We thank the referees for their careful and thoughtful comments on the manuscript. We also thank the editor are giving us additional time to respond to the comments. We will begin with a general response to all referees, followed by a more detailed response to the comments from each referee.

Referees #2 and #4 both expressed concern with referring to the flux inversion approach as 4D-Var. Referee #2 suggested that it is a 3D-Var problem, whereas Referee #3 suggested we "drop the "4D" and "use only variational" to describe the approach. We would like to stress that this approach has been widely referred to as "4D-Var" by the inverse modeling community for quite some time. This has been the case in inversion analyses to estimate CO_2 fluxes (e.g. Basu et al., 2013, Deng et al., 2016; Zheng et al., 2018), CH_4 emissions (e.g., Meirink et al., 2008; Bergamaschi et al., 2010; Turner et al. 2015), CO emissions (e.g., Kopacz et al., 2010, Jiang et al., 2017), NOx and SO_2 emissions (Qu et al., 2019), and emissions of SOx, NOx, and NH_3 (e.g., Henze et al., 2009). We agree with the referees that the 4D-Var methodology in the flux inversion context is different from that in numerical weather prediction (NWP), but it is unclear what terminology would be better. It is not a 3D-Var scheme as the fluxes are time dependent. In our analysis we solve for monthly CH_4 emission over a four-month period (February – May). This means that the February fluxes are being influenced by observations from February to May, relying on the adjoint model to propagate information from observations in the future back in time to update the February emissions. If we were solving for mean fluxes over the whole four-month period, it would be more similar to a 3D-Var approach. In addition, just referring to the approach as a variational scheme is vague. It might be a useful exercise for the inverse modeling community to develop new terminology for these methods that have been adapted from the NWP community, but that is not an objective of this study. Introducing new terminology for the flux inversion community would require a methodology study that clearly highlights the similarities and differences in how the 4D-Var technique is used in the flux inversion and NWP communities so that it can justify the proposed terminology.

Another concern raised by Referees #2 and #4 is our statement that the model is assumed perfect in strong constraint 4D-Var. The referees stated several times that this is not true. We should have been more precise in our wording. In strong constraint 4D-Var it is assumed that the model perfectly evolves the state in time. As noted by Trémolet (2007), "current operational implementations of 4D-Var rely on the assumption that the numerical model representing the evolution of the atmospheric flow is perfect, or at least that model errors are small enough (relative to other errors in the system) to be neglected." The full 4D-Var cost function is

$$J(\mathbf{x}) = \sum_{i=0}^{N} \frac{1}{2}(\mathbf{y}_i - \mathbf{H}_i\mathbf{x}_i)^T\mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{H}_i\mathbf{x}_i) +$$

$$\sum_{i=1}^{N} \frac{1}{2}[\mathbf{x}_i - M_i(\mathbf{x}_{i-1})]^T\mathbf{Q}_i^{-1}[\mathbf{x}_i - M_i(\mathbf{x}_{i-1})] + \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b)$$

where $\mathbf{x}$ is the model state, $\mathbf{y}_i$ are the observations, $\mathbf{H}$ is the observation operator, and $\mathbf{R}$, $\mathbf{B}$, and $\mathbf{Q}$ are is the observation, background, and model error covariance matrices, respectively. Here M is the forecast model that evolves the state in time from time t to $t+1$ as

$$\mathbf{x}_{i+1} = M_{i+1}(\mathbf{x}_i) + \mathbf{u}_{i+1}$$

with model error $\mathbf{u}$. In the cost function above, we are solving for the full 4D state vector. This is weak constraint 4D-Var. However, by assuming that the model evolution is perfect we can neglect

the second term in the cost function and optimize only the initial state ($\mathbf{x}_0$) (since this initial state is evolved forward in time without error). Thus, the cost function becomes

$$J(\mathbf{x}_0) = \sum_{i=0}^{N} \frac{1}{2}(\mathbf{y}_i - \mathbf{H}_i\mathbf{x}_i)^T \mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{H}_i\mathbf{x}_i) + \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b)$$

where $\mathbf{x}_i = M_{i,0}(\mathbf{x}_0)$ is the state at time i based on the perfect evolution of the initial condition $\mathbf{x}_0$ forward in time. This is strong constraint 4D-Var. We have modified the text to make it clear that the model evolution is assumed to be perfect in strong constraint 4D-Var.

It was suggested by the referees that the **R** matrix accounts for model errors in strong constraint 4D-Var. In theory, **R** accounts for errors in the observations and for representative errors in the observation operator. It does not account for errors in the propagation of the state in time. In the strong constraint approach, one could adjust **R** to capture errors in the evolution of the state, but that would be inconsistent with the framework. Nevertheless, this is what is done in practice. It was also suggested by the referees that there are alternative means of estimating the model bias, and we agree. For example, one could incorporate the bias **b** into the first term of the cost function (i.e. [$\mathbf{y}_i - \mathbf{H}(\mathbf{x}_i - \mathbf{b}_i)$]) and solve for the bias together with $\mathbf{x}_0$ (i.e., J becomes a function of $\mathbf{x}_0$ and **b**). A similar approach was proposed by Dee and DaSilva (1998) in the context of a sequential assimilation scheme. We are not claiming that our approach is the only means of estimating model bias. We have adapted the weak constraint 4D-Var framework, as described by Trémolet (2007), for the GEOS-Chem 4D-Var scheme because in the context of a 4D-Var assimilation framework, the weak constraint scheme provides a means of estimating the model bias that is consistent with the 4D-Var formalism.

References

Basu, S., et al., Global CO_2 fluxes estimated from GOSAT retrievals of total column CO_2, Atmos. Chem. Phys., 13, 8695–8717, https://doi.org/10.5194/acp-13-8695-2013, 2013.

Bergamaschi, P., et al., Inverse modeling of European CH4 emissions 2001–2006, J. Geophys. Res., 115, D22309, doi:10.1029/2010JD014180, 2010.

Deng, F., et al., Combining GOSAT XCO2 observations over land and ocean to improve regional CO2 flux estimates, J. Geophys. Res. Atmos., 121, 1896–1913, doi:10.1002/2015JD024157.

Dee, D. P., and Da Silva, A. M., Data assimilation in the presence of forecast bias, Q. J. R. Meteorol. Soc., 124, 269–295, 1998.

Henze, D. K., Seinfeld, J. H., and Shindell, D. T., Inverse modeling and mapping US air quality influences of inorganic PM2.5precursor emissions using the adjoint of GEOS-Chem, Atmos. Chem. Phys., 9, 5877–5903, https://doi.org/10.5194/acp-9-5877-2009, 2009.

Kopacz, M., et al., Global estimates of CO sources with high resolution by adjoint inversion of multiple satellite datasets (MOPITT, AIRS, SCIAMACHY, TES), Atmos. Chem. Phys., 10, 855–876, https://doi.org/10.5194/acp-10-855-2010, 2010.

Jiang, Z., et al., A 15-year record of CO emissions constrained by MOPITT CO observations, Atmos. Chem. Phys., 17, 4565–4583, https://doi.org/10.5194/acp-17-4565-2017, 2017.

Meirink, J. F., Bergamaschi, P., and Krol, M. C.,Four-dimensional variational data assimilation for inverse modelling of atmospheric methane emissions: method and comparison with synthesis inversion, Atmos. Chem. Phys., 8, 6341–6353, https://doi.org/10.5194/acp-8-6341-2008, 2008.

Qu, Z., Henze, D. K., Theys, N., Wang, J., & Wang, W., Hybrid mass balance/4D-Var joint inversion of NOx and SO_2 emissions in East Asia. Journal of Geophysical Research: Atmospheres, 124, 8203–8224. https://doi.org/10.1029/2018JD030240

Trémolet, Y., Model-error estimation in 4D-Var, Q. J. R. Meteorol. Soc. 133: 1267–1280, 2007.

Turner, A. J., et al.: Estimating global and North American methane emissions with high spatial resolution using GOSAT satellite data, Atmos. Chem. Phys., 15, 7049–7069, https://doi.org/10.5194/acp-15-7049-2015, 2015.

Zheng, T., French, N. H. F., and Baxter, M.: Development of the WRF-CO2 4D-Var assimilation system v1.0, Geosci. Model Dev., 11, 1725–1752, https://doi.org/10.5194/gmd-11-1725-2018, 2018.

Responses to Individual Referees

The comments from the referees are in **bold Calibri** font and our responses are in plain Times New Roman font.

Anonymous Referee #2

The authors propose a novel method to quantify errors in transport modelling. It uses GOSAT measurements within a new assimilation framework. In particular, the authors propose to use part of the knowledge provided by the adjoint of the transport model to constrain model errors. I am personally a great enthusiast of using the adjoint in a more comprehensive way than what is currently done by the community. Therefore, I recommend supporting the authors in their efforts to do so, and their work should make an interesting contribution to the community.

However, the present state of their manuscript requires significant changes, clarifications and rearranging before suitable for publications.

1   General comments

1.1   Structure

With the current structure and organization, it is hard to filter the take away messages. Some details are missing, other are not necessary. I have some doubts on the choice made by the authors in the way they split the content of their work between the present manuscript and the sister paper in GMD (gmd-2019-248). At lot of details about the method itself is given here, while it would be more suitable for a model description paper in GMD. Also, many details about the method, in particular the tuning choices, are set aside, while they are critical for validating the reliability of the approach. Similarly, all the OSSEs would make more sense in a GMD paper.

Only the part directly concerning real data would be suitable for an ACP content in my opinion, as well as the discussion on model resolution?

I suggest the author dramatically reconsider the way they organize their presentation of their work to help the reader navigate through the results.

We agree that the manuscript was long and consequently it was difficult to navigate through the results. We have significantly shortened the OSSE section. We have removed one of the four OSSE experiments (the perturbed flux experiment), based on the comments from Referee #4, and we have removed the GOSAT observational coverage sensitivity experiment. Overall, we have removed 6 figures from the manuscript. We have also improved the description of the remaining experiments so that it is easier to follow the results.

1.2   Weak constraint vs Strong constraint

The authors insist on sticking to the 4D-VAR formulation of the surface flux inversion problem. Such an approach, following the work from the numerical weather forecast community, is artificial and ends up in clumsy and artificial formulations. Even though the equations are correct, they are uselessly complicated. The surface flux inversion problem is a 3D-VAR problem, the time step of the NWF community having no meaning in our case (the author implicitly acknowledge this fact by putting the emissions in the model parameters p).

I may be wrong, but the author's formalism could be easily replaced by the classical inversion equation, simply adding model bias in the control vector. Thus, the matrix Q would only be a sub matrix of B, and the Lagrangian terms would not be needed; they would be implicitly solved for in the problem as a 4D CH4 atmospheric source/sink in the transport model.

Please see our general response above regarding the widespread use of the 4D-Var approach by the inverse modeling community. In addition, the emission optimization problem in our analysis is not 3D, as the emissions are time dependent and rely on the model transport to use observations in the future to quantify the fluxes in a given month. We are taking advantage of the fact that the 4D-Var scheme is a smoother to quantify the monthly mean fluxes over the entire assimilation period.

Yes, one can use the classical inversion equation and add the model bias to the control vector, as we noted in our general response above. However, the weak constraint 4D-Var approach is an alternate approach, in which we are adding the bias term to the control vector, but in a manner that is consistent with the 4D-Var formalism.

1.3   Period of interest

The authors chose a very short period of interest (4 months) in 2010. This seems to be guided by the availability of validation data.

Such a duration is very short considering the global atmospheric transport patterns. The author show that the biases in the initial conditions can be corrected quite quickly, which would excuse the short period. However, what about long-term biases?

In the present conditions, the WC inversion seems to only allow for short term corrections, at the cost of a loss in the mass balance. It would then limit the inversion conclusions to very regional patterns, reducing the interest of running global models...

The main goal of the inversion analysis is to quantify regional sources of CH_4 (using regional and global models). We are trying the use the imprint of fresh emissions on atmospheric CH_4, as measured by GOSAT, to infer the surface emissions. Long simulations are not helpful in this regard as the emission signals become well mixed into the background on long timescales. However, biases associated with transport on long timescales can adversely impact the inversions. For example, a bias in the Brewer-Dobson circulation or in the representation of the polar vortex can result in a bias in CH_4 in the lower stratosphere, which will produce a bias in the modeled XCH_4. Such a bias cannot be mitigated by a correction in the surface emissions on short timescales, given the age of air in the lower stratosphere, but that would not be desirable in any case since our interest is in estimating emissions that are free of the signature of model biases. Our results show that the weak constraint approach can mitigate such biases in the stratosphere on emission-relevant timescales.

Also, we note that there is no loss in mass balance. The model is mass conserving. The weak constraint is adding sources and sinks throughout the atmosphere to correct for the model errors, as opposed to putting these sources and sinks only at the surface. In doing so, it is redistributing the mass of CH_4 in the atmosphere to better fit the CH_4 mass suggested by the GOSAT measurements.

1.4   Uncertainty matrices

It feels that your results are so dependent to the subjective choice of Q that they are hardly exploitable. It is already rather challenging to find a balance between R and B in a classical inversion. The final taste of the work as it is describe is that the method does not really fit its purpose. Quantifying biases would be as efficient with simple forward simulations as it is presented...

I am convinced that the use of the adjoint to quantify model errors is a good approach, but with the author's framework and no additional data, it seems quite impossible to deduce any conclusive results...

The motivation for the study was the fact that there is increasing evidence that the vertical transport in the GEOS-Chem model is too weak, and that latitudinal dependent biases in XCH_4 in the model are problematic for CH_4 inversion analyses using the model (e.g., Turner et al., 2015). Therefore, we decided to apply the weak constraint approach, as described by Trémolet (2007), to try to better characterize these biases. A major concern was that assimilation of GOSAT XCH_4 data, which do not provide vertical profile information, would be inadequate for capturing the model errors. Consequently, we chose to assume the simplest form for **Q** to see what structures in the model error can be identified given the GOSAT column observations. Imposing prior structure in **Q** would have made it more challenging to determine if the XCH_4 data, when assimilated into a weak constraint 4D-var framework, can provide constraints in the vertical structure of the model error. Surprisingly, the results show that the estimated bias field is consistent with our developing

understanding of the vertical transport bias in GEOS-Chem. For our follow-up application of the approach, we are now trying to develop a better **Q** matrix to further improve the performance of the system.

Reference
Turner, A. J., et al.: Estimating global and North American methane emissions with high spatial resolution using GOSAT satellite data, Atmos. Chem. Phys., 15, 7049–7069, https://doi.org/10.5194/acp-15-7049-2015, 2015.

2   Specific comments

- **p.2 l.24: regional and global scale**

Transport can have global impacts, but the focus is on the regional impacts that confound efforts to quantify regional emission estimates of CH_4.

- **p.2 l.33: "assume that the model is perfect": this statement is misleading; the so- called strong-constraint inversion never assumes that the model is perfect. Errors are represented in the R observational error matrix. Of course, in most inversion framework, the R matrix is too simple and misses most of the error patterns, but that is only a technical choice in the application.**

This should have stated "assume that the model evolution is perfect." We have corrected the text. Please see our general response above regarding the nature of the errors captured by the **R** matrix.

- **p.4 l.7: EDGAR 2004 is quite outdated; could the author justify such a choice?**

This is an older inventory, but it does provide a reasonable a priori for the assimilation of GOSAT data for 2010 conditions. It would be difficult to justify the use of the 2004 inventory for simulation of 2020 conditions, for example.

- **p.4-6 Sect. 2.2.1: This section mixes observation information with preliminary studies and side conclusions. Shorten and put results in the results section if really needed or in Supplement, or in a GMD style paper**

We have removed Figures 1 and 2 and shortened the section.

- **p.6 Sect. 2.2.2: a lot of information is given about the instrument precision and techniques; are such details really needed in a OSSE paper?**

This is not an OSSE paper. We start with the OSSEs to assess the performance of the system, but the focus is on the assimilation of the GOSAT data. The description of the data used for validation of the assimilation results is important to help the reader interpret our results.

- **p.7 l.20: p vary over time? it is not clear from equation (1)**

The fluxes do vary over time. We solve for monthly fluxes over the assimilation period. We have added text to make this point clear on Page 7, lines 16–17.

- **p.7 l 26: it is not clear at all what are the dimensions and spaces of the object presented here**

We have added text to specify the dimension of the variables.

- **p.8 eq.3: are Q and R always the same for each i? by design, you make it impossible to have temporal correlations in the observation space. It is often the case in practice in classical inversions, but should be highlighted as a limitation of this formulation**

In theory, $\mathbf{R}$ and $\mathbf{Q}$ change for each i, but in practice, we keep Q constant in our analysis. We now explicitly remind the reader of this on Page 8.

- **p.8 l.10: 4D-VAR artificial and makes it hard to understand. Justified in NWF where the state is directly optimized with respect to observations, but our interest is p (surface emissions). We rather do 3D-VAR!!!**

Please see our general comment above. We are not using a 3D-Var approach.

- **p.8 eq.4: should H be Hi? should be different at each so-called time step?**

Yes, $\mathbf{H}$ is different for each observation that is assimilated. We have corrected that.

- **p.8 eq.4: The Lagrangian factor part would be automatically solved with a 3D-var SC including a 3D source-sink atmospheric term...**

That is an alternative inversion approach using 3D-Var. Inversion analyses using the 4D-Var scheme benefit from the fact that 4D-Var is a smoother, which allows us to more effectively exploit observations that are distributed in time.

- **p.9 l.11: What is exactly the size of ui? and what are i standing for? days? minutes? It is not fully clear from the text neither as it changes over the course of sections...**

The size can be equal to or smaller than the size of $\mathbf{x}$, depending on the spatial extent over which the model errors are quantified. We have added text to clarify this on Page 8 (line 1).

- **p.10 l.1: It is a big problem to assume diagonal B (as well as diagonal Q) as you give too many degrees of freedom to your inversion compared to the number of observations.**

The chemical data assimilation problem, whether we are solving for the state or emissions, is an underdetermined problem. We agree that imposing structure in the covariance matrices will help constrain the solution. However, as we noted in our response to General Comment 1.4, we chose the simplest representation of Q since it was unknown to what extent we would be able to capture the spatial patterns in the model error using the GOSAT XCH$_4$ data (given that we use only data over land and are not assimilating CH$_4$ profiles). Without good knowledge of the structure of Q, the most conservative approach is to assume that it is diagonal. In a similar vein, in the flux inversion community it is not uncommon to assume that B is diagonal (e.g., Turner et al. 2015, Jiang et al., 2017). Estimating the covariances for different emission types (e.g. for wetlands, agricultural, fossil fuels) is challenging, and assuming no structure is safer than imposing incorrect structure in the analysis.

- **p.10 l.17: the choice of u is very shortly justified (and unconvincingly); consider extending such justification or all the results appear untrustworthy**

As we explained in the manuscript, the scaling ũ adjusts the direction of the gradient descent, between optimizing the emissions and the state. We believe that it would impair the reader's ability to navigate through the results if we were to expand the manuscript and present a detailed discussion of the sensitivity of the optimization algorithm to ũ.

- **p.10 l.20-30: very clearly and accurately written paragraph stating the limitations of the method. But later sections contradicts the acknowledgement of the limitations.**

It is unclear which sections the referee thinks contradict our discussion here. We were careful in highlighting the limitations, and our results were presented in the context of these limitations.

- **p.11 l.3-6: important sensitivity results! should be more extensively detailed, either in the result part, or supplement, or GMD companion paper...**

As we described in the manuscript, we conducted a series of sensitivity analyses using a range of values of q between 0.05 ppb and 2000 ppb, and found that the validation results (similar to what is shown in Figure 15 in the original manuscript) did not change for values of q larger than about 50 ppb. As a result, we assumed a value of 50 ppb for q. We do not believe that adding individual plots of the validation results for a range of values of q to the already long manuscript would be any more informative than the description of the results already in the manuscript.

- **p.11 eq.4: is there a link with the cost function? the use of J is misleading; the equation is artificial and does not make sense. If the purpose is to introduce the total sensitivity in eq.12, the author should rather change both eq 11 and 12 and write them with the adjoint of the model, evaluated at an increment observation vector equal to 1 at every GOSAT obs.**

Our description of this experiment was confusing. The cost function here is different from the weak constraint cost function. In the interest of shortening the OSSE section, we have removed this experiment.

- **p.12 l.15: with less extreme situations, the balance between Q and B is expected to be even more subtle...**

The objective with the OSSEs was to assess whether the XCH_4 data is able to provide information to help mitigate the model bias. Since these were OSSEs in which the pseudo-observations were generated with the same model, we chose extreme biases for the evaluation. The results showed that even when convection was turned off, the model was able to partially mitigate the model bias. They suggested that XCH_4 data should be able to correct more modest model biases that may be present in real inversion analyses. Yes, the balance between B and Q would be more subtle with smaller errors. This will be an issue in the case where the model errors mimic emissions errors, such as a bias in the chemical sink for CH_4. However, it will not be an issue for transport errors, which cannot be corrected for by adjusting surface emissions. If the biases mimic emission errors, additional information will be needed, such as by assimilating formaldehyde (HCHO) data to provide constraints on the hydroxyl radical (OH). And this will be an issue for any inversion analysis, regardless of the bias correction scheme employed.

- **p.12 l.27: not clear what tuning you are talking about; please consider adding a table detailing all the OSSE and inversion set-ups to help the reader**

This was poorly worded. We meant that we conducted a number of experiments to evaluate the impact of the selected window length on the analysis. We have rewritten this section to more clearly describe the experiments.

- **p.12 l.35: negativity bound: ad-hoc unjustified choice; might be reasonable, but needs some details**

It is ad hoc. But, as we noted in the text, it was previously shown that the GEOS-Chem CH_4 simulation was positively biased in the extratropical stratosphere at a resolution of 4° x 5°. We chose the negativity bound to speed up convergence in the extratropical stratosphere, to remove this known bias. The L-BFGS-B algorithm provides a means of doing this as part of the optimization.

- **p.13 sect.3.1: more than 2 full pages, 8 figures, a lot for one section... consider splitting**

As we mentioned above in our response to general comment 1.1, we significantly shortened this section. In doing so, we have removed 4 figures.

- **p.13 l.27: fig.3 may be misleading; it shows the integrated 'footprint' of GOSAT observations, but the inversion uses increments depending on the deviation from the prior and observations (or truth)**

Yes, the inversion uses the increment, but it is unclear how one would show the spatial distribution of the sensitivity for each increment over the whole 4-month period, that is easy for the reader to interpret. The integrated sensitivity over the four months seemed to be the best approach. Regardless, we have removed the figure to shorten the manuscript.

- **p.14 l. 7: I disagree. It only shows that the matrix Q you chose is incorrect for that set-up... it feels that your results are so dependent to the subjective choice of Q that they are hardly exploitable**

We respectfully disagree with the referee. The specification of Q is not the issue here. As Referee #4 noted, "in this setup the strong constraint assimilation does better than the weak constraint because it puts all the measurement information into the cause of the difference, the emissions, by design." Because we have put all of the bias in the emissions, at the surface, it is not surprising that the strong constraint assimilation, that assumes all of the bias is in the emissions, outperforms the weak constraint assimilation. In retrospect, given the comment from Referee #4, we realize that this experiment is not a meaningful evaluation of the weak constraint assimilation. We have therefore removed this experiment.

- **p.14 l. 31: again, it only shows that Q is ill specified**

- **p.14 fig.6: Right column: Due to ill specified Q, WC applies correction upwind in the Atlantic ocean to improve the situation over the Amazon basin. This could be highly misleading for diagnosing model errors!**

This discussion on Page 14 (starting from Line 25) was regarding the results of the OSSE in which we turned off convection. As shown in Figure 6, turning off convection results in an accumulation of CH_4 near the surface over the continental source regions, and a deficit aloft, downwind of the source regions. The positive correction to the deficit over the Atlantic and the negative correction to the excess CH_4 over the Amazon (and central Africa) is exactly what we hope the bias correction would do. It is not an indication of an ill-specified Q. Clearly, changing Q will impact the small-scale features in the corrections, but the large-scale corrections to the excess CH_4 near the source regions and to the deficit aloft is an indication that the scheme is working.

- **p.15 fig 6-7-8: why not including SC in these figures for comparison?**

We did not impose any flux errors in these experiments, consequently there would be no value in using the strong constraint assimilation in them. Our objective here is simply to evaluate the performance of the weak constraint scheme to mitigate the convection bias.

- **p.15 fig.11 and last OSSE: I didn't really get this last OSSE. why initial conditions at different dates? In the end, it is probably on of the most important OSSEs as it show the capability of the WC to correct for long-term errors; but the way it is presented makes it hard to understand**

Figure 11 (in the original manuscript) was not well-described. We only introduce a bias in the initial conditions for 1 February 2010. Figure 11 showed how that initial bias decreased in time in different regions of the atmosphere. We have removed this figure and instead briefly summarize the main result in the text.

- **p.18 l.34: is it really necessary to extent ACE-FTS profiles to compare with GEOS-CHEM? can't you produce real equivalents?**

To map the ACE-FTS data into XCH_4 space we do need to extend the ACE-FTS profile to construct the full column. However, when looking at the differences between the modeled and ACE-FTS XCH_4, the tropospheric component of the column from GEOS-Chem is removed, leaving only the real differences in the stratosphere and UTLS. To use real observations to extend the ACE-FTS data, we need global profile observations of CH_4 that have near spatial and temporal coincidence with the ACE-FTS measurements. We are not aware of such observations for the period of interest.

- **p.19 l.30: such results is quite obvious with a one month assimilation window; please comment accordingly...**

Please see our response to General Comment 1.3 above.

- **p.20 l.5-10: Gives the impression of fitting pre-conceived conclusion at all costs... the results are not convincing in that direction**

In all data assimilation applications, if there is a known bias in the model it is important to try to mitigate that bias. It is not an issue of fitting a pre-conceived conclusion in the assimilation. The positive bias in the GEOS-Chem stratosphere at the 4° x5° resolution was previous identified, as we noted in the text. In their inversion analysis, Maasakkers et al. (2019) accounted for this bias by fitting a second order polynomial to the modeled background as a function of latitude. We

showed that the weak constraint scheme can mitigate this bias in the context of the assimilation. That is a desirable outcome. In addition, as we noted, at 2° x 2.5°, the bias changed sign between 30°–40°N, and the weak constraint scheme was able to reduce both the positive and negative bias in the stratosphere. No one had previously documented the difference in the bias at 2° x 2.5°, yet the assimilation was able to mitigate it.

Reference:
Maasakkers, J. D., et al.: Global distribution of methane emissions, emission trends, and OH concentrations and trends inferred from an inversion of GOSAT satellite data for 2010–2015, Atmos. Chem. Phys., 19, 7859–7881, https://doi.org/10.5194/acp-19-7859-2019, 2019.

Anonymous Referee #3

The effort at assessing the errors in the (chemistry-)transport model before going into flux inversions is necessary and the issue remains too rarely treated adequately in in- version papers. It is therefore a very good idea to present a methodology to tackle this issue. The explanations of the methodology and most of the interpretations of the results, regarding mainly biases in the transport, are clear and interesting. Nevertheless, parts of the study seem convoluted and misleading:

1. **what is the purpose of the "comparison" to the so-called "strong constraint" (SC) inversion? SC is said to consider that the model is perfect but this is not true since the covariance matrix R can contain errors in the model as well as errors in the measurement, representativity errors, etc. Moreover, it is always possible to invert transport variables/parameters along with the emission fluxes and boundary conditions - which may require assimilating more data, as is suggested but in the case of WC and only for glint measurements. Finally, the comparison between two data assimilating systems would be "fair" and useful only if the same information is provided to both. Here, it is not very clear what the configuration of the SC inversion is but it seems that it includes a lot less information than WC: for example, WC uses data to estimate the noise on GOSAT XCH4 (p. 9) but it is not stated how this knowledge is used in SC. I would suggest to simply drop this "comparison" between SC and WC: the inconsistency between assessing errors in the model (WC) and inverting fluxes without taking these errors into account (as seems to be the case in SC) is too strong.**

The comparison between the weak constraint and strong constraint assimilation is intended to highlight the impact of not accounting for the errors in the strong constraint assimilation. The two assimilation schemes are ingesting exactly the same information. There is no difference in the GOSAT data assimilated by the two schemes. The only difference is that the strong constraint scheme adjusts the surface emissions, whereas the weak constraint adjusts the surface emissions as well 3D sources and sinks in the model state. We have expanded the description of the assimilation schemes in Section 2.3 to better explain the strong constraint method.

Regarding the issues as to whether the model is considered perfect in strong constraint 4D-Var, and the possibility of capturing model errors in the **R** matrix, please see our general comment above.

