes over years, months, and days.
Concatenation after Convolutional Layers: Concatenating the feature maps from three scales, followed by a 1x1 convolution to mix these features, aids the model in integrating information across scales, enhancing its expressive power.

Transformer Module:
Multi-head Attention: The multi-head attention mechanism allows the model to learn the internal dependencies of the input sequence from different representational sub spaces in parallel. This is highly effective for understanding complex relationships within sequences.
Dropout: The use of dropout in the attention mechanism prevents over fitting and improves the model's generalization ability.
Summation of Self-attention: This operation allows the model to integrate new information learned through the self-attention mechanism while retaining the original input information.
Decoder Network:

Linear Layers and GELU Activation Function: This design maps the output of the Transformer back to the prediction target. The GELU activation function, a relatively newer nonlinear function, is thought to perform better than the traditional ReLU in models predicting turbulent heat flux.

Loss Function for Contrastive Learning:
Application of Shared Weights and Dropout: By invoking the model's forward method twice for the same input Xtrain, generating two different views (F1 and F2). Though not directly mentioned in the code, the comment suggests that due to dropout, the outcomes of these two forward passes may differ. In deep learning, dropout is a regularization technique to reduce overfitting by randomly dropping a fraction of neurons during the training process.
Definition and Weighting of the Loss Function: Utilizing SmoothL1Loss as the loss function without reduction (reduction='none') allows for the weighting of each sample's loss. This approach emphasizes those samples with higher weights during training, improving the model's predictive performance on specific samples, especially in predicting turbulent heat flux where extreme or unusual data points may be involved.
Application of Contrastive Learning: The loss between F1 and F2 and Ytrain is calculated using the weighted loss function, and the sums are added. Moreover, a contrastive loss term (0.1 * torch.dist(F1, F2, p=2)**2) is introduced, aiming to bring the two views of the same sample closer in the feature space. Contrastive learning is a powerful technique, enhancing the model's generalization and recognition abilities by learning to differentiate between samples.
Prediction Process: In the non-training mode, the model returns the average of F1 and F2 as the final prediction result. This practice may help reduce prediction variability caused by dropout, thereby stabilizing the model's predictions.

6- Abstract: the RF is not part of the methods evaluated with SVM and so on. Please reformulate.

Upon reviewing your comment, I realized that mentioning RF alongside Support Vector Machine (SVM) and other methods was indeed an error in my writting. I want to clarify that RF was not intended to be compared directly with SVM and the other methods in the evaluation section of our study. The mention of RF was a mistake and has been corrected in the subsequent version of the document. I appreciate your attention to this detail, and I apologize for any confusion it may have caused. I have taken steps to thoroughly review the entire manuscript to prevent similar issues and ensure the accuracy of our presented methods and results. Thank you for your understanding and for helping us improve the quality of our work.

7- What is the sampling of the variables used to reconstruct the fluxes? I guess hourly samples (l. 210 and for the study of the diurnal cycle), but daily values are probably used in Fig. 2, Fig 6 and Fig 9? Please specify.

I assume you're talking about Figures 8
The temporal resolution of the data from the QOMS site is hourly. In Figures 2 and 8, daily average data are utilized because the use of hourly data would result in a cluttered appearance in the overall graph. However, hourly data is used in Figure 6.

8- Figure 2: the legend is flawed and incomplete, please complete.

We have completed the legend for Figure 2 in subsequent sections of the article.

9- Table 1: please discuss the statistics and trends provided here. Are these figures useful for the main analysis presented in the paper ?

The mention of the basic meteorological conditions and their changing trends at the QOMS site is intended to provide readers with a better understanding of the site's situation. Many readers may find this information of interest. We believe that this will create a complete narrative, making the article more comprehensive.

