Response to Referee #2

We would like to thank the reviewer for the comments and suggestions, which are all valuable and very helpful for improving our paper. We have made revisions and a point-to-point response is present in the following.

Comments:

Hou et al. present a study about "surface global and diffuse solar radiation over China acquired from geostationary Multi-functional Transport Satellite data". The following questions should be satisfactorily answered before consideration for publication:

1. The topic is not innovative enough, which has been done by many researchers, for example Tang et al (2016) has published an article named "Retrieving high-resolution surface solar radiation with cloud parameters derived by combining MODIS and MTSAT data". The input data in your model, the spatial and temporal resolution of your output GSR values are similar to that in Tang' study. Only using a artificial intelligence model could not be an innovation idea. I would strongly advise the author(s) of this paper to rewrite their introduction section to give more explanation of the research background. A very general sentence is not enough to demonstrate the research significance.

Response:

Thank you for your advice.

Yes, the topic of radiation is not new and has been done by many researchers in view of its general interests in many fields and supports to various basic models (e.g., JULES, FöBAAR, YIBs, SWAP) and applications (e.g., modelling radiation-use efficiency of wheat, early yield assessment of soybean, wheat and sunflower), as described in the Introduction of our revised manuscript.

In theory, previous studies (including Tang et al., 2016) are based on an independent pixel approximation which assumes a plane-parallel horizontally homogeneous cloud. Thus, their radiation retrievals are pixel-based (from point to point), in other words, only (multi-band) satellite signals or multi-source information corresponding to the specific ground location is used for surface radiation estimation. However, it has been demonstrated that multiple reflections and scattering events off the sides of clouds or on the surface would lead to significant horizontal photon transport, so that adjacent pixels within a certain spatial extent also influence the measured radiation at a specific location on the ground. To the best of our knowledge, the traditional pixel-based retrieval cannot handle such spatial adjacent effects of surface radiation and none of operational methods has been proposed currently. Thus our innovation idea is to deal with image blocks of 16*16 pixels through CNN blocks to infer the surface radiation at the location corresponding to the central point of the input image block. We expect the CNN blocks can approximate the spatial adjacent effects thereby improving the final accuracy of radiation estimation. The CNN blocks allow for identical treatment of adjacent satellite pixels, and can be stacked to construct deep residual structure to extract hierarchical features from low-level details (e.g., geometric shapes, sizes, orientations, edges and distribution) to high-level comprehensive abstract representations (e.g., intrinsic physical and optical properties of mixed clouds). It is believed that such hierarchical architecture of spatial features can fully expose the scattering effects, absorption effects as well as their interactions in the atmosphere, thus can be considered as substitutes for various input parameters representing atmospheric state in radiative transfer models. Meanwhile, the MLP can be utilized to link extracted features of CNN and additional auxiliary information to target measurements of hourly surface radiation through implicit non-linear expressions. More explanations have been added into the Introduction of revised manuscript as "These algorithms mainly include two categories: constructing empirical mathematical relationships between top of atmosphere (TOA) and surface radiative fluxes (Linares-Rodriguez et al. 2013; Lu et al. 2011) and driving complex radiative transfer models utilizing satellite-derived atmospheric parameters (Greuell et al. 2013; Huang et al. 2011). These methods are in theory based on an independent pixel approximation which assumes a plane-parallel horizontally homogeneous cloud. Thus, surface radiation retrievals from satellite imagers are pixel-based (from point to point), in other words, only (multi-band) satellite signals corresponding to the specific ground location are used for surface radiation estimation. However, in reality this idealized situation does not always exist, or even is uncommon. For example, in the presence of broken clouds, multiple reflections and scattering events off the sides of clouds or on the surface would lead to significant horizontal photon transport (Madhavan et al. 2017; Oreopoulos et al. 2000; Schewski and Macke 2003), which makes significant differences when the spatial resolution increases to several kilometres where the surface radiation of an individual footprint under inhomogeneous clouds is relevant to multiple adjacent satellite pixels (Huang et al. 2019). In fact, large biases and uncertainties occur frequently under broken clouds when comparing current high-resolution surface radiation products to quality-controlled ground observations (Deneke et al. 2009; Huang et al. 2016).

