Review of **Investigating differences in DOAS retrieval codes using MAD-CAT campaign data**, by E. Peters *et al*.

AMTD, doi:10.5194/amt-2016-358, 2016

General Evaluation

Peters et al. present a comparison study on NO2 slant column retrievals from a range of different DOAS retrieval codes on the same set of MAX-DOAS observations acquired during the 2013 MAD-CAT campaign. Results of NO2 columns and RMS values from the retrieval codes, which are run with a basic set of harmonized settings for molecular absorption cross-sections and closure polynomials, are compared and a range of possible sources for their differences is investigated. Based on this study, a short list of recommendations is given as general guidance for DOAS retrievals to obtain high confidence/low RMS retrievals.

The paper is solidly written, and there is little to criticize in methodology and overall quality of presentation. My main criticism is that, at 27 journal pages, the manuscript is overly long for a study that concludes with five basic recommendations. The paper provides excellent insight into the workings of DOAS retrievals, and as such is valuable for both data providers and data users, but this reviewer strongly suggests that the discussion be tightened and the main part of the message conveyed more concisely.

The meat of the paper is straight forward and relatively simple:

1. A set of MAX-DOAS observation from MAD-CAT was selected for a retrieval algorithm comparison.
2. A common set of basic spectral fitting settings was prescribed with which to run the retrieval codes.
3. Differences in NO2 slant columns and RMS were found, relative to a reference retrieval.
4. The reference retrieval code was run with modified settings for five essential code elements - radiance reference spectrum, slit function, offset correction, I0 correction, matrix inversion - to investigate their effect on the retrievals and to possibly explain the differences in the results.
5. The modified retrieval runs lead to the final recommendations, while the attempt to attribute differences between the codes to the investigated five sources is only moderately successful.

The majority of the "take home messages" comes from Bullet 4 above, but the attribution of those effects in the actual differences observed between the results from the various retrieval codes remains qualitative at best. With this in mind, any figures and discussions relating primarily to relationships between the results from different codes - in particular figures 4 and 6 and their discussion - are non-essential and should be marked for removal.

Below are some more specific comments. Very few of these are copy-editorial, since the level of presentation of this paper is very high.

Recommendation

The manuscript is acceptable for publication, but should undergo some tightening and add a few clarifications. Since there are no basic problems with methodology or presentation, a second round of review is not necessary.

Specific Comments

Retrieval Uncertainties

While the paper compares NO2 slant columns and retrieval RMS, no NO2 slant column uncertainties are shown. Purely spectral minimization-based uncertainties are a combination of RMS and fitting covariances, and thus provide important information on the quality of the retrieved slant columns beyond the RMS.

Reference Cross-Section Wavelength Scale
Three (admittedly very basic) questions regarding wavelength registration:

1. Does the IUPB MAX-DOAS instrument measure in vacuum or air?
2. Which wavelength registration (vacuum or air) was used for the retrievals?
3. Was it assured that all molecular and solar reference spectra were on the same type of wavelength registration as the MAX-DOAS spectra?

Slit Function
The measured slit function as shown in Figure 8 is slightly asymmetric. Yet, no attempts are reported of having fit an asymmetric Gaussian to the measurement for use in the retrievals. At least part of the comparison exercise utilized pre-convolved molecular absorption cross-sections, so this should have been an easy case to include. It is not very surprising that results from original and re-centered original slit function are virtually identical: the asymmetry should mainly manifest as a spectral shift, which is taken care of by the shift parameter during the retrieval process.
Regarding the differences introduced by removing the offset of 0.001: was the resulting slit function renormalized to the same area as the one with the offset?

Section 4 "Understanding differences between retrieval codes"
Ultimately, this is the most important section of the manuscript since it systematically investigates the effect of different fit settings on the retrieved slant columns and the resulting RMS. It is also here that the five recommendations in the Summary are derived. In principle, this exercise is independent of the MAD-CAT comparison. While the differences in results from the various retrieval codes are a good motivation to perform these tests, they are valuable in their own right, and more emphasis should be placed on this. By referring to this part of the study as "differences between retrieval codes", this reviewer believes that the importance of these tests is somewhat muddled and degraded. The reader would benefit from a clear statement of the type "differences between the harmonized MAD-CAT retrieval results prompted the team to systematically investigate effects of the non-harmonized aspects of the retrievals, with the aim to derive a key set of Best Practices recommendations". Since quantitative attribution of "what part of the differences originates from which non-harmonized retrieval setting" turns out to be unfeasible/unsuccessful, more emphasis should be placed on the derived recommendations for DOAS retrievals.

Editorial Comments
Line 122: suggest to reword as "real data without cross-instrumental bias", to avoid confusion with measurements free of instrumental bias.

Figure 1 caption: suggest to include "(90° = Zenith)" for the benefit of readers less familiar with MAX-DOAS observation methodology.

Figure 2: Use a different color for the fitted spectrum. Green and Blue are hard to distinguish.

Figure 2: What is the definition of "differential cross section", and is it optical "density", "depth", or "thickness"? None of these quantities would be expected have negative values, thus there has to be a reference point.

Line 138: "However, these are normally the ones of interest".

Line 171: "*r*, the root mean square (RMS) of the fit residual, is an important quantity used within this study to identify and evaluate differences between the DOAS retrieval codes."

Line 229: either "groups participating in MAD-CAT" or "participating MAD-CAT groups".

Line 305: delete "one" after "WCRS".