



## Peer Review Report

## Peer review report 1 on “CARBON NEUTRAL OR A SINK? UNCERTAINTY CAUSED BY GAP-FILLING LONG-TERM FLUX MEASUREMENTS FOR AN OLD-GROWTH BOREAL BLACK SPRUCE FOREST”

### 1. Original Submission

#### 1.1. Recommendation

Major Revision

### 2. Comments to Author

In the manuscript, the authors investigate the carbon balance of the Northern Old Black Spruce (NOBS) site based on a 15-year data record. First, they want to test the hypothesis that this old growth forest is carbon neutral and second, they want to analyse the uncertainty caused by different gap-filling approaches.

In my view, both aspects are interesting and legitimate questions and worth publishing. The manuscript is well written and follows a clear logic. Nevertheless, in my view, the manuscript lacks novelty for two reasons. First, the data from that site have been partly published before (Dunn et al., 2007). Now a longer time series is provided, but little is done to understand better the longer time series and its inter-annual variation. Second, the main focus of the paper is on comparing four different gap filling approaches. While this is interesting in principle, we don't really learn why there are differences. At the end, the differences are taken as a measure of uncertainty, but we don't gain any insights in the reasons.

#### Specific comments

1) I suggest that the authors expand their analysis and provide a better understanding on why the four different gap-filling methods

are different. Why are night time NEP more negative for FCRN and Harvard compared to MPI and Mean Data? Why are summer NEP larger for MPI and Mean Data? Why is R smallest for MPI? Is this because of data selection, the statistical model, or the size of data windows? Are the differences specific to the NOBS data set or are they to be expected in general.

2) The Harvard gap-filling dataset seems to be based on an earlier calculated dataset by Dunn et al., 2007. I am wondering why the Harvard gap-filling was not recalculated based on the exact same dataset. Then, it would also cover the full time period (incl. 2007 and 2008) and it would become more clear if differences are attributable to the gap-filling approach or to some other data difference, e.g. due to flux calculation, data quality control, etc. In line 496–498, the authors state that there were “slight differences in how the variables were derived” and that “Harvard is more sensitive to [environmental] conditions”.

3) The authors could easily save some space by removing Fig. 2. I don't think it provides any information that could not be written just in 1–2 sentences.

4) Methods: it is unclear which software was used to calculate the fluxes. Any of the standard packages, e.g. EddyPro? Or a custom-made software? Which corrections were applied? Where all years calculated with the same software? Was the Harvard dataset (1994–2004) calculated with the same software as the Amiro dataset (2005–2006)?

Anonymous

Available online 5 December 2016