Atmos. Chem. Phys. Discuss., 15, C1–C17, 2015
www.atmos-chem-phys-discuss.net/15/C1/2015/

Atmospheric Chemistry and Physics Discussions

Open Access

ACPD

15, C1–C17, 2015

Interactive Comment

Interactive comment on "Comparing the CarbonTracker and TM5-4DVar data assimilation systems for CO_2 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 CO_2 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.

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.

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 $\left(\Delta\vec{x}^{1}_{\mathrm{b},t}, \Delta\vec{x}^{2}_{\mathrm{b},t}, \ldots, \Delta\vec{x}^{E}_{\mathrm{b},t}\right)$ 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

ACPD

15, C1–C17, 2015

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.

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

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

ACPD
15, C1–C17, 2015

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 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

ACPD
15, C1–C17, 2015

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

ACPD
15, C1–C17, 2015

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 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

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 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-

ACPD
15, C1–C17, 2015

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).

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.

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.

ACPD
15, C1–C17, 2015

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 CO_2 fluxes estimated from GOSAT retrievals of total column CO_2, 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/C1/2015/acpd-15-C1-2015-supplement.pdf

---

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

ACPD

15, C1–C17, 2015

Interactive Comment

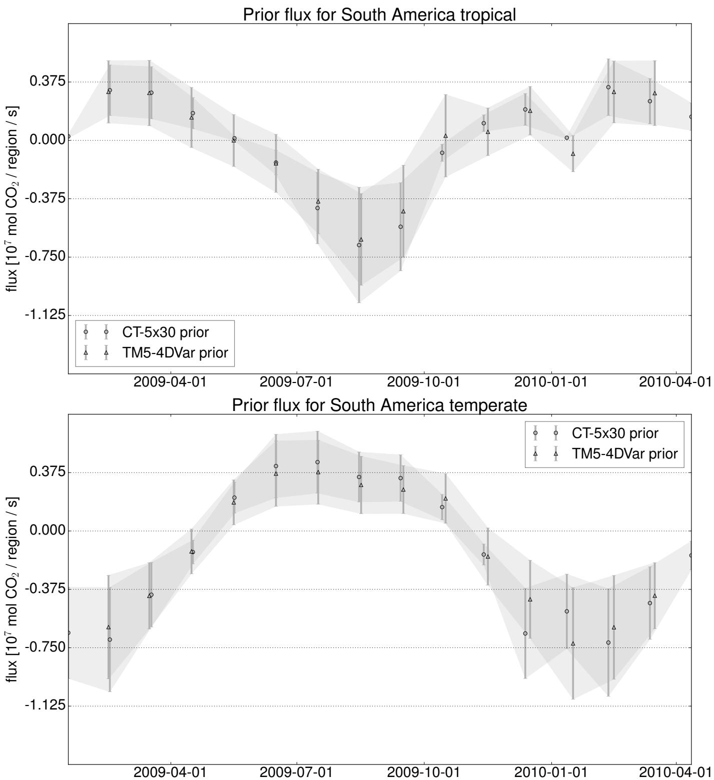

Fig. Si.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

Manuscript prepared for Atmos. Chem. Phys. Discuss.
with version 2014/09/16 7.15 Copernicus papers of the LATEX class copernicus.cls.
Date: 13 July 2015

Comparing the CarbonTracker and TM5-4DVar data assimilation systems for CO_2 surface flux inversions

A. Babenhauserheide[1], S. Basu[2], S. Houweling[3], W. Peters[4,5], and A. Butz[1]

[1]IMK-ASF, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
[3]Netherlands Institute of Space Research (SRON), Utrecht, the Netherlands
[4]Dept. of Meteorology and Air Quality, Environmental Sciences Group Wageningen University, Wageningen, the Netherlands
[5]Univ. of Groningen, Centre for Isotope Research, Groningen, the Netherlands
[2]Earth System Research Laboratory, Global Monitoring Division, National Oceanic and Atmospheric Administration, Boulder, Colorado, USA

Correspondence to: A. Babenhauserheide (arne.babenhauserheide@kit.edu)

Abstract

Data assimilation systems allow for estimating surface fluxes of greenhouse gases from atmospheric concentration measurements. Good knowledge about fluxes is essential to understand how climate change affects ecosystems and to characterize feedback mechanisms. Based on assimilation of more than one year of atmospheric in-situ concentration measurements, we compare the performance of two established data assimilation models, CarbonTracker and TM5-4DVar, for CO_2 flux estimation. CarbonTracker uses an Ensemble Kalman Filter method to optimize fluxes on ecoregions. TM5-4DVar employs a 4-D variational method and optimizes fluxes on a $6^\circ \times 4^\circ$ longitude/latitude grid. Harmonizing the input data allows analyzing the strengths and weaknesses of the two approaches by direct comparison of the modelled concentrations and the estimated fluxes. We further assess the sensitivity of the two approaches to the density of observations and operational parameters such as ~~temporal and spatial correlation lengths~~the length of the assimilation time window.

Our results show that both models provide optimized CO_2 concentration fields of similar quality. In Antarctica CarbonTracker underestimates the wintertime CO_2 concentrations, since its 5-week assimilation window does not allow for adjusting the far-away surface fluxes in response to the detected concentration mismatch. Flux estimates by CarbonTracker and TM5-4DVar are consistent and robust for regions with good observation coverage, regions with low observation coverage reveal significant differences. In South America, the fluxes estimated by TM5-4DVar suffer from limited representativeness of the few observations. For the North American continent, mimicking the historical increase of measurement network density shows improving agreement between CarbonTracker and TM5-4DVar flux estimates for increasing observation density.

1 Introduction

Sources and sinks of atmospheric carbon dioxide (CO_2) largely control future climate change (Schimel, 2007). Anthropogenic emissions release roughly 10 Gt carbon into the atmosphere per year (Peters et al., 2013), part of which gets taken up by the biosphere and the oceans. The fraction of emitted CO_2 which remains in the atmosphere is the largest driver of climate change (Stocker et al., 2013, chapter 8.5.1), but the distribution and strength of carbon sources and sinks on the surface is hard to measure directly. Methods for observing the fluxes directly require either eddy covariance measurements at multiple height levels (Foken et al., 2012) or measurements of concentration changes in a sealed volume of air. But such bottom-up approaches are only representative for a given collection of vegetation types in a limited geographic area.

Inverse modelling therefore uses CO_2 concentration gradients observed in the Earth's atmosphere to quantify the spatio-temporal distribution of the net CO_2 surface fluxes (e.g. Enting, 2000; Peters et al., 2007; Chevallier et al., 2010; Feng et al., 2011; Peylin et al., 2013). To this end, various data assimilation (DA) techniques have been developed. These DA approaches differ in four main characteristics: first, they ingest different observational constraints, for example in-situ concentration measurements at different sites. Second, they represent sources and sinks of carbon differently, for example by binning them by by vegetation type or on a latitude/longitude grid. Third, they relate sources and sinks to observed atmospheric abundances using different air-mass transport models (Gurney et al., 2004, estimate their impact on fluxes). And fourth, they use different inverse methods that find the best estimate of the source-sink distribution using the transport model, the observational constraints, the representation of sources and sinks and a prior estimate of the sources and sinks. 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.

There are two main classes of assimilation techniques for complex inversions, variational methods and ensemble methods (Lahoz et al., 2007; Lahoz and Schneider, 2014). Both approaches are approximate variants of the general Bayesian optimal estimation scheme (e.g. Rodgers, 2000) which aims at balancing prior or background information with actual measurement information to derive robust parameter estimates. Approximations are necessary to render the inverse problem computationally feasible since real-world CO_2 surface flux inversions typically involve thousands of concentration measurements and millions of unknown flux parameters. Both schemes can either treat the entire considered assimilation period at once or divide it into shorter periods to be treated sequentially. Ensemble methods approximate the exact solution from an ensemble of model runs, while variational methods approach the optimal solution step-by-step (e.g. Juhász and Bölöni, 2007; Gilbert and Lemaréchal, 1989).

The performance of ensemble methods and variational methods has been evaluated previously for numerical weather prediction (e.g. Kalnay, 2005; Fairbairn et al., 2013) and direct optimization of atmospheric gas abundances (Skachko et al., 2014). Chatterjee and Michalak (2013) are the first to evaluate the performance of the two methods for the purpose of CO_2 surface flux estimation. They use a synthetic setup with simulated observations and a 1-dimensional transport model which has the advantage of knowing the true fluxes and for which a direct Bayesian inversion is computationally feasible. In particular they find that under constraints on model runtime and resource use, the estimated surface fluxes are more realistic with their variational implementation than with their ensemble method, and that for both models small-scale fluxes (flux aggregation spanning up to 5 % of the model size) are very sensitive to the data coverage and distribution.

Here, we focus on evaluating the performance of an ensemble method and a variational method used for real atmospheric CO_2 flux inversion problems. We focus on a case study for the period from 2009 to 2010, and use observational constraints collected by an in-situ measurement network and compiled by the NOAA Environmental Sciences Division and Oak Ridge National Laboratory (2013, exact version: obspack PROTOTYPE v1.0.2 2013-01-28). Our ensemble method is the Ensemble Square Root Filter (EnSRF, Whitaker and Hamill, 2002) as employed by the CarbonTracker modelling system (Peters et al., 2007), a variant of the Ensemble Kalman Filter. The variational method is the TM5-4DVar package described by Meirink et al. (2008) and Basu et al. (2013).

~~Beside~~ Besides the mathematical treatment of the inversion, CarbonTracker and TM5-4DVar differ in the design of the state vector. CarbonTracker optimizes fluxes binned by regions with similar vegetation – like cropland or boreal forest – and separated by geographic regions following the transcom basemap (Gurney et al., 2000). TM5-4DVar adjusts the fluxes on a grid ($6° \times 4°$ longitude $\times$ latitude) with correlations which decay exponentially in time and space.

Both ~~models~~ methods are used in a number of studies. CarbonTracker studies include estimates of global CO_2 fluxes (Peters et al., 2007, 2010), European fluxes (Meesters et al., 2012), Asian fluxes (Zhang et al., 2014) as well as ^{13}C isotope studies (van der Velde et al., 2014). Studies with TM5-4DVar include CO_2 flux estimation (Basu et al., 2013), CO estimation (Hooghiemstra et al., 2011) and CH_4 emission estimates (Meirink et al., 2008; Bergamaschi et al., 2010; Houweling et al., 2014). Additionally ~~both~~ the models were employed in several multi-model comparison studies (e.g. Schulze et al., 2009; Peylin et al., 2013; Thompson et al., 2014).

Our goal is to evaluate the impact of the inverse method (including the flux representation) on the ~~accuracy of the~~ estimated surface fluxes. Therefore, we must make sure that the other components of the DA systems – the observations to be assimilated, the transport model and the prior assumptions – are the same. After a short summary of the general CarbonTracker and TM5-4DVar methodology in Sect. 2, Sect. 3 describes how we harmonize these other components of the two DA systems, mostly focusing on the observation input and the prior assumptions, since CarbonTracker and TM5-4DVar both operate on the same transport model, the tracer model 5 (TM5, Krol et al., 2005). In Sect. 4 we compare the performance of the two inverse methods by evaluating the mismatch between modelled and measured concentration fields. The comparison to assimilated observations verifies that the schemes work as expected. The comparison to non-assimilated observations yields an estimate of how the DA systems succeed in modelling CO_2 concentration fields in regions where the ~~models~~ methods do not assimilate observations. Building on these results, Sect. 5 analyzes the estimated surface fluxes and tests their sensitivity to observation density.

2 Inverse methods and ~~model~~ setup

The DA systems aim at inferring a state vector $\boldsymbol{x}$ that contains spatially and temporally binned surface fluxes or a related quantity such as scaling factors for an initial guess flux field. To this end, the systems exploit measurements of the atmospheric concentration chained into an observation vector $\boldsymbol{y}$. Fluxes and measured concentrations are linked through the transport and observation operator **H** which is linear for the case of our CO_2 flux inversions, but in general could be non-linear such as for CH_4 flux inversions. Typically, the inverse problem of estimating $\boldsymbol{x}$ from a set of observations $\boldsymbol{y}$ is ill posed. Due to sparse observational coverage, measurement errors or measurement configuration, the observations contain insufficient information to determine all components of $\boldsymbol{x}$ independently. A background flux estimate $\boldsymbol{x}_\mathrm{b}$ from biosphere and ocean models is used to provide a constraint that fills the null-space where measurement information is insufficient. Accordingly, the state vector of fluxes $\boldsymbol{x}$ is determined by minimizing a cost function J that typically consists of two terms, the mismatch between measured and modelled observations and the mismatch between the fluxes to be estimated and the background estimate,

$$\begin{aligned} J = \ & (\boldsymbol{y}-\mathbf{H}\boldsymbol{x})^T\mathbf{R}^{-1}(\boldsymbol{y}-\mathbf{H}\boldsymbol{x}) + \\ & + (\boldsymbol{x}-\boldsymbol{x}_\mathrm{b})^T\mathbf{B}^{-1}(\boldsymbol{x}-\boldsymbol{x}_\mathrm{b}) \end{aligned} \tag{1}$$

with **R** the observation covariance and **B** the background flux covariance. **R** and **B** define the relative weights of the measurement and background mismatch.

