

## ***Interactive comment on “Surface global and diffuse solar radiation over China acquired from geostationary Multi-functional Transport Satellite data” by Hou Jiang et al.***

### **Anonymous Referee #1**

Received and published: 6 December 2019

#### Summary and comments

Jiang et al. produced total and diffuse surface radiation over China using the geostationary satellite data and the deep learning network. I find the dataset of general interest. I have some concerns related to this manuscript.

(1) the title suggests it is surface global solar radiation over China. I think “global” here should represent the whole earth. The title needs to be changed.

(2) scientific relevance: I do not think the authors provide enough details about how the new technique can address previous research problems/scientific questions. For example, what do the authors mean “no consideration of spatial collocation of surface

C1

radiation”, and how the proposed deep learning algorithm addresses this consideration remains unclear. What spatial patterns have been extracted from satellite images and how they are linked to target hourly radiation values are also not clear. The introduction gives an impression that the dataset is probably “over-produced” with no clear clue that how these datasets will be used specifically. Although they first paragraph in the introduction mentioned some possible applications of these datasets, further digging of these literatures would tell you the produced dataset here would be useless. I would suggest the authors further clearly identify, for example, how diffuse radiation are being used to forecast crop yield (if there is a crop model uses it, please name it), and how diffuse radiation are being used to simulate carbon dynamics (if there are Earth system models requiring these datasets, please list them in the paper) or any other specific applications that use this type of datasets either partly or over the whole part of China (rather than talk them general). Based on current descriptions, I think the produced datasets would be useless.

(3) Methods and Comparisons with other products: I think throughout the manuscript, there is no scientific explanation that why we should correlate satellite images of five bands with total/diffuse solar radiation (I think the author also need to provide a definition in this manuscript what exactly data they are providing). I believe what you mean solar radiation here should be the integration of radiation over the whole wavelength rather just a few wavelengths. Although the authors use the deep learning algorithm, it is necessary to explain what the mechanisms behind this correlation. Particularly, as the authors mentioned previously that spatial patterns are being extracted and correlated with target data from point locations, the authors should also explain what patterns are being used. Without sufficient explanations, the readers may concern about what are the error estimates and sources which the authors do not provide at all. For example, how reliable can we use the solar radiation estimates under cloudy conditions? Can the authors provide a quality flag for these areas and the confidence interval for us to use the dataset for regions under cloudy conditions? Does the authors separate comparisons/validation in accuracy for point locations between clear-sky and

C2

cloudy conditions? What is the accuracy level under clear-sky conditions and what is the accuracy level under cloudy conditions? IF the comparison in accuracy between this produced product and the other product under clear-sky conditions is the same or similar, what are the advantages of this produced product compared to previous ones? In addition, for the deep learning part, I do not understand why we use 16\*16 pixels. How you get the size determined? Plus, do the radiation measurements from point locations are across all these 16 pixels or just belong to only one of them? I have the impression that because the authors would like to use deep learning, they need to have an image used as input. So, this is a point to area comparison? If so, I would like to see the authors providing error estimates related to this misrepresentation of points as areas for the produced datasets.

(4) My last concern is that I think the manuscript has duplication issues: some of the figures have been seen previously in your other manuscripts. for example, figure 1, 2, etc (in your recent publications in Renewable and Sustainable Energy Reviews, Volume 114, October 2019, 109327). Also, I would like to mention here that previously you used the ResnetTL, and in this manuscript, it seems that you use a different network structure. So, what is the difference between them? what are the improvements?

The authors also need to pay attention to the grammar of the manuscript and the language needs to be further edited.

---

Interactive comment on Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2019-209>, 2019.