Reply to Referee #1' s comments

Title: A flux tower site attribute dataset intended for land surface modeling
No.: essd-2024-77

We would like to thank you for your careful reading, helpful comments, and constructive suggestions, which have significantly improved the presentation of our manuscript.

All comments are addressed on a point-by-point basis below. The comments are laid out below in italicized font and specific concerns are numbered. Our response is given in normal font. The list of all related changes is given in blue text.

Comment 1: *While I understand the focus of the paper is on presenting the new dataset, I believe a short description of the treatment of water and energy fluxes at the land surface in the model used here (CoLM) would be very helpful. Since the model results are an important part of the manuscript, this would help the readers in interpreting the improvement shown due to the improved data sources.*

Response1: Thank you for your careful evaluation of this manuscript. Following your suggestions, we have added a description of how water and energy fluxes are treated at the land surface in the model (CoLM) used here. The comparison before and after modification is as follows:

Origin:

"The impact of collected attributes on carbon, water, and energy fluxes is assessed through single-point simulations using the Common Land Model (CoLM) (Dai et al., 2003). We used its latest version, CoLM202X (https://github.com/CoLM-SYSU/CoLM202X/tree/master, last access: 21 November 2023)."

Revised:

"The impact of collected attributes on carbon, water, and energy fluxes is assessed through single-point simulations using the latest version of the Common Land Model (Dai et al., 2003) (CoLM202X, https://github.com/CoLM-SYSU/CoLM202X/tree/master, last access: 21 November 2023). CoLM202X incorporates processes related to biogeophysics, biogeochemistry, ecological dynamics and human activities, and also offers optional processes and schemes which can be customized by the user. In our experiments, vegetation is modeled using a set of time-invariant parameters (optical properties: leaf optical properties; morphological properties: canopy height, vegetation root depth and profile, leaf size and angle distributions; and physiological properties). The dynamic vegetation module is turned off and the time-variant LAI and stem area index (SAI) values are prescribed from the reprocessed MODIS LAI data (Lin et al., 2023; Yuan et al., 2011). The two-big-leaf model (Dai et al., 2004) is employed to calculate processes such as radiative transfer (Yuan et al., 2017), photosynthesis (Collatz et al., 1992; Farquhar et al., 1980), and stomatal conductance (Ball et al., 1987). Surface turbulent exchange is simulated using similarity theory (Brutsaert, 1982; Zeng and Dickinson, 1998). Total evapotranspiration includes evaporation from stems, leaves, and the ground, as well as vegetation transpiration. Surface and subsurface runoff consider factors such as terrain, groundwater level, precipitation, and infiltration rate. Additionally, the model accounts for processes including precipitation phase and intensity, canopy interception, vertical movement of water in snow and soil, and snow compaction (Dai et al., 2003).

Comment 2: *It seems that when vegetation or soil properties data are not available for the sites, the authors use the "default data" instead in order to fill the missing data. I believe the authors should provide some information about potential inhomogeneities in the final dataset resulting from this choice. If I understand correctly, the default data used to fill missing data here are those also shown in Figure 3 as comparison. I would recommend the authors use the sites for which both data sources are available to provide some quantification of the difference between new in-situ data and default data, thus quantifying the resulting inhomogeneities in the final data product. Some of this information may already been shown in Figure 3, but I recommend the authors quantify this explicitly as it is an important feature of the data produced here.*

Response2: We completely agree with your suggestion. It is necessary to account for potential inhomogeneities in the final dataset resulting from data filling. Figure 3 shows the discrepancies between site data and default data to demonstrate the importance of site data. After careful consideration, we believe that Sect. 3.2 describes the flux tower site attribute dataset. Therefore, the quantification of discrepancies between site data and filled data has been added to Sect. 3.2, illustrating the inhomogeneities in the final dataset due to data filling. The added information is as follows.

