

Interactive comment on “Water, Energy, and Carbon with Artificial Neural Networks (WECANN): A statistically-based estimate of global surface turbulent fluxes using solar-induced fluorescence” by Seyed Hamed Alemohammad et al.

Anonymous Referee #2

Received and published: 2 January 2017

This manuscript is well written and deserves consideration for publication in this journal. However, I have the following issues that need to be addressed.

The paper proposes an empirical machine learning ‘meta-model’ to try to learn from different existing datasets to combine their strengths and factor out their limitations. On one hand, I appreciate this effort to bring together different datastreams and somehow harmonize them through this new consolidated product, but on the other, I am wary of this approach of blindly adding further algorithmic layers without really trying to un-

C1

derstand mechanistically why the initial datasets have shortcomings. If all products are equally off in some parts, combining them just gives the false impression we are going in the right direction while reality is still off. Also, the FLUXNET-MTE used as training is already a machine learning product driven by various input variables, very much like WECANN is. Furthermore, there is quite some circularity in the work since the FLUXNET-MTE and MODIS GPP are both strongly based on the same fluxtowers used here for validation. I deem that all these points need to be acknowledged clearly and discussed thoroughly.

Could you specify why you use the SYN products (Level 3) from CERES instead of the EBAF ones (Level 3B)? The later have been energy balanced according to the product specifications. Wouldn't this be an advantage in your case?

In the construction of the ANN, I would welcome to have some justification of why tangent sigmoid transfer functions are used instead of linear ones. I know this is often done, but it seems very arbitrary. Also, I did not quite understand how the 20% of ‘testing’ data is used. I clearly see that 60% are used for training and 20% for validation, but how exactly do you use the other 20%? Perhaps this just needs some rephrasing in the text for clarification.

Comparison with fluxtower measurements is not appropriate as the difference in spatial support is just too different (1 squared degree vs <1km²). Saying that WECANN performs better than other products based on individual towers while all these products cover such a larger area (by several orders of magnitude) just does not make much sense (even if it has been done in other studies). The authors would need to do some filtering of the towers to select only those that can be considered representative (e.g. <http://doi.org/10.1016/j.rse.2016.04.027>), although I doubt this would leave many valid towers for pixels of 1 squared degree. Another option that may be more feasible would be to make an evaluation at a larger aggregation scale, such as for clusters of similar climates and plant functional types. Making such averages from the fluxtowers on one side and from all pixels that are comparable in this respect on the other would reduce

C2

the number of measurements for validation, but would render them more credible. I would also suggest to exploit more of the available towers in the Fluxnet2015 dataset instead of only 21.

The part pretending to demonstrate the value of SIF is also inadequate as the authors only test the effect of removing this one input. By doing so, any information of the actual vegetation phenology is lost, which would necessarily reduce the performance. What would be interesting would be to show that SIF provides better information than the classical vegetation indices like NDVI or EVI. To do so, the SIF input of the ANN should be replaced by one of these and then a judgement on the pertinence of SIF can be made.

Finally, the manuscript is often too long and too descriptive in several parts describing the graphs and maps. This needs to be reduced drastically. Most of what is being said can be easily inferred from the reader by looking at the graphs, while deeper discussion on why discrepancies occur between products and fluxtowers would be more welcome. Also, please remove the extensive references to different parts of the text and the description of the structure of the paper (e.g. page 3 lines 10-20), I think they are lengthening the text needlessly.

Interactive comment on Biogeosciences Discuss., doi:10.5194/bg-2016-495, 2016.