Atmospheric Chemistry and Physics Discussions

Atmos. Chem. Phys. Discuss., referee comment RC1
https://doi.org/10.5194/acp-2021-631-RC1, 2021

Comment on acp-2021-631

Anonymous Referee #1

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-RC1, 2021

The study by Yao et al. tested and evaluated a suite of machine-learning algorithms in filling the data gaps of eddy-covariance CO2 fluxes at a sagebrush site. They claimed that artificial neural networks and random forest algorithms perform better than the k-nearest neighbors and support vector machine algorithms. Last, they proposed a two-layer framework based on random forest algorithms and suggested providing a more reliable and robust alternative when filling extremely long data gaps.

The research topic is essential and attracts much attention in the science community. The manuscript is generally well-structured and written. I think the manuscript can be considered for publication in Atmospheric and Chemistry and Physics, after addressing a few general and specific comments.

[1] There have been studies, including several cited in the current manuscript, that tested and explored the applications of machine-learning algorithms in filling the data gaps of eddy covariance measurements. I suggest the authors summarize the previous studies' findings and highlight this study's uniqueness or innovative aspect (e.g., dryland ecosystem or the proposed two-layer approach). Below are two recent relevant studies:

Mahabbati, A., J. Beringer, M. Leopold, I. McHugh, J. Cleverly, P. Isaac, and A. Izady (2021), A comparison of gap-filling algorithms for eddy covariance fluxes and their drivers, Geosci. Instrum. Method. Data Syst., 10(1), 123-140, doi:10.5194/gi-10-123-2021.

Irvin, J., et al. (2021), Gap-filling eddy covariance methane fluxes: Comparison of machine learning model predictions and uncertainties at FLUXNET-CH4 wetlands, Agric For Meteorol, 308-309, 108528, doi: 10.1016/j.agrformet.2021.108528.

[2] Presentation of technical details: Certain parts of technical information are not clearly explained or only presented later in the Result and Discussion sections. I suggest reorganizing the texts and moving technical parts forward to the Materials and Methods section when feasible. It will also improve the readability by adding an overview subsection in the M&M, summarizing the study design and entire workflow.

Specific Comments:

[3] Line 38: Since MDS is specifically called out, the original paper (Reichstein 2005) should be cited here.

[4] Line 42: MDS also may be the most widely applied method for gap-filling eddy covariance data and has been used extensively in FLUXNET data (e.g., Pastorello et al. 2020).

Pastorello, G., et al. (2020), The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data, Scientific Data, 7(1), 225, doi: 10.1038/s41597-020-0534-3.

[5] Line 50-52: The sensitivity of dryland ecosystems to water availability is essential and maybe less addressed in previous gap-filling studies. This could be a unique contribution of this study. Yet, it is unclear to me whether and how this current study addresses this knowledge gap. Soil moisture or groundwater table seem not used as input variables. I'd suggest the authors considering exploring additional input variables for water availability.

[6] Line 78-79: Certain variables (e.g., soil) may be spatially varied among the stations. Consider briefly explaining whether or how the spatial bias is corrected.

[7] Line 96-97: I think 10-fold cross-validation already implies resampling and grouping data for model training and validation. It doesn't need to state "repeated ten times". Would you please clarify it?

[8] Section 2.5: It may also be informative to explore the relative importance of input variables. For example, random forest allows the calculation of the relative importance of input variables, and such a feature has been utilized in previous studies to help interpret the results (e.g., Irvin et al. 2021). Other metrics have also been proposed to explain the variable importance, e.g., Knox et al. 2016; Kim et al. 2020.

Knox, S. H., J. H. Matthes, C. Sturtevant, P. Y. Oikawa, J. Verfaillie, and D. Baldocchi (2016), Biophysical controls on interannual variability in ecosystem-scale CO2 and CH4 exchange in a California rice paddy, Journal of Geophysical Research: Biogeosciences, 121, 978-1001, doi: 10.1002/2015JG003247.

Kim, Y., M. S. Johnson, S. H. Knox, T. A. Black, H. J. Dalmagro, M. Kang, J. Kim, and D. Baldocchi (2020), Gap-filling approaches for eddy covariance methane fluxes: A comparison of three machine learning algorithms and a traditional method with principal component analysis, Global change Biol, doi:10.1111/gcb.14845.

[9] Line 135-141: Some technical details need to be explained here. (1) For "10% of the total data length", does it mean that an additional 10% of gaps (i.e., the total number of missing points) are created, or does it mean that artificial gaps are applied to 10% of the data records (i.e., some data points already missing)? I think it's likely the latter case since it's impossible to locate two months without any missing point. Would you please clarify it? (2) I assume the performance evaluation is done based on comparing score-0 observed data and estimated values. Following the previous comment, what are the actual number of data points that are used for each comparison? Would unequal sample sizes affect the performance evaluation or statistic metrics used?

[10] Line 141-143: Some of these metrics seem redundant. Several previous studies used Taylor diagrams, which may be considered, but I don't insist on it.

[11] Line 157-163: As commented earlier, there may be spatial variability among the stations. Additional uncertainties may be introduced to flux gap-filling through these filled meteorological drivers. I suggest considering at least discuss the potential uncertainties.

[12] Figure 2: Please add units to the y-axis.

[13] Section 3.3: I think it's more accurate to call these estimated (or empirical) probability density functions since they are estimated based on data. I'd suggest being more specific about how they are calculated (in M&M). For example, the kernel density function may be the most commonly used. Also, there are more robust statistic tests for comparing density functions, e.g., Z-test.

[14] Section 3.4: I suggest briefly explaining and justifying the use of random measurement errors as a reference, e.g., Why? How to interpret it? In my opinion, gap-filling uncertainties are more like systematic errors, unlike random errors resulting from measurements or turbulence's stochastic nature. I'm not sure it's suitable to compare these two types of errors directly.

[15] Figure 5 and relevant texts: I think it needs a reference (e.g., pure observation or best-case prediction) for the performance evaluation here. I don't understand how the performance is evaluated here. Line 271-272 and analyses in Figure 6 seem to be a better option.

[16] Line 291-293: I suggest providing a brief justification or discussion of the proposed approach.