Response to the RC1

RC1: 'Comment on essd-2024-9'

The manuscript presents a comprehensive dataset detailing land-atmosphere interactions over the Tibetan Plateau, derived from 12 field stations covering a range of landscapes. This dataset encompasses hourly measurements of surface energy balance components, soil hydrothermal properties, and near-surface micrometeorological conditions for up to 17 years (2005-2021). However, I have several major concerns that the authors should address.

Response: We are grateful to reviewer #1 for the effort reviewing our paper and the constructive feedback provided. Here below we tried our best to address all the concerns and suggestions raised by the reviewer #1. We hope that the modification made on the revised manuscript will cover the reviewer expectation. Changes highlighted in red have been made accordingly in the revised manuscript. The revised sentences are highlighted in blue in the following replies.

1) Section 2 provides extensive detail on the observation infrastructure and data post-processing workflow, including data processing, quality control, gap filling, and archiving procedures. The authors should include more explicit information on the calibration of instruments across different stations and the rationale behind the selection of specific quality control algorithms. Comparisons with standard practices in the field could help in benchmarking the dataset's reliability.

Response: Thank you very much for your careful review and insightful comment. The quality control procedures implemented in this study are standard practices following the guidelines described by Zahumensky (2004) as a proposal submitted to WMO. These quality control procedures are currently widely adopted to detect errors and ensure quality of meteorological data, despite varying degrees of modifications have been made to deal with site difference (e.g., different climate conditions, measured meteorological variables, and sensor hardware specifications) and specific issues. The design of the entire data quality control workflow in this study and some of the thresholds (for example, the plausible rate of change) used in the testing protocols are also followed the WMO's recommendation. Those standard quality controls of meteorological data, like checks for range limits and temporal consistency checks, are adjusted to the requirements of the instrumental and geographical settings of the research sites. This is due to the fact that generic methods frequently failed to deal with site-specific issues and unique problems that emerged from the field observations. The following texts were added to the original manuscript based on your suggestion.

- Calibration instruments:

Line 214-228: Calibration of instruments is critical for ensuring accurate measurements. It is important to note, however, calibrating in a particularly harsh environment such as the TP is challenging. As a result, for meteorological and soil observations, both of which are relatively stable, calibrated reference instruments were used on a regular basis to perform field calibration across multiple stations, or the calibration was performed in a laboratory setting when instruments were returned for repair. In the case of turbulent observations, the measurement accuracy of the gas analyzer (i.e., LI-7500 and LI-7500DS) depends upon the cleanliness of the instrument lenses, it needs to be calibrated at regular intervals (once every six months at the five sites affiliated with the ITPCAS) due to signal attenuation for CO_2/H_2O. The calibration consists of two major components: 1) determining the values of the calibration coefficients, and 2) adjusting zero and span to align the gas analyzer's actual response with the previously determined factory response. In addition, we conduct monthly inspections of the operational status of all observational equipment (Ma et al., 2023), as well as semi-annual on-site instrument maintenance for all stations, which includes instrument cleaning, checking the level of commissioned instruments, and checking instrument cables and connectors.

■ Selection of specific quality control algorithms:

Line 254-261: To provide the best level of accuracy feasible, an automatic processing scheme was specifically designed for each type of variable, following the guidelines described by Zahumensky (2004). Despite a wide array of methods has been proposed to obtain plausible micrometeorological data series, those methods share similar processing flow but varying degrees of modifications were made to deal with site-specific concerns and unique problems that emerged from the field observations. This is due to the fact that generic methods frequently failed to resolve these issues. This scheme is specifically adapted aimed at verifying the reliability of observations and detecting errors and suspicious values. The automatic data processing chain was built up as a series of sequential checks recommended by Zahumensky (2004), with emphasis on continuity and inter-consistency of meteorological fields to detect suspect observations.

2) The authors should provide a comprehensive and detailed explanation of the data collection methods and quality control procedures employed in their study. Instead of merely listing various methodologies, it is crucial to elaborate on how data was gathered, the criteria used for data selection, and the specific steps taken to ensure the integrity and accuracy of the data.

Response: We completely agree with the comment made by the Reviewer and really appreciate for pointing this issue out. We have supplemented the Section 2.2 with a detailed explanation of the data collection method used in our current field practice.

Line 225-228: To the maximum extent feasible, qualified personnel will take over and rectify any instrument malfunctions found during routine inspection (on-site or remote) to ensure the accuracy and integrity of the observations. Data logger (e.g., CR6, Campbell Scientific, USA) recordings are first temporarily stored on the memory card before being routinely transmitted to our Data Processing Center by wireless transmission or on-site collection for processing, analysis, and archiving.

