

General Comments

This paper presents for the first time the time series of the vertical profiles and total column amounts of one of the most important gases on the atmospheric chemistry, ozone (O₃), for Latin America. The strategic location of the ground-based stations used and the methodology proposed, using high-resolution and medium-resolution Fourier Transform infrared (FTIR) solar absorption spectra, provide high confidence to the results obtained. Especially interesting is the presentation of an improved tropospheric ozone product from the combination of two remote sensing FTIRs data, which allows for a better monitoring of tropospheric ozone concentrations. The paper is well-written and concise. Thereby, I suggest this paper may be suitable for publication after addressing the specific comments listed below.

Specific Comments

Section 2: Ground-based FTIR remote sensing

a) The authors mention that the microwindows used to retrieve the O₃ concentrations are the same windows as presented in Schneider and Hase (2008), but this is not strictly true. Schneider and Hase (2008) suggest a broad microwindow between 1000 and 1005 cm⁻¹, which has been split into two microwindows in the current work. I guess that the authors did this to avoid the interference of water vapour (H₂O). If so, please clarify in the text. Also, the authors include the 1005-1006cm⁻¹ microwindow, not present in Schneider and Hase (2008). So, I think it would be more appropriate to say that the selection of the O₃ microwindows is based on Schneider and Hase (2008) and clarify why the authors modify the original microwindows.

Yes, the referee is right. We will correct and clarify this in the revised version. Thanks!

b) The error budgets clearly show that the atmospheric temperature profile is an important error source for Alzomoni and UNAM FTIRs as well as for the combined product. Simultaneous temperature retrieval with O₃ profile has widely demonstrated that improves theoretically and experimentally the quality of the FTIR O₃ products (e.g. Schneider and Hase, 2008, Schneider et al., 2008, García et al., 2012). Have the authors considered performing this temperature retrieval? Why not?

Simultaneous temperature retrievals can improve the quality of the retrieved O₃ data using high resolution spectra and spectral windows in the 1000cm⁻¹ region. This has been shown almost 10 years ago (Schneider and Hase, 2008) and has been confirmed by other studies (e.g., García et al., 2012). There are two reasons why for this study we decided not performing simultaneous temperature retrievals. First, for the retrieval setup that uses the medium resolution spectra, so far no simultaneous temperature retrieval study has been performed. So we would need to do such study in this paper (which would be an extra paper and it is out of the scope of the here presented work). Second, simultaneous temperature retrievals mean an advanced retrieval methodology (for instance, the simultaneous temperature fit interferes with ILS uncertainties) and for a first study with the Alzomoni data, we prefer using a more common and less sophisticated retrieval methodology. Actually, despite the clearly demonstrated benefit of the simultaneous temperature retrievals, it has so far only been used in a rather limited number of studies, mainly those made with the FTIR instrument at the Izana Observatory. The

advanced retrieval methodology has for instance not been used for most of the retrievals, whose data are presented in the trend studies as shown in Vigouroux et al. (2015).

c) The authors have assumed the same error in the ILS for both FTIRs (5% for the modulation efficiency and 0.1 rad for the phase error). But, as for the measurement noise, I guess that the ILS's error of the medium-resolution FTIR should be a bit higher than the high-resolution one. Also, the ILS's errors could explain in part the large systematic/statistical errors observed in the combined product. Why the authors use the same value? Why the authors do not consider the ILS as a possible explanation for the large errors found in the combined product? Could the authors show a plot with the time series of the ILS for each station? What is the frequency of the cell measurements at each station?

The medium resolution spectra are measured with a maximal optical path difference (OPDmax) of 9cm, whereas the high resolution spectra are measured with an OPDmax of 180cm. So the requirement of the modulation efficiency being stable within 5% along 180cm is significantly stricter than the requirement of being stable within 5% along 9cm, only. So we actually assume that the high resolution FTIR is much more stable than the medium resolution FTIR. The uncertainty values of 5% are estimated according to different cell measurements made at Altzomoni and UNAM.

