Satellite-inferred European carbon sink larger than expected

M. Reuter[1], **M. Buchwitz**[1], **M. Hilker**[1], **J. Heymann**[1], **O. Schneising**[1], **D. Pillai**[1], **H. Bovensmann**[1], **J. P. Burrows**[1], **H. Bösch**[2], **R. Parker**[2], **A. Butz**[3], **O. Hasekamp**[4], **C. W. O'Dell**[5], **Y. Yoshida**[6], **C. Gerbig**[7], **T. Nehrkorn**[8], **N. M. Deutscher**[1], **T. Warneke**[1], **J. Notholt**[1], **F. Hase**[3], **R. Kivi**[9], **R. Sussmann**[10], **T. Machida**[6], **H. Matsueda**[11], and **Y. Sawa**[11]

[1]Institute of Environmental Physics (IUP), University of Bremen, Bremen, Germany
[2]Department of Physics and Astronomy, University of Leicester, Leicester, UK
[3]IMK-ASF, Karlsruhe Institute of Technology, Karlsruhe, Germany
[4]Netherlands Institute for Space Research, Utrecht, The Netherlands
[5]Colorado State University, Fort Collins, CO, USA
[6]National Institute for Environmental Studies, Tsukuba, Japan
[7]Max Planck Institute for Biogeochemistry, Jena, Germany
[8]Atmospheric and Environmental Research Inc., Lexington, USA
[9]Arctic Research Center, Finnish Meteorological Institute, Sodankylä, Finland
[10]IMK-IFU, Karlsruhe Institute of Technology, Garmisch-Partenkirchen, Germany
[11]Meteorological Research Institute, Tsukuba, Japan

Correspondence to: M. Reuter (maximilian.reuter@iup.physik.uni-bremen.de)

Abstract

Current knowledge about the European terrestrial biospheric carbon sink, from the Atlantic to the Urals, relies upon bottom-up inventory and surface flux inverse model estimates (e.g., 0.27 ± 0.16 GtC/a for 2000–2005 (Schulze et al., 2009), 0.17 ± 0.44 GtC/a
5 for 2001–2007 (Peters et al., 2010), 0.45 ± 0.40 GtC/a for 2010 (Chevallier et al., 2014), 0.40 ± 0.42 GtC/a for 2001–2004 (Peylin et al., 2013)). Inverse models assimilate in situ CO_2 atmospheric concentrations measured by surface-based air sampling networks. The intrinsic sparseness of these networks is one reason for the relatively large flux uncertainties (Peters et al., 2010; Bruhwiler et al., 2011). Satellite-based CO_2 measurements
10 have the potential to reduce these uncertainties (Miller et al., 2007; Chevallier et al., 2007). Global inversion experiments using independent models and independent GOSAT satellite data products consistently derived a considerably larger European sink ($1.0 - 1.3$ GtC/a for 09/2009–08/2010 (Basu et al., 2013), $1.2 - 1.8$ GtC/a in 2010 (Chevallier et al., 2014)). However, these results have been considered unrealistic due to potential retrieval
15 biases and/or transport errors (Chevallier et al., 2014) or have not been discussed at all (Basu et al., 2013; Takagi et al., 2014). Our analysis comprises a regional inversion approach using STILT (Gerbig et al., 2003; Lin et al., 2003) short range (days) particle dispersion modelling, rendering it insensitive to large scale retrieval biases and less sensitive to long-range transport errors. We show that the satellite-derived European
20 terrestrial carbon sink is indeed much larger (1.02 ± 0.30 GtC/a in 2010) than previously expected. This is qualitatively consistent among an ensemble of five different inversion set-ups and five independent satellite retrievals (BESD (Reuter et al., 2011) 2003–2010, ACOS (O'Dell et al., 2012) 2010, UoL-FP (Cogan et al., 2012) 2010, RemoTeC (Butz et al., 2011) 2010, and NIES (Yoshida et al., 2013) 2010) using data from two different instruments
25 (SCIAMACHY (Bovensmann et al., 1999) and GOSAT (Kuze et al., 2009)). 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). The highest gain in information is obtained during

the growing season when satellite observation conditions are advantageous, a priori uncertainties are largest, and the surface sink maximises; during the dormant season, the results are dominated by the a priori. Our results provide evidence that the current understanding of the European carbon sink has to be revisited.

1 Introduction

Global anthropogenic CO_2 emissions are estimated to be $9.3 \pm 0.6\,\text{GtC/a}$ (2002–2011), of which $4.3 \pm 0.1\,\text{GtC/a}$ remain in the atmosphere with the difference being taken up by land $(2.6 \pm 0.8\,\text{GtC/a})$ and ocean $(2.5 \pm 0.5\,\text{GtC/a})$ (Le Quéré et al., 2013). However, large uncertainties remain in our knowledge of the distribution of the land sink. This arises, e.g., from the sparseness of the surface measurements assimilated by inverse models to infer the surface fluxes (Peters et al., 2010; Bruhwiler et al., 2011). This is also the case in Europe, where surface stations are limited to member states of the EU (EU28) and prevailing westerly winds poorly constrain the carbon fluxes in the largest part, i.e., the Russian part of Europe up to the Urals.

Inverse models are optimised to derive global or regional surface fluxes of CO_2 from surface-based in situ measurements. The high accuracy of these measurements allow the models to analyse small gradients over long distances and to apply strict mass-conservation so that (to some extent) information about surface fluxes can be inferred in regions remote from measurement sites. Such models are particularly sensitive to long-range transport errors and large scale biases of the measurements (Miller et al., 2007; Chevallier et al., 2014); e.g., measurement biases in North Africa can corrupt the inferred fluxes in Europe or elsewhere.

Satellite measurements have entirely different strengths and weaknesses compared to surface in situ measurements. They have lower accuracy and precision, but much better spatial coverage. However, regional biases of the satellite retrieved dry-air column-average

3

mole fraction of CO_2 (XCO_2) of a few tenths of a ppm can already hamper an inversion with mass-conserving global inversion models (Miller et al., 2007; Chevallier et al., 2007). Achieving this accuracy is challenging for current satellite retrievals (Reuter et al., 2010, 2013; Buchwitz et al., 2013b) and even for ground based validation measurements
5 (Wunch et al., 2011).

Spatial gradients in the satellite data are more reliable over small scales than over large scales because potential retrieval biases are minimal when similar meteorology, surface characteristics, and observation geometry exist. By allowing global land/sea XCO_2 biases (Basu et al., 2013) or monthly, latitudinally varying XCH_4 biases (Bergamaschi et al.,
10 2013), attempts have been made to adapt the inverse models to the characteristics of the satellite data but mass-conservation can still transport errors over long distances.

2 Inversion technique and satellite data sets

In this study we perform a regional surface flux inversion using only satellite measurements
15 within the European TRANSCOM (Gurney et al., 2002) region (from the Atlantic to the Urals, area $= 1.0 \cdot 10^{13} \, \text{m}^2$, Fig. A1a), thus ensuring that any potential retrieval biases in other regions do not impact on the results. Taking the satellites' averaging kernels into account, we analyse the differences between CarbonTracker CT2011_oi (Peters et al., 2007) model simulations and five independent satellite XCO_2 retrievals
20 (BESD v02.00.08 2003–2010, ACOS v3.4r03 2010, UoL-FP v4.0 2010, RemoTeC v2.11 2010, and NIES v02.xx 2010) of two different instruments (GOSAT and SCIAMACHY). For each sounding (17400/a for BESD, 4000 for ACOS, 4900 for UoL-FP, 3100 for RemoTeC, and 3800 for NIES), we calculate the accumulated European surface influence function (Jacobian) by using the Stochastic Time-Inverted Lagrangian Transport model
25 (STILT). Potential issues arising from long-range transport are reduced because air masses leave the analysis-region typically within a few days. If the XCO_2 difference to CarbonTracker depends on the European surface influence, we infer by how much

4

the CarbonTracker fluxes (being the basis for CarbonTracker's concentrations) would have to be modified in order to bring measurement and model in better agreement. As an example, if the XCO_2 difference (satellite-model) decreases with increasing surface influence, the model fluxes are assumed to be too large (Fig. A1b). A systematic offset is interpreted as retrieval (or model) bias. This results in the inversion being solely dependent on regional (medium scale) gradients, which is a strength of the satellite retrievals. The inversion yields monthly optimised fluxes and utilises the optimal estimation formalism with CarbonTracker fluxes as a priori and first guess. As CarbonTracker assimilates surface in situ measurements, our inversion can be considered to be a stepwise inversion of satellite and surface in situ measurements. For more details see Appendix A.

3   Error analysis and ensemble set-up

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. Additionally, it seems unlikely that five independently developed retrieval algorithms optimised for two different sensors produce consistent erroneous surface fluxes. The analysed potential uncertainties add up to a total uncertainty of $0.30\,\mathrm{GtC/a}$ for the annual fluxes. If not otherwise noted, all uncertainty estimates of annual fluxes within this paper include this additional uncertainty; monthly uncertainty estimates correspond to un-modified a posteriori error estimates. More details about the error analysis and the specific error components can be found in Appendix B.

5

4 Results

Fig. 1 (top) shows the annual European biospheric surface fluxes (land excluding fossil and fire) of CarbonTracker and five satellite data inversions. Note that all uncertainties within this publication correspond to 1-sigma and that CarbonTracker uncertainties
5 have been scaled (see Appendix A). CarbonTracker fluxes, here representing current knowledge, imply that the European carbon sink is 0.41 ± 0.36 GtC/a (multi-year average, Fig. 1, top). For some of the years (especially 2003–2005) it cannot be concluded with high confidence whether Europe is a sink or a source. The flux inversion with BESD using SCIAMACHY data indicates that Europe's biosphere is indeed likely a sink in all
10 analysed years (2003–2010) and very likely in the period 2004–2010. The satellite-derived multi-year average of the European sink is 0.95 ± 0.33 GtC/a.

The year-to-year variations are somewhat larger for BESD than for CarbonTracker, which may imply a larger ecosystem sensitivity. In this context, note that Schneising et al. (2014) found that the seasonal cycle amplitude tends to have a larger temperature
15 sensitivity on the northern hemisphere compared with CarbonTracker, even though the difference is not statistical significant. The smallest sink is being found in 2003 (due to the European heat-wave and drought, Ciais et al., 2005) and the largest sink in 2006.

