Robust inversion of carbon dioxide fluxes over temperate Eurasia in 2006–2008

P. S. Swathi[1,*], N. K. Indira[1], P. J. Rayner[2], M. Ramonet[3], D. Jagadheesha[4], B. C. Bhatt[5] and V. K. Gaur[1,5]

[1]CSIR Centre for Mathematical Modelling and Computer Simulation, NAL Belur Campus, Bangalore 560 041, India
[2]University of Melbourne, Melbourne, Australia
[3]LSCE, Gif sur Yvette, France
[4]Atmospheric Science Programme, ISRO Headquarters, Bangalore 560 231, India
[5]Indian Institute of Astrophysics, Bangalore 560 034, India

Carbon dioxide (CO_2) fluxes over the temperate Eurasian region have been estimated robustly for the period 2006–2008 using the time-dependent inversion component of the TRANSCOM protocol. Data from six CO_2 measuring stations in temperate Eurasia, including the Indian station at Hanle along with 75 others from different parts of the world have been used. A careful analysis of a posteriori errors and model resolution shows that temperate Eurasia is a major residual sink of CO_2 with a value of 1.5 Gigatonnes of carbon/year.

Keywords: Bayes theorem, carbon flux, inversion, sources and sinks, temperate region, TRANSCOM.

ESTIMATES of CO_2 fluxes over the temperate Eurasian region have been poorly constrained because of the absence of high-quality measurements of atmospheric CO_2. Since reliable quantification of the sources and sinks of greenhouse gases (GHGs) provides credible bases for debating mitigative options, the lack of robustness in currently available estimates[1–3] is clearly a matter of concern. While the first reliable data on atmospheric CO_2 concentrations from India were produced by Common Wealth Scientific and Industrial Research Organisation (CSIRO), Australia from weekly flask samples collected at Cape Rama, near Goa[4] during 1993–2003, the first continuous CO_2 station became operational in September 2005 at the Indian Astronomical Observatory at Hanle (32.75°N, 79°E and 4500 m amsl) in Ladakh, North India. It was equipped with a non-dispersive infrared spectroscopy-based instrument with an accuracy of 0.01 parts per million (ppm), which was secured by careful calibration and microprocessor control of temperature, pressure and flow rate. Details of the instrument along with an analysis of data quality will be described elsewhere. It should also be noted that these measurements are distinct from direct flux measurements or eddy covariance measurements, which aim to understand CO_2 fluxes from ecosystems on small spatial scales and specific ecosystem types. This study measures the CO_2 signals which are representative of large geographic areas. These are called baseline measurements and a classic example is the Mauna Loa Observatory, which was responsible for relating global warming to secular increase in global CO_2 concentrations. Clean sites far away from local sources such as cities or active biospheric effects such as forests are preferred. The ultimate aim of these measurements is to infer robust sources and sinks on large spatial scales, which could help close the global carbon budget.

The protocols for forward simulation of atmospheric transport and time domain inversion of TRANSCOM Level 3, a multi-institutional, multi-model comparison exercise have been described in detail in a series of reports[1,5] and will be outlined here only briefly. The major aim of the exercise was to compare the performance of inverse models to determine sources and sinks of CO_2. While the participants employed their own transport models in a carefully controlled protocol to simulate the transport of CO_2 from sources, they employed a common approach for inverting residuals (misfit between the a priori model and data) to obtain residual fluxes. The forward a priori transport simulations were made for four background fluxes (average fossil fuel emissions for 1990–1995 and 1995–2000, a neutral biosphere and ocean fluxes of Takahashi)[1]. The earth was aggregated into 11 land and 11 ocean regions (Figure 1) and a monthly pulse of 1/12 gigatonnes of carbon (GTC) was emitted from each region for each calendar month to serve as bases functions. The pulses were kept on for a month and turned off for the next 35 months, and the transport model tracked their concentrations over this period. This involves the simulation of 22∗12 (264) tracers along with the four described above. From these forward simulations using MOZART[6], a versatile atmospheric tracer transport and chemistry model, the inverse problem was constructed as will be described later.

The background fossil fuel emission was estimated to increase at approximately 2% per year since the 1995–2000 estimate, yielding a global sum of 7.68 GTC in

*For correspondence. (e-mail: swathi@cmmacs.ernet.in)