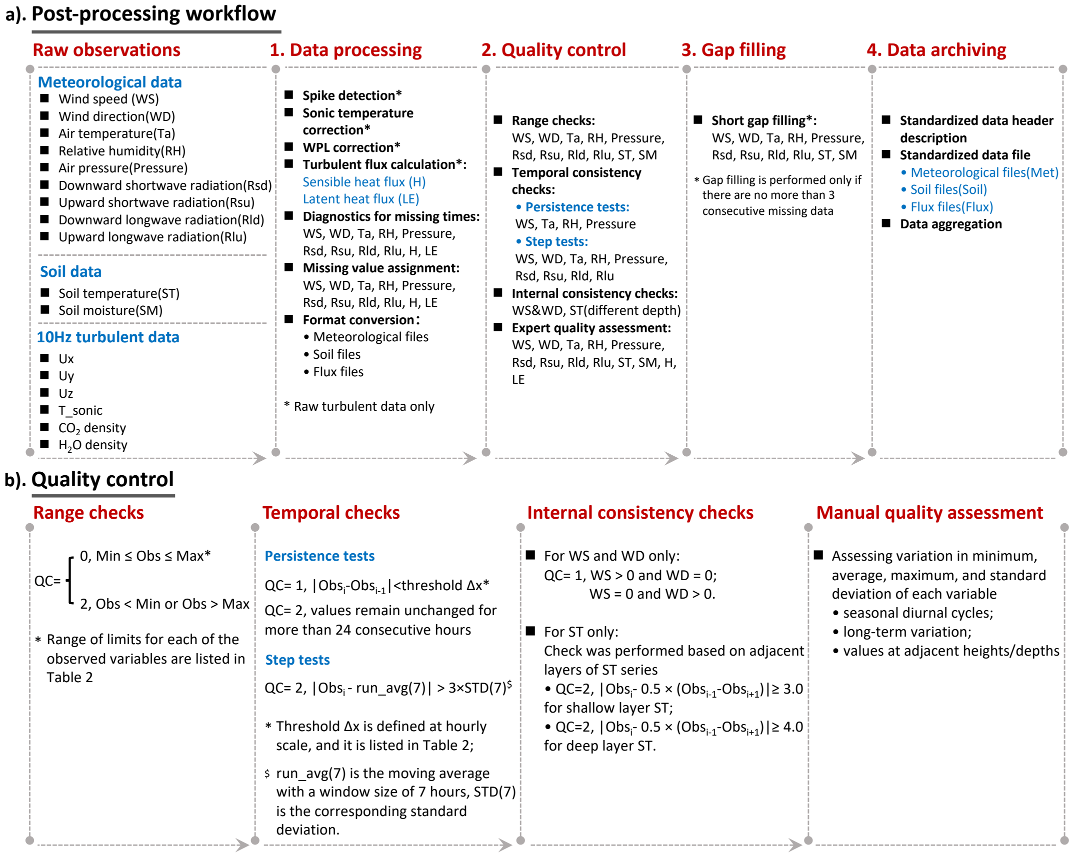
We do understand your concerns regarding the quality control procedures. We tried to introduce the data quality control procedures in a better way by referring to relevant literatures (e.g., Fiebrich et al., 2010; Rollenbeck et al., 2016; Cerlini et al., 2020) on data quality control and assessment of meteorological observations, the following revisions have been made to the original manuscript to address the Reviewer's concerns.

- **Redesigning the Figure 2a.** To clearly demonstrate the workflow of the quality control procedures employed in this study, and for a detailed description of the criteria used to resolve the concerns raised from the Reviewer, we redesign the Figure 2a.
- **Adding the Figure 2b.** Figure 2b was added to clearly illustrate the key formulas, and

quantitative metrics used in each procedure.

- **Expanding description.** The following text was added in the revised manuscript to provide additional information on how the data flows through the procedures.

  Line 262-266: The program first reads the data file for each station to be processed, checks were performed sequentially from left to right, and only when all the prescribed check procedures for each variable completed before moving on to the next one. The quality control procedures are arranged in a deliberate sequence, and ignore values flagged as errors by preceding checks in the sequence because the checks each have specific data requirements (e.g., running average and corresponding standard deviation should be calculated based on correct data).