One point of clarification is that the default data and the data used to fill the missing data (Filled data) are not exactly the same. The default data is the data commonly used in the LSMs. Filled data is used to fill the missing site-observed attributes in the final dataset. The details are shown in the table below:

Attribute	Default data	Filled data	Consistency
PCT_PFT	IGBP classification	PFT_{local} maps (Harper et al., 2023)	Inconsistent
LAI	Reprocessed MODIS 6.1 LAI (Lin et al., 2023)		Consistent
Canopy height	Lookup table from CoLM	PLUMBER2 (Ukkola et al., 2022)	Inconsistent
Soil texture	GSDE soil dataset (Shangguan et al., 2014)		Consistent

Add:

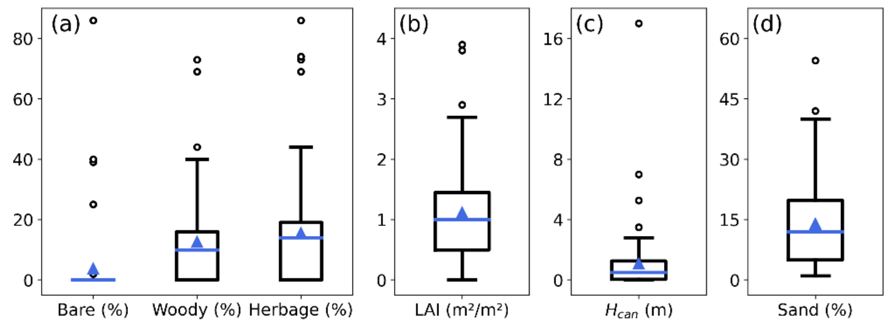

Figure 4. Quantification of discrepancies between site data and filled data of (a) PCT_PFT, (b) maximum LAI, (c) canopy height, and (d) the percentage of sand (at all sites for which both data sources are available). The 16 PFTs were divided into three main categories (bare soil, woody, and herbage) to be quantified separately,

Figure 4 quantifies the differences between site data and filled data at all sites for which both data sources are available, illustrating the inhomogeneities in the final dataset due to data filling. The differences in vegetation cover (including bare soil, woody, and herbaceous vegetation) generally fall within 20%, with a minority of sites exceeding 40%. The mean and median LAI differences are about 1 m2/m2. Canopy height deviations are primarily within 2 m, although a few sites exceed 4 m. Differences in sand content typically remain within 30%, with both mean and median differences below 15%. The quantification indicates that the filled data are relatively reliable across most sites.

Comment 3 (L20): *Which model? Or do you mean "models"?*

Response3: Thank you for pointing this out. I'm sorry for the ambiguity. What we are trying to express here is the data commonly used by LSMs. Therefore, it should be 'models' instead of 'the model'. And we changed the wording to express it more clearly. The comparison before and after modification is as follows:

Origin:

"the attribute data observed at the site and the defaults of the model"

Revised:

"the attribute data observed at the site and commonly used by LSMs"

Comment 4 (L369): *"Using CoLM at 36 sites": Is there a specific reason the model was run at 36 sites out of 90 and not at all? In particular, at line 378 it is stated that all selected sites used for the modelling experiment have fairly large LAI values, but a large sensitivity to LAI is expected at sites characterized by lower LAI.*

Response4: Thank you for your question. In our opinion, the basis for simulation differences lies in differences in attribute values, with greater disparities in attributes values typically leading to more pronounced differences in model results. Therefore, we selected 10 sites with the largest differences between site data and default data for LAI, tree height, and soil texture, respectively. Specifically, for vegetation cover, sites with IGBP types that are a combination of tree and grasses (OSH, WSA, SAV) were chosen, resulting in six available sites. Thus, a total of 36 sites were used for modeling assessment.

In line 377, we note that variations in unit LAI elicit more substantial fluctuations in fluxes at lower LAI values (usually less than 2 m2/m2), indicating greater sensitivity of fluxes to LAI. Consider that the modeling assessment of attribute data has focused primarily on the magnitude of the impact of the attribute data and has not addressed specialized sensitivity analyses. We believe that this passage may cause some misunderstanding. Therefore, after careful consideration, we removed this part of the argument from the manuscript.

Delete:

Notably, unit LAI variations elicit more substantial fluctuations in fluxes at lower LAI values (usually less than 2 m2/m2), according to Launiainen et al. (2016). In light of that, all of the sites we chose have LAI values greater than 2 m2/m2, except US-GLE, the impact of LAI obtained here are relatively minor.

Comment 5 (L40): *Maybe "for testing and validating LSMs"?*

Response5: Thanks for your suggestion. We have revised it according to your suggestion.

Origin:

"flux tower data was not originally designed for LSMs"

Revised:

"flux tower data was not originally designed for testing and validating LSMs"

Comment 6 (L41): *It suffers -> these datasets suffer*

Response 6: Thanks for your suggestion. We have revised it according to your suggestion.

Origin:

"it suffers from poor data quality and a deficiency of attribute data"

Revised:

"these datasets suffer from poor data quality and a deficiency of attribute data."

