

## Data and modelling requirements for CO<sub>2</sub> inversions using high-frequency data

By R. M. LAW\*, P. J. RAYNER, L. P. STEELE and I. G. ENTING, *CSIRO Atmospheric Research, PMB 1, Aspendale, Victoria 3195, Australia*

(Manuscript received 14 January 2002; in final form 25 September 2002)

### ABSTRACT

We explore the future possibilities for CO<sub>2</sub> source estimation from atmospheric concentration data by performing synthetic data experiments. Synthetic data are used to test seasonal CO<sub>2</sub> inversions using high-frequency data. Monthly CO<sub>2</sub> sources over the Australian region are calculated for inversions with data at 4-hourly frequency and averaged over 1 d, 2.5 d, 5 d, 12.17 d and 1 month. The inversion quality, as determined by bias and uncertainty, is degraded when averaging over longer periods. This shows the value of the strong but relatively short-lived signals present in high-frequency records that are removed in averaged and particularly filtered records. Sensitivity tests are performed in which the synthetic data are 'corrupted' to simulate systematic measurement errors such as intercalibration differences or to simulate transport modelling errors. The inversion is also used to estimate the effect of calibration offsets between sites. We find that at short data-averaging periods the inversion is reasonably robust to measurement-type errors. For transport-type errors, the best results are achieved for synoptic (2–5 d) timescales. Overall the tests indicate that improved source estimates should be possible by incorporating continuous measurements into CO<sub>2</sub> inversions.

### 1. Introduction

Estimating regional CO<sub>2</sub> sources and sinks from atmospheric concentration data is important both for understanding natural carbon fluxes and as a confirmation of 'bottom-up' estimates of fluxes being compiled for international protocols. This estimation of fluxes by inversion was first attempted 40 years ago (Bolin and Keeling, 1963), but progress has always been hampered by the under-determined nature of the problem; there have been insufficient atmospheric data available to reduce uncertainties to a useful level, especially for tropical land regions (Gurney et al., 2002). Current inversions tend to estimate monthly mean sources (e.g. Bousquet et al., 2000) from monthly concentration data. The monthly concentrations are typically derived from a curve fit to marine-boundary-layer data either from a continuous analyser or from flask sam-

ples taken at sub-weekly to monthly intervals (e.g. GLOBALVIEW-CO<sub>2</sub>, 2001). The sampling times are usually chosen so that samples represent large air-masses (so-called clean or baseline air) and avoid local (often land) signals. In this context, outlying data are viewed as noise and are removed in the curve-fitting procedure and hence from the inversion.

Law et al. (2002) (referred to subsequently as L02) took a first step to examine some of the issues involved in using non-baseline data in CO<sub>2</sub> inversions. They inverted time series of 4-hourly synthetic concentration data to test whether known sources could be successfully retrieved with low uncertainty. Synthetic data allowed them to include data at times or locations where none currently exists and hence to examine the potential improvements to source estimation if flask sampling at sites was replaced by continuous sampling. Their results showed that source uncertainties were much reduced for 4-h compared to monthly data, but that the source estimates may be severely biased when the sources are only estimated for large (up to

\*Corresponding author.  
e-mail: Rachel.Law@csiro.au

continental-scale) regions. In an Australian case study, sources were estimated for transport model grid-cell size regions (5.6° longitude by 2.8° latitude). In this case the source bias for the total Australian region was reduced to less than the estimated uncertainty and useful source information was also retrieved at grid-cell scale. The synthetic concentration data used by L02 provided a relatively simple first test, since the same transport model was used both to create the data and to invert them. This paper aims to extend the work of L02 by presenting a series of sensitivity tests designed to mimic some measurement-related data limitations such as calibration errors between records, as well as some modelling limitations. In addition, we present inversions that average the 4-hourly data used in L02 to periods ranging from daily to monthly.