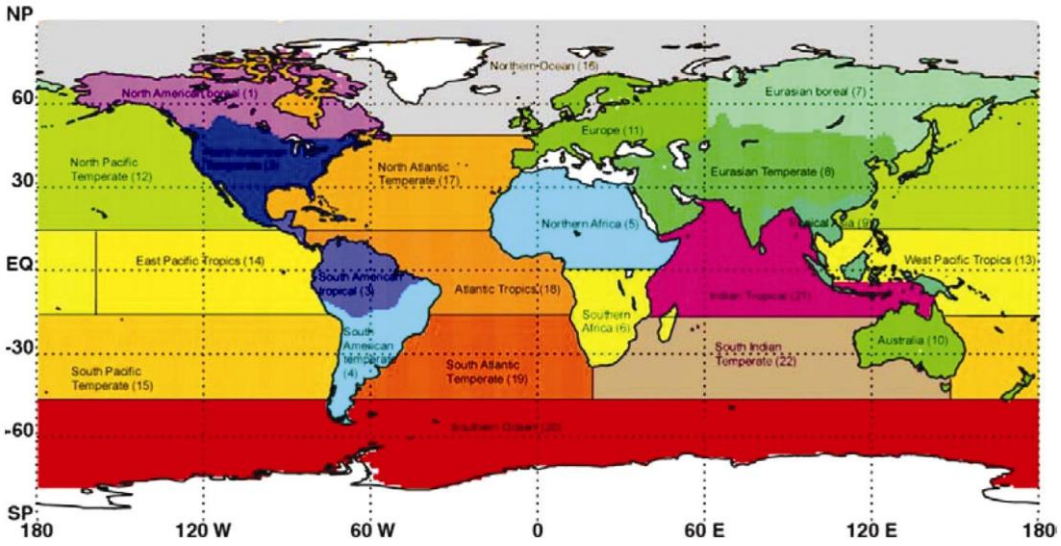

Figure 1. Geographical distribution of Transcom regions (http://transcom.project.asu.edu/transcom03_protocol_basisMap.php).

2006, 7.874 GTC in 2007 and 8.04 GTC in 2008. These estimates compare well with those compiled by The Netherlands Environmental Agency[7], if we exclude aircraft emissions which are not attributable to geographic locations. The basic idea is to see where these anthropogenic emissions go; how they are partitioned between the land and ocean, geographically and temporally. Surface maps of the net ecosystem productivity (NEP) and background ocean fluxes were taken from the TRANSCOM protocol with no inter-annual variability (IAV)[5].

The inverse formulation follows the Bayesian approach[8] of updating an a priori model (we pose the transport problem as $JS = D$) by the model-data misfit described as

$$\text{Cost} = (S - S_0)^T C(S_0)^{-1} (S - S_0) + (JS - D)^T C(D)^{-1} (JS - D), \quad (1)$$

where Cost is the cost function to be minimized, S_0 is the a priori flux model, $C(S_0)$ is its covariance, D is the data and $C(D)$ is covariance of data. In eq. (1) we try to find a compromise between the fits to the a priori model and data. In data-poor regions, the a priori model will have a larger influence, while in data-rich regions it will be the other way. The formal solutions to eq. (1) can be written as

$$S = S_0 + (J^T(C(D)^{-1} J + C(S_0)^{-1})^{-1} J^T C(D)^{-1} (D - JS_0), \quad (2)$$

$$C(S)^{-1} = J^T C(D)^{-1} J + C(S_0)^{-1}, \quad (3)$$

$$R = I - C(S)C(S_0)^{-1}, \quad (4)$$

where $C(S)$ is the a posteriori covariance. From eq. (3), it is guaranteed that the a posteriori covariance will be less than the a priori and the $J^T\text{C(D)}^{-1}J$ term determines the amount of information that the data adds to the problem. The matrix R in eq. (4) is the model resolution matrix which is a useful diagnostic of the inversion and I is the identity matrix. We describe later how eqs (1–4) are implemented in our study.

Unlike the cyclo-stationary[5] case, the size of the J matrix here is much larger. If monthly mean fluxes from M regions in N years have to be determined using monthly mean observations from Q stations, J has $(12NQ)$ rows and $(12NM)$ columns, where the factor 12 comes from the number of months in a year. In order to construct this matrix, we take the full 4D output of the monthly means of pulsed simulations with MOZART[6] at T42 resolution. We then extract the CO_2 signal at each station from each pulsed run corresponding to a region and the month of the active pulse. The J matrix constructed from this subset, thus accounts for signal decay from pulses in the past.

Data

The main source of CO_2 data is the GLOBALVIEW[9] dataset of Earth System Research Laboratory (ESRL), National Oceanographic and Atmospheric Administration (NOAA). At ESRL, raw CO_2 data are processed using the curve-fitting procedures of Thoning *et al.*[10] to obtain monthly means from both weekly/biweekly flask data as well as continuous data. In addition to the monthly means, GLOBALVIEW also provides residuals from the fit which we take as a measure of uncertainty (square root of the covariance of D, $C(D)$ in eq. (1)). We used the same procedure as GLOBALVIEW to process the data from Hanle. From the GLOBALVIEW database we selected data from stations which had more than 70% real data during the period 2006–2008. Further, we excluded

Table 1. Stations used in inversion (see also Figure 2 and Table 2)

