First of all, we thank Peter Rayner (in the following referred to as reviewer 1) for his efforts in carefully reviewing our manuscript and his constructive comments.

Point-by-point answers to the comments of reviewer #1

1 General discussion

Reviewer 1: *This paper presents a new look at the European carbon balance from the viewpoint of satelite measurements of xCO2. Its result, if correct, is striking indeed, carrying a strong reminder of the controversy following the publication of the Fan et al. paper in 1998 ... Finally I offer one caution from the Fan et al. controversy ...*
Authors: We note the controversy but conclude that there are not many parallels with the discussion of the Fan et al. 1998 publication. Fan et al. analysed the same in situ measurements, which were also used by others at the same time but reported significantly different results. It was thus relevant to question whether the differences with other their results were an artefact introduced by the used method. We, however, invert satellite data sets and find that our results are consistent among an ensemble of inversion set ups. More importantly, our results are also similar to the results of others (Chevallier et al., 2014; Basu et al., 2013; Takagi et al., 2014) inverting (partly) the same satellite data sets but with (more complex) models which were usually used for flask inversions. The authors of these studies did not discuss the European sink or considered their findings unrealistic and suspected potential retrieval biases and/or transport errors as explanation. For this reason, we set up an inversion system, which was particularly designed to be insensitive (or less sensitive) to the suspected error sources. In our view taking the consistency with others into account, the performed sensitivity studies, and the validation, it appears very likely that our results are driven by the data rather than an artefact of the method. In this context, please note also the interactive comment SC C8037 (Nassar et al., 2014) which implies consistency of the results of Nassar et al. (2011) and our interpretation of the carbon sink in Europe.

Reviewer 1: *Controversial, of course, does not mean wrong but, as with the Fan paper, there are enough simplifications here to worry about.*
Authors: Without suggesting that the presented results are wrong, reviewer 1 basically comments on three important issues for this or any equivalent regional inversion technique: 1) the suitability of the used a priori covariance, 2) the horizontal boundary conditions, and 3) the analysis of the aggregation error. All three topics are discussed and explained in the manuscript. The corresponding sensitivity studies do not indicate potential problems with the used method. We have attempted in our sensitivity studies to quantify potential issues in a logical manner, but it is potentially feasible to undertake much more comprehensive studies on the individual technical points. Each

of such study could easily be subject of one or more extensive additional publication. Within the revised manuscript, we conclude our sensitivity studies / error analysis as follows: "By means of an ensemble of five different inversion set-ups (25 ensemble members in total) and a comprehensive error analysis, we can find no indications that i) the used background model providing reference concentrations and a priori fluxes (CarbonTracker), ii) the used convection scheme, iii) the used meteorology, iv) aggregation errors, or v) persistent, inner-European retrieval biases in mean wind direction explain the observed carbon sink."

Reviewer 1: *Another way of stating the paper's conclusion for me is that in situ and remotely-sensed measurements of CO2 suggest very different things about the European net sink.*
Authors: We only partly agree. There is a difference of more than 0.5 GtC/a between our results and the majority of in situ inversions (Peylin et al., 2013). This difference, whilst large with respect to the magnitude of the mean reported total European carbon sink, is similar to the uncertainty or error reported using in situ data. Based on in situ measurements, Chevallier et al. (2014) specifies the European carbon sink by 0.45 ± 0.40 GtC/a for 2010 and Peylin et al. (2013) by 0.40 ± 0.42 GtC/a for 2001–2004. This means only if the satellite derived best estimate was above 1.2 GtC/a, one would (usually) consider it statistically significantly different. Note also that Peylin et al. (2013) found that in situ inversions for Europe in 2001–2004 range from about 0.2 GtC/a to -1.4 GtC/a (assuming an area of $1.0 \cdot 10^{13}\,\mathrm{m}^2$). Within the manuscript, we note this explicitly: "Peylin et al. (2013) performed an inter-comparison study of an ensemble of eleven global inversion models which showed that European CarbonTracker fluxes (0.30 GtC/a for 2001–2004, assuming an area of $1.0 \cdot 10^{13}\,\mathrm{m}^2$) are similar to the ensemble mean (0.40 GtC/a). However, the ensemble spread is 0.42 GtC/a (1-sigma) and individual models estimate the European biospheric carbon sink to be in the order of 1 GtC/a, which is similar to our findings and it should be noted that the analysed period (2001–2004) includes 2003 with little uptake.". In this context, we added to the abstract of the revised manuscript: "The difference to in situ based inversions (Peylin et al., 2013), whilst large with respect to the mean reported European carbon sink (0.4 GtC/a for 2001–2004), is similar in magnitude to the reported uncertainty (0.42 GtC/a)."

