

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

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 learning, 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.

Page 1, Line 10-11: “However, several issues, including low spatial and temporal resolution 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.

C1

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 decorrelation length (variability) in time? Why not 1*day?

Page 3, Lines 24-26: “In addition, OCO-2 can only generate a gridded monthly dataset at relatively coarse spatial resolution, typically at $1^{\circ} \times 1^{\circ}$, which limits its application in small regions. “

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

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 contiguous. 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.

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

Reviewer: This statement is questionable, 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.

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

C2

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

Page 6, Line 1:” For prediction, we first aggregated the daily reflectance to 4 days. “

Reviewer:Why 4?

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.

Page 6, Line 16:” The rectified linear unit (ReLU) was used as the activation “

Reviewer:Is there any particular reason for this choice?

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 can 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.

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 parametrization? How many epochs? Did you use test/validation sets approach or only test?

Page 10. Line 8:” Figure 10 shows the difference between instantaneous clear-day

C3

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.

Interactive comment on Biogeosciences Discuss., <https://doi.org/10.5194/bg-2018-255>, 2018.

C4