2. **what is the final aim of the characterization of the model errors? Is it to improve the model? In this case, the study lacks suggestions on how to do this (change convection schemes? use only high resolutions?). Is it to gain insights on how to invert fluxes? In this case, two questions are to be answered, which are not discussed as such in the paper:**

   a) **are the model errors too large compared to the errors due to the emission fluxes to allow meaningful inversions of the fluxes?**

   b) **if inversions are to be run with the model as it is, how can the errors in the model be taken into account so that the optimized fluxes are actually meaningful (e.g recommendations on the building of the R matrix for a so- called SC set-up)?**

Previous work by Yu et al. (2017), based on the GEOS-Chem simulation of ^{222}Rn, ^{210}Pb, and ^{7}Be, suggested that the vertical transport in the model is too weak at the coarse resolution. We therefore applied the weak constraint assimilation here to characterize these biases in the context of the CH$_4$ simulation. In our companion paper in GMD we used the insight gained from the assimilation work presented here to better understand the source of the CH$_4$ biases. As we stated in Stanevich et al. (2020) in GMD, that paper "complements Yu et al. (2017), with a specific focus on the impact of model resolution on the CH$_4$ simulation and the goal of better understanding the source of the biases identified in Stanevich et al. (submitted)." Consequently, it is in the companion paper where we discuss the implications of the bias for flux inversion analyses and ways of improving the model. One conclusion in the companion study was that because of the magnitude of the errors at 4° x 5°, "we do not recommend the 4° x 5° GEOS-Chem model for CH$_4$ inverse modelling." If one wants to use the 4° x 5° model for flux estimation, the weak constraint approach would offer the best means of mitigating the bias, instead of trying to do so through the **R** matrix. In the companion paper we also suggested that one way to improve GEOS-Chem and reduce the vertical transport bias was "by archiving and globally remapping the native resolution horizontal [air mass fluxes] in order to drive advection at the coarse resolution instead of calculating the horizontal [air mass fluxes] from the coarse-resolution wind fields."

3. **what does the proposed methodology bring to the assessment of the model errors? In the discussion of the results of this study, it is not clear what it brings compared to previous studies: the errors assessed in this paper seem to be already well-known through other methodologies.**

As we discussed above, previous work had identified a stratospheric bias in the CH$_4$ simulation, but previous work in the troposphere had focused on the simulation of ^{222}Rn, ^{210}Pb, and ^{7}Be. The weak constraint scheme enabled us to characterize the structure in the model bias in CH$_4$, which motivated the analysis presented in the companion paper, and in which we quantified the contribution to the bias that arises from diagnosing the vertical transport from the horizontal winds and from the loss of eddy mass flux in the low-resolution meteorological fields. The focus in this manuscript is just on the assimilation results. We refer the referee to the companion paper for the discussion of the implications of the bias identified here.

Specific comments

General

12

Beware of the use of 4D-Var: it should be used only in cases when the problem is actually 4 dimensional. This is the case in meteorology, where the problem is on the initial conditions. Usually, it is not the case for flux inversions, which are problems on the boundary conditions: the relation between the 3D maps of fluxes through time is not taken into account in the model (only the relation of concentration fields through time is). It would be best to use only "variational" and drop the 4D in these cases of flux inversions.

Please see our general response above regarding the issue as to whether this approach should be called 4D-Var.

Abstract

- **p.1, l.6-7: "capable of differentiating the vertical distribution of model errors": what does differentiating mean here?**

We changed this to read "capable of providing information on the vertical structure of model errors."

- **p.2, l.2: "indicating the presence of resolution-dependent model errors": do you mean errors directly linked to the parameterizations in the physical core of the model or representativity errors also?**

These errors are due to the way the vertical transport is diagnosed from the low-resolution wind fields. A detailed discussion of this is presented in the companion paper.

- **p.2, l.4: "However, a major limitation of this approach is the need to better characterize the specified model error covariance in the assimilation scheme": this is true for all data assimilating systems in our domain. What does this study bring to this issue?**

The referee is correct. This is an issue for all data assimilation systems. We highlighted this in the abstract and in the conclusions to make it clear that this remains an issue. The weak constraint scheme offers no novel insights into characterizing the covariance matrices.

Section 1 Introduction

- **p.2, l.21-22: "the impact of biases in chemistry and transport are often neglected": this is not exactly true, a lot of studies try and deal with biases in various ways e.g. debiasing previous to the inversion itself, specifying adapted R matrices, etc. Please look deeper into the available studies.**

The referee is correct. We have modified the text to read that "the impact of biases in chemistry and transport are often neglected or accounted for using various ad hoc approaches."

- **p.2, l.32: " In contrast to the traditional "strong constraint" (SC) 4D-Var method, the WC scheme does not assume that the model is perfect": same remark as above, the so-called SC method does not imply that the model is assumed to be perfect, it implies that all errors are taken into account in the R (and B) matrix so that what is not is "perfect".**

Please see our general comment above regarding the assumption that the model is perfect in strong constraint 4D-Var.

- **p.3, l.16-17: "Highly accurate aircraft CH4 profile measurements would be an ideal source of information, but they are limited in space and time.": what about aircores?**

AirCores provide excellent information on the vertical distribution, but they are still limited in space and time. We have included AirCores in the revised text.

Section 2 Data and Methods

- **p.4, l.11-12: "due to the dependence of wetland emissions on the meteorological fields": do you mean that wetland emissions are actually recomputed from the references provided above with the regridded GEOS meteorological fields? Or computed on-line in GEOS-Chem?**

The wetland emissions are computed online in the version of GEOS-Chem used in the analysis, but, as described in the manuscript, this results in slight differences in the emissions between the 2° x 2.5° and 4° x 5° resolutions (due to small differences in the regridded meteorological fields). To ensure consistency in the emissions between the two resolutions in our analysis we regridded the emissions from the 4° x 5° simulation to 2° x 2.5°.

- **p.4, l.21: "the period of 1 February 2010 to 31 May 2010": this is a very short period of time, which does not allow for one full seasonal cycle for mid-latitudes. Why not work on a full year?**

We are interested in assessing the utility of the weak constraint 4D-Var scheme in capturing the model bias. To do this there is no need to run the model over the whole seasonal cycle. However, a longer analysis would provide a means of examining seasonally varying biases and evaluating the impact of the seasonally varying observation coverage of GOSAT on the ability of the scheme to capture the biases. Such a study is broader in scope than the work presented here, but would be an interesting follow-up study.

- **p.4, l.22-25: "5.5 years ... initial condition for the analysis period": this seems a bit convoluted. Why not run a spin-up of about 9 years (life time of CH4) OR assimilate data to obtain initial conditions?**

We did assimilate GOSAT data to obtain optimized emissions for the initial conditions. As we stated in the manuscript, we ran without data assimilation for "5.5 years until July 2009. From July 2009 to January 2010 we assimilated the GOSAT Proxy XCH₄ retrievals (Parker et al., 2015) to obtain monthly mean emission estimates at 4° x 5° resolution. The optimized emissions were then regridded and used to perform forward model simulations at 2° x 2.5° resolution for the same period from July 2009 to January 2010. The updated model fields on 1 February 2010 at both model resolutions were taken as initial condition for the analysis period." Thus, the initial conditions on 1 February 2010 are based on the a posteriori CH₄ from optimized emissions obtained from the assimilation of GOSAT data from 1 July 2009 to 31 January 2010.

- **p.5, l.22-25: "The use of the alternative CO2 fields did not change any of the findings about model errors in our study. There may still be unidentified biases in both retrieval products. However, the fact that both CO2 fields were obtained using different methodologies gives us confidence in our results." Were both fields used for all the following results? But only one set of results is presented and there is no comment about a sensitivity test. Also, why**

not use the default field? How was it not satisfying?

Our assumption was that the use of CO_2 fields from GEOS-Chem would be preferable for assimilation of the proxy XCH_4 in GEOS-Chem, as the structures in the modeled CO_2 would be consistent with those in CH_4. However, as we discussed in the manuscript, the differences in the resulting XCH_4 fields were small, with differences typically less than 3 ppb. Not surprisingly, we found that these small differences had no consequential impact of the initial assimilation with the two data sets. It demonstrated that the proxy approach works well. Instead of reverting back to the original CO_2 fields, we used the GEOS-Chem CO_2 fields for all of the analyses presented in the manuscript.

- **p.5, l.30-31: "However, we expect vertical structure to emerge from atmospheric transport patterns": what about the OH field patterns (e.g its vertical structure)?**

Because the lifetime of CH_4 is long, we would not expect much vertical structure in the errors in the CH_4 field due to OH. This can be seen in Figures 8 and 9 in the original manuscript (which are Figures 3 and 4 in the revised manuscript) for the OSSE in which we turned the OH sink during the assimilation period. The chemistry related bias was fairly uniform throughout the troposphere.

- **p.6, l.4-6: "Such precision could be enough in many regions of the world to improve knowledge about CH4 a priori surface emissions. However, the presence of potential model errors significantly undermines this assumption." This is not clear: which regions? Improve how? By decreasing the uncertainties by how much? What would be the required ratio between the model errors and the precision and the expected improvement of knowledge?**

Regions such as North America, where Sheng et al (2018) showed that enhancements in XCH4 above the background in North American are about 10 ppb (as measured at Lamont). With a random error of about 12 pbb on each retrieval, assimilation of the GOSAT data should provide constraints on regional emissions when they are aggregated in space and time. We now explain this in the revised manuscript. It is unknown what would be the "required" ratio, but clearly one would expect the emission signal to exceed the measurement noise (which will depend on the degree to which the data are aggregated).

Reference
Sheng, J.-X., et al.: 2010–2016 methane trends over Canada, the United States, and Mexico observed by the GOSAT satellite: contributions from different source sectors, Atmos. Chem. Phys., 18, 12257–12267, https://doi.org/10.5194/acp-18-12257-2018, 2018.

- **p.6, l.19-23: why use only flasks and not continuous measurements?**

The flask data are widely used by the GHG inverse modeling community, so we chose to use them in our analysis. Given our interest in correcting the vertical distribution of CH_4 in the assimilation, we focused more on incorporating the HIPPO, TCCON, and ACE-FTS data in the evaluation, rather than additional surface data.

- **p.7, l.2: "Retrievals are bias corrected based on comparisons with calibrated aircraft and AirCore profiles." Which aircraft profiles? Be sure not to use them in the assimilation if the validation data is to be kept totally independent. Why not use also aircores (others than**

the ones used by TCCON) as validation data?

We refer the referee to the Wunch et al. (2015) paper for details of the TCCON validation. It is not within the scope of our modeling study to discuss the details of the TCCON validation. In our assimilation, we only assimilated GOSAT XCH$_4$ data.

As regards the aircore data, they are another useful validation data set. We did not consider them when we began the project because we were focused on finding validation data with a more global distribution. However, given the increasing availability of aircore data, if we were starting the project today, we would include aircore data in the validation data set.

- **p.7, l.27: "the adjoint forcing commonly used in 4D-Var": please clarify what "adjoint forcing" means.**

"Adjoint forcing" is often used to describe the gradient of the cost function. Here in the text, we drew attention to the fact that our use of the term "forcing" is different from the traditional use of the expression.

- **p.7, l.21-22: "This is the assumption that is employed in standard 4D-Var, which is also referred to as "strong constraint" 4D-Var because the model trajectory is used as a strong constraint in the optimization." Same remark as in the General comments about the use of "4D-Var": the concentrations are linked through time but not the fluxes.**

Please see our general comment above regarding the use of 4D-Var to describe the assimilation scheme.

- **p.7, l.29-30: "The 4D-Var problem to estimate surface emissions is transformed into a 3D sources and sinks estimation problem": same remark as in the General comments: flux inversions are generally not actual 4D-Var and the cost function used is the same as what is shown after, but for the Q term. Does anything prevents Q from being included in R or in B? Please clarify the mathematical and technical differences between SC and WC.**

Please see our general comment above regarding issue as to whether the flux inversion problem is a 3D problem.

We have added the strong constraint cost function to the manuscript and additional text in Section 2.3 to better explain the differences between weak and strong constraint 4D-Var as used in our analysis.

- **p.9, l.9-14: how long does it take to run for practical cases?**

It takes about 12 days to complete 25 iterations of the 4-month weak constraint assimilation at a resolution of 2° x 2.5°.

- **p.9, l.18-23: "For each WC inversion ... about 10 ppb": this is a good idea but the validation data are not independent anymore since information contained in them is actually used in the inversions. Please clarify how this issue is dealt with when comparing to the data for validation.**

We do not assimilate any of the validation data. We only assimilate the GOSAT XCH$_4$ data and compare the resulting a posteriori fields during the 4D-Var iterations (as the cost function is minimized) to the validation data (i.e., the in situ, TCCON, and HIPPO data).

- **p.10, l.3: "Therefore, we did not attempt to characterize global pattern of model errors on shorter time scales": even with a period of three days for GOSAT data, information is available at shorter time scales e.g. in the GOSAT data and in the meteorology. Would this not make it possible to characterize patterns of model errors at shorter time scales?**

Even with global coverage every three days, there are still significant gaps in the GOSAT observational coverage due to cloud cover, for example. It is possible that we could estimate the model errors at shorter time scales than three days, but we have not investigated this.

- **p.10, l.4-6: "Little is known about the a priori structure of the model errors": this is not what appears from discussions and references cited afterwards: at least some elements such as the role of the horizontal resolution or the tropo-strato gradient are known. Is this information not usable in the inversion framework?**

The stratospheric bias in CH$_4$ was known. However, it was unknown what was the impact on CH$_4$ of the weakened vertical transport identified by Yu et al. (2017) in their ^{222}Rn, ^{210}Pb, and ^{7}Be analysis.

- **p.10, l. 9-end and p.11, l.1-2: the issues described here are the same as when building the R and B matrices in the so-called SC case. Therefore, the exploration of the nature of errors in the modelled CH4 uses the strong assumption of a diagonal Q. What are then the advantages of this methodology compared to the usual R and B matrices in the so-called SC, with various set-ups for R for example?**

As we discussed above, this weak constraint method of capturing the model error is consistent with the 4D-Var formalism. We have not explored the utility of trying to account for the model error using the R and B matrices in the context of the strong constraint assimilation.

- **p. 11, l.9-11: "Therefore, we considered a uniform structure for Q to be   a satisfactory assumption for this initial assessment of model errors in the context of the WC 4D-Var analysis": is this statement justified by "expert- knowledge"? Is it not possible to design sensitivity tests to assess the impact of this strong assumption?**

As a follow-up analysis, we are examining the impact of a non-uniform Q on the weak constraint assimilation.

- **p.12, l.9: "we believe that the OSSEs should reveal the best performance of the WC method": it is a bit dangerous to show the performances of a methodology only in best-case scenarios since the application to realistic cases may show the tool to be very limited.**

The text here was referring to the fact that most of the OSSE results are presented for March, which is the middle of the assimilation window, when one would expect the 4D-var scheme to provide the best estimate of the state. This would the case even with real data. We actually chose extreme model biases for the OSSEs that we expected would be challenging for XCH$_4$ data to help mitigate.

- **p.12, l.15: "for the real world applications, we expect less extreme model errors": would the method proposed here be able to characterize the errors if they are smaller? See also comment above.**

Please see our response to Referee #2 above (on Page 8).

- **.12, l.21-22: "We also conducted SC 4D-Var assimilation experiment for comparisons with the WC approach in the OSSE with biased surface emissions." see General comments. Without more details, it is not possible to understand the differences between WC and SC. If SC contains less information than WC, the comparison is not really meaningful and interesting.**

The weak constraint and strong constraint assimilations are assimilating exactly the same observations. We have added a description of the strong constraint assimilation (including the SC cost function) in Section 2.3 to make this clear. Nevertheless, we have removed this experiment since it was not too informative, as noted by Referee #4.

Section 3 Results

- **p.13, l.26 - p.14, l.4: how do the sensitivities in the model compare to the errors in the actual GOSAT data?**

The sensitivity is of XCH_4 with respect the CH_4 in the model state. We cannot directly compare them to the GOSAT precision. One would have to define a given perturbation in the state and ask how that perturbation, projected through the sensitivity, compares to the GOSAT errors.

- **p.14, l.7-8: "when using the SC method, we implicitly supply the assimilation with knowledge about the source of the bias": how?**

In this OSSE, as noted by Referee #4, we are imposing all of the bias in the surface emissions, and by design, the strong constraint assimilation assumes that all of the bias is at the surface emissions. As a result of this aspect of the OSSE, and to shorten the manuscript, we have removed this OSSE from the manuscript.

- **p.14, l.11-12: "Due to weak vertical sensitivity of the pseudo-data, it is difficult for the WC 4D-Var method to mitigate strong localized vertical bias." The low sensitivity is a problem for data assimilation in general. What is particular to WC here?**

We are not claiming that this is particular to the weak constraint assimilation. We are only noting that because of the limited vertical information in the XCH_4 data, the assimilation cannot mitigate strong localized vertical bias.

- **p.14, l.13-14: "Instead, it compensates for the bias by applying relatively weak CH4 state adjustment of the opposite sign in the column of the atmosphere above, particularly in the stratosphere": what are the consequences of the creation of such a dipole?**

We have not examined the consequence of this in the OSSE. We anticipate that it would not be too consequential since the largest bias is in the middle stratosphere, which will have a minimal impact on the total column.

- **p.14, l.15-20: the additional information explicitly provided to WC could also be used in SC in the covariance matrices R and B.**

If one insists in using strong constraint 4D-Var, given the identified model bias, one could try to tune the R and B matrices to improve the performance of the assimilation, but it would be preferable to fix the model to remove the identified model bias.

- **p.14, l.33-34: "GOSAT retrievals possess sensitivity to biases in vertical transport and can distinguish...": this is not a property of the GOSAT data as such but of the whole data assimilation framework. Maybe "GOSAT retrievals contain information on the vertical transport, which can be used in our set-up to distinguish..." would be clearer.**

We have changed this to "GOSAT retrievals contain information to enable us to capture vertical transport bias even when…"

- **p.15, l.11-12: "we do not expect chemical biases to be as strongly localized as the biases associated with emissions and vertical transport": why?**

Because of the long lifetime time of CH_4, we do not expect vertical transport to provide localized biases. This can be seen in Figure 8 in the original manuscript (Figure 3 in the revised manuscript), in which we turned off the chemical sink of CH_4.

- **p.15, l.30-32: "The perfect observing system would completely remove the initial condition bias at the start of the assimilation period (on February 1). However, what is shown on February 1 is just an 8% reduction in the bias in each of the eight regions, relative to the a priori, with the rest of the bias propagated onto the assimilation period". What are the consequences of this?**

The results suggest that a "spin up" time of about two months would be desirable to avoid the impact of initial condition biases on the analysis. As a result, as noted at the beginning of Section 4.1, we only examine the stratospheric bias in May 2010, to avoid any influence from the initial conditions. To shorten the OSSE section, as requested by Referee #2, we have removed Figure 11 from the revised manuscript and instead briefly summarize the key result in the text.

- **p.16, l.3-4: " This suggests that additional vertical correlation between forcing terms in the stratosphere would be beneficial to accelerate convergence in the stratosphere." Why not test the sensitivity to such correlations?**

As we discussed in our response to Referee #2, we chose the simplest form of Q with which to evaluate the initial performance of the scheme. We are now doing experiments to quantify the structure in Q, and that will be the focus of a future manuscript.

- **p.16, l.11: "residual high latitude bias, which resembles noise or bias in the GOSAT observations." Is this due to the period of interest being winter in the Northern hemisphere?**

It is unclear what is the source of this residual bias. One would expect larger errors in the GOSAT retrievals at high latitude in winter/spring, but we have not investigated this.

- **p.16, l.19-20: "the SC assimilation leaves significantly larger residual biases." What is the impact of these on the optimized fluxes and on the uncertainty reduction?**

The larger residual biases are due to the inability of the flux adjustments to mitigate the model bias. We have not tried to quantify what is the impact of these residuals on the fluxes since we can mitigate the biases using the weak constraint scheme.

- **p.17, section 3.2.1: the TCCON and NOAA data are not actually independent from the WC (see above). How is this dealt with? If they are actually independent from the SC, the comparison between SC and WC against the fit to these data is not meaningful.**

We do not assimilate the TCCON or NOAA data. These data are used only to evaluate the assimilation. We assimilate only GOSAT data, and both the weak constraint and strong constraint schemes assimilate exactly the same data.

- **p.18, l.11: "similar errors": similar to what? The sentence is not very clear to me.**

We mean that the errors may be similar in both the 2° x 2.5° and 4° x 5° simulations. We have reworded this on Page 16 (line 21) in the revised manuscript.

- **p.18, l.9-13: these sentences seem a bit convoluted. Is the idea to state that a station such as IZO sees the upper troposphere and is best compared to the model's upper troposphere?**

We are arguing that the improved model agreement at IZO with the weak constraint 4D-Var scheme is due largely to improvements in the column in the upper troposphere and lower stratosphere, since neither model resolution can capture the topography for the island, which will result in lower tropospheric model transport errors in the vicinity of the island. We have rewritten the text here on Page 16 in the revised manuscript.

- **p. 18, l.27: what about the impact on GOSAT retrievals of the still long nights around the North pole during the period of interest?**

We are unsure as to how polar night impacts high latitude GOSAT retrieval just outside the terminator. This is beyond the scope of the manuscript.

Section 4 Discussion of Model Biases

- **p.19, l.22-23: "a stratospheric bias introduced in the system through the initial conditions ": is there no other way than the ICs to introduce a bias in the stratosphere?**

There are numerous ways the stratosphere could be biased. We intentionally introduced an initial condition bias to assess the ability of the system to remove such a bias.

- **p.19, l.29-30: "it shows that the SC assimilation attempts to correct the positive high-latitude stratospheric CH4 bias at the expense of surface emissions": this seems to be a direct consequence of the set-up of the SC inversion. More details on this inversion are required to discuss its results.**

It is a direct consequence of the fact that the strong constraint assimilation assumes that all the model bias can be mitigated through adjustment of the surface fluxes. We have added a more detailed description of the strong constraint assimilation in Section 2.3.

- **p.21, l.6-14: can you deduce some recommendations for improvements in the model? If not, how can this knowledge be used for setting up inversions for fluxes?**

See our response to general comments 2 and 3 above.

- **p.21, l.15-23: same questions as above: recommendations for improvements in the model? Information for flux inversion?**

See our response to general comments 2 and 3 above.

- **p.22, l.25-26: what could be concluded? Recommendations for improvements in the model? Information for flux inversion?**

See our response to general comments 2 and 3 above. The assimilation results presented here motivated the work in our companion GMD paper, in which we examined the sources of the model errors and made recommendations for improving the model.

- **p.23, l.13-14: "Finally, the results strongly suggest that the WC assimilation and the GOSAT observations have the potential to diagnose transport errors at both model resolutions" What would prevent the assimilation of data to bring information on errors at any resolution (if the data are relevant)?**

The referee is correct. If the data have sufficient information it should not matter. We have removed the statement from the manuscript.

Section 5 Conclusions

- **p.24, l.1-2: " However, characterizing these correlations will be challenging": is it possible at least to design sensitivity tests to explore various possibilities for these error correlations?**

Yes, and that will be the focus of a future manuscript. The current manuscript, together with its companion GMD paper, represents a significant effort to adapt the weak constraint scheme for CH_4 assimilation. Developing these correlations and expanding the manuscript to incorporate sensitivity tests with the correlations would have made the manuscript unwieldy.

- **p.24, l.5-6: "Initial comparisons suggested that GEOS-Chem was affected by biases not solely related to discrepancies in surface emissions." This is not very informative as it is well known and flux inversion would be relatively straightforward otherwise.**

Yes, it is well known that models have biases, and in the subsequent sentences we discussed the nature of these biases in GEOS-Chem and the fact that the weak constraint 4D-Var scheme was able to mitigate them.

- **p.24, l.13-14: "Meanwhile, the results showed that running the a priori model at 2x2.5resolution produced better agreement with TCCON observations than the a posteriori**

fields from the SC 4D-Var surface emission optimization at 4x5." What can be deduced from this result?

As we concluded in the companion paper in GMD, without correcting for the biases, "we do not recommend the 4° x 5° GEOS-Chem model for CH_4 inverse modelling."

- **p.24, l.31-32: "glint measurements": a lot of other data could be assimilated, not only satellite data related to the concentrations of a given species.**

Yes, there are a lot of other data, but satellite observations give the best global observational coverage. We stressed the need for glint data here, in the context of the shortwave infrared measurements, to obtain a more global coverage, particularly over the oceans.

- **p.25, l.2: "if the model were assumed to be perfect, as is the case in SC 4D-Var": see General comments: this is not so simple.**

Please see our general response regarding the assumption that the model is perfect. We now state that "…if the evolution of the model state were assumed to be perfect."

- **p.25, l.7: "Potentially, any CTM may be improved if the signal from the surface emissions can be separated from other model errors." This is only relevant if the objective is to optimize fluxes. A given CTM and set up may work very well with errors compensating one another for other objectives (e.g forecasting). Be more specific on what improvements could be made, how and to which purposes.**

Our overall interest here is in optimizing fluxes. We have added text to make this clear.

- **p.25, l.14: "regional-scale analysis at higher spatial resolution": how in practice? The link to what it is used for in this study is not plain for me.**

By optimizing the state, we can provide a 4D field of CH_4 that can be used as boundary conditions for a regional model.

Technical corrections

- **p.10, l.33: "the errors in the model CH4 simulation" -> the errors in the modelled CH4 concentrations?**

We have changed this to "explore the structure of the errors in the model."

- **p.11, l.19: "The sensitivity of the GOSAT observations to the modelled state" -> The sensitivity of the equivalent of GOSAT observations to the modelled state**

We have removed this text to shorted the OSSE section, as requested by Referee #2.

- **p.15, l.13: "in the Fig.8" -> remove "the"**

Removed.

- **p.18, l.10: "inland" -> island?**

Fixed.

- **p.18, l16: homogenize including/excluding Sodankyla.**

Fixed.

- **p.22, l.9: "lofted" -> lifted**

Changed.

Tables and Figures

- **Tables 1, 2: please use consistent names for the columns in the legend and table.**

Fixed.

- **Fig. 4: the black boxes are not so easy to see: maybe use a very different color (try pink?)**

We have removed this figure from the manuscript.

- **Fig. 15: do not repeat in the legend what appears in the graph itself so that is it easier to read and may be kept on one page.**

We have shortened the description in the figure caption so that it is easier to read.

- **Fig. 16: reduction (third column) may be easier to evaluate in % and absolute value.**

We believe that the change in the absolute bias provides the reader with a meaningful sense of the improvement that the assimilation offers.

Anonymous Referee #4

General Review Comments:

This is a significant and important body of work and I believe the community will find it very interesting. The main limitations I see are in terms of the figures which don't always back up the assertions in the text. Sometimes, in the case of Fig6/Fig7, they clearly support the text but seem to fail because of lack of difference plots or poor choice of color scales. The text is sometimes hard to follow because of the multiple terms being used, eg. state/forcings/3D CH4 adjustments and parameters vs. surface fluxes. After providing the theory, it might be wise to work with common terminology.

Furthermore, I particularly found the assumed variability of the "forcings" in time and space, somewhat hard to follow. In summary though, the work is quite interesting and novel in its handling of the main source of error in atmospheric trace gas inversions (transport).

We have shortened the manuscript by deleting some of the material in the OSSE section. We have also tried to improve the description of the experiments to make it easier for the reader to follow the discussion.

Specific Comments:

caption Figure 2: "and smoothed with the GOSAT averaging kernels" The authors should give the exact equation they use to calculate modeled XCH4 from their model: is the prior CH4 profile assumed in the retrieval used in this calculation, in addition to the averaging kernel, or not?

We have added the equation (Eq. (1) in the revised manuscript) describing the GOSAT retrievals in terms of the averaging kernels and a priori profile.

Figure 6: Fig 7 provides a motivating summary of Fig 6. However, looking at Fig 6 relative to its colorbar, it is hard to see that the two even provide the same information. I'd really consider adopting a more dynamic color palette for this image.

Our intention with Figure 6 was to give the reader a sense of the spatial patterns of the imposed model bias and the resulting correction. Because we are showing the differences at the surface and in the upper troposphere, and vertically in the tropics and mid-latitudes, it was difficult to find a scale that will allow the reader to quantitatively compare the plots. Thus, we included Figure 7. We have added text to emphasize that the focus of Figure 6 (Figure 1 in the revised manuscript) is the spatial patterns of the model error corrections.

p6 l10: "...with smaller [modeled] XCH4 in the SH..." A difference plot would be helpful in Figure 2. I am having a very hard time seeing how XCH4 is lower in the SH in the model (top panel) than in the measurements (bottom panel) – the opposite appears to be the case, to my eye.

See the response below.

p6 l11: I agree that China, India, and equatorial Africa are higher in the model, but South America appears to be lower, at least over the Amazon. Again, a difference plot in Figure 2 would help.

We apologies for the confusion here. The discussion regarding Figure 2 was about the spatial patterns of the XCH_4 field in both the model and observations. The XCH_4 is smaller in the southern hemisphere than in the northern hemisphere in both the model and observations. Similarly, there is enhanced XCH_4 over South America and equatorial Africa in both the model and observations. We have removed Figure 2 to shorten the discussion in this section, as requested by Referee #2.

p7 l29-on: "The 4D-Var problem to estimate surface emissions is transformed into a 3D sources and sinks estimation problem..." I think you ought to be a bit more precise with the wording here. The 4Dvar method in general can be used to estimate both surface sources/sinks and 3D fields, either using the strong- or weak-constraint dynamics. I think what you want to say is that for your application, which in the past was solving only for surface fluxes and not 3D fields, you are now allowing this 3D forcing term to be solved for, as well. You might want to mention here that you are not too concerned with what this 3D field represents in your application – it could be a chemical source/ sink, a correction to the 3D concentration field, or a dynamical error. That comment would help those who are

accustomed to thinking of this term as being solely a dynamical error term in the weak-constraint approach.