## 2. Method

The details of the method are given in L02. Briefly, monthly sources are estimated by Bayesian synthesis inversion (Enting, 2000). We calculate sources and uncertainties via

$$\vec{S} = \vec{S}_0 + \left[ \mathbf{C}^{-1}(\vec{S}_0) + \mathbf{J}^T \mathbf{C}^{-1}(\vec{D}_0) \mathbf{J} \right]^{-1} \times \mathbf{J}^T \mathbf{C}^{-1}(\vec{D}_0) (\vec{D}_0 - \mathbf{J} \vec{S}_0) \quad (1)$$

and

$$\mathbf{C}^{-1}(\vec{S}) = \mathbf{C}^{-1}(\vec{S}_0) + \mathbf{J}^T \mathbf{C}^{-1}(\vec{D}_0) \mathbf{J} \quad (2)$$

where  $\vec{S}$  represents sources,  $\vec{D}$  represents data, the quantity  $\mathbf{C}(x)$  represents the uncertainty on  $x$  in the form of a covariance matrix,  $\mathbf{J}$  is the matrix of responses at measurement locations to unit sources from the chosen regions and the subscript 0 represents prior estimates. Equation (1) shows that the estimated source is a correction to the prior source dependent on the mismatch between the actual observations and those calculated from the prior source. This mismatch is weighted by the confidence we have in the ability of our inversion to match the data. High values for  $\mathbf{C}(\vec{D}_0)$  could mean either that the data are highly scattered or that we do not place much trust in our inversion set-up. Equation (2) shows that our confidence in the source estimate (embodied in the inverse covariance) is a combination of our confidence in the prior estimate and information added by the data. Critically this does not involve any knowledge of the actual mismatch, allowing the possibility of misplaced confidence in our estimate.

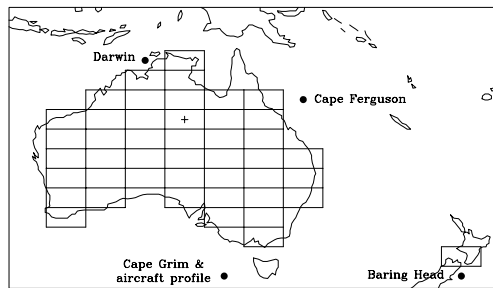


Fig. 1. Australian region showing grid cells for which the inversion estimates sources and locations of sites used in the inversion (labelled dots). The extra continental site used in the group 2 sensitivity tests is shown by the +.

The inversion used here is ‘cyclo-stationary’ in that it solves for the mean seasonal cycle of fluxes but not interannual variations. Sources are estimated for the 11 ocean and 11 land regions used in the TransCom intercomparison (Gurney et al., 2002). One of these regions encompasses Australia and New Zealand, and here we subdivide that region into 45 sub-regions as defined by the transport model grid-cells (Fig. 1). We invert for both the original large region and the 45 grid-cell regions. To reduce the computational demands of the response function calculations required by the inversion, only three-month responses are generated for the 45 grid-cell regions, compared to three-year responses for the 22 large regions.

The synthetic concentration data that we invert are generated by forward simulations combining specified fossil, biosphere and ocean emissions as used by TransCom (Andres et al., 1996; Randerson et al., 1997; Takahashi et al., 2002, describe the methodology for the construction of the ocean emissions, but here we used an earlier, 1999 version of the air-sea fluxes). The inversions use data from 83 sites distributed similarly to the current global flask-monitoring network. The synthetic data have been sampled every 4 h from the model simulation, allowing us to explore the possibilities of having continuous measurements at locations where currently only flask samples are available. Thus, this work explores what may be possible in the future should more continuous instruments be commissioned. The sites in the vicinity of the Australian region are marked on Fig. 1.

Uncertainties must be applied to each data point. These need to account for any potential differences between modelled and measured concentrations and as such are chosen to be substantially larger than measurement precision alone. Following tests in L02, we

use data uncertainties that are variable between sites but are constant in time. The uncertainties were calculated as 1.0 ppm plus the standard deviation of concentrations around an approximately 10-d running mean concentration. This gave uncertainties ranging from 1.1 to 3.6 ppm (compared to a typical measurement precision of around 0.2 ppm). These data uncertainties are included as the diagonal elements in  $\mathbf{C}(\bar{D}_0)$ . Off-diagonal terms are set to zero corresponding to an assumption of uncorrelated normally distributed uncertainties on the data.