We are aware that ILS monitoring is very important, especially for retrieving stratospheric trace gases with high precision. In this context the work done at Izana (see for instance the ILS time series as shown in García et al., 2012) is a nice reference and can serve as guideline for other stations. We strongly support the idea that publication of FTIR data should be accompanied by a documentation of the ILS. In the Appendix we will show ILS retrievals and respective averaging kernels and hope that this will become a standard for future publications of data from new stations (the following Fig. 1 will be added in an Appendix of the revised paper).

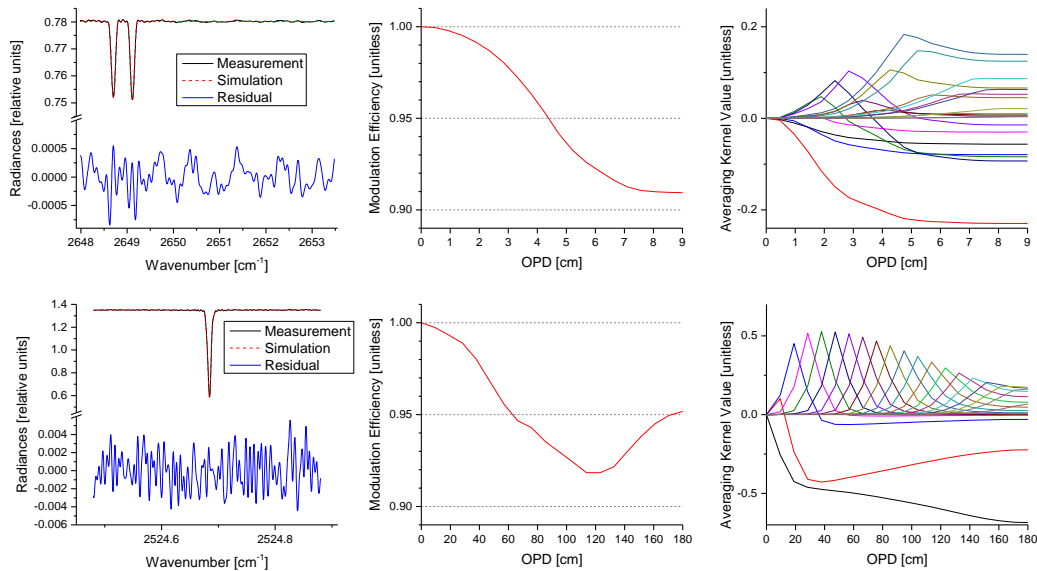


Fig. 1: Cell measurements (examples of spectral windows) and corresponding ILS retrieval results (modulation efficiency) for the medium resolution instrument at UNAM (top) and the high resolution instrument at Altzomoni (bottom).

Referee #2 also commented on the ILS at UNAM. For simplicity we have used nominal ILS during the retrieval process, however, the referees are right and the loss of modulation efficiency is also not negligible for the medium resolution instrument. So we will redo the UNAM medium resolution retrievals using the measured ILS.

d) How the layers that are detected by each FTIR are defined? Are the consecutive levels added until a DOF of one? There is a high signal in the UNAM averaging kernels at about 35 km, what is the reason?

We are not sure if we correctly understand this comment. We look on the full averaging kernels matrix not only on the diagonal elements (the diagonal elements of the averaging kernel matrix are used for calculating the DOFS).

Does the referee ask about our criteria for choosing the altitudes 4, 17, 28, and 42 km (for Altzomoni) and 2.3, 17, and 32 km (for UNAM)? We simply chose these altitudes by looking on the row averaging kernels (Figs. 3 and 5), because the row kernels reveal that the data retrieved for these altitudes reflect rather different altitudes (there is no significant overlap of the respective kernels).

The averaging kernels depend on the quality of the measured spectra (spectral resolution and signal-to-noise ratio) and the kind and strength of the constraint. They document how the remote sensing system interprets real atmospheric variability.

Section 3: Free tropospheric and stratospheric O₃

a) I am wondering if smoothing the Altzomoni averaging kernels with the UNAM ones makes sense when the difference of the total DOFs is only of one. Would it be more appropriate to estimate the altitudes that are well comparable analyzing the square root of the diagonal of Scmp as presented in Wiegeler et al. (2014) and then compare the original volume mixing ratios? How are the correlations shown in Fig. 7 for the not-smoothing data?

