Biogeosciences Discuss., 12, C6067–C6069, 2015
www.biogeosciences-discuss.net/12/C6067/2015/

Biogeosciences
Discussions
Open Access

Interactive comment on "Automation of soil flux chamber measurements: potentials and pitfalls" *by* C.-M. Görres et al.

Anonymous Referee #3

Received and published: 2 October 2015

Görres and colleagues present a comparison of two automated systems for the measurement of soil CO2 efflux and discuss many methodological issues related to the measurement of trace gas fluxes from soil. I have two major concerns, followed by a few other concerns.

Major concerns:

(1) The authors make several claims that are not supported by data.

a. The authors make conclusions regarding collar artefacts, particularly the alteration of root growth by the 8100A collars. This is even mentioned as one of the "main reasons for the "observed differences in the performance of the two systems" (iv; line 17) in the abstract. However they present no data to support this claim. A single photo in

supplementary material does not constitute data. b. The authors say that "the impact of the automated chamber systems on the environmental conditions increased with the size of the chamber itself and additionally with the size of the frame. . ." (line 14, page 14713). Presumably the authors are referring to the soil temperature and moisture data presented in Figure 2, as the bulk density, C content, and DOC content were not different (line 6, page 14707). However the authors present no in-situ (i.e., non-chamber affected) measurements of soil temperature or moisture, as such it is not possible to conclude whether the AGPS or the Licor systems altered these values relative to in-situ conditions. The authors can only compare the temperature and soil moisture from the AGPS and Licor systems (e.g, Fig. 2c-d). Remarkably, the authors do not quantitatively or statistically compare these data in any way.

(2) The authors make no statistical comparison of the flux estimates provided by the two automated systems apart from the integrated temporal sums reported on line 14, and this statistical test appears to be based on the 95% confidence intervals of integrated predictions from a Loyd and Tylor equation. I find the lack of other statistical comparisons of the two methods to be a striking omission for a manuscript purporting to compare the two methodologies. I suggest the authors consider a robust statistical comparison, possibly such as repeated-measures ANOVA on daily mean flux estimates, with fixed effects of method, date, and method $\times$ date and a random chamber (collar) term. Other methods such as time-series analyses, spectral analyses, or generalized additive models may also be appropriate (see comment below for generalised additive model information). Such methods could identify particular dates or periods when the flux estimates diverged, which could usefully focus the manuscript around methodological issues specific to those periods. Neither autochamber method is quantitatively compared to the CO2 concentration gradient method. I suggest the authors consider removing this method from the manuscript.

Other concerns:

(3) It is difficult to compare the methods in the time series plots (Figs. 4-5) given the

issue of overplotted points. I suggest the authors explore heat maps, density clouds, or even simple running averages to visualize the central tendency of these datasets. Better yet, generalized additive models (GAMs) with random chamber effects could be used to display the estimated mean and 95% confidence intervals of these datastreams over time, and any statistical difference between the methods could be inferred via the 95% confidence intervals. See the "mgcv" R package and associated articles (Wood, 2011).

(4) The authors should more fully illustrate how well the Lloyd and Taylor models describe the observations, particularly if the summed predictions of these models will be used for inference of methodological differences, as is currently done. Model predictions plotted on top of the data vs. temperature, or observed vs. predicted plots, would be useful to assess potential bias. The authors do present the residual standard errors and the parameter standard errors in Table 3, but these numbers are of limited utility to assess bias.

(5) The units in Table 3 for "Average cSR" appear to be incorrect. Efflux rates of 897 μmol CO2 m-2 s-1 (for example, AGPS, Wide, not filtered) are a bit high. The units for these cumulative sums are likely incorrect. I also hope and expect the authors intended to refer to the 95% confidence interval, rather than the 5% confidence interval.

Reference Wood S.N. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society Series B-Statistical Methodology, 73, 3-36.

Interactive comment on Biogeosciences Discuss., 12, 14693, 2015.