

Interactive comment on “Comparing the CarbonTracker and TM5-4DVar data assimilation systems for CO₂ surface flux inversions” by A. Babenhauserheide et al.

A. Babenhauserheide et al.

arne.babenhauserheide@kit.edu

Received and published: 13 July 2015

1 Anonymous Referee #1

We thank referee #1 for the focussed review which concentrates on getting a more refined understanding of the differences between the two approaches to surface CO₂ flux inversion.

p8887 Line 15. Exactly speaking, neither TM5-4DVar nor Carbontracker took part in the intercomparison by Gurney et al, 2004. It was TM3 model at that time.

C4834

The study by Gurney et al. 2004 does not compare Carbontracker and TM5-4DVar. The statement was intended as example for studies estimating the transport model part of the uncertainty for a paragraph which we decided to leave out to keep the paper shorter. We moved it into a note about uncertainty due to transport models.

p8893 Line 11 Author’s statement “the 4DVar method leaves the dimension of the state vector intact and instead approaches the minimum of the cost function step-by-step” is not accurate, as they use Lanczos method that implements truncated singular value decomposition with limited number of singular vectors, thus reducing a dimension of the problem. The state vector dimension is reduced to number of reconstructed singular vectors. The statement should be reformulated accordingly.

Thank you for catching this inaccuracy! We adjusted the statement to “Approximates the solution using a limited set of search directions, corresponding to the dominant singular vectors of the inverse problem”.

p8893 Line 14 Equivalence between conjugate gradient algorithm and the Lanczos method is not trivial. Fisher and Courtier (1995) who worked on the code used in Meirink et al (2008) had to come up with a proof of the equivalence. Adding reference to Fisher and Courtier (1995) or similar text would help

We added the reference.

p8894 Line 19 Length scale choice of 200km for biosphere fluxes was referred to as standard setting, but it is different from one used in other studies with the same model. Pandey et al 2015 used 1000 km, and Basu et al 2013 used 500 km. If there is a reason to use 200 km it is worth mentioning. The choice of using relatively short distance is related to the conclusion which states a potential benefit of reusing in TM5-4DVar the correlation structures found in Carbontracker. That suggests indirectly that the selected flux correlation length in TM5-4Dvar may be too short.

C4835

The length scale in the default setup of TM5-4DVar was 200km when we did the comparison. This is shorter than the setup used in Basu et al. (2013) and at the time was the best known setup of TM5-4Dvar. We compared our setup to runs with correlation lengths of 500km for the biosphere and 3000km for the ocean, as used by Basu et al. (2013), and found no significant impact on the results.

p8886 line 25. Add “and” between ESRL and Oak Ridge

Fixed.

p8890 line 19 Expression in Eq. (12) was introduced without explaining the notation. Reader may guess the expression in brackets is actually a matrix of size E by dimension of x, but it is better to explain.

We added “with each of the vectors $(\Delta \vec{x}_{b,t}^1, \Delta \vec{x}_{b,t}^2, \dots, \Delta \vec{x}_{b,t}^E)$ defining the deviations from the mean state.”

Thank you again for your review.

2 Anonymous Referee #2

We thank Referee #2 for the detailed review!

The referee analyzes the study in depth and correctly asserts that using systems which are in operational use limits the questions which can be answered. The challenge is to make the analysis largely independent of differences in external input parameters provided to the data assimilation (DA) systems while maintaining a behavior representative of the actual operational use.

The authors recognize that it is necessary to harmonize the inputs to both systems in order to interpret differences in the resultant flux estimates and assess the relative strengths and weaknesses of the two DA approaches. Sensitivity

C4836

tests related to the assimilation window length for CT and the observation coverage are carried out to further evaluate the response of the two DA approaches to these parameters. The manuscript is limited in its scope, however. The final conclusions do not add any new knowledge about: (a) the performance of ensemble or variational systems for carbon flux estimation at high resolutions, (b) why the carbon community should (or should not) prefer a particular system, especially with the recent availability of remote-sensing data, and (c) the global/continental carbon budget, associated uncertainty reduction from the two DA approaches, and more importantly, which approach actually provides more accurate estimates? These are the types of questions, for example, that the community is interested in.

We decided to use published and well-established inversion systems, building on the study by Chatterjee and Michalak (2013) which presented a synthetic comparison of methods similar to those used in Carbontracker and TM5-4DVar.

