Response to reviewers for "The ACOS CO2 retrieval algorithm – Part 1: Description and validation against synthetic observations"

We'd like to thank both anonymous reviewers for their detailed observations and suggestions on our manuscript. We have responded to all questions, and implemented many of the requested changes. We have explained our reasoning where we have not implemented the changes. Overall, the changes to the manuscript were fairly minor, consisting of small changes to the text, as well as changes to some axis labels in figures 7 and 9.

Anonymous Reviewer 1

Comments

p.6106,l.8: While fitting the logarithm of the aerosol parameters avoids negative values, it makes the forward model significantly nonlinear. Did you check if this trick deteriorates convergence behavior or makes the retrieval get stuck in local minima of the cost function?

We've checked this to some degree. In general, the model does not do well using linear (rather than logarithmic) aerosol amounts, especially for our aerosol profile retrieval. Convergence behavior does not seem to suffer, though often the model converges to an aerosol regime relatively close to the prior. Overall, the logarithmic treatment seems to work reasonably well.

p.6114,l.22: Section 3.1 and the discussion of the retrieval performance in section 3.3 would benefit from a more detailed summary of the differences between simulation and retrieval method. How do the radiative transfer methods differ? How do the aerosol models differ (height distributions, sizes, refractive indices, non-spherical particles)?

This is a great comment – we've expanded the discussion in section 3.1 to highlight the most important forward model differences between the retrieval and simulation. The primary differences are the surface treatment (unpolarized Lambertian vs. polarized non-lambertian) and cloud+aerosol (4 simple types in the retrieval vs. dozens of cloud types and 6 aerosol types in the simulations). There are minor differences in the radiative transfer and treatment of Rayleigh scattering, and number of atmospheric levels (20 in the retrieval vs. greater than 100 in the simulations). In terms of attributing retrieval errors to forward model differences, this is in general difficult, but where it is possible, this has already been done in the text.

p.6118,l.25: What are the sources of the non-vanishing XCO2 and surface pressure error for retrieval test 1, "which is to be expected on simple theoretical grounds"?

The non-vanishing X_{CO2} error is directly caused from the surface pressure error. Because XCO_2 is essentially the ratio of the CO_2 column to dry air column, errors in surface pressure (which is nearly proportional to the dry air column) directly lead to errors in X_{CO2}:

$$\delta X_{CO2} = -\frac{X_{CO2}}{P_{surf}} \delta P_{surf}$$

Thus, errors are not linearly related, but for typical values of XCO2 (390 ppm) and surface pressure (1000 hPa), the proportionality constant is -0.4. It is larger in magnitude for lower surface pressures. Given that the surface pressure errors account for 75% of the variance in

the X_{CO2} errors, the main question for test 1 is, therefore, what leads to its surface pressure bias and errors? We don't know exactly the cause, but they seem to be related to minor differences in the radiative transfer and treatment of Rayleigh scattering between the simulation and retrieval; there could also be some minor interplay of the non-Lambertian surfaces in the simulator with the A-band Rayleigh scattering at play, though dedicated tests would be required to uncover this. Because the errors are much smaller than those associated with clouds and aerosols, we have not yet run tests targeting these specific differences.

We've added a couple of sentences in this section to make this clearer.

p.6121,l.10: The authors claim that there is a strong correlation between retrieved and true AOD for values <0.3. Fig. 8c does not support this and might hint at the retrieved aerosol parameters actually being pure correction parameters. Consider to use less strong wording eg. by replacing "strong correlation" by "some correlation".

Agreed. We've implemented this recommendation.

p.6121,l.26: The residual errors detected for test 5 come from the differences between the "true" and the "retrieved" scattering scenario. Positive bias over dark surfaces could for example be explained by the retrieval finding cirrus at higher altitudes than in the simulation. The paragraph reads like an explanation for test 4 errors ie. for a retrieval that entirely neglects aerosol and cloud scattering. Consider to refine the reasoning here.

The reviewer is absolutely right. We've modified the text in this section to read:
"The dominant atmospheric scattering mechanisms driving these biases are not yet evident. It is complicated by the fact that the algorithm simultaneously retrieves \chem{CO_2} and dry air column (via the surface pressure), so reasoning involving path-shortening vs. path lengthening effects must be broken down in terms of the different NIR bands. Biases can only occur when path shortening or lengthening effects are different between the \chem{CO_2} bands versus the \chem{O_2} A band. Future Monte-Carlo simulations may shed some light on the dominant mechanisms at work."

p.6124,l.16: The study finds that more than 10% of the accurate XCO2 retrievals correspond to scenes with true AOD>0.3. One of the quality filters screens all retrievals with retrieved AOD > 0.15 ie. retrieved and true aerosol scenario differ a lot. I would conclude that in these cases some lucky combination of surface albedo and mismatch between retrieved and true aerosol parameters yields small XCO2 errors. Thus, I suggest not to highlight this as a peculiar achievement of the retrieval method.