When data are averaged for inversions with lower data frequency the uncertainty is reduced as

$$\sigma_n = \frac{\sigma_1}{\sqrt{n}}, \quad (3)$$

where  $\sigma_1$  is the data uncertainty applied to the 4-hourly data and  $\sigma_n$  is the uncertainty for the average of  $n$  data values. This assumes that the data uncertainties are independent in time, which is unlikely to be realistic. In reality, we would expect to reduce the uncertainty by less than  $\sqrt{n}$ . However, this would mean that the high data frequency inversions would produce lower source uncertainties due only to the greater amount of data available to the inversion. Since we are interested in exploring the reasons for the increased power of inversions using high-frequency data we wanted to avoid this trivial cause. We perform inversions with data at 4-hourly frequency and averaged to daily, 2.5-d, 5-d, 12.167-d (73 4-h values) and 1-month periods.

Bayesian synthesis inversion requires prior sources to be specified. For the large regions we use  $0 \pm 10$  Gt C yr<sup>-1</sup>, while for the grid-cell regions the zero priors had a variable uncertainty of five times the net primary production (NPP) in the grid cell. This gave uncertainties ranging from 0.02 to 0.45 Gt C yr<sup>-1</sup>. This choice of prior source uncertainty gave the best results of three cases tested in L02.

### 2.1. Assessing inversion quality

The task of an inversion, in statistical terms, is to generate a confidence interval that contains the true solution. For the inversion to be useful and not misleading, the true solution must lie within the confidence interval described by the estimate and uncertainty produced by the inversion. In a perfect inversion case, the relationship between the true value, the estimate and the uncertainty is simple. By construction, if we carried out many experiments in which the data was perturbed about its input value according to the distribution de-

scribed by the data uncertainty, we would generate a distribution of sources equivalent to that suggested by the true source and its uncertainty. In an imperfect set-up this will not be true: the distribution of source estimates will be offset with respect to the truth; we describe this offset as the bias of the inversion. Since no random perturbations based on the data uncertainty are added in these experiments, this offset or bias may be computed directly as the difference between the true and estimated sources. Note that these biases may vary in sign and magnitude from month to month and region to region. If we have reason to expect that the magnitude of the bias is, on average, larger than the source uncertainty then our estimate is significantly different from the truth, which is clearly undesirable. Since we generate source uncertainties algorithmically from data uncertainties [eq. (2)], we must adjust these data uncertainties until this problem is avoided. Synthetic data experiments, for which we know the true solution, provide a test-bench since biases and uncertainties can be compared for any given inversion set-up. Thus we can, for example, assess whether the chosen data uncertainties are large enough to ensure that any bias in the estimates is smaller than the uncertainty.

We require some global statistics describing overall uncertainty and bias. Here we express these as the root mean-square bias (RMSB) and root mean-square uncertainty (RMSU) of the monthly estimates:

$$\text{RMSB} = \sqrt{\frac{1}{N} \sum_n (C_n - E_n)^2} \quad (4)$$

where  $C_n$  is the correct source for month  $n$ ,  $E_n$  is the estimated source and  $N$  is the number of months (12);

$$\text{RMSU} = \sqrt{\frac{1}{N} \sum_n U_n^2} \quad (5)$$

where  $U_n$  is the uncertainty for month  $n$ . RMSB and RMSU can be calculated for all regions for which sources are estimated but we focus only on RMSB and RMSU for the total Australian region (i.e. the combined sources from the original large Australian region and the 45 grid-cells).