10- Please use the same acronyms for the variables throughout the text, tables and figure legends (e.g. table 1 and table 2);

Below are the corrected tables 1 and 2

Table 3 MK Statistics and Fitting Equations

Indicator	MK	P-value	Fitting Equation	Trend
Wind speed	-0.036	4.48e^{-51}	$Y = -5.19\times10^{-8}X + 3.01$	Downward
Air temperature	0.023	3.90e^{-23}	$Y = 1.19\times10^{-5}X + 3.69$	Upward
Humidity	0.017	7.19e-^{13}	$Y = 2.26\times10^{-5}X + 42.43$	Upward
Precipitation	0.023	4.60e^{-16}	$Y = 1.75\times10^{-8}X + 0.03$	Upward
Soil temperature	0.017	3.14^{-10}	$Y = 4.99\times10^{-5}X + 7.49$	Upward
Soil water content	-0.19	0	$Y=-1.44\times10^{-5}X+3.18$	Downward

Table 4 Installation Heights and Burial Depths of the Observation Instruments

Variables	Sensor models	Manufacturers	Heights	Units
Air temperature	HMP45C-GM	Vaisala	1.5,2,4,10and20m	^{o}C
Wind speed and direction	034B	MetOne	1.5,2,4,10and20m	$ms^{-1}/^{o}$
Humidity	HMP45C-GM	Vaisala	1.5,2,4,10and20m	%
Pressure	PTB220A	Vaisala	-	hPa
Radiations	CNR1	Kipp&Zonen	-	Wm^{-2}
Precipitation	RG13H	Vaisala	-	mm
Soil temperature	Model107	Campbell	0,0.1,0.2,0.4,0.8and1.6m	^{o}C
Soil water content	CS616	Campbell	0,0.1,0.2,0.4,0.8and1.6m	v/v%
Soil heat flux	HFP01	Hukseflflux	0.05m	Wm^{-2}
H	CSAT3	Campbell	3.25m	Wm^{-2}
LE	LI-7500	Li-COR		

11- Please define MAE, RMSE, FC

We assume you're talking about R^2, Fc stands for carbon dioxide flux.

Root Mean Square Error (RMSE): RMSE represents the square root of the mean of the squared errors, which is the average of the differences between the simulated and observed values. The lower the RMSE value, the better the model's fit.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

Mean Absolute Error (MAE): MAE calculates the average of the absolute differences between the observed and predicted values. A lower MAE value indicates a better fit of the model..

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

Coefficient of Determination(R^2): R^2 measures the proportion of the variance in the dependent variable that is predictable from the independent variables. This metric indicates how close the data are to the fitted regression line. The closer R^2 is to 1, the more effectively the model explains the data's variability.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2}$$

12- Some references are not correctly cited in the text (l. 39)

I will correct my references to make them compliant with GMD publication requirements

13- 181: please explain where this value of 159 comes from.

The value of 159, based on ten-fold cross-validation and grid search algorithms, sets the number of estimators for the Random Forest model to 159, reflecting a method to find the optimal balance point of model parameters. Ten-fold cross-validation divides the dataset into ten parts, using nine of them in rotation for model training and the remaining one for testing, thereby effectively evaluating the model's generalization ability on unknown data. The grid search algorithm, by traversing a predefined set of parameter combinations, such as the number of estimators, seeks to find the parameter setting that provides the best predictive performance. In this case, choosing 159 as the number of estimators represents a balance found in experiments, considering both the model's predictive performance and computational efficiency.

14- 378-388: this analysis is interesting and could be completed /made more specific by making explicit reference to the variables influencing the flux variations (Fig. 3).