It is interesting to note that the results reported by Nassar et al. (2011) support
20 a strong European sink in 2006, which they derived from global inversions of TES (Tropospheric Emission Spectrometer) satellite measurements. The TES CO_2 retrieval conceptually differs from SCIAMACHY or GOSAT XCO_2 retrievals because the instrument measures thermal infrared radiation and averaging kernels for CO_2 peak in the mid-troposphere. In the study, solely soundings above oceans between $40° S - 40° N$ were used. Remapping
25 their results yields for the European TRANSCOM region 1.33 ± 0.20 GtC/a (Nassar et al., 2014) which agrees well with our result for 2006 (1.33 ± 0.33 GtC/a).

6

From the perspective of the European carbon sink, 2010 was an average year (CarbonTracker $0.38 \pm 0.35\,\mathrm{GtC/a}$) and it was the first complete year in which GOSAT data were gathered. Analysing an ensemble of five independently developed satellite retrieval algorithms and five different inversion set-ups, our best estimate of the European carbon sink in 2010 is $1.02 \pm 0.30\,\mathrm{GtC/a}$ (see Appendix B). All 25 ensemble members agree reasonably well and suggest a larger carbon sink than estimated by CarbonTracker. The best agreement among the satellite retrievals is found for the baseline inversion (see Appendix B). The largest deviations from the baseline inversion are observed when changing the background model or the meteorology.

Peylin et al. (2013) performed an inter-comparison study of an ensemble of eleven global inversion models which showed that European CarbonTracker fluxes ($0.30\,\mathrm{GtC/a}$ for 2001–2004, assuming an area of $1.0 \cdot 10^{13}\,\mathrm{m}^2$) are similar to the ensemble mean ($0.40\,\mathrm{GtC/a}$). However, the ensemble spread is $0.42\,\mathrm{GtC/a}$ (1-sigma) and individual models estimate the European biospheric carbon sink to be in the order of $1\,\mathrm{GtC/a}$, which is similar to our findings and it should be noted that the analysed period (2001–2004) includes 2003 with little uptake.

The middle panel of Fig. 1 compares monthly surface fluxes and shows that BESD's larger annual sink originates mainly from stronger CO_2 uptake during the growing season and to a lesser extent from weaker CO_2 release during the dormant season. This pattern is stable over the years and consistent among the satellite derived fluxes in 2010 with the exception of the RemoTeC fluxes which are similar to CarbonTracker fluxes in June and July but which are the lowest in January–April. As a result, the annual fluxes for RemoTeC show the weakest sink ($0.74 \pm 0.33\,\mathrm{GtC/a}$). Note that RemoTeC has also the lowest number of soundings which can result in sparsely sampled regions. All satellite retrievals gain the largest uncertainty reduction during the growing season (Fig. 1, bottom) when CarbonTracker uncertainties are largest. A relatively large fraction of satellite observations are made within this period because of advantageous solar zenith

7

angles and cloud conditions. The poor sampling during the dormant season does not allow for a larger error reduction and it cannot be completely excluded that CarbonTracker underestimates respiration and/or decomposition within this period, which would result in a weaker annual average sink. However, it should be noted that this is, in principle,

5   accounted for by error propagation into the uncertainty of the annual averages assuming that the a priori fluxes in the dormant season are unbiased. Due to the lower activity of the biosphere during the dormant season over Europe, the a priori flux uncertainties in this season are smaller, which is consistent with results from an ensemble study of global inversion models showing the smallest inter-model spread in this season (Peylin et al.,

10   2013).

The phase of the seasonal cycle seems consistent among the satellite inversions and CarbonTracker but this agreement should not be over interpreted because of the coarse monthly resolution and the month-to-month correlations regularising the inversion (see

15   Appendix A).

By means of a regional inversion study using 15 CarboEurope stations, Broquet et al. (2011) estimated the European summer uptake (June-September) to be in the range of about $1.4 - 3.6\,\mathrm{GtC/a}$ for 2003–2007. Our results indicate an uptake of $3.5 - 5.0\,\mathrm{GtC/a}$

20   within the summer months for the same period. As the cited study concentrated on a much smaller domain in western Europe of area $3.9 \cdot 10^{12}\,\mathrm{m}^2$, we have scaled their results to $1.0 \cdot 10^{13}\,\mathrm{m}^2$.

5   Discussion

25   In order to investigate the origin of the difference between the satellite $(0.90 \pm 0.33\,\mathrm{GtC/a}$ BESD baseline inversion in 2010) and CarbonTracker $(0.38 \pm 0.35\,\mathrm{GtC/a}$ for 2010) fluxes, we limited our analysis to satellite (BESD) soundings falling in a 350 km surrounding of

all surface stations assimilated by CarbonTracker (Fig. A1a). In this case we estimate a European carbon sink of $0.58 \pm 0.37\,\mathrm{GtC/a}$, which agrees with CarbonTracker (within the error bars). The satellites retrieve column-averages in contrast to the surface sites performing point measurements in the boundary layer. Therefore, the satellite measurements

5 still represent a larger part of Europe because footprints are wider spread, i.e., a one-to-one comparison with surface sites is not appropriate. Nevertheless, this experiment indicates that the underestimation of the European carbon sink in some inverse models may result from the sparse sampling of surface sites with no stations outside EU28 countries. This is consistent with the inversion studies of Bruhwiler et al. (2011) who

10 showed that the derived European carbon budget and its seasonal cycle can critically depend on the spatial coverage of the surface sites.

Some inversion models, assimilating only surface in situ measurements, find a large sink in the Eurasian boreal TRANSCOM region and a weaker sink in Europe. The

15 models shift the sink towards Europe when assimilating satellite XCO_2 measurements (Basu et al., 2013). A potential explanation is that the sparseness of surface sites hinder the models to discriminate between the European and Eurasian regions. This hypothesis is supported by a study of Schneising et al. (2011) using satellite XCO_2 data suggesting that the Eurasian boreal forests are a weaker sink than expected from CarbonTracker.

20

For two of the aggregation experiments (see Appendix B) we divided the European domain in two equally large parts i) at $27.4°E$ and ii) at $52.3°N$. The inferred fluxes indicate that eastern Europe $(53 \pm 31\%)$ may contribute more than western Europe $(47\pm30\%)$ and northern Europe $(66\pm31\%)$ more than southern Europe $(34\pm35\%)$ to the

25 overall European carbon sink. However the error estimates, which consider a posteriori error and ensemble spread, show that the differences are not significant.

Inverse modelling studies are the focus of this paper. However, recent findings reveal that carbon accumulation increases continuously with tree size (Stephenson et al., 2014).

9

This potentially contributes to explaining the discrepancy with bottom-up inventories. In this context, it should also be noted that the flux estimates of Schulze et al. (2009) concentrated on the period 2000–2005 including years (e.g., 2003) with little uptake in Europe (Ciais et al., 2005).

6 Validation

In order to validate our results, we use the optimised fluxes of the baseline inversion of BESD satellite data to simulate optimised concentrations. These concentrations as well as CarbonTracker concentrations are then compared with independent measurements, which have neither been inverted by CarbonTracker nor by us. For more details see Appendix C. The comparison with TCCON (Wunch et al., 2011, Total Carbon Column Observing Network) ground based FTS (Fourier transform spectrometer) column measurement shows that the optimised concentrations slightly improve the standard deviation of the difference and the seasonal cycle amplitude (Fig. C1). This is also the case when comparing with CONTRAIL (Machida et al., 2008, Comprehensive Observation Network for TRace gases by AIrLiner) in situ measurements aboard commercial aircraft. In an altitude range corresponding to $700-300$ hPa optimised and CarbonTracker concentrations are similar because the European surface influence is small. However, we observe an improvement of the seasonal cycle in the lower troposphere at $950-700$ hPa with differences of up to 0.86 ppm between optimised and CarbonTracker concentrations during the growing season (Fig. C2). A comparison with ground based in situ measurements of NOAA's cooperative air sampling network cannot be considered a validation because these measurements have been assimilated into CarbonTracker so that improved agreement cannot be expected. Nevertheless, this comparison is valuable to assess potential inconsistencies between the assimilated data sets. The overall agreement with the surface measurements slightly degrades. The seasonal cycle of the optimised concentrations shows marginal improvements in most of the months but also the largest discrepancy in one of the

10

months (Fig. C3).

7 Conclusions

In summary, our study reveals that the European terrestrial carbon sink appears considerably
5 larger than expected from bottom-up estimates and the majority of inverse models
assimilating in situ CO_2 atmospheric concentration measurements (Schulze et al., 2009;
Peylin et al., 2013; Peters et al., 2010; Chevallier et al., 2014). The addition of surface
measurement sites in the eastern part of Europe may in the future help to confirm our
findings with global mass-conserving inversion models. In addition, these models have
10 the potential to identify the origin of the carbon absorbed in Europe. New satellite
missions with more measurements, higher spatial resolution, precision, and accuracy
(e.g., OCO-2 or CarbonSat / CarbonSat Constellation, Crisp et al., 2004; Bovensmann et al.,
2010; Buchwitz et al., 2013a) have the potential to reduce the remaining uncertainties
especially during the dormant season. They will enable flux estimations at high spatial
15 resolution and contribute to improved process understanding on local, regional, and
global scales.

Appendix A: Inversion technique

This section describes our baseline inversion set-up corresponding to the bars and solid
20 lines within Fig. 1. Deviations from the baseline are described in Appendix B.

We use the optimal estimation formalism to infer optimised surface fluxes for the
European TRANSCOM region (from the Atlantic to the Urals, area $= 1.0 \cdot 10^{13}\,\mathrm{m}^2$,
Fig. A1a) during the period 2003–2010 on a monthly basis by minimising the cost

11

function:

$$\chi^2 = (\boldsymbol{y} - \boldsymbol{F}(\boldsymbol{x}))^T \mathbf{S}_\epsilon^{-1} (\boldsymbol{y} - \boldsymbol{F}(\boldsymbol{x}))$$
$$+ (\boldsymbol{x} - \boldsymbol{x}_a)^T \mathbf{S}_a^{-1} (\boldsymbol{x} - \boldsymbol{x}_a) \tag{A1}$$

The state vector $\boldsymbol{x}$ (containing the parameters of interest) consists of 96 monthly fluxes
5 at European scale and 96 monthly biases (2003–2010). The bias elements ensure that
the flux information is coming solely from inner-regional gradients and is insensitive to
seasonal, region-wide (constant) biases.

