Anonymous Referee #2

Received and published: 9 August 2018

In this study authors demonstrate the possibility of generating contiguous, high resolution estimates of SIF utilizing machine leaning, using as inputs sparse available OCO- 2 SIF retrievals and ancillary satellite data (MODIS). Authors provided extensive error statistics and demonstrated applicability of the new approach in identifying/studying effects of drought. The study is well written and suffers only from minor issues.

Response: We thank the reviewer for the clear summary and positive comments.

Page 1, Line 10-11: "However, several issues, including low spatial and temporal reso- lution of the gridded datasets and high uncertainty of the individual retrievals, limit the applications of SIF.

Reviewer: Binned/averaged datasets are not the only option, there is entire family of products based on geostatistics/kriging (i.e. Tadic et al, 2017), so it could be nice to compare weaknesses/advantages to those products as well.

Response: Thanks for your suggestion. Indeed, geo-statistics is another option to generate high resolution contiguous SIF dataset. However, considering that we already have quite a few contents in this study, and the comparison between different methods in generating high resolution SIF dataset can foster an independent study, we decided not adding the comparison in this paper. Yet, we mention that some other family of products could be used.

Page 1, Lines 14-15: "....we generated two global spatially continuous SIF (CSIF) datasets at moderate spatio-temporal resolutions (0.05 degree 4-day). . ."

Reviewer: How did you choose the ST resolution? Why 4-day? Is it based on the expected decoorelation lenght (variability) in time? Why not 1*day?

Response: The spatial resolution of 0.05 is fine enough for comparison with site level observations (compared with swath based OCO-2 SIF), it also provides global coverage without dividing the Earth surface into tiles (like the 500m MODIS dataset), which simplifies the global application. In addition, many input datasets are available at the 0.05 degree (BESS PAR for example).

We chose this temporal resolution to reach a balance among applications requirements, information redundancy, and dataset sizes. For most GPP dataset, they are either 8-day or monthly temporal resolution. For widely used SIF datasets, GOME-2 for example, they are mostly at the monthly or semi-monthly resolution. Since SIF is often used as a reliable proxy of APAR or GPP, and both of which do

not change abruptly, to get the seasonal or spatial variation, the 4-day temporal resolution is adequate. Although higher temporal resolution may be obtained using the MODIS MCD43C4 dataset, larger spatial gaps caused by the low data quality (cloud, high aerosols) need to be filled. The information redundancy increases if excessive interpolation is applied. In addition, the MODIS MCD43C4 uses 16 days of input with the day of interest emphasized. This reduced the reliability of using MCD43C4 to represent the actual reflectance for the target date. Using 4-day temporal resolution will also decrease the dataset size compared to using a higher temporal resolution, and it is easy to aggregate and compare with other GPP dataset at 8-day or monthly temporal resolution.

Page 3, Lines 24-26: "In addition, OCO-2 can only generate a gridded monthly dataset at relatively coarse spatial resolution, typically at1°×1°, which limits its application in small regions. "

Reviewer: This is not quite correct. It is correct only if we limit our approach to bin- ning/averaging, and ignore spatial autocorrelations. However, if we take autocorre- lations into account, we get have estimates at much higher spatio-temporal resolu- tions (see Tadic et al, 2015, doi:10.5194/gmd-8-3311-2015âA˘ land Tadic et al, 2017, doi:10.5194/gmd-10-709-2017)

Response: We agree that using geo-statistical methods, we can get higher spatial-temporal dataset from low resolution dataset. These methods are effective when the spatial autocorrelations are high (e.g., for atmospheric gases or atmospheric temperature). However, the surface vegetation heterogeneity is high especially in the presence of land use and land cover changes, making this method less applicable. Considering the large gaps between OCO-2 swaths (~100 km), using statistical method without additional information to generate high resolution SIF dataset from OCO-2 SIF would suffer from high uncertainty. Nevertheless, we revised this sentence to make it accurate, we also discussed the challenges of using statistical method to downscale the OCO-2 SIF dataset and cite the references suggested by the reviewer.

"Although several statistical methods are proposed to downscale satellite observations to finer spatial-temporal resolutions (Tadić et al., 2015, 2017), considering the land surface heterogeneity and wide gaps between OCO-2 swaths (~ 100 km), it could be challenging to apply these methods to OCO-2 SIF."

We also discussed the possibility of using geostatistical method to generate spatially contiguous drought monitoring dataset in Section 4.2:

"The spatial coverage issues can be further improved using the geostatistical based method (Tadić et al., 2017), but this may need further investigation."

Page 4, Line 17:" In this study, we aim to generate a global continuous SIF (CSIF) product. . ."

Reviewer: Here and throughout the text, perhaps better choice of word would be con- tiguous. Continuous implies that there is an infinite number of estimation locations, while in practice your estimation interval is determined by the granularity of input data, in this case MODIS retrievals.

Response: Thanks for your suggestion, we use the word continuous to distinguish from the swath based, 16-day revisit cycle OCO-2 SIF data. We agree that "contiguous" is more precise and we have revised the title and other occurrences throughout the manuscript.