Aside from interpolation data with short-term gaps, we did not take any specific steps to adjust the data series during the post-processing phase. Only short gaps were filled because the performance of the reconstruction method is strictly dependent on the length of the data gap, long-term gaps may greatly affect the reliable and accuracy of the observations. The main purpose of our data quality control is to identify and locate problems in the data series and flag them so that data users can base their research on reliable observations. To ensure the integrity, continuity, and reliable of the observations, we primarily implement targeted actions. For instance, we pay close attention to the operational status of the equipment throughout field observations and promptly address any issues that arise, such as instrument malfunctions or abnormal data. The following content has been added in the revised manuscript to describe the efforts we have made to ensure the integrity and accuracy of the data.

Line 222-226: In addition, we conduct monthly inspections of the operational status of all observational equipment (Ma et al., 2023b), as well as semi-annual on-site instrument maintenance for all stations, which includes instrument cleaning, checking the level of commissioned instruments, and checking instrument cables and connectors. To the maximum extent feasible, qualified personnel will take over and rectify any instrument malfunctions found during routine inspection (on-site or remote) to ensure the accuracy and integrity of the observations.

We think these revisions now provide a more comprehensive and detailed explanation of the data collection methods and quality control procedures, and we hope that the revision is acceptable and the Reviewer feel satisfied with this revision.

[1] Cerlini P B, Silvestri L, Saraceni M. Quality control and gap-filling methods applied to hourly temperature observations over central Italy[J]. Meteorological Applications, 2020, 27(3): e1913.

[2] Fiebrich C A, Morgan C R, McCombs A G, et al. Quality assurance procedures for mesoscale meteorological data[J]. Journal of Atmospheric and Oceanic Technology, 2010, 27(10): 1565-1582.

[3] Rollenbeck R, Trachte K, Bendix J. A new class of quality controls for micrometeorological data in complex tropical environments[J]. Journal of Atmospheric and Oceanic Technology, 2016, 33(1): 169-183.

3) While the approach for handling missing data through linear temporal interpolation is mentioned in 2.3.3 Gap filling, a discussion on the impact of these interpolations on the dataset's overall quality and potential biases introduced should be mentioned. Including statistical metrics to quantify the robustness of the gap-filled data could enhance the dataset's credibility.

Response: You have raised an important question. The accuracy of the linear temporal interpolation based gap filling technique was assessed based on filling artificially generated data gaps with different lengths. The performance of the gap filling method and the robustness of the gap-filled data for gap lengths of 1, 2, and 3 hours was evaluated. We randomly select 5,000 records of wind speed, wind direction, air temperature, relatively humidity, downward shortwave radiation, upward shortwave radiation, soil temperatures (at depths of 0.1 m and 0.8 m), sensible heat flux and latent heat flux, respectively. These variables were selected for assessment because they exhibit varying degrees of variability in the observed values over relatively short intervals. For example, wind speed and wind direction vary significantly over 1-3 hours, but soil temperature exhibits litter variation in the same time frame. This gives a good illustration of the impact of the interpolation scheme on the variables with varying degrees of variability. The mean error, mean absolute error, root mean square error, p value from the t test, and r square were calculated and provided in the new added Table 4. The following content has been added in the revised manuscript based on your suggestion.

Line 322-335: Series of random gaps (5,000 records for each variable) with different lengths were artificially created to quantify the overall performance of the gap filling method used and the robustness of the gap-filled data produced. The performance in filling gaps in wind speed, wind direction, air temperature, relatively humidity, downward/upward shortwave radiation, soil temperatures (at depths of 0.1 and 0.8 m), sensible heat flux, and latent heat flux for gap lengths of 1, 2, 3 hours were evaluated. These variables were selected for assessment because they exhibit varying degrees of variability in the observed values over relatively short intervals. For example, wind speed and wind direction vary significantly over 1-3 hours, whereas soil temperature changes less during that time. Table 4 shows the mean error, mean absolute error, root mean square error, p value from the t test, and r square. Results suggest that the gap length is one of the key factors influences the performance. This is demonstrated by the fact that the longer the gap length, the greater the error (ME, MAE, and RMSE) and the lower the coefficient of determination of the regression between the real values and the gap filled values, as well as the relatively larger errors of the variables with a higher degree of variability in a short period of time (wind direction, for example, is the most unreliable to interpolate). The interpolated upward shortwave radiation series with three hours gaps differs significantly ($p<0.05$) from the true values, for other variables evaluated, the difference is not significant. These findings suggest that the gap filling method used in this study can reasonably reconstruct the gaps within one to three hours.