We use monthly CarbonTracker CT2011_oi (Peters et al., 2007) fluxes (resulting from
10 an inversion assimilating surface in situ measurements) as flux a priori and zero as
a priori biases ($\boldsymbol{x}_a$). Usually, monthly averages cannot be obtained directly because of
large regional (e.g., plant type) and temporal (e.g., day/night) flux variations. Such
variations are modelled by CarbonTracker so that it is sufficient for our application to
derive the monthly average flux deviation from CarbonTracker.
15

The a priori error covariance matrix $\mathbf{S}_a$ is constructed from (scaled) monthly CarbonTracker
flux uncertainties and a 100 ppm uncertainty for the monthly biases (rendering them
basically un-constrained). CarbonTracker uses a Kalman filter technique with a five-week
assimilation window which results in monthly flux uncertainties considered unrealistically
20 large. Consequently, 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. The resulting monthly
uncertainties (Fig. 1, bottom), with lowest values during the dormant season and largest
values during the growing season, agree reasonably well with the inter-model spread of an
25 ensemble of atmospheric CO_2 inversions (Peylin et al., 2013). Potential seasonal biases
are assumed to vary only slowly during the year and we assume an error correlation
length l of three months between the bias elements of $\mathbf{S}_a$ (the correlation between two
months i and j is computed by $e^{-|i-j|/l}$). In order to better constrain the inversion, we

add the same correlations to that part of $\mathbf{S}_a$ corresponding to the fluxes. A priori error correlations between the bias and flux elements are not assumed.

The measurement vector $\boldsymbol{y}$ comprises all XCO_2 measurements falling in the time period 2003–2010 and in the European TRANSCOM region (Fig. A1a). Averaging kernels have been applied to the measurements using co-located CarbonTracker profiles (Reuter et al., 2013). The forward model $\boldsymbol{F}$ is represented by the co-located CarbonTracker XCO_2 values.

The measurement error covariance matrix ($\mathbf{S}_\epsilon$) is constructed from the retrieval uncertainties. However, as some GOSAT data sets include unrealistically small errors, we scale them to match (on average) the single measurement precision determined from a validation exercise using the European TCCON sites shown in Fig. C1. We use the same validation approach as that of Reuter et al. (2013) and find the following scaling factors (single measurement precision from the validation / average reported error, both in ppm): BESD 2.25/2.39, ACOS 1.98/1.00, UoL-FP 2.50/1.06, RemoTeC 2.16/0.77, NIES 2.19/0.88. GOSAT measurements are regularly sampled with 150 km (260 km since 08/2010) distance and we assume no error correlations between them. In contrast, SCIAMACHY measurements can have direct neighbours thus error correlations become more likely (e.g., as a result of similar meteorological conditions). In order to make the error characteristics comparable with those of GOSAT measurements, we assume an error correlation length of 200 km (GOSAT's approximate average sampling distance in 2010) for measurements within one orbit. This introduces off-diagonal elements in $\mathbf{S}_\epsilon$; a matrix which can have more than $1.5 \cdot 10^5 \times 1.5 \cdot 10^5$ elements. Therefore, sparse matrix arithmetic and block wise inversion is required for the calculations.

Within this study, we concentrate on European fluxes, so that correlations with fluxes in other regions will not affect our results. Additionally, we can assume that the CarbonTracker fluxes already provide a good estimate. For these reasons, we use

13

CarbonTracker fluxes as a priori and first guess and assume linearity of the forward model. The a posteriori solution $\hat{x}$ minimising the cost function then results from a single correction to the first guess:

$$\hat{x} = x_a + \hat{S}\left[K^T S_\epsilon^{-1}\left(y - F(x_a)\right)\right] \tag{A2}$$

$$\hat{S} = \left(K^T S_\epsilon^{-1} K + S_a^{-1}\right)^{-1} \tag{A3}$$

Here $\hat{S}$ includes the a posteriori errors and their correlations and K is the Jacobian matrix which is calculated with the STILT (Gerbig et al., 2003; Lin et al., 2003) Lagrangian particle dispersion model. The XCO_2 increment expected from the optimised fluxes can be calculated from $K(\hat{x} - x_a)$ (Fig. A1c).

We use STILT to calculate global column-average footprints $\overline{R}$ (surface sensitivities in $ppm/(\mu molC\,m^{-2}\,s^{-1})$) with a resolution of $0.5° \times 0.5°$ for each satellite sounding (Fig. A2a-d shows an example). For this purpose, we split the atmospheric column into 40 sub-layers with variable (pressure) width Δp from which about 25 are placed in the planetary boundary layer, 12 in the free troposphere, and 3 in the stratosphere. One STILT receptor is placed at the centre of each layer. Each receptor is the starting point of 25 particles for which back-trajectories of 480 h are calculated forming the basis for the receptors footprint R. In order to avoid unnecessary calculations, we terminate particles leaving a $10°$ bounding box around Europe. Given the surface pressure p_0 and the column averaging kernel a, vertical integration is performed by:

$$\overline{R} = \sum_{i=1}^{40} \frac{\Delta p_i}{p_0} a_i R_i \tag{A4}$$

The European influence of a sounding, i.e., the corresponding element of the Jacobian matrix K (in $ppm/(GtC\,a^{-1})$), is calculated by integration over the European TRANSCOM region and unit conversion (Fig. A2e shows an example). We use

14

NCEP/NCAR reanalysis as driving meteorological fields.

The annual flux average of year y is calculated from the inversion results by $\bar{\boldsymbol{x}}_y = \boldsymbol{w}^T \hat{\boldsymbol{x}}$. Here the elements of the weighting vector $\boldsymbol{w}$ are $1/12$ for the twelve monthly fluxes corresponding to y and zero elsewhere.

Appendix B: Error analysis and ensemble set-up

A posteriori error. The a posteriori error estimates are calculated from Eq. A2 accounting for the error estimates of the satellite retrievals $\mathbf{S}_a$ and the a priori flux uncertainties $\mathbf{S}_\epsilon$. The uncertainty of the annual flux average is given by $\sigma_{\bar{\boldsymbol{x}}_y} = \left(\boldsymbol{w}^T \hat{\mathbf{S}} \boldsymbol{w} \right)^{1/2}$ and amounts to $0.15 \, \text{GtC/a}$ for BESD (multi-year average), $0.14 \, \text{GtC/a}$ for ACOS (2010), $0.15 \, \text{GtC/a}$ for UoL-FP (2010), $0.14 \, \text{GtC/a}$ for RemoTeC (2010), and $0.19 \, \text{GtC/a}$ for NIES. However, this uncertainty estimate is incomplete and additional error terms are considered via an ensemble approach as explained in the following.

Background model. 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, ♣). The resulting annual fluxes are consistent with the results based on CarbonTracker to which they have a root mean square difference (RMSD) of $0.22 \, \text{GtC/a}$.

Convection. Inaccuracies in the parametrisation of convection in STILT are an additional potential error source. In order to assess this potential uncertainty, we modified the baseline set-up and used the Convective Available Potential Energy

15

(CAPE) parametrisation instead of the modelled vertical wind speeds and re-processed the footprints for all five satellite retrievals in 2010 (Fig. 1, ♦). The resulting annual fluxes deviate by a RMSD of 0.09 GtC/a. Thus, convection is unlikely to explain the observed carbon sink.

Stilt set-up. Other STILT set-up parameters may also influence the results. Experiments using finer receptor grids and/or more particles showed negligible influence on the accumulated European surface influence. The used integration time also seems to be sufficient because flux results level off well before 480 h (Fig. B1a). This justifies the assumption that the vast majority of particles have left Europe within this time.

Meteorology. Global inversion studies showed that the atmospheric transport model can significantly impact the inferred surface fluxes (Gurney et al., 2002). We expect that our regional set-up does not critically depend on long-range transport errors for two main reasons: i) Air masses leave the analysis-region typically within a few days. ii) The analysis relies on accumulated European surface influences, i.e., the exact pattern of the surface influence is less important. The inversion results of all five satellite retrievals in 2010 using the ERA Interim reanalysis (Fig. 1, ♥) have a RMSD of 0.32 GtC/a compared to the results based on the NCEP/NCAR reanalysis.

Aggregation error. As described before, we derive Europe-wide, monthly flux increments. This could be interpreted as "hard constraint" possibly resulting in temporal and/or spatial aggregation errors (Kaminski et al., 2001). Engelen et al. (2002) estimated this effect for the European TRANSCOM region to be $0.13 - 0.31$ GtC/a depending on the used flux fields when inverting sparsely sampled in situ measurements. We expect that aggregation errors are less pronounced in our case because i) spatio-temporal patterns of the used a priori fluxes (CarbonTracker) are assumed to be relatively realistic and ii) the inverted satellite data are considerably more densely

16

sampled than in situ measurement sites. Nevertheless, we analysed this potential error component by undertaking three experiments. Each of which divided the domain into two equally large parts i) at $27.4°E$, ii) at $52.3°N$, and iii) at the middle of each months. Within each sub-domain, we derived the surface flux and the bias and aggregated the sub-domains afterwards. The a priori error covariance has been apportioned accordingly so that the a posteriori error statistics of the aggregated domains remained similar. The inversion results differ from the baseline by $0.12\,\mathrm{GtC/a}$ (RMSD, longitude split), $0.06\,\mathrm{GtC/a}$ (RMSD, latitude split), and $0.04\,\mathrm{GtC/a}$ (RMSD, temporal split), respectively. The average of the three experiments differs by a RMSD of $0.08\,\mathrm{GtC/a}$ from the baseline (Fig. 1, ♠).