Page 4, Line 32:" The reasons for using this resolution include: (1) it is directly compa- rable to the OCO-2 SIF footprint s

Reviewer: This statememnt is questinable, 5x5 gives 25km2 footprint, and OCO-2 footprint is less than 3km2 in size. 8 times difference might or might not be viewed as significant.

Response: We agree that, in terms of the area, 8 times difference can be regarded as significant. However, compared to other OCO-2 SIF aggregations (to 1 degree or 2 degree), they are still in the same order of magnitude. And OCO-2 retrievals can be much more representative for the gridcells SIF values of 0.05*0.05 than 1*1 or 2*2 degree. If we reduce the pixel sizes, we may get very few observations within each pixel and would not effectively reduce the uncertainty. We clarified this statement and have rewritten it as:

"it is directly comparable (in the same order of magnitude) to the OCO-2 SIF footprint size (around 1.3km×2.25km) and the samples within each gridcell can be more evenly distributed and, thus, more representative of the gridcell SIF values than using much coarser $1° \times 1°$ *or* $2° \times 2°$ *grids"*

Page 5, Lines 3-4:" assuming independent estimates and homogeneous SIF value within each gridcell. . ."

Reviewer: Here an additional assumption is required - the SIF has to be not only homogeneous (spatial dimension) but also constant in time.

Response: Since the aggregation is conducted for each day, that is, all the OCO-2 SIF retrievals used to calculate the average of the 0.05 by 0.05 gridcell are obtained with in a very short period of time (usually within several seconds). As long as we assume that SIF is homogeneous, this aggregation process can be regarded as multiple measurements of the same SIF source.

Page 6, Line 1:" For prediction, we first aggregated the daily reflectance to 4 days. " Reviewer:Why 4?

Response: We chose 4-day temporal resolution for the CSIF dataset as a balance between application requirements, information redundancy, and dataset sizes. We refer the reviewer to the previous responses for details.

We added a brief explanation why 4-day temporal resolution is used in this study.

Page 6, Lines 10-11:" A feedforward neural network (NN) is a number of computational nodes (called neurons) structured in a multi-layer architecture."

Reviewer:In principle, NN can be a single layer structure.

Response: Thanks for pointing this out, we have revised this sentence to

"A feedforward neural network (NN) is a number of computational nodes (called neurons) structured in a single or multi-layer architecture"

Page 6, Line 16:" The rectified linear unit (ReLU) was used as the activation " Reviewer:Is there any particular reason for this choice?

Response: The ReLU is frequently used as activation function for scientific research, although in computer sciences, sigmoid functions are also used for classification problems. Overall, ReLU has shown better performance than sigmoid functions in our application but also in most other ones. We now clarify this point further.

Page 7, Lines 14-15:" RSIFGOME-2 (Gentine and Alemohammad, 2018a) uses a similar machine learning technique approach to CSIF but the 15 training is based on the bi-weekly gridded SIF product from GOME-2, and 8-day MYD09A1 reflectance dataset. "

Reviewer: This choice is surprising, as GOME-2 Level 3 products cn be obtained at much higher temporal resolutions, even daily, like it was demonstrated at Tadic et al., 2017. In this case, an unnecessary degradation of information content is induced, as temporal SIF variations during biweekly periods are converted into noise. Given large footprint on GOME-2 retrievals, ML processing here played the role of the downscaling as well, which itself is a challenging process.

Response: We believe that the RSIF dataset has its merits of reducing the uncertainties in raw GOME-2 SIF dataset and downscale GOME-2 SIF to higher spatial resolution. We agree that during this process, the within month variation are partly regarded as the noises, but this has limited effects since the GOME-2 SIF has relative high uncertainties for each individual observation. Since this was performed

in another study we think it is beyond the scope of the current manuscript, but we clarify the point suggested by the reviewer.

Page 8, Lines 1-6

Reviewer: More details are needed here, for the sake of reproducibility. Did you use regularization? What kind of regularization? L2, dropout, their parametrizatuon? How many epochs? Did you use test/validation sets approach or only test?

Response: We did not use dropouts nor other regularization method since the network was not very deep and there was no sign of overfitting. We used 50 epochs with a batch size of 1024 for the training. Only test approach is used. These are now clarified in the text.

Page 10. Line 8:" Figure 10 shows the difference between instantaneous clear-day OCO-2 SIF and CSIFclear-inst. "

Reviewer: Using contiguous Level 3 products based on OCO-2 data and spatio- temporal kriging would yield a figure equivalent to Fig 10, but contiguous.

Response: We thank the reviewer for the suggestion, as we have discussed previously, the application of spatio-temporal kriging to OCO-2 SIF may be challenging and beyond the scope of this study. Here we are only exploring the possibility of using the difference between SIF and CSIF to detect drought. Nevertheless, it could be a very interesting study for the future to map drought impact contiguously using the spatial-temporal kriging. We now acknowledge this in the text.