In general, minimization of Eq. (1) can be solved by means of matrix algebra (Rodgers, 2000) yielding optimized fluxes and their error covariances,

$$
\begin{aligned}
\hat{\boldsymbol{x}} &= \boldsymbol{x}_\mathrm{b} + \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}(\boldsymbol{y} - \mathbf{H}\boldsymbol{x}_\mathrm{b}) && (2)\\
&= \boldsymbol{x}_\mathrm{b} + (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{B}^{-1})^{-1}\mathbf{H}^T\mathbf{R}^{-1}(\boldsymbol{y} - \mathbf{H}\boldsymbol{x}_\mathrm{b}), && (3)\\
\hat{\mathbf{B}} &= \mathbf{B} - \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{B} && (4)\\
&= (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{B}^{-1})^{-1}, && (5)
\end{aligned}
$$

with $\hat{\boldsymbol{x}}$ the a posteriori state vector and $\hat{\mathbf{B}}$ as the respective covariance matrix. Equivalence of equation pairs – Eqs. (2) and (3), Eqs. (4) and (5) – can be shown (Rodgers, 2000, Eqs. 4.11 and 2.27).

While theoretically the minimization of Eq. (1) reduces to a matrix inversion for linear systems like CO_2 flux inversion (e.g. Rodgers, 2000), the large number of parameters to be estimated and the amount of measurements to be ingested requires approximate methods such as ~~EnSRF~~ EnSRF and 4DVar which are numerically efficient.

2.1 CarbonTracker: ~~EnSRF~~ EnSRF based data assimilation

CarbonTracker is an inverse modelling framework based on the Ensemble Square Root Filter (~~EnSRF~~EnSRF) developed by Peters et al. (2005). Instead of solving the minimization problem in one step, the ~~EnSRF~~ EnSRF determines optimized surface fluxes sequentially in a time stepping approach with $\boldsymbol{x}_t$ defining a subset of $\boldsymbol{x}$ for a certain time window. In our standard setup $\boldsymbol{x}$ contains scaling factors for the surface fluxes for 96 weeks, while $\boldsymbol{x}_t$ only spans 5 weeks.

Commonly, a gain matrix $\mathbf{G}$ is defined as

$$\mathbf{G} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \tag{6}$$

$$= (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{B}^{-1})^{-1}\mathbf{H}^T\mathbf{R}^{-1}. \tag{7}$$

Equations (2) and (4) then read

$$\hat{\boldsymbol{x}}_t = \boldsymbol{x}_{\mathrm{b},t} + \mathbf{G}_t\left(\boldsymbol{y}_t - \mathbf{H}_t\boldsymbol{x}_{\mathrm{b},t}\right), \tag{8}$$

$$\hat{\mathbf{B}}_t = \mathbf{B}_t - \mathbf{G}_t\mathbf{H}_t\mathbf{B}_t \tag{9}$$

with the Gain Matrix

$$\mathbf{G}_t = \mathbf{B}_t\mathbf{H}_t^T\left(\mathbf{H}_t\mathbf{B}_t\mathbf{H}_t^T + \mathbf{R}_t\right)^{-1} \tag{10}$$

where subscript t indicates quantities of reduced dimensions, for the time step under investigation. Once Eqs. (8) and (9) are solved for time slice t, the solution of the scaling factors $\hat{\boldsymbol{x}}_t$ is used as the background estimate $\boldsymbol{x}_{\mathrm{b},t+1}$ for the next time slice $t+1$, assuming that a simple persistence forecast is adequate for our CO_2 flux inversion problem,

$$\boldsymbol{x}_{\mathrm{b},t+1} = \hat{\boldsymbol{x}}_t. \tag{11}$$

The covariance $\mathbf{B}_{t+1}$ is prescribed at each time step as described in Peters et al. (2005). Given an initial guess for the first background state, this strategy allows for sequentially calculating the complete state vector $\hat{\boldsymbol{x}}$.

To estimate the gain matrix $\mathbf{G}_t$, the ~~EnSRF~~ EnSRF uses an ensemble approach. The ensemble members $\boldsymbol{x}^i_{\mathrm{b},t} = \boldsymbol{x}_{\mathrm{b},t} + \Delta\boldsymbol{x}^i_{\mathrm{b},t}$ ($i = 1 \ldots E$, with E the ensemble size) of the background state are drawn such that their mean and covariance is consistent with the background state $\boldsymbol{x}_{\mathrm{b},t}$ and background covariance matrix $\mathbf{B}_t$, respectively, so that

$$\mathbf{B}_t \approx \frac{1}{E-1}\left(\Delta\boldsymbol{x}^1_{\mathrm{b},t}, \Delta\boldsymbol{x}^2_{\mathrm{b},t}, \ldots, \Delta\boldsymbol{x}^E_{\mathrm{b},t}\right) \cdot \left(\Delta\boldsymbol{x}^1_{\mathrm{b},t}, \Delta\boldsymbol{x}^2_{\mathrm{b},t}, \ldots, \Delta\boldsymbol{x}^E_{\mathrm{b},t}\right)^T \tag{12}$$

with each of the vectors $\left(\Delta\boldsymbol{x}^1_{\mathrm{b},t}, \Delta\boldsymbol{x}^2_{\mathrm{b},t}, \ldots, \Delta\boldsymbol{x}^E_{\mathrm{b},t}\right)$ defining the deviations from the mean state.

Then, the terms $\mathbf{H}_t\mathbf{B}_t\mathbf{H}_t^T$ and $\mathbf{B}_t\mathbf{H}_t^T$ required for calculating $\mathbf{G}_t$ following Eq. (10) can be approximated using the results from an ensemble run of the possibly non-linearized transport model $\mathcal{H}$

$$\mathbf{H}_t\mathbf{B}_t\mathbf{H}_t^T \approx \frac{1}{E-1}\left(\mathcal{H}_t\Delta\boldsymbol{x}^1_{\mathrm{b},t}, \mathcal{H}_t\Delta\boldsymbol{x}^2_{\mathrm{b},t}, \ldots, \mathcal{H}_t\Delta\boldsymbol{x}^E_{\mathrm{b},t}\right) \cdot \left(\mathcal{H}_t\Delta\boldsymbol{x}^1_{\mathrm{b},t}, \mathcal{H}_t\Delta\boldsymbol{x}^2_{\mathrm{b},t}, \ldots, \mathcal{H}_t\Delta\boldsymbol{x}^E_{\mathrm{b},t}\right)^T \tag{13}$$

$$\mathbf{B}_t\mathbf{H}_t^T \approx \frac{1}{E-1}\left(\Delta\boldsymbol{x}^1_{\mathrm{b},t}, \Delta\boldsymbol{x}^2_{\mathrm{b},t}, \ldots, \Delta\boldsymbol{x}^E_{\mathrm{b},t}\right) \cdot \left(\mathcal{H}_t\Delta\boldsymbol{x}^1_{\mathrm{b},t}, \mathcal{H}_t\Delta\boldsymbol{x}^2_{\mathrm{b},t}, \ldots, \mathcal{H}_t\Delta\boldsymbol{x}^2_{\mathrm{b},t}\right)^T, \tag{14}$$

where the approximation becomes more exact with increasing ensemble size E. The ~~EnSRF~~ EnSRF method yields robust results with non-linear transport operators $\mathcal{H}$ as long as the transport model is close to linear for small perturbations ($\mathcal{H}(\boldsymbol{x}+\Delta\boldsymbol{x}) \approx \mathbf{H}\boldsymbol{x}+\mathbf{H}\Delta\boldsymbol{x}$). Using Eqs. (13) and (14), the gain matrix $\mathbf{G}_t$ can be calculated from Eq. (10), finally to update the state estimate $\hat{\boldsymbol{x}}_t$ via Eq. (8). Peters et al. (2005) describe in detail how to estimate the state covariance $\hat{\mathbf{B}}_t$ by separately updating the ensemble deviations $\Delta\boldsymbol{x}^i_{\mathrm{b},t}$ while avoiding the costly evaluation of Eq. (10) and circumventing spurious underestimation of $\hat{\mathbf{B}}_t$. Overall, CarbonTracker's ~~EnSRF~~ EnSRF approach requires running the transport model $\mathcal{H}$ for E ensemble members over the time period covered by all time steps t. At each time step t the transport model is sampled at all measurement instances within the time step and the above methodology is followed.

CarbonTracker uses a refined approach for stepping through the entire time period considered. CarbonTracker's state vector $\boldsymbol{x}_t$ is subdivided into five one-week bins (five cycles) resulting in an assimilation window of five weeks (Peters et al., 2005, chapter 2.3). At each optimization step the oldest cycle at the "end" of the state vector drops out of the state vector and is used as a posteriori flux estimate while a new cycle is added to the "beginning" of the state vector according to Eq. (11). As such, each one-week cycle experiences a number of optimization steps equal to the number of weeks in the assimilation time window. The choice of assimilation time window, here five weeks, also implies that CarbonTracker can adjust surface fluxes only when their effects are observed at a site within five weeks of atmospheric transport. In the zonal direction, this limitation is of little consequence, because typical global transport times scales are on the order of weeks. But in meridional direction and especially for interhemispheric transport where the transport timescales are on the order of months, this choice needs to be taken into account when interpreting flux results. The time stepping also defines the temporal binning of one-week fluxes.

The spatial binning of CarbonTracker's state vector follows the transcom regions (Gurney et al., 2000), further categorized into land regions with similar ecosphere following Olson et al. (1992) and ocean regions following the Ocean Inversion Fluxes (Jacobson et al., 2007b) as described in the documentation of CarbonTracker North America[1]. In total, there are 240 flux ecoregions to be optimized, which is signficantly less than the number of grid cells of the transport model operating on $6^\circ \times 4^\circ$ (longitude $\times$ latitude). The fluxes to be optimized are further separated into 3 categories: biosphere/ocean, fire and fossil fuel. Only the category biosphere/ocean is optimized, the others are imposed from their priors following the assumption that fossil fuel fluxes are known with much higher precision than biosphere and ocean fluxes and that fire fluxes cannot easily be distinguished from biosphere fluxes, so they could not be interpreted separately. Altogether, temporal and spatial binning results in a state vector $\boldsymbol{x}_t$ with $240 \times 5 = 1200$ elements.

The structure of the background covariance $\mathbf{B}_t$ in the Northern Hemisphere is a diagonal matrix with a variance of 0.64 (80 % standard deviation) in units of dimensionless flux scaling factors. In tropical and many Southern Hemisphere regions, the ecosystems are coupled with exponentially decreasing covariance, selected such that the total covariance in the transcom region matches the variance in Northern Hemisphere regions. The covari-

[1]CarbonTracker 2011_oi results and documentation are provided by NOAA ESRL, Boulder, Colorado, USA from the website esrl.noaa.gov/gmd/ccgg/carbontracker/CT2011_oi/. The site builds on the work from Peters et al. (2007).

ance for ocean regions uses the results of the ocean inversion by Jacobson et al. (2007a). Temporal covariance in CarbonTracker stems from processing observations multiple times in the timestepping approach. The observation covariance **R** is assumed diagonal.

The version of CarbonTracker used here is derived from version 1.0 of the code maintained by Wageningen University with the same state vector as CarbonTracker North America (as used in Peters et al., 2007) and without a zoom region.