Station no.	Station name	Latitude	Longitude	Station no.	Station name	Latitude	Longitude
1	alt_01D0	82.45	–62.52	42	pal_01D0	67.97	24.12
2	alt_02D0	82.45	–62.52	43	palmbc_30C0	67.97	24.12
3	alt_06C0	82.45	–62.52	44	poc000_01D1	0.00	–163.00
4	alt_06D0	82.45	–62.52	45	pocn05_01D1	5.00	–158.00
5	asc_01D0	–7.92	–14.4	46	pocn10_01D1	10.00	–152.00
6	ask_01D0	23.18	5.42	47	pocn15_01D1	15.00	–147.00
7	azr_01D0	38.77	–27.38	48	pocn20_01D1	20.00	–140.00
8	bal_01D1	55.50	16.67	49	pocn25_01D1	25.00	–134.00
9	bhd_15C0	–41.41	174.87	50	pocn30_01D1	30.00	–126.00
10	bme_01D0	32.37	–64.6	51	pocs05_01D1	–5.00	–168.00
11	bmw_01D0	32.27	–64.88	52	pocs10_01D1	–10.00	–174.00
12	brw_01C0	71.32	–156.60	53	pocs15_01D1	–15.00	–178.00
13	brw_01D0	71.32	–156.60	54	pocs20_01D1	–20.00	–178.50
14	cba_01D0	55.20	–162.72	55	pocs25_01D1	–25.00	174.00
15	cfa_02D0	–19.28	147.06	56	pocs30_01D1	–30.00	169.00
16	cgo_01D0	–40.68	144.68	57	psa_01D0	–64.92	–64.00
17	cgo_02D0	–40.68	144.68	58	psa_04D0	–64.92	–64.00
18	cgo_04D0	–40.68	144.68	59	rpb_01D0	13.17	–59.43
19	chr_01D0	1.70	–157.17	60	ryo_19C0	39.03	141.83
20	cpt_36C0	–34.35	18.49	61	sbl_06C0	43.93	–60.02
21	crz_01D0	–46.45	51.85	62	sbl_06D0	43.93	–60.02
22	eic_01D0	–27.15	–109.45	63	sey_01D0	–4.67	55.17
23	esp_06D0	49.38	–126.55	64	shm_01D0	52.72	174.10
24	gmi_01D0	13.43	144.78	65	smo_01C0	–14.25	–170.57
25	hba_01D0	–75.58	–26.50	66	smo_01D0	–14.25	–170.57
26	ice_01D0	63.25	–20.15	67	smo_04D0	–14.25	–170.57
27	izo_27C0	28.30	–16.48	68	spo_01C0	–89.98	–24.80
28	izo_01D0	28.30	–16.48	69	spo_01D0	–89.98	–24.80
29	jbn_29C0	–62.23	–58.82	70	spo_04D0	–89.98	–24.80
30	key_01D0	25.67	–80.20	71	stm_01D0	66.00	2.00
31	kum_01D0	19.52	–154.82	72	stmebc_01D0	66.00	2.00
32	kum_04D0	19.52	–154.82	73	sum_01D0	72.58	–38.48
33	maa_02D0	–67.62	62.87	74	syo_01D0	–69.00	39.58
34	mhd_01D0	53.33	–9.90	75	tap_01D0	36.73	126.13
35	mid_01D0	28.22	–177.37	76	tdf_01D0	–54.87	–68.48
36	mlo_01C0	19.53	–155.58	77	uum_01D0	44.45	111.10
37	mlo_01D0	19.53	–155.58	78	wlg_01D0	36.29	100.90
38	mlo_02D0	19.53	–155.58	79	yon_19C0	24.47	123.02
39	mlo_04D0	19.53	–155.58	80	zep_01D0	78.90	11.88
40	mnm_19C0	24.30	153.97	81	hle_11C0	32.8	75.57
41	mqa_02D0	–54.48	158.97				

aircraft data and all those which had large residuals. Table 1 shows the list of stations that were used in the inversion and Figure 2 displays their geographical spread. Notice that some stations (alt, brw, cgo, izo, kum, mlo, pal, psa, sbl, smo, spo and stm; see Table 1) have multiple data reflecting the availability of measurements from different labs as part of their inter-comparison exercise. There are a total of 81 measurements at 60 distinct sites. Table 2 shows the results when the measurements are aggregated to the TRANSCOM regions.

However, this aggregation is only notional as several measurements fall at regional boundaries. Data from stations which are immediately downwind of a given region, are expected to be quite representative of the region upwind of the station. Figure 3 shows the data and residuals of the 81 measurements used in this analysis. There is a preponderance of northern hemisphere stations as can be seen from the lows during northern summer months. This study includes as many as six stations in temperate Eurasia and should expectedly yield higher fidelity flux estimates over the region. The CO_2 trend is quite variable from a low of 1.6 ppm/year to a high of 2.1 ppm/year with an average of 1.94 ppm/year. This corresponds to an average global atmospheric increase of 4.04 GTC/year in 2006–2008. (Compare this with the average fossil-fuel emission of 7.87 GTC/year, which implies that 51% of the emission is retained in the atmosphere). The uncertainty (also called residuals, square root of $C(D)$; Figure 3), is generally smaller over oceanic sites as expected, because of the absence of biospheric effects. Some land stations, on the other hand, have much larger residuals (bal[8], key[30], numbers in square brackets correspond

