# *Interactive comment on* "A global spatially Continuous Solar Induced Fluorescence (CSIF) dataset using neural networks" *by* Yao Zhang et al.

**Anonymous Referee #1**

Received and published: 24 July 2018

Review for "A global spatially Continuous Solar Induced Fluorescence (CSIF) dataset using neural networks" by Yao Zhang et al.

In this work the authors produce three datasets of CSIF by filling spatial and temporal gaps of SIF soundings by OCO2 using MODIS surface reflectances and machine learning. The resulting datasets in 0.05deg and 4-day resolution represent gap-filled instantaneous SIF under cloud-free conditions, cloud-free SIF integrated to a daily value and daily SIF under all-sky conditions. To illustrate the advantages and the usefulness of these high-resolution datasets they compare to another downscaled fluorescence product (RSIF, based on GOME2 and different MODIS reflectance datasets), GOME2 SIF, EC GPP and OCO2-SIF itself based on drought occurrences.

The authors convincingly argue why a new down-scaled SIF product - based on OCO2

as a new factor – is needed. At several points I do see, however, need for clarification (where the authors partly contradict themselves in my opinion), further discussion and analysis.

The main points are:

1) Does CSIF represent SIF or APARgreen (p. 3 l. 30)? At several occasions in the manuscript (e.g. p.8 l.14, p.11 l.26) the authors state that based on the reflectances SIFyield cannot be reproduced by NN. They base the drought event analysis on this assumption by comparing CSIF to OCO2 SIF (p.10 l.5). At other moments, however, they stress the close relationship of CSIF to SIF (p.8 l. 18, p.11 l. 31) and call it CSIF.

2) Related to point 1, why not compare CSIF to other estimates of APAR? Greenness index * PAR?

3) Temporal resolution: The final data sets are claimed to have 4day temporal resolution. What is left unmentioned in the manuscript is that the MCD43C4 reflectance data have daily sampling, but each value for a given day still represents a 16-day period (weighted to the central date). So the 4-daily temporal resolution might in reality represent periods of 19days length and potentially affect the downscaling and all comparisons. Please consider this in your evaluations and discussion.

4) Although data processing is described in detail, at points clarification is necessary (e.g. regarding temporal aggregation for training, quality filters of reflectance and EC GPP data, see below).

5) Spatial splitting for training-validation in addition to the temporal one, as extrapolation is not only done in time but also in space. I put it here as a recommendation.

6) I miss plots of time series throughout the manuscript both regarding training/validation as well as comparisons to other datasets. Such plots could contribute to supporting CSIF as a very useful dataset.

7) Related to point 2: An interesting comparison next to the one to RSIF and GOME2

SIF v27 would also be the one to SIF* by Duveiller and Cescatti 2016 that you cite several times.

ad 4)

a) Which period exactly do the OCO2 SIF data cover? It was launched only in September 2014, so only a few months of this year are available. Do you use the full year of 2017?

b) Aggregation of OCO2 SIF (p.4 l. 29-p.5 l.18): How do you aggregate in time? Daily, 4-daily?

c) Figure 1 nicely shows the spatial distribution of the training and validation data. What does the temporal distribution and representativeness look like? A lat-time plot might be useful here. Why are there no data for validation in Alaska and eastern Siberia?

d) You might consider removing barren areas in Sahara and central Asia from the analysis to potentially obtain clearer signals as many data points are obtained from these areas (Fig.1) and might affect the relationships in Fig.2-4.

e) p.5 ll. 12-17: If you first aggregate several soundings to 0.05deg and only afterwards integrate to a daily value, which SZA of the measurement do you use?

f) p.5 l. 21 and at several other occasions in the manuscript: which is the period covered? 2000-2017 or 2001-2016 (abstract) for CSIF?

g) Processing of reflectance data: Do you apply any quality filters? You mention the 'best atmospheric conditions' (p. 6 l. 7), what did you do? I am wondering how can you obtain 'more realistic prediction of SIF during winter' if the reflectances do not represent vegetation but snow? MCD43C4 is sampled daily but values represent 16 days worth of data.

h) You might consider adding a few sentences on BESS PAR in addition to citing Ryu et al. 2018. The quality of BESS PAR is not discussed in the discussion part.

i) In the comparisons to GOME2 SIF v27 and RSIF, is there any accounting for overpass time and wavelength necessary between GOME2 and OCO2? A comparison to SIF* by Duveiller and Cescatti would complete the suite of products.

j) EC GPP: Please add a bit more information: Did you use FLUXNET 2015 or the LaThuile data set? Is there any reason for choosing nighttime partitioning? l. 28-30: I would think you retain those data?

Training and validation (p. 8 ll. 1-26):

-Time series of the CSIF during validation would be nice to see, with validation points overlaid, for example to illustrate the point of SIFyield in SV and GRA.

-Also crops are strongly biased.

-l.17-19: At other points it is argued that CSIF cannot accurately reproduce SIFyield effects based only on reflectances. Please clarify and discuss.

Drought monitoring application:

-Please consider the different temporal information in instantaneous OCO2 SIF versus CSIF based on 19 days worth of data and take it into account and/or discuss it.

-It would also be interesting to see here a comparison (although I see that the different temporal resolutions would mean more work) also to an estimate of APAR, RSIF, OCO2 SIF/ APAR and SIF* by Duveiller & Cescatti 2016.

A few questions when reading the Discussion:

4.1. What is the advantage of CSIF compared to using vegetation index * PAR?

Is there a trend in RSIF?

4.2. Although I fully agree, vegetation indices might not be completely blind. Sims et al. 2006 (RSE,

10.1016/j.rse.2005.01.020) argue that for some ecosystems vegetation indices can

contain information on photosynthetic yields.

4.3. In far-red SIF reabsorption should play a very minor role compared to scattering.

I am afraid I do not understand the last paragraph (specifically p.14 ll. 12-15). Please rephrase.

Minor:

P.6 l. 3: In the end four bands are used, not seven.

Fig.6a) There are also low values in 2007 and 2010 and high values in 2011 compared to 2010 and 2012 during La Nina in Australia at 40S.

Fig.9 Maybe exclude SAA from the plot as GOME2 v27 will be affected by it.

p.10 l. 27: somehow there are too many numbers.

Fig. 10: I recommend a clearer white for the zero values, the gray is difficult to distinguish from the red and blue for small values.

Discussion of uncertainties in BESS PAR is missing.

Conclusion:

MCD43C4 is based not only on Aqua.

p. 15 l. 19: 0.5 deg or 0.05deg?

p. 15 l. 27: beam radiation

subscripts sb and sd are sometimes incorrect/the same

For me, calling the SIF multiplied by a daily correction factor ' daily average SIF' is confusing. It is rather an integration over the day, 'daily integrated SIF'.

Are there any applications or areas for which you would not recommend the use of CSIF?

C5