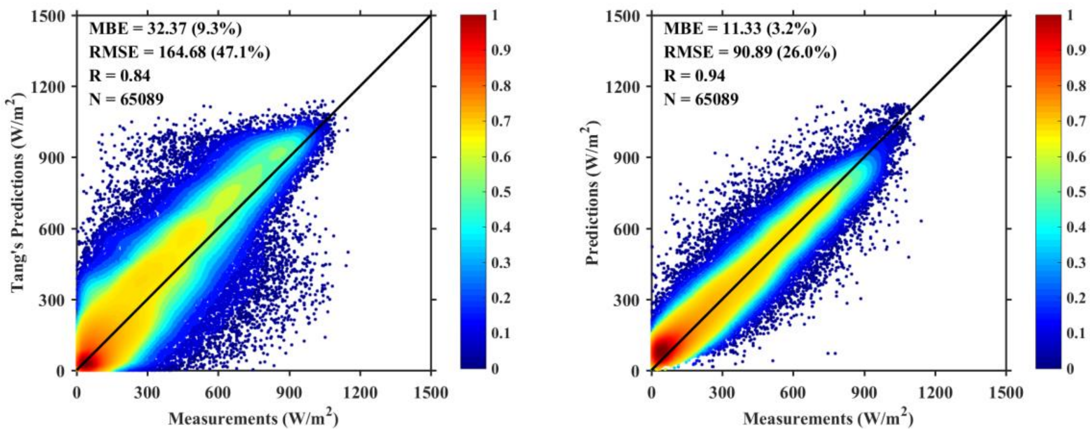
Therefore, it seems that area-to-point retrievals are the optimal solutions. From this point of view, a practical effort has been made in our previous works, where a hybrid deep network mainly consisting of convolutional neural network (CNN) blocks and multi-layer perceptron (MLP) is built to retrieve hourly GSR/DIF from geostationary satellite data (Jiang et al. 2019). The CNN blocks takes image blocks as inputs thereby allowing for identical treatment of adjacent satellite pixels, and are further stacked to construct deep residual structure to extract hierarchical features from low-level details (e.g., geometric shapes, sizes, orientations, edges and distribution) to high-level comprehensive abstract representations (e.g., intrinsic physical and optical properties of mixed clouds). It is believed that such hierarchical architecture of spatial features can fully expose the scattering effects, absorption effects as well as their interactions in the atmosphere, thus can be considered as substitutes for various input parameters representing atmospheric state in radiative transfer models. The MLP is utilized to link extracted features of CNN and additional auxiliary information defining the state in time and space to target measurements of hourly surface radiation through implicit non-linear expressions, whose parameters are learnt from pre-prepared training samples in supervised manner. The deep network is demonstrated to be effective in handling spatial adjacent effects and simulating complicated radiative transfer processes, and successful in achieving superior accuracy of GSR estimates."

Here, we would like to point out the differences between our works and Tang's:

1) The inputs of our algorithm involve only the visible band of MTSAT data along with additional attributes including the local time (month, day and hour) and location (longitude, latitude and altitude) without other cloud parameters.

2) The MTSAT data are input into the model as a whole to handle the spatial adjacent effects of surface radiation.

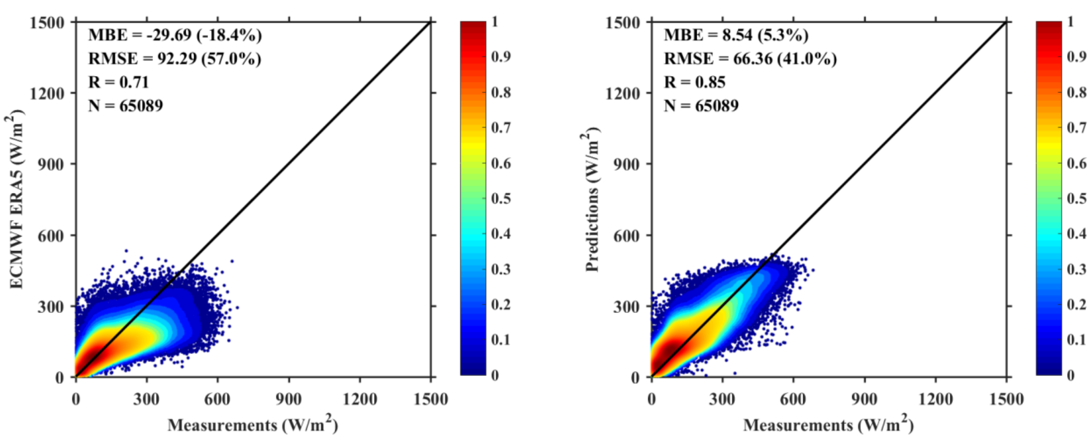
3) Much higher accuracy of radiation data has been achieved in our work as shown in Figure R1, which further demonstrate the importance to take spatial adjacent effects into consideration during satellite-based radiation retrievals.