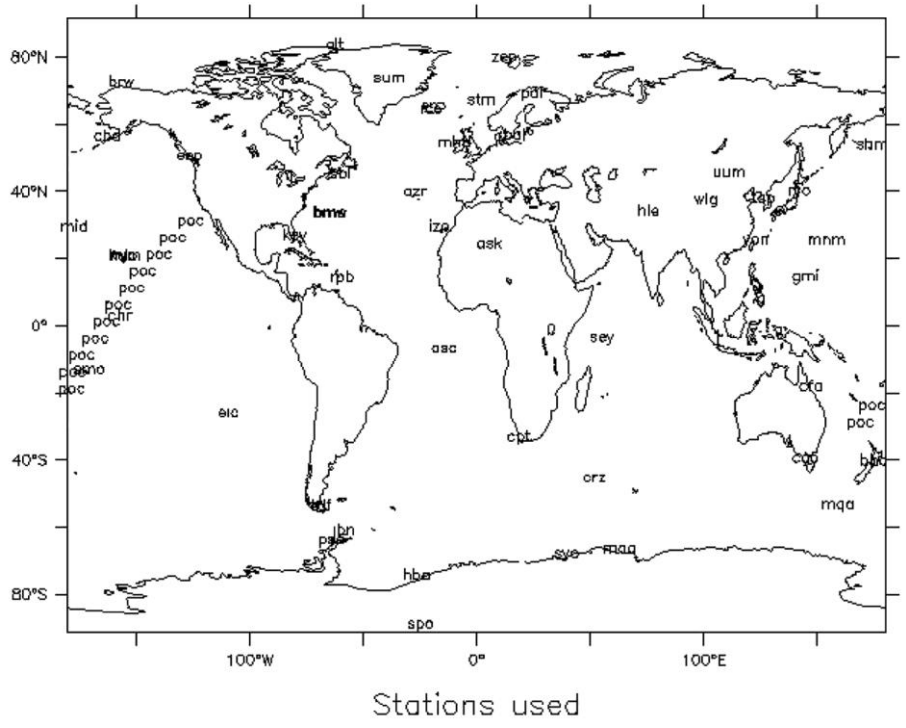

Figure 2. Stations used in flux inversions. There are 81 measurements at 60 distinct sites.

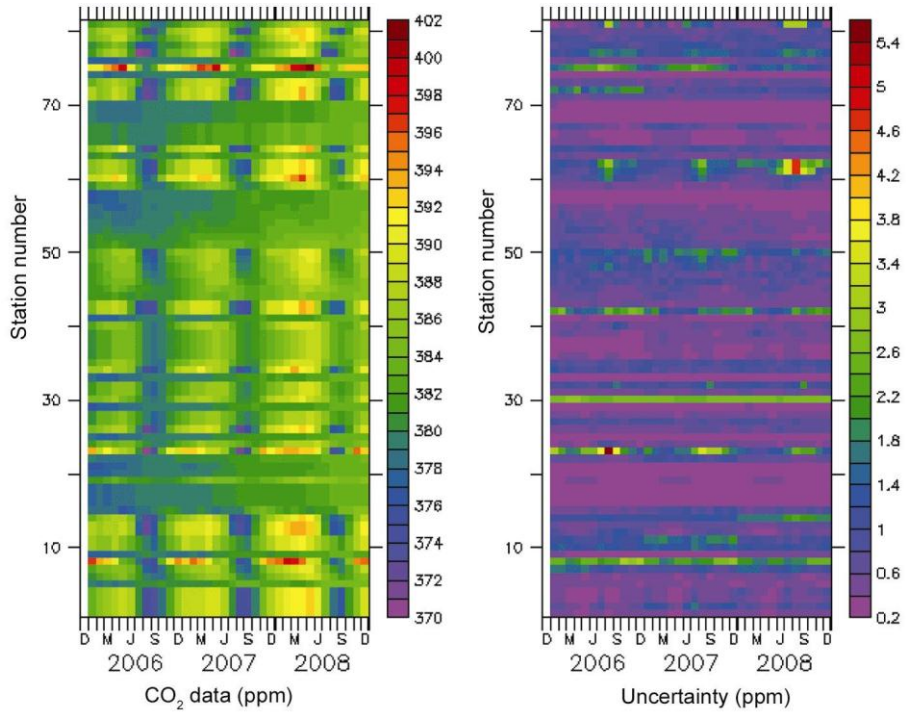

Figure 3. CO_2 data (left) and standard deviation (square root of diagonal elements of $C(D)$; right) of the 81 stations during 2006–2008. See Table 1 and Figure 2 for the geographic location of the stations.

to the listing in Table 1 and the *y*-axis of Figure 3), but their influence on inverted estimates is comparatively reduced because of being inversely weighted by their higher variances. The minimum residual was fixed at 0.3 ppm as in the TRANSCOM inversions.

The data from Hanle are shown separately in Figure 4. Some missing data in August 2006 and July–September 2008 were filled during curve-fitting and the corresponding residuals were set to a maximum of 3 ppm. The reason for this is to ensure that interpolated data values

Table 2. Stations aggregated by regions (see also Figures 1 and 2)

Region no.	Land	Number of stations	Region no.	Ocean	Number of stations
1	Boreal America	7	12	North Pacific	14
2	Temperate North America	4	13	West Equatorial Pacific	2
3	Tropical America	0	14	East Equatorial Pacific	3
4	South America	1	15	South Pacific	8
5	North Africa	1	16	Arctic	8
6	Southern Africa	1	17	North Atlantic	5
7	Boreal Eurasia	0	18	Equatorial Atlantic	1
8	Temperate Eurasia	6	19	South Atlantic	0
9	Tropical Asia	0	20	Southern Ocean	10
10	Australia	4	21	Northern Indian Ocean	1
11	Europe	4	22	Southern Indian Ocean	1
Total		28		Total	53

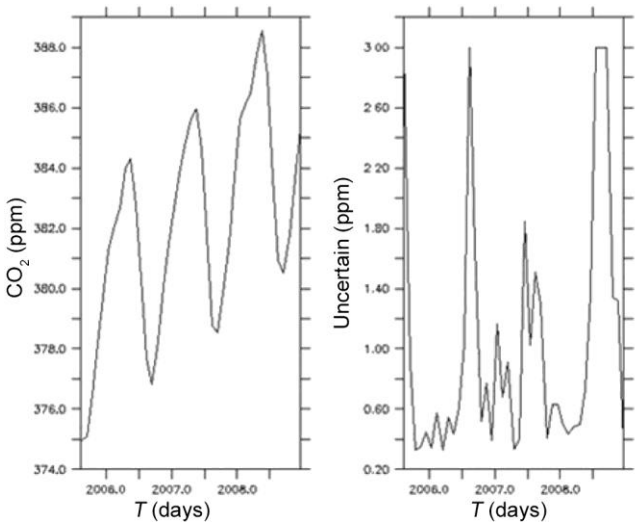

Figure 4. CO_2 data (left) and standard deviation (right) of Hanle.

