Geosci. Instrum. Method. Data Syst. Discuss.,
https://doi.org/10.5194/gi-2020-21-AC1, 2020

Geoscientific Instrumentation Methods and Data Systems Discussions
Open Access

Interactive comment on "A comparison of gap-filling algorithms for eddy covariance fluxes and their drivers" *by* Atbin Mahabbati et al.

Atbin Mahabbati et al.

atbin.m@hotmail.com

Received and published: 10 December 2020

General Comments:

RC01: The paper by Mahabbati et al. presents an updated comparison of gap-filling algorithm, which are an important tool in the analysis of data from eddy-covariance sensors and understanding the ecosystem functioning. Their methodology is oriented at the Australian version of the data processing chain taking into account information in addition to the eddy-stations from weather forecasting models and from BIOS2 model data integration environment. For gap-filling of meteorological drivers, they corroborate previous findings of complex methods being not much better than simple methods. Contrary, for the carbon fluxes itself they find a better performance of the machine

learning (ML) based approaches. This study is a valuable contribution to the Australian setup. However, their findings are difficult to transfer to other processing setups and other sites. Hence, the paper is a quite special application and in the current form better suited for an Australian journal.

AC: Even though the data used in this paper came from Australia, the focus was to find out whether ML algorithms other than ANNs can provide more robust results regarding gap-filling of drivers and fluxes. That being said, the towers are selected just as samples to compare the performance of different algorithms. In that sense, the paper is algorithm-oriented rather than Australian-style oriented, and the output is suitable for all members of the FLUXNET. Please note that the diversity amongst the towers has been wide, and there is less likely that an algorithm like RF, which has consistently provided a robust performance in all the five different climates and sites, perform poorly in other parts of the world, or with different input features. Keep in mind that the initiation of this study is to compare different algorithms.

RC01: I encourage the authors for a major revisions to extent their study to setups that are comment also applicable at other sites for submission to GI.

I have several major concerns, which I state here and explain below. First, I propose to add a comparison with fitting the models to only data that are commonly available at other sites. Second, the methodology needs to be updated to introduce gaps at random positions in time instead of all starting at 1st of January to avoid confounding of gap-length with seasonality. Third, I propose to include the MDS algorithm that was simple but well performing at previous gap-filling comparisons and a "business as usual" for gap-filling NEE at many sites.

AC: For the first suggestion, since the main goal of the study was to compare different gap-filling algorithms, we do not believe changing the input data leads to a difference in the relative performance of the algorithms. Moreover, as mentioned in the materials and methods, a variety of climates is involved in this study (Beringer et al. 2016), which makes the results useful for different types of audiences. Please note that the area of Australia is almost as twice as big as the Western Europe, and it has a large variety of climates.

As for the second suggestion, we think we need to clarify the scenario by which we filled the gaps. Since in each gap-filling round the entire 2013 data have been covered by multiple steps for the gap windows shorter than 365 days. For instance, when the gap window is 30 days, the script does the training and testing process 12 times in a row so that it fills the entire 2013, where in each step the model is trained using the data of the previous month. We know that it might not be the best way to fill the gaps, but please note that this paper is the first one of a series of papers which the corresponding author has been working on for his PhD. He has been undertaking the second part of his research by superimposing the gaps randomly and he is writing it down as the second paper of his thesis. So the suggestion will be fulfilled in the second paper wherein the corresponding author has used the same data. However, if the reviewer insists on changing the gap-filling scenario, we would be happy to do the process again, and applying random gaps, instead.

For the third suggestion, including the MDS, we accept the suggestion and will do so.

Specific comments:

RC01: In order to be usable at other sites, the methods should be compared in addition to the presented setup by using only data commonly available at eddy-covariance sites, which are the measurements themselves (Fc, Fh, Fe) together with ancillary measurements (Rg, VPD, rH, Tair, Tsoil, Ustar, precip, wind speed, and wind direction), and maybe another comparison using in addition more detailed radiation measurements and ground heat flux and soil water storage (Table 2).

AC: We believe that it is a useful suggestion. However, as mentioned earlier, the main goal was to compare different gap-filling algorithms and it is less likely that changing the input features makes any change in the comparative performance of the models.

For instance, Kim (Kim et al 2019) compared ANNs, RF, SVR and MDS to fill the gaps of Methane flux with different input features than this study, and the performance ranking amongst the ML methods was quite similar to this paper: RF outperformed the rest, and ANNs outperformed SVR. Besides, the data used in that research came from North America.

RC01: In the current comparison setup, the larger gap-lengths comprise a larger proportion of other seasons, while the short gap-lengths only comprise summer records. Hence, the conclusions on gap-lengths are confounded with seasonality. I suggest to randomly distribute gaps in the portion of the entire data series with sufficiently high proportion of non-missing original data. Moreover, most data-processing setups will not fit a model for each gap tailored at the gap-length. Hence, I suggest to introduce several gaps (of a given length) across the entire dataset (say of proportions of 40% and 70% of the data according to p6L215) and let each methods fill all these gaps and compute the statistics across all the gaps but also of the aggregated annual value. In this way a recommendation can be presented that is closer to the gap-filling as applied at many sites. The decision to adjust the training window to the gap-length is very difficult to compare to other gap-filling of real time series where gap-lengths vary. Most investigators will not effort to fit a model around each gap. I suggest training the methods on a shifting window and filling all gaps inside this window, and for efficiency use only few increasing window lengths of the training.