Regional biases. Even though our regional inversion scheme is insensitive to retrieval biases outside Europe, it could in principle still suffer from retrieval biases within Europe arising, e.g., from persistent aerosol or cloud patterns, surface albedo, or chlorophyll fluorescence. However, this would only be the case if biases were correlated with the surface sensitivity. Europe's weather is complex and characterised by alternating high and low pressure systems induced by Rossby waves with west winds dominating on average. This means, measurements in eastern Europe will tend to have a larger sensitivity to the European surface flux. Therefore, a hypothetical retrieval bias with east/west gradient would correlate with the surface sensitivity and affect the inversion results. A retrieval bias producing larger XCO_2 values in eastern Europe could be misinterpreted as a CO_2 source for west wind conditions. Conversely it would be misinterpreted as sink under east wind conditions. Therefore, we built two sub-samples of the BESD data set containing only measurements with planetary boundary layer winds in west or east direction, respectively. The resulting annual average surface fluxes of both sub-samples have only a small difference of $0.06 \pm 0.19\,\mathrm{GtC/a}$. Sub-sampling for boundary layer winds in north and south direction results in a small difference of $0.03 \pm 0.19\,\mathrm{GtC/a}$. The uncertainty estimates represents the standard deviation over the analysed years. In case of persistent bias patterns, the sign of the flux error

changes with wind direction so that it appears twice in the calculated differences. We estimate that the annual surface flux error due to persistent biases amounts $0.02\,\mathrm{GtC/a}\,(\approx 1/4\,(0.03\,\mathrm{GtC/a} + 0.06\,\mathrm{GtC/a}))$. Therefore, retrieval biases in mean wind direction are unlikely to explain the observed carbon sink.

5

Differences of the retrieval algorithms. Additionally, it is valuable to recall the differences in the retrieval algorithms used for this study: i) The retrievals use different spectral fitting windows and spectroscopy, ii) different cloud and aerosol screening techniques are used resulting in different samplings, iii) light scattering related errors
10 are corrected differently by the retrievals' full physics schemes being optimised for clouds and/or aerosols, iv) the surface albedo is handled differently, v) chlorophyll fluorescence is explicitly accounted for by only one of the retrievals (ACOS), vi) one retrieval uses no empirical bias correction (NIES). See also Reuter et al. (2013) for a summary of differences. Overall it seems unlikely that five independently developed
15 retrieval algorithms optimised for two different sensors produce the same bias patterns.

Statistical set-up. Inherent to Bayesian inversion systems, the statistical set-up (a priori errors and their covariances) influences the a posteriori solution. We generate the a priori error statistics by scaling monthly CarbonTracker flux uncertainties with a
20 factor of $1/3$ and assuming error correlation lengths (for bias and fluxes) of three months. The resulting seasonal cycle flux uncertainty and the annual average flux uncertainty agree reasonably well with values found by Basu et al. (2013); Chevallier et al. (2014); Peylin et al. (2013) (see discussion in Appendix A). Nevertheless, we examined the influence of error scaling (Fig. B1b) and correlation lengths (Fig. B1c) on the results
25 of the baseline inversion. The analysed error scaling factors range from 0.20 to 0.45 so that the corresponding a priori uncertainties vary over a relatively large range from about $\pm 0.2\,\mathrm{GtC/a}$ to $\pm 0.5\,\mathrm{GtC/a}$. All inverted annual fluxes deviate by less than $\pm 0.1\,\mathrm{GtC/a}$ from the baseline. The same is true for the vast majority of fluxes inferred

18

for error correlation lengths which range from one to five months.

Combining the error contributions. Fig. 1 (top) shows the results of an ensemble of different inversion set-ups. The ensemble members quantify the expected departures of the annual fluxes from the baseline inversion with respect to those properties which are believed to have the largest influence on the inversion results (i.e., background model, meteorology, convection parametrisation, and aggregation set-up). The individual aspects are analysed separately so that the ensemble spread may not be a reliable measure for the uncertainty because permutations are missing. We approximate the missing permutations (16 permutations times 5 satellite retrievals in total) by linear combinations of the individual departures and calculate the median carbon sink (1.02 GtC/a) and its standard deviation (0.28 GtC/a). Using only the 25 ensemble members (and ignoring all permutations) would result in an average of 0.93 GtC/a and a standard deviation of 0.19 GtC/a. Additional error components such as regional biases, statistical set-up, etc., are assumed to contribute to a lesser extent. For convenience, we consider them by adding 0.10 GtC/a (via summation of variances) so that the overall uncertainty of the ensemble median is estimated to be 0.30 GtC/a (Fig. 1, dotted area). The uncertainty due to different samplings of the satellite retrievals is implicitly accounted for because part of the inter-algorithm differences. If not otherwise noted, all uncertainty estimates of annual fluxes within this paper include this additional uncertainty, e.g., the annual flux uncertainties of the individual satellite algorithms (Fig. 1, error bars) include the a posteriori error and the estimated additional uncertainty. Monthly uncertainty estimates correspond to un-modified a posteriori error estimates.

Appendix C: Validation

Based on the optimised fluxes $\hat{f}$ of the baseline inversion of BESD satellite data, we simulate optimised concentrations $\hat{c}$ and compare them and CarbonTracker concentrations c_a with independent measurements, which have neither been inverted
5 by CarbonTracker nor by us. For this purpose, we calculate the European influence k, i.e., the Jacobian for each measurement and multiply it with the derived flux increment $\Delta f = \hat{f} - f_a$ where f_a represents CarbonTracker fluxes.

$$\hat{c} = c_a + k\,\Delta f \tag{C1}$$

Except for a potential offset, the simulated concentrations ideally agree better with the
10 measurements than the CarbonTracker concentrations. Vertical integration (in case of column measurements) is performed as described in Appendix A.

The Total Carbon Column Observing Network (Wunch et al., 2011, TCCON) uses ground based Fourier transform spectrometers (FTS) to derive the dry-air
15 column-average mole fraction of CO_2 (and other gases). Six TCCON sites are located within the European TRANSCOM region (Fig. C1g) namely Białystok (Poland), Bremen (Germany), Garmisch-Partenkirchen (Germany), Karlsruhe (Germany), Orléans (France), Sodankylä (Finland). We limit the validation period to 2010 because four of the six sites started operation in 2009. In order to reduce the amount of
20 computational expensive footprint calculations, we only use those TCCON measurement per day and site being closest to 12h UTC. Fig. C1a–f shows these measurements overlaid by CarbonTracker and optimised concentrations. The differences between both are generally small (0.22 ppm standard deviation) and largest in summer when the flux increment is maximal. The overall standard deviation of the difference to
25 TCCON marginally improves from 1.13 ppm to 1.11 ppm. Station-to-station biases (Fig. C1h) tend to improve but the differences are below 0.1 ppm, which is smaller than TCCON's network accuracy of 0.4 ppm (Wunch et al., 2011). Consistent with the studies of Reuter et al. (2011); Keppel-Aleks et al. (2012), CarbonTracker has a too

20

small seasonal cycle amplitude at TCCON sites in northern mid latitudes (Fig. C1i). The seasonal cycle of the optimised concentrations is in slightly better agreement with TCCON: eleven months show an improvement by about 0.2 ppm (difference to CarbonTracker).

The Comprehensive Observation Network for TRace gases by AIrLiner (Machida et al., 2008, CONTRAIL) uses commercial aircraft as platform for highly accurate and precise in situ measurements of atmospheric CO_2 and other species. In order to prevent the instruments from damage caused by polluted air masses, no measurements are performed below approximately 2000 ft. Within the period 2003–2010 the cities London, Milan, Moscow, and Paris were frequent destinations in Europe. Level flights are typically performed in altitudes above 10 km. The footprints of measurements in high altitudes are wide spread and back-trajectories usually do not collect significant European surface influence within 480 hours. Therefore, we concentrate on ascents and descents; Fig. C2i shows the position of all analysed measurements. Fig. C2a-d compares CarbonTracker as well as optimised concentrations with CONTRAIL measurements near London, Milan, Moscow, and Paris within an altitude range corresponding to $700 - 300$ hPa. The optimised concentrations agree well with CarbonTracker concentrations and CONTRAIL concentrations (0.10 ppm standard deviation). This is not surprising because the European surface influence is already small in these altitudes. Lower in the atmosphere, the differences between optimised and CarbonTracker concentrations become larger (Fig. C2e-h, 0.42 ppm standard deviation). Station-to-station biases (Fig. C2j) as well as the seasonal cycle in $700 - 300$ hPa (Fig. C2l) are very similar with differences always below 0.07 ppm. However, we observe a distinct improvement of the seasonal cycle in the lower troposphere at $950 - 700$ hPa (Fig. C2k) with differences of up to 0.86 ppm between optimised and CarbonTracker concentrations during the growing season.

NOAA's cooperative air sampling network performs highly accurate and precise ground based in situ measurements of atmospheric CO_2 (and other species) which

21

are the backbone of the CarbonTracker assimilation system. During the inversion procedure, CarbonTracker modifies its surface fluxes so that an optimal (Bayesian) fit is achieved between simulated and measured atmospheric concentrations. Any additional constraints (e.g., the satellite data) have only the potential to degrade the agreement
with the surface concentrations; an improvement is almost impossible. Therefore, a fair validation is only possible with independent measurements, which have not been assimilated. Nevertheless, a comparison with in situ measurements is valuable to assess potential inconsistencies between the assimilated data sets. Fig. C3i shows the position of the European measurement sites assimilated in CarbonTracker in 2010. Compared
to the relatively large seasonal cycle amplitude Fig. C3a-h, the difference between optimised and CarbonTracker concentrations is small (0.80 ppm standard deviation). The largest concentration increments can be found at sites with relatively large European surface influence (e.g., Hegyhatsal, Hungary, Fig. C3e) while the smallest increments are found upwind (e.g., Mace Head, Ireland). As expected, the overall
agreement with the surface measurements slightly degrades (5.88 ppm vs. 5.96 ppm, standard deviation of the difference). Station-to-station biases of the optimised and the CarbonTracker concentrations are very similar and the most pronounced feature is the large underestimation at the Black Sea site in Romania (Fig. C3j). The seasonal cycle at this site seems to be overlaid by frequent positive outliers which may explain its
discontinuation (Fig. C3b). The average seasonal cycle of the optimised concentrations is in marginally (0.05 − 0.66 ppm, difference to CarbonTracker) better agreement with the surface sites during eight months (Fig. C3k). However, the remaining months are in less good agreement and the largest departure from CarbonTracker concentrations (1.22 ppm) is found in June.