(a) (b)

Figure R1: The density plots of predicted hourly global solar radiation versus measured values for (a) Tang's data and (b) our datasets in 2007.

4) The diffuse radiation is also estimated and provided in our datasets, which is comparable to the widely-used ECMWF-ERA5 data as shown in R2.



(a) (b)

Figure R2: The density plots of predicted hourly diffuse solar radiation versus measured values for (a) ECMWF ERA5 and (b) our datasets in 2007.

<u>Comments:</u>

2. In Tang's study (Ditto), Tang et al., innovatively correlated the cloud optical properties to the satellite data that were used in your study. The main radiation dumping processes including Rayleigh scattering, aerosol extinction, ozone absorption, water vapor

absorption, permanent gas absorption, and cloud extinction are considered in Tang's study. What is the scientific correlation between satellite signals in five bands with GSR and DIF values in your study? Author should explain the mechanisms in your model.

Response:

Thank you for your advice. I am so sorry for the ambiguous explanation of our model and its mechanisms.

Different from Tang's work, we directly correlate the ground measured GSR or DIF values to the visible channel of MTSAT satellite data, and we expect the deep network to build their relationships through hierarchical feature representations of CNN and continuous mapping function approximation of MLP by learning from pre-prepared training samples in a supervised manner. The radiation dumping processes such as Rayleigh scattering, aerosol extinction, ozone absorption, water vapor absorption, permanent gas absorption, and cloud extinction are avoided as it is believed that hierarchical architecture of spatial features can fully expose above-mentioned effects, thus can be considered as substitutes for various input parameters representing atmospheric state in radiative transfer models. The utilization of other channels such as IR1-4 of MTSAT might add useful information about water vapor, cloud temperature etc. for radiation estimation, but herein to assure the cross-sensor applications, only the visible channel of MTSAT data are used as it is available for almost all satellite images. This is reasonable as the visible channel provides the most proportion of information on aerosols, clouds and other atmospheric properties. More explanations on the mechanisms in our model are added into the Introduction and Section 2.2 of the revised manuscript. The related parts in the Introduction are as "The CNN blocks takes image blocks as inputs thereby allowing for identical treatment of adjacent satellite pixels, and are further stacked to construct deep residual structure to extract hierarchical features from low-level details (e.g., geometric shapes, sizes, orientations, edges and distribution) to high-level comprehensive abstract representations (e.g., intrinsic physical and optical properties of mixed clouds). It is believed that such hierarchical architecture of spatial features can fully expose the scattering effects, absorption effects as well as their interactions in the atmosphere, thus can be considered as substitutes for various input parameters representing atmospheric state in radiative transfer models. The MLP is utilized to link extracted features of CNN and additional auxiliary information defining the state in time and space to target measurements of hourly surface radiation through implicit non-linear expressions, whose parameters are learnt from pre-prepared training samples in supervised manner. The deep network is demonstrated to be effective in handling spatial adjacent effects and simulating complicated radiative transfer processes, and successful in achieving superior accuracy of GSR estimates." The related parts in Section 2.2 are as "Satellite image is regarded as a vivid portrayal of the atmosphere and the surface state, and its recorded signals usually contain information on cloud-radiation interactions and impacts among adjacent locations. Traditional physical algorithms retrieve surface radiation from satellite signals on the basis of various radiative transfer models or their simplified versions, where geometric conditions, atmospheric conditions, and aerosol types should be strictly defined, complex processes such as atmospheric absorption and scattering, and their interactions are needed to be precisely simulated, or clear-sky and cloudy retrieval modes are independently developed. Herein, we utilize deep learning technique to directly build the implicit correlations between satellite signals and surface radiation in view of its powerful approximation ability of continuous mapping function. Except that all-sky situations are under a unified framework and tedious intermediate simulations are avoided, different from classical pixel-based retrievals, the CNN blocks are able to deal with spatial adjacent effects of surface radiation, that is, the influence of neighbouring pixels on the central point can be taken into account … In addition, for the convenience of cross-sensor applications, it is better to only depend on the visible channel which is available for nearly all satellite images. This is reasonable as the visible channel provides the most proportion of information on aerosols, clouds and other atmospheric properties (Lu et al. 2011)"