AC: This is a good suggestion, and this is a better approach in general for a realistic gap-filling process. However, we want to point a few things out: First, It seems that the paper explanation about the gap-filling approach is not clear enough. we think we need to clarify the scenario by which we filled the gaps. Since in each gap-filling round the entire 2013 data have been covered by multiple steps for the gap windows shorter than 365 days. For instance, when the gap window is 30 days, the script does the training and testing process 12 times in a row so that it fills the entire period of 2013, where in each step the model is trained using the data of the previous month. We confirm that this approach might not be the best way to fill the gaps, but please note that this paper is the first one of a series of papers which the corresponding author has been working on for his PhD thesis. The second paper of his research is been doing by superimposing the gaps randomly. So the suggestion will be fulfilled in the second paper anyway.

Second, the main goal of this study has been comparing the performance of different ML-based gap-filling algorithms. In that regard, it is less likely to be a considerable relative performance difference between the scenarios whereby the gap-fillings are carried out.

According the points above, I believe that changing the gap-filling scenarios would not make any significant change in the relative performance of the algorithms. Considering the main goal of the current study, which to compare different algorithms, although the suggestion is generally constructive, changing the gap scenarios does not seem to add that much of a value to this paper. Nonetheless, I am happy to redo the study with the suggested approach if the reviewer or the editor insists on that.

RC01: Moffat et al. (2007) concluded that the quite simple and widely applied MDS algorithm for filling Fc, i.e. NEE time series, which is using only the common variables NEE, Rg, Tair, and VPD as predictors. What are the reasons to omit this for many sites "business as usual"-algorithm? The computation can even be outsourced to the online tool provided by the MPI-BGC Jena.

AC: We accept the suggestion, and we would include the MDS.

RC01: P7L226: Were all the eight drivers used or a subset of them, maybe different by method? What is q? The formulation "by trial and error" needs more explanation.

AC: All the eight drivers were used for all methods. Symbol q is the specific humidity, which should have mentioned on table 2. Here "trial and error" was made based on applying feature importance analysis using random forest, and then feeding the al-

gorithms with the different combinations of the suggested features to find out which combination provide the best performance metrics. We will explain that in the revised version.

RC01: P10L308: Here it does not become clear what cross-sections have been used. I imaged some categories based on similar environmental conditions or day/night time. This only becomes clear in the discussion, in that data from other sites have been used with site as cross-section. This cross-site gap-filling is hard to transfer to other studies. In what respect does the PD model differ from a classical mixed effects models?

AC: For each tower, we used the four rest towers as its cross-sections. Now that we know how much important the similarity of the cross-sections are, it is obvious that the method can be used for the regions where the density of towers are high enough, e.g. central Europe. Nonetheless, the computational problem is also a big concern, making the method not to be feasible, at least as long as our computational power has not been dramatically changed. Regarding the difference of PD from classical mixed effect models, it should be noticed that PD can be considered as a combination of a classical mixed effect model with a time series model ,e.g. ARIMA models. More information will be provided in the methods, accordingly.

RC01: P27L720 Conclusions 4 and 5 are mere speculations given the results presented in the paper. They should be moved to the discussion. Contrary to the suggestion 4, I hypothesize that using net radiation as a predictor should handle this case already well (at least with RF). Otherwise, I suggest first trying to add a nighttime/daytime flag to the set of predictors before splitting the dataset.

AC: We would move conclusions 4 and 5 to the discussion. As for the reviewer's hypothesis, it is a good idea to be tested out. However, as mentioned earlier, this study is the first ongoing series of papers the corresponding author is going to prepare for his thesis. Thus, it is a good idea to include the reviewer's hypothesis in the second paper, since the later should be the logical consequence of what has been found in the first paper.

RC01: P1L35: Currently, I was confused reading the abstract. It was hard for me to spot the distinction between filling of environmental drivers and filling of fluxes. This can be formulated more clearly.

AC: We do agree with the suggestion. The abstract would be revised thoroughly.

Technical corrections:

RC01: P2L43: This formulation does not become clear to me.

AC: Right point. The sentence is needed to be edited.

RC01: Tab 2: I suggest indicating the commonly used abbreviation for the fluxes in parentheses in addition to the notation of the paper (NEE, LE, H). P11L331: typo: "non-periodic" eq 12: one bar too much.

AC: We do agree.

RC01: P23L583: I suggest to provide another table with method abbreviations or repeat the abbreviations at the beginning of the discussion. By this way you do not force your readers to study the methods section first.

AC: Sounds useful. This will be done.

Interactive comment on Geosci. Instrum. Method. Data Syst. Discuss., https://doi.org/10.5194/gi-2020-21, 2020.