2.2 TM5-4DVar: variational data assimilation

Whereas the ~~EnSRF~~ EnSRF in CarbonTracker reduces the dimension of the minimization problem of Eq. (1) by solving sequentially for time-sliced state vectors, the 4DVar method in TM5-4DVar leaves the dimension of the state vector intact and ~~instead approaches the~~ approximates the solution using a limited set of search directions, corresponding to the dominant singular vectors of the inverse problem to approach the minimum of the cost function step-by-step. The iterative minimization of Eq. (1) in TM5-4DVar is described in detail by Chevallier et al. (2005) and Meirink et al. (2008). It employs the conjugate gradient algorithm (Navon and Legler, 1987) which is equivalent to the Lanczos method (Lanczos, 1950; Fisher and Courtier, 1995) and requires calculation of the cost function gradient

$$\nabla_{\boldsymbol{x}} J = \mathbf{B}^{-1}(\boldsymbol{x}_n - \boldsymbol{x}_{\mathrm{b}}) - \mathbf{H}^T\mathbf{R}^{-1}(\boldsymbol{y} - \mathbf{H}\boldsymbol{x}_n) \tag{15}$$

where subscript n indicates the nth iterative step. The adjoint formulation of TM5 allows calculating the cost function gradient by a single run of the transport model and its adjoint (Errico, 1997; Chevallier et al., 2005). The conjugate gradient algorithm further provides the leading eigenvalues and eigenvectors of the preconditioned Hessian

$$\nabla_{\boldsymbol{\chi}}(\nabla_{\boldsymbol{\chi}} J) = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}, \tag{16}$$

which is the second derivative of the cost function J with respect to the dimensionless preconditioned state $\boldsymbol{\chi}$ defined as $\boldsymbol{x} = \mathbf{L}\boldsymbol{\chi} + \boldsymbol{x}_{\mathrm{b}}$, where **L** is the preconditioning matrix with $\mathbf{B} = \mathbf{L}\mathbf{L}^T$. This can be used to construct the inverse of the state covariance $\hat{\mathbf{B}}^{-1}$ as defined in Eq. (4). After n steps, corresponding to n runs of the forward and the adjoint model, the minimization algorithm yields an optimized state estimate $\hat{\boldsymbol{\chi}}_n$ and the first n eigenvalues λ_i ($\lambda_i > 1$) and eigenvectors $\boldsymbol{v}_i$ $(i = 1, \ldots, n)$ for the eigensystem of the preconditioned Hessian. The latter can be used to construct an approximate error covariance matrix,

$$\hat{\mathbf{B}}_n \approx \mathbf{B} + \sum_{i=1}^{n} \left(\frac{1}{\lambda_i} - 1 \right) (\mathbf{L}\boldsymbol{v}_i)(\mathbf{L}\boldsymbol{v}_i)^T. \tag{17}$$

With increasing number of iterations, the optimized state vector $\hat{\boldsymbol{x}}_n$ approaches the optimal state vector $\hat{\boldsymbol{x}}$ at the minimum of the cost function and the approximate state covariance $\hat{\mathbf{B}}_n$ approaches $\hat{\mathbf{B}}$ from above, so that the estimated uncertainty is always larger than the analytical value(Basu et al., 2013). For practical purposes the iteration is stopped when the gradient norm reduction exceeds a threshold, i.e.

$$|\boldsymbol{\nabla}_{\boldsymbol{x}} J(\boldsymbol{x}_n)| \leq \eta \cdot |\boldsymbol{\nabla}_{\boldsymbol{x}} J(\boldsymbol{x}_0)| \tag{18}$$

with the constant chosen to be $\eta = 10^{-9}$ here.

TM5-4DVar's state vector $\boldsymbol{x}$ is binned temporally in monthly fluxes and spatially on the transport model grid scale, i.e. $6^\circ \times 4^\circ$ longitude $\times$ latitude. Fluxes are categorized into biosphere, ocean, fire and fossil fuel. To create a setup comparable to CarbonTracker, only biosphere and ocean fluxes are optimized. The background covariance **B** of the state vector is characterized by a global temporal and spatial correlation length. By default TM5-4DVar uses an exponential decay with a temporal and spatial length scale of 1 month and 200 km for biosphere fluxes and 3 months and 1000 km for ocean fluxes. As such, the temporal binning of TM5-4DVar's state vector containing monthly bins is about a factor 4 coarser than the temporal binning of CarbonTracker's weekly bins. TM5-4DVar's spatial binning has a different overall structure. Whereas CarbonTracker's prior fluxes are fully correlated inside the 240 ecoregions and mostly uncorrelated between different ecoregions, the correlation of TM5-4DVar's fluxes ~~exponentially falls off~~ falls off exponentially around each grid box. The exponential decay in TM5-4DVar's temporal background correlation limits the effects of observations in time. However, TM5-4DVar has no strict limit on the time window during which observations can be linked to fluxes but rather reduces the strength of the influence with temporal lag. TM5-4DVar can adjust surface fluxes in response to any observation during the entire considered time period given that the transport model reveals a link between fluxes and observations. As for CarbonTracker, the observation covariance **R** is assumed diagonal.

3 Setup of the comparison

Given the setup of the CarbonTracker and TM5-4DVar modelling systems, we aim at comparing the performance of their data assimilation concepts for the purpose of CO_2 surface flux estimation when assimilating atmospheric CO_2 concentration records. To avoid affecting conclusions about the inverse methodology, care must be taken that model input such as transport parameters, background estimates, initial concentration fields and assimilated observations are harmonized as far as possible. However, as outlined in Sect. 2, conceptual differences between the models prevent us from making the model setup literally identical.

3.1 Transport model and observation operator

To connect concentration measurements and surface fluxes, CarbonTracker and TM5-4DVar use a transport model which transports the CO_2 tracer using meteorological fields. Both models use the Tracer Model 5 (TM5) as described by Krol et al. (2005) which utilizes meteorological data from the European Centre for Medium-Range Weather Forecasts (ECMWF, 2013). For CarbonTracker, we follow the setup used by Peters et al. (2007). For TM5-4DVar our setup differs from the setup used by Basu et al. (2013) in one main aspect to be consistent with CarbonTracker: the CO_2 concentration field is sampled in the second model layer ($\approx 980\,\mathrm{hPa} \approx 170\,\mathrm{m}$) or higher instead of in the first model layer ($\approx 994\,\mathrm{hPa} \approx 50\,\mathrm{m}$) or higher. Except for these adjustments and some minor differences due to different interfaces of the inverse methods, the versions of TM5 used by the CarbonTracker and TM5-4DVar systems we are using are the same.

3.2 Background flux and initial guess

CarbonTracker and TM5-4DVar use the same background fluxes and initial concentration fields. The biosphere fluxes are taken from the Simple Biosphere model using the ~~Carnegie-Ames-Standord~~ Carnegie-Ames-Stanford Approach (SIBCASA as by Schaefer et al., 2008). SIBCASA is a carbon cycle model that represents the uptake of CO_2 by different types of vegetation and its subsequent transfer back to the atmosphere through autotrophic and heterotrophic respiration. Its mechanistic description of the processes involved is driven by a combination of high-resolution weather data and satellite remote sensing products and includes interactions between the carbon, water, and energy cycles of the land-surface. For the oceans both models use Ocean Inversion Fluxes (oif), the output from an ocean inversion which assumes that the uptake of antropogenic CO_2 increases proportional to the mismatch between atmospheric and oceanic CO_2 partial pressure. Fire fluxes are taken from the Global Fire Emissions Database version 2 (GFEDv2, van der Werf et al., 2010). Fossil fuel fluxes are taken from the Miller dataset as described in Peters et al. (2007) and its Supplement.

The initial concentration field is generated from the output of a previous CarbonTracker run which ended on 1 January 2007. The field for 2009 is derived by increasing the concentration by 1.9 parts per million (ppm) per year. The value 1.9 ppm was chosen based on tests of the fit to observation sites in the first month of 2009.

The covariance of the fluxes is defined in the models as described in Sects. 2.1 and 2.2. We harmonize the overall covariance by adjusting the prior flux uncertainty in TM5-4DVar to 172.59% of the flux for ocean grid boxes and to 199.17% for land grid boxes to match uncertainty of a CarbonTracker run with a monthly cycle for global and continental aggregates. Due to the different ways of specifying the state vector x ~~in CarbonTracker and TM5-4DVar~~and its covariance **B** in CarbonTracker (weekly with ecoregions) and TM5-4DVar (monthly gridded with global covariance parameters), it is not possible to get an exact match of the flux ~~uncertainies. We harmonize the overall covariance by adjusting the~~ uncertainties. This is a result of comparing real-world systems used for flux estimation to not only capture theoretical effects but also differences which show in practical use. While making the comparison more complex, this choice allows getting a better understanding of the uncertainties due to the large amount of implementation decisions which have to be taken for a production system. The remaining mismatches in the prior flux uncertainty ~~in TM5-4DVar to match uncertainty of a CarbonTracker run with a monthly cycle for global and continental aggregates~~can have an effect on the estimated fluxes. This effect has to be taken into account for interpreting a posteriori flux differences. Section 5.1.1 includes an example of such an analysis. The remaining mismatches in the flux uncertainty per transcom region and month are provided in the supplement.

3.3 Observations and observation errors

Both DA systems use the same observations from the "obspack" compilation of in-situ CO_2 concentration measurements (Masarie et al., 2014; NOAA Environmental Sciences Division and Oak Ridge National Laboratory, 2013, version: PROTOTYPE v1.0.2 2013-01-28). Discrete (e.g. one sample per week) measurements from surface flask sites, in situ continuous (and semi-continuous) measurements from surface sites and towers, and aircraft campaign measurements are collected, aggregated and quality screened to make them suitable for inverse flux estimation. At many but not all of the continuous measurement sites, the measurements are averaged to provide afternoon or nighttime averages (depending on the type of site, e.g., continental planetary boundary layer site or mountain site), using intra-day averaging periods representative of large scale fluxes and discarding single measurements outside the respective averaging periods. For our baseline CarbonTracker and TM5-4DVar runs, we exclude 21 measurement sites from the assimilation to use them as validation sites.

Additionally we take out 5 sites which have more than 1000 measurements in the assimilation period. This is to keep the TM5-4DVar results representative of TM5-4DVar runs which use the native TM5-4DVar input. When using these 5 sites with the CarbonTracker preprocessing, TM5-4DVar shows strong gradients between neighboring grid cells in North America which it does not show when processing its native set of observations. In addition to these 26 excluded sites, there are 24 further sites from which the default run of CarbonTracker uses no data or only a subset of the observations. Reasons for not using some of the observation data of a site include that the data is assumed not representative of its grid-cell or recorded in aircraft campaigns.

Measurement uncertainty is set to a fixed value for each site accounting for the measurement errors and for representativeness errors. The latter originate from using the in-situ samples to represent the CO_2 concentration in a transport model grid box of 6° longitude and 4° latitude. Concentration uncertainties range from 0.75 ppm for marine boundary layer sites over 2.5 ppm for land sites up to 7.5 ppm for sites which experience variable meteorological conditions. Table 1 in the Supplement lists the observation records used in our study. Figure S1 shows the global distribution of observation sites together with a visual representation of their weight due to sampling frequency and representativeness error. In our setup CarbonTracker and TM5-4DVar use the same representativeness errors.

4 A posteriori concentration fields

As a first step, we compare and validate the performance of CarbonTracker and TM5-4DVar by evaluating the difference between measured and modelled CO_2 concentration fields at the location of various ground sampling stations. Comparing concentration fields at the assimilated sites in Sect. 4.1 provides a check to verify that data assimilation works in both systems. Comparing measured and modelled concentrations at non-assimilated sites in Sect. 4.2 demonstrates to what extent the data assimilation approaches yield improvements where observational constraints are distant in space and/or time. CarbonTracker and TM5-4DVar are both run with the baseline setup (as described in Sect. 3) for a 23 month period starting on 1 February 2009.

4.1 Assimilated sites

As an example for an assimilated site, Fig. 1 shows a time series of measured and modelled CO_2 concentrations at Mauna Loa (MLO), Hawaii, located 3399 m a.s.l. in the Pacific.

For the period from 1 February 2009, to 30 December 2010, the models assimilate 94 weekly flask measurements. We compare the observations to a posteriori and a priori model concentrations. The a posteriori concentrations are sampled using the a posteriori surface fluxes estimated by CarbonTracker or TM5-4DVar. The prior model concentrations are sampled using the background (prior) flux estimate common to both models. The Mauna Loa record demonstrates that the a posteriori concentrations produced by both models match the observations within the uncertainty estimate and that the match is substantially better than for the prior concentration fields. Differences between CarbonTracker and TM5-4DVar are much smaller than the representativeness error of the measurements at Mauna Loa (0.75 ppm) over the entire period. This is consistent with the results at other sites.

The mismatch between measured and modelled CO_2 concentrations for all assimilated measurements is shown in Fig. 2, with the prior concentrations, the a posteriori concentrations optimized by CarbonTracker, and the a posteriori concentrations optimized by TM5-4DVar. The concentration mismatch is normalized by the representativeness error of the observations such that a (unitless) mismatch of 1 corresponds to a mismatch with the magnitude of the representativeness error. Unlike the time series for Mauna Loa, the histograms only integrate over the 1 year period 3 April 2009 to 2 April 2010 in order to be consistent with the analysis of the a posteriori surface fluxes in Sect. 5. This time period gives the models sufficient spin-up and spin-down time, given that the initial concentration is already well-optimized by a previous CarbonTracker run.

