

## ***Interactive comment on “On the statistical optimality of CO<sub>2</sub> atmospheric inversions assimilating CO<sub>2</sub> column retrievals” by F. Chevallier***

**F. Chevallier**

frederic.chevallier@lsce.ipsl.fr

Received and published: 3 July 2015

I thank the referee for his helpful comments. I have addressed all the issues he has raised in the following. The full review is copied hereafter and my responses are inserted where appropriate.

This work presents essentially two different, though related, pieces of research.

There is indeed a theoretical part and a practical one, but they are not two different  
C4323

pieces of research: they are not more different than a conclusion is from an introduction. This misunderstanding may actually originate from the way the abstract was built (the two parts were just linked by "More generally") and I will make it reflect the structure of the paper better in the revised version.

The first argues that the current pipeline used to derive optimal flux estimates from satellite measurements of column CO<sub>2</sub> (XCO<sub>2</sub>) are fundamentally flawed. They are flawed because different prior assumptions are used in the retrieval as compared to the inversion, and the author argues that this inconsistency could bias inversion results. The author then argues that using a strong prior constraint (as most inverse models would suggest) in the GOSAT retrieval algorithm seems to yield better agreement between the XCO<sub>2</sub> in his MACC (v13.1) model, than do comparisons with the standard ACOS (v3.5) XCO<sub>2</sub> retrievals. He also states that ACOS - MACC XCO<sub>2</sub> differences appear to be correlated with surface albedo, though only upon visual inspection of difference maps.

This paper, while certainly thought-provoking, suffers from a severe logical deficiency that must be addressed before publication.

The reviewer's remarks give me the opportunity to clarify a few points, but not to change the paper logic, as I will explain below.

Regarding the first point, of the basic inconsistency between the GOSAT retrieval's prior CO<sub>2</sub> covariance assumption and that of the model, it is worth stating that retrieval groups use a loose prior primarily because they want to be maximally consistent with any model prior covariance. A sufficiently loose covariance is always consistent with a tighter one, but not necessarily the other way around.

This statement is actually a misconception that this paper tries to correct. The requirement of statistical consistency expressed by Eq. (7) in the paper indicates that a loose covariance is *never* consistent with a tighter one within an assimilation system that uses averaging kernels. Actually, we can find the same requirement in the alternative (and heavier) approach of Migliorini (2012, doi:10.1175/MWR-D-10-05047.1, bottom right of p. 263), a paper that I found after the present one was published in ACPD. I will add a paragraph in Section 3 to summarise his findings.

Therefore, it is not clear to me that using a tighter covariance is required to yield formal mathematical consistency upon assimilation of the satellite-retrieved XCO<sub>2</sub>, assuming the averaging kernels are fairly applied.

The way the averaging kernels are applied makes the retrieval prior disappear from the equation, but not the retrieval prior error statistics. Intuitively, we can understand that retrieval prior error statistics that are different from what the inversion system assumes, prevent the inversion equation from looking like radiance assimilation and therefore corrupt the optimality of the system.

My strongest concern, however, regards the author's evaluation of the GOSAT XCO<sub>2</sub> retrieval quality via the comparison to a single model. Disagreement does not necessarily mean the GOSAT retrievals are biased. Models have many sources of error: transport model error, imperfect prior fluxes, and the assimilation of datasets that are sparse in many regions of the world.

This remark has been anticipated and is actually written in many places of the paper (e.g., abstract, l. 7-8; p. 11, l. 24-25; conclusion, p. 17, l. 12-13).

The author's only serious argument  
C4325

The reviewer dismisses the other parts of the discussion in Section 4.1, but it would have been helpful to substantiate this opinion.

is that the difference map between the model-predicted and satellite-retrieved XCO<sub>2</sub> should not have sharp spatial gradients because these should be smoothed out by transport effects (page 1900, line 8). But this argument problematic for at least two reasons:

- He does not specifically demonstrate that there is no way such a spatial gradient can be supported by transport, even if the underlying flux was large and itself contained a strong spatial boundary, as of course happens in some ecotones as well as at land/ocean interfaces; and

We are discussing here column-integrated CO<sub>2</sub> concentrations, not surface concentrations. For instance, a megacity like Los Angeles forms an emission hotspot that enhances  $\chi_{CO_2}$  by 3.2 ppm on average only (Kort et al. 2012, doi:10.1029/2012GL052738). Land/ocean interfaces cannot have a comparable effect in magnitude. For ecotones, the paper already discusses the impact of the Corn Belt on  $\chi_{CO_2}$ , which is found indeed very small.

