Review of the submitted article:

Towards an understanding of surface effects: Testing of various materials in a small volume measurement chamber and its relevance for atmospheric trace gas analysis

Ece Satar, Peter Nyfeler, Céline Pascale, Bernhard Niederhauser, and Markus Leuenberger

General comments:

This paper describes a series of experiments aimed at comparing the adsorption of some atmospheric trace gases in various materials. Experiments were planned in a very structured way to allow meaningful observations. The study was part of a larger study on adsorptions, and it is clear that authors chose to limit this paper to one main variable: the surface material. A good number of different materials were chosen, and all of them appear to be of interest to the community. The paper is generally well written, well-structured, clear, and provides a number of details on the instruments and methods, with some further details missing.
However the discussion part of the paper is quite limited. The results need to be put in perspective with other published work, in particular on CO_2 with aluminium. It also misses explanation and assumptions on the phenomena at work. Previous work by Leuenberger included a complete model with an attempt to fit the results during similar experiment in large cylinders. This paper should at least summarise this effort and explain if such attempt was also made here, and why it does not appear.
Considering the type of comments provided below, I recommend a major revision before the paper can be published.

Specific comments by section:

Section 1. Introduction:

- Cylinders volume in this study compared to others: the introduction mentions this difference but does not state the potential impact on observations. In Schibig 2018 it is explained that cylinders smaller than 30 L should present larger effects, due to the surface to volume ratio. This should be observed and reflected through the introduction and the rest of the paper, in terms of the magnitude of observed effects compared to cylinders commonly used as standards.

Section 2.1:

- Small chambers in aluminium and steel cylinders were designed, but this study only reports observations with the aluminium cylinder. The rationale behind this choice should be added. Was it after the conclusions of the other paper?
- The analyser is mentioned line 27 without a description. Please add the model and the performances which are relevant to the study. In particular one needs to pay attention to the sensitivity for the compounds measured, to demonstrate that observations are meaningful (or not). The claimed repeatability of the instrument appears sometimes on the same order as the changes measured during the study.
- Compressed air used for the study: more details on the composition are clearly needed, at least nominal values provided by the company. The water content in particular is under question, as some of the observed differences are of the order of 70 µmol/mol. Does this mean the water amount fraction was even higher than this? This is important to clarify, considering that the work of Brewer et al. showed how water can be preferably adsorbed on surfaces, decreasing the adsorption of other compounds such as CO_2.

Section 2.2:

- Flow rate: previous studies of Schibig et al. and Brewer et al. mentioned an influence of the flow rate at which cylinders are being emptied. How was this taken into account? What was the flow rate during the measurements? Some consideration on this point should be provided.
- Pressure values during temperature studies: table 1 shows that the pressure could change when changing the temperature. Consider explaining the reason and potential impact on the results.
- It is explained that in this study, all reported values are in amount fraction difference. It can be assumed that this choice was made to plot all data together and be able to compare different observations. This might be a good reason, but absolute values should also be provided, at least once, to be able to compare the results in this study with others.
- Temperature cycle: please clarify that the container was refilled at the beginning of each new temperature step. This information could also be added on figure 3.

Section 3:

- It is said several times that changes observed with CO_2 are significant only for Dursan. However differences of the order of 0.15 µmol/mol were observed with other materials and this is comparable with the compatibility goal within GAWG. In other studies similar changes were not considered negligible. Some analysis in view of already published work should be added and made more consistent.
- The "empty" case needs further clarification. First on the term itself which is badly chosen as the container is always filled with gas. Second on the values compared to the other paper of the authors. They are apparently those of the case "aluminium, filled at 14 bar, after heating". This should be clarified and related to the choice of this material (best results?).
- The "steel" results can be confusing when compared to the other paper, where a difference of 0.5 µmol/mol was observed. The setup is of course not the same, but this would need some consideration and some assumptions to explain those discrepant results.
- Results on water: figure 4 shows up to 60 µmol/mol changes, which would mean quite large water content to start with. Was it the case? If not, where does the water come from?

Comments on figures:

Figure 5: consider splitting in different figures to allow a larger front. This is currently far too small.

Figures 6: the *x*-axes is very disturbing, even after the highlight in the text noting that it indicates the temperature cycle, which is why the scale is not linear. There is some logic in this choice, but it discards the possibility to clearly see the temperature effect. Consider plotting with a linear temperature scale using a color code or different shapes to show the time. Another option is to use time as x-axes and superpose the temperature cycle.