We do not provide unambiguous answers to the questions from the referee, in particular since using real measurements (instead of synthetic data as in previous assessments) faces practical constraints such as the true state to be estimated being generally unknown. However, our manuscript approaches these questions given the practical constraints of real-world DA systems.

Regarding new knowledge at high resolution (a), our study provides estimates of the minimum uncertainty due to differences between the methods for different levels of observational density.

Fig. 13 shows that obtaining a robust result on the scale of Transcom regions requires a density of observations similar to the in-situ ground network in the USA in 2010. This lower limit to the uncertainty is consistent with the uncertainties reported by Carbontracker North America and underlines the danger of interpreting differences of lower magnitude than these estimates.

C4837

Further, the case resampling of non-assimilated sites shows that with an increased assimilation window for Carbontracker the uncertainty due to limited validation data is larger than the overall difference between the models. With the current in-situ network in obspack (ground stations and aircraft measurements) the overall quality of the methods is therefore indistinguishable. The impact of the selection of validation sites is larger than the difference between both models. So, evaluating (real-world) performance on even finer scales is limited by the availability of independent validation data.

However we identified specific characteristics in which the two methods differ significantly – namely the estimated source in South America and the concentration in Antarctica. Some of these can be mitigated by adjusting model parameters, like the assimilation window length in Carbontracker where we could show that a doubled assimilation window improves the overall match to non-assimilated observations.

Regarding remote sensing data, comment (b), we indeed do not use or assess the usefulness of remote sensing data. However our analyses (section 5.2, Fig. 13) suggest that data density is key to improving model agreement. Therefore enhanced data density expected from remote sensing tools is expected to have a large impact on flux estimates, provided that remote sensing achieves adequate accuracy. We leave the assessment of remote sensing data to future studies.

Related to comment (c), the carbon budget, we show that both models provide similarly accurate fluxes, validated by the match of modelled concentrations to non-assimilated sites, with consistent results on the global scale but lesser agreement on scales of the transcom regions.

Global scale fluxes are consistent within the uncertainty estimated by TM5-4DVar throughout the different comparisons of methods and assimilated sites, and also consistent with the 2013B estimates from CarbonTracker North America run by NOAA, ESRL ¹. CT2013B sees a global sink of 6.79 ± 6.86 PgC for 2009 while we see values

¹The 2013B release of the estimated fluxes of Carbontracker North America (NOAA, ESRL) is available from

C4838

between 6.37 and 7.03 PgC for April 2009 to April 2010, depending on the observation data we assimilate and the model we use. One of the key goals of designing this study was to know whether one system performed significantly better than the other system in global inversions. We found however that they yield consistent posterior fluxes with a consistent match to non-assimilated measurements. As such the accuracy of the fluxes cannot drive the choice of the model when using the current in-situ measurement network from obspack. A choice between the two systems may therefore boil down to practical considerations, such as (a) Carbontracker is easily parallelisable because of the ensemble structure, but (b) TM5-4DVar yields defined uncertainties over long time flux integrals which have to be approximated in Carbontracker, or (c) TM5-4DVar requires an adjoint of the transport model, Carbontracker does not, etc.

This result was indeed missing from the manuscript. To ensure that its visibility matches its relevance for the scientific community, we added it to the conclusions.

The other main contribution to the field is an estimate of the uncertainty due to differences in the inverse method and the structure of the optimized flux state, which includes the setup of the **B** matrix. The question from the referee shows that this was unclear. To improve on this, we added the following to the conclusions:

“Differences between the a posteriori fluxes provide a lower estimate of the uncertainty due to the choice of the optimization method.”

This lower estimate of the uncertainty complements the results from earlier studies on the spread of results due to differences in the transport model and provides missing estimates of uncertainties due to the characteristics of the method, as described in the introduction (observational constraints, flux representation and inverse method).

Also we found a high sensitivity of the fluxes estimated by TM5-4DVar in South America to the measurements in Arembepe, Brazil, along with compensation fluxes in the

esrl.noaa.gov/gmd/ccgg/carbontracker/CT2013B/fluxtimeseries.php?region=Global

C4839

oceans. This sensitivity persisted with longer temporal and spatial correlation length. It creates far reaching effects within strongly underconstrained regions. These effects show that regionally limited sets of measurements which deviate from the mean concentration in the region have stronger non-local effects in TM5-4DVar than in CarbonTracker.

This result is also relevant to comment (b), because measurements with such regional structure are seen in the continuously changing coverage from remote sensing instruments.