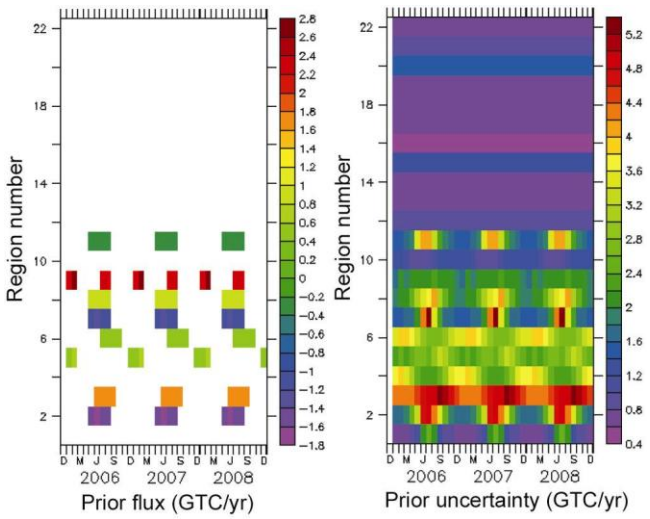

Figure 5. Prior values (left) and square root of the variance (right) of the fluxes in the TRANSCOM regions. Only non-zero values are shown. See Table 2 and Figure 1 for a description of the TRANSCOM regions. Units are GTC/year.

do not have a large impact on the cost function. The seasonal cycle has amplitude of approximately 8 ppm with a maximum during April–May, when the prevailing winds are northeasterly. The minimum, on the other hand, falls in September, reflecting a combination of the SW monsoon winds from the ocean and the biospheric effect of the Indian subcontinent.

Before inversion, we augment the J matrix with rows and columns corresponding to the yearly fossil-fuel emissions, net ecosystem productivity, ocean fluxes and a constant offset. The size of the J matrix in our case is 2916*805 (81 stations*3 years *12 months = 2916 data points, 22 regions *3 years *12 months sources + 12 background + 1 offset = 805 sources). We use the Math Kernel Library (MKL) routines for singular value decomposition of the J matrix and subsequent analysis.

The priors S_0, used in the Bayesian inversion are shown in Figure 5. These are assumed to be uncorrelated amongst the regions, which implies that $C(S_0)$ is a diagonal matrix. The left panel in Figure 5 shows non-zero S_0 and the right panel, the uncertainty in S_0, i.e. the square root of the diagonal entries of $C(S_0)$. The justification for using them can be found in Gurney *et al.*[1]. Temperate Eurasia (region #8) is taken to be a mild positive source during the summer months. The uncertainties on land are reasonably high to reflect the uncertainty in this flux. The oceanic regions have non-seasonal uncertainty with much lower values, except in the southern ocean. As temperate Eurasia has six stations in this study, it is expected that a posteriori fluxes will be determined to a large extent by the data and not by the priors.

Results and discussion

We recognize at the outset that some regions are poorly constrained and therefore focus our interest on those that can be trusted. We attempt hereunder a rigorous analysis of the a posteriori errors to elucidate the robustness of our results.

The unfiltered results of the inversion are shown in Figure 6 with S on the left and the posteriori uncertainty (square root of the diagonal entries of $C(S)$) on the right. As expected, the inversion is quite noisy, but quite a few features are still readily apparent. Positive values indicate sources, while negative indicate sinks. Temperate Eurasia (region #8) and Europe (region #11) clearly emerge as strong sinks with reduced uncertainties (compare Figures 5 and 6). Land regions with few observations (tropical and South America, Africa) still have large errors. Interestingly, the Southern Ocean shows a positive anomaly with a small uncertainty. The prior and posteriori variances of all the regions for all the months are shown in Figure 7. (This has to be divided in blocks of 36 for each region, 1–36 covering boreal North America, January 2006–December 2008 and so on.) The sharp fall in the land regions is noticeable, but this is not a proper measure of robustness. A good measure, although not unique, are the diagonal entries of the model resolution matrix, which is described later.