Comments:

3. Please check the unit of GSR, DIF in Figure 4. The unit are different throughout your article.

Response:

Thank you for your advice. We have checked it carefully throughout our article. In this paper, two units (i.e., W/m^2 and MJ/m^2) are used. In Section 3, all figures are shown in the unit of MJ/m^2 corresponding to the unit of the published datasets. To compare the spatial distribution of other products with our results in Figure 4, the figures in Figure 9 also shown in the unit of MJ/m^2. In contrast, the figure 6-8 during validation are displayed in the unit of W/m^2 to keep identical with other studies of hourly radiation estimation (e.g., Tang et al., 2016) for the convenience of comparisons.

Comments:

4. Check the label of the color-bar in Figure 6.

Response:

Thank you for your advice. We have checked it carefully and added color-bar for each sub-plot as well as labels for each color-bar. More explanations about the density of color-bar are added in the figure title as "Figure 6: The density plots of predicted hourly GSR (first column) and DIF (second column) versus measured values at (a-b) all sites and the top of Mt. Everest and (c-d) 5 independent validation sites in 2008. The black solid lines are 1:1 lines and the red dashed lines represent the fitted regression lines, whose expressions are labelled at the lower right corner. Gaussian kernels are used for density plots and the density values are normalized to the 0–1 range through min-max normalization."

Comments:

5. Many statistical indicators (RMSE, MAE, rRMSE, R2) are used to evaluate the model accuracy. How to evaluate the overall model performance of your model?

Response:

Thank you for your advice. We have chosen four indicators, i.e., R^2, MBE, RMSE, rRMSE, to evaluate the model accuracy from different aspects as any single index cannot demonstrate the comprehensive quality of datasets. In fact, it's almost impossible to develop a single indicator to assess the comprehensive model performance. For example, there are issues that the predicted hourly values are highly correlated to the measurements (large R^2 values) but show significant overestimation or underestimation (large MBE values). In contrast, the utilization of multiple indicators could avoid such situation. The R^2 measures the linear correlation between the observations and predictions, RMSE is measure of their overall differences and MBE indicates whether predictions are overestimated or underestimated. For the convenience of comparison, rRMSE is also adopted as it is usually used in similar researches. The overall model performance of our model should be the comprehensive inspects of these four indicators. In addition, I wonder whether the word "overall" leads to misunderstanding. Herein, we use the word "overall" to mean that the model performance is validated on all measured records, not for different months or seasons, or for clear-sky and cloudy conditions. To avoid such misunderstanding, we have changed the title of Section 4.1 as "Validation against ground measurements"

Comments:

6. On the 14th page of your article, you noted "on the whole, estimates from our production correlate well with ground observations at sites with high probability of cloud-free skies". As well known, the northern China and northwestern China are the area with the highest of dust aerosol particles in China, especially in summer. How do you detect clear-sky? Author should evaluate the model accuracy in clear-skies and cloudy skies, otherwise author could not get this conclusion above. As well known, the southern and southeastern China are the areas with abundant precipitable water vapor and dense cloud, which would strongly affect the accuracy of your model. How do explain the accuracy of the estimated DIF are higher in cloud weather conditions? Further sufficient explanation should be given for these questions.

Response:

Thank you for your advice.

The conclusion that estimates from our production correlate well with ground observations at sites with high probability of cloud-free skies is incorrect. We have corrected as "On the whole, estimates from our production correlate well with ground observations at sites with low probability of cloudy conditions, for instance, the north and northwest China." Besides, more explanations are added as "Low R^2 and large rRMSE are likely to occur at sites located in regions with more cloudy days, such as the south and southwest China, especially the Szechwan Basin perennially covered by clouds. As we know, both dust aerosol particles in the north and the northeast China and abundant precipitable water vapor and dense cloud in the south and southeast China lead to non-clear skies, but the model performance is opposite in these areas. This might indicate that deep network does better in emulating the radiation effect of aerosols, but slightly worse in handling that of water vapor."