2.1 Major Comments

1. Specification of background error covariance (B) matrix – By the authors' own admission (Pg. 8896, Line 24), it is not possible to get an exact match of the flux uncertainties. This statement is unclear - it is imperative to clarify that this maybe because the authors chose to implement "out-of-the-box" versions of the two approaches. From a data assimilation standpoint, one can and should specify the same initial B matrix (i.e., same spatial and temporal correlation length, same uncertainties) for both the ensemble and the variational system. Once the structure of the B matrix is prescribed, I agree that it may not be a trivial task to revise it to specific lat/long grids (for TM5-4DVar application) or aggregate to broad-scale ecoregion/vegetation types and then generate the ensemble members (for CarbonTracker application). However, the background error covariance plays a critical role in filtering and spatially spreading the information from the observations. Discrepancies between the structure and setup of this matrix impacts interpretation of the differences between the flux estimates. For example, a potential reason for the South American flux anomaly (Section 5.1.1) may be due to the misspecification of prior flux uncertainties in case of TM5-4DVar, which results in a high weight being given to the set of observations from ABP. The

C4840

authors acknowledge this in an indirect way by highlighting the outlier-detection framework of CT; but again the fundamental basis for that outlier-detection criterion is related to the spread in the ensemble, and thereby the background error covariance matrix. To resolve this issue, the authors should consider either of the following – (a) (ideal scenario) attempt to specify the same initial background error covariance for both systems and do a sensitivity analysis (for one year) to evaluate the impact on the flux estimates, or (b) (practical scenario) if it is realistically not possible to modify the initial B, then clearly state that as a drawback of the way the systems were implemented and make an argument as to how differences in B may manifest in the differences seen in the flux estimation results. In its current form, the study completely overlooks the role that the B matrix plays even though it is one of the important inputs that should have been harmonized for such an inter-comparison study.

The referee describes the importance of the a priori covariance matrix for the attribution of measured differences to flux regions. To minimize this effect, the study employs a harmonization procedure which avoids the unspecified temporal aggregation of uncertainties from CarbonTracker. As written in p8896, Lines 25–27, we use the flux uncertainties of a CarbonTracker run with a monthly instead of a weekly cycle and a TM5-4DVar run to harmonize the a priori covariance matrices of the models. Due to the different specification of fluxes, only global flux uncertainties can be matched exactly, while regional uncertainties have to be approximate, though as Fig. Si illustrates they are very close except for months with very small prior fluxes.

To check whether this can be the cause of the South America mismatch, we need to examine the prior flux time series of the two transcom regions in South America.

The timeseries in Fig. Si shows that there is a mismatch in the flux uncertainty definition, but this mismatch occurs in April and May 2009 and in Winter 2009/2010, while as figure 9 shows, the flux difference in TM5-4DVar from assimilating the site in Arembepe extends from April to August. So the mismatch of the prior flux uncertainty is

C4841

unlikely to be the source of the difference in the estimated fluxes. April and May are the months, however, where the outlier rejection of Carbontracker makes a difference in the estimated weekly fluxes shown in figure 9. To ensure that readers are aware of this, we added the following to the discussion of the South America flux anomaly in section 5.1.1:

“Carbontracker specifies the flux uncertainty relative to the total flux, which in April and May 2009 yields a lower uncertainty than that from TM5-4DVar, which causes smaller changes to the flux, leading to the strong reaction of the outlier rejection. But as shown in Fig. 9, Carbontracker does not show the additional source seen in TM5-4DVar between July and August 2009, where the flux uncertainty of both models differs by less than 10%“

To make the statement clearer, which says that it is not possible to get an exact match of the flux uncertainty due to the different ways of specifying the state vector, we added “(weekly with ecoregions)” and “(monthly gridded with global covariance parameters)” to the respective mentions of Carbontracker and TM5-4DVar in the statement (Pg. 8896, Line 24).

Also we added a caveat to the conclusions:

“A caveat applies since prior fluxes and prior flux uncertainties cannot be made identical due to differences in how the state vectors of the two methods are setup: Carbontracker optimizes weekly ecosystem-wide fluxes while TM5-4DVar optimizes monthly fluxes on a regular longitude-latitude grid.”

The referee says that the outlier rejection framework highlights the impact of these discrepancies. However, enabling or disabling this framework does not have a significant effect on the estimated yearly fluxes. The effect of the prior flux uncertainty on the outlier rejection in Carbontracker is very indirect: the outlier rejection uses the prescribed representativeness errors of observations, not the estimated uncertainty of the modelled concentration fields to decide whether to reject a measurement. The harmonized

C4842

prior flux uncertainty can have an effect on the outlier rejection, because the extent of the flux adjustment due to other, non-rejected measurements is subject to the prior flux uncertainty.