The fact that there are cases which satisfy all our post-screening criteria, and have good XCO_2 errors despite high aerosol optical depths, is by itself interesting and worth noting. You are right, there is definitely a lucky combination of retrieval parameters at play. We checked the results for these same ~100 soundings for the no-scattering retrieval, and the results were terrible. So a lucky combination of retrieved aerosol is certainly involved. However, we can turn it around and say, why aren't there cases that pass our screens, have high AOD, and have very poor XCO_2 results? The worthwhile thing noting is that our screens seem effective at getting rid of such cases, though how this is accomplished is unclear. We've added a couple sentences in the paper to this effect.

Technical comments

p.6099,l.25 (or at a more appropriate place): I suggest to refer to other retrieval algorithms that already demonstrated highly accurate CO2 (and/or CH4) retrievals from GOSAT, in particular since results from SCIAMACHY are actually cited. Consider Morino et al., 2011, Butz et al., 2011, Oshchepkov et al., 2011, and potentially Parker et al., 2011.

These references (except the Parker paper which is about CH4 exclusively) have been added to this section with the following sentence:
"That said, several of these algorithms have recently shown relatively good agreement X_{CO2} agreement as compared with simultaneous, colocated TCCON observations \citep{morino2011, butz2011, wunch2011b} or models \citep{oshchepkov2011}."

p.6102,l.9: This rest -> The rest

Fixed.

p.6105,l.29: Could you classify aerosol types "2b" and "3b" by some descriptive wording eg, industrial, soot, marine, absorbing?

Yes. We've added this sentence:
"Type ``2b'' is a mixture of course and fine-mode dust, while type ``3b'' is a carbonaceous mixture; both mixture types contain some sulfate and sea salt."

p.6107,l.16: been -> has been; p.6113,l.22: chi2R ,>2.3 -> chi2R >2.3; p.6115,l.23: ranged ranged -> ranged

Fixed.

Anonymous Reviewer 2

• p.6098, l. 14: the last words of the paragraph about surface reflectance and radiative transfer assumptions are not so well explored in the paper and may not deserve a place in the abstract.

This is a fair statement. We've changed the last line of the abstract to read: "Overall, systematic errors due to imperfect characterization of clouds and aerosols dominate the error budget, while errors due to other simplifying assumptions, in particular those related to the prior meteorological fields, appear small. " Further tests would be required to fully quantify the errors associated the retrieval's simplifying assumptions associated with surface reflectance and radiative transfer.

• p. 6099, l. 15. I was curious to read the reference given about 'inversions that ingest these data' and actually only found theoretical simulations in it. To my knowledge, the only published studies that have ingested real data are Chevallier et al. (2005, 2009) and Nassar et al. (2011)

Chevallier's work (2005, 2009) is the primary reason for this statement, but it is peer-reviewed and currently stands alone in terms of the forefront of obtaining flux inversions from AIRS. Nassar's recent TES work showed that TES had at best a marginal impact on constraining the surface fluxes. We've added the Chevallier 2009 reference after the review

paper of Breon and Ciais (2010) to this statement.

• p. 6104, l. 21: 'prior errors' may replace 'profile' here.

Actually, what it said is what is meant. The correlations in the prior covariance matrix do impose constraints on the level of "wiggles" allowed in the retrieved profile; the stronger the prior correlations, the smoother the retrieved profile is. We have changed "vertical correlation" to "smoothness constraints" to be clear.

• p. 6104, l. 27: I guess that the 12 ppm refers to a standard deviation. If it does, it is about 4 times the variability of XCO2. The quality of the prior profile should be much better than that. I understand that the authors want to minimize the weight of the prior XCO2 information in their retrieval system, but doing that, the balance between the various prior errors of the state vectors cannot be correct.

12 ppm does refer to the 1-sigma standard deviation. Your question explores the "art vs. science" when using prior constraints on retrieved variables; ie, much of this work is indeed an art. In general, we have opted to give retrieved variables as small a prior constraint as possible; this is commonly done throughout the remote sensing field though is perhaps contradictory to the moniker "optimal estimation". We've added a sentence to this effect.

• p. 6105, l. 14: do the authors mean 'precision' rather than accuracy?