- One certainly cannot make this argument on maps that contain variable spatio-temporal sampling all plotted on the same map. For example, in the seasonally dry African Sahel region, the satellite has strong seasonality in its ability to monitor this region (namely due to wet vs. dry seasons), and this in and of itself could cause apparent spatial gradients because, in fact different times are plotted on the same map.

This is exactly why monthly maps are also shown in Fig. 3. The reviewer can see that the patterns are indeed robust.

C4326

Secondly, the author states that in certain regions of large (1-2 ppm) model-satellite disagreement, the fault likely lies in the satellite data. While this is certainly possible, the reverse is of course also possible in the lack of additional information.

The former sentence is correct, but not the latter. The regions where the fault is attributed to the satellite data are the regions where models are very unlikely wrong that way, as discussed in the paper. The discussion clearly states that the models can be wrong for other patterns.

Even though the author admits a few times in the text that the model may be imperfect, he does not comment about the general agreement (or disagreement) between the XCO<sub>2</sub> of different carbon inverse model systems.

Indeed this is not the purpose of the paper. The maps of Figs. 1-3 aim at showing and discussing some regional suspicious patterns of the retrievals before the misfits are binned by more abstract retrieval increment size.

These differences exist and they have been shown to be notable especially in regions where the models are not well constrained by in-situ data. For example, Kulawik et al. (AMTD, 2015) and Lindqvist et al. (ACPD, 2015) have recently shown that inversion models can have major differences in the seasonal magnitude of their optimized XCO<sub>2</sub> values both latitudinally and longitudinally. Most of the regions with large retrieval-to-model differences in Fig. 2a are, interestingly, the same regions where also model-to-model differences in XCO<sub>2</sub> can be notable: for example, in the African savannas, in seasonally dry forest/grassland regions in South America, in

C4327

India, and in the high northern latitudes there can be up to 1-3 ppm differences in monthly averages between different inverse models constrained by in-situ measurements.

Kulawik et al. and Lindqvist et al. indeed show large differences between models but they do not show that models can reproduce some of the gradients that are discussed as unphysical in my paper. Incidentally, I note that MACCv13r1 has the best latitudinal fit to ACOS-GOSAT in the Tropics and in the high latitudes (Lindqvist et al., Fig. 7, which they comment with "ACOS is in excellent agreement to MACC from 0 to 50°N"): I interpret this feature as a likely low noise level of this product in these latitudes, that should help isolating local retrieval errors.

Ultimately, of course, we would like to know what is driving these persistent model differences. Nevertheless, the author's conclusions would be on much more solid ground if independent model data sets were shown to support the author's arguments both about the surface albedo effect on retrievals and over-fitting of the radiances.

The reviewer's recommendation would be on much more solid ground if it came after a discussion of the detailed arguments of Sections 4. Analysing one model is already challenging. Suggesting analysing two or more models without further motivation may be a side step.

The author argues that the differences between the model and the retrieval over land at high latitudes are likely due to retrieval errors over dark surfaces. While this argument might have some truth to it (as retrieved XCO<sub>2</sub> is indeed sensitive to the surface albedo in all three bands, and to its changes within each band), it is not entirely supported by the figures

C4328

shown: the map of the mean surface albedo (Fig. 4) shows that the darkest land regions are in Scandinavia and the westernmost Russia while the largest positive differences are most continuous and consistent in central and eastern Russia.

The paper does not claim that the bias is a monotoneous function of the surface albedo, at least because surface albedo is not the only variable that interferes with the retrieval of  $\chi_{CO_2}$ .

Moreover, the author says that the regions with the largest positive differences correspond to the evergreen needle leaf forest biome type, which is not true especially for central Russia where differences in June vary from -1.5 to 1.5 ppm inconsistently (Fig. 3b) and parts of Alaska.

