Dear authors,

the referees made only minor comments, which you have addressed in a satisfying way. However, I have identified an issue related to the changed statistical analysis in the revised version (see Comment 5). In addition there are some other minor issues that need to be addressed before the manuscript is ready for publication. They are listed below.

Best wishes
Christof Ammann
AMT Associate Editor

Dear Dr. Ammann,

Thank you for the constructive feedback on our paper. Below we list our responses to your comments.

EDITOR COMMENTS

1) Line 45: Was the mass flow controller specifically calibrated for CO2?

The mass flow controller (MFC, Aalborg GFC model) was newly purchased, configured for CO_2, and factory calibrated. As a gross check on the MFC accuracy, we monitored the weights of the CO_2 cylinders used for the gas release study (measured before and after use). Over the full-study, the total MFC based gas release total agreed with that calculated from the cylinder weight loss to within 5%.

We have added more information about the MFC on line 46:

> *"… passed through a mass flow controller (GFC57 configured for CO_2, Aalborg Instruments and Controls, Inc. Orangeburg, NY, USA) …"*

2) Line 47: What was the pressure in the manifold? Was it really much above ambient pressure?

The inlet pressure to the MFC was 150 kPa, and the maximum pressure drop across the MFC as given by the manufacturer is approximately 75 kPa for the flow rates used. Based on simple engineering calculations for a similar synthetic source, we assume the pressure loss across the manifold, hoses, and piping is small (as mentioned in Flesch et al. 2004). We thus estimate the manifold pressure as greater than 75 kPa (or 10 psid).

On a related point, we have modified the text on line 48-49:

> *"We assumed equal flow rates from each outlet due to the high head loss across each outlet relative to the manifold pressure (following the argument made by Flesch et al., 2004)"*

3) Line 65: What was the value range of the background fluxes? Are they negligible for the uncertainty of the results? Some of the measured EC fluxes for the large source distance were also quite small (according to the data in the supplemental material).

The background EC fluxes varied from -0.8 to +0.9 umol m^{-2} s^{-1}. During gas release periods the measured fluxes ranged from 0.3 to 118 umol m^{-2} s^{-1}. In most cases the measured flux during gas release was more than an order of magnitude larger than the background flux. But as you mention, this was not always the case with the larger fetches (as seen in the supplemental file). So yes, in these cases there will be a relatively larger level of measurement uncertainty in Q_{LS} or Q_{KM}. But these are also cases where the actual release rate is large, so that the uncertainty in the Q_{LS}/Q or Q_{KM}/Q ratios should still be relatively small.

4) Line 86: Unfortunately the link provided in Neftel et al. (2008) to access the ART footprint tool software is obsolete. You could add the following updated DOI link to access the tool: http://doi.org/10.5281/zenodo.816236

Thank you for this information. Reference to the ART footprint software (Spirig et al. 2007) was added to the text on line 87, and the DOI link to access the tool was included in the list of References.

5) Line 103-106: I agree that the geometric mean and corresponding asymmetric confidence interval are likely more accurate for the present data than the originally used artithmetic mean. However, the provided argument is not really appropriate. It is true that the ratio of two quantites with Gaussian error distributions shows an asymmetric error distribution (F-distribution). But in the present case, the error (uncertainty) of Q_KM/Q and Q_LS/Q are almost fully determined by the error of the nominator, because the denominator (Q) presumably has a negligible random error. This means that Q_KM and Q_LS themselves already must have an asymmetric error distribution, which is reasonable because they are calculated as a ratio with the footprint fraction (of the artificial source) in the nominator and the measured EC flux in the denominator. In addtion, the modelled footprint fraction may itself have an asymmetric error distribution because of the limitation to positive values and the special shape of the footprint function. Please reconsider and improve the reasoning/explanation for the use of the geometric mean. Apart from theoretical considerations, it could simply be argued that the values of Q_model/Q show a clearly asymmetric distribution and thus a logarithmic transformation is useful. It may additionally be useful to test whether the geometric mean and the median of Q_model/Q show similar results.

