Geosci. Instrum. Method. Data Syst. Discuss.,
https://doi.org/10.5194/gi-2020-21-RC2, 2020

Geoscientific Instrumentation Methods and Data Systems Discussions

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

Anonymous Referee #2

Received and published: 20 October 2020

GENERAL COMMENTS

The paper presents a detailed evaluation of eight algorithms for gap-filling time series data, using eddy covariance data as a target for the comparisons. The content about the algorithms and the metrics for comparisons are a strong feature of the paper. However, it is more limited in advancing the knowledge of best practices for eddy covariance and micro-meteorological data gap-filling. In other words, the evaluation of the algorithms against each other is of interest, but the chosen test domain is not clearly impacted.

It seems that to really benefit the knowledge of methods for gap-filling eddy covariance data, longer time series and more representative gap scenarios would be necessary, as well as a clear comparison to more established methods. Multi-year datasets are

key to properly evaluate these algorithms. Such datasets are now widely available, so it is unclear why only 2013 was used. With this aspect in mind, it seems clear that in the evaluation of the first objective of the paper longer gaps led to disproportional increases in uncertainty. This might not have happened if other years without gaps for the same season were available, for instance. More direct comparisons to "classic" gap-filling algorithms would have helped in this evaluation. Implementations of algorithms such as MDS are now widely available, including as part of OZFlux's own OzFluxQC software package. The comparison of newer methods is informative, but unless compared to currently used solutions, it's hard to assess the improvement. Although the authors are correct, and performance of the MDS algorithm was shown to be comparable to ANNs before, parameterizing MDS is much simpler (no choices in layers, nodes, iterations, or window sizes) and would lead to a more robust and clear comparison.

Should the authors choose to really focus on the comparison among the methods presented, I would suggest adding all the comparison metrics RMSE, R2, MBE, etc., for all sites individually and combinations thereof as supplementary materials, making this a valuable and thorough comparison of methods, and reducing the focus from the application to eddy covariance. If the intention really is to show the impact on EC, longer time series and more direct comparisons to current methods would be necessary.

SPECIFIC COMMENTS

On the ancillary datasets, it seems they introduce some entanglement to this evaluation. One of the key advantages of purely empirical methods, such as the ones presented in the paper, is that they will not be biased by predefined models (like the reanalysis datasets) or atmospheric interferences (like the MODIS data). After an evaluation without datasets such as these, adding them to improve the methods would be a natural choice. However, without the unbiased evaluation it is hard to qualify the sources of uncertainty in the paper's evaluation.

Although the performance criteria selected for the paper work well, it is curious to see

that the methods all seem to represent high variabilities but fail to capture the extremes, as the authors point out for CO2 and latent heat fluxes – and this doesn't seem to be the case for sensible heat flux. Could this be an issue of the underlying data requiring further quality control before the gap-filling methods are applied? Or maybe this is an artifact of the period selected in the examples?

The following claim requires either more details or a reference, otherwise it's not possible to know what concerns/challenges the authors are referring to and what aspects of gap-filling the paper is aiming to address: "...there are some serious concerns regarding the challenges associated with the technique, e.g. data gaps and uncertainties."

The +/-25gCm-2y-2 (Moffat et al. 2007) and +/-30gCm-2y-2 (Richardson & Hollinger 2007) are dependent on the underlying datasets used for the evaluation. These numbers should not be taken as general benchmarks.

In the sentence "Nevertheless, one of the concerns regarding this algorithm is that the independent variables, here meteorological drivers, might be auto-correlated." it is unclear why this would be a concern, since the meteorological drivers being auto-correlated is one of the assumptions that allow the MDS method to work.

The sentence "This challenge becomes acute when the gaps happen within a period when the ecosystem behaviour is changing and thereby showing different response under similar meteorological conditions." is another reason why multi-year datasets should be used to compare these algorithms.

The gap scenarios and training windows selected are highly structured and rigid. It's unclear how the evaluation over these scenarios would translate into real-world examples, which have both structured gaps (e.g., from sensor failures) and arbitrary gaps (e.g., from data filtering). It seems it would be important to use are least one scenario with gaps and training data both randomized, and also combinations of lengths for gap windows and training windows.

TECHNICAL CORRECTIONS

— Abstract

- The acronyms RF and CLR were referenced before being defined

- "...RF provided more consistent results with less bias, relatively." It would be clearer to describe "relatively" to what in this sentence.

- This sentence is a bit unclear "In each scenario, the gaps covered the data for the entirety of 2013 by consecutively repeating them, where, in each step, values were modelled by using earlier window data." Were measured and modeled data used simultaneously in training?

— Introduction

- "...and not measured at the point." Maybe could be "not measured at a point scale"?

- A more classic reference for FLUXNET is: Baldocchi et al. 2001. FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem-Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities. BAMS, 11: 2415-2434.

- And more appropriate references for EUROFLUX and AmeriFlux are: Aubinet, M. et al. 1999. Estimates of the Annual Net Carbon and Water Exchange of Forests: The EUROFLUX Methodology. Advances in Ecological Research, pp. 113–175. Law, B. 2007. AmeriFlux Network aids global synthesis. Eos, 88, 286–286. Novick, K. A. et al 2018. The AmeriFlux network: A coalition of the willing. AFM, 249:444-456.

- "Despite the capability of EC to frequently validate process modelling analyses..." might be more precisely phrased as something like "Despite EC data being frequently used to validate process modelling analyses..."

- "[...] Moffat et al. (2007) compared a couple of different commonly-used gap-filling algorithms"; in fact, Moffat et al. 2007 compared 15 gap-filling techniques.

— Materials and Methods

- "and Tumbarumba form 2011 to 2013..." form -> from

- "Each algorithm was tuned up individually using gird search,..." gird -> grid

— Results

- Even with a maximum zoom in the PDF file, it is rather hard to read the axis for Figures 3 and 4

— Discussion

- This sentence is unclear: "That is because ANNs have been checking out for a long time in different locations and considered as one of the most reliable algorithms in the field for more than a decade"

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