Review of "Assessing 5 years of GOSAT proxy XCH_4 data and associated uncertainties" by R.J. Parker et al

July 10, 2015

In this manuscript, the authors have presented a coherent picture of the UoL GOSAT proxy XCH_4 retrievals over five years, their uncertainties, and where they may contribute information to existing CH_4 inverse models. This work is within the scope of this journal and of relevance to the community of users of GOSAT XCH_4, and therefore I recommend its publication. I have however, a few comments and suggestions which I will detail below.

- **p5943, line ~20** The authors say that they substitute the stratosphere of the MACC-II CH_4 inversion with the TOMCAT stratospheric chemistry model. As far as I know, the the TM5 4DVAR CH_4 setup for MACC-II uses stratospheric OH, Cl and O^1D loss rates from a stratospheric chemistry model already. Therefore, it is not clear to me that substituting TOMCAT CH_4 gains anything in the stratosphere. Have the authors compared the MACC-II stratospheric CH_4 with TOMCAT CH_4 to see how different they are, and whether one is better than the other when compared to (say) stratospheric CH_4 limb sounding measurements?

- **p5944, line ~1** The tolerance of 30 hPa between the retrieved surface pressure and the ECMWF surface pressure seems awfully large. The authors explain that this is necessary because the ECMWF orography at $0.75° \times 0.75°$ is quite coarse. While I agree that $0.75° \times 0.75°$ is too coarse compared to the GOSAT footprint, I am surprised at the implication that the UoL retrieval just uses that orography and surface pressure without any refinement. From talking to colleagues who do GOSAT retrievals, I know that at least ACOS and RemoteC assume the ECMWF surface pressure to be the average over the grid cell, and then derive the average surface pressure over a GOSAT footprint by using a high resolution digital elevation map and the local temperature and humidity (for the density of air). This way, topography at scales finer than $0.75° \times 0.75°$ does not automatically screw up the estimate of average surface pressure across a footprint derived from ECMWF. Doesn't the UoL retrieval do something similar?

- **p5945, line ~19** I'm not sure how a correlation coefficient of 0.54 can be considered "good". More importantly, I'm not sure that the correlation coefficient between GOSAT XCH_4 and TCCON XCH_4 is the right quantity to quote in order to demonstrate the "goodness" of GOSAT retrievals. After all, a high correlation coefficient only means that the *phasing* between two time series is consistent, and says nothing about the amplitudes. For example, the seasonal cycle amplitude of

GOSAT XCH_4 could be several times that of TCCON XCH_4 (I'm not saying that it is), and still yield a high correlation coefficient as long as the phases are consistent. So I think quoting the biases per station and plots such as Figure 2 are enough to show how good (or bad) the retrieval is, and correlation coefficients can be left out.

- **p5945, line ~23** It seems to be a common practice in GOSAT retrieval literature to calculate the standard deviation of the individual site biases and call it the station-to-station bias, with the understanding that a lower value is better. However, I'm still mystified as to the significance of this quantity, or why this should be a measure of relative accuracy (relative to what?). Let's say that a retrieval of GOSAT XCH_4 has a bias of 20 ppb with respect to *all* TCCON stations, while another retrieval has biases varying from -1 ppb to 1 ppb. In that case, the station-to-station bias of the first retrieval is much lower than that of the second one. Does that mean that the first retrieval is better, despite having worse biases?

- **p5946, line ~1** The authors compare the ratio XCH_4/XCO_2 from their retrieval to that from TCCON. However, it is not clear to me whether a bias of 0.014 ppb/ppm and a single-sounding precision of 0.033 ppb/ppm is good or bad, or more importantly, any better than the performance of full physics XCH_4. If the authors want to prove that the ratio is indeed a "better" product with less aerosol-related bias, I would like to see scatter plots of XCH_4/XCO_2(GOSAT) – XCH_4/XCO_2(TCCON) versus retrieved aerosol parameters. In all full physics retrievals I've seen, such scatter plots have non-zero slopes, meaning that the retrieved value of XCH_4 depends (spuriously) on other retrieved parameters. If the authors show such scatter plots for their full physics XCH_4 as well as for their XCH_4/XCO_2 ratio, and if the slopes of those scatter plots turn out to be smaller for the latter product, then I'll be convinced that the ratio is indeed a better/less biased product than the full physics XCH_4.