We have added text clarifying this point on Page 8 (lines 2-3) in the revised manuscript.

p8 L15: There is a spurious extra term in the cost function here that should lead to   some "double counting" of the dynamical errors. To be specific, the last two terms in the cost function both involve the dynamical mismatches, while only one is needed, and that term ought to be: ... + [-Gu_i]ˆT [Q_i]ˆ-1 [-Gu_i] , which equals ... + (x_i - M(x_i-1,p))ˆT [Q_i]ˆ-1 (x_i - M(x_i-1,p))

If we note that lambda_i = [Q_i]ˆ-1 (x_i - M(x_i-1,p)), then there is no need to carry around the additional variable u_i, and u_i can be calculated from lambda_i as

G u_i = - [Q_i] lambda_i

Similarly, it is not necessary to have equations for the partial of L with respect to both lambda_i and u_i.

In the Lagrangian, the first term is the observational mismatch, the second term is the a priori constraint on the emissions, the third term is the a priori constraint on the state corrections, and the fourth term is the constrain from the dynamics. We are not "double counting" the influence of the dynamics.

The Referee is correct, we do not need the partial of the Lagrangian with respect to λ_i. We have removed this equation.

p10 L5: "... a priori estimates of the model errors were set to zero." Just to be sure here, you are setting the actual initial model errors themselves equal to zero (u_i=0) and not the assumed uncertainty in those same errors (Q_i=0), correct? Because setting Q_i=0 would be equivalent to reverting back to the hard dynamical constraint. Writing this out with the mathematical symbols would eliminate any uncertainty the reader might have on that score.

Yes, we are setting u = 0. We now write this out mathematically on Page 10 (line 6).

p10 second paragraph: This discussion of the relative weighting of the forcing versus emission parts of the solved-for control vector points to something that could be simplified in this approach. It can be shown that the forcing vector -Gu_i at any iteration is simply [Q_i] lambda_i – that is, the adjoint state vector times the assumed model error covariance matrix. The forcing vector does not need to be solved for in the control vector – it is already solved for in the strong-constraint 4Dvar, essentially, since lambda_i is solved for. All that needs to be done to implement the weak-constraint version is to add the forcing vector (=[Q_i] lambda_i) onto the state at each step during the forward runs (although saving lambda_i at fine temporal resolution, at the timestep of the model, for example, can take a lot of memory and I/O time). Once this is recognized, it is clear that the full parameter space of the forcing vector can be solved for, up to the temporal resolution limits just mentioned. Since it is not too clear what the magnitude of [Q_i] ought to be, some experimentation with the relative

weighting between the forcing and emissions parts of the cost function is probably still needed, even with the simplification just noted.

We thank the referee for the suggestion. We agree that once the sensitivities have been calculated in the strong constraint case, it is straightforward to extend the model to estimate the forcing terms in weak constraint 4D-Var. We will consider the implications of the referee's suggestion for simplifying the way we have implemented the scheme in the model.

p10 L32-34: "Instead, we use the WC 4D-Var method to optimally constrain the CH4 state and explore the nature of the errors in the model CH4 simulation." Actually, you   are not really solving for an optimal estimate of the CH4 state (the 3D CH4 field at each timestep), because you have not put this in the control vector that you are solving for. This state evolves according to your model M and has dynamical constraints upon it. What you are really solving for is some sort of error term on the state, which might be thought of as a dynamical error, or a 3D source/sink (e.g. a chemical one) if one is not modeled, or an error on a modeled 3D source/sink. Since the measurements have error on them, this also allows large measurement errors to be given less importance in the problem (the measurement error is turned into a large forcing error and its effect is on the actual parameters solved-for is lessened).

The referee is correct. This is a more accurate description of what we are doing. We have modified the text to state that we are using "the WC 4D-Var method to constrain the 3D corrections to the CH$_4$ state."

p10 L33-34: "We performed two types of inversions: "full state assimilation" and "flux+state assimilation"." Point back to your description of these two cases in Section 2.3 here. I had already forgotten that you described them earlier by this point. Also, it wouldn't hurt to say explicitly that the "full state" case solves only (or mainly) for u, while the "flux+state" case solves for both u and p. Spell it out, especially since the state is defined as "x" earlier, and you don't actually solve for the full 3-D field "x" in your control vector, no? I guess you get the 3D CH4 fields by adding on the dynamical errors that you do solve for onto the state using equation (2), right?

Thank you for the suggestion. We have modified the text as suggested.

p11 L13: This would be a good place to mention that, in all these OSSEs, the same transport model was used to generate the truth as was used in the inversion. In other words, these are all "perfect model" experiments in all respects except the isolated error signal looked at in each experiment. This is important to say, because most readers will be thinking of the weak constraint as a way to handle general model errors, and might be thinking that you used a different transport model for the generating the truth and doing the inversion (I will actually advocate adding an OSSE of this sort below in the emissions case.)

We have added text stating this on Page 11.

p13 L29-31: Isn't the greater sensitivity in the upper troposphere / lower stratosphere also due to the fact that the GOSAT Xch4 averaging kernel is weighted most heavily in the mid- to

upper-troposphere (it being a thermal IR measurement)? A small figure showing what the GOSAT XCH4 averaging kernel actually looks like might be valuable here.

We are using the shortwave infrared retrievals (see Section 2.2.1), so the averaging kernels are heavily weighted toward the lower and middle troposphere. Since the manuscript is already too long (with 19 figures in the submitted version), and because plots of the averaging kernels are in the published literature, we would prefer not to include such a plot in the manuscript. We added the Yoshida et al. (2011) reference in Section 2.2.1, which contains plots of the averaging kernels.

p14 L9: "The results confirm that the SC 4D-Var method better removes CH4 biases due to emissions." This is misleading. Yes, in this setup the SC does better than the WC because it puts all the measurement information into the cause of the difference, the emissions, by design (it doesn't solve for the 3D forcing corrections) as you have noted. But in general case of when there are model errors, the WC case should solve for emissions more accurately than the SC case. It would be really useful if you could have done an additional OSSE here in which you introduced emissions errors AND dynamical errors in the truth, then tried to estimate both using the SC case and the WC "flux+state" setup. That ought to show the WC case doing better, since it would correctly partition the dynamical and emissions errors to the forcing terms and emissions parameters, whereas the SC would incorrectly attribute the dynamical errors solely as emissions errors. You should use a completely different transport model in generating the truth from what is used in the inversion to make this case realistic. The dynamical errors in Q should be based on the differences between the two transport models. I think that that new OSSE would demonstrate the benefit of the WC approach for that case that a lot of readers care about: how much their surface emissions estimates are degraded by model error when using the SC approach.

The referee is correct. In retrospect it is not surprising that the strong constraint performs better when we impose all of the bias in the surface fluxes. Since this OSSE is not that informative, we have removed it from the manuscript, in the interest of shortening the manuscript. The suggestion of the referee to use a different transport model to show that the strong constraint assimilation will incorrectly attribute the transport errors to flux errors would have been helpful earlier in the analysis. In our companion paper we effectively show this. We found larger flux corrections with the strong constraint assimilation at 4° x 5° than at 2° x 2.5°, which is consistent with the strong constraint assimilation incorrectly adjusting the fluxes to compensate for the larger model errors at 4° x 5°.

p16 L15-18: I cannot tell from Fig 12 and 13 that the high latitude a priori bias is smaller in 2x2.5 compared to 4x5. Similar question for the model bias reduction. It may just be that the color scale saturates so quickly in the images? Are they zonal avg summary stats you could provide to back this up?

This can be seen more clearly in Figure 2 and 3 in our companion paper in GMD. In Figure 3 in the GMD paper we plot the zonal mean differences of the 4° x 5° and 2° x 2.5° model with respect to the ACE-FTS data.

p17 L10: "The results suggest that GEOS-Chem a priori CH4 simulation suffered from biases that were not related only to incorrect surface emissions." Since GEOS-Chem is not perfect (i.e. the real world has different transport) this is the most obvious source of the errors.

We agree that this is obvious, but we felt it necessary to point it out nevertheless.

p18 L5: I find it a bit perplexing that 2x2.5 sims are generally argued to provide better atmospheric transport but perform more poorly against in situ obs, which one would think would be the most sensitive to things like vertical transport. Would you consider this an indication of residual error in surface CH4 fluxes?

Yes, we believe that this is an indication of residual error in the surface fluxes. Because the constraints that the GOSAT data provide on regional fluxes is limited, the posterior flux field at $2° \times 2.5°$ is noisier than at $4° \times 5°$. For example, Maasakkers et al. (2019) found that their inversion of GOSAT data to estimate CH_4 fluxes for 2010-2015 had only 128 degrees of freedom for signal on the global $4° \times 5°$ GEOS-Chem grid.

Maasakkers, J. D., et al., Global distribution of methane emissions, emission trends, and OH concentrations and trends inferred from an inversion of GOSAT satellite data for 2010–2015, Atmos. Chem. Phys., 19, 7859–7881, https://doi.org/10.5194/acp-19-7859-2019, 2019.

Tables 1 & 2: These comparisons to independent data are powerful demonstrations that using the weak constraint approach on the GOSAT XCH4 data really does improve the estimated 3D CH4 fields better than the SC approach. Apparently either the transport or the measurements are bad enough that the WC results in a big improvement. If the GOSAT data are badly biased, adding flex in the state trajectory would help things, even if the transport were not too bad.

Yes, it would help even if the GOSAT data were badly biased. However, based on previous work, we believe that the problem is with the model.

p22 L30: I get the averaging out of resolved eddy motions causing problems but what are you referring to by "incorrect" regridding?

This was poorly worded. When we degrade the model resolution, the air mass flux is not conserved from one resolution to another in the regridding. The text now states that the vertical transport is weakened due to "loss of eddy mass flux and air mass flux in the regridding of the meteorological fields."

p23 L28: "Despite having almost flat averaging kernels in the troposphere..." Show a plot of the GOSAT XCH4 averaging kernel – since GOSAT measures it in a thermal IR band, I would have thought that the averaging kernel would be mainly sensitive to the upper troposphere.

As we mentioned above, we are using the CH_4 retrievals based on the shortwave infrared (SWIR) measurements, so the sensitivity is relatively flat throughout the lower and middle troposphere. We have added the Yoshida et al. (2011) reference in Section 2.2.1 that the reader can consult for a description of the SWIR averaging kernels.

Minor Comments: p5 line 13: should this read "and the GHG_CCI group"??

We have corrected the text.

p5 L20: "produced a comparable fit...": to what, the old three-model suite used in the retrievals?

Comparable to the agreement that we obtain using the XCH_4 data with the original CO_2 fields. We now state that "…a posteriori inversion results using the new CO_2 fields and the original fields generally produced comparable fits to independent…"

p5 L24: you should say GOSAT "XCH4" here, since that is what you are looking at; "CH4" does better than that, I think.

We do mean CO_2 fields here; the original LMDZ/MACC-II Co2 fields and the GEOS-Chem a posteriori fields.

P21 L25-26: Reword these lines, I don't think the CH4 budget is conserved because there is bias in the stratosphere induced by transport error.

The referee is correct. A transport bias can change the CH_4 distribution and thus the overall loss of CH_4. We have removed this text.

P23 l8: add a comma before "are".

We added the comma.

In general, I would prefer full descriptions like "the difference between A and B" as opposed to the "bias in A". The term "bias" means a lot of different things, and can relative to another unspecified quantity, and using the full description is always going to be less confusing albeit more wordy.

We have tried to include the full description where possible, without making the text tedious since the main focus of the analysis is on biases.

Characterizing model errors in chemical transport modelling of methane: Using GOSAT XCH4 data with weak constraint four-dimensional variational data assimilation

Ilya Stanevich[1], Dylan B. A. Jones[1], Kimberly Strong[1], Martin Keller[1], Daven K. Henze[2,3], Robert J. Parker[4,5], Hartmut Boesch[4,5], Debra Wunch[1], Justus Notholt[6], Christof Petri[6], Thorsten Warneke[6], Ralf Sussmann[7], Matthias Schneider[8], Frank Hase[8], Rigel Kivi[9], Nicholas M. Deutscher[10], Voltaire A. Velazco[10], Kaley A. Walker[1], and Feng Deng[1]

[1]Department of Physics, University of Toronto, Toronto, Ontario, Canada
[2]Department of Mechanical Engineering, University of Colorado Boulder, Boulder, CO, USA
[3]California Institute of Technology, Pasadena, CA, USA
[4]Earth Observation Science, Department of Physics and Astronomy, University of Leicester, Leicester, UK
[3]National Centre for Earth Observation (NCEO), University of Leicester, Leicester, UK
[6]Institute of Environmental Physics, University of Bremen, Bremen, Germany
[7]Karlsruhe Institute of Technology (KIT), Institute of Meteorology and Climate Research (IMK-IFU), Garmisch-Partenkirchen, Germany
[8]Karlsruhe Institute of Technology (KIT), Institute of Meteorology and Climate Research (IMK-ASF), Karlsruhe, Germany
[9]Finnish Meteorological Institute, Sodankylä, Finland
[10]Centre for Atmospheric Chemistry, School of Chemistry, University of Wollongong, Wollongong, NSW, Australia

Correspondence to: Ilya Stanevich (stanevich@atmosp.physics.utoronto.ca)

Abstract.

We examined biases in the global GEOS-Chem chemical transport model for the period of February-May 2010 using weak constraint (WC) four-dimensional variational (4D-Var) data assimilation and dry-air mole fractions of CH_4 (XCH_4) from the Greenhouse gases Observing SATellite (GOSAT). The ability of the observations and the WC 4D-Var method to mitigate model errors in CH_4 concentrations was first investigated in a set of observing system simulation experiments (OSSEs). We then assimilated the GOSAT XCH_4 retrievals and found that they were capable of ~~differentiating the vertical distribution~~ providing information on the vertical structure of model errors and of removing a significant portion of biases in the modelled CH_4 state. In the WC 4D-Var assimilation, corrections were added to the modeled CH_4 state at each model time step to account for model errors and improve the model fit to the assimilated observations. Compared to the conventional strong constraint (SC) 4D-Var assimilation, the WC method was able to significantly improve the model fit to independent observations. Examination of the WC state corrections suggested that a significant source of the model errors was associated with discrepancies in the model CH_4 in the stratosphere. The WC state corrections also suggested that the model vertical transport in the troposphere at mid- and high-latitudes is too weak. The problem was traced back to biases in the uplift of CH_4 over the source regions in eastern China and North America. In the tropics, the WC assimilation pointed to the possibility of biased CH_4 outflow from the African continent to the Atlantic in the mid-troposphere. The WC assimilation in this region would greatly benefit from glint observations over the ocean to provide additional constraints on the vertical structure of the model errors in the tropics. We

also compared the WC assimilation at the $4° \times 5°$ and $2° \times 2.5°$ horizontal resolutions and found that the WC corrections to mitigate the model errors were significantly larger at $4° \times 5°$ than at $2° \times 2.5°$ resolution, indicating the presence of resolution-dependent model errors. Our results illustrate the potential utility of the WC 4D-Var approach for characterizing model errors. However, a major limitation of this approach is the need to better characterize the specified model error covariance in the assimilation scheme.

1 Introduction

Atmospheric concentrations of methane (CH_4), the second most important anthropogenic greenhouse gas, have been rapidly raising since 1850 (Etheridge et al., 1992). However, atmospheric measurements in recent decades show that the rate of CH_4 increase in the atmosphere has varied and its behaviour is not well understood (Dlugokencky et al., 2009). Significant effort has been put into characterizing surface emissions of CH_4 in order to attribute its recent trends. In this context, a number of satellites have been launched to measure atmospheric CH_4 in order to constrain its sources. These include Envisat carrying the Scanning Imaging Absorption Spectrometer for Atmospheric Cartography (SCIAMACHY) (Schneising et al., 2011), the Greenhouse Gases Observing Satellite (GOSAT) carrying the Thermal And Near-infrared Sensor for carbon Observation Fourier Transport Spectrometer (TANSO-FTS) (Kuze et al., 2009), Sentinel-5p with the Tropospheric Monitoring Instrument (TROPOMI) on-board (Veefkind et al., 2012), and the Greenhouse Gases Satellite (GHGSat). Proposed missions include the Methane Remote Sensing Lidar Mission (MERLIN) (Kiemle et al., 2014), GOSAT-2 (Nakajima et al., 2017), the Geostationary Carbon Cycle Observatory (GeoCARB) (Polonsky et al., 2014) and the recently-announced MethaneSat. However, current regional CH_4 emissions remain largely uncertain (e.g., Saunois et al., 2016). One of the biggest challenges for reducing uncertainty on emission estimates is the relatively weak signal of emissions in the atmospheric column of CH_4, which puts tight requirements on the accuracy of satellite measurements. However, while future satellite instruments and improved spectroscopy are expected to provide better CH_4 measurements, errors in the atmospheric models used to simulate CH_4 remain poorly characterized. While random model errors can be accounted for in flux inversion analyses, the impact of biases in chemistry and transport are often neglected or accounted for using various ad hoc approaches. In the case of CH_4, which is a relatively long-lived gas with an atmospheric lifetime of about 9 years (Prather et al., 2012), chemistry plays a critical role in long-term trends (McNorton et al., 2016), whereas transport, alone or coupled with chemistry, defines how total surface emissions are distributed on a regional scale. Therefore, transport errors, such as those produced by numerical advection schemes, biases and uncertainties of meteorological fields, and parametrization of sub-grid scale processes may significantly undermine our ability to use models to relate emissions to atmospheric observations, and thus our ability to improve CH_4 emission estimates (Prather et al., 2008; Locatelli et al., 2015; Patra et al., 2011).

One potential solution is to apply a bias correction to the model in the context of the inversion analysis. Simple bias correction schemes with uniform or latitudinally dependent bias estimates have been attempted before (Bergamaschi et al., 2009; Fraser et al., 2013; Monteil et al., 2013; Alexe et al., 2015; Locatelli et al., 2015), mostly to correct poor description of the modeled stratosphere. Here we explore the utility of a "weak constraint" (WC) four-dimensional variational (4D-Var) data assimilation

method to characterize forward model errors. In contrast to the traditional "strong constraint" (SC) 4D-Var method, the WC scheme does not assume that the model evolution is perfect. The WC 4D-Var method was introduced by Sasaki (1970) and used in numerical weather prediction (NWP) models by Derber (1989), Zupanski (1997) and Trémolet (2006, 2007). It was first applied by Keller (2014) in the GEOS-Chem simulation of atmospheric carbon monoxide (CO) to characterize model bias.

5   One of the first attempts to apply bias correction in chemical data assimilation was done in the framework of the suboptimal Kalman filter by Lamarque et al. (2004), who used the bias estimation approach of Dee and Da Silva (1998) to constrain the CO state using measurements from the Measurement of Pollution in the Troposphere (MOPITT) instrument. The study pointed to the possibility of errors in the model vertical transport, however most of the estimated biases were attributed to poor a priori estimates of CO surface emissions in the model. The major challenge for this type of analysis for CH_4 is the

10   limited information available about the global vertical distribution of CH_4 in the atmosphere. There are satellite observations that contain information about the CH_4 distribution in the middle and upper troposphere, such as the thermal infrared CH_4 retrievals from the Tropospheric Emission Spectrometer (TES) on-board the NASA Aura satellite (Worden et al., 2012), or in the stratosphere, such as the solar occultation measurements from the Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE-FTS, Bernath et al., 2005) on-board SCISAT. However, the accuracy of these measurements, based on

15   validation studies (for example, De Mazière et al., 2008; Wecht et al., 2012), may not be sufficient to detect model errors. The most accurate satellite measurements are those of the total dry-air mole fraction of CH_4 in the total atmospheric column (XCH_4) obtained by TANSO-FTS on-board GOSAT. However, these measurements provide less vertical information on CH_4 than those from TES or ACE-FTS, although the latter are less sensitive to surface emissions. Highly accurate aircraft or AirCore CH_4 profile measurements would be an ideal source of information, but they are limited in space and time. We explore here the

20   information content of GOSAT CH_4 observations and show that despite being designed to constrain surface emissions, they contain sufficient information to help characterize possible model errors. We assimilate the GOSAT observations using the WC 4D-Var data assimilation approach to estimate biases in GEOS-Chem. This approach is shown to provide a valuable tool for diagnosing and determining the origin of model errors.

This paper is organized as follows. Section 2 gives an overview of the forward model, the observations, and the WC 4D-

25   Var method. It also contains a description of the various sensitivity studies conducted through a series of Observing System Simulation Experiments (OSSEs). In Sect. 3, we present the results of the sensitivity experiments, as well as the results of the assimilation of real GOSAT observations. Sect. 4 provides an interpretation of the pattern of model biases estimated from the GOSAT assimilation. Finally, conclusions are given in Sect. 5.

2   Data and Methods

30   ### 2.1   The GEOS-Chem Model

For all assimilation experiments we use version v35 of the GEOS-Chem adjoint, which is based on version v8-02-01 of the forward model, with updates up to v9-02 (Henze et al., 2007). The GEOS-Chem CTM (www.geos-chem.org) is driven by archived meteorological fields from the Goddard Earth Observing System (GEOS-5.2.0) produced by the NASA Global Modelling and

Assimilation Office (GMAO). The meteorological fields are regridded from their native resolution of $0.5° \times 0.67°$ with 72 vertical levels to $4° \times 5°$ and $2° \times 2.5°$ with 47 vertical levels. The vertical grid spacing in the troposphere varies from about 150 m in the lower part to about 1 km in the upper part. CH_4 is advected using the multi-dimensional flux-form semi-lagrangian (FFSL) scheme by Lin and Rood (1996). Convection is implemented based on the relaxed Arakawa-Schubert scheme (Moorthi and Suarez, 1992). The model uses a simple treatment of turbulent mixing in the boundary layer by instantaneously mixing species from the surface to the top of the planetary boundary layer (PBL). The GEOS-Chem CH_4 sources and sinks used here are described in detail in Wecht et al. (2014). Anthropogenic CH_4 sources include emissions from natural gas and oil extraction, coal mining, livestock, landfills, waste water treatment, rice cultivation, biofuel burning and other minor sources and are based on the 2004 anthropogenic inventory from the Emission Database for Global Atmospheric Research (EDGAR) v4.2 (European Commission Joint Research Centre/Netherlands Environmental Assessment Agency, 2009). Natural CH_4 sources include wetland emissions after Kaplan (2002) and Pickett-Heaps et al. (2011), termite emissions (Fung et al., 1991) and open fire emissions from the daily Global Fire Emissions Database Version 3 (GFED3) (van der Werf et al., 2010; Mu et al., 2011). The CH_4 emissions at $4° \times 5°$ and $2° \times 2.5°$ resolutions are slightly different due to the dependence of wetland emissions on the meteorological fields. Therefore, for consistency of the analysis of model errors, the emissions were regridded from the coarser to the finer resolution. The main loss of CH_4 (about 90% of the total loss) in the atmosphere is due to oxidation by OH, with the remaining 10% sink mainly due to soil absorption and oxidation in the stratosphere. CH_4 chemistry is performed in off-line mode in which changes in CH_4 concentrations do not feed back on other species. Tropospheric OH fields in the model are prescribed as a three-dimensional monthly mean climatology from a tropospheric chemistry simulation in GEOS-Chem v5-03 (Park et al., 2004). Stratospheric CH_4 loss frequencies are from archived climatology of the NASA Global Modelling Initiative (GMI) model (Murray et al., 2012).

The adjoint model is described by Henze et al. (2007) and has been used for assimilation of CH_4 observations by Wecht et al. (2012, 2014), Turner et al. (2015), Bousserez et al. (2016), and Tan et al. (2016). For the analysis presented here, we focus on the period of 1 February 2010 to 31 May 2010. The CH_4 fields were spun up at a resolution of $4° \times 5°$ and $2° \times 2.5°$ for about 5.5 years until July 2009. From July 2009 to January 2010 we assimilated the GOSAT Proxy XCH_4 retrievals (Parker et al., 2015) to obtain monthly mean emission estimates at $4° \times 5°$ resolution. The optimized emissions were then regridded and used to perform forward model simulations at $2° \times 2.5°$ resolution for the same period from July 2009 to January 2010. The updated model fields on 1 February 2010 at both model resolutions were taken as initial condition for the analysis period. As a result, the initial conditions at both resolutions contain similar amounts of CH_4 in the atmosphere. However, CH_4 is distributed differently, reflecting the balance between emissions and transport at each model resolution.

2.2 Measurements

2.2.1 GOSAT

We obtain information about the CH_4 distribution in the atmosphere from XCH_4 retrievals from the TANSO-FTS on-board GOSAT, which has a three-day repeat orbit period. The instrument has a surface footprint of 10.5 km in diameter and records

spectra at about 13:00 local time. We use version 5.2 of the University of Leicester (UoL) GOSAT Proxy XCH_4 data product. The retrieval algorithm is explained in detail in Parker et al. (2011, 2015). In this algorithm, simplified spectral retrievals of XCO_2 and XCH_4 are obtained in spectral bands centred at 1.65 μm and 1.61 μm, respectively. The final total column-averaged dry-air mole fraction of CH_4 is obtained by multiplying the retrieved XCH_4/XCO_2 ratio by modelled XCO_2 fields. This is useful for cancelling out common spectral features caused by light path modifications due to thin clouds, aerosol scattering, and instrumental artefacts in close spectral bands. However, a reliable knowledge of the XCO_2 data is required. The Proxy method provides significantly greater observational coverage, especially in tropical areas, compared to the "full-physics" retrievals. The weakness of the approach is in the fact that the modeled CO_2 fields may still contain biases that are not accounted for in the final XCH_4 product. Version 5.2 of the XCH_4 data does not include retrievals from spectra recorded over oceans (glint observations). This is in contrast to the later versions 6 and 7 which, however, use the same algorithm for XCH_4 retrievals over land. Furthermore, in our analysis we exclude all retrievals over Greenland and poleward of $75°$ (including retrievals over snow).

The original XCH_4 retrievals utilized XCO_2 fields based on the median of three models: GEOS-Chem (from the University of Edinburgh), LMDZ/MACC-II, and CarbonTracker (National Oceanic and Atmospheric Administration (NOAA)), that was smoothed with GOSAT CO_2 averaging kernels (Parker and the GHG-CCI the group, 2016). CO_2 fields in all three models were produced by assimilating in-situ surface CO_2 observations. In this work, we replaced the original modeled CO_2 fields with optimized CO_2 fields from a GEOS-Chem CO_2 surface flux assimilation analysis that used GOSAT XCO_2 retrievals over land (Deng et al., 2014). For the period of interest (February-May 2010), the XCH_4 retrievals using both proxy CO_2 fields are unbiased against each other with a scatter of 3 ppb and a correlation of $R = 0.99$. ~~Figure ?? shows the mean difference between the two products and points to some systematic regional discrepancies of up to 8 , however, in 90% of all 4× 5grid cells covered by the measurements, the differences are smaller than 3 .~~ Sensitivity tests that were conducted showed that a posteriori inversion results using the new CO_2 fields ~~generally produced a comparable fit~~ and the original fields generally produced comparable fits to independent CH_4 measurements from the Total Carbon Column Observing Network (TCCON, Wunch et al., 2011) and from the NOAA Earth System Research Laboratory (ESRL) global cooperative air sampling network (Dlugokencky et al., 2016). The use of the alternative CO_2 fields did not change any of the findings about model errors in our study. ~~There may still be unidentified biases in both retrieval products. However, the fact that both CO_2 fields were obtained using different methodologies gives us confidence in our results.~~

GOSAT CH_4 retrievals contain about 1 degree of freedom for signal (DOFS) and have relatively flat averaging kernels in the troposphere that slowly decrease in the stratosphere (Yoshida et al., 2011). Therefore, they contain little vertical information about the atmosphere at the time of measurement. We use these averaging kernels to smooth the GEOS-Chem CH_4 fields and map them into the measurement space of the GOSAT retrievals using the expression

$$XCH_4 = XCH_4^a + \mathbf{a}^T(\mathbf{z}_{mod} - \mathbf{z}_a) \tag{1}$$

where $\mathbf{z}_{mod}$ is the GEOS-Chem CH_4 profile, $\mathbf{z}_a$ is the GOSAT a priori profile, $\mathbf{a}^T$ is the GOSAT column averaging kernel, and XCH_4^a is the a priori XCH_4 based on $\mathbf{z}_a$. The absence of vertical information in the measurements is a challenge for constraining

the 3D structure of model errors. ~~However, ,~~ but we expect vertical structure to emerge from atmospheric transport patterns. ~~For example, the majority of the CH₄ mass enters the North American domain through the western boundary in the jet stream in the upper troposphere. Therefore, XCH₄ observations over North America would be more sensitive to past CH₄ concentrations in the upper troposphere upwind over the Pacific and Asia.~~

5 Errors in GOSAT Proxy XCH_4 retrievals with the original XCO_2 data were assessed against co-located TCCON ground-based measurements by Hewson et al. (2015). That validation study found that GOSAT retrievals contain random errors of 12.55 ppb and systematic errors of 4.8 ppb (although per-site biases ranged from -2.15 ppb (Wollongong) to 13.44 ppb (Garmisch)). However, errors away from TCCON sites could be larger. Overall, GOSAT and TCCON were highly correlated with a correlation coefficient of 0.86. Buchwitz et al. (2017) obtained similar results with random errors of 11.9 ppb and sys-

10 tematic errors of 5.7 ppb for GOSAT Proxy XCH_4 retrievals against co-located TCCON retrievals. Such precision, combined with spatial and temporal aggregation of the data, could be enough ~~in many regions of the world~~ to improve knowledge about CH_4 a priori surface emissions in regions such as North America, where the XCH_4 enhancements above the background are about 10 ppm (Sheng et al., 2018). However, the presence of potential model errors significantly undermines this assumption. Therefore, here we explore the potential utility of the weak constraint 4D-Var scheme to discern model biases using the XCH_4

15 data.