Using these measures a good inversion has two requirements. The first is that RMSU is small; the smaller the RMSU, the more confidence we expect to have in the source estimates. The second requirement is that the RMSB is less than the RMSU. If we were inverting real data, where the sources are unknown, we are unable to calculate RMSB. Thus we are reliant on the RMSU to indicate the reliability of our source

estimates. If synthetic data tests, set up in a similar manner to the real inversions, consistently give RMSB larger than RMSU, then RMSU is not providing a realistic measure of source uncertainty and we should look at ways that RMSU can be increased.

Source estimates will be incorrect for a number of reasons. Some of these, such as imprecise measurements, are random in nature. They occur because we only have one realisation of a dataset. If we had many realisations to invert (and all other aspects of the inversion were perfect) then we could expect the mean estimated sources to be unbiased and the mean RMSB to equal the RMSU that would be determined from data uncertainties set at the measurement precision value alone (e.g. 0.2 ppm). Other reasons for incorrect source estimates are systematic, such as limitations in inversion set-up. We allow for these in the inversion by increasing the data uncertainty from measurement precision alone to a value that can accommodate the other reasons that modelled and measured concentrations might not perfectly agree. However, putting a numerical value on this data uncertainty increase is difficult. Kaminski et al. (2001) demonstrated how this is done for ‘aggregation’ errors associated with inverting for large regions, but other contributions to the data uncertainty have not been evaluated. Thus our data uncertainty choices can be rather arbitrary. Using the requirement of statistical consistency in synthetic data experiments is one check on these choices.

## 2.2. Sensitivity tests

A range of sensitivity tests are performed, which we divide into three groups. The first two groups investigate issues surrounding data intercalibration while the final group explores how the inversion responds if transport is imperfectly modelled. The sensitivity tests are compared to inversions of ‘clean’ data, i.e. unmodified concentrations generated from the forward simulations with known sources. The different tests are summarised in Table 1.

**2.2.1. Group 1.** Many different laboratories around the world make high-quality measurements of the mixing ratio of atmospheric CO<sub>2</sub>, but the seamless merging of such data for interpreting some aspects of the global carbon cycle has proved elusive. It is elusive partly because finding the geophysical signals in the data places severe demands on the quality of the data. In a recent paper, Masarie et al. (2001) provide a candid evaluation of a long-running (1992–1998) experiment designed to assess the level of consistency

Table 1. *Sensitivity tests*

| Test | Description                                                               |
|------|---------------------------------------------------------------------------|
| 1a   | Offset every 2 months up to $\pm 0.2$ ppm                                 |
| 1b   | Offset every 2 months up to $\pm 0.5$ ppm                                 |
| 2a   | Addition of Australian continental site                                   |
| 2b   | Addition of Australian continental site with unknown offset               |
| 2c   | Perturbed (by 0.2, 0.5, 1.0, 2.0 ppm) additional site with unknown offset |
| 2d   | All (84) sites with unknown offsets                                       |
| 3a   | Timeseries shifted temporally by up to $\pm 1$ d                          |
| 3b   | Timeseries peaks amplified or reduced by up to 100%                       |

achieved by two independent laboratories in measurements of several atmospheric trace species (including CO<sub>2</sub>) in flask air samples. One conclusion is that achieving the required data quality remains an ongoing and difficult challenge. Another view of how different laboratories compare with their measurements of atmospheric CO<sub>2</sub> is provided by the results of the blind round-robin experiments conducted periodically under the auspices of the World Meteorological Organization (WMO). In these experiments, suites of three standards (CO<sub>2</sub>-in-air, with a low, medium and high CO<sub>2</sub> mixing ratio, contained in high-pressure cylinders) have been circulated to participating laboratories, who report their results to a neutral referee, prior to publication of all results (e.g. Pearman, 1993; Peterson et al., 1997). Results show significant differences ( $>0.05$  ppm) between laboratories, with evidence that these differences are sometimes dependent upon the CO<sub>2</sub> mixing ratio (see discussion by Francey et al., 2001).