2 The used a priori

Reviewer 1: *The Carbontracker posterior estimate is a reasonable starting point for an inversion like this. As the authors point out, it could allow a stepwise inversion of surface and xCO2 data. for this, though, the uncertainty from Carbontracker must be correctly fed through from the observations and*

must, itself, be correctly derived from the observations for the new step of the inversion.
Authors: We agree with the reviewer that a key issue is that the uncertainty or error used in the inversion describes the "true" or realistic uncertainty of the prior sufficiently well. As a result of CarbonTracker's Kalman filter technique, the original monthly flux uncertainties are considered unrealistically large ("CarbonTracker uses a Kalman filter technique with a five-week assimilation window which results in monthly flux uncertainties considered unrealistically large"). This is the reason why we use a scaling factor. Assuming that CarbonTracker results are not significantly better or worse compared to the results of Chevallier et al. (2014) and Basu et al. (2013), we chose a scaling factor of 1/3, which results in annual uncertainties being similar to those of Chevallier et al. (2014) and Basu et al. (2013) ("Therefore, we apply a scaling of 1/3 so that the uncertainties of CarbonTracker's annual averages become similar to uncertainties estimated by Basu et al. (2013) and Chevallier et al. (2014) inverting surface in situ measurements."). Also the resulting monthly uncertainties, with lowest values during the dormant season and largest values during the growing season, agree reasonably well with the inter-model spread of an ensemble of atmospheric CO_2 inversions (Peylin et al., 2013). Additionally, a sensitivity study showed that our results do not critically depend on the exact number of the scaling factor (Fig. B1, middle).

Reviewer 1: *I am not confident of this for a couple of reasons. Firstly the Ensemble Kalman Filter used for Carbontracker is a fine technique but a weakness is the specification of the posterior uncertainty. Limited ensemble size and the finite assimilation window make this difficult. furthermore, aggregating the uncertainty to the region used for this study is doable but not trivial ... hopefully it was directly generated from the ensemble members rather than from the estimated posterior covariance.*
Authors: We used the uncertainties as provided by NOAA for the TRANSCOM regions. These uncertainties comprise the "internal" variance (the Kalman filter uncertainty variance) and an "external" variance representing across-model variability.

Reviewer 1: *There are also likely to be temporal uncertainty correlations among the estimates from month to month since the influence of observations in the Carbontracker system is not limited to month boundaries. Correlations have been added but the choice of correlation structure is not very clearly motivated. Certainly doing so to regularize an inversion which is already regularized and constrained by a previous inversion is difficult to justify.*
Authors: As for the state vector elements describing potential biases, we assumed month-to-month correlations with a correlation length of 3 month. For convenience, we used a simple exponential relation to compute consistent covariance matrices for the bias and flux elements (temporal correlations were not provided by NOAA). Our choice of month-to-month correlations may not perfectly represent reality. However, Fig. B1 (right) shows that our

results do not critically depend on the exact choice. Note that only the correlation length has been modified for Fig. B1 (right). This means that the choice of the correlation length influences also the annual uncertainty estimate (dark grey) and the total "weight" of the prior. Simultaneously adapting the scaling factor to obtain a constant annual uncertainty would probably result in an even smaller dependency of our results on the choice of the correlation length.

Reviewer 1: *In summary I think the problem of using the Carbontracker posterior estimate as a prior is harder than it might appear. I even wonder why the authors did this? their formalism makes it fairly easy to start from the same prior as Carbontracker but use both the in situ and xCO2 measurements in the inversion. this would directly test the consistency of the measurements.*
Authors: We used CarbonTracker a posteriori fluxes and corresponding concentrations for multiple reasons: i) The used concentrations have to be "consistent" with the used fluxes. We had no CarbonTracker concentrations at hand which corresponded to the CarbonTracker a priori fluxes. ii) We decided to start from fluxes which are as close to reality as possible so that the flux increments are also as small as possible. We assume that the optimized fluxes represent the spatial and temporal variability better than the prior fluxes. Both helps to minimise potential aggregation errors. iii) We consider that assimilation of in situ measurements can be complex and wanted to profit from NOAA's long-time expertise in the field. As an example, Fig. C3 shows that biases between in situ measurements and CarbonTracker optimised concentrations at individual sites may exist, which are much larger than the measurement accuracy.

3 Boundary conditions

Reviewer 1: *Another concern is the role of horizontal boundary conditions in conditioning determining inversion results. I understand that the use of a global bias removes the impact of the absolute value of the boundary conditions but I dont believe the sensitivity studies rule out a major role for the east-west difference in boundary conditions in determining the integrated flux. this is especially important when there might be considerable uncertainty in these conditions. I recommend carrying out an ensemble of inversions with an ensemble of prior flux and boundary conditions from Carbontracker.*
Authors: Our inversion analyses the relationship between model minus satellite differences and the European surface influence. As a result, differences of gradients in mean wind direction are interpreted as source or sink increments. In other words, our inversion is designed to be sensitive to such gradients. We agree that the particular boundary conditions may influence the results. To address this issue, the ensemble shown in the original manuscript includes five ensemble members using concentrations and a priori fluxes from the MACC

model instead of CarbonTracker (Fig. 1, MACC background). Within the manuscript we discuss the results of these inversion runs: "Even though the inversion solely relies on inner-European gradients, the choice of the background model (CarbonTracker) may introduce potential uncertainties to the inferred fluxes. To investigate this issue, we derived fluxes for all five satellite retrievals in 2010 using the MACC (Chevallier et al., 2014) model (version 11.2) for reference concentrations and a priori fluxes (Fig. 1, MACC background). The resulting annual fluxes are consistent with the results based on CarbonTracker to which they have a root mean square difference (RMSD) of 0.22 GtC/a."