The concentrations from the Prior Forward Run in Fig. 2 reveal an overall bias in the normalized (unitless) mismatch of 0.37 with a standard deviation of 1.09. Tentatively, the prior fields show a dipole pattern with peaks around -1 and 1 which can be traced back to the Northern Hemisphere prior generally overestimating the observations and the Southern Hemisphere prior generally underestimating the observations. The CarbonTracker and TM5-4DVar histograms show small biases of 0.006 and 0.025 with standard deviation of 0.727 and 0.650, respectively. Compared to the prior, both DA systems improve the overall bias and they substantially reduce the spread of the observation-model mismatch. Normalized standard-deviations smaller than 1 indicate that the mismatch is on average smaller than the estimated representativeness error, which points to a conservative choice of representativeness errors and consequently a stronger than optimal influence of the prior flux estimate. However, avoiding this would require using the output of the assimilation systems to adjust their input parameters which could lead to transient errors in the result.

The histograms for CarbonTracker and TM5-4DVar a posteriori concentrations reveal some non-Gaussian behavior with long tails toward greater mismatch and with a narrow peak at the center. The tails most likely stem from temporally varying contributions to the representativeness error which our input data assumes constant in time. The narrow peak likely stems from two sources: first, sites with high frequency measurements are assumed uncorrelated in the models and as such provide a stronger constraint than sites with low frequency measurements. Second, an already well-optimized prior which is close to the observations causes the models to stick to the prior in a sparse observation network.

In summary, both models show similar performance for assimilated sites, and the assimilation substantially reduces the mismatch between modelled and measured concentrations at assimilated sites.

4.2 Non-assimilated sites

Next, we evaluate the performance of the DA systems for sites whose observations are not assimilated~~by the models~~. These sites provide independent validation of the results. Figure 3 shows a time series of flask measurements in Guam on Mariana Islands (GMI), West Pacific. In contrast to Mauna Loa, the measurements are taken at sea level, and are not assimilated by the CarbonTracker and TM5-4DVar inverse models. The observation error in Guam is 1.5 ppm, and the ~~models~~ modeled concentrations agree well with measurements taken at the site. Both, CarbonTracker and TM5-4DVar, reproduce the measurements similarly well with a respective bias of 0.12 and 0.02 ppm. Their standard deviation of 0.79 and 0.82 ppm, are greater than the standard deviation at Mauna Loa, our selected example for assimilated sites. The prior concentrations on the other hand deviate substantially from the measurements with a bias and standard deviation of 0.89 and 1.24 ppm, respectively.

The histograms of model-observation mismatch, are shown in Fig. 4, for the concentrations of a Prior Forward Run and from the a posteriori CarbonTracker and TM5-4DVar runs. Many of the non-assimilated measurements come from continuous sampling sites and aircraft campaigns which provide a high number of measurements. Normalized bias and standard deviation of the prior mismatch aggregated for all sites are 0.66 and 1.03, respectively. The normalized biases of the mismatch for CarbonTracker and TM5-4DVar are 0.097 and 0.004, respectively, and the standard deviation of the histograms are 0.835 and 0.839, indicating that assimilating observations with the DA systems substantially improves the match to independent data when compared to the prior performance. The spread of the a posteriori model-observation mismatch, however, is somewhat greater than for the comparison to assimilated measurements. This is as expected and indicates a slightly worse performance of both ~~models~~ methods for the non-assimilated than for assimilated sites.

4.2.1 Robustness of the result

CarbonTracker a posteriori concentrations show a larger bias for non-assimiliated measurements (0.097) than for assimilated measurements (0.006). TM5-4DVar biases are more similar for non-assimiliated (0.004) and assimilated measurements (0.025). In order to investigate whether these differences are likely to be an artefact of our selection of validation sites, we conduct a resampling experiment. Out of the 50 sites for which there are non-assimilated observations – our ~~2~~ 26 validation sites, aircraft measurements and sites for which only a given measurement method is assimilated – we randomly select subsets of 25 sites and recalculate the statistical model-observation bias for non-assimilated measurements. Then we repeat the exercise ~~10~~ 9 times and examine the distribution of the resampled CarbonTracker and TM5-4DVar biases. Figure 5 shows that the normalized biases for the CarbonTracker baseline run consistently scatter around 0.08 with a standard deviation of 0.04 while the TM5-4DVar average bias and standard deviation are −0.04 and 0.07, respectively.

So, while CarbonTracker a posteriori concentrations appear offset from the (non-assimilated) observations, TM5-4DVar does not show a significant overall bias but greater station-to-station variability for the model-observation mismatch.

4.2.2 Impact of the CarbonTracker assimilation window length

In order to investigate whether the robust bias our resampling found for CarbonTracker can be due to the choice of the ~~EnSRF~~ EnSRF assimilation time window, we vary CarbonTracker's lag and cycle parameters. Figure 6 illustrates the effect of the window length on the model-observation mismatch at Syowa (SYO), Antarctica. Syowa is located far from any major sources or sinks to be adjusted by the DA systems. Therefore, the DA systems cannot match the Syowa measurements by flux adjustment unless they account for far- and long-reaching correlations between concentrations and fluxes. While TM5-4DVar allows for such connections, CarbonTracker's baseline assimilation window strictly limits these to 5 weeks, which is shorter than the transport timescales from strong flux regions to Antarctica. Therefore, the baseline CarbonTracker run shows a small but systematic underestimation of the CO_2 concentration by up to 0.5 ppm observed in Syowa in summer and fall 2009 while TM5-4DVar a posteriori concentrations match well (not shown). Increasing or decreasing CarbonTracker's assimilation window length respectively improves or deteriorates the match to Syowa observations, showing that the assumed temporal correlations play a role. For sites which are closer to biosphere regions, this effect could manifest as flux misattribution, which would show as a mismatch to non-assimilated stations. Figure 5 illustrates the resulting biases for our resampling assessment when CarbonTracker is run with an assimilation window of 10×7 days or 5×20 days instead of 5×7 days. For 10×7 the average normalized bias reduces to 0.03 with a standard deviation of 0.03 and for 5×20 the average normalized bias reduces to -0.01 with a standard deviation of 0.03. Both are consistent with TM5-4DVar's performance and better than the run with 5×7 days. This suggests that a longer assimilation window adds valuable information to CarbonTracker's DA system. It is unclear, though, whether this improved match to validation measurements translates into improved flux estimates since transport model errors might have a larger impact for the longer assimilation windows. In Sect. 5.1 we discuss additional effects from a larger bin size which may make a long assimilation window undesirable, despite the better match to validation measurements.

5 Comparison of a posteriori surface fluxes

Section 4 shows that the ~~models~~ methods are of similar quality when comparing the a posteriori concentrations with assimilated and non-assimilated observations. Here, we turn to evaluating the a posteriori surface fluxes delivered by CarbonTracker and TM5-4DVar.

As first step we describe the results of the baseline runs. Then we analyze detectable features and the effect of a longer assimilation window in CarbonTracker.

5.1 Surface fluxes of the baseline run

For the baseline CarbonTracker and TM5-4DVar runs, Table 1 shows the globally aggregated a posteriori fluxes for the biosphere and oceans from 3 April 2009, to 2 April 2010. CarbonTracker and TM5-4DVar estimate a global carbon sink (due to the biosphere and oceans) which is stronger than the prior estimate by 1.42 and 1.35 $\mathrm{Pg\,C\,a^{-1}}$, respectively. We only show the uncertainty for the prior and TM5-4DVar which is calculated as described by Basu et al. (2013), because for CarbonTracker the aggregation of uncertainties from weekly to yearly scale requires using assumptions about the temporal correlation of the uncertainties. Due to these assumptions, the yearly uncertainties of TM5-4DVar and CarbonTracker would not be comparable, even if we adopted existing schemes as for example the one employed by Peters et al. (2005). The differences in the uncertainties would not be representative of actual differences in the models. Therefore we use the uncertainties from TM5-4DVar as a metric for comparisons.

Different from the Monte-Carlo based uncertainty calculation which Chatterjee and Michalak (2013) used, the error propagation employed in TM5-4DVar always approaches uncertainties from above~~, so~~: the aggregated errors are larger than the analytical uncertainties at the exact minimum of the cost function. Due to this we expect our uncertainties to overestimate the real uncertainties ~~due to~~ from measurement and representativeness errors. With this caveat, the sink estimates of the two models are consistent within the TM5-4DVar uncertainties and also match previous findings for CarbonTracker (Peters et al., 2007). Examining the time series of globally aggregated surface fluxes in Fig. 7 confirms that the two DA systems are consistent on the global scale, both showing stronger summer uptake than the prior.

Figure 8 illustrates the a posteriori biogenic and oceanic fluxes aggregated over the one-year time period on continental scale regions. Agreement between CarbonTracker and TM5-4DVar is found for North America, Africa, Europe, and Australia, as well as for all the oceans except for the Indian Ocean. The optimized fluxes in these regions differ by less than the yearly uncertainties estimated from TM5-4DVar's statistical error aggregation (see Basu et al., 2013). On the other hand, the modelled fluxes from CarbonTracker and TM5-4DVar differ by more than their uncertainty in South America, Asia and the Indian Ocean. In South America they differ by roughly two times the estimated uncertainty, therefore we take a more detailed look at this discrepancy.

5.1.1 TM5-4DVar's flux anomaly in South America

The time series of South American surface fluxes in Fig. 9 reveals that the flux differences in South America stem from particularly large emission estimates in summer 2009 by TM5-4DVar. The temporal structure of TM5-4DVar fluxes for the Indian Ocean as well as the Pacific Ocean, suggest that ocean uptake compensates for the large South America source to match the hemispheric flux budget.

South America suffers from sparseness of observational constraints such that validation of the estimated surface fluxes via comparison of measured and modelled atmospheric CO_2 concentrations is difficult. Aircraft measurements regularly conducted in South America do not provide deeper insight, because they have a data gap in the critical time between June and August 2009. The only other site that is close to the South America flux region is Arembepe in Brazil (ABP, 12.77° S, 38.17° W), a ground sampling station which is used as constraint within our data assimilation exercise.

To check its impact on the fluxes, we perform a sensitivity run without assimilating Arembepe. In this run both models are similarly good at matching modelled a posteriori and measured CO_2 concentrations in Arembepe and mostly follow the prior (see Fig. 10). When assimilating observations from Arembepe however, TM5-4DVar closely follows the observations in spring 2009 while CarbonTracker only moves halfways from the prior to the observations. This can be explained by the ~~outlier-detection~~ outlier-rejection in CarbonTracker: when the difference between the model and a measurement is more than three times the estimated representativeness error of the measurement, CarbonTracker ignores the measurement as outlier. As marine boundary layer site, Arembepe is assigned a representativeness error of only 0.75 ppm, so CarbonTracker ignores most measurements before May 2009.

The aggregated fluxes in Fig. 8 show that assimilating the measurements in Arembepe has a significant effect on the a posteriori fluxes of TM5-4DVar. When taking out Arembepe from the baseline run, TM5-4DVar's attribution of fluxes shifts: the sinks in the Pacific and the Indian Ocean weaken while the strong source in South America disappears. The time series in Fig. 9 provide a temporal fingerprint of the flux difference due to removing Arembepe from the assimilation which identifies the changes in the Pacific and the Indian Ocean as compensation for the removal of the strong source in South America.

The flux changes in CarbonTracker with assimilating Arembepe are within the estimated uncertainties, in the yearly aggregated fluxes as well as in the time series. Disabling the outlier rejection in CarbonTracker causes the modelled concentrations to follow the observations much ~~closer, but~~ more closely. 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 can cause the flux to change less than in TM5-4DVar in those months, leading to the strong reaction of the outlier rejection. But as shown in Fig. 9 ~~it~~ Carbontracker does not show the additional source seen in TM5-4DVar between ~~June~~ July and August 2009~~and neither the~~, where the flux uncertainty of both models differs by less than 10%. Also it does not show the compensation fluxes TM5-4DVar gives in the oceans.

The fluxes induced by assimilating Arembepe show that TM5-4DVar is more susceptible than CarbonTracker to the effect of single measurement sites in regions with very low observation density.

5.1.2 CarbonTracker with longer assimilation window

Figure 8 shows that ~~the difference in the Asian flux estimates is not affected by removing Arembepe from the assimilated sites. When~~ when increasing the assimilation time window of CarbonTracker to 5×20 days ("5×20"), ~~however,~~ CarbonTracker yields roughly the same aggregated flux as TM5-4DVar.

The time series in Fig. 11 suggests that the change in the CarbonTracker estimate of Asian fluxes when going to the longer assimilation window originates from high frequency corrections to the prior fluxes. If the biosphere model needs to be corrected for only one week, the run with weekly flux bins can adjust that week separately while the run with 20 day flux bins has to adjust a full 20 day period. To test this theory, we verified that a run with an assimilation window consisting of ten one-week cycles yields a similar Asian sink as the run with five one-week cycles (1.84 instead of 1.61 $\mathrm{Pg\,C\,a^{-1}}$) which does not increase further when going to fifteen one-week cycles (not shown), while a run with three 20 day cycles yields a similar Asian sink as the run with five 20 day cycles (2.22 instead of 2.25 $\mathrm{Pg\,C\,a^{-1}}$).