~~Figure ?? shows the mean XCH₄ fields from February to May 2010 modeled by GEOS-Chem at 4× 5resolution (top panel) and as measured by GOSAT (bottom panel). A number of features in modeled XCH₄ can be identified. There is a clear inter-hemispheric difference with smaller XCH₄ in the Southern hemisphere. Enhanced XCH₄ concentrations can be observed over China, India, equatorial Africa and South America, with weaker signals also present over Europe and the eastern US,~~

20 ~~which, to first approximation, are related to local surface CH₄ emissions. Major features generally agree between the modelled and observed XCH₄ fields, however there are also a number of discrepancies that are discussed in Section 3.~~

2.2.2 Validation Data

The a priori and constrained model CH_4 fields are validated against in situ NOAA-ESRL CH_4 measurements (Dlugokencky et al., 2016) and measurements from the third HIAPER Pole-to-Pole Observations (HIPPO-3) aircraft campaign (Wofsy et al.,

25 2011), against TCCON ground-based XCH_4 retrievals (Wunch et al., 2011), and ACE-FTS space-based CH_4 retrievals (Boone et al., 2005).

The NOAA network operates by collecting air flask samples which are later analysed by gas chromatography with flame ionization detection. At stationary sites, samples are collected once per week. Shipborne samples from sites in the Pacific Ocean and the South China Sea are collected once every three weeks and weekly, respectively, per latitude band. Measurements

30 are reported relative to the NOAA X2004A CH_4 scale. The absolute uncertainty of the scale is 0.2% (about 3 ppb), and measurements are reproducible to within 1-3 ppb.

Airborne data are provided by the HIPPO-3 aircraft campaign which took place between 20 March 2010 and 20 April 2010. The campaign sampled the atmospheric curtain from the North Pole to the coast of Antarctica through the central Pacific Ocean and from the surface to 14 km altitude. We used CH_4 measurements performed by a quantum cascade laser

spectrometer (QCLS) at 1 Hz frequency. QCLS measurements have precision of 0.5 ppb and accuracy of 1 ppb, while the mean bias relative to simultaneous flask-based measurements is 0.44 ppb (Santoni et al., 2014). We exploited the Merged 10-second Meteorology, Atmospheric Chemistry, and Aerosol Data product (Wofsy et al., 2012), which was derived from 1-sec measurements by applying a median filter.

TCCON is a global network of ground-based high-resolution Fourier transform infrared (FTIR) spectrometers retrieving XCH$_4$ from solar absorption spectra in the near-infrared band. We used the GGG2014 version of TCCON XCH$_4$ data from multiple stations around the globe (Kivi and Heikkinen, 2016; Kivi et al., 2017; Blumenstock et al., 2017; Griffith et al., 2017; Hase et al., 2017; Notholt et al., 2017; Sherlock et al., 2017; Sussmann and Rettinger., 2017; Warneke et al., 2017; Wennberg et al., 2017b, a). The estimated accuracy and precision of XCH$_4$ retrievals are less than 0.5% and 0.3%, respectively (Wunch et al., 2015). Retrievals are bias corrected based on comparisons with calibrated aircraft and AirCore profiles.

ACE-FTS on-board SCISAT performs solar occultation measurements over a range of tangent heights. The satellite makes 15 occultations for both sunrise and sunset per day separated by about 24° in longitude. Measurements cover an altitude range from the cloud tops in the upper troposphere up to 150 km. Spectra are recorded continuously during 2-sec scans, which implies that the altitude and tangent point changes slightly during the scan. As a result, the instrument has low horizontal resolution of about 300 km in the limb direction. The vertical resolution determined by the instrument field-of-view is about 3 km at a tangent point 3000 km away from the satellite. However, vertical sampling ranges from 2 to 6 km depending on viewing geometry. In this study, we use the most recent v3.6 CH$_4$ retrievals with geolocation information (Boone et al., 2013; Waymark et al., 2014). Version 3.6 only differs from version 3.5 in that a local computer was used to process v3.5 while a shared supercomputing system was used for v3.6. Olsen et al. (2017) compared ACE-FTS v3.5 and MIPAS CH$_4$ vertical profiles coincident with TANSO-FTS measurements, and found small differences above the tropopause except in the tropics. The mean differences were larger than 20% below about 450 hPa, within 5% between 450 and 40 hPa, and larger than 5% above 40 hPa.

2.3 The Weak Constraint 4D-Var Approach

The estimation of surface emissions of CH$_4$ using the strong constraint 4D-Var scheme is achieved by minimizing the strong constraint cost function

$$J(\mathbf{p}) = \sum_{i=0}^{N} \frac{1}{2}(\mathbf{y}_i - \mathbf{H}_i\mathbf{x}_i)^T \mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{H}_i\mathbf{x}_i) + \frac{1}{2}(\mathbf{p} - \mathbf{p}_a)^T \mathbf{B}^{-1}(\mathbf{p} - \mathbf{p}_a),$$ (2)

where N is the number of one hourly time steps, $\mathbf{y}_i$ is the vector of XCH$_4$ observations during the time step i, $\mathbf{x}_i \in \mathbb{R}^n$ is the model state at time step i that is represented by a 3D field of CH$_4$ concentrations, $\mathbf{p} \in \mathbb{R}^p$ is the vector of surface emissions of CH$_4$, and $\mathbf{p}_a \in \mathbb{R}^p$ is the a priori estimate of the CH$_4$ emissions. Here, $\mathbf{H}$ is the observation operator that maps the modeled CH$_4$ state into the measurement space at the location of the GOSAT XCH$_4$ observations, $\mathbf{R}_i$ represents the observation error covariance matrix, and $\mathbf{B}$ is the a priori error covariance matrix. In minimizing J, we solve for monthly mean emission estimates over the specified assimilation period. The evolution of the model state in Eq. 2 is performed by the GEOS-Chem

~~forward model~~model, which can be represented by an operator M ~~which~~that acts on the model state $\mathbf{x}_i$ and ~~model parameters~~ emissions $\mathbf{p}$ at time step i to produce a new model state $\mathbf{x}_{i+1}$ at the next time step as follows:

$$\mathbf{x}_{i+1} = M(\mathbf{x}_i, \mathbf{p}). \tag{3}$$

~~The model state at each time step is represented by a 3D field of CH$_4$ concentrations, and model parameters usually include~~ 5 ~~CH$_4$ emissions from each model surface grid cell.~~ In Eq. 3, it is assumed that there are no errors in propagating the state forward in time. This is the assumption that is ~~employed in standard 4D-Var, which is also~~ implicit in Eq. 2, and thus the optimization is referred to as "strong constraint" 4D-Var~~because~~; the model trajectory is used as a strong constraint in the optimization. ~~However, as~~

As described by Trémolet (2006), Eq. 3 can be modified to account for model errors by adding corrections $\mathbf{u}_{i+1}$ to the CH$_4$ 10 state at time step $i+1$ so that the model forecast becomes

$$\mathbf{x}_{i+1} = M(\mathbf{x}_i, \mathbf{p}) + \mathbf{G}\mathbf{u}_{i+1}, \tag{4}$$

where $\mathbf{G}$ is an operator that maps corrections ~~$\mathbf{u}_i$~~ $\mathbf{u} \in \mathbb{R}^m$ into the model state. Here~~the corrections $\mathbf{u}_{i+1}$~~, the corrections $\mathbf{u}$ are referred to as forcing terms, which is distinct from the adjoint forcing commonly used in 4D-Var. The operator $\mathbf{G}$ can also be understood as a mask that defines the spatial regions in the 3D model state where corrections need to be applied. Hence, the 15 second term in Eq. 4 represents additional sources and sinks of CH$_4$ in the region of the atmosphere defined by $\mathbf{G}$. ~~The 4D-Var problem to estimate surface emissions transformed into a~~ In the case where $\mathbf{G}$ represents the whole atmosphere, $m = n$ and $\mathbf{u}$ will have the same dimension as $\mathbf{x}$. The sources and sinks could arise from errors in the model transport or chemistry. In minimizing Eq. 2 we solve only for the surface emissions, however, because of Eq. 4 we have the means of solving for the surface emissions as well as the 3D distribution of sources and sinks~~estimation problem in which a cost function~~. In this case, 20 the 4D-Var cost function, which is minimized with respect to both ~~model parameters~~ surface emissions ($\mathbf{p}$) and state corrections ($\mathbf{u}$~~. In this case, the WC 4D-Var cost function~~), is expressed as

$$J(\mathbf{p}, \mathbf{u}_i) = \sum_{i=0}^{N} \frac{1}{2}(\mathbf{y}_i - \mathbf{H}_i\mathbf{x}_i)^T \mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{H}_i\mathbf{x}_i) +$$
$$+ \frac{1}{2}(\mathbf{p} - \mathbf{p}_a)^T \mathbf{B}^{-1}(\mathbf{p} - \mathbf{p}_a) + \sum_{i=1}^{N} \frac{1}{2}\mathbf{u}_i^T \mathbf{Q}_i^{-1}\mathbf{u}_i, \tag{5}$$

where ~~N is the number of one hourly time steps, $\mathbf{y}_i$ is~~ $\mathbf{Q}_i$ defines the a priori model error covariance matrix. This is the weak constraint 4D-Var cost function, which is similar to Eq. 2, except for the addition of the ~~vector of XCH$_4$ observations during the~~ 25 ~~time step i, $\mathbf{H}$ is the observation operator that maps the modelled CH$_4$ state into measurement space at the location of GOSAT XCH$_4$ observations, $\mathbf{R}_i$ represents the observation error covariance matrix, $\mathbf{p}_a$ is the a priori estimates of model parameters, $\mathbf{B}$ is the a priori error covariance matrix, and $\mathbf{Q}_i$ defines the a priori model error covariance matrix~~third term that accounts for the errors in the evolution of the model state. As described by Trémolet (2006), $\mathbf{u}_i$ can be considered to represent model errors on time scales as short as each model time step or as long as the full assimilation period, and is assumed to be constant over

the appropriate interval. In the case where the forcing is estimated over the full assimilation window, the optimized forcing will represent a constant model bias over the whole model trajectory. For the results presented here, $\mathbf{u}_i$ changes in time, but we assume that $\mathbf{Q}$ is constant.

The WC 4D-Var approach was implemented into the GEOS-Chem model by Keller (2014) and here we describe that approach. The cost function (Eq. 5) is minimized subject to the equality constraints in Eq. 4 by adding the model constraints to the cost function to create the following Lagrangian function:

$$
\mathcal{L}(\mathbf{p}, \mathbf{x}_i, \boldsymbol{\lambda}_i, \mathbf{u}_i) = \frac{1}{2}(\mathbf{p} - \mathbf{p}_a)^T \mathbf{B}^{-1}(\mathbf{p} - \mathbf{p}_a) +
$$
$$
\sum_{i=0}^{N} \frac{1}{2}(\mathbf{y}_i - \mathbf{H}_i \mathbf{x}_i)^T \mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{H}_i \mathbf{x}_i) + \sum_{i=1}^{N} \frac{1}{2}\mathbf{u}_i^T \mathbf{Q}_i^{-1} \mathbf{u}_i - \tag{6}
$$
$$
\sum_{i=1}^{N} \boldsymbol{\lambda}_i^T [\mathbf{x}_i - M(\mathbf{x}_{i-1}, \mathbf{p}) - \mathbf{G}\mathbf{u}_i],
$$

where λ_i are the Lagrange multipliers. We define gradients of the Lagrangian $\mathcal{L}$ with respect to $\boldsymbol{\lambda}_i, \mathbf{x}_i, \mathbf{p}$ and $\mathbf{u}_i$ by the following system of equations:

$$
\frac{\partial \mathcal{L}}{\partial \boldsymbol{\lambda}_i} = \mathbf{x}_i - M(\mathbf{x}_{i-1}, \mathbf{p}) - \mathbf{G}\mathbf{u}_i,
$$

$$
\frac{\partial \mathcal{L}}{\partial \mathbf{x}_i} = -\mathbf{H}_i^T \mathbf{R}_i^{-1} \left[\mathbf{y}_i - \mathbf{H}_i \mathbf{x}_i \right] - \boldsymbol{\lambda}_i + \left(\frac{\partial M}{\partial \mathbf{x}_i} \right)^T \boldsymbol{\lambda}_{i+1}, \tag{7}
$$

$$
\frac{\partial \mathcal{L}}{\partial \mathbf{x}_N} = -\mathbf{H}_i^T \mathbf{R}_i^{-1} \left[\mathbf{y}_N - \mathbf{H}_i \mathbf{x}_N \right] - \boldsymbol{\lambda}_N, \tag{8}
$$

$$
\frac{\partial \mathcal{L}}{\partial \mathbf{p}} = \mathbf{B}^{-1}(\mathbf{p} - \mathbf{p}_a) + \sum_{i=1}^{N} \left(\frac{\partial M}{\partial \mathbf{p}}(\mathbf{x}_{i-1}, \mathbf{p}) \right)^T \boldsymbol{\lambda}_i, \tag{9}
$$

$$
\frac{\partial \mathcal{L}}{\partial \mathbf{u}_i} = \mathbf{Q}_i^{-1} \mathbf{u}_i + \mathbf{G}^T \boldsymbol{\lambda}_i, \tag{10}
$$

where $\mathbf{M}^T = (\frac{\partial M}{\partial \mathbf{x}_i})^T$ is the adjoint of the tangent linear model $\mathbf{M}$. At the minimum, the $\mathcal{L}$ gradients are equal to zero. In this case, Eq. ?? transforms into Eq. 4, whereas Eqs. 7-8 give the adjoint model equations

$$
\boldsymbol{\lambda}_N = -\mathbf{H}_i^T \mathbf{R}_i^{-1} \left[\mathbf{y}_N - \mathbf{H}_i \mathbf{x}_N \right],
$$
$$
\boldsymbol{\lambda}_i = \left(\frac{\partial M}{\partial \mathbf{x}_i} \right)^T \boldsymbol{\lambda}_{i+1} - \mathbf{H}_i^T \mathbf{R}_i^{-1} \left[\mathbf{y}_i - \mathbf{H}_i \mathbf{x}_i \right]. \tag{11}
$$

Values of λ_i are derived from the forward and adjoint model integrations and are substituted into Eqs. 9-10. In general, $\frac{\partial \mathcal{L}}{\partial \mathbf{u}_i}$ and $\frac{\partial \mathcal{L}}{\partial \mathbf{p}}$ do not equal zero as the minimum has yet to be reached by iteratively minimizing the Lagrangian function $\mathcal{L}$. In GEOS-Chem this is done using the L-BFGS-B algorithm (Byrd et al., 1995). Finally, the entire optimization algorithm consists of the following steps:

5   1. Run the forward model (Eq. 4) from time t_1 to t_N using the current estimates of $\mathbf{p}$ and $\mathbf{u}_i$.

2. Run the adjoint model and simultaneously accumulate the estimate of λ_i based on Eq. 11.

3. Calculate the gradients of $\mathcal{L}$ with respect to $\mathbf{p}$ and $\mathbf{u}_i$ using Eqs. 9-10 and estimates of λ_i.

4. Update the estimates of $\mathbf{p}$ and $\mathbf{u}_i$ using the L-BFGS-B optimization algorithm based on the descent direction defined by $\frac{\partial \mathcal{L}}{\partial \mathbf{u}_i}$ and $\frac{\partial \mathcal{L}}{\partial \mathbf{p}}$.

10   5. Repeat steps 1-4 until convergence is reached.

Generally, at some point during the convergence process the inversion will start fitting the noise in GOSAT observations. This can be prevented by stopping the iterative algorithm when the reduced chi-squared value for the fitted model approximately equals unity. In practice, the real uncertainty on GOSAT XCH$_4$ retrievals is unknown due to unaccounted errors in the CO$_2$ fields, for example, so we used a different approach. For each WC inversion that was performed, we monitored the evolution

15   of the optimized model fields and compared them to independent observations (from TCCON, the NOAA in situ network, and the HIPPO-3 aircraft campaign). The iterative process was terminated when the fit to independent observations did not improve any further or started to get worse, based on the assumption that after this threshold the optimization began to fit noise in GOSAT observations. On average, the level of noise was estimated to correspond to GOSAT XCH$_4$ uncertainty of about 10 ppb.

20   We utilized the reported uncertainty on the GOSAT XCH$_4$ retrievals (with the median value of approximately 10 ppb) and inflated it to match the GOSAT scatter against TCCON observations (approximately 13 ppb). It was assumed that the observation errors are uncorrelated, so that $\mathbf{R}$ was assumed to be diagonal. The a priori error covariance matrix $\mathbf{B}$ was also assumed to be diagonal, with 50% uncertainty on CH$_4$ emissions in each surface grid box. Emissions were not split into separate categories but optimized as monthly totals in each surface grid box. GOSAT provides global coverage with a period

25   of three days. Therefore, we did not attempt to characterize global pattern of model errors on shorter time scales and explored keeping the forcing terms constant over a time interval that varied from a minimum of three days up to one month. Little is known about the a priori structure of the model errors, so in the design of the cost function, a priori estimates of model errors were set to zero ($\mathbf{u} = 0$ at the beginning of the assimilation).

The WC algorithm optimizes scaling factors (SFs) for both the forcing terms and the model parameters (surface emissions).

30   Emission SFs are ratios of optimized emissions to a priori emissions, while forcing SFs are the ratios of optimized forcing terms to a constant scaling parameter $\tilde{u}$. The WC inverse method becomes sensitive to the choice of scaling parameter when working with multidimensional problems. This choice does not affect the Lagrangian $\mathcal{L}$ (Eq. 6), however, it does change the relative

magnitude of $\mathcal{L}$ gradients with respect to forcing terms $\frac{\partial \mathcal{L}}{\partial \mathbf{u}_i}$ (Eq. 10) and to surface emissions $\frac{\partial \mathcal{L}}{\partial \mathbf{p}}$ (Eq. 9). The state vector of the WC inversion is largely dominated by the number of forcing SFs as opposed to the emission SFs (with a ratio of up to 500:1). Due to the high dimensionality of the problem, the L-BFGS-B optimization algorithm can search only a fraction of parameter space in the direction of the largest gradient descent. Therefore, it becomes sensitive to the relative magnitude of the forcing

5   gradients $\frac{\partial \mathcal{L}}{\partial \mathbf{u}_i}$ versus the emission gradients $\frac{\partial \mathcal{L}}{\partial \mathbf{p}}$. For large values of $\tilde{u}$ (for example, $\tilde{u} > 50$ ppb), the algorithm descends in the direction of the forcing gradient and the WC inversion is transformed into the so-called "full state assimilation". Meanwhile, small values of $\tilde{u}$ (for example, $\tilde{u} < 0.05$ ppb) force the algorithm to minimize the cost function in the direction of emission gradients ("flux assimilation"). The value of $\tilde{u} = 1.0$ ppb was empirically chosen to perform simultaneous optimization of the emissions and forcing terms ("flux+state assimilation").

10   Application of the WC 4D-Var method is sensitive to the specification of the covariance matrix $\mathbf{Q}$, which is difficult to characterize (Trémolet, 2007). We adopted a diagonal structure of matrix $\mathbf{Q}$ as our standard option. This implies there was no explicit temporal or spatial correlation assumed between model errors. However, some correlation is implicitly present in the model and emerges from both atmospheric transport patterns and the definition of the constant forcing time window. Still, assigning adequate model error uncertainty is one of the major challenges for using the WC method. Generally, there is no

15   single recipe for that, as model errors come from a variety of sources, with different characteristics and, moreover, vary on daily to seasonal time scales. Additionally, in practice, there is usually no way to properly validate whether the inversion correctly attributed biases in CH_4 fields as being caused by surface emissions, model errors, or observational biases. This later statement is related to the fact that surface emission, observational bias, and some model errors may leave similar signatures in the CH_4 fields that would not be easy to distinguish even with perfect observational coverage. The situation may even be worse for CH_4

20   biases if incorrect emissions and model errors mask each other and do not show up in the model comparison with the GOSAT data.

Given these issues, our focus here is not on estimating surface emissions of CH_4. Instead, we use the WC 4D-Var method to optimally constrain the 3D corrections to the CH_4 state and explore the ~~nature~~ structure of the errors in the model~~CH_4 simulation~~. We performed two types of inversions, as described in Sect. 2.3: "full state assimilation", in which we estimate

25   only the 3D corrections ($\mathbf{u}$) to the model state, and "flux+state assimilation", in which we estimate the surface emissions ($\mathbf{p}$) and the 3D corrections. Given that little is known about the distribution of model errors in CH_4 in the troposphere, in both cases we chose a uniform spatial and temporal structure of model error uncertainty q so that the model error covariance is defined as $\mathbf{Q} = q^2 \mathbf{I}$.

We conducted a series of parameter tuning experiments where the WC 4D-Var analysis was performed using values of q

30   ranging from 0.05 ppb to about 2000 ppb, and optimized CH_4 fields were validated against independent observations. The experiments showed that for values of q above 50 ppb, the fit of optimized CH_4 fields to independent observations did not change noticeably. However, for values of q below 50 ppb, the fit deteriorated as q became smaller. Therefore, q was set to 50 ppb. It is important to note that the magnitude of estimated forcing terms changes with changing q, but the general pattern of positive and negative corrections was not significantly affected by the choice of q. In the experiments described in Sect. 2.4, we

35   found that the WC method was still able to significantly improve the model and capture the bias in the CH_4 state. Therefore, we

considered a uniform structure for $\mathbf{Q}$ to be a satisfactory assumption for this initial assessment of model errors in the context of the WC 4D-Var analysis.

2.4 Configuration of ~~Sensitivity~~ the OSSE Experiments

~~To evaluate the performance of the WC 4D-Var scheme, we conducted a series of OSSEs. We began with an analysis of the sensitivity of the GOSAT observational coverage to the CH$_4$ state in the model. Following Liu et al. (2015) and Byrne et al. (2017), we constructed the sensitivity function~~

$$J = \sum_{n=1}^{N} (\mathrm{XCH}_4^{model})_n,$$

~~where XCH_4^{model} are modelled CH$_4$ dry-air mole fractions sampled at the times and locations of the GOSAT observations and convolved with GOSAT scene-dependent averaging kernels. The sensitivity function is summed over all N observations available over the assimilation period. The sensitivity of the GOSAT observations to the modelled state is obtained by taking the derivative of J with respect to the state, using the model adjoint~~

$$\frac{\partial J}{\partial x_{i,j,k,t}} = \sum_{n=1}^{N} \frac{\partial (\mathrm{XCH}_4^{model})_n}{\partial x_{i,j,k,t}},$$

~~where $x_{i,j,k,t}$ is the CH$_4$ at longitude i, latitude j, altitude k, and time t. The sensitivities are expressed in units of and can be understood as the propagation backward in time of the GOSAT averaging kernels weighted by the value proportional to the local surface pressure. The sensitivities can be summed in space or time to give an aggregated view of the sensitivity of the GOSAT observational coverage to the modelled CH$_4$ state.~~

~~We conducted four~~ We conducted three OSSEs in order to evaluate the performance of the WC 4D-Var method in regards to mitigating artificially introduced model errors for February-May 2010. In particular, we investigated model biases ~~from four different sources: surface emissions,~~ due to vertical transport, chemical loss, and initial conditions. The "true" model state was defined as optimized CH$_4$ global fields obtained from an inversion analysis to constrain estimates of monthly CH$_4$ fluxes using GOSAT XCH$_4$ Proxy retrievals during the same time period. We also refer to these constrained fluxes as "true" CH$_4$ surface emissions. The CH$_4$ initial conditions are described in Sect. 2.1. This "true" model state was used to produce pseudo GOSAT XCH$_4$ measurements by sampling it at the corresponding times and locations of the real GOSAT measurements and then convolving them with GOSAT averaging kernels. The perturbed model was defined by introducing bias in the "true" model from one of the four specified sources of model bias. Then the pseudo-observations were used to constrain and mitigate biases in the perturbed model CH$_4$ state. It should be noted that these are "perfect model" experiments since we are using GEOS-Chem to simulate the pseudo data as well as for the inversions. The performance of the pseudo-inversion was evaluated by comparing the recovered CH$_4$ fields to the "true" ones. The analyses were conducted for the standard period of four months (February-May 2010), but most of the results are presented for the second month of the assimilation period, March 2010. This gives the model errors time to accumulate during February, and provides two months of pseudo-data, in April and May,

to constraint the CH_4 state in March. No noise was added to pseudo-observations. Given that and the fact that, usually, the state is most optimally constrained in the middle of the assimilation period, we believe that the OSSEs should reveal the best performance of the WC method.

The ~~emission bias was introduced by replacing the "true" CH_4 emissions with the original a priori emissions. Convection~~ bias in vertical transport and chemistry were ~~artificially biased by completely turning them off~~ introduced by turning off convection and chemistry, respectively, in the model for the duration of the assimilation period. ~~Finally~~In the third OSSE, a bias in initial conditions was introduced by replacing the "true" initial conditions with the ones obtained by running the forward model without convection and with 70% of the a priori emissions from 1 July 2009 to 1 February 2010, the beginning of the assimilation period. The applied biases for these ~~four~~ three OSSEs were intentionally designed to be extreme; for the real world applications, we expect less extreme model errors.

We configured the WC method to carry out "full state assimilation" (as described in Sect. 2.3) and have the freedom to determine independently the location of the bias. The constant forcing time window was set equal to three days and the forcing terms were optimized throughout the entire atmosphere (the mask $\mathbf{G}$ equals unity everywhere). This particular configuration may not be optimal to mitigate a specific type of bias in a real assimilation with limited observational coverage. Here, we intend to investigate the performance of the measurements and the assimilation method when no information is given about the sources and magnitude of model errors. We also conducted SC 4D-Var assimilation experiment for comparisons with the WC approach in the OSSE with biased surface emissions.

2.5 Configuration of the Assimilation with Real GOSAT Data

For the assimilation of the real GOSAT CH_4 data, ~~we used the same WC 4D-Var configuration as in the OSSEs, but with some modifications.~~ the model error corrections to the CH_4 state were constrained during the standard four-month period of February-May 2010. The CH_4 initial conditions are as described in Sect. 2.1. ~~The CH_4 state was constrained during the standard four-month period of February-May 2010 using GOSAT observations during the same period. Modifications~~ We conducted four sets of experiments, which are described below, to assess the sensitivity of the results to the WC 4D-Var configuration~~included tuning the length of the constant forcing time window *T* and the horizontal and vertical structure of the forcing mask $\mathbf{G}$. We also performed additional tuning of the method in order to explore the nature of the model errors. The quality of the constrained CH_4 fields was evaluated against independent observations.~~. Additionally, we compared results of the WC inversions with results of the SC surface flux assimilation.

The a priori model validation presented in Sect. 3.2.2, as well as the results of Saad et al. (2016), pointed to the fact that the stratosphere in GEOS-Chem at the $4° \times 5°$ resolution, particularly, at high latitudes, may be positively biased. The OSSE results also suggested that the WC assimilation may benefit from additional constraints on stratospheric forcing terms. Therefore, for the assimilation of the real GOSAT data we imposed a negativity bound in the L-BFGS-B algorithm for the optimization of the forcing terms in the extra-tropical stratosphere (above about 210 hPa and poleward of 44°) to remove the known bias at $4° \times 5°$ resolution. No bound was imposed on forcing terms in the $2° \times 2.5°$ resolution assimilation.

In ~~one~~ the first set of experiments, we performed "full state assimilation" and changed the length of the time window over which the forcing terms are held constant in the assimilation. ~~The~~ In these experiments, the forcing mask **G** comprised the entire atmosphere, and biases in the CH_4 state potentially induced by incorrect surface emissions were treated as just another source of model errors included in forcing terms. The length for the forcing window was varied from three to 30 days. Short time windows would be more appropriate if the model were affected by temporally changing biases such as those related to transient mesoscale eddies. However, the observations may not be able to constrain the short time scales. Also, for short temporal correlation length scales, there is a higher risk that the inversion will fit noise or possible biases in observations. In contrast, the use of long time windows introduces additional temporal correlations between forcing terms that may be suitable only for mitigation of stationary systematic biases in the model, such as those related to surface emissions, chemistry or stationary transport errors.