Different from the previous parameterization schemes, we didn't develop independent clear-sky and cloudy retrieval modes separately. The deep network estimates solar radiation under all-sky conditions in a unified manner, and provides reliable results as indicated by Figure 6. As labels indicating a clear-sky or cloudy condition are unavailable for hourly measurements, we didn't carry out separate comparisons during validation.

To investigate the causes for the contradiction that the accuracy of the estimated DIF are higher in cloudy weather conditions, we carry out an overfitting-test, in which the early-stopping mechanism is removed deliberately and the model for DIF estimate is trained repeatedly until it reaches an obvious over-fitting state where all training samples are intended to be well-fitted. If the model capability is responsible for the contradiction, changes in spatial distribution of R^2 and rRMSE will be in our expectation. However, the results in figure R3-4 show that the model performance in the northwest China is improved but the spatial distribution keeps consistent with that in Figure 8c-d. This evidence points out that the low-quality diffuse measurements in the northwest China results in the apparently worse performance of diffuse estimation. In theory, on the premise that GSR model has proved its effectiveness in arid areas, the worse performance of DIF estimation under the same framework can only be attributed to the data quality. Further evidence also comes from the fact that measurements of diffuse radiation in the western China are not in a full-automatic tracking manner but involves manual operations, of which the nonstandard ones usually lead to measurement errors. More explanations are added into Section 4.3 in the revised manuscript as "With regard to DIF, the correlation between predictions and measurements is much worse than that of GSR, in agreement with the results in section 4.2. Contrary to the GSR, predications of DIF behave well in humid areas (southern China) rather than arid areas (northwest China), which is against our common sense that cloudy weather conditions in the southern China strongly affect the accuracy of radiation estimation. On the premise that GSR model has proved its effectiveness in arid areas, the worse performance of DIF estimation under the same framework might be attributed to the poor data quality. To further investigate whether the model capability or data quality leads to such contradiction, we carry out an overfitting-test, in which the early-stopping mechanism is removed deliberately and the model for DIF estimation is trained repeatedly until it reaches an obvious over-fitting state where all training samples are intended to be well-fitted. If the model capability is responsible for aforementioned contradiction, changes in spatial distribution of R2 and rRMSE will be in our expectation. The results show that the model performance in the northwest China is improved but the spatial distribution keeps consistent with that in Figure 8c-d, supporting the judgement that the low-quality diffuse measurements in the northwest China bear the responsibility for the apparently worse performance. Another evidence comes from the fact that measurements of diffuse radiation in the western China are not in a full-automatic tracking manner but involves manual operations, of which the nonstandard ones often lead to measurement errors. Howbeit, this contradictory phenomenon on the contrary proves the outstanding robust of deep network whose performance would not be easily affected by a small proportion of problematic ground measurements.”

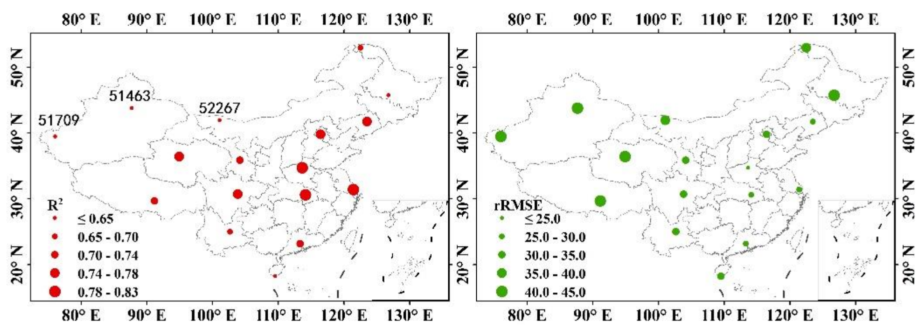
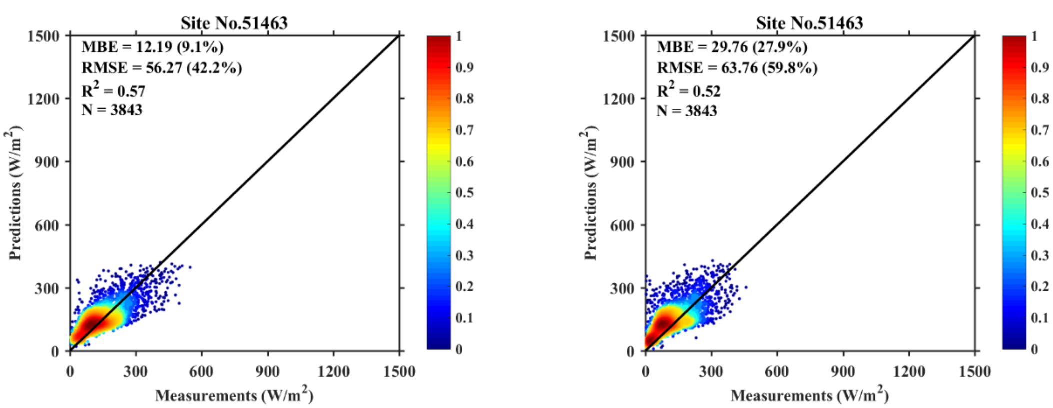


