Reply to the referee comments by Auke van der Woude

We wish to thank Auke van der Woude for his careful review and helpful suggestions to improve the manuscript. Our answers and proposed changes are given below with the original comments printed in black and our replies in blue.

General comments

I think this is a very interesting and much needed research, helping the scientific measuring and modelling community. The paper discusses the potential uses of flask measurements for C14, measurements that are getting more and more important in climate research and proposes a sampling strategy for these measurements. The proposed sampling strategy for these flask measurements is justified with extensive research. To my knowledge, such a sampling strategy has not been created in such detail and with such justification before. Therefore, I recommend that this paper is published with minor revisions.

Specific comments

Especially in the introduction, the English is unclear and needs a revision.

We will check the English language (again) in the revised version of our manuscript and re-write the introduction.

Throughout the paper, the aims are discussed shortly thrice. One, more elaborate explanation would be better for me.

There are many different aspects concerning the sampling strategy in the ICOS network and we therefore feel that it may be helpful for the reader to remind him/her shortly of the specific aim before we discuss the particular aspect in detail. We can understand that this reviewer may like to have a more elaborate general introduction into the topic. For this we would, however, like to point him to the related literature, which we will extend.

Nevertheless, we will expand the first part of Sect. 3 in response to the reviewer's request.

The introduction is unclear. In line 40-45, it is unclear why the measurements by Keeling are mentioned. Also a link between marine and terrestrial measurements is needed. The first two paragraphs should therefore be rewritten.

We do not understand why we should NOT give credit here to the pioneering work of Dave Keeling !

We will re-write the introduction and make a better link between marine and terrestrial measurements.

In Section 2.3, only STILT is mentioned. For stack emissions, STILT could introduce biases due to the representation on a grid. Therefore, a plume model might be needed for emissions from e.g. power plants. I would like to see a few sentences added that discuss this, either in Sect. 2.3 or in a discussion section.

We mention here only STILT because this is the model we used for this study. We will add a short discussion on the problem of simulating correctly the influence of elevated point sources in Sec. 4.3.

In line 246, the '1/t filling approach' is mentioned. What is meant is unclear and should be explained or a reference should be added.

The 1/t filling approach is detailed in the reply to referee 1. We will give more information on this in the revised manuscript.

Line 290-295: Nighttime transport models are very erroneous. However, the integrated flask samples are filled for two weeks. Does the limited capability of the nighttime transport not limit the usefulness of these integrated samples for modelers (as described in the aims), if they are also filled during night. (c.f. Line 294-295). I would like to see explained why the flasks are not only filled during well-mixed conditions.

This is a very good point. Currently, we do the two-week integrated sampling day and night because this gives us a representative "continuous integrated" measure of the REAL MEAN ffCO2 concentration at the stations, with the prospect to observe potential changes in the future. This sampling scheme was our historical way to monitor 14CO2 at European stations and also globally, and we wanted to be compatible with these earlier measurements. It was also not our primary aim to collect these two-week integrated high volume samples in a way to best serve the modelling community, but rather for long-term monitoring purposes (although we are well aware of the fact that at sampling levels below 100 m height, the results would be biased towards night-time concentrations). We are still very hopeful that atmospheric transport models will improve in the near future and will then be able to "digest" also night-time observations. For these future times we will then have our observations ready to be used for validation of emissions changes. Still we always welcome the input from the experts, who tell us, e.g. via Observing System Simulation Experiments, how to optimise ICOS 14CO2 sampling strategies.

Line 420 needs additional explanation: why are conditions with low ambient variability best suited to meet aim 1? It states that this is explained in the previous section, but this is unclear. Therefore, I would like to see this explicitly explained in the previous section.

We would like to point the reviewer to Figs. 3 and 4 (left sides), which explicitly show that the agreement between 1-hour integrated flasks and in-situ measurements decreases with ambient variability. The reason for this is most probably that synchronization and weighting of flask filling and in-situ measurements (both collecting air through different intake lines and with different flow rates) is not perfect. These parameters are not so important when ambient variability is low; this was explained in section 4.1 (lines 379 ff).