In ~~a different~~ the second set of experiments, we carried out WC 4D-Var "source+state assimilation' and explored the ~~optimal design of the~~ sensitivity of the results to vertical extent of the forcing mask **G**. Here, the forcing window was set equal to three days. ~~First, we explored the vertical structure of G.~~ The algorithm was configured to optimize forcing terms in 1) the whole atmosphere, 2) above 750 hPa, 3) above 500 hPa, and 4) above 200 hPa. Then ~~the horizontal structure~~ in the third set of experiments, the horizontal extent of **G** was modified~~as well~~. Forcing terms were applied 1) globally throughout the stratosphere and 2) in the troposphere only over ~~four separate~~ the following four regions: the three regions defined by the boundaries of the GEOS-Chem nested model domains (North America (NA), Europe (EU), and China with South-East Asia (CH)) and over Equatorial Africa (EQAf). In these experiments we also attempted to identify the origin of the biases affecting the model at the location of the TCCON and NOAA measurement sites.

All the above experiments were conducted at the $4° \times 5°$ model resolution. In ~~one additional~~ the fourth experiment, we applied the WC 4D-Var "full state assimilation" to constrain errors in GEOS-Chem at $2° \times 2.5°$ resolution. We used the standard configuration with a forcing time window of three days. The only difference between the $4° \times 5°$ and the $2° \times 2.5°$ assimilation was in the initial conditions, which are described in Sect. 2.1.

3 Results

3.1 OSSE Experiments

~~We began with the analysis of the sensitivity of the GOSAT observation coverage to the CH_4 state in the model (see Eq. ??). Shown in Fig. ??, as a function of altitude and latitude, is the total zonal adjoint sensitivity (upper panel), which is a sum of adjoint sensitivities over time and longitude. Additionally, we included a vertical slice of the total adjoint sensitivity across 34N latitude, which is a sum of adjoint sensitivities over time (lower panel). As suggested by the upper panel in Fig. ??, the entire GOSAT observing system in February to May 2010 has the greatest sensitivity to CH_4 changes in the upper troposphere and the lower stratosphere (UTLS) in the northern hemisphere. This can be explained by the fact that winds in the UTLS region are stronger than in the lower troposphere, hence any change in the CH_4 fields in the former region will eventually affect a larger number of GOSAT measurement locations in the model. The sensitivity in the tropics is approximately half that in~~

mid-latitudes. The lower panel in Fig. **??** also shows that over the oceans (which are not covered by the GOSAT observations used in this study), such as over the Pacific Ocean, between 120E and 130W), the sensitivity is reduced near the surface and is increased in the UTLS where most of CH_4 mass flux from Asia enters the North American domain. Increased sensitivity between approximately 30W and 30E is due to the large number of GOSAT measurements over the Sahara desert.

In the first OSSE, we tested the ability of both the SC and WC 4D-Var schemes to reproduce mean CH_4 atmospheric concentrations for the case of biased emissions. It is expected that the SC 4D-Var method will produce better results than the WC 4D-Var due to the fact that, when using the SC method, we implicitly supply the assimilation with knowledge about the source of the bias. The results of the OSSE are presented in Fig. **??**, which shows the mean difference between the recovered state and the "true" CH_4 fields. The results confirm that the SC 4D-Var method better removes CH_4 biases due to emissions.

As shown in Fig. **??**, the major challenge for the WC 4D-Var was to constrain CH_4 fields in the boundary layer below about 800 above large emission sources. Here the method failed to properly correct the vertical structure of the model biases. Due to weak vertical sensitivity of the pseudo-data, it is difficult for the WC 4D-Var method to mitigate strong localized vertical bias. Instead, it compensates for the bias by applying relatively weak CH_4 state adjustment of the opposite sign in the column of the atmosphere above, particularly in the stratosphere (see Fig. **??**).

In order to improve the WC 4D-Var performance, additional information is required about the location (which can be specified using forcing mask **G**) and properties of the model errors. For example, the equivalent of perfect temporal correlation can be accounted for using a constant forcing time window (one month instead of three days for emissions). Improvement in the performance can also be achieved by providing additional information on the structure of the model errors, such as by assigning a forcing error covariance matrix **Q** with non-uniform vertical error structure, exhibiting larger uncertainty in the boundary layer and smaller errors in the free troposphere. A simple example of the gain in the performance by using a 30-day forcing time window is shown in Fig. **??**. In another example, shown in the same figure, we changed the forcing mask **G** so that model errors are estimated only in the troposphere (approximately, from the surface to 200). However, this is a rather loose constraint on the potential location of model errors and did not result in a significant improvement in CH_4 state in the troposphere beyond the previous experiments.

In the ~~second~~ first OSSE, we investigated the ability of the WC 4D-Var method to mitigate errors in vertical transport by turning off convection in the model. ~~This~~ The spatial patterns of the estimated model corrections are shown in Fig. 1. As can be seen, the assimilation resulted in enhanced CH_4 concentrations in the lower troposphere and reduced CH_4 in the upper troposphere over the main source regions~~as seen in Fig. 1.~~. Furthermore, the positive CH_4 anomalies in the lower troposphere were partly advected downstream. For example, over Equatorial Africa and South America, instead of being convectively lofted over the continent, CH_4 emissions were transported westward in the lower and middle troposphere (see Fig. 1, first column, third row). As shown in the figure, the state corrections capture the general horizontal and vertical structure of the a priori bias. The largest corrections are co-located with the regions of deep convection. Positive corrections are found in the upper troposphere and negative corrections in the lower. Still, this was not sufficient to fully mitigate the extreme bias associated with turning off convection, but the results show that GOSAT retrievals ~~possess sensitivity to biases in vertical transport and can~~

distinguish them contain information to enable us to capture vertical transport bias even when the sources and magnitude of model errors are unknown.

Figure 2 shows the mean vertical distribution of the a priori and a posteriori residual biases in the CH_4 state over equatorial South America, equatorial Africa, equatorial Southeast Asia, and Europe. In mid-latitudes over Europe, the convection bias was much weaker than over the tropics and reached just about 16 ppb near the surface. At altitudes above 600 hPa the WC 4D-Var method was able to strongly mitigate this bias, and below 800 hPa it reduced the bias by more than a factor of two. The worst results in terms of the fractional reduction of the bias were achieved over equatorial Southeast Asia, most likely due to fewer GOSAT retrievals over this region and limited constraints on the CH_4 distribution in the outflow region over the ocean. The assimilation also removed a large fraction of the bias in the CH_4 fields over Equatorial Africa and South America, particularly in the middle and upper troposphere over Africa and in the lower troposphere over South America.

In the third OSSE, The second OSSE, in which a chemistry bias was created by turning off the reaction of CH_4 with OH. This, was the least challenging bias for the WC 4D-Var scheme to mitigate. This bias was rather smooth in the troposphere and did not contain small-scale features. Although the actual chemistry bias in the model may have more complex vertical structure, we do not expect chemical biases to be as strongly localized as the biases associated with emissions and vertical transport. The a priori and a posteriori residual biases, as well as WC forcing terms, are shown in the Fig. 3. The WC state optimization performed best over land where the a priori biases were almost completely removed. The optimization was least successful over the oceans in the lower troposphere. This situation is consistent with the distribution of the adjoint sensitivities shown in Fig. ??, which showed lower GOSAT fact that the assimilation of GOSAT data has lower sensitivity to variations in CH_4 in the lower troposphere as compared to the upper troposphere, due in part to the absence of GOSAT observations over oceans in our analysis as well as to different transport patterns and stronger winds in the upper troposphere. Shown in Fig. 4 are the mean vertical profiles of the prior and posterior bias over the same four regions considered in Fig. 2. The model does indeed successfully mitigate the bias. Over the convection regions in the tropics, there is some compensatory corrections in the lower troposphere and in the UTLS, which is probably due to the fast vertical transport in these regions and the limited vertical information in the GOSAT retrievals.

In the final OSSEwe biased the initial conditionsby introducing biases in the vertical distribution of CH_4 (by turning off convection) and in the total CH_4 mass in the atmosphere (by running the model with different surface emissions in the previous seven months). The third OSSE, with biased initial conditions, the initial condition bias is shown on the left panel in Fig. 5. The stratosphere and southern troposphere were positively biased, whereas the northern troposphere was negatively biased. The right panel shows the structure of the a posteriori bias after the WC assimilation, on the last day of the assimilation window, May 31st. It shows that the CH_4 state converged to the "true" concentrations everywhere except in the upper stratosphere; the positive upper stratospheric bias was compensated for in the column by a small negative CH_4 bias in the troposphere and the lower stratosphere.

In Fig. ??, we show Examination of the evolution of the initial condition bias , relative to the total atmospheric CH_4 mass, in four altitude bins in each hemisphere: 1000-700 (not shown), 700-400 , 400-200 , and 200-0 . What the figure does not show is how much the bias was adjusted in the actual initial conditions. The perfect observing system would completely remove

16

the initial condition bias at the start of the assimilation period (on February 1). However, what is shown on February 1 is just an 8% reduction in the bias in each of the eight regions , relative to the a priori, with the rest of the bias propagated onto the assimilation period. The different regions indicates that different regions of the atmosphere converged to the "true" CH$_4$ mass at different rates. The tropospheric CH$_4$ burden in both hemispheres (in the 1000-700 , 700-400 , and 400-200 bins) converged mainly during the first month, however, convergence was slower near the surface in the SH. Above , with levels above 200 , the convergence rate was slow hPa converging the slowest, such that by the third month the CH$_4$ mass had not fully recovered , particularly in the SH where there is reduced sensitivity due to the limited GOSAT observational coverage. The slower convergence above 200 hPa (compared to the troposphere) is expected due to weak vertical transport in the stratosphere. This suggests that additional vertical correlation between forcing terms in the stratosphere would be beneficial to accelerate convergence in the stratosphere. at these levels.

3.2 Assimilation of Real GOSAT Retrievals

The bias between the GOSAT data and the $4° \times 5°$ a priori and a posteriori model is shown in Fig. 6. Here we will refer to the a posteriori results as the **WC_4x5** assimilation, which is our standard WC 4D-Var assimilation at $4° \times 5°$ resolution with a three-day forcing time window and a forcing mask **G** comprising the entire vertical extent of the atmosphere. As can be seen, there are large positive a priori biases at high latitudes in the northern hemisphere and in some low-latitude regions, such as Equatorial Africa and eastern China. The **WC_4x5** assimilation successfully reduces the a priori bias. There is some residual high latitude bias, which resembles noise or bias in the GOSAT observations. In a companion analysis, Stanevich et al. (2020), in which we examine the impact of model resolution on the modelled CH$_4$ distribution, we showed that the large positive a priori CH$_4$ bias over China may partly be explained by weakening of the vertical transport in the model due to the coarse $4°$ $\times 5°$ resolution. In Stanevich et al. (2020), we also showed that a significant fraction of the high-latitude bias comes from the stratosphere and is a consequence of running the model at $4° \times 5°$ resolution. As a result, here we repeated the GOSAT WC assimilation at the higher resolution of $2° \times 2.5°$. The results, which are shown in Fig. 7, reveal that the high latitude a priori bias is indeed smaller in the $2° \times 2.5°$ model. At the higher resolution, the WC assimilation also successfully reduces the model bias. For comparison, we repeated the assimilation at $4° \times 5°$, but optimized the emissions instead of the CH$_4$ state. The results for this experiment, referred to as **SC_4x5**, are shown in Fig. 8. As can be seen, the SC assimilation leaves significantly larger residual biases. The pattern of the residual bias indicates that there were other biases that the assimilation could not fit at the expense of the emissions. We will investigate possible sources for these biases in the sections below.

The signal of surface emissions is mixed with possible model errors in the troposphere, such as those related to vertical transport. Biases in the CH$_4$ fields caused by incorrect surface emissions will in some cases have identical structure to those caused by biased vertical transport, which may complicate the interpretation of WC 4D-Var state corrections in the troposphere. On the other hand, it takes much longer for the surface emissions signal to mix into the stratosphere. We therefore assumed that, on the short (four-month) time scale of the simulation, optimized forcing corrections $\mathbf{u}_i$ in the stratosphere can be considered independent from the influence of surface emissions. The third column in Figs. 6 and 7 shows the actual mean monthly bias in the a priori CH$_4$ fields that was corrected by the stratospheric forcing terms. The bias corrections in the $2° \times 2.5°$ CH$_4$

simulation are smaller than for the $4° \times 5°$ simulation, which is consistent with Stanevich et al. (2020), who suggested that part of the stratospheric bias at $4° \times 5°$ resolution is due to the model resolution itself. The WC inversion results suggest that the $4° \times 5°$ model is positively biased in the stratosphere at the high latitudes and weakly negatively biased in the tropics. In contrast, the $2° \times 2.5°$ model is mainly negatively biased in the stratosphere, particularly, around 30-40°N, except for few high latitude regions, possibly related to the polar vortex.

3.2.1 Evaluation With TCCON and NOAA Data

Table 1 presents the results of the evaluation of the **SC_4x5** and the **WC_4x5** assimilation with the in situ and TCCON data, whereas Table 2 gives the comparison results at individual TCCON sites. Based on the OSSE results in Sect. 3.1, and provided that the only model bias is due to incorrect surface emissions, we would anticipate the WC assimilation to produce generally worse fits to the surface measurements than the SC assimilation. The comparisons show that both approaches produced similar improvements in the fit to the NOAA in situ observations, with slightly better performance from the WC method. The WC assimilation had a significant impact on the overall fit to the TCCON XCH$_4$ retrievals, whereas the SC assimilation had a much more limited impact. Table 2 shows the benefits of using the WC method at the individual TCCON sites. With the exception of Park Falls and Lamont, the WC assimilation significantly improved the correlation and reduced the bias between the model and the TCCON observations. The results suggest that GEOS-Chem a priori CH$_4$ simulation suffered from biases that were not related only to incorrect surface emissions.

The evaluation of the WC ~~tuning~~ sensitivity experiments is summarized in Fig. 9. The series of WC experiments described in Sect. 2.5 were organized into four groups. The most sensitive indicator of the quality of the model-observations fit is the correlation. The scatter was close to the level of the GOSAT measurement noise and did not change much among the different assimilation experiments. In the ~~first group~~ set of experiments (first panel in Fig. 9) ~~.~~, in which we changed the vertical extent of the forcing mask **G**. ~~We~~, we found that restricting the optimized forcing to the stratosphere (altitudes above 200 hPa) resulted in correlation statistics that were only slightly worse than when we optimized the forcing throughout the whole atmosphere. This suggests that a significant part of model errors above all TCCON stations may be related to the representation of the stratosphere in the model. In addition, the bias and scatter plots show that optimization of forcing terms above 200 hPa produced the best fit to NOAA surface observations. In the ~~second group of~~ experiments (second panel in Fig. 9) ~~.~~, in which we modified the horizontal extent of the forcing mask **G**. ~~We~~, we found that optimization of the forcing throughout the stratosphere and only over North America, Europe, China, and Equatorial Africa in the troposphere, as described in Sect. 2.5, produced almost identical fits to the case of the "full state assimilation". These four regions are major sources of CH$_4$ and our results suggest that at the TCCON sites the model was likely affected by errors in emissions and the transport of the emission signal over these regions. Henceforth, we refer to these assimilation results as **WC_4REG_4x5**. In the ~~third group of~~ experiments (see the third panel in Fig. 9) ~~.~~, in which we varied the length of the forcing window from 3 days to 7 days, 14 days, and 30 days. ~~We~~, we found that the agreement at some of the stations, such as Lamont, Park Falls, and Sodankylä, were generally insensitive to increasing the length of the forcing window, which could suggest that the model above these stations was affected by slowly varying biases. The model fit at other stations, particularly, Bialystok, Bremen and Karlshure, degraded when the

18

window length was increased. The three later stations are located close to each other and are, probably, affected by the similar model errors on synoptic times scales of about one week.

In the last group of experiments (see the fourth panel in Fig. 9), we compared the performance of the two 4D-Var assimilation modelling approaches (WC "full state assimilation" and SC "flux assimilation") at the two model resolutions, ($4° \times 5°$ and $2° \times 2.5°$). The comparison suggested that, in the absence of a priori bias correction, the SC method brings limited improvements to the a prior CH_4 fields at both resolutions. Indeed, we conclude that the SC assimilation at the $4° \times 5°$ resolution is futile as the a priori model at $2° \times 2.5°$ resolution produces a better fit to the TCCON observations than the SC $4° \times 5°$ assimilation. The performance of the SC assimilation at the $2° \times 2.5°$ resolution was similar to but was surpassed by the "best fit" WC state assimilation at the $4° \times 5°$ resolution in term of its fit to TCCON and NOAA in situ measurements. Overall, the WC state assimilation at $2° \times 2.5°$ resolution generated the best model fit to TCCON observations. However, in all $2° \times 2.5°$ resolution experiments the model bias against NOAA surface measurements was larger compared to the $4° \times 5°$ experiments. For example, the smallest WC a posteriori bias at $4° \times 5°$ was about 10 ppb, whereas at $2° \times 2.5°$ it was about 17 ppb.

Another important conclusion can be drawn from the fact that the WC assimilation at both model resolutions significantly improved the model fit to Izana measurements (see the fourth panel in Fig. 9). The Izana station is located at an altitude of 2370 m above sea level on a small island near the coast of Africa that has no local CH_4 emission sources. The model at $2° \times 2.5°$ and $4° \times 5°$ resolutions is not able to resolve the ~~inland~~topography of the island. Therefore, the model transport in the vicinity of ~~a~~ this high-altitude station, particularly, in the lower troposphere, may be subject to similar errors at both the $2° \times 2.5°$ and $4° \times 5°$ resolutions. Hence, the improvement in the assimilated CH_4 fields ~~should mainly~~ may be related to the corrected ~~transport~~ model errors in the upper troposphere and the stratosphere, ~~which also supports the conclusions drawn from the first group of experiments~~rather than in the lower troposphere where topography-related errors would be dominant.

The WC full state assimilation at $4° \times 5°$ leaves a weak positive biases in the GEOS-Chem fields against the TCCON observations (excluding Sodankylä) in most of the experiments. Mean a posterior inter-station bias at $4° \times 5°$ ($2° \times 2.5°$) resolution is 3.4 (4.0) ppb (excluding Sodankylä), while the scatter is 8.6 (7.3) ppb (~~including~~ excluding Sodankylä). It is not clear if the GOSAT data is positively biased or if this could be caused by differences between the GOSAT and TCCON averaging kernels in the stratosphere and the fact that, for example, the stratospheric model bias was not fully recovered by the assimilation, particularly, during the first couple of months of the assimilation period~~(see Fig. ??).~~. The results also do not indicate the presence of a latitudinal bias between TCCON and GEOS-Chem and, hence, between TCCON and GOSAT.

There is a larger positive XCH_4 bias between the model and Sodankylä measurements, 12.6 ppb and 11.2 ppb for the WC assimilation at $4° \times 5°$ and $2° \times 2.5°$ resolution, respectively, however, the correlation is also high, 0.81 and 0.93, respectively. Tukiainen et al. (2016) and Ostler et al. (2014) pointed to the fact that polar vortex conditions at high-latitude stations may induce biases in TCCON XCH_4 retrievals. It has been claimed that a priori profiles in the retrievals do not account for and are not adjusted to these dynamic conditions, hence, they significantly deviate from the real CH_4 profiles. When there is not enough information in the spectra to correct for such discrepancies, the XCH_4 retrievals can be systematically biased. It is possible that both the GOSAT and TCCON could have been affected by the polar vortex conditions during some days in February-April 2010 so that the biases in co-located retrievals are partially cancelled. It should also be noted that the negative a priori correlation

19

between the model and Bialystok XCH_4 measurements is partly caused by the limited number (84) of measurements during the four-month assimilation time window.

3.2.2 Evaluation With ACE-FTS and HIPPO-3 Data

Figures 10 and 11 show the results of the GEOS-Chem comparison with the ACE-FTS and HIPPO-3 data. Model versus ACE-FTS data is shown only in the stratosphere in order to exclude potentially biased data due to interference with clouds in the upper troposphere. The mean XCH_4 difference between GEOS-Chem and ACE-FTS that is shown was obtained by artificially extending the ACE-FTS CH_4 profiles down into the troposphere using the GEOS-Chem fields and then applying the GOSAT column averaging kernels. Consistent with ?Saad et al. (2016), the CH_4 differences reveal that the a priori $4° \times 5°$ model has a positive stratospheric bias that can be as large as 250 ppb averaged zonally (see Fig. 10). HIPPO-3 comparison also showed that the $4° \times 5°$ model is positively biased in the stratosphere and slightly negative in the troposphere. Wang et al. (2017) showed that similar positive CH_4 biases in mid- and high latitudes exist in TM3, TM5 and LMDz CTMs. The $4° \times 5°$ WC assimilation reduced the positive stratospheric bias with respect to both HIPPO-3 and ACE-FTS, but it did not remove it completely. For example, the maximum model minus ACE-FTS XCH_4 bias due to the stratosphere was reduced from about 40 ppb to 30 ppb. The average negative tropospheric CH_4 bias relative to HIPPO-3 was reduced. It is possible that the WC method was not able to properly localize the stratospheric bias. However, the validation analysis may also reflect the influence of the slow recovery of the stratospheric CH_4 fields from the bias in the initial conditions. Therefore, discrepancies in the stratospheric CH_4 field from the initial conditions in the first two months of the WC assimilation could be contributing to the observed HIPPO-3 and ACE-FTS bias. Unfortunately, the measurements are either too sparse or limited in space and time to verify this assumption.

The positive a priori stratospheric bias relative to ACE-FTS and HIPPO-3 was significantly smaller at $2° \times 2.5°$ than at the $4° \times 5°$ resolution (see Fig. 11), however, it was not completely removed. Stratospheric CH_4 fields in the NH above 200 hPa even became negatively biased at $2° \times 2.5°$, particularly around 30°N–40°N, where the absolute bias became larger than at $4° \times 5°$. The WC assimilation at $2° \times 2.5°$ further corrected the positive biases and significantly reduced the negative bias around 30°N–40°N. As can be inferred from Fig. 7, the latter covered the entire latitudinal band but was particularly pronounced over the Himalayas. Despite the reduction of the stratospheric bias, the $2° \times 2.5°$ WC assimilation introduces a positive CH_4 bias relative to HIPPO-3 in the NH lower troposphere.

4 Discussion of Model Biases

4.1 Stratospheric Bias

The sensitivity experiments carried out in Sect. 2.4 suggested that a stratospheric bias introduced in the system through the initial conditions has the slowest correction rate. However, by the start of the last month of the assimilation, May 2010, the bias is either removed or does not change much with time. Therefore, we focus the discussion here on the stratosphere in the month of May 2010, with the assumption that the model is free of the influence of the initial conditions. Figure 12 compares the a

priori CH_4 fields to the optimized fields from the **WC_4x5** and **SC_4x5** assimilations. The top panel shows that corrections in the stratospheric CH_4 abundance are the most pronounced feature of the WC optimized CH_4 fields, and that changes are smaller in the zonal mean tropospheric fields. The bottom panel is presented to contrast the behaviour of the two 4D-Var approaches. It shows that the SC assimilation attempts to correct the positive high-latitude stratospheric CH_4 bias at the expense of surface emissions. This results in a negative CH_4 bias in the lower troposphere, while the surface signal hardly impacts the stratosphere. In the WC assimilation, stratospheric CH_4 was significantly reduced at high latitudes and increased in the tropics relative to the a priori, which is consistent with the correction of the biases shown in Fig. 10 and 11. The changes are more substantial in the northern hemisphere due to the asymmetrically larger number of GOSAT measurements in the northern hemisphere(see the adjoint sensitivity in Fig. ??)..

Large biases in the stratosphere were previously identified in GEOS-Chem (?) (Saad et al., 2016) and in other chemistry transport models (Strahan and Polansky, 2006; Patra et al., 2011; Ostler et al., 2016). The problem was mainly linked to biases in the meridional Brewer-Dobson circulation in the stratosphere and in the rate of troposphere-stratosphere exchange. However, neither mechanism was analysed in detail. Indeed, the observed changes in Fig. 12 may partly reflect discrepancies in the Brewen-Dobson circulation projected from the initial conditions. In particular, too-rapid meridional overturning in the months prior to the assimilation would have transported excess of CH_4 from the tropics and to the high latitudes. In the companion study, Stanevich et al. (2020) show that the stratospheric bias in GEOS-Chem can also be due to increased numerical diffusion at the coarse horizontal model resolution. This leads to additional unphysical horizontal mixing between the troposphere and the stratosphere and between the high latitudes and the tropics in the stratosphere.

4.2 Tropospheric Bias

4.2.1 Pattern of forcing terms

The forcing terms are corrections applied to the CH_4 fields at each model time step. This time step is equal to 30 min and 15 min for the $4° \times 5°$ and $2° \times 2.5°$ simulations, respectively. In order to compare the forcing terms in the two simulations, we added together the state corrections at two successive $2° \times 2.5°$ time steps. Therefore, all forcing terms discussed in this section are presented for 30 min time intervals. The first column in Fig. 13 presents forcing terms in the troposphere optimized by the **WC_4x5** assimilation. The observed structure of the forcing terms simultaneously mitigated model errors from multiple sources. In this section, we attempt to give the most likely explanation of the retrieved pattern of the state correction and identify sources of regional biases.

In general, the original a priori CH_4 fields can be affected by model errors that either occurred during the assimilation period or have been projected onto the assimilation window from the initial conditions. Here, we investigate the former case. Given the results of the OSSE with biased initial conditions in Sect. 3.1, we focus in Fig. 13 on the mean forcing terms in the last three months of the assimilation (March-May 2010) as they are much more likely to be related to recent model errors rather than to biases in the initial conditions. The temporally averaged structure also gives insight into systematic model errors and is easier to interpret. Figure 13 (first column) shows that negative forcing terms dominate near the surface and in the lower

troposphere, particularly over Europe, Equatorial Africa and East Asia. The CH_4 reduction at the surface is consistent with NOAA observations. Positive state corrections are more frequently found in the upper troposphere, mainly in mid-latitudes over the Pacific and Atlantic oceans as well as over Europe and significant part of Russia. There are also several regions, such as eastern China and equatorial Africa, where the forcing terms are negative throughout the entire tropospheric column.

5 Vertical slices over mid-latitudes (bottom right panel) show that strong negative corrections over the east coast of Asia and North America are accompanied by positive corrections in the upper troposphere downwind of the continents. Forcing terms are generally weaker in the lower troposphere over the oceans where we lack GOSAT observations.

Generally, corrections of one sign with monotonically decaying magnitude from the surface to the upper troposphere could be associated with biases in the surface emissions, while the dipole structures with corrections of the opposite sign in the upper

10 and lower troposphere could be related to errors in vertical transport. However, it is not feasible to uniquely identify the origin of model errors from the pattern of forcing terms because model errors from separate sources are mixed in the atmosphere and the estimation of the forcing terms is an under-constrained inverse problem.

Still, we may try to identify possible sources of model errors. For example, initial assessment of the state corrections pointed to potential issues in vertical transport. Indeed, the dipole structure of the forcing terms could indicate that upward transport

15 of CH_4 in mid-latitudes may be insufficient, particularly, over regions with strong vertical CH_4 gradients that are present over large sources of CH_4. In NH mid-latitudes the major CH_4 source regions are China, the US, and Europe. Moreover, the eastern parts of China and North America are located in regions of significant extra-tropical cyclone activity (Stohl, 2001; Shaw et al., 2016), where CH_4 emitted from the surface is being lifted into the free troposphere in warm conveyor belts associated with these cyclones (Kowol-Santen et al., 2001; Li et al., 2005; Sinclair et al., 2008; Lin et al., 2010). Moist convection over land

20 could also contribute to the total transport bias, however convective transport is not strong over these mid-latitude regions during the months of February-May.

Similar vertical structure in the forcing terms was identified above and downwind of eastern North America and China (see the first column, forth row of Fig. 13). The WC method applied negative corrections over the land, from the surface to the upper troposphere, and large positive corrections in the upper troposphere and weakly negative correction in the lower troposphere

25 over the oceans downwind of the continents. The WC method may suggest that vertical transports over eastern parts of the continents has to be stronger. In such case, more CH_4 emitted from local sources reaches the middle to upper troposphere and is transported away from the continents by strong westerly winds. Meanwhile, CH_4 concentrations in the entire atmospheric column over land and in the lower troposphere over the adjacent oceans are reduced. Therefore, the large positive a priori bias between the model and GOSAT over China shown in Fig. 6 (first column) may partly be attributed to weak local uplift of CH_4.