Figure R3 The spatial distribution of R^2 and rRMSE in 2008 during the overfitting-test.



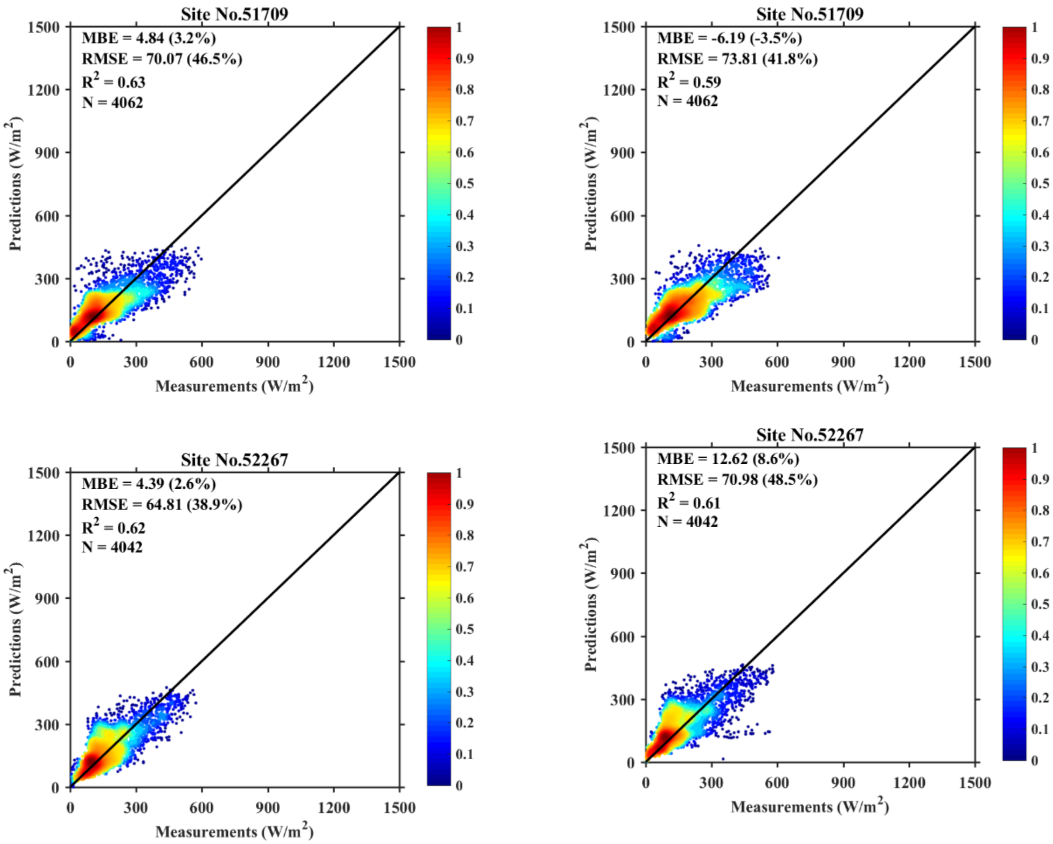


Figure R4 Comparisons of model performance during the overfitting-test (left column) and the standard one (right column) at three sites whose locations are labelled in Figure R3.

Comments:

7. Syntax check in the whole manuscript should be done.

Response:

Thank you very much for your advice. We have checked the grammar throughout the manuscript and corrected related errors.

