Atmospheric Chemistry and Physics Discussions

Atmos. Chem. Phys. Discuss., referee comment RC2
https://doi.org/10.5194/acp-2021-631-RC2, 2021

Comment on acp-2021-631

Anonymous Referee #2

Referee comment on "Technical note: Uncertainties in eddy covariance CO_2 fluxes in a semiarid sagebrush ecosystem caused by gap-filling approaches" by Jingyu Yao et al., Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2021-631-RC2, 2021

Review report of "Technical note: Uncertainties in eddy covariance CO2 fluxes in a semi-arid sagebrush ecosystem caused by gap-filling approaches" by Jingyu Yao, Zhongming Gao, Jianping Huang, Heping Liu, and Guoyin Wang

General Comments:

This study reports the result of applying different gap-filling approaches to access the change of NEE over the dry land ecosystems in the western US. Several types of artificial gaps were designed and put into the model to test the capability of using the various gap-filling approach. The authors noted that the performance among these available gap-filling approaches was silimiar, but all of these appraoches fail to fill large gaps over a period longer than two months. Among all selected gap-filling models, the ANN and RF approaches show a better model performance than others approaches, and the RF is relatively cheap in the cost of computational resources. The authors suggested that using RF to fill the gaps is the most efficient way to fill small gaps, such as the gap between hours and several days. In order to deal with the issue of data gaps over two months, the authors develop the strategy by adapting the information from the gap-filled dataset at a daily time scale. Again, the RF gap-filling approach was applied as defined as the second layer of the RF approach to avoid the issue of bias estimation (see Figure 2).

I aggreed with the point of view for dealing with the EC gaps suggested by the authors, and the strategy of using a two-layer RF approach is robust as the results shown in this study. I think the idea proposed by the authors is good, but there is no need to stick to the RF approach. The same idea can also be applied to other gap-filling approaches to avoid bias estimation from long-term data gaps. As reported by the authors, the RF gap-filling approach shows a relatively good and stable result for filling short-term data gaps. Readers may be interested in understanding the importance of the potential variables

used in the RF model. How does the RF model deal with the problems of collinearity among these variables? I suggested the authors report this part of the information to reads in order to support the conclusion made by the autors. Besides, the structure of this manuscript is a bit confusing. Therefore, the idea of using two layer model can move to the section of methodology. The issue of bias estimation from long-term data gaps should also be emphasized in the introduction section. Base on the evaluation mentioned above, I support the publication of this manuscript as a technical note in ACP.

Technical Comments:

L19: …with this framework, the model performance is improved significantly, especially for the nighttime data.

Shall we separate the dataset into daytime and nighttime because the mechanism of producing the CO2 ere quite different both for daytime and nighttime?

L50: The motivation of this gap-filling practice was driven by the fact that dryland ecosystems are very sensitive to water availability,….

A short review of the global coverage of the dryland ecosystem is suggested. Readers may have an overall view of the importance of the dryland ecosystem under the current pace of global warming. Is the ecosystem going to be enlarged or reduced?

L66: …197mm, …

How about the precipitation during the wet years and dry years?

L74: These data were sampled at a rate of 1 Hz…..

The frequency of 1Hz is too low for applying the eddy covariance approach to determine the surface exchange for the grassland ecosystem. Therefore, a 10Hz sampling rate is usually applied to determining the eddy flux for the surface exchange. I hope this is simply due to a typo. However, if the 1Hz is the actual system acquisition rate, I recommend conducting a spectrum analysis to be examined the contribution of high frequency to the total eddy flux by a theoretical correction.

L90: … a score of 2 ….

I have no idea about the score. How to define the score in processing the EC data?

L104: … we use two hidden layers with 12 and 10 nodes in the first and second hidden layers…

Why 10 and 12 layers. Any literature to support these values?

L111: Here, the optimized k value is 9…

Again, how to determine this value?

L265: …. RFs denotes the proposed two-layer RF based gap-filling framework, and MDS is the marginal distribution sampling algorithm….

The abbreviation of RF makes readers confused. Therefore, I suggested using RF-2L to represent the two-layer random forest approach.

L287: Section: A two-layer RF based gap-filling framework for extremely long gaps

I was confused about the two-layer RF approach while reading the manuscript for the first time. Did you mean that combining the approaches from two layers is for ANN + RF, but there is no information on how to combine these two approaches? Therefore, I suggested moving this section to methodology.

L309-L310: The RF algorithm outperforms the other 310 algorithms in terms of the overall performance.

I would like to see the importance of the potential variables that were applied in this study. How to deal with this issue of the collinearity problem?