I am referring to the classification of <http://www.esa-landcover-cci.org/>, but do not claim that the correspondance is systematic. The pattern is just strikingly similar (see the land cover map viewer at <http://maps.elie.ucl.ac.be/CCI/viewer/index.php>).

The author finds substantial model-to-retrieval differences in the African savanna/Sahel region, and attributes these differences to "systematic errors in the retrievals", speculating about averaging kernels not peaking low enough in the atmosphere due to too loose retrieval prior error variances. However, the author does not speculate more about the reason for such regionally constrained errors: why would the prior error variances have more impact in that particular region compared to elsewhere?

My statement was not rigorous enough and I apologize for it. Rather than loose retrieval prior error variances, some of the fault should lie in inappropriate prior error correlations. I will correct the sentence.

C4329

He suggests that CO<sub>2</sub> from fires inaccurately represented in the MACC model might be another cause for the differences but considers this unlikely.

Indeed, "if the model was underestimating the intensity of the fire, we would expect the mean difference to take the shape of a plume, i.e. to spread downstream the source region, but this is not the case." (p. 12, l. 20-25).

However, a look at this particular region's optimized, natural CO<sub>2</sub> fluxes inverted by different models reveals extremely large differences in the fluxes, and also that similar differences are reflected in that region's XCO<sub>2</sub>. As long as the model differences in this region are unexplainably large, one of the models cannot be fairly used to speculate about biases in the satellite retrievals in that region without some kind of additional information.

The reviewer does not explain how any model would have so high values of  $\chi_{CO_2}$  over this specific land region. Actually the same difference pattern between ACOS-OCO-2 and the GEOS-5 model was shown at the first OCO-2 Science Team Meeting (Pasadena, CA, USA, February 2015) by Baker et al.

Page 11901 line 21. It is mentioned that "boreal forests are covered with needle-leaved trees". It is safer to say "are largely covered". Apart from the widespread light coniferous larch and pine forests, dark coniferous needle-leaved trees can not dominate the landscape and often appear in mosaic patches with broad-leaved trees mostly due to post-fire successional dynamics (eg Shvidenko and Nilsson, Tellus, 2003).

I will make the change.

C4330

The author presents in Figs. 5-8 an interesting metric for evaluating overfitting in the retrievals (i.e., too tight a prior), and shows that increasing the weight of the prior XCO<sub>2</sub> could make the retrievals statistically more consistent with the model. However, he does not show any spatial patterns of this metric; therefore it remains unclear if the suggested change in the retrieval prior errors would lead to worse misfits in some currently well-matched regions in addition to the likely improvements in the model-retrieval misfits in the regions where the differences are large.

The reviewer seems to suggest re-drawing the mean difference map of Fig. 2 with  $\check{x}^{a,r}$ . However,  $\check{x}^{a,r}$  is not bias-corrected and this map could not be interpreted in terms of better or worse misfits. The argument made with Fig. 6 relies on the standard deviation of the misfit distribution, not on its mean. For this reason, the conclusion states that "the optimal-estimation retrieval process and, *consequently, its posterior bias correction* need retuning".

And even if he did, it would still suffer from the problem of comparing to a single model,

As explained in p. 14, l. 17-19, what matters here, while discussing about random differences (standard deviations) is that the model errors are uncorrelated with the retrieval errors and with the retrieval-prior errors. This hypothesis is further discussed in p. 15, l. 10-13. I will complete the discussion by excluding the possibility that subgrid scale variability plays a role in the results.

and the fact that it couldn't be accounted for by faithfully using the column averaging kernel in the assimilation.

This fact just comes from the maths (Eq. (7) of Section 3).

C4331

Overall, by counting too much on the results obtained by this metric, we risk the possibility of both the model and the prior XCO<sub>2</sub> being wrong

Again, what matters is that they are not similarly wrong (correlated errors), which is the case.

and the satellite observations the truth. The satellite retrievals are certainly not (yet) completely free of retrieval biases, but it is fruitful to remind oneself why they are being carried out: because neither our prior knowledge nor our models are perfect.