Comments:

8. The main contents of this article have been published previously in another journal. This is a serious academic moral issue. This article is highly repetitive with your previous articles on Renewable and sustainable Energy Reviews (https://doi.org/10.1016/j.rser.2019.109327). The Figure 1, Figure 2 have been used in your previously published article. Even the main method (CNN and MLP) and the main framework of this article are the same as that in previously published article.

Response:

Thank you very much for your advice.

First of all, we would like to declare that this manuscript is not a copy but a further development of previous works:

1) The previous network for GSR estimation is extended to fit the estimation of diffuse radiation through transfer learning, an approach to reuse already gained knowledge to solve different but analogous problems. A new deep network for DIF estimation is obtained by fine-tuning the GSR network using new training samples consisting of ground measured diffuse radiation and the corresponding satellite image block.

2) The trained DIF network and the previous GSR network are used to simultaneously generate global and diffuse solar radiation over China based on the visible channel of MTSAT data. Herein, time series from 2007 to 2018 are generated to observe long-term variations of surface solar radiation, and a simple example is given in Figure 5.

3) Spatiotemporal errors of our datasets and their potential causes are analyzed to verify the expansion capability of deep network in radiation estimation, provide new insights for future model improvement and give suggestions for rational use of our datasets.

In the revised manuscript, we have made the relations to our previous work clear in the Introduction as "From this point of view, a practical effort has been made in our previous works, where a hybrid deep network mainly consisting of convolutional neural network (CNN) blocks and multi-layer perceptron (MLP) is built to retrieve hourly GSR/DIF from geostationary satellite data (Jiang et al. 2019)… In this paper, we extend the previous network for GSR to fit the estimation of diffuse radiation through transfer learning, an approach to reuse already gained knowledge to solve different but analogous problems. A new deep network for DIF estimation is obtained by fine-tuning the GSR network using new training samples consisting of ground measured diffuse radiation and the corresponding satellite image block. After complete learning and optimization, the trained DIF network in combination with previous GSR network is used to generate radiation datasets including GSR and DIF over China based on Multi-functional Transport Satellites (MTSAT) data. The datasets, covering a period from 2007 to 2018, with a spatial resolution of 1/20 degree, reproduce the spatial distribution and diurnal/seasonal variations of solar radiation in fine scales." and also in Section 2.2 as "In

the previous work (Jiang et al. 2019), we have built a hybrid deep network for GSR estimation. Herein, we further optimize the GSR model to fit the estimation of diffuse radiation by fine-tuning (refer to Section 2.3). The structure is shown in figure 1b and the detailed configurations are listed in table 1. There are two input pipes: *Input1* for MTSAT image blocks and *Input2* for additional attributes including the local time (month, day and hour) and location (longitude, latitude and altitude) corresponding to the central point of Input1. The Output can be either GSR or DIF associated with the central point of *Input1*."
Transfer learning process is also added into Section 2.3 as "4) Train the network for GSR prediction. After initialization, the Adagrad optimizer (Duchi et al. 2011) is used to iteratively find the optimal weights and biases that minimize the mean-squared error (MSE) between the network's predictions and the training targets. An early-stopping mechanism which relinquishes on further optimization when the performance ceases to improve sufficiently or even degrades, is utilized to relieve overfitting by monitoring the network's performance on the validation part, randomly selected 20% of the GSR training set. The model with the best performance is preserved for subsequent predictions, i.e., gaining spatially continuous radiation maps or estimates for other years. 5) Fine-tune the network for DIF prediction. The gained knowledge of the trained network in 4) is from GSR labels through supervised learning, making it not completely suitable for DIF estimation. Thus, DIF training set is used to fine-tune the network through the transfer learning technique. The parameters for CNN blocks are initialized from the pre-trained GSR model while that for FC layers are reset to zero. Other processes are the same to that in 4). In this way, the best model for DIF estimation can be obtained in short time as CNN blocks have mastered the knowledge to extract abstract spatial pattern from image blocks."

The datasets present in this paper are related to the previous work. The main framework of deep network (consisting of CNN and MLP) and the basic data used for model input are similar to the published ones. To avoid the duplication issues, we have revised related figures, for example, the background of figure 2 is changed as land cover types that are a perfect way to demonstrate the representativeness of our stations, the graphical structure of the deep network (figure 1b) is expressed in the way similar to the widely-used framework *Keras*. These changes are added into the revised manuscript as follows.