This is a good point and we were also concerned about that. So we compared the raw row kernels and the row kernels after smoothing and found that smoothing is an advantage by making the kernels better comparable (see also the following Figure).

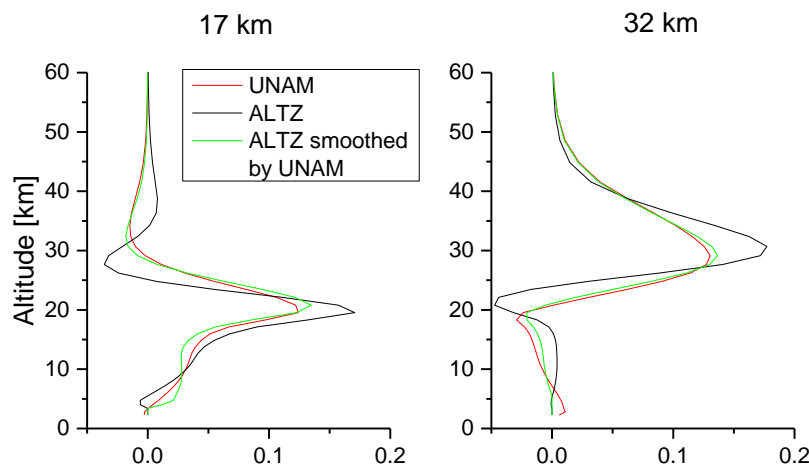


Fig. 2: Row averaging kernel, examples for 4, 20 and 30km, for UNAM (red), Altzomoni (black), and Altzomoni smoothed by UNAM kernels (green).

b) The authors use the works of Thompson et al. (2008) and Emmons et al. (2010) to explain in part the observed O₃ annual cycle of the free troposphere. These works point out the importance of the transport of the urban emissions on the free tropospheric O₃ levels. But, according to the authors in page 7, line 32, “the Altzomoni solar absorption spectra are only very weakly affected by the polluted boundary layer”. So, I find both statements contradictory, but maybe I am missing something. Have the authors considered that the maxima in spring-summer could indicate the importance of photochemical production of tropospheric ozone?

Yes, the increase of free tropospheric O₃ background in spring is partly connected to O₃ pollution transported from the boundary layer of Mexico City (see discussions in Thompson et al., 2008 and Emmons et al. 2010). In the data of our paper we see an increase in the peak values of the surface in-situ data between January and May (see Fig. 8). This is consistent to the increase observed in the Altzomoni retrievals at 4km (see Fig. 6). However, the dominating boundary layer signal has a diurnal time scale. For this reason we observe no significant correlation between the coinciding surface in-situ data and the Altzomoni retrievals at 4km (see left panel in Fig. 9). So the Altzomoni data do not reflect this strong diurnal cycle, especially since the measurements are made mainly before 14:00, when the boundary layer top altitude is below or only slightly above the altitude of Altzomoni. Nevertheless the 4km O₃ background levels likely depend on the outflow of O₃ pollution from the boundary layer. We will improve the respective discussion in the text.

c) The O₃ annual cycles at Altzomoni are compared with other NDACC stations such as Izaña or Jungfraujoch. But, there is another NDACC site at equal latitude and about the same altitude, Mauna Loa. Have the authors compared the Altzomoni and Mauna Loa FTIR O₃ products? This could be a very nice cross-validation of the Altzomoni O₃ data, especially for the stratospheric values.

In Sect. 3.1 we mention in one sentence that there are similarities in the annual cycles between Altzomoni (on the one hand) and Izaña and Jungfraujoch (on the other hand). The idea is to give references to studies that make scientific analyses of similar FTIR O₃ data. We mention these two stations because there are a lot of publications available. We can in addition search for publications with Mauna Loa FTIR O₃ data and cite it in the same sentence.