We agree however that it is essential to ensure that readers are aware of the limitations due to using established optimization methods for estimating surface fluxes which represent the state in different ways. Therefore we followed suggestion (b) of the reviewer and expanded the final paragraph in section 3.2 to explicitly state that having small mismatches in the covariance matrix is unavoidable when comparing real systems with different representation of the fluxes and that this can cause some differences in the flux attribution which have to be taken into account when interpreting the results.

Also we added this discussion of the effects of the harmonization of the prior flux covariance matrix and plots of the monthly prior flux covariance to the supplement.

2. Motivation of the study and novelty – This study compares carbon fluxes estimated from two different DA systems (CarbonTracker and TM5-4DVar) at aggregated scales. The authors need to make a better case for motivating why such a study is necessary and what new information it provides in terms of improving our understanding of the applicability of DA systems for carbon flux estimation purposes. By reporting the comparison of the posterior flux estimates (but not the associated uncertainties) at aggregated spatial and temporal scales, it is difficult to judge the performance of the two systems at finer spatiotemporal scales, for e.g., biomes/ecoregions, Transcom-scales. Expectedly at aggregated scales both the DA approaches provide similar estimates, and it is not clear what new knowledge, if any, the carbon science community stands to gain from this study.

The authors claim that their primary goal is to evaluate the impact of the inverse method on the “accuracy of the estimated fluxes” (Pg. 8887, Line 17-18). Without any comparison of posterior uncertainties, however, it is difficult to back this statement. Given that the authors have used the in situ network, there may be

C4843

value in comparing these flux estimates to existing studies from the literature. Have the authors compared their global/continental estimates to other studies over the same time period? Alternatively, the authors can report results at the Transcom3 regions, which may highlight additional regional differences between the estimates from the two DA systems. These are a few possible additional analyses/tests that will add value to the manuscript and make it scientifically relevant to the carbon science community.

We avoid comparing the uncertainty reduction of the two systems on temporally aggregated scales, because such a comparison would require assumptions on the temporal structure of the a posteriori covariance in Carbontracker. As such, we do not analyze the accuracy of the fluxes, so our statement was inexact. To ensure that readers do not stumble over the choice of words, we removed the word “accuracy” from the introduction. We are comparing the effect of the choice of the method on the fluxes and are using the result to estimate a lower limit on the accuracy due to effects from the choice of the assimilation method and the flux representation which are not captured by the individual methods. In addition we show the uncertainties at the native timescales of the methods where they require no temporal aggregation (weekly for Carbontracker and monthly for TM5-4DVar) in figure 7, 9 and 11.

Also the referee notes that “it is difficult to judge the performance of the systems at finer spatiotemporal scales” due to reporting aggregated values. In the baseline study we show the regions where we observe significant divergence of the models – namely South America, Asia and the Indian Ocean – and trace these discrepancies to their causes: The high sensitivity of TM5-4DVar to the measurements in Arembepe, Brazil and a combination of the large temporal binsize of TM5-4DVar and the limited assimilation window of Carbontracker. We compare the estimated fluxes to the fluxes reported by Carbontracker North America CT2013B (NOAA, ESRL) and find agreement within the flux uncertainty reported there. And while the uncertainties reported in CT2013B on continental scales may seem large, the differences we see between Carbontracker

C4844

and TM5-4DVar justify uncertainties of that magnitude.

To ease later comparison we now include the estimated annual flux per transcom region in the supplement – with the exception of region 5 and 6 (Africa) which are combined, because their definition differs in TM5-4DVar and Carbontracker.

An important finding is that the quality of the concentration field from both models is similar, as estimated by comparing with non-assimilated measurements (the only independent validation available). As such differences between the estimated fluxes provide a lower limit for the accuracy which can be reached with the current ground based in-situ network. The analysis of the effects of increasing observational coverage in North America shows that estimating fluxes on the scale of Transcom Regions is already feasible for North America, while e.g. an assimilation in Europe needs to assimilate additional observations to allow for robust finer scale inversions.

Therefore we actually do believe that this study provides a significant contribution to the understanding of the carbon cycle.

1. Pg. 8884, Lines 13-14 – In the abstract the authors claim that one of the sensitivity tests will include impact of operational parameters “such as temporal and spatial correlation lengths”. However no sensitivity tests are presented in the manuscript to justify this statement. Revise.