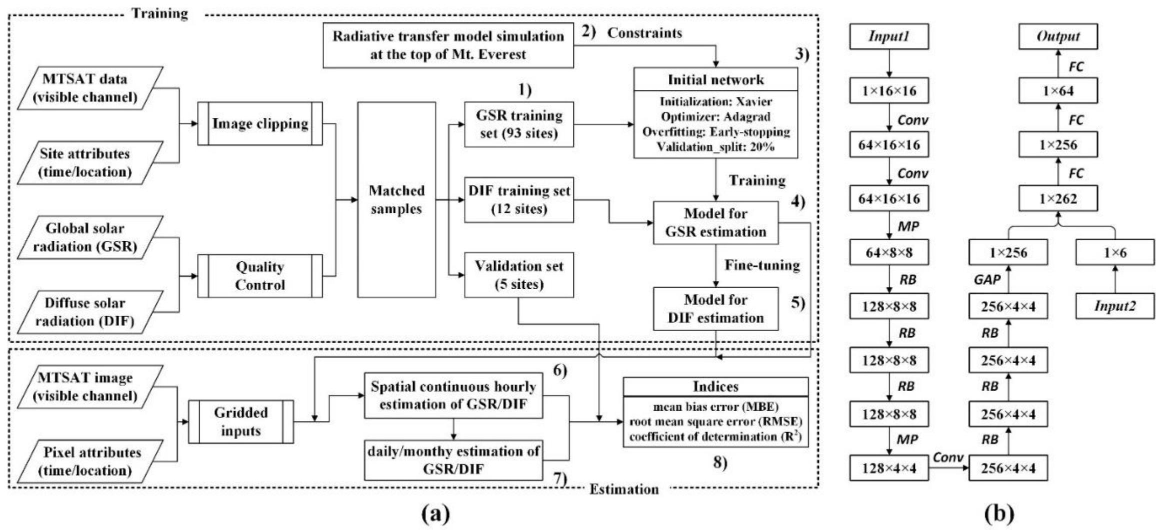


Figure 1: The algorithm used to generate radiation datasets. (a) The flowchart to generate GSR and DIF. Numbers 1-8 correspond to the main procedures listed in section 2.3. (b) The structure of the hybrid deep network. Conv represents convolution operation, MP means max-pooling operation, RB is the abbreviation of residual blocks, and GAP stands for global average-pooling operation. The size of three-dimensional blocks is labelled below as channels × width × height.

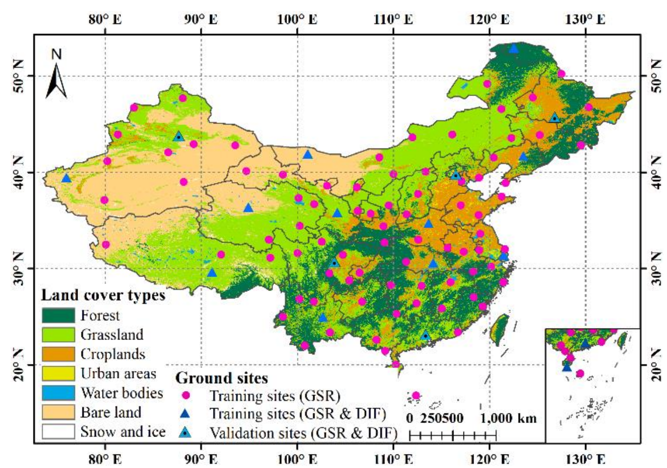


Figure 2: Locations of radiation stations used in our study. Triangles with central point are used for model validation while others are for model training or fine-tuning. The background land cover types are a reclassification of MODIS MCD12Q1 products in 2008.

Comments:

In all, we think that this article is not prepared and should be rejected for publication on ESSD.

Response:

Thank you very much for your comments and suggestions. In the revised manuscript, we have further improved the content of the article and more explanations about the research background and the mechanisms of our model are added. We hope that all the revisions will meet your expectations and you can reconsider our paper and datasets. In addition, we think that our dataset really deserves more attention, especially the provided diffuse radiation which reveals a similar spatial distribution to that of National Renewable Energy Laboratory (NERL), but is significantly different from that of ECMWF-ERA5. Such difference also calls for more attention to the verification and analysis of diffuse radiation from different datasets.