- **p5946, line ~15** Could you say if all three models assimilated the same set of surface observations? If not, how different where those observations, in terms of spatiotemporal coverage?

- **p5946, line ~15** CT2013B only generated CO_2 field till the end of 2012. How did the authors do their analysis for the year 2013, as their figures suggest?

- **p5947, line ~1** If the authors' goal is to prove that the models agree with TCCON, why not sample the models *at* the TCCON stations whenever there's a TCCON sounding, instead of within ±2° of the station? Also, were the model fields smoothed with the TCCON averaging kernels?

- **p5949, line ~8** The sentence "This allows us to calculate the average uncertainty related to the XCO_2 model for each grid box and 8 day time step is a redundant repetition of the previous sentence. Please delete it. A suggested modification to the text around it would be "In order to assess the importance of the uncertainty of model XCO_2 we bin the three model fields in 4° × 5° boxes over eight day chunks, and calculate the inter-model spread. We convert this spread in model XCO_2 into an uncertainty in XCH_4 by ..."

- **p5949, line ~17** Here (and elsewhere) when the authors say "a posteriori error", they mean the posterior error in XCH_4 from the full physics retrieval, correct? In that case, that should be mentioned explicitly here, and perhaps a couple of other places just to solidify it by repetition. Since there are two retrievals here (proxy and full physics), I kept getting confused about which one the a posterior error was referring to.

- **p5950, line ~14** From the way the authors set up the manuscript, I was expecting that the proxy XCH_4 would be compared to the three-model median. Instead, they compared it to MACC-II XCH_4. Why? I was under the impression that the multi-model median was more accurate than any individual model.

- **p5951, line ~22** 2010 was a high biomass burning year over tropical S America. Could there be a connection between this and the fact that all three models agreed better with GOSAT XCH_4 this year than other years?

- **p5952, line ~6** The proxy method only minimizes biases due to aerosols, because of similar aerosol scattering properties. I have yet to see proof that any bias due to the M gain mode is similar for XCO_2 and XCH_4. So the proxy method will not necessarily do away with M gain related biases.

- **p5953, line ~1** This is a general comment about using $\sigma_{\Delta XCH4}$. This quantifies the variation of XCH_4 about a model field over a $4° \times 5°$ grid box, and the the authors suggest that a large value for $\sigma_{\Delta XCH4}$ means that the model does not capture variations observed by GOSAT, and therefore GOSAT XCH_4 can supply new information to the inversion used to generate the model field. However, I wonder how much of the $\sigma_{\Delta XCH4}$ is driven by the finite resolution of the (inverse) model. In other words, say an inverse model estimates fluxes at $4° \times 5°$ and simulates atmospheric transport at $4° \times 5°$. Then no matter how much information one pumps into it (to the point where the $4° \times 5°$ fluxes are "perfect"), it will not be able to reproduce the sub-$4° \times 5°$ variability seen by GOSAT, and can potentially have a large $\sigma_{\Delta XCH4}$. This large $\sigma_{\Delta XCH4}$ will not mean, however, that (a) the fluxes at $4° \times 5°$ are wrong, or (b) XCH_4 can add more information to the inversion to lower $\sigma_{\Delta XCH4}$.

- **p5954, line ~6** The sentence reads as if biomass burning contributes to the model uncertainty over Central Africa *and* over Eurasia during boreal summer. Please rewrite to make it clear that biomass burning is the factor over Central Africa and large uptake (which is underconstrained due to lack of observations) is the factor over Eurasia.

- **p5955, line ~3** More recently, Pandey et al (doi:10.5194/acpd-15-8801-2015) have also used the ratio XCH_4/XCO_2 directly in a flux inversion. Perhaps mention that as well along with Fraser et al?

- **Figure 6** Why do the $\mu_{\Delta XCH4}$ and $\sigma_{\Delta XCH4}$ lines stop at the end of 2012, but the posterior and model error lines continue for another year? If it's due to model availability, how can the green model error line be there for 2013?