Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2018-22-RC2, 2018

Atmospheric Measurement Techniques Discussions

Interactive comment on "NDACC harmonized formaldehyde time-series from 21 FTIR stations covering a wide range of column abundances" *by* Corinne Vigouroux et al.

Anonymous Referee #3

Received and published: 26 March 2018

This study presents the retrieval settings of formaldehyde from ground-based FTIR solar spectra, which has been harmonized to allow for consistent retrievals at various stations, under various conditions (remote area, polluted sites, high-altitude sites...). An error budget is presented for each station. The formaldehyde times-series are then presented along with a preliminary investigation of trend and diurnal cycles. Finally, the consistency of the FTIR products is evaluated via comparison with formaldehyde columns simulated by a chemistry-transport model.

Developing harmonized formaldehyde retrieval settings through the NDACC (and future affiliated stations) is quite challenging because of the weak absorptions of HCHO

in the infrared and the many interferences. This work is therefore valuable in the framework of validation efforts of model simulations and of current and future satellite instruments. The topic developed here fits perfectly the scope of AMT. The paper is globally well written and the structure is clear. Nonetheless, some results/figures are not adequately presented, which somehow impedes a proper evaluation of the results (see major comments here below). Therefore, I recommend publication of this study after addressing the comments listed hereafter.

Major comments

Additional effort is needed to present the results more synthetically and to make some figures easier to read. Fig. 5-8 are particularly difficult to read because of the numerous small panels. It is really unfortunate because these figures present the main results of the study. I assume that the large number of subplots makes them difficult to display, but I really think that such figures deserve a reshape. In particular:

- The various x-axes in Fig. 5 and 7 don't help the reader. Some seasonal cycles appear completely squeezed because these panels encompass >15 years, while others represent 2-3 years only on a panel of the same size. Please display the time series on an x-axis common to all the stations.

- For Fig. 5 and 7, I also suggest to gather some time-series within the same panels, e.g., following the subdivision in the text (i.e. clean, intermediate, polluted sites), using different colour lines. The interest to gather time series within the same panels would also be to help the reader to appraise the large panel of HCHO columns covered by the FTIR sites.

- For Fig. 6, perhaps it is not needed to display all the diurnal cycles in the manuscript. A solution would be to keep here only a few representative of those to support the discussion. The others can be moved to supplementary material.

- In Fig. 7-8, it is very hard to distinguish the raw model data from the smoothed ones.

I also find the discussion on the basis of Table 4 quite “raw”. The authors made huge efforts to harmonize the retrievals and to produce a consistent pattern of HCHO measurements worldwide. There could be, along with Table 4, a map including in colour background the mean HCHO from the model over 2003-2016, and the mean FTIR HCHO in dots filled following to the same colour bar, at the location of each station. In one glance, the reader would have a good overview of the pattern of FTIR measurements as well as of the overall consistency with the model.

A single HCHO a priori profile is used at each station for the retrievals. This assumes that not only the shape, but also the HCHO concentration simulated by the model, are quite reliable. What is the impact of another HCHO a priori profile on the retrieved columns? e.g., an a priori from another model that would be significantly different, or again an a priori that would be derived from other measurements (like ACE-FTS)? For a very weak absorber like HCHO, with little retrieved vertical information, I expect the impact to be, if not substantial, at least not negligible. For example, it is clear from the very similar shapes of the a priori and retrieved profiles in Fig. 4, that the retrievals are dependent on the a priori. It is important to discuss this point and to add this component to the error budget.

The error budget is established for each station on the basis of a single measurement. Why one measurement only per station? I find this very reductive, especially that it is not even said whether this single measurement is representative for the whole data set (DOFS, residuals, total column...). Hence, one can easily casts doubts about such error budget. This should be made ideally with a representative subset of FTIR measurements, covering different seasons, different zenith angle, etc.

Page 14, Lines 9-16: Since you know that models usually underestimate the natural variability of HCHO, and since you know the impact of such underestimation on the smoothing error estimation, why wouldn’t you increase the variability of the model to be more representative for the real variability? Knowing the difficulty for the models to simulate highly-variable reactive gases like HCHO, a model variability multiplied by

e.g., 2, would still be conservative.