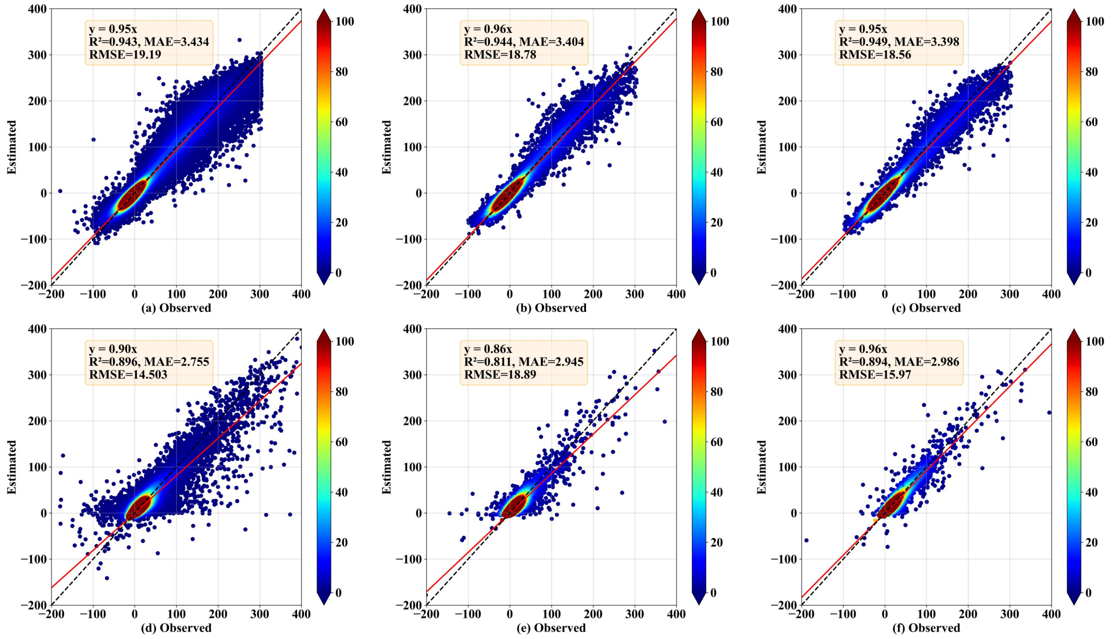
Through the analysis of Figure 3, we observe that the most significant factor affecting sensible heat flux is downward shortwave radiation (Rsd), while latent heat flux is primarily influenced by soil water content (Swc). Specifically, in the pre-monsoon season, the value of Rsd is high, reaching the annual peak. Correspondingly, the sensible heat flux increases with the rise in Rsd, also achieving the highest value of the year. As the monsoon season approaches, with an increase in rainfall, the Rsd value gradually decreases, and simultaneously, the soil water content progressively increases. During this process, the latent heat flux exhibits an upward trend, while the sensible heat flux decreases. After the monsoon season, Rsd gradually rebounds but does not reach its previous peak; the sensible heat flux rises, reaching a smaller peak. Meanwhile, a reduction in soil water content also leads to a decline in latent heat flux. This analysis clearly reveals the impact of changes from the pre-monsoon season to the monsoon season on sensible and latent heat fluxes, further confirming the critical role of downward shortwave radiation and soil water content in regulating the surface energy balance process.

15- Tables 4 and 5, and Fig. 5 are somehow redundant. Figure 5 can probably be moved to an appendix.

Thank you very much for your insightful feedback regarding Tables 4 and 5, and Figure 5. I agree that this adjustment will enhance the clarity and conciseness of our manuscript. We have changed it in the revised version.

16- Figure 6 is ill-designed. Please use the same scale for x and y axes and enlarge to be sure to include all the data.

The changes in Figure 6 are shown below：



17- Figure 8 is not very clear: I guess that the reconstructed values (red) are masked by the observations (purple) when both are present. Plotting the reconstructed values above the observed ones would probably make it clearer.

The modification as shown in Figure 8 represents the complete dataset formed after imputation, where red indicates reconstructed values and purple indicates actual values. The aim of this paper is to create the most complete and accurate dataset possible by imputing missing parts of the original dataset with reconstructed values, while the non-missing parts of the original data are supplemented with precise observational data. Thus, the dataset formed is the most complete and accurate dataset.

Once again, we sincerely appreciate your insightful comments, which have undoubtedly strengthened the quality of our work. We have made the necessary revisions based on your suggestions, and the improved manuscript now better meets the journal's requirements.

Best regards,

Sincerely

Quanzhe Hou, Zhiqiu Gao, Zexia Duan, and Minghui Yu

March 30, 2024