We adjusted the statement to name the operational parameter we show: The assimilation time window. For the study we also tested different temporal and spatial correlation lengths in TM5-4DVar (harmonized to similar effective prior uncertainty), but these did not have a significant effect on the aggregated fluxes at the scale of transcom regions.

2. Pg. 8885, Line 16- Differences in assumptions about error covariances (both model-data mismatch and prior) contribute significantly to the differences in flux estimates from different studies. This point needs to be acknowledged here.

The harmonization resulted in an increase of the prior ocean flux uncertainty in TM5-

C4845

4DVar by 10% compared to the values used before and a reduction of the land flux uncertainty by 10%. For each gridbox that results in a prior land flux uncertainty of 199.17% of the prior flux and a prior ocean flux uncertainty of 172.59% of the prior flux.

We added these numbers in section 3.2 and the following in the introduction: “Differences in these characteristics contribute to the differences in flux estimates from different studies. To analyze the impact from the representation of sources and sinks and from the inverse method, it is therefore necessary to harmonize the observational constraints, the transport model and the prior concentration, flux and flux covariance estimates between the approaches which are compared.”

3. Pg. 8887, Lines 15-16 – This statement and the associated references need to be revised. For e.g., TM5-4DVar nor CarbonTracker as used in this study took part in the Transcom experiments (Gurney et al. [2004] had TM3 though). Similarly, TM5-4DVar as used in this study wasn't part of the suite of atmospheric inversions used in the Schulze et al. [2009] study. Kindly check the use of references here, and throughout the manuscript.

Gurley et al. (2004) was misplaced – thank you for catching that! Schulze et al. (2009) however used results from a pre-publication of Peters et al. (2009) which used CarbonTracker.

We clarified the text to make it clear that the distinction we draw is between studies based mostly on a single model and studies which use multiple models. The word “both” implied that each of the studies used both models. We replaced “both models” by “the models”.

4. Pg. 8893, Lines 14-15 – For all purposes, the appropriate reference here should be Fisher and Courtier [1995] for showing the feasibility of eigenvector based approximation methods (see old.ecmwf.int/publications/library/ecpublications/_pdf/tm/001-300/tm220.pdf).

C4846

For implementation purposes, the appropriate references are Meirink et al. [2008] and Chevallier et al. [2005].

We adjusted the references.

5. Pg. 8903, Lines 24-25 – It is not clear what the authors mean by “..approaches uncertainties from above, ...”. Clarify.

We clarified the uncertainty aggregation by adding “the aggregated errors are larger than the analytical uncertainties at the exact minimum of the cost function”.

6. Pg. 8906, Line 1-2 – This statement is unnecessary for this portion. Delete.

This was not needed here, yes. Removed.

7. Pg. 8908, Lines 19 – The word ‘adjustment’ is misspelled.

Typo fixed.

8. Pg. 8908, Lines 22-23- This statement is unclear. Do the authors mean to say that potential flux adjustment for ecoregions can be constrained well by a single site? And hence CT does better than TM5-4DVar? Overall the discussion in this paragraph was difficult to follow.

Due to binning the fluxes in ecoregions, the flux adjustment is based on a larger number of sites, because ecoregions have a larger latitudinal spread than grid boxes. We added an example to make it easier to come back to the practical effects of the ecoregion approach: “For example adjusting the flux in the corn belt yields concentration changes all over North America (downwind of the corn belt ecoregion).”

However such a long latitudinal correlation length would be unrealistic for latitudinally constrained regions like the tropical forests in central Africa and many smaller ecoregions in Europe. A grid with a uniform spatial correlation length has to strike a balance between both kinds of regions.

C4847

9. Pg. 8909, Line 7 – Replace “observations” by “observational network”.

Adjusted.

10. Figures 7, 9, 11 – For the benefit of the reader, it would be better to stick to a single flux unit (such as PgC/region/year) throughout the manuscript. Note that this will require edits throughout the text as well.

We prefer to use two flux units for two reasons: The first reason is that when we tried using PgC/a for these weekly or monthly fluxes multiple people mixed up temporally aggregated and non-aggregated fluxes, in particular since putting units of PgC/a on weekly fluxes results in big numbers. The graphs seemed easy to understand, but led to misinterpretations. The second reason is to underline that the fluxes in these graphs have not been aggregated temporally, so their uncertainty can be compared.