Accuracy is what is meant. From Salstein et al. (2008):
"Typical RMS differences between analyses and observations are around 2–3 hPa but larger differences tend to occur in high latitude and high topography regions, in which the analyses are expected to be in general less well constrained."
RMS differences between observations and ECMWF imply accuracy, not precision.

• p. 6105, l. 15: I guess that the 4hPa figure refers to a standard deviation. This should be clarified.

We added "1σ" for clarification, as in Table 2.

• p. 6105, l. 13-16: the text here is ambiguous in that it refers to a publication that actually contains much better figures than the one used here.

This statement is not clear. The text refers to a publication that attempts to quantify the accuracy of the ECMWF analysis surface pressure by comparing to observations. The only figure relating to surface pressure accuracy in our paper, Fig 10a, is something different; it is the difference between actual NCEP surface pressure corresponding to simulated sounding locations, and ECMWF surface pressures in our simulator. The latter have additional errors associated with our simulator model, not the ECMWF data itself. So, comparing the two sets of figures (Salstein et al with our Fig10a) is not an apples-to-apples comparison at all, if that is indeed what is meant by this statement.

• p. 6110, l. 2 and 5: the acronym RT is defined after it is used. It may actually be removed since it is not used elsewhere.

We've moved the few instances of RT and replaced them with "radiative transfer".

• p. 6112, l. 4-6: it is not clear whether this approximation is used in the results presented. If it is not, it would be interesting to show its impact on the results, in particular those presented in Section 3.4. Again, the errors are assigned in a very ad hoc manner while a more rigorous setting could be used.

We used this seemingly ad-hoc formula only to apply the most realistic noise for GOSAT possible. This formula is an empirical fit to the actual GOSAT instrument noise, which is found to be a weak function of signal. In fact, we believe this is the most accurate representation of GOSAT noise available in the literature.

In the simulations, this formula is used twice. For all retrieval experiments, it is used to define the measurement error covariance matrix, $\mathbf{S_\varepsilon}$. For simulations where artificial noise was added (tests 3 through 6), it was used to define the artificial noise properties. The only issue is in tests 1 and 2, where no noise was added, so there is a bit of an inconsistency there. However, these two tests are highly idealized, and the only real penalty is that the chi-squared values of the fits to the spectra are much less than one. Otherwise, we believe the results from these tests are quite valid. We have modified the wording in this section to make this clearer.

• p. 6113: the symbol ΔP_{cld} is not very intuitive and could be replaced by some- thing like ΔP_s to help the reader.

We've changed this symbol to $\Delta P_{s,cld}$. This is to differentiate the cloud-screener retrieved surface pressure, and the L2 retrieved surface pressure (now called ΔP_s). We've also added a footnote in this section to explain this to the reader.

• p. 6116, l.1: missing 's' in particles.

Fixed.

• p. 6125, l. 19: A standard deviation of 4.4 hPa is unrealistically large. There seems to be more here than just interpolation problems. In other words, the authors should not say that such differences can be caused by interpolation.

We actually say they are driven by "vertical binning effects in the orbit simulator", not interpolation problems. Specifically, the surface level is artificially moved in the orbit simulator to match the closest CloudSat level at or below the surface, which occur every 240 meters. So the surface pressure in the orbit simulator is often moved "underground" to a higher pressure, but unfortunately there seems to be a bug in the orbit simulator which does not correspondingly adjust the surface elevation. This is what specifically drives the difference, but this is largely irrelevant for the purposes of the paper. All that matters is that the errors are somewhat realistic (and ideally, a bit on the large side, which the reviewer notes is the case). We've added the word "spurious" to "vertical binning effects" to try to make this more clear that it is a simulator effect. We've also modified the next sentence to include the phrase "and therefore represents a relatively difficult test for the retrieval".

• p. 6126, l. 3: With an RMS error of 1.7hPa, the inverted surface pressure does not look better than NWP analyses. Does this variable deserve to be in the state vector?

Surface pressure is not in the state vector to improve upon NWP (though actually 1.7 hPa is quite competitive with NWP, given the results from the Salstein paper); it is included to

correct for path-length errors induced by scattering, as well as other effects that may be common to the three NIR instrument bands. We've added a sentence to this effect in the state vector section (2.1).

• p. 6128, l. 26: the numbers given here (1-2 ppm) are important and it would be appropriate to explain how they are computed from the results presented. Also, do the authors mean precision rather than accuracy?

We mean accuracy. This was actually described 3 paragraphs above, in the statement: "…the filtered retrievals have RMS X_{CO2} errors of ~1 ppm, relative to more than 4 ppm for the unfiltered retrievals." However, we've modified the last paragraph to better reflect this.