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,
[20] 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, http://www.atmos-chem-phys.net/13/8695/2013/, 2013.

Bergamaschi, P., Houweling, S., Segers, A., Krol, M., Frankenberg, C., Scheepmaker, R. A., Dlugokencky, E., Wofsy, S. C., Kort, E. A., Sweeney, C., Schuck, T., Brenninkmeijer, C., Chen,
[25] H., Beck, V., and Gerbig, C.: Atmospheric CH_4 in the first decade of the 21st century: Inverse modeling analysis using SCIAMACHY satellite retrievals and NOAA surface measurements, Journal of Geophysical Research: Atmospheres, 118, 7350–7369, doi:10.1002/jgrd.50480, http://dx.doi.org/10.1002/jgrd.50480, 2013.

Bovensmann, H., Burrows, J. P., Buchwitz, M., Frerick, J., Noël, S., Rozanov,
[30] V. V., Chance, K. V., and Goede, A.: SCIAMACHY – Mission Objectives

23

and Measurement Modes, Journal of the Atmospheric Sciences, 56, 127–150, http://dx.doi.org/10.1175/1520-0469(1999)056<0127:SMOAMM>2.0.CO;2, 1999.

Bovensmann, H., Buchwitz, M., Burrows, J. P., Reuter, M., Krings, T., Gerilowski, K., Schneising, O., Heymann, J., Tretner, A., and Erzinger, J.: A remote sensing technique for global monitoring of power plant CO_2 emissions from space and related applications, Atmospheric Measurement Techniques, 3, 781–811, doi:10.5194/amt-3-781-2010, http://www.atmos-meas-tech.net/3/781/2010/, 2010.

Broquet, G., Chevallier, F., Rayner, P., Aulagnier, C., Pison, I., Ramonet, M., Schmidt, M., Vermeulen, A. T., and Ciais, P.: A European summertime CO_2 biogenic flux inversion at mesoscale from continuous in situ mixing ratio measurements, Journal of Geophysical Research, 116, doi:10.1029/2011jd016202, http://dx.doi.org/10.1029/2011JD016202, 2011.

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, http://dx.doi.org/10.5194/bg-8-1309-2011, 2011.

Buchwitz, M., Reuter, M., Bovensmann, H., Pillai, D., Heymann, J., Schneising, O., Rozanov, V., Krings, T., Burrows, J. P., Boesch, H., Gerbig, C., Meijer, Y., and Löscher, A.: Carbon Monitoring Satellite (CarbonSat): assessment of atmospheric CO_2 and CH_4 retrieval errors by error parameterization, Atmospheric Measurement Techniques, 6, 3477–3500, doi:10.5194/amt-6-3477-2013, http://www.atmos-meas-tech.net/6/3477/2013/, 2013a.

Buchwitz, M., Reuter, M., Schneising, O., Boesch, H., Guerlet, S., Dils, B., Aben, I., Armante, R., Bergamaschi, P., Blumenstock, T., Bovensmann, H., Brunner, D., Buchmann, B., Burrows, J., Butz, A., Chédin, A., Chevallier, F., Crevoisier, C., Deutscher, N., Frankenberg, C., Hase, F., Hasekamp, O., Heymann, J., Kaminski, T., Laeng, A., Lichtenberg, G., Mazière, M. D., Noël, S., Notholt, J., Orphal, J., Popp, C., Parker, R., Scholze, M., Sussmann, R., Stiller, G., Warneke, T., Zehner, C., Bril, A., Crisp, D., Griffith, D., Kuze, A., O'Dell, C., Oshchepkov, S., Sherlock, V., Suto, H., Wennberg, P., Wunch, D., Yokota, T., and Yoshida, Y.: The Greenhouse Gas Climate Change Initiative (GHG-CCI): Comparison and quality assessment of near-surface-sensitive satellite-derived CO_2 and CH_4 global data sets, Remote Sensing of Environment, doi:http://dx.doi.org/10.1016/j.rse.2013.04.024, http://www.sciencedirect.com/science/article/pii/S0034425713003520, 2013b.

Butz, A., Guerlet, S., Hasekamp, O., Schepers, D., Galli, A., Aben, I., Frankenberg, C., Hartmann, J. M., Tran, H., Kuze, A., Keppel-Aleks, G., Toon, G., Wunch, D., Wennberg, P., Deutscher, N., Griffith, D., Macatangay, R., Messerschmidt, J., Notholt, J., and Warneke, T.:

24

Toward accurate CO_2 and CH_4 observations from GOSAT, Geophysical Research Letters, 38, doi:10.1029/2011GL047888, http://dx.doi.org/10.1029/2011GL047888, 2011.

Chevallier, F., Bréon, F.-M., and Rayner, P. J.: Contribution of the Orbiting Carbon Observatory to the estimation of CO_2 sources and sinks: Theoretical study in a variational data assimilation framework, Journal of Geophysical Research, 112, D09307, doi:10.1029/2006JD007375, http://dx.doi.org/10.1029/2006JD007375, 2007.

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, http://dx.doi.org/10.1002/2013GL058772, 2014.

Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogée, J., Allard, V., Aubinet, M., Buchmann, N., Bernhofer, C., Carrara, A., Chevallier, F., de Noblet, N., Friend, A. D., Friedlingstein, P., Grünwald, T., Heinesch, B., Keronen, P., Knohl, A., Krinner, G., Loustau, D., Manca, G., Matteucci, G., Miglietta, F., Ourcival, J. M., Papale, D., Pilegaard, K., Rambal, S., Seufert, G., Soussana, J. F., Sanz, M. J., Schulze, E. D., Vesala, T., and Valentini, R.: Europe-wide reduction in primary productivity caused by the heat and drought in 2003, Nature, 437, 529–533, doi:10.1038/nature03972, http://www.nature.com/nature/journal/v437/n7058/abs/nature03972.html, 2005.

Cogan, A. J., Boesch, H., Parker, R. J., Feng, L., Palmer, P. I., Blavier, J. F. L., Deutscher, N. M., Macatangay, R., Notholt, J., Roehl, C., Warneke, T., and Wunch, D.: Atmospheric carbon dioxide retrieved from the Greenhouse gases Observing SATellite (GOSAT): Comparison with ground-based TCCON observations and GEOS-Chem model calculations, Journal of Geophysical Research: Atmospheres, 117, doi:10.1029/2012JD018087, http://dx.doi.org/10.1029/2012JD018087, 2012.

Crisp, D., Atlas, R. M., Bréon, F.-M., Brown, L. R., Burrows, J. P., Ciais, P., Connor, B. J., Doney, S. C., Fung, I. Y., Jacob, D. J., Miller, C. E., O'Brien, D., Pawson, S., Randerson, J. T., Rayner, P., Salawitch, R. S., Sander, S. P., Sen, B., Stephens, G. L., Tans, P. P., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Yung, Y. L., Kuang, Z., Chudasama, B., Sprague, G., Weiss, P., Pollock, R., Kenyon, D., and Schroll, S.: The Orbiting Carbon Observatory (OCO) mission, Advances in Space Research, 34, 700–709, 2004.

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, http://dx.doi.org/10.1029/2002JD002195, 2002.

25

Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Andrews, A. E., Stephens, B. B., Bakwin, P. S., and Grainger, C. A.: Toward constraining regional-scale fluxes of CO_2 with atmospheric observations over a continent: 2. Analysis of COBRA data using a receptor-oriented framework, Journal of Geophysical Research: Atmospheres, 108, doi:10.1029/2003JD003770, http://dx.doi.org/10.1029/2003JD003770, 2003.

Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, P., Fan, S., Fung, I. Y., Gloor, M., Heimann, M., Higuchi, K., John, J., Maki, T., Maksyutov, S., Masarie, K., Peylin, P., Prather, M., Pak, B. C., Randerson, J., Sarmiento, J., Taguchi, S., Takahashi, T., and Yuen, C.-W.: Towards robust regional estimates of CO_2 sources and sinks using atmospheric transport models, Nature, 415, 626–630, doi:10.1038/415626a, 2002.

Kaminski, T., Rayner, P. J., Heimann, M., and Enting, I. G.: On aggregation errors in atmospheric transport inversions, Journal of Geophysical Research, 106, 4703, doi:10.1029/2000jd900581, http://dx.doi.org/10.1029/2000JD900581, 2001.

Keppel-Aleks, G., Wennberg, P. O., Washenfelder, R. A., Wunch, D., Schneider, T., Toon, G. C., Andres, R. J., Blavier, J. F., Connor, B., Davis, K. J., Desai, A. R., Messerschmidt, J., Notholt, J., Roehl, C. M., Sherlock, V., Stephens, B. B., Vay, S. A., and Wofsy, S. C.: The imprint of surface fluxes and transport on variations in total column carbon dioxide, Biogeosciences, 9, 875–891, doi:10.5194/bg-9-875-2012, 2012.

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

Le Quéré, C., Andres, R. J., Boden, T., Conway, T., Houghton, R. A., House, J. I., Marland, G., Peters, G. P., van der Werf, G. R., Ahlström, A., Andrew, R. M., Bopp, L., Canadell, J. G., Ciais, P., Doney, S. C., Enright, C., Friedlingstein, P., Huntingford, C., Jain, A. K., Jourdain, C., Kato, E., Keeling, R. F., Klein Goldewijk, K., Levis, S., Levy, P., Lomas, M., Poulter, B., Raupach, M. R., Schwinger, J., Sitch, S., Stocker, B. D., Viovy, N., Zaehle, S., and Zeng, N.: The global carbon budget 1959-2011, Earth System Science Data, 5, 165–185, doi:10.5194/essd-5-165-2013, http://www.earth-syst-sci-data.net/5/165/2013/, 2013.

Lin, J. C., Gerbig, C., Wofsy, S. C., Andrews, A. E., Daube, B. C., Davis, K. J., and Grainger, C. A.: A near-field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT)

26

model, Journal of Geophysical Research: Atmospheres, 108, doi:10.1029/2002JD003161, http://dx.doi.org/10.1029/2002JD003161, 2003.