For a quantitative discussion of the propagation of aggregation errors see Turner and Jacob (2015). Our findings suggest that there is an impact of roughly 0.5 $\mathrm{Pg\,C\,a^{-1}}$ from high frequency mismatches between the prior model and the measured concentrations during the Asian summer which cannot be corrected accurately with a binsize of 20 days or more.

In summary we see good agreement for the baseline fluxes between CarbonTracker and TM5-4DVar on a global scale and for most continents and oceans. The mismatch of the fluxes in South America, the Indian Ocean and Asia can be traced back to two distinct effects: a different flux response in regions with very limited observation coverage and using weekly (CarbonTracker) or monthly (TM5-4DVar) adjustments to account for mismatches on shorter time scales.

5.2 Sensitivity to observation coverage

In order to assess the importance of data density and coverage on the two DA systems, we follow the approach which Bruhwiler et al. (2011) used to analyze the performance of their initial version of a fixed-lag Ensemble Kalman Smoother (Bruhwiler et al., 2005). We carry out 5 "historical" model-runs where we stepwise increase the number of assimilated observation sites, mostly following the historical availability of data. The first run, termed "2/cont", assimilates observations from up to 2 stations per continent. It represents an extremely sparse observation network with different sampling frequencies per site. The runs "1988" and "2000" assimilate observations from all sites that were active in the years 1988 and 2000, respectively. The "2000" run assimilates roughly the same number of observations as our baseline run. The run "2010" uses all stations which were active in the year 2010 except for Arembepe. We exclude Arembepe from the "2010" run, because as shown in Sect. 5.1 the different treatment of the observations there would dominate the flux changes and as such mask other effects. Figure S2 illustrates the observation density and coverage for the different historical runs while Table S1 lists the sites included for all the historical runs.

Figure 12 shows the globally aggregated prior and a posteriori fluxes for the baseline setup and each of the historical runs. All the historical runs for both models, CarbonTracker as well as TM5-4DVar, yield consistent estimates of the global (biospheric and oceanic) carbon sink. The results differ by a few tenth's of a $\mathrm{Pg\,C a^{-1}}$ which is well below the TM5-4DVar uncertainty estimate of about 1 $\mathrm{Pg\,C\,a^{-1}}$. This consistency is expected since the global carbon sink is well constrained by the trend in global background concentrations. Compared to the prior, all runs indicate a stronger sink by more than 1 $\mathrm{PgCa^{-1}}$. The global flux estimate is robust against changes in the observation coverage and against the choice of the inverse ~~model.~~ method. Global scale fluxes are also consistent with the 2013B estimates from CarbonTracker North America (NOAA, ESRL) [2]. NOAA shows a global sink of 6.79 ± 6.86

[2] The 2013B release of the estimated fluxes of Carbontracker North America (NOAA, ESRL) is available from esrl.noaa.gov/gmd/ccgg/carbontracker/CT2013B/fluxtimeseries.php?region=Global

PgC for 2009 while we see values between 6.37 and 7.03 PgC for April 2009 to April 2010, depending on the observation data we assimilate.

On the continental scale we take a closer look at North America, since changes in observation density are historically most pronounced there. Figure 13 shows that TM5-4DVar and CarbonTracker fluxes for North America become more similar the denser the observation network becomes, with almost the same flux estimate in the "2010" setup in which the ~~models~~ DA systems assimilate more than 15 sites on the North American continent (see Fig. 14). This good match of both ~~models~~ methods suggests that the density of observation sites in North America suffices to optimize continental scale fluxes with some degree of certainty. Separating the fluxes of the two North America Transcom regions (Fig. 13) shows that for the more homogeneous transcom region in Boreal North America the results from both ~~models~~ methods are already converged with the observation coverage in the "1988" run, while in the more heterogeneous North American Temperate region with many agricultural regions, the ~~models~~ methods only converge in the "2010" setup.

The stronger land sink seen by TM5-4DVar for "2/cont" stems from assimilating only two sites: a site in West Branch in Iowa, USA (WBI, 41.7° N, 91.4° W), in the US corn belt, and a site on Sable Islands, Nova Scotia, Canada (WSA, 43,9° N, 60.0° W). In TM5-4DVar, the strong summer sink near West Branch dominates the North America fluxes and increases the sink from roughly 1 $\mathrm{Pg\,C\,a^{-1}}$ in the "2010" run to more than 1.6 $\mathrm{Pg\,C\,a^{-1}}$ in the "2/cont" run. CarbonTracker is less susceptible to this effect than TM5-4DVar, because its ecoregion approach enforces a correlation between the fluxes for all regions in the corn belt as well as for all regions with grassland – both region-types span the area from the southern parts of North America up to the border of Canada. This makes it more likely that a potential flux ~~adjustement~~ adjustment is constrained by more than one site which gives it a stronger meridional coupling. Since meridional mixing is much slower than zonal mixing, stronger meridional coupling forces a larger region to change in the same way. ~~As such the ecoregion approach makes it more likely that a potential flux adjustment is constrained by more than one site~~For example adjusting the flux in the corn belt yields concentration changes all over North America (downwind of the corn belt ecoregion).

On the other hand, the overall North American sink of $0.65\,\mathrm{Pg\,C\,a^{-1}}$ estimated by CarbonTracker in the "1988" run are 30 % lower than the sink of $0.95\,\mathrm{Pg\,C\,a^{-1}}$ in the "2010" run, while in TM5-4DVar the "1988" and the "2010" run differ only by 10 % ($0.1\,\mathrm{Pg\,C\,a^{-1}}$). The difference between the results for the "2000" and the "2010" runs in North America is on the order of $0.1\,\mathrm{Pg\,C\,a^{-1}}$ for both models, but in different ~~directons~~directions. So with low observation coverage, the quality of the inversion in either system depends on the exact distribution of the observations. This suggests that with the coverage from "2000", we need to assume a minimum uncertainty of $0.25\,\mathrm{Pg\,C\,a^{-1}}$ from only the choice of the inverse method. For "2010" this is down to less than $0.1\,\mathrm{Pg\,C\,a^{-1}}$.

The strong reduction of the uncertainty estimate in the North America fluxes of TM5-4DVar in the "2/cont" run, despite assimilating only 2 sites in North America, shows the sensitivity of these estimates to the raw number of assimilated observations. It proves that the actual structure of the ~~observations~~ observational network has to be taken into account when interpreting the reduction of model-estimated uncertainty.

Overall our results show that the current observation coverage in North America allows estimating robust fluxes on continental scales and on the scales of transcom regions. The historically improving agreement between both models for the aggregated North American fluxes and the two transcom regions in North America suggests that increasing the observation coverage allows getting robust fluxes on even smaller scales.

6 Conclusions

Our study evaluates the performance of the data assimilation models CarbonTracker and TM5-4DVar by comparing their a posteriori CO_2 concentration fields to measurements and by comparing their a posteriori surface fluxes. We test the sensitivity of the a posteriori CO_2 fluxes to model parameters and data coverage. To analyze the impact of the inverse method and the flux representation, the models run in setups which are close to their production settings but use harmonized input data, tracer transport model~~and prior estimates.~~, prior flux and prior flux covariance estimates. 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.

Both inverse models yield CO_2 concentration fields of comparable quality. We show that increasing the length of the assimilation time window of CarbonTracker to five bins of twenty days or ten bins of seven days gives a good agreement to observations in Antarctica which are underestimated in summer when using the default setup with an assimilation window of only five weeks. With these longer windows, the difference of the bias of the models at non-assimilated measurement sites is lower than the uncertainty of the bias due to the limited number of non-assimilated sites. This has two implications: first, the differences between the a posteriori fluxes provide a lower estimate of the uncertainty due to the choice of the optimization method, and second, a choice between the two systems may reduce 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.

The a posteriori fluxes from both models are in good agreement on a global scale, but on continental ~~scale~~ scales they show significant differences, most noticably in South America which has very sparse coverage of observation sites. Investigating the flux time series allows tracing these differences back to spurious flux adjustments in TM5-4DVar for South America due to assimilating observations from a single site in Arembepe, Brazil, along with compensating fluxes in the oceans. Also we see a difference in the adjustment of Asian fluxes, but an additional CarbonTracker run with a coarser temporal flux adjustment bin size of 20 days gives similar fluxes in Asia as TM5-4DVar. Here, the flux time series reveal that part of the weaker sink in CarbonTracker with smaller bin size stems from high frequency changes which cannot be represented with the monthly binning of flux-adaptions in TM5-4DVar and the CarbonTracker run with bins of 20 days. The impact of this effect on the fluxes in Asia is 0.5 $\mathrm{Pg\,C\,a^{-1}}$.

To better analyze the sensitivity of both models to the observation coverage, we run the models with collections of measurement sites selected by historical availability. In North America, where the change of observation coverage is most pronounced, fluxes estimated with the observation network from 2000 differ by 0.25 $\mathrm{Pg\,C\,a^{-1}}$, which can serve as lower limit for the uncertainty due to changing the ~~inverse system~~method. With the measurement network from 2010, the difference reduces to 0.1 $\mathrm{Pg\,C\,a^{-1}}$.

TM5-4DVar has a stronger response to the data coverage than CarbonTracker. This shows that the ecoregion approach in CarbonTracker with its stronger meridional coupling of fluxes and observations makes CarbonTracker less susceptible to changes in the distribution and density of observations than the simple global flux covariance in TM5-4DVar. As such it might be useful to reuse CarbonTracker's spatial flux correlation structure in TM5-4DVar.

Generally, we see sensitivity of the optimized fluxes to the density and distribution of observations which might be particularly important for using satellite data, in which the coverage of observations changes with cloud cover. The improved agreement between both models when adding observation sites indicates that the coverage of observation sites in North America should be sufficient to yield robust fluxes on a continental scale when only considering the uncertainty from the inverse methods and the flux representation.

The Supplement related to this article is available online at doi:10.5194/acpd-0-1-2015-supplement.

Acknowledgements. André Butz and Arne Babenhauserheide are supported by the Emmy-Noether program of the Deutsche Forschungsgemeinschaft (DFG) through grant BU2599/1-1 (RemoteC). This study would not have been possible without the work of Maarten Krol (SRON Netherlands Institute for Space Research, Utrecht, the Netherlands; Institute for Marine and Atmospheric Research (IMAU), Utrecht University, Utrecht, the Netherlands; Department of Meteorology and Air Quality (MAQ), Wageningen University and Research Centre, Wageningen, the Netherlands) who maintains the Transport Model 5 (TM5). Also the authors are indebted to the many individuals and institutions who have contributed original data to the ObsPack project, specifically version v1.0.2 2013-01-28 used in this study. They are J. A. Morgui (IC3); Ernst Brunke (SAWS); Samuel Hammer (UHEI-IUP); Shuji Aoki (NIPR); Markus Leuenberger (KUP); Harro Meijer (RUG); Hidekazu Matsueda (NIES-MRI); Juha Hatakka (FMI); Paul Krummel (CSIRO); Marcel van der Schoot (CSIRO); Margaret Torn (LBNL); Michel Ramonet (LSCE/RAMCES); Doug Worthy (EC); Angel J. Gomez-Pelaez (AEMET); Takakiyo Nakazawa (NIPR); Shinji Morimoto (NIPR); Luciana V. Gatti (IPEN); Ken Masarie (NOAA); Arlyn Andrews (NOAA); Ray Langenfelds (CSIRO); Sebastien Biraud (LBNL); Colm Sweeney (NOAA); Britton Stephens (NCAR); Paul Steele (CSIRO); Ed Dlugokencky (NOAA); Ralph Keeling (SIO); Eckhart Scheel (SAWS); Toshinobu Machida (NIES-MRI); Tuula Aalto (FMI); Hiroshi Koide (JMA); Steve Wofsy (HU); Martina Schmidt (LSCE/RAMCES); Yousuke Sawa (NIES-MRI); Laszlo Haszpra (HMS); Ingeborg Levin (UHEI-IUP); Kirk Thoning (NOAA); Casper Labuschagne (SAWS) and Pieter Tans (NOAA). Also we thank the CarbonTracker NOAA ESRL, Boulder, Colorado, USA for providing the 2011_oi results and documentation. Finally we thank Paul Tol from SRON for his work on colorschemes which work for colorblind people, O. Tange for easy ways to save time using GNU parallel (;login: The USENIX Magazine, February 2011:42-47) and E. Jones and J. D. Hunter for their tools for convenient data analysis with Python.