Table 4. Mean error (ME), mean absolute error (MAE), root mean square error (RMSE), p-value from t test, and coefficient of determination calculated based on gap-filled artificially created missing data series and true values for gap lengths of 1, 2, and 3 hours, respectively.

	WS_1.5m	WD_1.5m	Ta_1.5m	RH_1.5m	Rsd	Rsu	ST_0.1m	ST_0.8m	H	LE
Mean Error, ME										
Gap_1	-0.011	1.439	0.001	0.077	0.501	0.180	-0.002	0.0005	1.514	0.371
Gap_2	0.011	1.139	-0.012	0.123	-1.331	0.588	-0.003	0.0007	2.749	-3.407
Gap_3	-0.009	2.331	-0.042	0.143	-1.358	2.158	0.006	0.0014	4.302	-1.433
Mean Absolute Error, MAE										
Gap_1	0.816	57.342	0.596	3.013	46.515	15.917	0.069	0.0099	20.639	12.177
Gap_2	0.929	63.531	0.894	4.023	67.617	20.714	0.206	0.0114	26.460	18.827
Gap_3	1.159	71.068	2.350	8.423	148.017	39.256	0.754	0.0162	46.514	21.081
Root Mean Square Error, RMSE										
Gap_1	1.128	90.107	0.879	4.775	99.612	34.119	0.108	0.0267	33.564	27.244
Gap_2	1.275	96.611	1.305	6.366	122.297	38.740	0.291	0.0281	40.256	212.681
Gap_3	1.569	100.235	3.000	11.861	227.072	61.860	0.994	0.0388	64.444	70.102
P value from t test										
Gap_1	0.798	0.506	0.994	0.885	0.941	0.921	0.991	0.9961	0.571	0.837
Gap_2	0.716	0.459	0.921	0.726	0.778	0.639	0.977	0.9926	0.158	0.453
Gap_3	0.728	0.068	0.659	0.624	0.721	0.034	0.948	0.9822	0.005	0.569
Coefficient of determination, R^2										
Gap_1	0.757	0.237	0.990	0.966	0.921	0.863	1.000	1.0	0.896	0.789
Gap_2	0.692	0.167	0.977	0.940	0.881	0.821	0.999	1.0	0.861	0.259
Gap_3	0.546	0.144	0.876	0.793	0.593	0.562	0.982	1.0	0.632	0.893

4) Section 3 on different datasets are well-detailed but the authors should add specific examples of data validation against external measurements or models, if available. This could include inter-comparison with satellite data, other observational networks, or model outputs to validate the spatial and temporal accuracy of the dataset.

Response: Thank you for this valuable suggestion. Unfortunately, in-situ field observations are extremely rare in such an extreme environment as the Tibetan Plateau. The closest automatic weather station operated by China Meteorological Administration (CMA) to those in our dataset are tens to hundreds of kilometers away. Additionally, the ASWs typically record only conventional meteorological elements, soil hydrothermal and turbulent fluxes are not available. Due to the difference in topography and subsurface features, observations between field stations and CMA operational AWSs may differ dramatically. Our stations included in the dataset represent currently the only observational network in the TP region that is able to conduct comprehensive observational measurements of land-atmosphere interactions. It is not suggested to validate the in-situ observations against model results, reanalysis products, or remote sensing products because of their poor accuracy in this area, which can be attributed to several factors like resolution. Besides, a great deal of work has been done on the assessment of model results and remote sensing products based on the observations provided in our dataset (e.g., Minola et al., 2024; Yao et al., 2023; Tong et al., 2023). Taking these factors into account, we therefore did not use external measurements or models to validate the in-situ observations. But your suggestion has motivated us to investigate the possibility of using multi-source data from satellite products, model outputs, and reanalysis products to filling the gap data in the field observations in future work. This is a better approach than the linear interpolation method we used in this study, and it can be applied to a wider range of scenarios. We hope you could understand our concern. Thank you very much.

[1] Minola L, Zhang G, Ou T, et al. Climatology of near-surface wind speed from observational, reanalysis and high-resolution regional climate model data over the Tibetan Plateau[J]. Climate Dynamics, 2024, 62(2): 933-953.

[2] Yao T, Lu H, Yu Q, et al. Uncertainties of three high-resolution actual evapotranspiration products across China: Comparisons and applications[J]. Atmospheric Research, 2023, 286: 106682.

[3] Tong L, He T, Ma Y, et al. Evaluation and intercomparison of multiple satellite-derived and reanalysis downward shortwave radiation products in China[J]. International Journal of Digital Earth, 2023, 16(1): 1853-1884.