The chi-squared fit of the deviation from the priors as well as data is shown in Figure 8. On the left is chi-squared fit of the model (i.e. $(S_i - S_{0,i})^2/C(S_{0,i})$) and on the right is chi-squared fit of the data (i.e. $(D_i - JS_i)^2/C(D_i)$). Clearly, there is considerable data misfit in the first month (January 2006), but this is understandable as the sources are represented only from the year 2005 with no consideration given to the years before that. Besides, there are a few other stations (ask[6], cfa[15], hba[25], izo[28], mhd[34], pal[42], spo[68–70], stm[71], sum[73], tap[75]; the number in square brackets indicates the station number) where the fit is not good, indicating the inability of the MOZART to simulate these stations well. However, the fit at Hanle is particularly good. From the chi-squared fit to priors, we notice that regions 1, 9 and 12 (boreal North

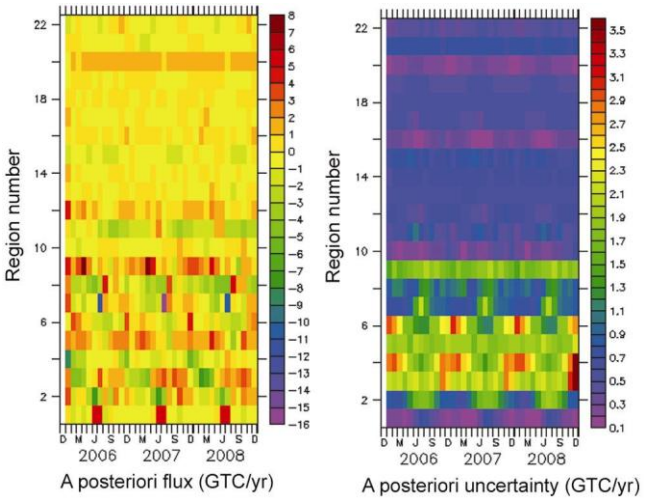

Figure 6. Unfiltered results from flux inversion: a posteriori values (left) and square root of a posteriori variance (right). Units are GTC/year.

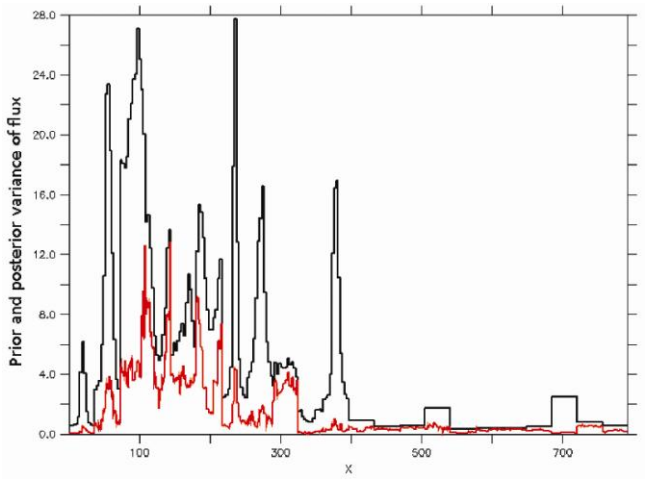

Figure 7. Prior (black) and a posteriori (red) variance of the inversion in $GTC^2/year^2$. They are in blocks of 36 contiguous values corresponding to each TRANSCOM region, i.e. 1–36 is boreal North America, 37–72 is temperate North America, and so on.

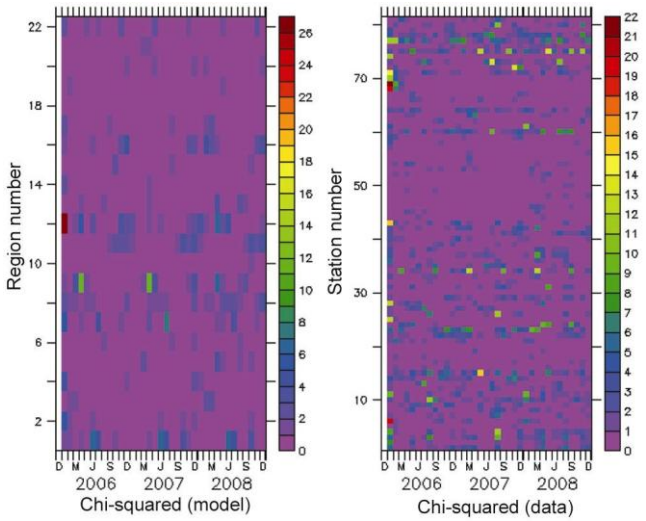

Figure 8. Chi-squared fit of sources $[(S_i - S_{0,i})^2/C(S_{0,i})$, left] and data $[(D_i - JS_i)^2/C(D_i)$, right].

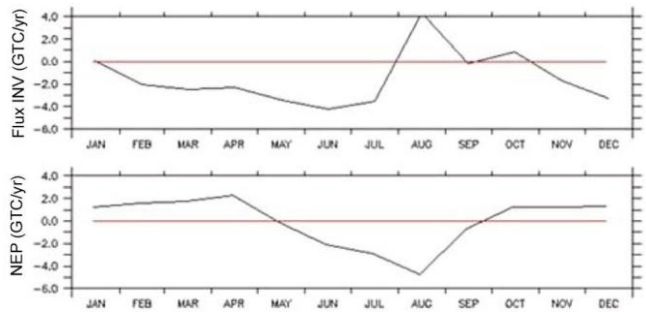

Figure 9. Climatology of CO_2 flux (residual) after inversion (top) and net ecosystem productivity (NEP, bottom) of temperate Eurasia. Red line represents zero.