The article processing charges for this open-access publication have been covered by a Research Centre of the Helmholtz Association.

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 CO_2 fluxes estimated from GOSAT retrievals of total column CO_2, Atmos. Chem. Phys., 13, 8695–8717, doi:10.5194/acp-13-8695-2013, 2013.

Bergamaschi, P., Krol, M., Meirink, J. F., Dentener, F., Segers, A., van Aardenne, J., Monni, S., Vermeulen, A. T., Schmidt, M., Ramonet, M., Yver, C., Meinhardt, F., Nisbet, E. G., Fisher, R. E., O'Doherty, S., and Dlugokencky, E. J.: Inverse modeling of European CH_4 emissions 2001–2006, J. Geophys. Res.-Atmos., 115, D22309, doi:10.1029/2010JD014180, 2010.

Bruhwiler, L. M. P., Michalak, A. M., Peters, W., Baker, D. F., and Tans, P.: An improved Kalman Smoother for atmospheric inversions, Atmos. Chem. Phys., 5, 2691–2702, doi:10.5194/acp-5-2691-2005, 2005.

Bruhwiler, L. M. P., Michalak, A. M., and Tans, P. P.: Spatial and temporal resolution of carbon flux estimates for 1983–2002, Biogeosciences, 8, 1309–1331, doi:10.5194/bg-8-1309-2011, 2011.

Chatterjee, A. and Michalak, A. M.: Technical Note: Comparison of ensemble Kalman filter and variational approaches for CO_2 data assimilation, Atmos. Chem. Phys., 13, 11643–11660, doi:10.5194/acp-13-11643-2013, 2013.

Chevallier, F., Fisher, M., Peylin, P., Serrar, S., Bousquet, P., Bréon, F.-M., Chédin, A., and Ciais, P.: Inferring CO_2 sources and sinks from satellite observations: method and application to TOVS data, J. Geophys. Res., 110, 24309, doi:10.1029/2005JD006390, 2005.

Chevallier, F., Ciais, P., Conway, T. J., Aalto, T., Anderson, B. E., Bousquet, P., Brunke, E. G., Ciattaglia, L., Esaki, Y., Fröhlich, M., Gomez, A., Gomez-Pelaez, A. J., Haszpra, L., Krummel, P. B., Langenfelds, R. L., Leuenberger, M., Machida, T., Maignan, F., Matsueda, H., Morguí, J. A., Mukai, H., Nakazawa, T., Peylin, P., Ramonet, M., Rivier, L., Sawa, Y., Schmidt, M., Steele, L. P., Vay, S. A., Vermeulen, A. T., Wofsy, S., and Worthy, D.: CO_2 surface fluxes at grid point scale estimated from a global 21 year reanalysis of atmospheric measurements, J. Geophys. Res., 115, D21307, doi:10.1029/2010JD013887, 2010.

Enting, I. G.: Green's Function Methods of Tracer Inversion, vol. 114, American Geophysical Union, Washington, DC, 19–31, 2000.

Errico, R. M.: What is an adjoint model?, B. Am. Meteorol. Soc., 78, 2577–2591, 1997.

European Centre for Medium-Range Weather Forecasts: ERA Interim, via the ECMWF Data Server, available at: http://apps.ecmwf.int/datasets/ (last access: 20 March 2015), 2013.

Fairbairn, D., Pring, S. R., Lorenc, A. C., and Roulstone, I.: A comparison of 4DVar with ensemble data assimilation methods, Q. J. Roy. Meteor. Soc., 281–294, doi:10.1002/qj.2135, 2013.

Feng, L., Palmer, P. I., Yang, Y., Yantosca, R. M., Kawa, S. R., Paris, J.-D., Matsueda, H., and Machida, T.: Evaluating a 3-D transport model of atmospheric CO_2 using ground-based, aircraft, and space-borne data, Atmos. Chem. Phys., 11, 2789–2803, doi:10.5194/acp-11-2789-2011, 2011.

Fisher, M. and Courtier, P.: Estimating the covariance matrices of analysis and forecast error in variational data assimilation, ECMWF Technical Memorandum, Tech. Rep. 220, European Centre for Medium-Range Weather Forecast, 1995.

Foken, T., Aubinet, M., and Leuning, R.: The Eddy Covariance Method, Springer, the Netherlands, doi:10.1007/978-94-007-2351-1_1, 1–19, 2012.

Gilbert, J. and Lemaréchal, C.: Some numerical experiments with variable-storage quasi-Newton algorithms, Math. Programm., 45, 407–435, doi:10.1007/BF01589113, 1989.

Gurney, K., Law, R., and Rayner, P.: Transcom 3 experimental protocol, Department of Atmospheric Science, Colorado State University USA, 2000.

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 Biogeochem. Cy., 18, 1010, doi:10.1029/2003GB002111, 2004.

Hooghiemstra, P. B., Krol, M. C., Meirink, J. F., Bergamaschi, P., van der Werf, G. R., Novelli, P. C., Aben, I., and Röckmann, T.: Optimizing global CO emission estimates using a four-dimensional variational data assimilation system and surface network observations, Atmos. Chem. Phys., 11, 4705–4723, doi:10.5194/acp-11-4705-2011, 2011.

Houweling, S., Krol, M., Bergamaschi, P., Frankenberg, C., Dlugokencky, E. J., Morino, I., Notholt, J., Sherlock, V., Wunch, D., Beck, V., Gerbig, C., Chen, H., Kort, E. A., Röckmann, T., and Aben, I.: A multi-year methane inversion using SCIAMACHY, accounting for systematic errors using TCCON measurements, Atmos. Chem. Phys., 14, 3991–4012, doi:10.5194/acp-14-3991-2014, 2014.

Jacobson, A. R., Mikaloff Fletcher, S. E., Gruber, N., Sarmiento, J. L., and Gloor, M.: A joint atmosphere–ocean inversion for surface fluxes of carbon dioxide: 2. Regional results, Global Biogeochem. Cy., 21, GB1020, doi:10.1029/2006GB002703, 2007a.

Jacobson, A. R., Mikaloff Fletcher, S. E., Gruber, N., Sarmiento, J. L., and Gloor, M.: A joint atmosphere–ocean inversion for surface fluxes of carbon dioxide: 1. Methods and global-scale fluxes, Global Biogeochem. Cy., 21, B1019, doi:10.1029/2005GB002556, 2007b.

Juhász, Z. and Bölöni, G.: Tests with the CONGRAD minimization algorithm within the ALADIN/HU 3DVAR system, congrad requires about $2/3$rds of the CPU time required by M1QN3, University of Eötvös Loránd and Hungariam Meteorological Service (HMS), available at http://www.rclace.eu/File/Data_Assimilation/2007/jz_bg_congrad.pdf (last access: 20 March 2015), 2007.

Kalnay, E.: The Future of Data Assimilation: 4D-Var or Ensemble Kalman Filter?, Talk at SAMSI, available at: https://web.archive.org/web/20120915073405/http://www.atmos.umd.edu/~ekalnay/4DVarEnKFKalnaySAMSI.ppt.pdf (last access: 20 March 2015), 2005.

Krol, M., Houweling, S., Bregman, B., van den Broek, M., Segers, A., van Velthoven, P., Peters, W., Dentener, F., and Bergamaschi, P.: The two-way nested global chemistry-transport zoom model TM5: algorithm and applications, Atmos. Chem. Phys., 5, 417–432, doi:10.5194/acp-5-417-2005, 2005.

Lahoz, W. A. and Schneider, P.: Data assimilation: making sense of earth observation, Frontiers in Environmental Science, 2, 16, doi:10.3389/fenvs.2014.00016, 2014.

Lahoz, W. A., Errera, Q., Swinbank, R., and Fonteyn, D.: Data assimilation of stratospheric constituents: a review, Atmos. Chem. Phys., 7, 5745–5773, doi:10.5194/acp-7-5745-2007, 2007.

Lanczos, C.: An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, J. Res. Nat. Bur. Stand., 45, 255–282, 1950.

Masarie, K. A., Peters, W., Jacobson, A. R., and Tans, P. P.: ObsPack: a framework for the preparation, delivery, and attribution of atmospheric greenhouse gas measurements, Earth Syst. Sci. Data, 6, 375–384, doi:10.5194/essd-6-375-2014, 2014.

Meesters, A. G. C. A., Tolk, L. F., Peters, W., Hutjes, R. W. A., Vellinga, O. S., Elbers, J. A., Vermeulen, A. T., van der Laan, S., Neubert, R. E. M., Meijer, H. A. J., and Dolman, A. J.: Inverse carbon dioxide flux estimates for the Netherlands, J. Geophys. Res.-Atmos., 117, D20306, doi:10.1029/2012JD017797, 2012.

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, doi:10.5194/acp-8-6341-2008, 2008.

Navon, I. M. and Legler, D. M.: Conjugate-gradient methods for large-scale minimization, in meteorology, Mon. Weather Rev., 115, 1479–1502, 1990.

NOAA Environmental Sciences Division ~~,~~ and Oak Ridge National Laboratory: obspack CO_2 1 PROTOTYPE v1.0.4 2013-11-25, doi:10.3334/OBSPACK/1001, 2013.

Olson, J., Watts, J., and Allsion, L.: Major World Ecosystem Complexes Ranked by Carbon in Live Vegetation: a Database, Tech. rep., Carbon Dioxide Information Analysis Center, U.S. Department of Energy, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, doi:10.3334/CDIAC/lue.ndp017, 1992.

Peters, G. P., Andrew, R. M., Boden, T., Canadell, J. G., Ciais, P., Le Quere, C., Marland, G., Raupach, M. R., and Wilson, C.: The challenge to keep global warming below 2 °C, Nature Climate Change, 3, 4–6, doi:10.1038/nclimate1783, 2013.

Peters, W., Miller, J. B., Whitaker, J., Denning, A. S., Hirsch, A., Krol, M. C., Zupanski, D., Bruhwiler, L., and Tans, P. P.: An ensemble data assimilation system to estimate CO_2 surface fluxes from atmospheric trace gas observations, J. Geophys. Res.-Atmos., 110, 24304, doi:10.1029/2005JD006157, 2005.

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M. P., Pétron, G., Hirsch, A. I., Worthy, D. E. J., van der Werf, G. R., Randerson, J. T., Wennberg, P. O., Krol, M. C., and Tans, P. P.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, P. Natl. Acad. Sci. USA, 104, 18925–18930, doi:10.1073/pnas.0708986104, 2007.

Peters, W., Krol, M. C., van der Werf, G. R., Houweling, S., Jones, C. D., Hughes, J., Schaefer, K., Masarie, K. A., Jacobson, A. R., 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., van der Laan, S., Neubert, R. E. M., Jordan, A., Rodó, X., Morguí, J.-A., Vermeulen, A. T., Popa, E., Rozanski, K., Zimnoch, M., Manning, A. C., 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 Biol., 16, 1317–1337, doi:10.1111/j.1365-2486.2009.02078.x, 2010.

Peylin, P., Law, R. M., Gurney, K. R., Chevallier, F., Jacobson, A. R., Maki, T., Niwa, Y., Patra, P. K., Peters, W., Rayner, P. J., Rödenbeck, C., van der Laan-Luijkx, I. T., and Zhang, X.: Global atmospheric carbon budget: results from an ensemble of atmospheric CO_2 inversions, Biogeosciences, 10, 6699–6720, doi:10.5194/bg-10-6699-2013, 2013.

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding: Theory and Practice (Series on Atmospheric Oceanic and Planetary Physics), World Scientific Publishing Company, available at: http://www.worldcat.org/isbn/981022740X (last access: 20 March 2015), 2000.

Schaefer, K., Collatz, G. J., Tans, P., Denning, A. S., Baker, I., Berry, J., Prihodko, L., Suits, N., and Philpott, A.: Combined Simple Biosphere/Carnegie-Ames-Stanford Approach terrestrial carbon cycle model, J. Geophys. Res.-Biogeo., 113, G03034, doi:10.1029/2007JG000603, 2008.

Schimel, D.: Carbon cycle conundrums, P. Natl. Acad. Sci. USA, 104, 18353–18354, doi:10.1073/pnas.0709331104, 2007.

Schulze, E. D., Luyssaert, S., Ciais, P., Freibauer, A., Janssens, I. A., Soussana, J. F., Smith, P., Grace, J., Levin, I., Thiruchittampalam, B., Heimann, M., Dolman, A. J., Valentini, R., Bousquet, P., Peylin, P., Peters, W., Rödenbeck, C., Etiope, G., Vuichard, N., Wattenbach, M., Nabuurs, G. J., Poussi, Z., Nieschulze, J., Gash, J. H., and the CarboEurope Team: importance of methane and nitrous oxide for Europe's terrestrial greenhouse-gas balance, Nat. Geosci., 2, 842–850, doi:10.1038/ngeo686, 2009.