Comment 7 (Figure 7 caption): *Do you mean "Precip" in the legend?*

Response 7: Thank you for your careful examination. We have revised it.

Origin:

"Pricip."

Revised:

"Precip."

Comment 8 (L117): *Please clarify sentence.*

Response 8: Thanks for the suggestion. We have described it more clearly.

Origin:

"The PLUMBER2 dataset got 170 sites by screening meteorological data."

Revised:

"The PLUMBER2 dataset got 170 sites by screening five key meteorological variables that have the largest influence on LSM simulations: incoming shortwave radiation, precipitation, air temperature, air humidity, and wind speed."

Comment 9 (L142): *because -> since;*

Response 9: Thank you for your correction. We have revised it.

Origin:

"Because FVC"

Revised:

"Since FVC"

Comment 10 (L142): *"they are close numerically" – could you be more precise and state how similar these data sources are?*

Response 10: Thank you for your suggestion. Here, in the absence of a description of fractional vegetation cover, the percentage of vegetation flux footprint contribution or dense forest canopy basal area is used as a proxy. These values are considered the closest numerical alternatives. Unfortunately, we don't have a reliable method or any citations to provide a precise evaluation.

However, it is clear that the fractional vegetation cover directly determines the percentage of vegetation flux footprint and dense forest canopy basal area. Theoretically, fractional vegetation cover equals the percentage of vegetation flux footprint under windless conditions; Basal area is defined as the total cross-sectional area of all stems in a stand. If the canopy width and stem cross-sectional area maintain a fixed ratio, fractional vegetation cover is equal to the percentage of dense canopy basal area.

Comment 11 (L144): *what are "site pictures"? satellite imagery? Could you please specify and indicate the data source?*

Response 11: Thank you for your question. In this context, 'site picture' refers to photographs taken at the site and does not involve satellite images. Their sources are the flux regional network and related publications. The specific sites where pictures were used for judgment and the sources are described in Table S1. We have clarified "site picture" based on your suggestion.

Origin:

"we referred to site pictures to make a judgment"

Revised:

"we referred to site pictures (photographs taken at the site) to make a judgment"

Comment 12 (L209, L311, L372, L392, L393, L399): *"between RUNS using…"; "but->however"; Remove "and"; A previous study found / discovered / stated ….; This study, however, …; Remove "And".*

Response 12: Thank you for your correction. We've revised these words and phrases based on your suggestion.

Origin:

"between using"; "But the effects of vegetation"; "And among the four attributes"; "A previous study viewed that"; "Its study, however,"; "And the attribute data we collected"

Revised:

"between runs using"; "However the effects of vegetation"; "Among the four attributes"; "A previous study stated that"; "This study, however,"; "the attribute data we collected"

Comment 13 (Eq. (1)): *Is n = 365 here?*

Response 13: Thank you for your question. 'n' stands for the number of days in different months, with a value of 28, 30 or 31, depending on the number of days in each month. Based on your comments, we have clarified Eq. (1). The comparison before and after modification is as follows:

Origin:

$$MD\ \% = \begin{cases} \dfrac{\left|\frac{1}{n}\sum_{i=1}^{n}(Mod_{site,i} - Mod_{default,i})\right|}{\frac{1}{365}\sum_{j=1}^{365} Obs_j}, & for\ LE, H, Rn, SWup, GPP, and\ Ustar \\ \dfrac{\left|\frac{1}{n}\sum_{i=1}^{n}(Mod_{site,i} - Mod_{default,i})\right|}{\frac{1}{365}\sum_{j=1}^{365} Mod_{default,j}}, & for\ SWC\ and\ TR \end{cases} \quad (1)$$

Revised:

$$MD\ \% = \begin{cases} \dfrac{\left|\frac{1}{n}\sum_{i=1}^{n}(Mod_{site,i} - Mod_{default,i})\right|}{\frac{1}{365}\sum_{j=1}^{365} Obs_j}, & for\ LE, H, Rn, SWup, GPP, and\ Ustar \\ \dfrac{\left|\frac{1}{n}\sum_{i=1}^{n}(Mod_{site,i} - Mod_{default,i})\right|}{\frac{1}{365}\sum_{j=1}^{365} Mod_{default,j}}, & for\ SWC\ and\ TR \end{cases} \quad n = days\ of\ month \quad (1)$$

We appreciate your warm work earnestly, and hope the correction made will meet with approval. These comments and suggestions have significantly improved the quality of our manuscript.