In Section 4.2, only the results using flask measurements at 13.00 LT are shown. However, afternoon mixing conditions persist through about 16h in winter. Why are the flasks only filled at 13h? An analysis with the footprints for other times in the afternoon would add much information. I would like to see a paragraph (possibly including a figure), explaining the differences between sampling at 13h and other afternoon hours, and an explanation on why 13h is chosen.

It is correct that afternoon mixing conditions persist longer than one hour and flasks could potentially be sampled any other afternoon hour. Choosing 13 h LT for footprint analysis throughout the ms was just for consistency reasons, but it could have been any other afternoon hour. We do not agree that footprint analysis for other afternoon hours would add much information to this sensitivity study and in view of the aims of flask sampling at ICOS sites. Station PIs are free to choose a different afternoon hour or change the time every day. We will add a respective remark.

In section 4.3, it is explained that mostly the winter C14 measurements are of interest. However, in winter time, the biosphere fluxes are small. Contrary, in summertime, the biosphere is very active and partitioning between biosphere and anthropogenic fluxes is very hard. C14 can help with this. Therefore, summertime flask samples are also very informative. I would like to see a sentence explaining, and possibly countering, this limitation.

The limitation of summertime 14C-based ffCO2 estimates is simply current measurement precision and signal strength during summer. We will add a respective sentence in the conclusions. (See also reply to referee 1 who raised the same concern.)

Technical corrections

r. 37: How often can these events be expected?

We will add a range here.

r. 53: The mentioned process understanding needs more elaboration.

We will elaborate this in the revised manuscript (e.g. biospheric functioning)

r. 50: These fluxes however, . . .

Fine with us.

r. 81: ICOS flask sampling strategy might change in the future

Fine with us.

r. 224: six

Fine with us.

r. 350: in-situ

Fine with us.

r. 369: suited to meet aim 1 (ongoing quality control at class 1 stations).

Fine with us.

r. 420-421: Replace the could

We will change the sentence to "In the preceding section we showed that low ambient variability situations...."

r. 427: As expected, the regional coverage . . .

Fine with us.

r. 493: In this summer month: What summer month?

July (still related to Fig. 8)

r. 518: The mix of abbreviations and full names is confusing. A table with full name, location and abbreviation would help

We will replace the station acronyms by names throughout the text and in particular in line 518.

r. 562: working successfully

Fine with us.

Figures and Tables:

Table 1: This table is very full and therefore it is hard to find the needed information. A histogram might be more intuitive. This, however, is up to the authors to decide.

Good suggestion, we will try a histogram.

Figure 1: For better overview, it might be useful to indicate the sites in this figure.

Good suggestion, this should be possible to add (at least station abbreviations)

Figure 2: The photograph of the sampler is superfluous. More explanation on the schematic is needed in the caption/text.

We would like to keep the photograph (it does not need to be large but gives a good impression how the thing looks like), but add more information in the Figure caption for the schematics.

Figures 3 and 4 should be combined.

Yes, can easily be done.

Figure 5: The title above the subfigures is very small and therefore unreadable.

Will be combined to one title per row.

Figure 6: The font and subplots are too small to read. It is also unclear what the main message of this figure is.

We are sorry that the reviewer did not get the main message of this Figure as it presents the overall outcome of our representativeness analysis (aim 2): Aiming only for low variability situations for flask sampling creates serious biases in the data as shown in red in the lowest panel for each site. (Besides that we wanted to present the first years of continuous ICOS data that we are all proud of!) We will change the Figure and use the empty space in the lower right field for an enlarged legend (as it is the same in all 5 panels).

Figure 7: The x-axis could do with only one scale, also increasing the amount of space for the figures. Same goes for the y-axis

We prefer not to change that Figure because individual panels are easier to read with individual axes. Space is not a problem here (contrary to Fig. 6).