Skachko, S., Errera, Q., Ménard, R., Christophe, Y., and Chabrillat, S.: Comparison of the ensemble Kalman filter and 4D-Var assimilation methods using a stratospheric tracer transport model, Geosci. Model Dev., 7, 1451–1465, doi:10.5194/gmd-7-1451-2014, 2014.

Stocker, T. F., Dahe, Q., and Plattner, G.-K.: Climate Change 2013: The Physical Science Basis, Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Summary for Policymakers (IPCC, 2013), IPCC Working Group I Technical Support Unit, University of Bern, Bern, Switzerland, 2013.

Thompson, R. L., Ishijima, K., Saikawa, E., Corazza, M., Karstens, U., Patra, P. K., Bergamaschi, P., Chevallier, F., Dlugokencky, E., Prinn, R. G., Weiss, R. F., O'Doherty, S., Fraser, P. J., Steele, L. P., Krummel, P. B., Vermeulen, A., Tohjima, Y., Jordan, A., Haszpra, L., Steinbacher, M., Van der Laan, S., Aalto, T., Meinhardt, F., Popa, M. E., Moncrieff, J., and Bousquet, P.: TransCom N_2O model inter-comparison – Part 2: Atmospheric inversion estimates of N_2O emissions, Atmos. Chem. Phys., 14, 6177–6194, doi:10.5194/acp-14-6177-2014, 2014.

Turner, A. J. and Jacob, D. J.: Balancing aggregation and smoothing errors in inverse models, Atmos. Chem. Phys. Discuss., 15, 1001–1026, doi:10.5194/acpd-15-1001-2015, 2015.

van der Velde, I. R., Miller, J. B., Schaefer, K., van der Werf, G. R., Krol, M. C., and Peters, W.: Terrestrial cycling of $^{13}CO_2$ by photosynthesis, respiration, and biomass burning in SiBCASA, Biogeosciences, 11, 6553–6571, doi:10.5194/bg-11-6553-2014, 2014.

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), Atmos. Chem. Phys., 10, 11707–11735, doi:10.5194/acp-10-11707-2010, 2010.

Whitaker, J. S. and Hamill, T. M.: Ensemble data assimilation without perturbed observations, Mon. Weather Rev., 130, 1913–1924, doi:10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2, 2002.

Zhang, H. F., Chen, B. Z., van der Laan-Luijk, I. T., Machida, T., Matsueda, H., Sawa, Y., Fukuyama, Y., Langenfelds, R., van der Schoot, M., Xu, G., Yan, J. W., Cheng, M. L., Zhou, L. X., Tans, P. P., and Peters, W.: Estimating Asian terrestrial carbon fluxes from CONTRAIL aircraft and surface CO_2 observations for the period 2006–2010, Atmos. Chem. Phys., 14, 5807–5824, doi:10.5194/acp-14-5807-2014, 2014.

Table 1. Yearly global CO_2 fluxes and uncertainty (standard deviation) from the Prior forward run and from the baseline runs of TM5-4DVar and CarbonTracker. The § column lists important notes.

	Biosphere + Ocean	Uncertainty	§
Prior forward run	$-5.34\,\mathrm{Pg\,C\,a^{-1}}$	$1.86\,\mathrm{Pg\,C\,a^{-1}}$	
TM5-4DVar	$-6.69\,\mathrm{Pg\,C\,a^{-1}}$	$1.07\,\mathrm{Pg\,C\,a^{-1}}$	
CarbonTracker	$-6.76\,\mathrm{Pg\,C\,a^{-1}}$	N/A	*

* CarbonTracker provides uncertainties on weekly scale. As discussed in Sect. 5.1, aggregating them to yearly scale is not clearly defined and would not be comparable to TM5-4DVar.

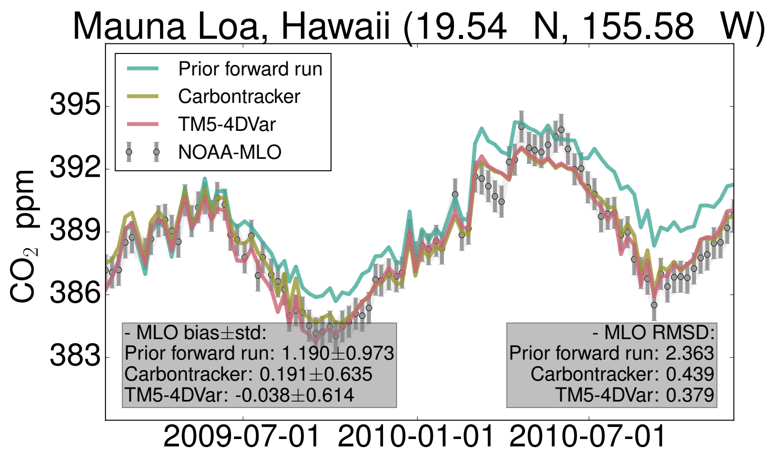

Figure 1. Time Series of measured and modelled CO_2 concentrations from CarbonTracker and TM5-4DVar at Mauna Loa, Hawaii, Pacific (assimilated weekly flasks), NOAA sitecode MLO. Also shown are the concentrations for obtained from a forward run of the transport model using the a priori background flux estimates.

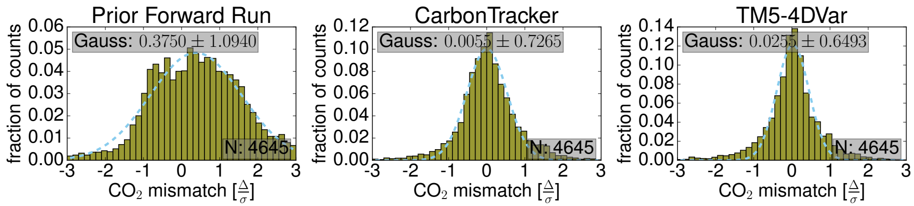

Figure 2. Histograms of the mismatch between measured and modelled CO_2 concentrations for all assimilated measurements using prior fluxes, CarbonTracker optimized fluxes and TM5-4DVar optimized fluxes. The histograms show residuals for one year (3 April 2009 to 2 April 2010) which are normalized by the estimated representativeness error. The line on top of the histograms is a fit of a Gauss function to the histogram. The parameters in the top left show the bias and standard deviation of the Gaussian. The bottom right shows the number of measurements which were accumulated into the histogram.

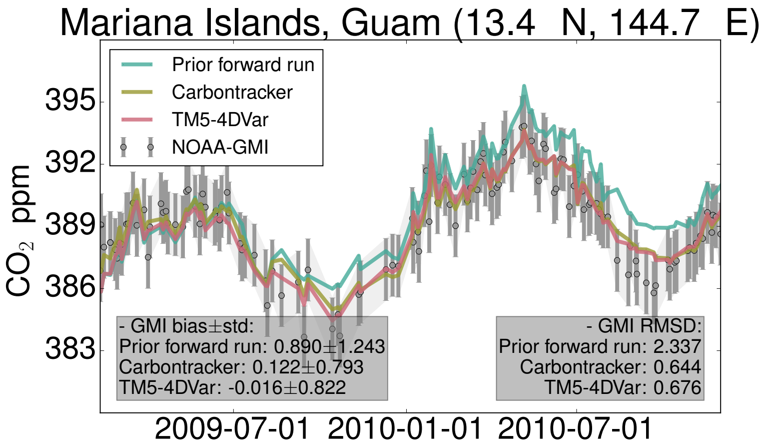

Figure 3. Time Series of measured and modelled CO_2 concentrations from CarbonTracker and TM5-4DVar at Guam, Mariana Islands, Pacific (non-assimilated). Also shown are the concentrations for obtained from a forward run of the transport model using the a priori background flux estimates.

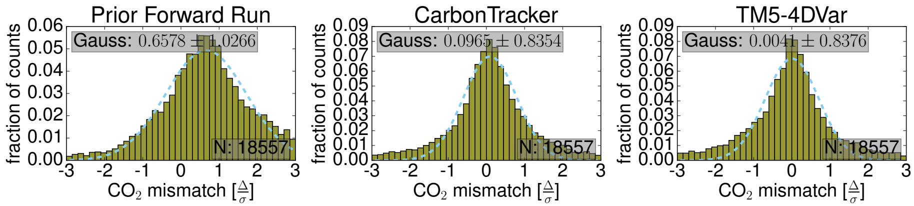

Figure 4. Histograms of the mismatch between measured and modelled CO_2 concentrations for all non-assimilated samples using prior fluxes, CarbonTracker opitmized fluxes and TM5-4DVar optimized fluxes. The histograms show residuals for one year (3 April 2009 to 2 April 2010) which are normalized by the estimated representativeness error. The line on top of the histograms is a fit of a Gauss function to the histogram. The parameters in the top left show the bias and standard deviation of the histogram. The bottom right shows the number of measurements which were accumulated into the histogram.

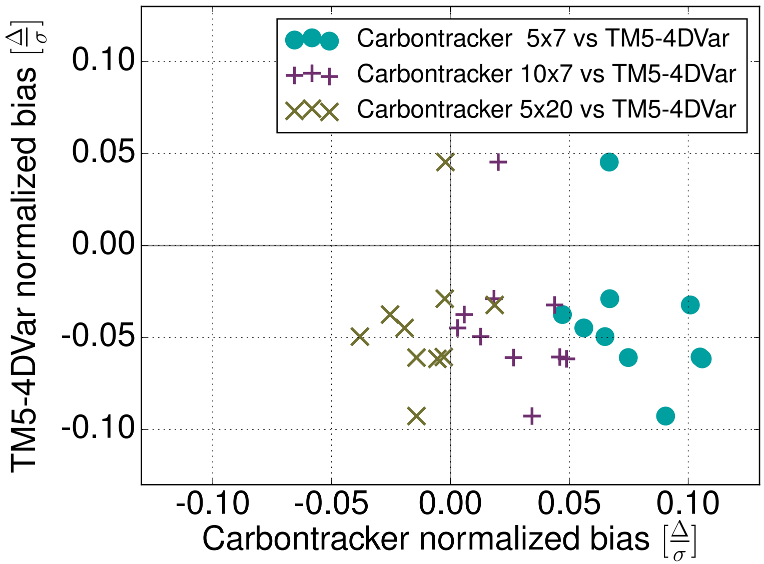

Figure 5. Model-measurement bias of TM5-4DVar against CarbonTracker for non-assimilated measurement sites. Each symbol corresponds to a case resampling excercise where the biases are calculated for 25 randomly drawn sites out of the total 50 resampling sites listed in Table 1 in the Supplement. The baseline run (dots) is compared to a CarbonTracker run with the assimilation period extended to 5×20 days ($\times$) instead of 5×7 days and 10×7 days (+).

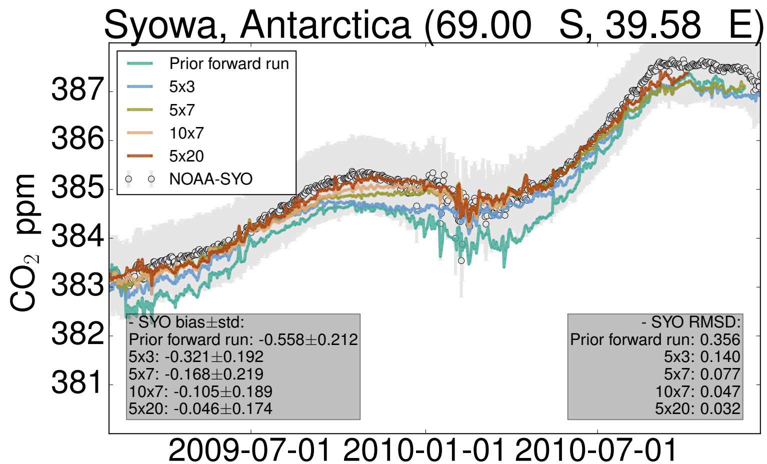

Figure 6. Time Series of measured and modelled CO_2 concentrations in Syowa, Antarctica, for CarbonTracker runs with different length of the assimilation time window. The baseline run uses an assimilation window of 5×7 days. Color coding of shorter and longer assimilation windows follows the legend (lag $\times$ cycle in days).