30 The observed structure of the forcing terms cannot be uniquely attributed only to biases in vertical transport. The WC method significantly reduced CH_4 in the stratosphere at high latitudes. If biases in the stratospheric CH_4 fields are induced by transport errors, the total CH_4 budget has to be conserved. Therefore, CH_4 removal from the high-latitude stratosphere has to be compensated for in either the tropical stratosphere or the upper troposphere. Hence, the positive forcing terms in the upper troposphere, particularly, in the vicinity of the westerly jet, may also be partly related to model errors in the

Another region of interest, as suggested by the WC assimilation (Fig. 13, first column, third row), is equatorial Africa. Similar to China, a large positive a priori model XCH$_4$ bias was found here. However, due to the observational coverage,
5 there are limited direct constraints on the CH$_4$ outflow from equatorial Africa except for sparse GOSAT observations over South America. While the African XCH$_4$ bias could be related to positively biased local a priori surface emissions, the WC assimilation also suggested another transport related explanation. The WC assimilation applied negative CH$_4$ forcing terms over central Africa and positive forcing terms downwind in the middle troposphere (between 400 and 800 hPa) over the Atlantic Ocean. Such a pattern of state correction could point to potential errors in CH$_4$ outflow from the African continent.
10 Southern Africa is characterized by a persistent high pressure system that drives easterly outflow from southern tropical Africa to the Atlantic in the lower to middle troposphere (Garstang et al., 1996). In their analysis of the sources of moisture in the Congo Basin, Dyer et al. (2017) showed that there is a strong export of moisture from southern tropical Africa to the Atlantic between 800-500 hPa. Furthermore, Arellano et al. (2006) found, in their inversion analysis of carbon monoxide (CO) data from the MOPITT instrument, a discrepancy between their a posteriori CO and observations at Ascension Island, which they
15 speculated could be due to errors in the altitude dependence of the outflow from Africa in the GEOS-Chem model. It is possible that too-much CH$_4$ is being convectively ~~lofted~~ lifted to the upper troposphere over central Africa and not enough is exported out over the Atlantic in the lower troposphere. Figure ~~??~~ 1 (first column) displays the bias in CH$_4$ fields when convection was turned off in the model. This caused CH$_4$ emitted over Africa to take a different transport pathway. Instead of being lifted up over the continent, more CH$_4$ was transported out to the Atlantic in the lower to middle troposphere between 500 and 900 hPa.
20 Under such conditions, CH$_4$ is simultaneously depleted over the continent and increased over the Atlantic, which is similar to what the WC forcing terms suggest. We cannot determine the exact origin of the XCH$_4$ bias over Africa, but the forcing terms do suggest the presence of a transport bias.

The estimation of the forcing terms is an under-constrained inverse problem. Consequently, here we evaluate the impact of reducing the dimensionality of inverse problem by limiting the region of the atmosphere where the forcing terms should be
25 applied. This was done in the **WC_4REG_4x5** assimilation, in which we restricted the forcing optimization to the stratosphere and only over the main CH$_4$ anthropogenic emission regions in the troposphere. The results presented in Sect. 3.2.1 suggested that the **WC_4x5** and **WC_4REG_4x5** assimilations produced similar fits to the independent observations. Therefore, errors affecting the model, at least, at the location of the validation stations could emerge from either the NA, CH, EU, EQAf, or STRAT regions. The second column in Fig. 13 presents the structure of optimized forcing terms from the **WC_4REG_4x5**
30 assimilation where the number of optimized variables was reduced using the forcing mask **G**. Over China and North America, the forcing terms acquired a better defined dipole structure with positive correction in the upper troposphere and negative correction in the lower troposphere. Over equatorial Africa, the region of positive corrections in the mid-troposphere moved closer to the continent.

4.2.2 Dependence of the forcing terms on model resolution

Coarsening the model resolution from $2° \times 2.5°$ to $4° \times 5°$ can be considered as equivalent to introducing errors in the finer resolution model. Yu et al. (2017) and Stanevich et al. (2020) showed that at coarse resolution vertical transport in GEOS-Chem is weakened due to loss of eddy mass flux and ~~incorrect regridding of~~ air mass flux in the regridding of the meteorological fields. Stanevich et al. (2020) also showed that the efficiency of transport barriers is reduced due to increased numerical diffusion, which causes unphysical mixing between the interior and the exterior of the polar vortex, too rapid mixing of CH_4 between the tropical and extratropical branch of the Brewer-Dobson circulation, and increased troposphere-stratosphere exchange. Thus, in Fig. 13 we compare the forcing terms from the $4° \times 5°$ assimilation (**WC_4x5**) with those from the $2° \times 2.5°$ WC assimilation (**WC_2x25**). Differences between the $2° \times 2.5°$ and $4° \times 5°$ forcing represent the response of the WC method to the resolution-induced transport errors. We found that the magnitude of the negative forcing term was reduced in the lower troposphere, particularly, over China. Similarly, the magnitude of positive forcing terms was reduced in the upper troposphere. The pattern of forcing terms on the vertical slice at mid-latitudes became significantly weaker. Comparison of Figs. 6 and 7 also suggested smaller stratospheric corrections at the $2° \times 2.5°$ resolution. At the same time, the structure and magnitude of forcing terms at the equator (particularly, over equatorial Africa) was not significantly affected by the increase of resolution.

Several conclusions follow from Fig. 13. First, the results suggest that a large fraction of model errors at $4° \times 5°$ resolution, particularly, in the stratosphere and over mid-latitudes in the troposphere are resolution-induced. Second, although the magnitude of the forcing terms at the $2° \times 2.5°$ resolution is smaller, the pattern remains similar, which implies that the $2° \times 2.5°$ resolution model may still be affected by the same type of transport errors. Third, the assumptions made about sources of model errors in the tropics, particularly, over equatorial Africa, still apply to the $2° \times 2.5°$ simulation as the structure and magnitude of forcing terms remained unresponsive to the model resolution. It is possible that these regions are dominated by discrepancies in moist convective transport that are large at $2° \times 2.5°$ and $4° \times 5°$. ~~Finally, the results strongly suggest that the WC assimilation and the GOSAT observations have the potential to diagnose transport errors at both model resolutions.~~

5 Conclusions

In this study, we assessed errors in the global GEOS-Chem chemistry transport model during the four-month period of February-May 2010 using the weak constraint 4D-Var data assimilation method at the model resolutions of $4° \times 5°$ and $2° \times 2.5°$. This was done by constraining simulated CH_4 fields with GOSAT XCH_4 retrievals. This represents the first application of WC 4D-Var scheme for assimilation of GOSAT XCH_4 retrievals to characterize model errors in a CTM.

An analysis of the sensitivity of the GOSAT measurements to the atmospheric CH_4 state found that the XCH_4 retrievals are most sensitive to CH_4 mass changes in the stratosphere and in the upper troposphere in the northern hemisphere, which was explained by the GOSAT observational coverage and stronger horizontal winds in the UTLS, allowing the CH_4 perturbations to be observed by a larger number of measurements. Sensitivity at the equator was about half that at northern mid-latitudes. In a series of OSSEs, the observations and the WC method were tested to determine the ability of the system to recover "unknown"

errors in CH$_4$ fields associated with artificially introduced biases in emissions, convection, chemistry, and initial conditions. We found that when not supplied with any information about the errors, the WC method was able to significantly mitigate biases in the CH$_4$ fields with slowly changing spatial structures, but was not able to correct strongly localized biases, particularly, those in the boundary layer. Despite having almost flat averaging kernels in the troposphere, our analysis showed that the GOSAT XCH$_4$ retrievals could help constrain the vertical distribution of model errors when convection was turned off in the model. The WC method needed about a month to recover the bias introduced in the initial condition in the troposphere and about two months to do so in the stratosphere. Generally, the method was successful in mitigating model errors of "unknown" origin and magnitude. However, more optimal performance could be achieved by supplying the method with additional information about model errors, such as their temporal and spatial correlation using the model errors covariance matrix **Q**. However, characterizing these correlations will be challenging.

The WC method was tuned in a set of experiments to diagnose real model errors in the GEOS-Chem CTM at the $4° \times 5°$ resolution. The a posteriori model fit to independent observations, such as ACE-FTS, HIPPO-3, TCCON and NOAA surface measurements, was used to evaluate the assimilation. Initial comparisons suggested that GEOS-Chem was affected by biases not solely related to discrepancies in surface emissions. Results suggested that the modelled CH$_4$ fields at the location of most NH TCCON stations were affected by slowly varying biases, however, a few stations, such as Bialystok, Bremen and Karlsruhe, were more likely influenced by errors varying on time scales of one week. The evaluations pointed to a large positive bias in the stratosphere relative to ACE-FTS and HIPPO-3 measurements, and weakly negative bias in the middle to upper troposphere relative to HIPPO-3 data. The WC assimilation was able to mitigate the negative tropospheric bias and partly removed the stratosphere bias. We found that the SC 4D-Var assimilation that optimized the surface emissions had only limited impact on the model fits. Furthermore, the WC assimilation at $4° \times 5°$ resolution performed better than the SC assimilation at $2° \times 2.5°$ resolution. Meanwhile, the results showed that running the a priori model at $2° \times 2.5°$ resolution produced better agreement with TCCON observations than the a posteriori fields from the SC 4D-Var surface emission optimization at $4° \times 5°$.

State corrections at the $4° \times 5°$ resolution also explicitly pointed to issues with vertical transport, suggesting that vertical transport of CH$_4$ in mid-latitudes over the large CH$_4$ source regions of eastern China and North America is too weak. In the tropics, the WC inversion corrected for large positive XCH$_4$ biases over equatorial Africa. From the pattern of forcing terms, it remained unclear whether the bias over Africa was related to surface emissions. However, the WC method suggested the possibility of biased CH$_4$ outflow from the African continent to the Atlantic Ocean in the mid-troposphere, which could be related to a discrepancy in the partitioning between deep convection transport to the upper troposphere and shallow outflow to the Atlantic Ocean.

In a companion analysis, Stanevich et al. (2020) examined the impact of model resolution on the CH$_4$ simulation and found larger model biases at $4° \times 5°$ compared to $2° \times 2.5°$. We found that assimilating the GOSAT data at the higher resolution of $2° \times 2.5°$ produced state corrections that were similar to those obtained at $4° \times 5°$, however, the magnitude of these corrections in the stratosphere and in the mid-latitude troposphere was significantly reduced at the higher resolution. This suggested that the model at both resolutions was affected by transport errors of similar origin, although less so at the $2° \times 2.5°$ resolution, and a significant fraction of these errors was induced by the model resolution itself. The WC assimilation also corrected for

the negative CH_4 bias relative the ACE-FTS and HIPPO in the northern mid-latitude stratosphere, found only at the $2° \times 2.5°$ resolution, and located this bias particularly over the Himalayas. However, the origin of this bias remained unclear.

In our analysis, we used only GOSAT CH_4 data over land. However, XCH_4 glint measurements over oceans could help better constrain the vertical structure of the model errors. The WC 4D-Var assimilation of shorter-lived species, such as CO, could also help better diagnose model errors, especially when transport and emission errors mask each other in CH_4 fields, although shorter-lived species may also be more strongly affected by errors in chemistry. The advantage of CH_4 is its longer memory of model transport, however shorter-lived gases are more strongly affected by and, hence, may be more sensitive to the same model errors. Clearly, the detected transport error at the $4° \times 5°$ resolution would have considerable impact on inferred emissions if the ~~model~~ evolution of the model state were assumed to be perfect, as is the case in SC 4D-Var. Instead of reducing positive high-latitude bias in the stratosphere, the $4° \times 5°$ SC 4D-Var surface flux assimilation negatively biased the lower troposphere. The SC inversion also significantly reduced Chinese CH_4 emissions by incorrectly attributing model errors in vertical transport. Some of the detected transport error were significantly smaller at the $2° \times 2.5°$ resolution, while others remained resolution-independent. The effect of these remaining errors at the $2° \times 2.5°$ resolution has to be further investigated.

~~Potentially,~~ In the context of optimizing fluxes, potentially, any CTM may be improved if the signal from the surface emissions can be separated from other model errors. This would be a rather challenging task for GOSAT XCH_4 measurements. Further analysis is needed on this problem, particularly on the design of the model error covariance matrix $\mathbf{Q}$. For example, Trémolet (2007) proposed a design based on statistics of model tendencies. The $\mathbf{Q}$ matrix had a rather primitive structure in our analysis, although sufficient for the objectives of this work. Based on our initial assessment of model errors, the structure of $\mathbf{Q}$ can be further improved. In the meantime, the WC 4D-Var method has a number of immediate useful applications. In general, it is a valuable instrument for diagnosing model errors. It can also be used as a tool to produce a better estimate of the CH_4 state in the model in order to provide boundary and initial conditions for forecasting purposes or regional-scale analysis at higher spatial resolution.

6   Data and code availability

The GOSAT satellite data are available at http://www.esa-ghg-cci.org/sites/default/files/documents/public/documents/GHG-CCI_DATA.html. The TCCON data are available at http://tccondata.org/. The NOAA-ESRL Global Greenhouse Gas Reference Network data are available at https://www.esrl.noaa.gov/gmd/dv/data/. The HIPPO aircraft data are available at http://hippo.ornl.gov/data_access/. The ACE-FTS data are available at https://database.scisat.ca/level2/ace_v3.5_v3.6/, and registration is required to download the data. The code for the GEOS-Chem model and its adjoint (with the weak constraint capability) is publicly available and can be downloaded from www.geos-chem.org. The output from the GEOS-Chem model simulations used in this analysis are available upon request.

References

Alexe, M., Bergamaschi, P., Segers, A., Detmers, R., Butz, A., Hasekamp, O., Guerlet, S., Parker, R., Boesch, H., Frankenberg, C., Scheepmaker, R. A., Dlugokencky, E., Sweeney, C., Wofsy, S. C., and Kort, E. A.: Inverse modelling of CH_4 emissions for 2010–2011 using different satellite retrieval products from GOSAT and SCIAMACHY, Atmospheric Chemistry and Physics, 15, 113–133, doi:10.5194/acp-15-113-2015, 2015.

Arellano, A. F., Kasibhatla, P. S., Giglio, L., van der Werf, G. R., Randerson, J. T., and Collatz, G. J.: Time-dependent inversion estimates of global biomass-burning CO emissions using Measurement of Pollution in the Troposphere (MOPITT) measurements, Journal of Geophysical Research: Atmospheres, 111, doi:10.1029/2005JD006613, 2006.

Bergamaschi, P., Frankenberg, C., Meirink, J. F., Krol, M., Villani, M. G., Houweling, S., Dentener, F., Dlugokencky, E. J., Miller, J. B., Gatti, L. V., Engel, A., and Levin, I.: Inverse modeling of global and regional CH_4 emissions using SCIAMACHY satellite retrievals, Journal of Geophysical Research: Atmospheres, 114, doi:10.1029/2009JD012287, 2009.

Bernath, P. F., McElroy, C. T., Abrams, M. C., Boone, C. D., Butler, M., Camy-Peyret, C., Carleer, M., Clerbaux, C., Coheur, P.-F., Colin, R., DeCola, P., DeMazière, M., Drummond, J. R., Dufour, D., Evans, W. F. J., Fast, H., Fussen, D., Gilbert, K., Jennings, D. E., Llewellyn, E. J., Lowe, R. P., Mahieu, E., McConnell, J. C., McHugh, M., McLeod, S. D., Michaud, R., Midwinter, C., Nassar, R., Nichitiu, F., Nowlan, C., Rinsland, C. P., Rochon, Y. J., Rowlands, N., Semeniuk, K., Simon, P., Skelton, R., Sloan, J. J., Soucy, M.-A., Strong, K., Tremblay, P., Turnbull, D., Walker, K. A., Walkty, I., Wardle, D. A., Wehrle, V., Zander, R., and Zou, J.: Atmospheric Chemistry Experiment (ACE): Mission overview, Geophysical Research Letters, 32, doi:10.1029/2005GL022386, 2005.

Blumenstock, T., Hase, F., Schneider, M., Garcia, O., and Sepulveda., E.: TCCON data from Izana, Tenerife, Spain, Release GGG2014.R0, doi:10.14291/tccon.ggg2014.izana01.R0/1149295, 2017.

Boone, C. D., Nassar, R., Walker, K. A., Rochon, Y., McLeod, S. D., Rinsland, C. P., and Bernath, P. F.: Retrievals for the atmospheric chemistry experiment Fourier-transform spectrometer, Applied Optics, 44, 7218–7231, doi:10.1364/AO.44.007218, 2005.

Boone, C. D., Walker, K. A., and Bernath, P. F.: Version 3 Retrievals for the Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE-FTS), in: The Atmospheric Chemistry Experiment ACE at 10: A Solar Occultation Anthology, edited by: Bernath, P. F., A, Deepak Publishing, Hampton, Virginia, USA, p. 103–127, 2013.

Bousserez, N., Henze, D. K., Rooney, B., Perkins, A., Wecht, K. J., Turner, A. J., Natraj, V., and Worden, J. R.: Constraints on methane emissions in North America from future geostationary remote-sensing measurements, Atmospheric Chemistry and Physics, 16, 6175–6190, doi:10.5194/acp-16-6175-2016, 2016.

Buchwitz, M., Reuter, M., Schneising, O., Hewson, W., Detmers, R., Boesch, H., Hasekamp, O., Aben, I., Bovensmann, H., Burrows, J., Butz, A., Chevallier, F., Dils, B., Frankenberg, C., Heymann, J., Lichtenberg, G., Mazière, M. D., Notholt, J., Parker, R., Warneke, T., Zehner, C., Griffith, D., Deutscher, N., Kuze, A., Suto, H., and Wunch, D.: Global satellite observations of column-averaged carbon dioxide and methane: The GHG-CCI XCO_2 and XCH_4 CRDP3 data set, Remote Sensing of Environment, 203, 276 – 295, doi:https://doi.org/10.1016/j.rse.2016.12.027, earth Observation of Essential Climate Variables, 2017.

Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C.: A Limited Memory Algorithm for Bound Constrained Optimization, SIAM Journal on Scientific Computing, 16, 1190–1208, doi:10.1137/0916069, 1995.

Byrne, B., Jones, D. B. A., Strong, K., Zeng, Z.-C., Deng, F., and Liu, J.: Sensitivity of CO_2 surface flux constraints to observational coverage, Journal of Geophysical Research: Atmospheres, 122, 6672–6694, doi:10.1002/2016JD026164, 2017.

De Mazière, M., Vigouroux, C., Bernath, P. F., Baron, P., Blumenstock, T., Boone, C., Brogniez, C., Catoire, V., Coffey, M., Duchatelet, P., Griffith, D., Hannigan, J., Kasai, Y., Kramer, I., Jones, N., Mahieu, E., Manney, G. L., Piccolo, C., Randall, C., Robert, C., Senten, C., Strong, K., Taylor, J., Tétard, C., Walker, K. A., and Wood, S.: Validation of ACE-FTS v2.2 methane profiles from the upper troposphere to the lower mesosphere, Atmospheric Chemistry and Physics, 8, 2421–2435, doi:10.5194/acp-8-2421-2008, 2008.

5 Dee, D. P. and Da Silva, A. M.: Data assimilation in the presence of forecast bias, Quarterly Journal of the Royal Meteorological Society, 124, 269–295, doi:10.1002/qj.49712454512, 1998.

Deng, F., Jones, D. B. A., Henze, D. K., Bousserez, N., Bowman, K. W., Fisher, J. B., Nassar, R., O'Dell, C., Wunch, D., Wennberg, P. O., Kort, E. A., Wofsy, S. C., Blumenstock, T., Deutscher, N. M., Griffith, D. W. T., Hase, F., Heikkinen, P., Sherlock, V., Strong, K., Sussmann, R., and Warneke, T.: Inferring regional sources and sinks of atmospheric CO_2 from GOSAT XCO_2 data, Atmospheric
10 Chemistry and Physics, 14, 3703–3727, doi:10.5194/acp-14-3703-2014, 2014.

Derber, J. C.: A Variational Continuous Assimilation Technique, Monthly Weather Review, 117, 2437–2446, doi:10.1175/1520-0493(1989)117<2437:AVCAT>2.0.CO;2, 1989.

Dlugokencky, E., Lang, P., Crotwell, A., Mund, J., Crotwell, M., and Thoning, K.: Atmospheric Methane Dry Air Mole Fractions from the NOAA ESRL Carbon Cycle Cooperative Global Air Sampling Network, 1983-2015, doi:Path:
15 ftp://aftp.cmdl.noaa.gov/data/trace_gases/ch4/flask/surface/, 2016.

Dlugokencky, E. J., Bruhwiler, L., White, J. W. C., Emmons, L. K., Novelli, P. C., Montzka, S. A., Masarie, K. A., Lang, P. M., Crotwell, A. M., Miller, J. B., and Gatti, L. V.: Observational constraints on recent increases in the atmospheric CH_4 burden, Geophysical Research Letters, 36, doi:10.1029/2009GL039780, 2009.

Dyer, E. L. E., Jones, D. B. A., Nusbaumer, J., Li, H., Collins, O., Vettoretti, G., and Noone, D.: Congo Basin precipitation: As-
20 sessing seasonality, regional interactions, and sources of moisture, Journal of Geophysical Research: Atmospheres, 122, 6882–6898, doi:10.1002/2016JD026240, 2017.

Etheridge, D. M., Pearman, G. I., and Fraser, P. J.: Changes in tropospheric methane between 1841 and 1978 from a high accumulation-rate Antarctic ice core, Tellus B, 44, 282–294, doi:10.1034/j.1600-0889.1992.t01-3-00006.x, 1992.

European Commission Joint Research Centre/Netherlands Environmental Assessment Agency: Emission Database for Global Atmospheric
25 Research (EDGAR), release version 4.0, http://edgar.jrc.ec.europa.eu, 2009.

Fraser, A., Palmer, P. I., Feng, L., Boesch, H., Cogan, A., Parker, R., Dlugokencky, E. J., Fraser, P. J., Krummel, P. B., Langenfelds, R. L., O'Doherty, S., Prinn, R. G., Steele, L. P., van der Schoot, M., and Weiss, R. F.: Estimating regional methane surface fluxes: the relative importance of surface and GOSAT mole fraction measurements, Atmospheric Chemistry and Physics, 13, 5697–5713, doi:10.5194/acp-13-5697-2013, 2013.

30 Fung, I., John, J., Lerner, J., Matthews, E., Prather, M., Steele, L. P., and Fraser, P. J.: Three-dimensional model synthesis of the global methane cycle, Journal of Geophysical Research: Atmospheres, 96, 13 033–13 065, doi:10.1029/91JD01247, 1991.

Garstang, M., Tyson, P. D., Swap, R., Edwards, M., Kållberg, P., and Lindesay, J. A.: Horizontal and vertical transport of air over southern Africa, Journal of Geophysical Research: Atmospheres, 101, 23 721–23 736, doi:10.1029/95JD00844, 1996.

Griffith, D. W. T., Velazco, V. A., Deutscher, N., Murphy, C., Jones, N., Wilson, S., Macatangay, R., Kettlewell,
35 G., Buchholz, R. R., and Riggenbach., M.: TCCON data from Wollongong, Australia, Release GGG2014.R0, doi:10.14291/tccon.ggg2014.wollongong01.R0/1149291, 2017.

Hase, F. a. T. B., Dohe, S., Gross, J., and Kiel., M.: TCCON data from Karlsruhe, Germany, Release GGG2014.R0, doi:10.14291/tccon.ggg2014.karlsruhe01.R0/1149270, 2017.

Henze, D. K., Hakami, A., and Seinfeld, J. H.: Development of the adjoint of GEOS-Chem, Atmospheric Chemistry and Physics, 7, 2413–2433, doi:10.5194/acp-7-2413-2007, 2007.

Hewson, W., Bösch, H., and Parker, R.: ESA Climate Change Initiative (CCI) Comprehensive Error Characterisation Report: University of Leicester proxy retrieval algorithm for XCH_4 CRDP-OCPR v5.2, 2015.

Kaplan, J. O.: Wetlands at the Last Glacial Maximum: Distribution and methane emissions, Geophysical Research Letters, 29, 31–34, doi:10.1029/2001GL013366, 2002.

Keller, M.: Mitigating model error in CO emission estimation, Ph.D. thesis, University of Toronto (Canada), 2014.

Kiemle, C., Kawa, S. R., Quatrevalet, M., and Browell, E. V.: Performance simulations for a spaceborne methane lidar mission, Journal of Geophysical Research: Atmospheres, 119, 4365–4379, doi:10.1002/2013JD021253, 2014.

Kivi, R. and Heikkinen, P.: Fourier transform spectrometer measurements of column CO_2 at Sodankylä, Finland, Geoscientific Instrumentation, Methods and Data Systems, 5, 271–279, doi:10.5194/gi-5-271-2016, 2016.

Kivi, R., Heikkinen, P., and Kyro., E.: TCCON data from Sodankyla, Finland, Release GGG2014.R0, doi:10.14291/tccon.ggg2014.sodankyla01.R0/1149280, 2017.

Kowol-Santen, J., Beekmann, M., Schmitgen, S., and Dewey, K.: Tracer analysis of transport from the boundary layer to the free troposphere, Geophysical Research Letters, 28, 2907–2910, doi:10.1029/2001GL012908, 2001.

Kuze, A., Suto, H., Nakajima, M., and Hamazaki, T.: Thermal and near infrared sensor for carbon observation Fourier-transform spectrometer on the Greenhouse Gases Observing Satellite for greenhouse gases monitoring, Applied Optics, 48, 6716–6733, doi:10.1364/AO.48.006716, 2009.

Lamarque, J.-F., Khattatov, B., Yudin, V., Edwards, D. P., Gille, J. C., Emmons, L. K., Deeter, M. N., Warner, J., Ziskin, D. C., Francis, G. L., Ho, S., Mao, D., Chen, J., and Drummond, J. R.: Application of a bias estimator for the improved assimilation of Measurements of Pollution in the Troposphere (MOPITT) carbon monoxide retrievals, Journal of Geophysical Research: Atmospheres, 109, doi:10.1029/2003JD004466, 2004.

Li, Q., Jacob, D. J., Park, R., Wang, Y., Heald, C. L., Hudman, R., Yantosca, R. M., Martin, R. V., and Evans, M.: North American pollution outflow and the trapping of convectively lifted pollution by upper-level anticyclone, Journal of Geophysical Research: Atmospheres, 110, doi:10.1029/2004JD005039, 2005.

Lin, M., Holloway, T., Carmichael, G. R., and Fiore, A. M.: Quantifying pollution inflow and outflow over East Asia in spring with regional and global models, Atmospheric Chemistry and Physics, 10, 4221–4239, doi:10.5194/acp-10-4221-2010, 2010.

Lin, S.-J. and Rood, R. B.: Multidimensional Flux-Form Semi-Lagrangian Transport Schemes, Monthly Weather Review, 124, 2046–2070, doi:10.1175/1520-0493(1996)124<2046:MFFSLT>2.0.CO;2, 1996.

Liu, J., Bowman, K. W., and Henze, D. K.: Source-receptor relationships of column-average CO_2 and implications for the impact of observations on flux inversions, Journal of Geophysical Research: Atmospheres, 120, 5214–5236, doi:10.1002/2014JD022914, 2015.

Locatelli, R., Bousquet, P., Saunois, M., Chevallier, F., and Cressot, C.: Sensitivity of the recent methane budget to LMDz sub-grid-scale physical parameterizations, Atmospheric Chemistry and Physics, 15, 9765–9780, doi:10.5194/acp-15-9765-2015, 2015.

McNorton, J., Chipperfield, M. P., Gloor, M., Wilson, C., Feng, W., Hayman, G. D., Rigby, M., Krummel, P. B., O'Doherty, S., Prinn, R. G., Weiss, R. F., Young, D., Dlugokencky, E., and Montzka, S. A.: Role of OH variability in the stalling of the global atmospheric CH_4 growth rate from 1999 to 2006, Atmospheric Chemistry and Physics, 16, 7943–7956, doi:10.5194/acp-16-7943-2016, 2016.

Monteil, G., Houweling, S., Butz, A., Guerlet, S., Schepers, D., Hasekamp, O., Frankenberg, C., Scheepmaker, R., Aben, I., and Röckmann, T.: Comparison of CH_4 inversions based on 15 months of GOSAT and SCIAMACHY observations, Journal of Geophysical Research: Atmospheres, 118, 11,807–11,823, doi:10.1002/2013JD019760, 2013.

Moorthi, S. and Suarez, M. J.: Relaxed Arakawa-Schubert. A Parameterization of Moist Convection for General Circulation Models, Monthly Weather Review, 120, 978–1002, doi:10.1175/1520-0493(1992)120<0978:RASAPO>2.0.CO;2, 1992.

Mu, M., Randerson, J. T., van der Werf, G. R., Giglio, L., Kasibhatla, P., Morton, D., Collatz, G. J., DeFries, R. S., Hyer, E. J., Prins, E. M., Griffith, D. W. T., Wunch, D., Toon, G. C., Sherlock, V., and Wennberg, P. O.: Daily and 3-hourly variability in global fire emissions and consequences for atmospheric model predictions of carbon monoxide, Journal of Geophysical Research: Atmospheres, 116, doi:10.1029/2011JD016245, 2011.