4 Aggregation error

Reviewer 1: *Finally I am not convinced that the sensitivity study to region size is sufficient to rule out aggregation error. the regions involved are still quite large and it may well be true that gradients near the edges of the region provide most of the constraint.*

Authors: For the following reasons we consider that our results are not an artefact due to aggregation errors (quotes indicate citations of the original manuscript): i) For our sensitivity studies (three aggregation experiments times five satellite retrieval data sets), we halved the grid size which is a drastic change in resolution. In spite of this, we find only a small influence on the inversion results: "The inversion results differ from the baseline by 0.12 GtC/a (RMSD, longitude split), 0.06 GtC/a (RMSD, latitude split), and 0.04 GtC/a (RMSD, temporal split), respectively. The average of the three experiments differs by a RMSD of 0.08 GtC/a from the baseline (Fig. 1, aggreg. exp.)". ii) "The inverted satellite data are considerably more densely sampled than in situ measurement sites" which contributes to minimise aggregation errors. iii) "The spatio-temporal patterns of the used a priori fluxes (CarbonTracker) are assumed to be relatively realistic" so that the (regionally constant) flux increment remains as small as possible. iv) "Engelen et al. (2002) estimated this effect for the European TRANSCOM region to be $0.13 - 0.31$ GtC/a". As they used sparsely sensed in situ measurements and a priori flux fields which were (partly) far away from reality, considerably smaller aggregation errors can be expected for our study. v) The used satellite data sets have different samplings resulting in different aggregation errors. If these were dominating our results we would expect to derive larger differences among the analysed ensemble of satellite retrievals. vi) Inversion models deriving surface fluxes on much finer grids by using "weak constraints" are less sensitive to aggregation errors (Engelen et al., 2002, e.g.). The models used for the studies of Basu et al. (2013) and Chevallier et al. (2014) are working in this way. The fact that they found similar values for the European carbon sink is another important evidence that our results are not dominated by aggregation errors (although cancellation of errors cannot be excluded from this point alone).

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, URL `http://www.atmos-chem-phys.net/13/8695/2013/`, 2013.

Chevallier, F., Palmer, P. I., Feng, L., Boesch, H., O'Dell, C. W., and Bousquet, P.: Towards robust and consistent regional CO_2 flux estimates from in situ and space-borne measurements of atmospheric CO_2, Geophysical Research Letters, 41, 1065–1070, doi:10.1002/2013GL058772, URL `http://dx.doi.org/10.1002/2013GL058772`, 2014.

Engelen, R. J., Denning, A. S., and Gurney, K. R.: On error estimation in atmospheric CO_2 inversions, Journal of Geophysical Research, 107, doi:10.1029/2002jd002195, URL `http://dx.doi.org/10.1029/2002JD002195`, 2002.

Nassar, R., Jones, D. B. A., Kulawik, S. S., Worden, J. R., Bowman, K. W., Andres, R. J., Suntharalingam, P., Chen, J. M., Brenninkmeijer, C. A. M., Schuck, T. J., Conway, T. J., and Worthy, D. E.: Inverse modeling of CO_2 sources and sinks using satellite observations of CO_2 from TES and surface flask measurements, Atmospheric Chemistry and Physics, 11, 6029–6047, doi:10.5194/acp-11-6029-2011, URL `http://www.atmos-chem-phys.net/11/6029/2011/`, 2011.

Nassar, R., Jones, D. B. A., and Kulawik, S. S.: Interactive comment on "Satellite-inferred European carbon sink larger than expected" by M. Reuter et al., Atmospheric Chemistry and Physics Discussions, 14, C8037–C8038, URL `www.atmos-chem-phys-discuss.net/14/C8037/2014/`, 2014.

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, URL `http://www.biogeosciences.net/10/6699/2013/`, 2013.

Takagi, H., Houweling, S., Andres, R. J., Belikov, D., Bril, A., Boesch, H., Butz, A., Guerlet, S., Hasekamp, O., Maksyutov, S., Morino, I., Oda, T., O'Dell, C. W., Oshchepkov, S., Parker, R., Saito, M., Uchino, O., Yokota, T., Yoshida, Y., and Valsala, V.: Influence of differences in current GOSAT XCO_2 retrievals on surface flux estimation, Geophysical Research Letters, 41, 2598–2605, doi:10.1002/2013gl059174, URL http://dx.doi.org/10.1002/2013GL059174, 2014.