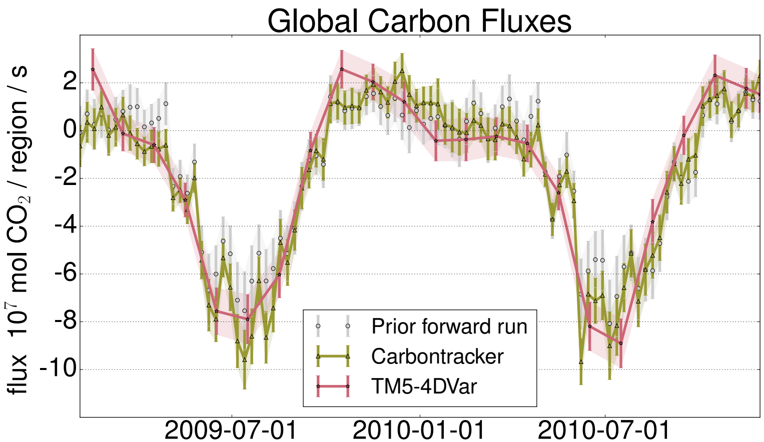

Figure 7. Global fluxes from the baseline runs of TM5-4DVar and CarbonTracker. The Prior is shown in the binning from CarbonTracker. The uncertainties shown for Carbontracker are aggregated spatially but not temporally. As such they represent the uncertainty of the estimated fluxes, calculated directly from the ensemble. These uncertainties are excluded from the anually 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.

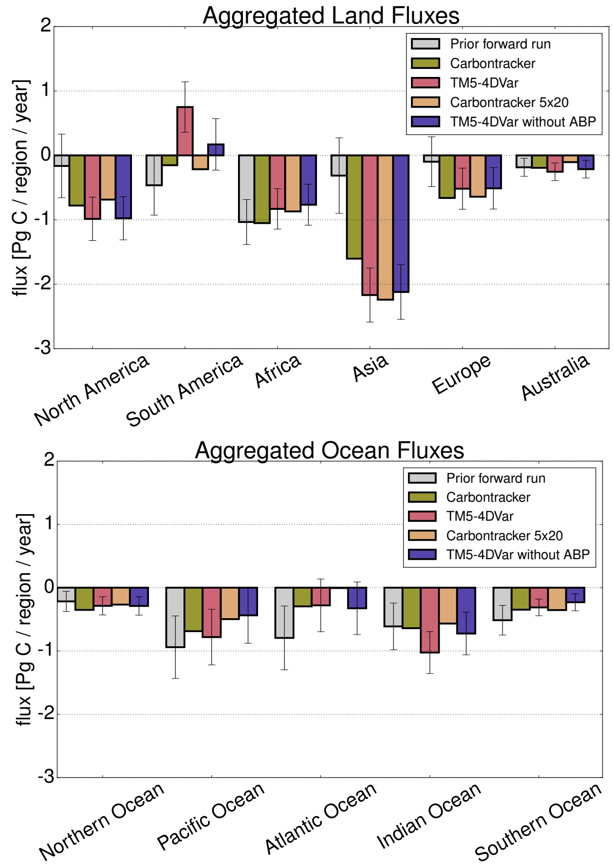

Figure 8. Fluxes from TM5-4DVar and Carbontracker aggregated on continental scale. The uncertainties for TM5-4DVar are calculated following Basu et al. (2013). The error bars for the prior are taken from TM5-4DVar. As written in Sect. 5.1 we show no uncertainties for CarbonTracker, because the aggregation of uncertainties from weekly to yearly scale is not clearly defined.

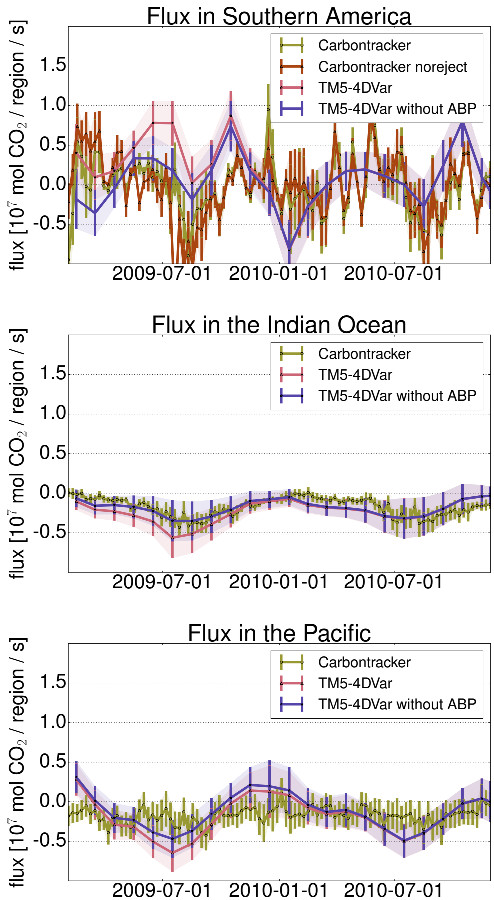
Flux in Southern America
Carbontracker
Carbontracker noreject
TM5-4DVar
TM5-4DVar without ABP
flux [10^7 mol CO2 / region / s]
1.5
1.0
0.5
0.0
-0.5
2009-07-01
2010-01-01
2010-07-01
Flux in the Indian Ocean
Carbontracker
TM5-4DVar
TM5-4DVar without ABP
flux [10^7 mol CO2 / region / s]
1.5
1.0
0.5
0.0
-0.5
2009-07-01
2010-01-01
2010-07-01
Flux in the Pacific
Carbontracker
TM5-4DVar
TM5-4DVar without ABP
flux [10^7 mol CO2 / region / s]
1.5
1.0
0.5
0.0
-0.5
2009-07-01
2010-01-01
2010-07-01

Figure 9. CO_2 surface fluxes from April 2009 to April 2010 in South America, the Indian Ocean and the Pacific. Only the timeseries for South America shows the CarbonTracker noreject, because it follows the CarbonTracker baseline in the other regions. The uncertainties shown for Carbontracker are aggregated spatially but not temporally. See the caption of figure 7 for details.

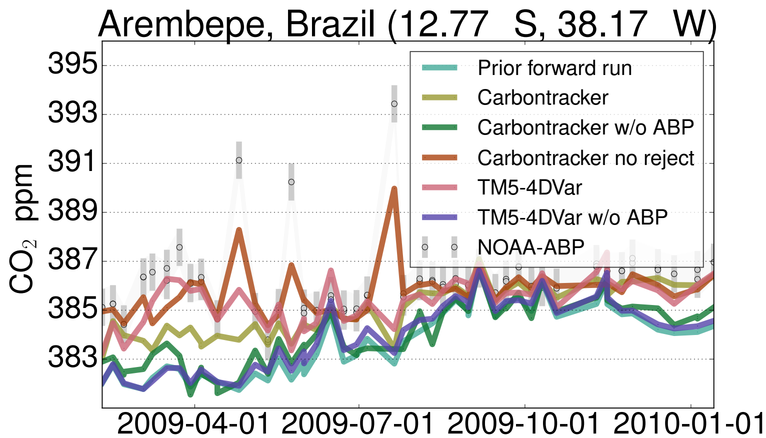

Figure 10. Time Series of CO_2 concentration in Arembepe, Brazil at the east coast of South America. The two "without ABP" runs show the concentrations when the models do not assimilate data from the Arembepe site. CarbonTracker noreject shows the concentrations for CarbonTracker with disabled outlier ~~detection~~rejection. The time series ends after January 2010, because data at Arembepe is only available in obspack PROTOTYPE v1.0.2 2013-01-28 from NOAA Environmental Sciences Division and Oak Ridge National Laboratory (2013) until then.

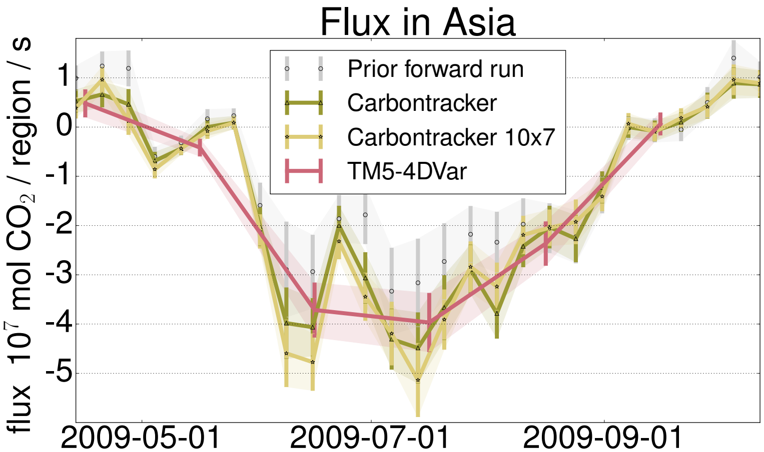

Figure 11. CO_2 surface fluxes during summer 2009 in Asia. The Prior Forward Run shows the prior fluxes aggregated to the binsize of the weekly Carbontracker scaling factors.

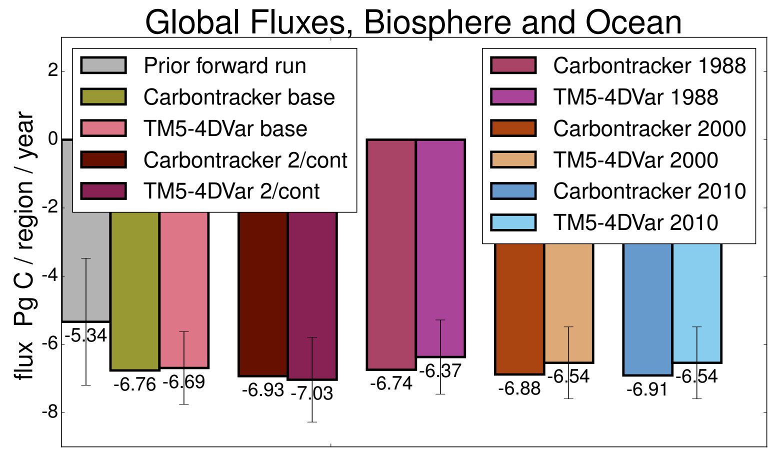

Figure 12. Globally aggregated surface fluxes estimated by the model runs indicated in the legend. In all aggregated flux bar charts, the uncertainties are estimated by TM5-4DVar.

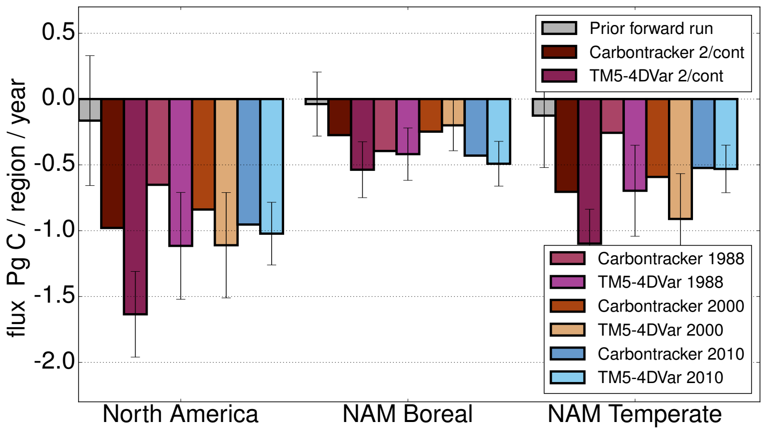

Figure 13. Fluxes for CarbonTracker and TM5-4DVar from April 2009 to April 2010 separated into the two Transcom Regions in North America.

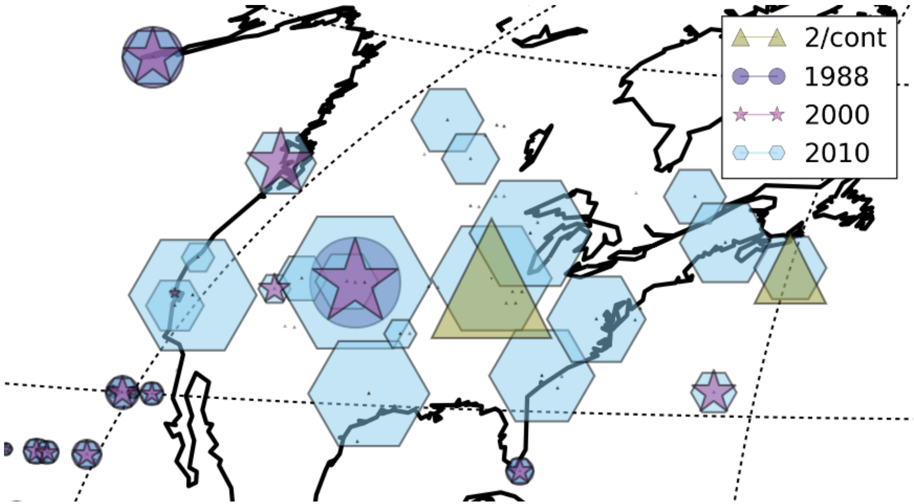

Figure 14. Visualization of the weight of the measurement sites which are assimilated in North America in the respective runs.