Murray, L. T., Jacob, D. J., Logan, J. A., Hudman, R. C., and Koshak, W. J.: Optimized regional and interannual variability of lightning in a global chemical transport model constrained by LIS/OTD satellite data, Journal of Geophysical Research: Atmospheres, 117, doi:10.1029/2012JD017934, 2012.

Nakajima, M., Yajima, Y., Hashimoto, M., Shiomi, K., Suto, H., and Imai, H.: The current status of the mission instruments of GOSAT-2, in: EGU General Assembly Conference Abstracts, vol. 19 of *EGU General Assembly Conference Abstracts*, p. 11324, 2017.

Notholt, J., Petri, C., Warneke, T., Deutscher, N., Buschmann, M., Weinzierl, C., Macatangay, R., and Grupe., P.: TCCON data from Bremen, Germany, Release GGG2014.R0, doi:10.14291/tccon.ggg2014.bremen01.R0/1149275, 2017.

Olsen, K. S., Strong, K., Walker, K. A., Boone, C. D., Raspollini, P., Plieninger, J., Bader, W., Conway, S., Grutter, M., Hannigan, J. W., Hase, F., Jones, N., de Mazière, M., Notholt, J., Schneider, M., Smale, D., Sussmann, R., and Saitoh, N.: Comparison of the GOSAT TANSO-FTS TIR CH_4 volume mixing ratio vertical profiles with those measured by ACE-FTS, ESA MIPAS, IMK-IAA MIPAS, and 16 NDACC stations, Atmospheric Measurement Techniques, 10, 3697–3718, doi:10.5194/amt-10-3697-2017, 2017.

Ostler, A., Sussmann, R., Rettinger, M., Deutscher, N. M., Dohe, S., Hase, F., Jones, N., Palm, M., and Sinnhuber, B.-M.: Multistation intercomparison of column-averaged methane from NDACC and TCCON: impact of dynamical variability, Atmospheric Measurement Techniques, 7, 4081–4101, doi:10.5194/amt-7-4081-2014, 2014.

Ostler, A., Sussmann, R., Patra, P. K., Houweling, S., De Bruine, M., Stiller, G. P., Haenel, F. J., Plieninger, J., Bousquet, P., Yin, Y., Saunois, M., Walker, K. A., Deutscher, N. M., Griffith, D. W. T., Blumenstock, T., Hase, F., Warneke, T., Wang, Z., Kivi, R., and Robinson, J.: Evaluation of column-averaged methane in models and TCCON with a focus on the stratosphere, Atmospheric Measurement Techniques, 9, 4843–4859, doi:10.5194/amt-9-4843-2016, 2016.

Park, R. J., Jacob, D. J., Field, B. D., Yantosca, R. M., and Chin, M.: Natural and transboundary pollution influences on sulfate-nitrate-ammonium aerosols in the United States: Implications for policy, Journal of Geophysical Research: Atmospheres, 109, doi:10.1029/2003JD004473, 2004.

Parker, R. and the GHG-CCI the group: ESA Climate Change Initiative (CCI) Algorithm Theoretical Basis Document Version 5 (ATBDv5) - The University of Leicester Full-Physics Retrieval Algorithm for the retrieval of XCO_2 and XCH_4, 2016.

Parker, R., Boesch, H., Cogan, A., Fraser, A., Feng, L., Palmer, P. I., Messerschmidt, J., Deutscher, N., Griffith, D. W. T., Notholt, J., Wennberg, P. O., and Wunch, D.: Methane observations from the Greenhouse Gases Observing SATellite: Comparison to ground-based TCCON data and model calculations, Geophysical Research Letters, 38, doi:10.1029/2011GL047871, 2011.

Parker, R. J., Boesch, H., Byckling, K., Webb, A. J., Palmer, P. I., Feng, L., Bergamaschi, P., Chevallier, F., Notholt, J., Deutscher, N., Warneke, T., Hase, F., Sussmann, R., Kawakami, S., Kivi, R., Griffith, D. W. T., and Velazco, V.: Assessing 5 years of GOSAT Proxy XCH_4 data and associated uncertainties, Atmospheric Measurement Techniques, 8, 4785–4801, doi:10.5194/amt-8-4785-2015, 2015.

Patra, P. K., Houweling, S., Krol, M., Bousquet, P., Belikov, D., Bergmann, D., Bian, H., Cameron-Smith, P., Chipperfield, M. P., Corbin, K., Fortems-Cheiney, A., Fraser, A., Gloor, E., Hess, P., Ito, A., Kawa, S. R., Law, R. M., Loh, Z., Maksyutov, S., Meng, L., Palmer, P. I., Prinn, R. G., Rigby, M., Saito, R., and Wilson, C.: TransCom model simulations of CH_4 and related species: linking transport, surface flux and chemical loss with CH_4 variability in the troposphere and lower stratosphere, Atmospheric Chemistry and Physics, 11, 12 813–12 837, doi:10.5194/acp-11-12813-2011, 2011.

Pickett-Heaps, C. A., Jacob, D. J., Wecht, K. J., Kort, E. A., Wofsy, S. C., Diskin, G. S., Worthy, D. E. J., Kaplan, J. O., Bey, I., and Drevet, J.: Magnitude and seasonality of wetland methane emissions from the Hudson Bay Lowlands (Canada), Atmospheric Chemistry and Physics, 11, 3773–3779, doi:10.5194/acp-11-3773-2011, 2011.

Polonsky, I. N., O'Brien, D. M., Kumer, J. B., O'Dell, C. W., and the geoCARB Team: Performance of a geostationary mission, geoCARB, to measure CO_2, CH_4 and CO column-averaged concentrations, Atmospheric Measurement Techniques, 7, 959–981, doi:10.5194/amt-7-959-2014, 2014.

Prather, M. J., Zhu, X., Strahan, S. E., Steenrod, S. D., and Rodriguez, J. M.: Quantifying errors in trace species transport modeling, Proceedings of the National Academy of Sciences, 105, 19 617–19 621, doi:10.1073/pnas.0806541106, 2008.

Prather, M. J., Holmes, C. D., and Hsu, J.: Reactive greenhouse gas scenarios: Systematic exploration of uncertainties and the role of atmospheric chemistry, Geophysical Research Letters, 39, doi:10.1029/2012GL051440, 2012.

Saad, K. M., Wunch, D., Deutscher, N. M., Griffith, D. W. T., Hase, F., De Mazière, M., Notholt, J., Pollard, D. F., Roehl, C. M., Schneider, M., Sussmann, R., Warneke, T., and Wennberg, P. O.: Seasonal variability of stratospheric methane: implications for constraining tropospheric methane budgets using total column observations, Atmospheric Chemistry and Physics, 16, 14 003–14 024, doi:10.5194/acp-16-14003-2016, 2016.

Santoni, G. W., Daube, B. C., Kort, E. A., Jiménez, R., Park, S., Pittman, J. V., Gottlieb, E., Xiang, B., Zahniser, M. S., Nelson, D. D., Mc-Manus, J. B., Peischl, J., Ryerson, T. B., Holloway, J. S., Andrews, A. E., Sweeney, C., Hall, B., Hintsa, E. J., Moore, F. L., Elkins, J. W., Hurst, D. F., Stephens, B. B., Bent, J., and Wofsy, S. C.: Evaluation of the airborne quantum cascade laser spectrometer (QCLS) measurements of the carbon and greenhouse gas suite − CO_2, CH_4, N_2O, and CO − during the CalNex and HIPPO campaigns, Atmospheric Measurement Techniques, 7, 1509–1526, doi:10.5194/amt-7-1509-2014, 2014.

Sasaki, Y.: Some basic formalisms in numerical variational analysis, Monthly Weather Review, 98, 875–883, doi:10.1175/1520-0493(1970)098<0875:SBFINV>2.3.CO;2, 1970.

Saunois, M., Bousquet, P., Poulter, B., Peregon, A., Ciais, P., Canadell, J. G., Dlugokencky, E. J., Etiope, G., Bastviken, D., Houweling, S., et al.: The global methane budget 2000-2012, Earth System Science Data, 8, 697, 2016.

Schneising, O., Buchwitz, M., Reuter, M., Heymann, J., Bovensmann, H., and Burrows, J. P.: Long-term analysis of carbon dioxide and methane column-averaged mole fractions retrieved from SCIAMACHY, Atmospheric Chemistry and Physics, 11, 2863–2880, doi:10.5194/acp-11-2863-2011, 2011.

Shaw, T., Baldwin, M., A. Barnes, E., Caballero, R., Garfinkel, C., Hwang, Y.-T., Li, C., A. O'Gorman, P., Rivière, G., R. Simpson, I., and Voigt, A.: Storm track processes and the opposing influences of climate change, Nature Geoscience, 9, 656–664, doi:10.1038/ngeo2783, 2016.

Sheng, J.-X., Jacob, D. J., Turner, A. J., Maasakkers, J. D., Benmergui, J., Bloom, A. A., Arndt, C., Gautam, R., Zavala-Araiza, D., Boesch, H., and Parker, R. J.: 2010–2016 methane trends over Canada, the United States, and Mexico observed by the GOSAT satellite: contributions from different source sectors, Atmospheric Chemistry and Physics, 18, 12 257–12 267, doi:10.5194/acp-18-12257-2018, 2018.

Sherlock, V., Connor, B., Robinson, J., Shiona, H., Smale, D., and Pollard., D.: TCCON data from Lauder, New Zealand, 125HR, Release GGG2014.R0, doi:10.14291/tccon.ggg2014.lauder02.R0/1149298, 2017.

Sinclair, V. A., Gray, S. L., and Belcher, S. E.: Boundary-layer ventilation by baroclinic life cycles, Quarterly Journal of the Royal Meteorological Society, 134, 1409–1424, doi:10.1002/qj.293, 2008.

5   Stanevich, I., Jones, D. B. A., Strong, K., Parker, R. J., Boesch, H., Wunch, D., Notholt, J., Petri, C., Warneke, T., Sussman, R., Schneider, M., Hase, F., Kivi, R., Deutscher, N. M., Walker, K. A., and Deng, F.: Characterizing model errors in chemical transport modelling of methane: Impact of coarse model resolution, Geoscientific Model Development, 13, 3839–3862, doi:10.5194/gmd-13-3839-2020, 2020.

Stohl, A.: A 1-year Lagrangian "climatology" of airstreams in the northern hemisphere troposphere and lowermost stratosphere, Journal of Geophysical Research: Atmospheres, 106, 7263–7279, doi:10.1029/2000JD900570, 2001.

10   Strahan, S. E. and Polansky, B. C.: Meteorological implementation issues in chemistry and transport models, Atmospheric Chemistry and Physics, 6, 2895–2910, doi:10.5194/acp-6-2895-2006, 2006.

Sussmann, R. and Rettinger., M.: TCCON data from Garmisch, Germany, Release GGG2014.R0, doi:10.14291/tccon.ggg2014.garmisch01.R0/1149299, 2017.

Tan, Z., Zhuang, Q., Henze, D. K., Frankenberg, C., Dlugokencky, E., Sweeney, C., Turner, A. J., Sasakawa, M., and Machida, T.: Inverse modeling of pan-Arctic methane emissions at high spatial resolution: what can we learn from assimilating satellite retrievals and using different process-based wetland and lake biogeochemical models?, Atmospheric Chemistry and Physics, 16, 12 649–12 666, doi:10.5194/acp-16-12649-2016, 2016.

Trémolet, Y.: Accounting for an imperfect model in 4D-Var, Quarterly Journal of the Royal Meteorological Society, 132, 2483–2504, doi:10.1256/qj.05.224, 2006.

20   Trémolet, Y.: Model-error estimation in 4D-Var, Quarterly Journal of the Royal Meteorological Society, 133, 1267–1280, doi:10.1002/qj.94, 2007.

Tukiainen, S., Railo, J., Laine, M., Hakkarainen, J., Kivi, R., Heikkinen, P., Chen, H., and Tamminen, J.: Retrieval of atmospheric CH_4 profiles from Fourier transform infrared data using dimension reduction and MCMC, Journal of Geophysical Research: Atmospheres, 121, 10,312–10,327, doi:10.1002/2015JD024657, 2016.

25   Turner, A. J., Jacob, D. J., Wecht, K. J., Maasakkers, J. D., Lundgren, E., Andrews, A. E., Biraud, S. C., Boesch, H., Bowman, K. W., Deutscher, N. M., Dubey, M. K., Griffith, D. W. T., Hase, F., Kuze, A., Notholt, J., Ohyama, H., Parker, R., Payne, V. H., Sussmann, R., Sweeney, C., Velazco, V. A., Warneke, T., Wennberg, P. O., and Wunch, D.: Estimating global and North American methane emissions with high spatial resolution using GOSAT satellite data, Atmospheric Chemistry and Physics, 15, 7049–7069, doi:10.5194/acp-15-7049-2015, 2015.

30   van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R. S., Jin, Y., and van Leeuwen, T. T.: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), Atmospheric Chemistry and Physics, 10, 11 707–11 735, doi:10.5194/acp-10-11707-2010, 2010.

Veefkind, J., Aben, I., McMullan, K., Förster, H., de Vries, J., Otter, G., Claas, J., Eskes, H., de Haan, J., Kleipool, Q., van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H., and Levelt, P.:
35   TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications, Remote Sensing of Environment, 120, 70–83, doi:http://dx.doi.org/10.1016/j.rse.2011.09.027, 2012.

Wang, Z., Warneke, T., Deutscher, N. M., Notholt, J., Karstens, U., Saunois, M., Schneider, M., Sussmann, R., Sembhi, H., Griffith, D. W. T., Pollard, D. F., Kivi, R., Petri, C., Velazco, V. A., Ramonet, M., and Chen, H.: Contributions of the troposphere and stratosphere to CH_4 model biases, Atmospheric Chemistry and Physics, 17, 13 283–13 295, doi:10.5194/acp-17-13283-2017, 2017.

Warneke, T., Messerschmidt, J., Notholt, J., Weinzierl, C., Deutscher, N., Petri, C., Grupe, P., Vuillemin, C., Truong, F., Schmidt, M., Ramonet, M., and Parmentier., E.: TCCON data from Orleans, France, Release GGG2014.R0, doi:10.14291/tccon.ggg2014.orleans01.R0/1149276, 2017.

Waymark, C., Walker, K., Boone, C., and Bernath, P.: ACE-FTS version 3.0 data set: validation and data processing update, Annals of Geophysics, 56, doi:10.4401/ag-6339, 2014.

Wecht, K. J., Jacob, D. J., Wofsy, S. C., Kort, E. A., Worden, J. R., Kulawik, S. S., Henze, D. K., Kopacz, M., and Payne, V. H.: Validation of TES methane with HIPPO aircraft observations: implications for inverse modeling of methane sources, Atmospheric Chemistry and Physics, 12, 1823–1832, doi:10.5194/acp-12-1823-2012, 2012.

Wecht, K. J., Jacob, D. J., Frankenberg, C., Jiang, Z., and Blake, D. R.: Mapping of North American methane emissions with high spatial resolution by inversion of SCIAMACHY satellite data, Journal of Geophysical Research: Atmospheres, 119, 7741–7756, doi:10.1002/2014JD021551, 2014.

Wennberg, P. O., Roehl, C., Wunch, D., Toon, G. C., Blavier, J.-F., Washenfelder, R., Keppel-Aleks, G., Allen, N., and Ayers., J.: TCCON data from Park Falls, Wisconsin, USA, Release GGG2014.R0, doi:10.14291/tccon.ggg2014.parkfalls01.R0/1149161, 2017a.

Wennberg, P. O., Wunch, D., Roehl, C., Blavier, J.-F., Toon, G. C., Allen, N., Dowell, P., Teske, K., Martin, C., and Martin., J.: TCCON data from Lamont, Oklahoma, USA, Release GGG2014.R0, doi:10.14291/tccon.ggg2014.lamont01.R0/1149159, 2017b.

Wofsy, S. et al.: HIPPO Merged 10-Second Meteorology and Atmospheric Chemistry and Aerosol Data (R_20121129), Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory and Oak Ridge and Tennessee.[Avaialble at http://dx.doi.org/10.3334/CDIAC/hippo_010,(Release 20121129)], 2012.

Wofsy, S. C., Team, H. S., Modellers, C., and Teams, S.: HIAPER Pole-to-Pole Observations (HIPPO): fine-grained, global-scale measurements of climatically important atmospheric gases and aerosols, Philosophical Transactions: Mathematical, Physical and Engineering Sciences, 369, 2073–2086, 2011.

Worden, J., Kulawik, S., Frankenberg, C., Payne, V., Bowman, K., Cady-Peirara, K., Wecht, K., Lee, J.-E., and Noone, D.: Profiles of CH_4, HDO, H_2O, and N_2O with improved lower tropospheric vertical resolution from Aura TES radiances, Atmospheric Measurement Techniques, 5, 397–411, doi:10.5194/amt-5-397-2012, 2012.

Wunch, D., Toon, G. C., Blavier, J.-F. L., Washenfelder, R. A., Notholt, J., Connor, B. J., Griffith, D. W. T., Sherlock, V., and Wennberg, P. O.: The Total Carbon Column Observing Network, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 369, 2087–2112, doi:10.1098/rsta.2010.0240, 2011.

Wunch, D., Toon, G. C., Sherlock, V., Deutscher, N. M., Liu, C., Feist, D. G., and Wennberg, P. O.: The total carbon column observing network's GGG2014 data version, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, 10, 2015.

Yoshida, Y., Ota, Y., Eguchi, N., Kikuchi, N., Nobuta, K., Tran, H., Morino, I., and Yokota, T.: Retrieval algorithm for CO_2 and CH_4 column abundances from short-wavelength infrared spectral observations by the Greenhouse gases observing satellite, Atmospheric Measurement Techniques, 4, 717–734, doi:10.5194/amt-4-717-2011, 2011.

Yu, K., Keller, C. A., Jacob, D. J., Molod, A. M., Eastham, S. D., and Long, M. S.: Errors and improvements in the use of archived meteorological data for chemical transport modeling, Geoscientific Model Development Discussions, 2017, 1–22, doi:10.5194/gmd-2017-125, 2017.

Zupanski, D.: A General Weak Constraint Applicable to Operational 4DVAR Data Assimilation Systems, Monthly Weather Review, 125, 2274–2292, doi:10.1175/1520-0493(1997)125<2274:AGWCAT>2.0.CO;2, 1997.

5

Table 1. Evaluation of a priori, **SC_4x5** and **WC_4x5** optimized CH_4 fields using TCCON XCH_4 from the stations listed in Table 2 and NOAA surface *in situ* observation (mean statistics for the period of February-May 2010). The first, second and third columns represent mean difference, standard deviation and correlation between the model and measurements, respectively. The fourth column represents the slope of the regression line with modelled data on the y-axis and measurements on the x-axis.

	Mean Difference [ppb]			Standard Deviation [ppb]			Correlation (R)			Slope of regression		
	Prior	SC	WC	Prior	SC	WC	Prior	SC	WC	Prior	SC	WC
TCCON	9.1	8.2	5.3	15.0	13.8	9.9	0.83	0.86	0.93	1.16	1.14	1.07
In situ	15.1	9.7	9.6	34.5	30.3	28.9	0.88	0.89	0.90	0.90	1.02	0.99

Table 2. Evaluation of a priori, **SC_4x5** and **WC_4x5** optimized CH_4 fields using TCCON XCH_4 (mean station-wise statistics for the period of February-May 2010). The first, second and third columns represent mean difference, standard deviation and correlation between the model and measurements, respectively.

	Mean Difference [ppb]			Standard Deviation [ppb]			Correlation (R)		
	Prior	SC	WC	Prior	SC	WC	Prior	SC	WC
Sodankylä (67.37°N, 26.63°E)	30.0	25.7	13.7	18.9	19.1	12.6	0.49	0.50	0.81
Bialystok (53.23°N, 23.03°E)	11.9	7.3	5.1	9.3	10.6	8.0	0.39	0.43	0.65
Bremen (53.10°N, 8.85°E)	6.3	3.2	0.7	14.3	15.2	10.5	-0.37	-0.28	0.47
Karlsruhe (49.10°N, 8.44°E)	6.4	4.4	0.8	9.7	9.9	8.9	0.33	0.29	0.49
Orleans (47.97°N, 2.11°E)	3.9	3.5	2.5	8.9	9.6	8.3	0.31	0.30	0.51
Garmish (47.48°N, 11.06°E)	9.9	10.0	5.7	9.0	9.7	8.4	0.46	0.56	0.65
Park Falls (45.95°N, 90.27°W)	1.9	3.6	2.3	9.7	10.6	8.5	0.37	0.47	0.65
Lamont (36.60°N, 97.486°W)	1.4	3.7	4.4	11.1	11.5	9.4	0.27	0.30	0.49
Izana (28.30°N, 16.5°W)	-5.8	-5.3	3.1	7.6	8.2	6.7	0.64	0.58	0.72
Wollongong (34.41°S, 150.88°E)	7.5	3.9	3.7	8.9	8.8	8.4	0.58	0.55	0.59
Lauder (45.04°S, 169.68°E)	9.6	9.2	5.9	5.6	5.7	5.4	0.72	0.72	0.73

Mean difference from 1 February - 31 May 2010 between GOSAT XCH$_4$ retrievals based on the original and new CO$_2$ proxy fields. The original CO$_2$ proxy is based on the median of the CO$_2$ distributions from the GEOS-Chem (from the University of Edinburgh), LMDZ/MACC-II and NOAA CarbonTracker models constrained by in-situ surface CO$_2$ observations. The new CO$_2$ proxy is from a CO$_2$ surface flux inversion analysis using the GEOS-Chem model constrained by GOSAT ACOS CO$_2$ retrievals over land. Mean XCH$_4$ fields from February to May 2010. Top panel: XCH$_4$ from the GEOS-Chem model at a resolution of 4$\times$ 5. The model was sampled at the locations and times of the GOSAT observations and smoothed with the GOSAT averaging kernels. Bottom panel: GOSAT XCH$_4$ retrievals based on the new CO$_2$ proxy fields. Distribution of total adjoint sensitivity in GEOS-Chem of GOSAT observations in February-May 2010 to the modeled CH$_4$ distribution (the model state) during the same period. The adjoint sensitivities have been summed over the time period. Shown are (top) the total zonal mean adjoint sensitivities and (bottom) the altitude-longitude cross section of the sensitivities along 34N.

Figure 1. Mean differences in the CH₄ distribution in March 2010 in the OSSE with biased convection. Left column: mean differences between the a priori CH₄ state and the "true" CH₄ state. Middle column: mean differences between the WC optimized CH₄ state and the "true" CH₄ state. Right column: the mean WC state corrections (the forcing terms), in units of ppb. Shown are the latitude-longitude differences at (top row) the surface and (second row) at 300 hPa, as well as the altitude-longitude differences (third row) along the equator and (bottom row) along 42°N. The black boxes indicate the four domains considered for the regional analysis discussed in the text and shown in Fig. 2.
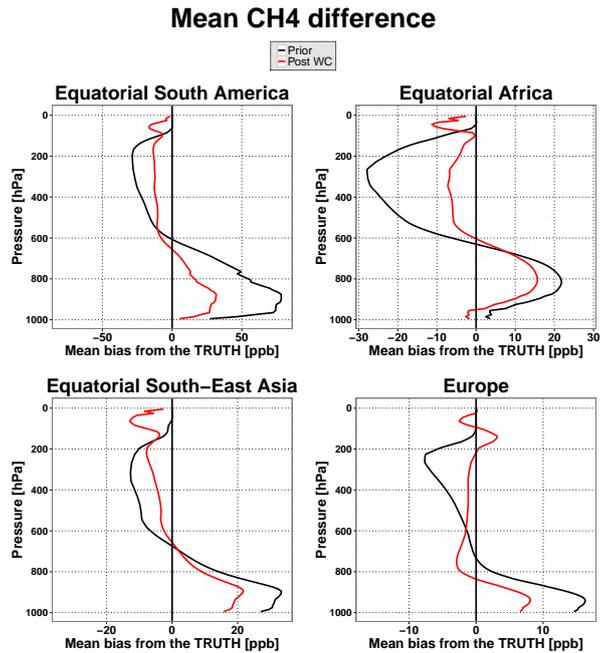
Figure 2. Mean vertical profiles of the CH$_4$ differences in March 2010 in the OSSE with biased convection for the four regions depicted in Fig. 1. The differences are between (black lines) the a priori and the "true" CH$_4$ state and between (red lines) the WC optimized state and the "true" CH$_4$ state.
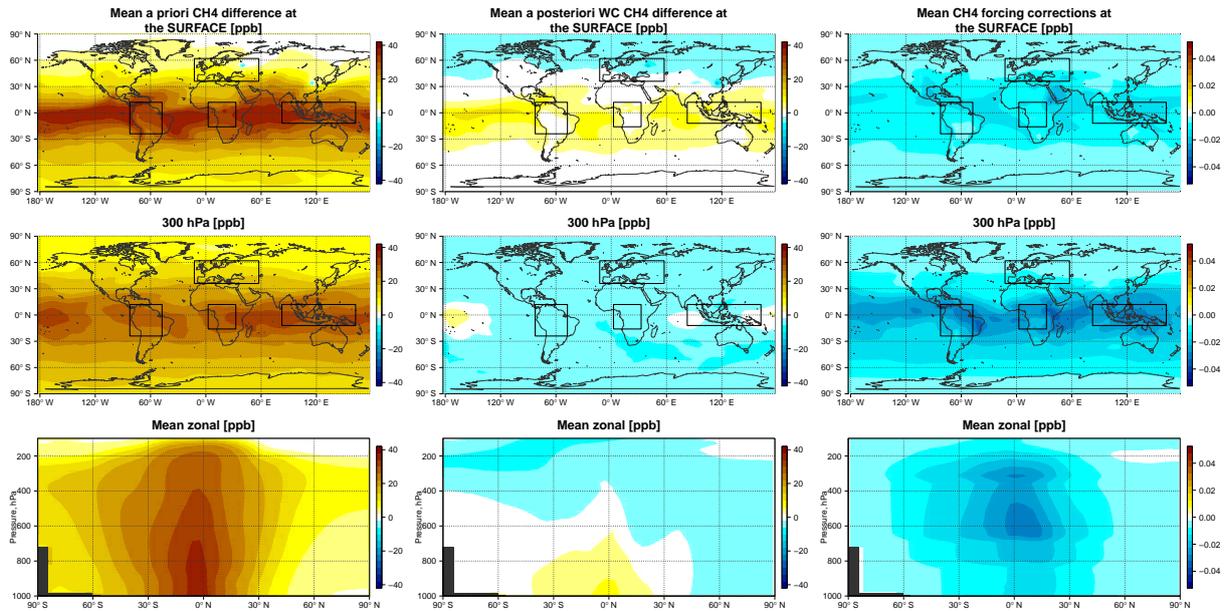
Figure 3. Mean differences in the CH$_4$ distribution in March 2010 in the OSSE with biased chemistry. Left column: mean differences between the a priori CH$_4$ state and the "true" CH$_4$ state. Middle column: mean differences between the WC optimized CH$_4$ state and the "true" CH$_4$ state. Right column: the mean WC state corrections (the forcing terms), in units of ppb. Shown are the latitude-longitude differences at (top row) the surface and (second row) at 300 hPa, as well as the altitude-longitude differences (third row) along the equator. The black boxes indicate the four domains considered for the regional analysis discussed in the text and shown in Fig. 4.