This question actually forms the first sentence of the introduction: "CO<sub>2</sub> surface fluxes at the Earth's surface can be inferred from accurate surface measurements of CO<sub>2</sub> concentrations, but the sparseness of the current global network still leaves the flux horizontal and temporal gradients, and even their latitudinal distribution, very uncertain (Peylin et al. 2013)".

Even if similar results were obtained based on comparisons to other models, this philosophical dilemma would still remain in the background but the reasons that support to change the current retrieval procedure would be stronger.

Adding models would not change the maths (Section 3). Our study with a single model, in particular Figs. 5-8, is just an illustration of this theoretical section.

Detailed comments:

C4332

- Page 11893, line 21. The author should state that the use of a rather loose prior CO2 covariance is not specific to ACOS, with some examples. For instance: 1) The RemoTeC retrieval has a formally unconstrained XCO2 (Butz et al., Applied Optics, 2009), and 2) the BESD retrieval uses a prior error on XCO2 of 15.6 ppm (Reuter et al, AMT, 2010). etc.

This section is about ACOS and the two references would not fit there, but I will add them in Section 2 after "This condition is not achieved by current satellite retrieval algorithms, at least because they artificially maximize the measurement contribution in the retrievals through the use of very large prior error variances".

- Page 11896, line 3: "H a linearized" → H is a linearized

I will correct it.

- Page 11896, line 12: "inversion window for the inversion" → inversion window

I will replace the first "inversion" by "assimilation".

- Page 11897, Eq. (4): might be more informative to simply show the derivation of Eq. (4) instead of describing it in the previous paragraph.

I will develop the demonstration in the revised text.

I have noticed a misplaced prime in Eq. (4), that induced missing primes in Eq. (6-7). Additionally, I have also noticed that, when we make Eqs. (5-6) consistent, the requirement of Equation (7) can be relaxed to  $\check{\mathbf{H}}\check{\mathbf{B}}\check{\mathbf{H}}^T = \check{\mathbf{H}}\hat{\mathbf{B}}\hat{\mathbf{H}}^T$ , which means that consistency needs only to be satisfied at the resolution (information content) of the retrieval. I will correct these equations and update the text accordingly.

C4333

- Page 11899, line 16: "long-tern" → long-term

I will correct this.

- Page 11899, lines 17-18: variability in the XCO2 field is ~ 8 ppm in Fig. 1, retrieval-to-model differences are most typically less than 1 ppm (Fig. 2a). Therefore, the retrieval-model difference is "much less" than the variability within the modeled or retrieved XCO2 field.

I will remove this statement.

- Page 11900, lines 10-11: it is incorrect to say that the local spatial gradients mostly reflect the retrieval gradients. For example, the gradients in Fig. 3a for South Africa, South America, and the latitudinal gradients in the oceans are not obviously wrong in the retrievals (Fig. 1a).

The text does not say that the retrieval gradients are wrong there, but that they explain the gradients of the differences there. Whether they are right or wrong is the topic of the rest of the section.

- Page 11900, lines 23-25: The surprising discontinuity in XCO2 on the NW coast of the U.S. compared to the adjacent ocean data is more clearly seen in the model (Fig. 1b) than in the retrieval.

I will remove the statement.

- The benefits of showing Tables 1 & 2 are not clear. Because the paper otherwise concentrates on the GOSAT data years 2009-2013, it might be

C4334

more helpful to the reader to see a map of where the in-situ data were collected during these years.

The tables follow a request from a station PI to have the name of his station appear publicly. It takes 1.5 pages on the "printer-friendly" version, which seems reasonable to me, but adding a map may be too much. I leave this question to the editor.

- Figures 5-8 need a more informative y-axis label. For example "XCO<sub>2</sub><sup>a</sup> - XCO<sub>2</sub><sup>b</sup> model (ppm), mean(—) or sigma (\_\_\_)", or something similar.

I will replace the label by "mean or  $\sigma$  of the  $\chi_{CO_2}$  misfits (ppm)".

- Figure 6: the two blue shades look very similar in the printed version. Consider colors with a larger contrast.

I will replace the light blue by gold.

---

Interactive comment on Atmos. Chem. Phys. Discuss., 15, 11889, 2015.