Machida, T., Matsueda, H., Sawa, Y., Nakagawa, Y., Hirotani, K., Kondo, N., Goto, K., Nakazawa, T., Ishikawa, K., and Ogawa, T.: Worldwide Measurements of Atmospheric CO_2
5 and Other Trace Gas Species Using Commercial Airlines, J. Atmos. Oceanic Technol., 25, 1744–1754, doi:10.1175/2008jtecha1082.1, http://dx.doi.org/10.1175/2008JTECHA1082.1, 2008.

Miller, C. E., Crisp, D., DeCola, P. L., Olsen, S. C., Randerson, J. T., Michalak, A. M., Alkhaled, A., Rayner, P., Jacob, D. J., Suntharalingam, P., Jones, D. B. A., Denning, A. S., Nicholls,
10 M. E., Doney, S. C., Pawson, S., Bösch, H., Connor, B. J., Fung, I. Y., O'Brien, D., Salawitch, R. J., Sander, S. P., Sen, B., Tans, P., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Yung, Y. L., and Law, R. M.: Precision requirements for space-based X_{CO_2} data, Journal of Geophysical Research, 112, D10314, doi:10.1029/2006JD007659, 2007.

Nassar, R., Jones, D. B. A., Kulawik, S. S., Worden, J. R., Bowman, K. W.,
15 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, http://www.atmos-chem-phys.net/11/6029/2011/, 2011.

20 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, www.atmos-chem-phys-discuss.net/14/C8037/2014/, 2014.

O'Dell, C. W., Connor, B., Bösch, H., O'Brien, D., Frankenberg, C., Castano, R., Christi,
25 M., Eldering, D., Fisher, B., Gunson, M., McDuffie, J., Miller, C. E., Natraj, V., Oyafuso, F., Polonsky, I., Smyth, M., Taylor, T., Toon, G. C., Wennberg, P. O., and Wunch, D.: The ACOS CO_2 retrieval algorithm - Part 1: Description and validation against synthetic observations, Atmospheric Measurement Techniques, 5, 99–121, doi:10.5194/amt-5-99-2012, http://www.atmos-meas-tech.net/5/99/2012/, 2012.

30 Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M. P., Petron, 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, Proceedings of

27

the National Academy of Sciences of the United States of America, 104, 18 925–18 930, doi:10.1073/pnas.0708986104, http://www.pnas.org/content/104/48/18925.abstract, updates documented at http://carbontracker.noaa.gov, 2007.

Peters, W., Krol, M. C., van der Werf, G. R., Houweling, S., Jones, C. D., Hughes, J.,
5 Schaefer, K., Masarie, K. A., Jacobson, A. R., Miller, J. B., Cho, C. H., Ramonet, M.,
Schmidt, M., Ciattaglia, L., Apadula, F., Helta, D., Meinhardt, F., di Sarra, A. G.,
Piacentino, S., Sferlazzo, D., Aalto, T., Hatakka, J., Strom, J., Haszpra, L., Meijer, H.
A. J., van der Laan, S., Neubert, R. E. M., Jordan, A., Rodo, X., Morgui, J. A.,
Vermeulen, A. T., Popa, E., Rozanski, K., Zimnoch, M., Manning, A. C., Leuenberger,
10 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, Global Change Biology, 16, 1317–1337, doi:10.1111/j.1365-2486.2009.02078.x,
http://dx.doi.org/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,
15 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,
http://www.biogeosciences.net/10/6699/2013/, 2013.

Reuter, M., Buchwitz, M., Schneising, O., Heymann, J., Bovensmann, H., and Burrows,
20 J. P.: A method for improved SCIAMACHY CO_2 retrieval in the presence of optically
thin clouds, Atmospheric Measurement Techniques, 3, 209–232, doi:10.5194/amt-3-209-2010,
http://dx.doi.org/10.5194/amt-3-209-2010, 2010.

Reuter, M., Bovensmann, H., Buchwitz, M., Burrows, J. P., Connor, B. J., Deutscher, N. M.,
Griffith, D. W. T., Heymann, J., Keppel-Aleks, G., Messerschmidt, J., Notholt, J., Petri,
25 C., Robinson, J., Schneising, O., Sherlock, V., Velazco, V., Warneke, T., Wennberg, P. O.,
and Wunch, D.: Retrieval of atmospheric CO_2 with enhanced accuracy and precision from
SCIAMACHY: Validation with FTS measurements and comparison with model results.,
J. Geophys. Res., 116, doi:10.1029/2010JD015047, http://dx.doi.org/10.1029/2010JD015047,
2011.

30 Reuter, M., Bösch, H., Bovensmann, H., Bril, A., Buchwitz, M., Butz, A., Burrows, J. P.,
O'Dell, C. W., Guerlet, S., Hasekamp, O., Heymann, J., Kikuchi, N., Oshchepkov, S.,
Parker, R., Pfeifer, S., Schneising, O., Yokota, T., and Yoshida, Y.: A joint effort to
deliver satellite retrieved atmospheric CO_2 concentrations for surface flux inversions: the

28

ensemble median algorithm EMMA, Atmospheric Chemistry and Physics, 13, 1771–1780, doi:10.5194/acp-13-1771-2013, http://www.atmos-chem-phys.net/13/1771/2013/, 2013.

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

Schneising, O., Reuter, M., Buchwitz, M., Heymann, J., Bovensmann, H., and Burrows, J. P.: Terrestrial carbon sink observed from space: variation of growth rates and seasonal cycle amplitudes in response to interannual surface temperature variability, Atmospheric Chemistry and Physics, 14, 133–141, doi:10.5194/acp-14-133-2014, http://www.atmos-chem-phys.net/14/133/2014/, 2014.

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., Roedenbeck, C., Etiope, G., Vuichard, N., Wattenbach, M., Nabuurs, G. J., Poussi, Z., Nieschulze, J., Gash, J. H., and Team, C.: Importance of methane and nitrous oxide for Europe's terrestrial greenhouse-gas balance, Nature Geoscience, 2, 842–850, doi:10.1038/ngeo686, http://dx.doi.org/10.1038/ngeo686, 2009.

Stephenson, N. L., Das, A. J., Condit, R., Russo, S. E., Baker, P. J., Beckman, N. G., Coomes, D. A., Lines, E. R., Morris, W. K., Ruger, N., Alvarez, E., Blundo, C., Bunyavejchewin, S., Chuyong, G., Davies, S. J., Duque, A., Ewango, C. N., Flores, O., Franklin, J. F., Grau, H. R., Hao, Z., Harmon, M. E., Hubbell, S. P., Kenfack, D., Lin, Y., Makana, J.-R., Malizia, A., Malizia, L. R., Pabst, R. J., Pongpattananurak, N., Su, S.-H., Sun, I.-F., Tan, S., Thomas, D., van Mantgem, P. J., Wang, X., Wiser, S. K., and Zavala, M. A.: Rate of tree carbon accumulation increases continuously with tree size, Nature, 507, 90+, doi:10.1038/nature12914, http://dx.doi.org/10.1038/nature12914, 2014.

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, http://dx.doi.org/10.1002/2013GL059174, 2014.

Wunch, D., Toon, G. C., Blavier, J.-F. L., Washenfelder, R. A., Notholt, J., Connor, B. J., Griffith, D. W. T., Sherlock, V., and Wennberg, P. O.: The Total Carbon Column Observing Network (TCCON), Philosophical Transactions of the Royal Society

of London, Series A: Mathematical, Physical and Engineering Sciences, 369, 2087–2112, doi:10.1098/rsta.2010.0240, http://dx.doi.org/10.1098/rsta.2010.0240, 2011.