41
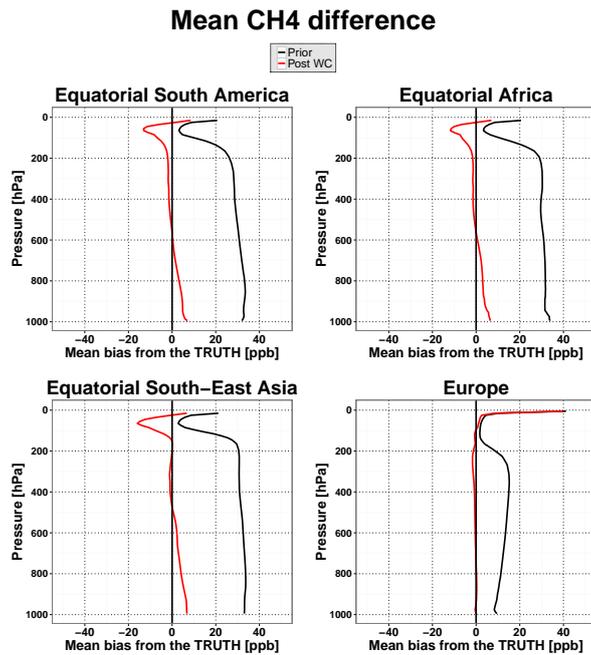
Figure 4. Mean vertical profiles of the CH$_4$ differences in March 2010 in the OSSE with biased chemistry for the four regions depicted in Fig. 3. The differences are between (black lines) the a priori and the "true" CH$_4$ state and between (red lines) the WC optimized state and the "true" CH$_4$ state.
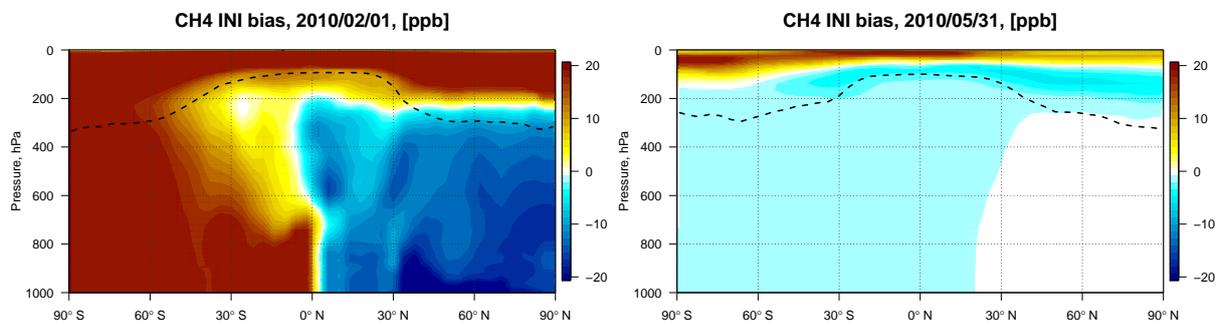
Figure 5. Results of the OSSE with biased initial conditions. Left: a priori bias in initial conditions. Right: a posteriori bias at the end of the assimilation window. The dashed line represents the mean tropopause height on 31 May 2010 taken from GEOS-5 meteorological fields.
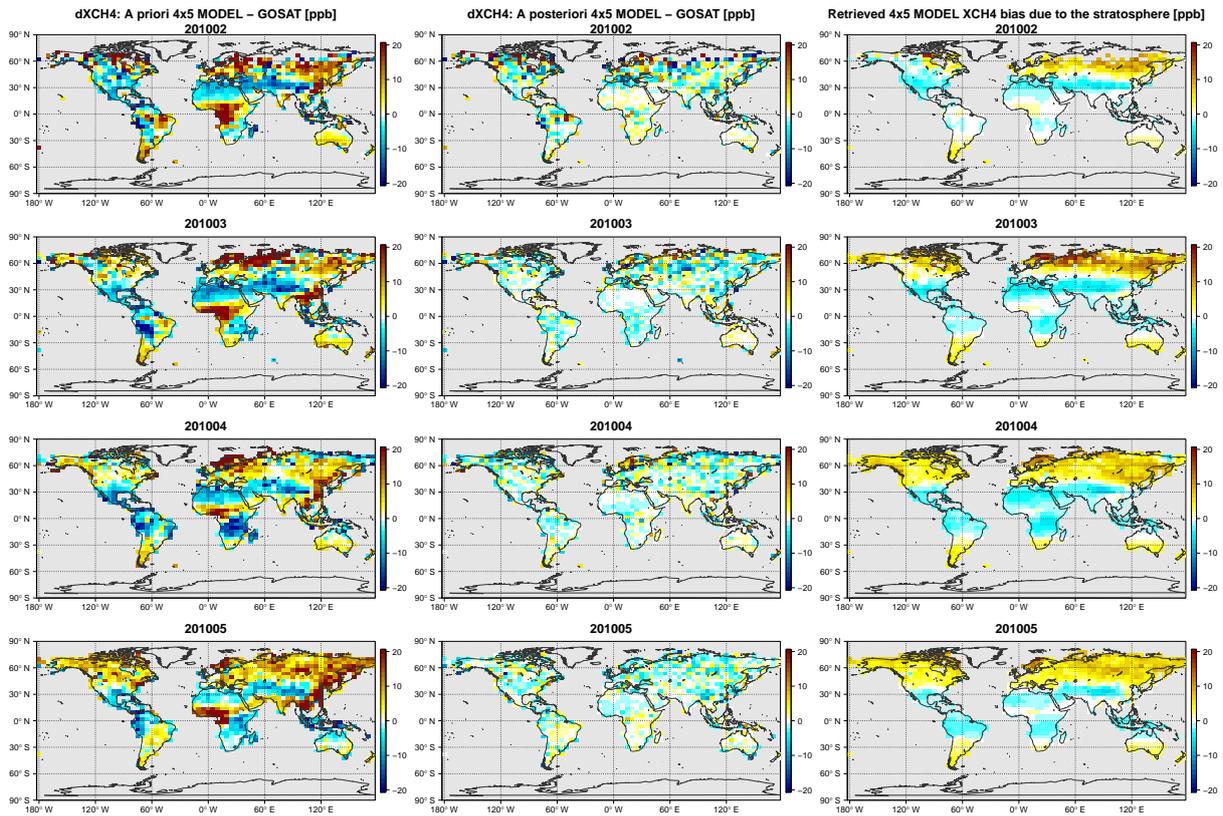
Figure 6. Monthly mean fields from the $4° \times 5°$ resolution model for February - May 2010. First column: differences between the GEOS-Chem a priori XCH_4 state and the GOSAT data. Middle column: differences between the a posteriori **WC_4x5** XCH_4 state and the GOSAT data. Right column: the optimized stratospheric XCH_4 bias correction, calculated as the difference between the model simulation with optimized forcing corrections everywhere and the model simulation with the forcing corrections estimated only in the troposphere. The rows represent results for (top row) February, (second row) March, (third row) April, and (bottom row) May 2010. All model simulations were sampled at the locations and times of the GOSAT observations and smoothed with the GOSAT averaging kernels.

~~Differences in the atmospheric CH_4 mass between the WC optimized state and the "true" CH_4 state, as a function of time, for the OSSE with biased initial conditions. The CH_4 mass was calculated over the eight different regions of the atmosphere as given in the legend. "SH" and "NH" indicate the southern hemisphere and northern hemisphere, respectively.~~
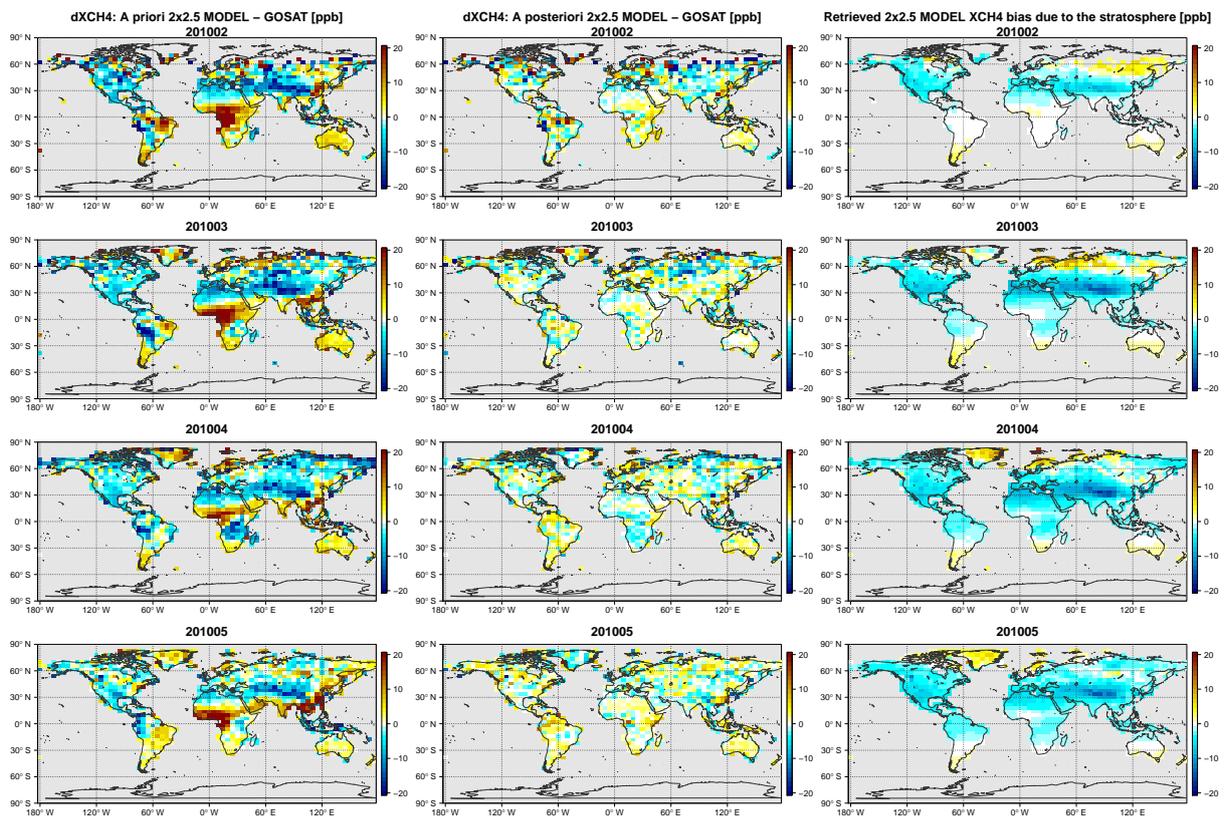
Figure 7. Same as Fig. 6 but for the $2° \times 2.5°$ resolution model.
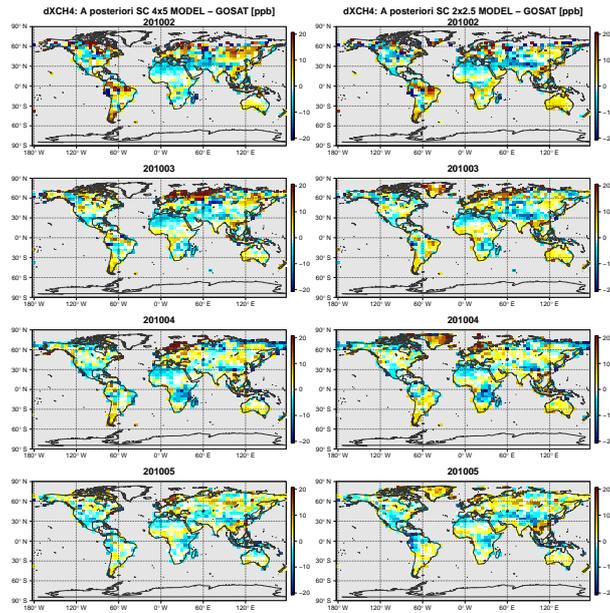
Figure 8. Monthly mean differences between the GEOS-Chem a posteriori XCH$_4$ state and GOSAT: Left column: differences between the a posteriori state from the SC "flux assimilation" at $4° \times 5°$ (**SC_4x5**) and GOSAT. Right column: differences between the a posteriori state from the SC "flux assimilation" at $2° \times 2.5°$ (**SC_2x25**) and GOSAT. The rows represent results for (top row) February, (second row) March, (third row) April, and (bottom row) May 2010.
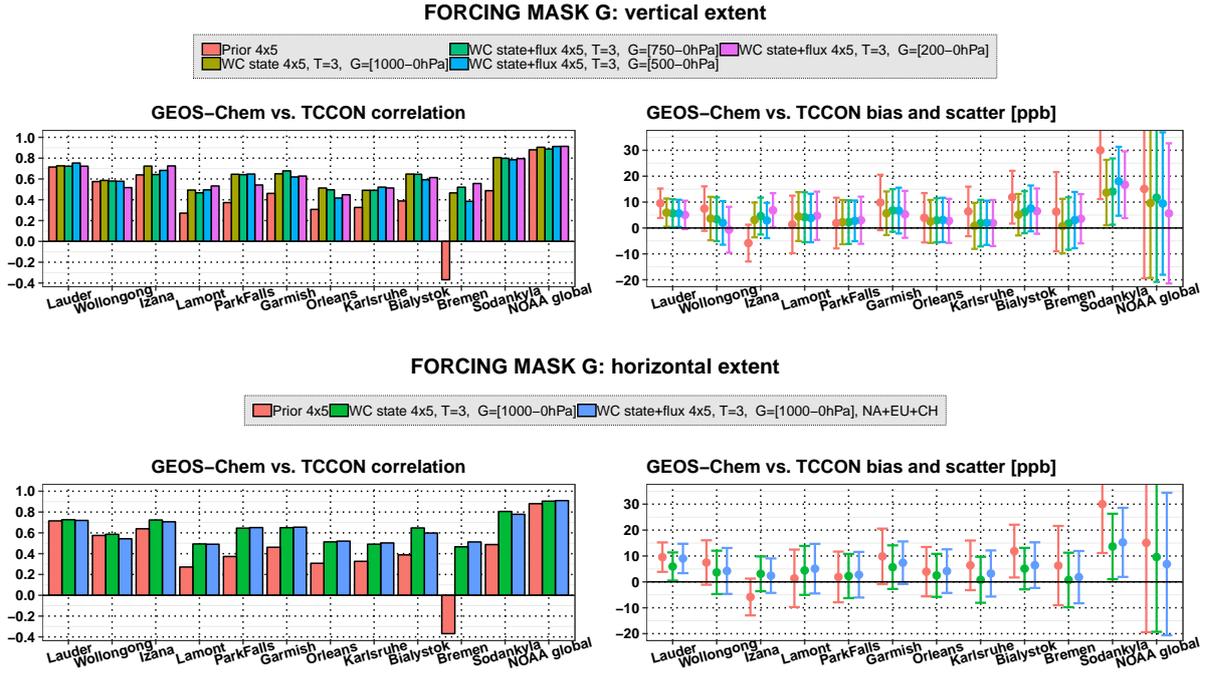
Figure 9. Evaluation of the mean (February-May 2010) a priori and optimized CH$_4$ fields using TCCON XCH$_4$ and NOAA surface in situ observation. Results are shown for ~~a set~~ the four sets of ~~inversions organized~~ experiments described in ~~separate characteristic groups for better representation~~Section 2.5. ~~The~~ For each set of experiment (each row), the left column shows the correlation with respect to the TCCON and NOAA data, whereas the right column shows the mean bias and scatter. Top row: comparison of (red) the a priori fields, (light green) the standard WC assimilation~~(with the forcing terms estimated throughout the atmosphere and with $T = 3$ days)~~, ~~(dark green)~~ and the WC assimilation with the forcing terms estimated at altitudes above (dark green) 750 hPa, (blue) ~~the WC assimilation with the forcing terms estimated above~~ 500 hPa, and (purple) ~~the WC assimilation with the forcing terms estimated above~~ 200 hPa. Second row: comparison of (red) the a priori fields, (dark green) the standard WC assimilation~~(with the forcing terms estimated throughout the atmosphere~~, and (blue) the WC assimilation with joint estimation of the state and surface emissions ~~, and (blue) the WC assimilation~~ with ~~the~~ forcing terms estimated only over North America, Europe, and Asia. Third row: comparison of (red) the a priori fields, (light green) the standard WC assimilation, ~~(dark green)~~ and the WC assimilation with a constant forcing window of (dark green) 7 days, (blue) ~~the WC assimilation with a constant forcing window of~~ 14 days, and (purple) ~~the WC assimilation with a constant forcing window of~~ 30 days. Bottom row: comparison of ~~(red)~~ the a priori fields ~~, at (red)~~ $4° \times 5°$ ~~and~~ (light ~~green~~blue) $2° \times 2.5°$, the ~~standard~~ SC assimilation ~~, at (dark~~ light green) ~~the standard WC assimilation, 4° × 5°~~ and (~~light~~ dark blue) ~~the a priori fields at~~ $2° \times 2.5°$, ~~(dark blue)~~ and the ~~SC~~ WC assimilation at ~~2~~(dark green) $4° \times 2.5$5°, and (purple) ~~the WC assimilation at~~ $2° \times 2.5°$. ~~Each plot shows the statistics for the comparison with each of the 11 TCCON sites considered as well as the global mean statistics for the comparison with the NOAA surface data.~~
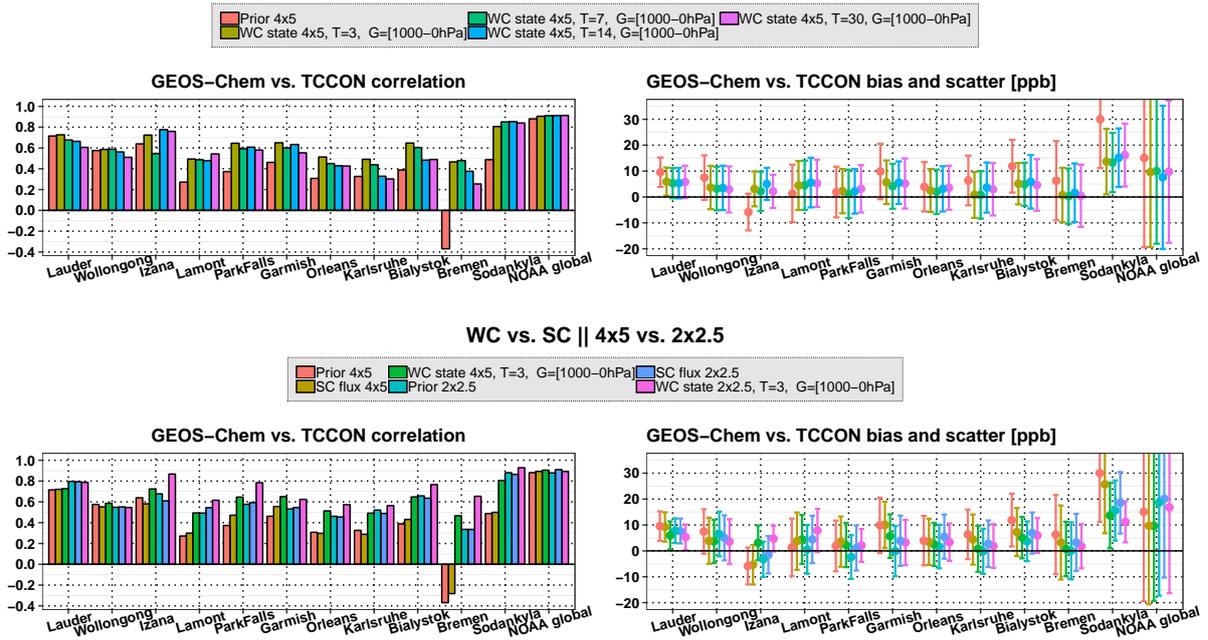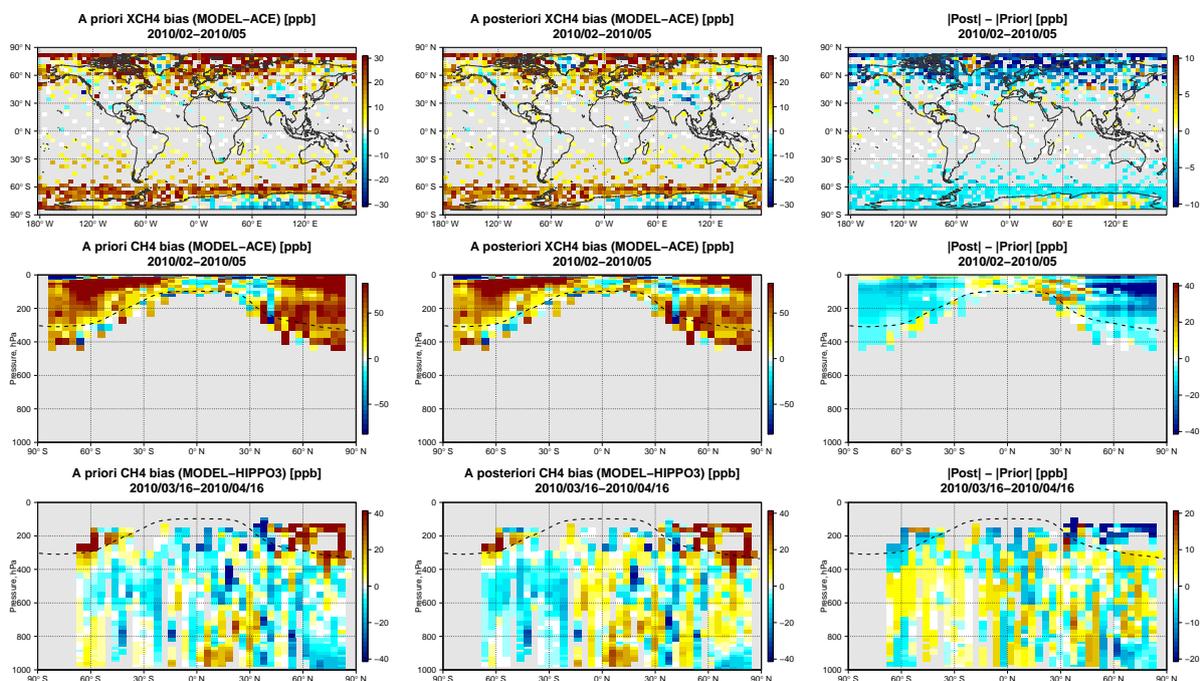
47

Figure 9. Continued.

Figure 10. Evaluation of the mean (February-May 2010) a priori and **WC_4x5** optimized CH_4 fields using ACE-FTS and HIPPO-3 CH_4 measurements. Shown is (left column) the a priori bias, (middle column) the a posteriori bias, and (right column) the reduction in absolute bias. Top row: XCH_4 bias between GEOS-Chem and ACE-FTS. Middle row: zonally averaged CH_4 bias between GEOS-Chem and ACE-FTS. Bottom row: CH_4 bias between GEOS-Chem and HIPPO-3. We used ACE-FTS retrievals only in the stratosphere. The XCH_4 bias between GEOS-Chem and ACE-FTS was obtained by augmenting the ACE-FTS profile in the stratosphere with the GEOS-Chem profile in the troposphere and smoothing the vertical CH_4 profile with mean meridional GOSAT averaging kernels. The dashed line represents the mean tropopause height.
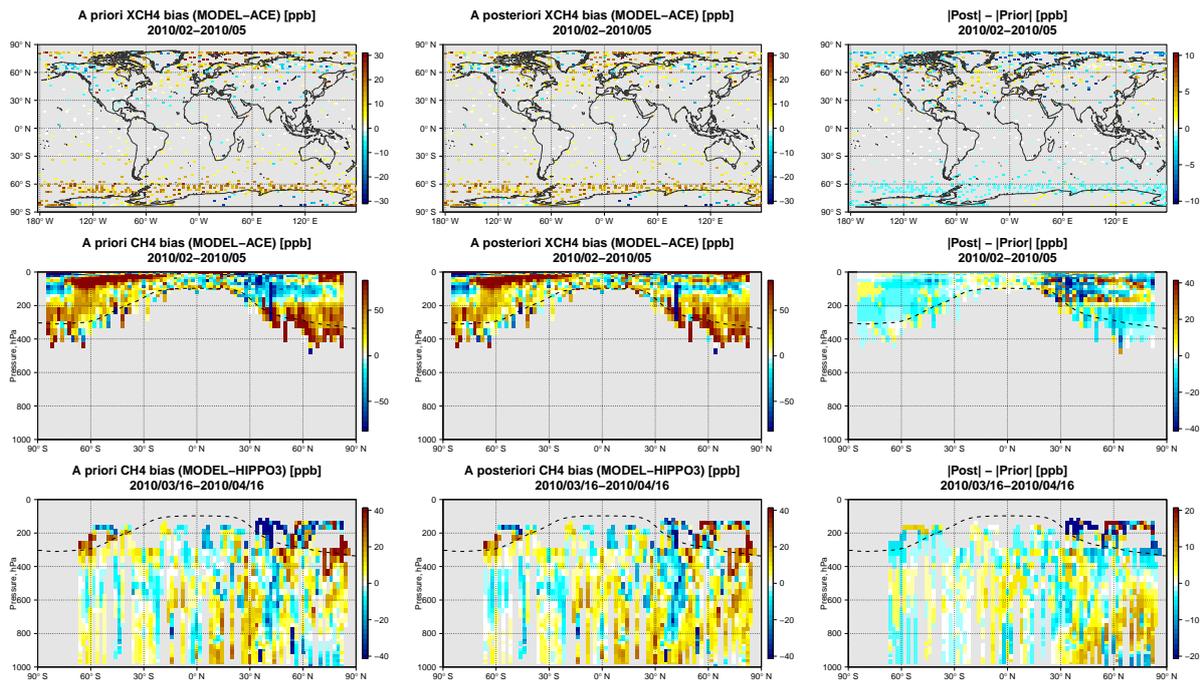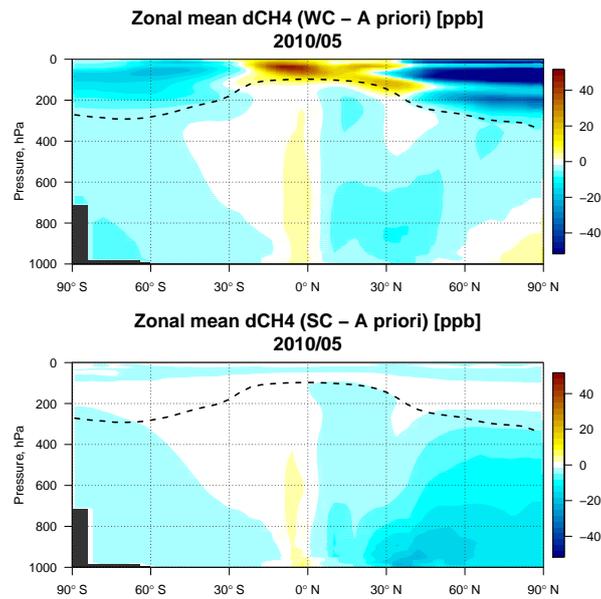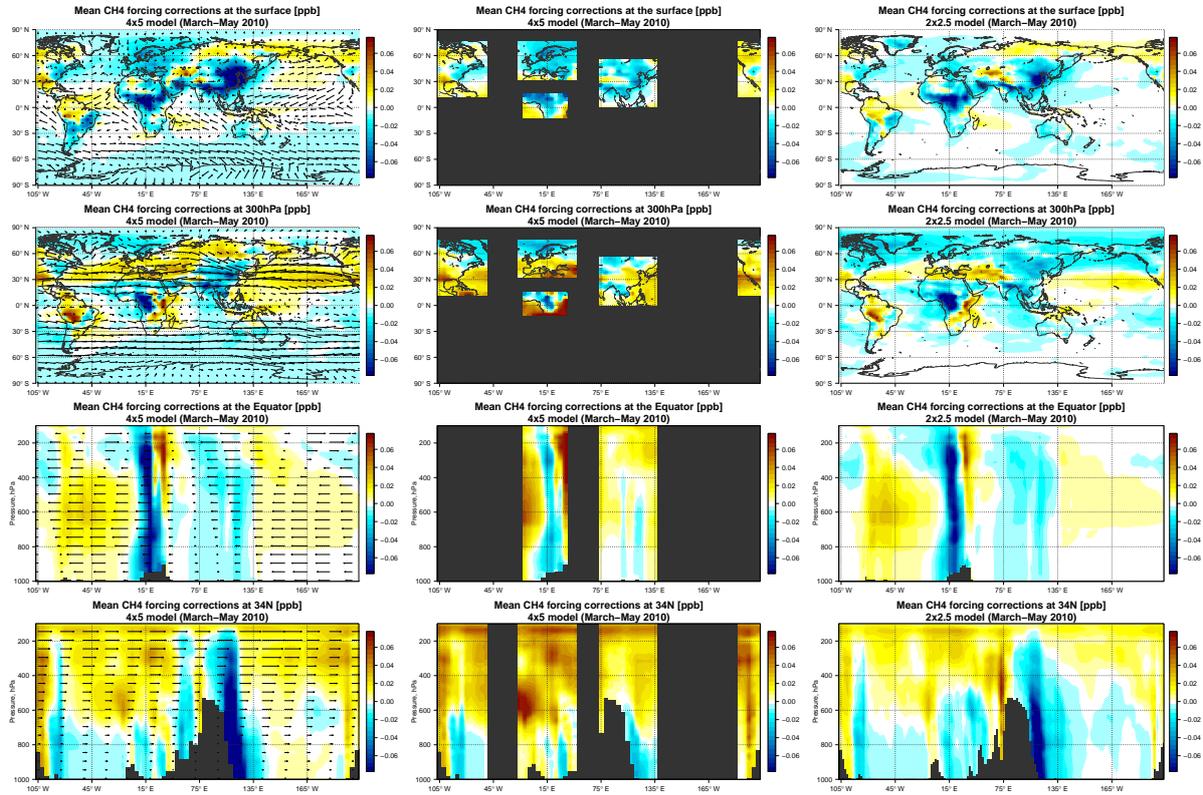
Figure 11. Same as Fig. 10 but for the $2° \times 2.5°$ resolution model.

Figure 12. Zonal mean CH$_4$ differences (in ppb) in May 2010 between (top panel) the **WC_4x5** optimized state and the a priori fields and (bottom panel) between the **SC_4x5** optimized state and the a priori fields. The dashed line represents the mean monthly tropopause height.

Figure 13. Mean optimized forcing terms (in ppb) for March-May 2010. Left column: **WC_4x5** assimilation at $4° \times 5°$. Middle column: **WC_4REG_4x5** assimilation at $4° \times 5°$. Right column: **WC_2x25** inversion at $2° \times 2.5°$. Top row: forcing terms at the surface. Second row: forcing terms at 300 hPa. Third row: altitude-longitude distribution of the forcing terms along the equator. Bottom row: altitude-longitude distribution of the forcing terms along $34°N$. In the plots in the left column, arrows represent the direction and relative magnitude of horizontal winds.