

Interactive comment on “Dynamic INtegrated Gap-filling and partitioning for OzFlux (DINGO)” by Jason Beringer et al.

Anonymous Referee #3

Received and published: 17 June 2016

The manuscript presents an overview of the software framework and methods for automated gap-filling and partitioning of OzFlux tower network data. Overall, the manuscript does a nice job presenting this overview with the right amount of detail and rationale, and is well-written. The gap-filling and partitioning methods employed are generally state-of-the-art community standards. That said, there is not much new information here. Perhaps the greatest novelty lies in the gap-filling of ancillary drivers, where to my knowledge the only other recent work on this topic is that of Vuichard and Papale (2015). Although that work is not referenced, the present manuscript expands upon it to incorporate data from nearby weather stations and satellite remote sensing observations, with additional soil- and radiation-tailored gap-filling methods.

I have two major recommendations: 1) The manuscript could be made more impactful by including an analysis of performance of the gap-filling and partitioning results

C1

across all or part of the OzFlux network. This would demonstrate the utility/flexibility of the framework, its ability to address the grand science challenges identified in the introduction, and highlight areas for continued improvement. 2) The paper would benefit from a discussion of how the methods employed in this paper conform to or push the envelope of current community practice. This would clearly communicate the significance of the software suite and its novel contributions.

Specific comments:

- The representation of meteorological quantities should make more effort to be consistent with community standards. I have found Reifsnyder et al. (1991) particularly useful for this purpose, presenting the symbols, units, and notation for use in the journal *Agricultural and Forest Meteorology*. Conforming to these where possible/practical will improve the readability and reach of the manuscript. Also, both R_e (pg. 3, line 31) and R_H (Fig. 1) are used in the manuscript to represent ecosystem respiration, and both F_c (pg. 6, line 14) and NEE (Fig. 1) are used to represent net ecosystem exchange of CO_2 .

- Pg. 5, line 17-20 & Pg. 6, line 9-11: Is there a threshold correlation below which the data are not gap-filled? A discussion of this topic is warranted.

- Pg. 5, line 17-20: Linear regression is a good start for using nearby station data for gap-filling. However, even from Fig. 2a (which looks to be a very clean example), the best correlation has a distinct non-linear component at the low end of values. Using this relationship to gap-fill the time series then extends the data into minimums not actually observed at the site, which has the potential to influence down-stream modeled physiological responses. Perhaps the manuscript could address this topic as a discussion point, and/or include a few different fitting functions in future iterations (maybe piecewise linear fits?).

- Pg. 8, line 9-12: It would be informative to show some stats addressing how well this procedure replicates actual variation in solar radiation under cloudy conditions across

C2

the OzFlux network (opposed to using nearby BoM site data). What about using the diurnal average approach employed for estimating incoming longwave radiation under cloudy conditions (pg. 9, line 6-8)?

- Pg. 10, line 4-5: How was the ANN hidden layer architecture of 24 and 16 nodes arrived at? Appropriate model complexity can significantly impact ANN model performance. Recent related works using ANNs for gap-filling have tested for ideal architectures (Papale and Valentini, 2003; Knox et al. 2015; Baldocchi et al., 2016).

- Pg. 10. ANN procedure: How are the data split into training and testing/validation sets?

- Pg. 12, line 12: “u* corrected output” implies that that data were somehow corrected for low u* conditions. I recommend rephrasing.

- Pg. 12, line 16: Please clarify which Fre variable was used (Fre_NN or Fre_Con).

- Pg. 12, line 16: What does “ve” mean? (as in +ve, -ve)

- Pg. 12, line 17-22: This explanation is confusing. Recommend rephrasing. Consider explaining that forcing GPP to zero at night removes positive and negative random error equally, but forcing any positive GPP values to zero during daytime would bias results because only negative random errors would remain. Also, I am not sure what is meant by “no random error in the Re ANN calculation”, since random error is relevant in any model fit to real data.

- Pg. 14, line 16-27: Since the ANN is employed for gap-filling, why not use the model residuals as an estimate of the random error? The model residuals of high performance gap-filling algorithms such as the ANN provide a good, if not conservative estimate of the random uncertainty (Moffat et al. 2007, Richardson et al. 2008). The daily differencing approach is much more conservative as it includes natural environmental variability as a result of variation in the flux footprint. This would help alleviate some of the double-counting of uncertainty mentioned in a later paragraph (pg. 15, line 7-12).

C3

- Pg. 14, line 19: Please indicate what lowercase delta means.

- Pg. 14, line 18-21: Why must random error be calculated over bins of u*? (rather than over bins of (or a regression with) flux magnitude)

- Section 2.6 Uncertainties: This section would benefit from a discussion on the uncertainty propagated along the entire processing chain described in the manuscript (i.e. using gap-filled ancillary drivers value to gap-fill fluxes).

- Pg. 17, line 16-18: What is the suggested turnaround time from data collection to quality output from the procedure outlined in the manuscript?

Technical corrections

- All bulleted lists: In some cases the lack of sentence case following the bullet point makes sense. However, in many cases the sentence following the bullet is a stand-alone sentence and should use sentence case.

- Pg. 10, line 29: “stable conditions” should specifically reference “atmospherically stable conditions”

- Pg. 13, line 26: I am confused by the sentence fragment “Once complete daily time series for the parameter estimates, . . .”. What about this: “Once daily estimates for the parameters are generated, . . .”

- Labels in figures need cleaning up (e.g. Fig. 4 y-axis – what is CABLE?, Fig. 5 y-axis – make this human readable, Fig. 9b legend – what is the difference between Fc and Fc?, use superscript and subscripts where appropriate)

- The manuscript needs thorough editing for typographical errors

References cited

Baldocchi, D. et al. The impact of expanding flooded land area on the annual evaporation of rice. *Agricultural and Forest Meteorology* 223, 181–193 (2016).

C4

Knox, S. H. et al. Agricultural peatland restoration: effects of land-use change on greenhouse gas (CO₂ and CH₄) fluxes in the Sacramento-San Joaquin Delta. *Global Change Biology* 21, 750–765 (2015).

Moffat, A. M. et al. Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. *Agric. For. Meteorol.* 147, 209–232 (2007).

Papale, D. & Valentini, A. A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization. *Global Change Biology* 9, 525–535 (2003).

Reifsnyder, W. E., McNaughton, K. G. & Milford, J. R. Symbols, units, notation. A statement of journal policy. *Agricultural and Forest Meteorology* 54, 389–397 (1991).

Richardson, A. D. et al. Statistical properties of random CO₂ flux measurement uncertainty inferred from model residuals. *Agricultural and Forest Meteorology* 148, 38–50 (2008).

Vuichard, N. & Papale, D. Filling the gaps in meteorological continuous data measured at FLUXNET sites with ERA-Interim reanalysis. *Earth System Science Data* 7, 157–171 (2015).

Interactive comment on Biogeosciences Discuss., doi:10.5194/bg-2016-188, 2016.