Table 3. Average inversion results for 2006–2008 (climatology). Columns 3–7 and 9 have units of GTC/year

Region number	Name	Prior	Post	Prior uncertainties	Post uncertainties	Fossil fuel ocean	Res	Total	Robust
1	Boreal North America	0.000	0.401	1.199	0.319	0.04	0.929	0.401	Y
2	Temperate North America	–0.534	–0.681	2.890	1.190	1.92	0.831	–0.681	Y
3	Tropical North America	0.547	–0.137	4.633	2.198	0.23	0.775	–0.137	Y
4	South America	0.000	–1.321	3.024	2.312	0.14	0.415	–1.321	N
5	North Africa	0.152	1.314	2.667	1.902	0.16	0.491	1.314	N
6	Southern Africa	0.148	–0.122	3.215	2.005	0.11	0.610	–0.122	Y
7	Boreal Eurasia	–0.396	–1.161	2.404	1.045	0.21	0.810	–1.161	Y
8	Temperate Eurasia	0.297	–1.485	2.741	1.018	2.38	0.862	–1.485	Y
9	Tropical Asia	0.806	1.887	2.091	1.776	0.64	0.278	1.187	N
10	Australia	0.000	–0.058	1.109	0.333	0.11	0.910	–0.058	Y
11	Europe	–0.100	–1.518	2.411	0.603	1.92	0.94	–1.518	Y
12	North Pacific	0.000	1.003	0.960	0.583	–0.50	0.631	0.500	Y
13	Equatorial West Pacific	0.000	–0.042	0.710	0.558	0.15	0.381	0.111	N
14	Equatorial East Pacific	0.000	–0.050	0.750	0.631	0.46	0.291	0.417	N
15	South Pacific	0.000	–0.615	1.320	0.686	0.23	0.730	0.845	Y
16	Arctic	0.000	0.352	0.560	0.270	–0.44	0.770	–0.086	Y
17	North Atlantic	0.000	–0.076	0.640	0.521	–0.29	0.338	–0.368	N
18	Equatorial Atlantic	0.000	0.036	0.640	0.557	0.13	0.24	0.165	N
19	Southern Atlantic	0.000	0.058	0.690	0.484	–0.13	0.508	–0.070	N
20	Southern Ocean	0.000	1.304	1.580	0.271	–0.90	0.971	0.417	Y
21	Northern Indian	0.000	–0.190	0.890	0.730	0.12	0.327	–0.072	N
22	Southern Indian	0.000	–0.205	0.740	0.456	–0.55	0.620	–0.759	Y
					Total	Fossil fuel = 7.86; Ocean = –2.19			

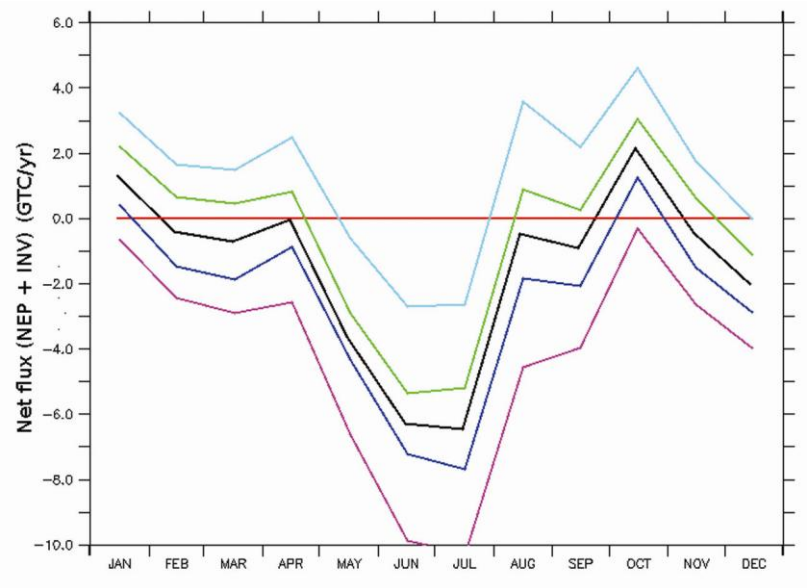

Figure 10. Climatology of net flux from temperate Eurasia with error bands. The outer envelope is the a priori and the inner is the a-posteriori. Red line represents zero and black line represents the climatology. The pink and cyan lines are the prior envelopes, whereas the blue and the green lines are the posterior envelopes.

America, tropical Asia and North Pacific) have the largest departures relative to their uncertainties.

We first present the averaged fluxes, uncertainties and resolution for the period 2006–2008 (all in GTC/year, except resolution (Res) which is dimensionless) in Table 3, before we discuss seasonality.

There are several measures to check the robustness of inversion[8], i.e. a comparison of prior and posterior uncertainties, correlations between estimated model parameters and the model resolution matrix R given in eq. (4). In the context of the standard least squares problem with no priors, the last quantity indicates the extent to which a parameter can be uniquely resolved (or if only a linear combination of parameters can be determined). We believe that the model resolution parameter is an appropriate indicator of robustness as it is a relative measure. In Table 3, the fourth column gives the final residual sources or sinks, the sixth column the posteriori uncertainties and the