These are very good points, and indicate the potential for a more complex analysis. And we seem to agree on the broad issue of how this data can be analyzed in a relatively simple way. We want to make a few general comments:

- It is intuitive that our ratio data (Q_{KM}/Q, Q_{LS}/Q) are not normally distributed: the ratios are bounded at zero and unbounded at the top end. It is intuitive that the arithmetic mean is not the ideal measure of central tendency.
- It is broadly accepted in statistical texts that for normalized data (like our ratios), the geometric mean is a better measure of central tendency than the arithmetic mean.

- The big-picture conclusions of our study (e.g., the large period-to-period variability in model accuracy, the lack of statistical differences between the LS and KM models, the poor performance of both models at large fetches) hold true regardless of whether we analyze the data in terms of arithmetic or geometric means (and associated uncertainties). This suggests insensitivity in our conclusions to the exact details of the statistical analysis.

With the goal of concisely expressing a simple, but scientifically reasonable statistical approach, we have taken your suggestion and simply said that our ratios are asymmetrically distributed, and have used a logarithmic transformation of our data (line 102):

> ***2.3 Statistical Analysis***
>
> *The accuracies of the footprint calculations are evaluated from the ratio of the model calculated emission rate to the actual release rate: Q_{KM} / Q and Q_{LS} / Q. These ratio data are asymmetrically distributed, and a logarithmic transform of the ratios is used when making our statistical comparisons. Thus, the geometric means of the emission ratios is our measure of central tendency. Confidence intervals for the geometric mean are calculated using the log-transformed ratio data, and then converted back to ratio units (Limpert et al., 2001). The confidence intervals (CI) are asymmetrical, and we report the upper and lower limits of the intervals.*

6) In the Introduction or Discussion sections, you may want to include the following references to other studies using an artificial source to test footprint models:

Heidbach et al.: Experimental evaluation of flux footprint models. Agricultural and Forest Meteorology, 246, 142-153, 2017.
Kumari et al.: Sensitivity of Analytical Flux Footprint Models in Diverse Source-Receptor Configurations: A Field Experimental Study. JGR Biogeosciences
125(8),e2020JG00569, 2020.

This is a good suggestion. It shows that our experimental design follows a well-accepted approach. We have added the Heidbach et al. reference as suggested, and have also added a reference to a field study conducted by some of our team:

> *"This field study compares the accuracy of the KM footprint model with a more rigorous LS model. The motivation for this study was the question of whether the accuracy of the LS model was sufficiently better than the KM model so as to justify a more complex LS application. In this experiment we released gas at a known rate from a small synthetic area source and measured the vertical gas flux at a downwind location using the eddy covariance technique. The KM and LS models were then used to calculate the source emission rate from the measured atmospheric flux. The accuracy of those calculations is examined in this report. This follows the approach of Heidbach et al. (2017) and Coates et al. (2017) in their experimental evaluation of footprint models."*

With the addition of the Heidbach et al. reference, we took the opportunity to acknowledge one of the results demonstrated by Heidbach et al. in our results section (line 122):

> *"Based on the calculations of Wilson (2015) and Heidbach et al. (2017), we had hypothesized that there would be substantial differences between the two models at the shorter fetch, with the LS model being more accurate than KM due to a better representation of horizontal turbulent transport, which is particularly important for defining the footprint at short fetches. However, this is not the case in this study."*

7) Figure 2: Indicate in the figure caption the number of data (n) in the three classes.

The caption for Fig. 2 now includes information about the number of data in each of the three classes, as noted below (lines 233-236).

> *"Figure 2: Agreement ratio of the footprint model calculated emission rate (Q_{model}) to actual release rate (Q), grouped by source fetch of 15 m (n = 26), 30 m (n = 9), and 50 m (n = 24). Calculations are from the LS and KM models. The columns show the geometric mean, and the error bars show the 95% confidence interval of the mean. The horizontal dashed line represents a Q_{model} / Q ratio of one, or a perfect model calculation."*

8) Add a (short) caption to the Table in the supplementary material

The following title has been added to the supplementary material.

> **"**Supplementary Table. Final data set used for footprint analyses following application of quality control criteria."