

Review of Energy-conservation datasets of global land surface radiation and heat fluxes from 2000-2020 generated by CoSEB

Summary and recommendation- In this paper, the authors apply a model of Coordinated estimates of land surface energy balance components (CoSEB) to generate estimates of surface radiation and heat fluxes from 2000 to 2020. An advantage of the CoSEB based approach is that estimates of radiation and heat are in “harmony” as opposed to generating independent estimates of each. The authors compare their estimates against observations from eddy covariance sites, other individual estimates and other individual observations. The paper is generally well written, and the results are presented clearly. However, I had several questions about the CoSEB framework itself and also the validations applied here in the manuscript. Hence I recommend major revisions. I have presented major comments and specific comments below.

Major comments-

1. **Explanation of updates to the CoSEB framework-** While reading the manuscript I realized that it is not only a paper that applies the existing CoSEB framework that is already published but also updates this framework to estimate to estimate radiation (previously this model estimated only land surface energy components and not short wave and long wave radiation). Therefore, authors need to discuss the effect of the addition of additional predicted variables on the equations and the results of the random forest. In particular, can the authors discuss which of the predictors were found to be the most important and also discuss how this differed with their previous publication? Also, can authors discuss generic details such as how many splits were generated by the random forest before and after the updates. Authors should also discuss the directionality of effects of different predictor variables based on the revised random forest.
2. **Multi-collinearity amongst predictor variables-** Authors should also discuss how multi-collinearity is handled amongst predictor variables given the large number of predictors. As far as I understand, random forests do not explicitly deal with multi collinearity unlike a PCA based approach for example. This can affect variable importance significantly. I would suggest authors explore this in detail.
3. **Effect of autocorrelation-** Given the temporal nature of several predictor variables, can authors confirm that autocorrelation does not exist or is minimized in their framework? What tests were performed to check for this? In particular I would recommend authors add lagged variables to the model to make sure that this is not the case. I believe several models constructed for earth system variables tend to ignore aspects such as autocorrelation and therefore this is an important point to address.
4. **Effect of downscaling ERA5- Land datasets-** The authors note on lines 195-197 that the ERA 5 land datasets used here have been downscaled from a resolution of ~9 kms to

~500m. This is a significant level of downscaling performed using a rather simple cubic convolution method. There are several variables related to the land cover (such as the LAI for example) that are used as predictor variables in the author's framework. Can the authors address the uncertainty caused by such large downscaling between scales on their results? On the one hand, based on the results, it seems that the model has produced reliable results compared to observations and other datasets even after such large downscaling. Is it that the land cover related variables do not play an important role in the predictions?

5. **In sample vs out of sample testing-** While the authors present significant comparisons with observations and other datasets to validate their model (e.g. Figure 3, Figure 4 and Figure 5), it seems the authors have not checked for overfitting of their approach by splitting the dataset into a training vs testing dataset. This is especially important since as mentioned in Major comment 1., the CoSEB framework itself has been updated. Authors should address this in detail. In fact, looking at Figure 3, it seems that the R squared values for G and H are on the lower side. I am curious as to what the values look like when out of sample testing is conducted?

Specific comments-

1. Abstract lines 31-36- The RMSEs presented here do not make any sense at this point since the reader has no sense of scale of values to expect. I recommend authors report the R squared values here instead. Also make sure to report whether the R squared is based on pooled data or just the testing data (See Major comment 5)
2. Introduction lines 74-75- Can the authors differentiate the citations between those for physical vs those for statistical methods.
3. Introduction line 92- "impending" is an awkward word here. I would just say "It was imperative".
4. Data lines 131-132- Why could a simple interpolation not be applied for missing half hourly data? Is the data extremely sensitive to time? Some clarification is needed here.
5. Data lines 138-139- Can the authors clarify why this criteria was applied for screening outliers?
6. Mainstream datasets/products for inter comparison- I was curious as to why the authors so not compare their estimates with heat and radiation estimates from popular earth system modelling systems such as CESM and CTSM (<https://www.cesm.ucar.edu/>). In fact, if the authors approach can produce estimates similar to earth system models, this would be a huge benefit to the community (since these models are laborious to run)
7. Methods lines 243-244- Once again the usage of RMSEs here does not make much sense. Can the authors just report the R squared values instead.

8. Methods lines 269-270- Just to confirm, the RF based uncoordinated models are models where only individual variables are estimated rather than the simultaneous calculation of several variables? This should be clarified.
9. Results Lines 306-309- I was curious looking at Figure 4 whether there were correlations or relationships between the EIR or RIR values and any of the other predictor variables? Is the shape of that distribution affected by any particular variables?
10. Results Lines 311-312- Can the authors clarify the differences between site-based validation vs sample-based validation?
11. Results lines 381-382- Once again, the RMSE values don't make a lot of sense here. Authors should report the R squared values instead.
12. Section 4.2- When discussing the differences between the CoSEB model estimates vs other estimates, can authors also describe why the differences occur? A detailed discussion is not warranted here. Rather, I was interested in the author's perspective as to why the author's approach produces some differences over existing approaches.