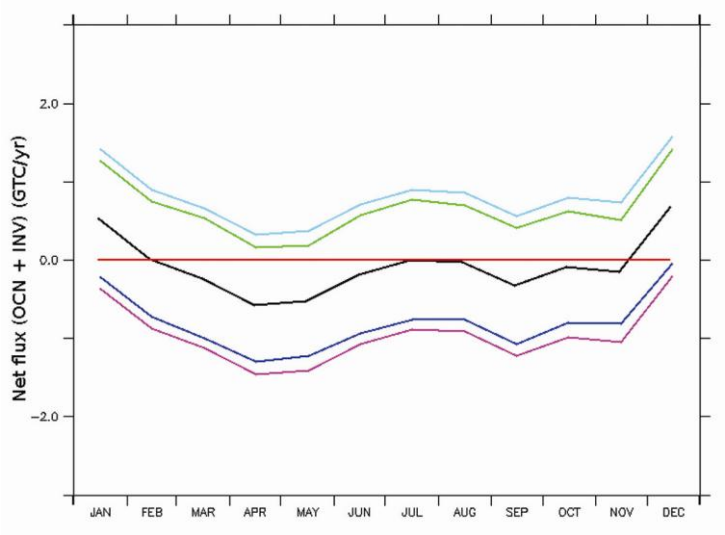

Figure 11. Climatology of the net flux from the northern Indian Ocean. The reduction in error is very small.

eighth column the square root of the diagonal elements in the model resolution matrix for the region. If we set a threshold of 0.6 for the last quantity as the acceptable lower limit for robustness, then 9 of the 22 regions fail to make the grade. Temperate Eurasia is fourth amongst the robust land inversions behind boreal North America, Australia and Europe. Of the 2.38 GTC/year of fossil-fuel emissions by Temperate Eurasia in 2006–2008, 1.485 GTC/year (ninth column, Table 3) is captured by the land sink leaving a net emission of 0.9 GTC/year.

The background neutral NEP flux (bottom) and the seasonal residual flux (top, 2006–2008 climatology) after inversion for temperate Eurasia are shown in Figure 9. The NEP flux indicates that the maximum CO_2 draw-down occurs in August, while the residual flux shows the presence of large sinks in June and July and a large source in August, effectively pushing the growing season to the earlier part of the boreal summer. The a priori (outer envelope) and a posteriori errors (inner) in the seasonal estimation are shown in Figure 10. Note the considerable reduction in errors, particularly during the growing season. The total flux for the northern Indian Ocean is shown in Figure 11. Notice that there has been very little reduction in the uncertainty as only one station (sey) is present in this region.

In conclusion, we affirm that the availability of the new Asian stations, especially Hanle, has led to considerable improvement in the robustness of inverted flux estimates for temperate Eurasia. To resolve fluxes at a smaller spatial scale (country-wise), we need a better station density than is currently available. Meanwhile, the closing of the global budget still remains a distant goal as several regions on the globe continue to have large data gaps.

1. Gurney, K. R. *et al.*, Towards robust regional estimates of CO_2 sources and sinks using atmospheric transport models. *Nature*, 2002, **415**, 626–630.
2. Baker, D. F. *et al.*, TransCom 3 inversion intercomparison: impact of transport model error on the interannual variability of regional CO_2 fluxes, 1988–2003. *Global Biogeochem. Cycles*, 2006, **20**, 1; doi: 10.1029/2004GB002439
3. Stephens, B. B. *et al.*, Weak northern and strong tropical land carbon uptake from vertical profiles of atmospheric CO_2. *Science*, 2007, **316**, 1732–1735.
4. Bhattacharya, S. K. *et al.*, Trace gases and CO_2 isotope records from Cabo de Rama, India. *Curr. Sci.*, 2009, **97**, 1336–1344.
5. Gurney, K. R., Law, R., Rayner, P. J. and Denning, S., TRANSCOM 3 experimental protocol. Department of Atmospheric Science, Colorado State University, Paper no. 707, 2000.
6. Horrowitz, L. W. *et al.*, A global simulation of tropospheric ozone and related tracers. *J. Geophys. Res. D*, 2003, **108**, 4784; doi: 10.1029/2002JD002853.
7. Olivier, J. G. J., Janssens-Maenhout, G. and Peters, J. A. H. W., Trends in global CO_2 emissions: 2012 report. The Netherlands Environmental Agency. 2012; http://edgar.jrc.ec.europa.eu/CO2REPORT2012.pdf
8. Tarantola, A., *Inverse Problem Theory*, SIAM, Philadelphia, 2005.
9. GLOBALVIEW-CO_2: Cooperative Atmospheric Data Integration Project – Carbon Dioxide. CD-ROM, NOAA ESRL, Boulder, Colarado (also available on Internet via anonymous FTP from ftp.cmdl.noaa.gov), 2011.
10. Thoning, K. W., Tans, P. P. and Komhyr, W. D., Atmospheric carbon dioxide at Mauna Loa Observatory: 2. Analysis of the NOAA GMCC data, 1974–85. *J. Geophys. Res. D*, 1989, **94**, 8549–8565.

ACKNOWLEDGEMENTS. We thank the Indo-French Centre for Advanced Research (IFCPAR) and the British High Commission for supporting two projects on which this study is based. We also thank late Prof. A. Tarantola for several lectures on the Bayesian inverse problem which provided a solid foundation for tackling this problem. We thank D. Angchuk and his colleagues at IIA, Hanle for monitoring the Caribou instrument; Prof. P. Seshu (C-MMACS) and the Director, IIA for their continued support and R. P. Thangavelu (C-MMACS) for computer assistance. We also wish to thank Marc Delmotte and Martina Schmidt, LSCE for data analysis and V. K. Shambhulinga (C-MMACS) for assistance with figures. P.J.R. is a recipient of an ARC Professorial Fellowship.

Received 15 January 2013; revised accepted 25 April 2013