Yoshida, Y., Kikuchi, N., Morino, I., Uchino, O., Oshchepkov, S., Bril, A., Saeki, T., Schutgens, N., Toon, G. C., Wunch, D., and et al.: Improvement of the retrieval algorithm for GOSAT SWIR XCO_2 and XCH_4 and their validation using TCCON data, Atmospheric Measurement Techniques, 6, 1533–1547, doi:10.5194/amt-6-1533-2013, http://dx.doi.org/10.5194/amt-6-1533-2013, 2013.
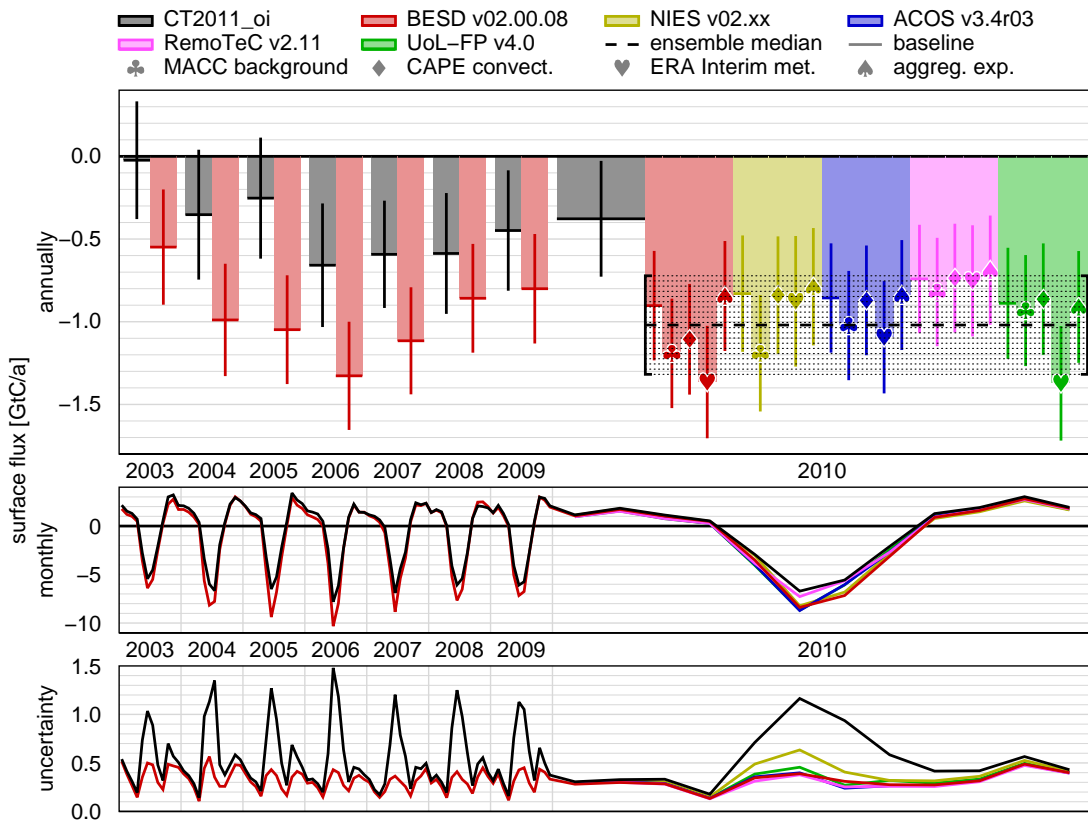
Figure 1. European biospheric surface fluxes (land excluding fossil and fire) from CarbonTracker and an ensemble of five satellite inversions (BESD, ACOS, UoL-FP, RemoTeC, and NIES) and five different inversion set-ups. The baseline uses CT2011_oi as background model, modelled vertical wind speeds as convection, NCEP/NCAR reanalysis meteorology, and an aggregation area which equals the European TRANSCOM region (Appendix A); each other inversion set-up differs by one of these properties (Appendix B). **top**: Annual averages and 1-sigma uncertainties (a posteriori and additional uncertainties, see Appendix B) as well as the ensemble median and its uncertainty (dotted area, see Appendix B). **middle**: Monthly averages of the baseline inversions; **bottom**: monthly uncertainties of the baseline inversion (1-sigma, as derived from the inversion scheme not including additional error components). Note that CarbonTracker uncertainties have been scaled (see Appendix A).
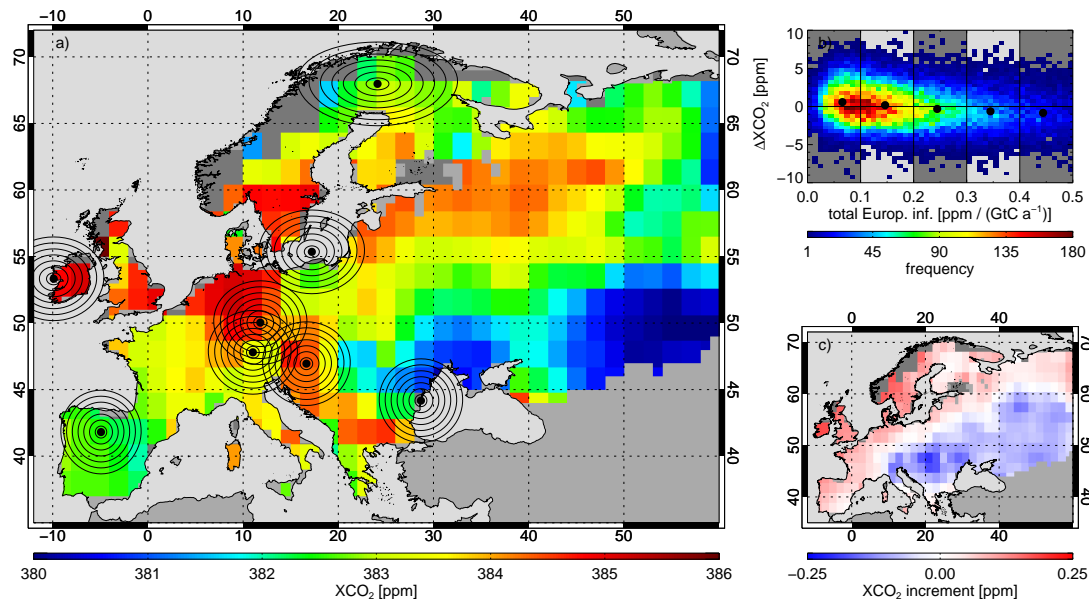
Figure A1. BESD multi-year (2003–2010) seasonal (June–August) statistics. **a)** Average satellite retrieved XCO_2 calculated from annual seasonal anomalies in order to minimise effects due to different annual samplings. Gridded to a regular $2° \times 2°$ grid and smoothed with a Hann function with $5° \times 5°$ effective width. Circles mark the European measurement sites of NOAA's air sampling network assimilated in CarbonTracker in 2010 and a 350 km surrounding. Areas in colour or dark grey represent the European TRANSCOM region, medium grey other land surfaces, and sea surfaces are light grey. **b)** Average relationship between European surface flux influence calculated with STILT and ΔXCO_2 (BESD-CarbonTracker). Black dots correspond to bin averages. Note that individual years and months may differ. **c)** Average XCO_2 increment anomaly expected from the optimised fluxes and calculated by $\mathbf{K}\,(\hat{\boldsymbol{x}} - \boldsymbol{x}_a)$ (see Appendix A).