As you mentioned, the manuscript still contains many unclear expressions. Other reviewers have also pointed out similar issues. We are very grateful for the partial corrections you have already provided. We will thoroughly check for language problems in the manuscript before the submission after the discussion phase concludes.

References:

Ball, J. T., Woodrow, I. E., and Berry, J. A.: A Model Predicting Stomatal Conductance and its Contribution to the Control of Photosynthesis under Different Environmental Conditions, in: Progress in Photosynthesis Research, edited by: Biggins, J., Springer Netherlands, Dordrecht, 221–224, https://doi.org/10.1007/978-94-017-0519-6_48, 1987.

Brutsaert, W.: Energy Budget and Related Methods, in: Evaporation into the Atmosphere, Springer

Netherlands, Dordrecht, 209–230, https://doi.org/10.1007/978-94-017-1497-6_10, 1982.

Collatz, G., Ribas-Carbo, M., and Berry, J.: Coupled Photosynthesis-Stomatal Conductance Model for Leaves of C4 Plants, Functional Plant Biol., 19, 519, https://doi.org/10.1071/PP9920519, 1992.

Dai, Y., Zeng, X., Dickinson, R. E., Baker, I., Bonan, G. B., Bosilovich, M. G., Denning, A. S., Dirmeyer, P. A., Houser, P. R., Niu, G., Oleson, K. W., Schlosser, C. A., and Yang, Z.-L.: The Common Land Model, Bulletin of the American Meteorological Society, 84, 1013–1024, https://doi.org/10.1175/BAMS-84-8-1013, 2003.

Dai, Y., Dickinson, R. E., and Wang, Y.-P.: A Two-Big-Leaf Model for Canopy Temperature, Photosynthesis, and Stomatal Conductance, J. Climate, 17, 2281–2299, https://doi.org/10.1175/1520-0442(2004)017<2281:ATMFCT>2.0.CO;2, 2004.

Farquhar, G. D., Von Caemmerer, S., and Berry, J. A.: A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species, Planta, 149, 78–90, https://doi.org/10.1007/BF00386231, 1980.

Harper, K. L., Lamarche, C., Hartley, A., Peylin, P., Ottlé, C., Bastrikov, V., San Martín, R., Bohnenstengel, S. I., Kirches, G., Boettcher, M., Shevchuk, R., Brockmann, C., and Defourny, P.: A 29-year time series of annual 300 m resolution plant-functional-type maps for climate models, Earth Syst. Sci. Data, 15, 1465–1499, https://doi.org/10.5194/essd-15-1465-2023, 2023.

Lin, W., Yuan, H., Dong, W., Zhang, S., Liu, S., Wei, N., Lu, X., Wei, Z., Hu, Y., and Dai, Y.: Reprocessed MODIS Version 6.1 Leaf Area Index Dataset and Its Evaluation for Land Surface and Climate Modeling, Remote Sensing, 15, 1780, https://doi.org/10.3390/rs15071780, 2023.

Shangguan, W., Dai, Y., Duan, Q., Liu, B., and Yuan, H.: A global soil data set for earth system modeling, J. Adv. Model. Earth Syst., 6, 249–263, https://doi.org/10.1002/2013MS000293, 2014.

Ukkola, A. M., Abramowitz, G., and De Kauwe, M. G.: A flux tower dataset tailored for land model evaluation, Earth Syst. Sci. Data, 14, 449–461, https://doi.org/10.5194/essd-14-449-2022, 2022.

Yuan, H., Dai, Y., Xiao, Z., Ji, D., and Shangguan, W.: Reprocessing the MODIS Leaf Area Index products for land surface and climate modelling, Remote Sensing of Environment, 115, 1171–1187, https://doi.org/10.1016/j.rse.2011.01.001, 2011.

Yuan, H., Dai, Y., Dickinson, R. E., Pinty, B., Shangguan, W., Zhang, S., Wang, L., and Zhu, S.: Reexamination and further development of two-stream canopy radiative transfer models for global land modeling, J Adv Model Earth Syst, 9, 113–129, https://doi.org/10.1002/2016MS000773, 2017.

Zeng, X. and Dickinson, R. E.: Effect of Surface Sublayer on Surface Skin Temperature and Fluxes, J. Climate, 11, 537–550, https://doi.org/10.1175/1520-0442(1998)011<0537:EOSSOS>2.0.CO;2, 1998.