11. Figures 7 and 9 –The CT simulations are represented in yellow lines with bars on both ends. Do these bars represent the posterior uncertainty? How are these posterior uncertainties calculated for the CT simulations? If these bars do not represent the uncertainty estimates, then I would suggest using a different symbol/marker to avoid confusion with the other figures.

These bars represent the uncertainty calculated from the posterior spread of the ensemble.

We added the following in the caption of Figure 7 and referenced it in Figure 9: “The uncertainties shown for Carbontracker are aggregated spatially but not temporally. As such they represent the uncertainty of the estimated fluxes, calculated directly from the ensemble. These uncertainties are excluded from the annually aggregated graphs, because there is no method for temporally aggregating the uncertainties in a way which is comparable to the uncertainties estimated by TM5-4DVar.”

Thank you again for your very detailed review.

C4848

3 Changes to the manuscript

For changes to the manuscript see the supplementary material of this comment.

References

- Basu, S., Guerlet, S., Butz, A., Houweling, S., Hasekamp, O., Aben, I., Krummel, P., Steele, P., Langenfelds, R., Torn, M., Biraud, S., Stephens, B., Andrews, A., and Worthy, D.: Global CO₂ fluxes estimated from GOSAT retrievals of total column CO₂, *Atmospheric Chemistry and Physics*, 13, 8695–8717, doi:10.5194/acp-13-8695-2013, <http://www.atmos-chem-phys.net/13/8695/2013/>, 2013.
- Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Pak, B. C., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, P., Fung, I. Y., Heimann, M., John, J., Maki, T., Maksyutov, S., Peylin, P., Prather, M., and Taguchi, S.: Transcom 3 inversion intercomparison: Model mean results for the estimation of seasonal carbon sources and sinks, *Global Biogeochemical Cycles*, 18, 1010, doi:10.1029/2003GB002111, 2004.
- Peters, W., Krol, M. C., van der Werf, G. R., Houweling, S., Jones, C. D., Hughes, J., Schaefer, K., Masarie, K., Jacobson, A., Miller, J. B., Cho, C. H., Ramonet, M., Schmidt, M., Ciattaglia, L., Apadula, F., Heltai, D., Meinhardt, F., Di Sarra, A. G., Piacentino, S., Sferlazzo, D., Aalto, T., Hatakka, J., Ström, J., Haszpra, L., Meijer, H. A. J., Jordan, A., Rodó, X., Morgui, J.-A., Vermeulen, A. T., Rozanski, K., Zimnoch, M., Manning, A., Leuenberger, M., Uglietti, C., Dolman, A. J., Ciais, P., Heimann, M., and Tans, P. P.: Seven years of recent European net terrestrial carbon dioxide exchange constrained by atmospheric observations, *Glob. Change Bio.*, accepted for publication, doi:10.1111/i.1365-2486.2009.02078.x, 2009.

Please also note the supplement to this comment:

<http://www.atmos-chem-phys-discuss.net/15/C4834/2015/acpd-15-C4834-2015-supplement.pdf>

Interactive comment on *Atmos. Chem. Phys. Discuss.*, 15, 8883, 2015.

C4849

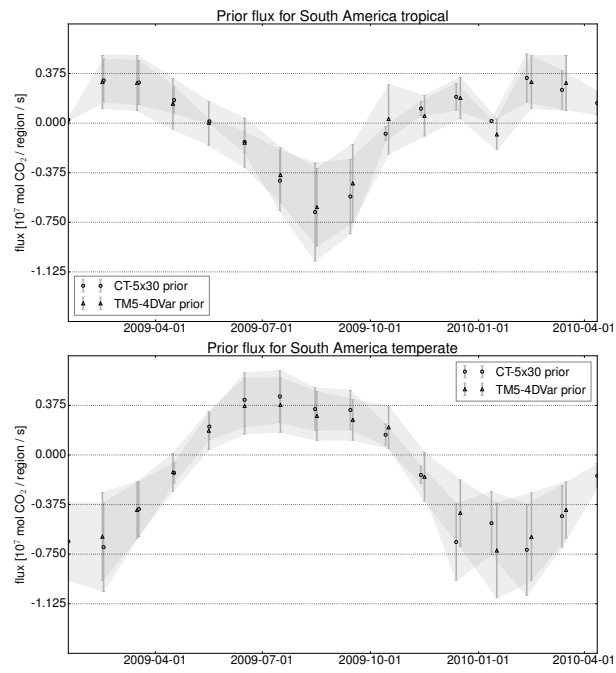


Fig. Si.