Page 15, Lines 15-27: The diurnal cycle is sometimes very weak. Furthermore, the midday observations (low zenith angle) probing less atmosphere, we can expect larger uncertainties associated with such measurements. Hence, owing to these larger midday uncertainties, are the diurnal cycles still significant? Or couldn’t some of these diurnal cycles (e.g., the midday minimum found at some sites) be just the effect of larger uncertainties and less sensitivity associated with the low zenith angle?

Section 3.3: The investigation of the trends are very preliminary, and there is no discussion of the results. If, as quoted in this section, a more comprehensive investigation is beyond the scope, I don’t see the interest of this section as it is currently. Or I recommend to add a bit of discussion, e.g., how do significant trends compare to other trends in the literature from, e.g.:

- De Smedt et al. (2010) Trend detection in satellite observations of formaldehyde tropospheric columns, Geophys. Res. Lett., 37, L18808, doi:10.1029/2010GL044245.

- De Smedt et al. (2015) Diurnal, seasonal and long-term variations of global formaldehyde columns inferred from combined OMI and GOME-2 observations, Atmos. Chem. Phys., 15, 12519–12545, doi:10.5194/acp-15-12519-2015.

- Franco et al. (2016) Diurnal cycle and multi-decadal trend of formaldehyde in the remote atmosphere near 46°âĂŕN, Atmos. Chem. Phys., 16, 4171-4189, https://doi.org/10.5194/acp-16-4171-2016.

- Jones et al. (2009) Long-term tropospheric formaldehyde concentrations deduced from ground-based fourier transform solar infrared measurements, Atmos. Chem. Phys., 9, 7131–7142, doi:10.5194/acp-9-7131-2009.

There have already been comparisons between previous HCHO columns from the FTIR and from UV-Vis instruments (satellites and MAX-DOAS), showing in overall a good agreement. However, it is obvious from this study that the HCHO retrievals from

the FTIR are very sensitive to the retrieval choices (spectroscopic database, micro-windows, a priori. . .). Biases up to 50 % are even mentioned between different retrieval approaches. This means that the harmonized retrievals presented here can potentially improve or deteriorate significantly the comparisons with the UV-Vis instruments. I think this point needs to be discussed, or at least mentioned in the conclusion.

Minor comments

Page 1, Line 3: “accurate and precise”. In light of the error budget and the large biases depending on the retrieval choices (“as large as 50%”, Line 5), this statement should be dampened.

Page 1, Line 8: stations. Most of them

Page 1, Line 11: Change “;” to “,”

Page 2, Line 1: Unclear. Is it for the systematic or the random uncertainties?

Page 2, Line 11: NOx is not defined yet

Page 2, Line 16: of only a few hours

Page 2, Line 17: and to test

Page 2, Line 32: at a few locations

Page 3, Line 3: Change “geographical” by “spatial”

Page 3, Line 3: A lot of efforts are

Page 3, Line 7: stations that will also be part

Page 3, Lines 13-16: Isn’t it because HCHO is so challenging to retrieve that it is not (yet?) a standard product from the NDACC FTIR?

Page 3, Line 27: monthly mean time-series

Page 4: Fig. 1 would deserve to be a bit larger due to e.g., the concentration of stations

in Europe

Page 4, Line 14: (1995) and/or Hase et

Page 4, Line 15: pressure- and temperature-dependent

Page 5, Lines 5-6: It is said elsewhere in the manuscript that the use of different retrieval parameters can substantially affect the retrieved columns. In particular, the use of either HITRAN 2004, 2008 or 2012 for the HCHO spectroscopic parameters leads to very large differences in the retrieved columns. I would have expected the authors to better motivate their choice of HITRAN 2012, especially that eventually some lines of interfering species needed to be empirically adjusted in this spectroscopic database.

Page 6, Line 3: that is distributed

Page 7, Line 8: Rodgers (2000)), is

Page 7, Lines 6-9: Is the little gain in information the only reason why you keep these two windows?

Page 7, Lines 10-11: Is this individual spectrum representative for the whole time-series? How do its DOFS and its fitting residuals compare to the other observations?