A comparison of time series obtained at different stations or a discussion of model measurement differences is certainly interesting. However, this should be in our opinion addressed in an additional ACP paper (instead of AMT). Here we focus on the retrieval technique and data validation.

d) The authors estimate the Scmp to filter out the data using a value of 10% for the whole range of altitudes. But, this threefold could be altitude-dependent as suggested in Fig. 7, where a worse correlation is observed for the upper level. Plots showing the Scmp vs correlation and the profiles of the square root of the diagonal of Scmp could help to analyze better the comparability of the two remote sensing instruments. See comment a) related to use the Scmp only for filtering the data and not to look for the altitudes for direct comparison.

We set up a Scov having $100\%^2$ in the diagonal. Then we require that for a meaningful comparison the diagonal elements of Scmp should be smaller than $10\%^2$. Thereby we require that as a maximum 10% of the variance between the two data can be explained by different smoothing characteristics. The rest of the scatter can be explained by errors. We apply no extra

filter for errors, because we want to compare all data that can be reasonably compared (by having similar smoothing characteristics) and not only data that have a particularly high quality (small errors).

Section 4: Boundary layer

a) The theoretical error budget for the combined product is not in agreement with the experimental comparison (higher statistical errors and lower systematic ones). Also, in the abstract the authors mention that the combined products offer theoretically and experimentally better results. This is true for the sensitivity, but not for the error estimation shown, which could cause some confusion to the readers. The authors provide some causes to account for these discrepancies, but they have considered re-doing the error estimations with more realistic values. For example, is it enough to assume 5% and 0.1 rad for the ILS errors when the UNAM ILS is assumed as ideal?

Already in the abstract we will better specify what can be improved by using two measurements/retrieval products instead of one measurement/retrieval product.

We think that the assumption of a 5% uncertainty in the ILS is realistic, also for the medium resolution instrument (recall that OPD_{max} is only 9 cm for the medium resolution instrument). However, this is only true if the retrievals work with the measured ILS. So far we used a nominal ILS for the medium resolution retrievals. Figure 1 reveals a loss of the modulation efficiency for our medium resolution instrument of 9% at OPD_{max}, which has so far not been correctly accounted for in the retrieval process. For this reason we will redo the retrievals by using the actually measured ILS. This retrieval improvement will reduce the bias between the in-situ data and the combined product and bring the observed bias in better agreement with the estimated systematic error.

In Section 4.3.2 we discuss three possibilities for the bias with respect to the in-situ data: O₃ pressure broadening parameter, O₃ continuum absorption, and the surface in-situ data being a poor reference for the whole boundary layer. We agree with the referee that we should in addition consider ILS uncertainties of the UNAM instrument as possible reason for the bias.

We think that the next step must be to perform O₃ radiosonde measurements in coincidence with FTIR measurements in order to understand whether the bias between the surface in-situ and the combined FTIR data is an error or if it is actually real and can be explained by the fact that the surface in-situ data do not correctly represent the boundary layer.

b) As the authors stated, the combined product is very promising. I am interested in other possible applications of this product. Could it be used the other way round, ie, to reduce the tropospheric signal in the Atizomoni ULTS data? Or to improve the tropospheric sensitivity of space-based sensors such as IASI?

We developed the method in line with the Rodgers formalism and it can be used whenever the combination of two remote sensing data increases the knowledge of the atmospheric state. An example could be to combine thermal nadir satellite data (low sensitivity close to the ground) with ground-based solar absorption data (good sensitivity close to the ground) or to combine thermal nadir infrared and short-wave infrared satellite data. This is a good point and we can add a brief comment in the "Summary and Outlook" Section.

An improvement of the Alzomoni data by using additional UNAM data seems not too promising, because the Alzomoni FTIR instrument is located at 4km and not affected by atmospheric variations below 4km (where the UNAM measurement can provide additional information). This means that the additional information given by the UNAM data will not be too useful for the Alzomoni data.

Technical Comments

Page 5, line 12. Please include reference for the WACCM model version 6.
OK!

Page 7, line 14. Replace “in temperature” by “temperature”.
OK!

Page 10, line 8. Replace “subtropical” by “sub-tropical”.
OK!

Page 12, line 5. Replace “in situ” by “in-situ”.
OK!

Page 16, line 11. Replace “Eq. 10” by “Eq.(10)”.
OK!