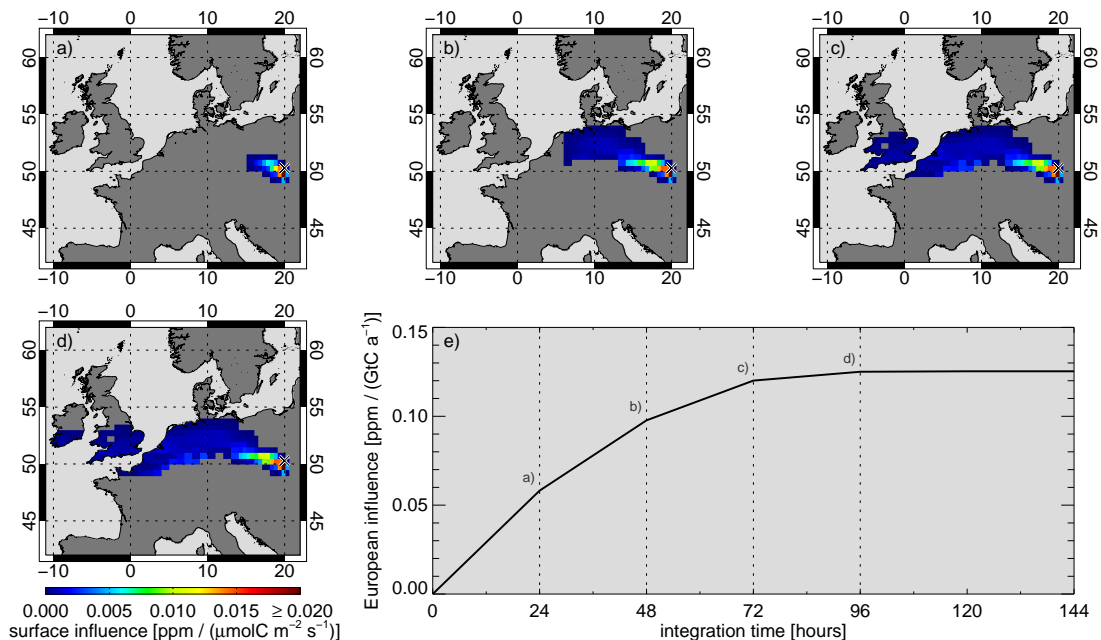**Figure A2.** Column-average footprint with a resolution of $0.5° \times 0.5°$ for a typical BESD sounding (August 21, 2010, 9:33 UTC, 20°E, 50°N) accumulated during **a)** 24h, **b)** 48h, **c)** 72h, and **d)** 96h before the sounding. Note, the colour scale is clipped at $0.02\,\text{ppm}/(\mu\text{molC\,m}^{-2}\text{s}^{-1})$ and the maximum value is $0.07\,\text{ppm}/(\mu\text{molC\,m}^{-2}\text{s}^{-1})$. **e)** Temporal evolution of the corresponding total accumulated European surface influence.
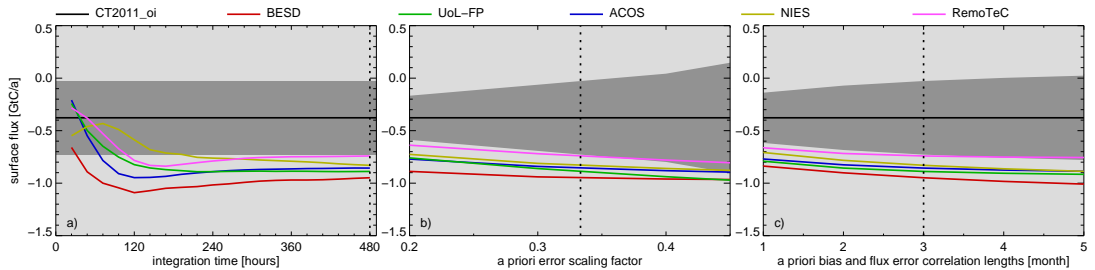
Figure B1. Integration time and statistical set-up. Annual average European surface fluxes for the five baseline satellite inversions (2003–2010 average for BESD, 2010 for ACOS, UoL-FP, RemoTeC, NIES, and CarbonTracker) versus **a)** integration time, **b)** a priori error scaling factor, **c)** a priori bias and flux correlation lengths. Areas in dark grey represent the 1-sigma a priori uncertainty, dotted lines represent values chosen for the ensemble inversions shown in Fig. 1.
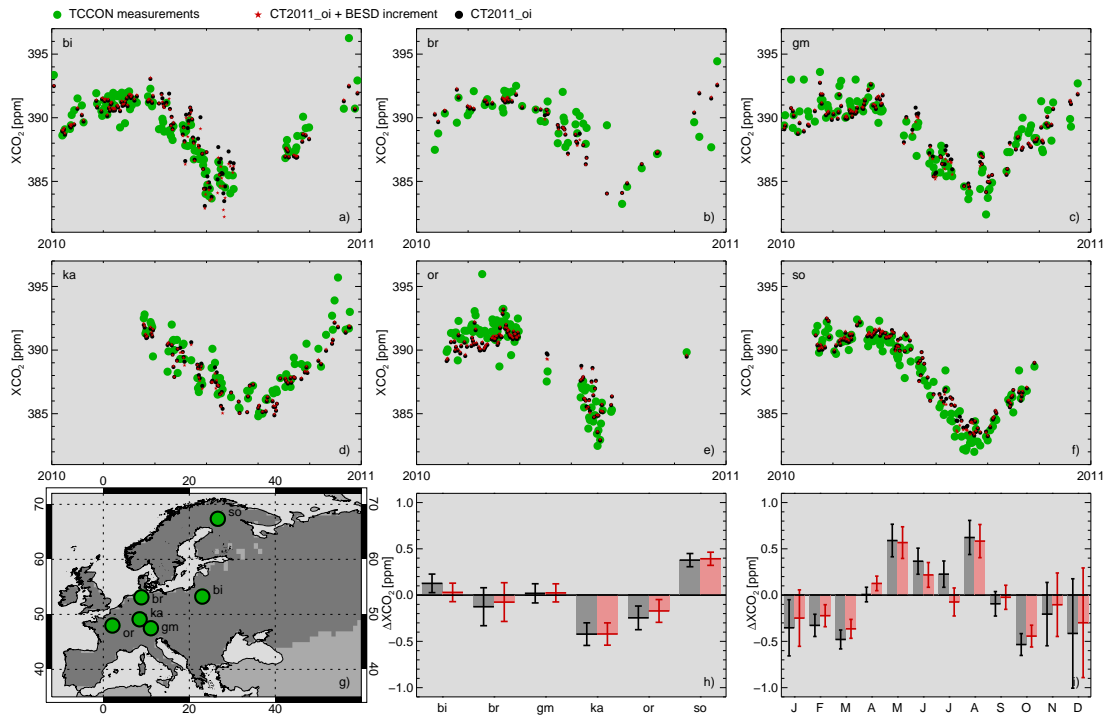
Figure C1. Validation with TCCON ground based FTS measurements. TCCON XCO_2 measurements and corresponding optimised (CT2011_oi + BESD increment) and CarbonTracker (CT2011_oi) concentrations at the European TCCON sites **a)** Białystok (Poland), **b)** Bremen (Germany), **c)** Garmisch-Partenkirchen (Germany), **d)** Karlsruhe (Germany), **e)** Orléans (France), and **f)** Sodankylä (Finland). A constant offset correction (by subtracting the mean difference) has been applied to the optimised and the CarbonTracker concentrations. **g)** Position of TCCON sites (green) within the European TRANSCOM region (dark grey). **h)** Station-to-station biases of the optimised (red) and the CarbonTracker (black) concentrations. **i)** Average seasonal biases of the optimised and the CarbonTracker concentrations. Error bars represent the standard error of the mean, i.e., the standard deviation of the difference divided by the square root of the number of measurements.
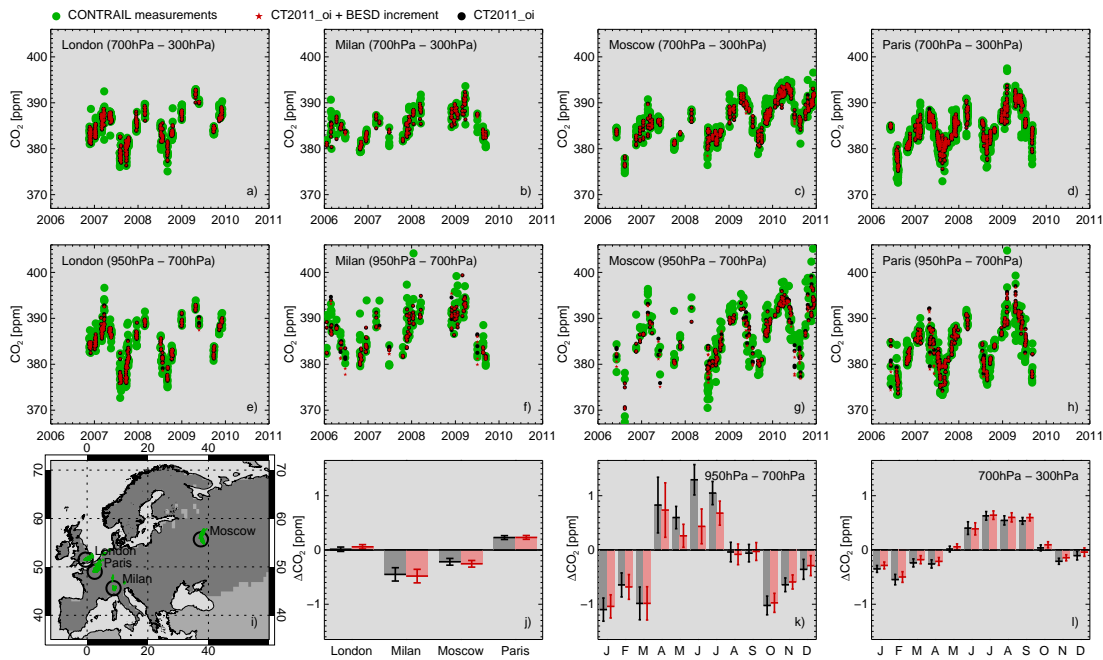
Figure C2. Validation with CONTRAIL aircraft based in situ measurements. CONTRAIL CO_2 measurements and corresponding optimised (CT2011_oi + BESD increment) and CarbonTracker (CT2011_oi) concentrations near the European cities **a, e)** London, **b, f)** Milan, **c, g)** Moscow, and **d, h)** Paris in the altitude range **a-d)** $700 - 300$ hPa and **a-d)** $950 - 700$ hPa. A constant offset correction (by subtracting the mean difference) has been applied to the optimised and the CarbonTracker concentrations. **i)** Position of CONTRAIL measurements (green) within the European TRANSCOM region (dark grey). **j)** Station-to-station biases of the optimised (red) and the CarbonTracker (black) concentrations. **k)** Average seasonal biases of the optimised and the CarbonTracker concentrations for the altitude range $950 - 700$ hPa. **l)** Average seasonal biases of the optimised and the CarbonTracker concentrations for the altitude range $700 - 300$ hPa. Error bars represent the standard error of the mean, i.e., the standard deviation of the difference divided by the square root of the number of measurements.
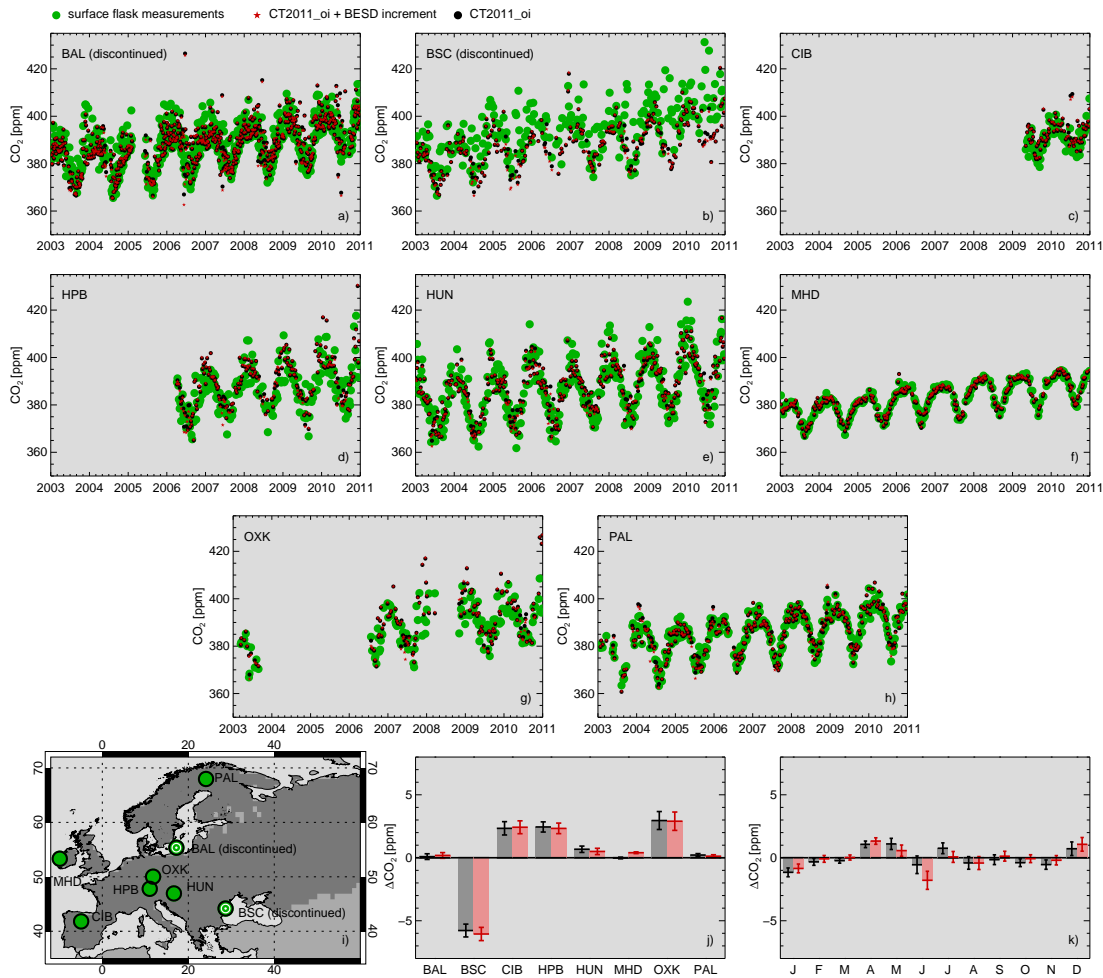
Figure C3. Comparison to ground based in situ measurements of NOAA's cooperative air sampling network. Surface flask CO_2 measurements and corresponding optimised (CT2011_oi + BESD increment) and CarbonTracker (CT2011_oi) concentrations at the European air sampling sites **a)** Baltic Sea (Poland), **b)** Black Sea (Romania), **c)** Centro de Investigacion de la Baja Atmosfera (Spain), **d)** Hohenpeißenberg (Germany), **e)** Hegyhatsal (Hungary), **f)** Mace Head (Ireland), **g)** Ochsenkopf (Germany), and **h)** Pallas-Sammaltunturi (Finland). A constant offset correction (by subtracting the mean difference) has been applied to the optimised (red) and the CarbonTracker (black) concentrations. **i)** Position of air sampling sites (green) within the European TRANSCOM region (dark grey). **j)** Station-to-station biases of the optimised and the CarbonTracker concentrations. **k)** Average seasonal biases of the optimised and the CarbonTracker concentrations. Error bars represent the standard error of the mean, i.e., the standard deviation of the difference divided by the square root of the number of measurements