

A flux tower site attribute dataset intended for land surface modeling

By Jiahao Shi et al.

This paper describes a dataset constructed by adding additional metadata variables to a set of existing data from flux tower observations. The additional variables added by the authors to this dataset include local vegetation properties, topographic parameters, soil properties, and wind measurement height. The purpose of this study is to provide a dataset to be used for development and validation of land surface models (LSMs). To evaluate the implication of using these updated variables with respect to default values commonly used in LSMs, the authors perform some experiments using a LSM. This analysis shows that these input variables can have significant impacts on the simulated fluxes of water and energy between land and atmosphere.

I believe the effort of enhancing existing datasets by adding additional variables and site characteristics as done here is certainly worthwhile, and therefore believe the manuscript would be of interest for the readership of the journal. The paper is overall concise and suitably organized to present the data and an application of their use in LSMs. However, the language used in the paper needs improvement. Several sentences throughout the paper should be rephrased in my opinion, as they make the text unclear and not very precise. I list several of these instances in the specific comments below, but I would recommend a thorough check of the language used in the paper. I believe the authors should address the comments below before the manuscript is considered for publication.

Comments

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.

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.

Minor comments and (non-exhaustive) suggestions on the language used in the paper

L20: Which model? Or do you mean “models”?

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.

L40: maybe “for testing and validating LSMs”?

L41: it suffers -> these datasets suffer

Figure 7 caption: Do you mean “Precip” in the legend?

L117: please clarify sentence.

L142: because -> since

L142: “they are close numerically” – could you be more precise and state how similar these data sources are?

L144: what are “site pictures”? satellite imagery? Could you please specify and indicate the data source?

L209 “between RUNS using...”

L311: but -> however

L372: Remove “And”

L392: A previous study found / discovered / stated

L393: This study, however, ...

L399: Remove “And”

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