Page 7, Line 26: WACCM v4 (Garcia et al., 2007).

Page 7, Line 29: not to use

Page 7, Line 33: Sussmann et al. (2011), and

Page 8, Fig. 2 caption: total column of

(the same in Fig. 3 caption)

And further, same line: Change “The figures in the lower panel are” to “The lower panels are”

Page 10, Line 31: DOFS, in Table 3

Same line: provide more than

Page 10, Line 33: (upper panels)

Page 11, Fig. 4 caption: Upper panels: averaging

And further: Lower panels:

Page 11, Line 5: associated with

Page 11, Line 6: (lower panels)

Page 11, Fig. 4: Do you have an explanation for the contribution from the high-altitude averaging kernels (in green)? It looks a bit odd to have a contribution from such layers to the HCHO retrievals. Still Fig. 4: From the shape and the high values of the total column averaging kernel, don't you "overfit" the HCHO retrievals?

Page 12, Table 3, as well as in the manuscript: Providing the random uncertainties in total column only is really misleading, especially in the discussions. Each time, it forces the reader to look at Table 3 and to calculate the percentage before knowing whether it is significant for the station that is considered. I recommend to provide all the uncertainties (also) in percentage of the total column (you already do it for the systematic uncertainties).

Page 13, Lines 25-29: Why such exceptions for these SFIT4 stations, and not for others? Isn't the error budget supposed to be fully harmonized among all the SFIT4 stations?

Page 13, Line 29: might have

Page 13, Line 33: stratosphere. This matrix

Page 13, Line 34: while for the PROFFIT users, these values

Page 14, Lines 3-4: rephrase as "considering the random uncertainty in Table 3 (4th column) is sufficient" to avoid misleading

Page 14, Line 24: HCHO line intensity.

Page 14, Line 26: the PROFFIT channelling source (from 7 to 17 %), which also has a systematic component. We see from Table 3

Page 15, Line 5: few 1 x 1013 molec/cm^2

Page 15, Lines 15-16: Bad sentence. "To reconcile the different ... (afternoon), it is crucial to have ground-based ...

Page 15, Line 21: mid-latitude cities

Page 18, Lines 1 and 5: The use of "time step" is here misleading with the computational time step of the model. I suggest to use "with outputs every 6 hours/ 20 minutes" instead.

Page 18, Line 11: delete one "the"

Page 18, Line 22: in Bauwens et al. (2016)

Page 19, Line 1: justified by the

Page 19, Lines 3-6: How do you deal with the model surface that is below the altitude of the station (which should be the case for most of the mountain sites), especially where there is a substantial altitude difference?

Page 19, Lines 7-9: Do you mean that you re-scale the model outputs at the time of the FTIR measurements? Or do you use the nearest model output to each FTIR data?

Page 19, Line 24: "within a month" is redundant

Page 19, Line 25: "The median of IMAGES and FTIR differences" or "The median of IMAGES and FTIR biases"

Page 19, Line 30: change ";" to ","

Page 21, Fig. 8 caption: "in coincidence with". Do you mean the same day, within 20

minutes of each FTIR data?

Page 23, Line 13: Is Boulder ($\sim$1600 m asl) really a urban site?

Page 23, Lines 29-30: At such a remote site, the dominant source of HCHO should be CH4 oxidation. I don't think that other sources from continental areas can be significantly at play here.

Page 24, Lines 5-9: Could it be also due to the FTIR technique, which measures in clear-sky conditions only? The FTIR would sample only air masses free of huge emissions and hence would underestimate the gas abundance in this region.

Page 25, Lines 14-15: We do not aimed at evaluating the model, but at showing that the FTIR

I have here an open question, which I think is relevant for a data set that is designed to be used for intense model and satellite validation efforts. Will this data set be made publicly available? And if yes, will there be an effort to fully harmonize the archives, the file format, and the way the data are saved? There are currently inconsistencies between FTIR data sets from different stations (especially for the AKs). Such inconsistencies sometimes refrain external users to use NDACC FTIR data, while such data sets deserve to be easily accessible and as user-friendly as possible for non-community users.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2018-22, 2018.