While differences between laboratories in their reported values for atmospheric CO<sub>2</sub> are challenging enough, differences between techniques or instruments within a laboratory have also been reported. At some measurement locations, continuous CO<sub>2</sub> analysers are operated in conjunction with regular flask air sampling, with subsequent measurement of the flask samples at a central laboratory. Comparison of the flask and continuous data reveal that significant periodic divergences ( $>0.2$  ppm) can occur [see, e.g. WMO (1995) for some examples from the NOAA/CMDL observatories]. More recently, in an experiment at the Cape Grim Baseline Air Pollution Station, it has been found that two different non-dispersive infrared CO<sub>2</sub> analyser systems, operating side-by-side, can yield significantly divergent (up to 0.15 ppm) values for atmospheric CO<sub>2</sub>. The pattern of differences

between the two analysers correlates well with the frequent (ca. every 2–3 months) changing of calibration tanks on the older of the CO<sub>2</sub> analysers, suggesting some systematic effects that have not yet been diagnosed.

We test the impact on the inversion of measurement errors of this kind by degrading the clean synthetic data as follows (test 1a). Random concentration perturbations are generated in the range  $-0.2$  to  $0.2$  ppm with a rectangular distribution. Each site record is divided into six sections of approximately two months in length. A different perturbation is added to each two-month section of the timeseries to mimic the impact of analyser tank changes, which would typically occur every couple of months. Additionally different sets of perturbations are applied to each site to allow for intercalibration differences between sites. Thus at each site there is a systematic difference between the degraded and the clean data which is constant for two months and then changes in magnitude and varies between sites. Finally, the 83 sets of perturbations were rotated through all sites, so that 83 inversions were performed to give a distribution of inversion impact. The inversions were also repeated (test 1b) with concentration perturbations in the range  $-0.5$  to  $0.5$  ppm.

**2.2.2. Group 2.** It is likely that future inversions may want to make use of both well calibrated CO<sub>2</sub> measurements and those that are not integrated onto international scales such as lower precision measurements being taken at flux tower locations (e.g. Potosnak et al., 1999; Running et al., 1999). To test this, we add a continental Australian site (Fig. 1) to the 83-site inversion using clean data. In the first case (test 2a), we assume that the extra site is perfectly calibrated with the rest of the network. In the second case (test 2b), we assume an unknown, but constant, offset between the extra site and all other sites, which we then estimate as part of the inversion. The unknown offset is given a prior estimate of  $0 \pm 2$  ppm. To check the ability of the inversion to retrieve this offset, the concentration data at the extra site were perturbed by a constant value across the whole year of  $0.2$ ,  $0.5$ ,  $1.0$  and  $2.0$  ppm (test 2c). A final case (test 2d) solves for unknown offsets at each of the 84 sites. This approaches a worst-case example in that we assume we know almost nothing about the intercalibration between sites except that any calibration difference remains constant in time.

**2.2.3. Group 3.** Since the synthetic data that we invert are created using the same transport model that we use for the inversion, this assumes that, in an inver-

sion of real data, we would be able to perfectly model atmospheric transport. Clearly, this is unrealistic. We explore the weakening of this assumption through two cases. In the first case (test 3a), we assume that we are able to model the variability in concentration successfully but that the phasing is wrong. Specifically, we temporally shift the time series of synthetic data at each site by a random shift of up to  $\pm 1$  d. We perform 40 inversions with different random shifts to provide a distribution of inversion impact. In the second case (test 3b), we assume that the transport model achieves the correct timing of ‘events’ but that it underestimates or overestimates the magnitude of non-baseline excursions. We approximate this as follows. An approximately 10-d running mean is removed from each time series and the standard deviation of the residuals is calculated. All residuals that are at least one standard deviation from the running mean are multiplied by a random number in the range  $-1.0$  to  $1.0$ . Thus, at the extremes, the non-baseline event could be removed or doubled in magnitude. The same factor is applied to the whole time series, but different factors are applied to different sites. Again 40 inversions are performed to give a distribution of inversion impact.

### 3. Results

First, we present inversion results using ‘clean’ data (i.e. the original unmodified synthetic data) averaged over different time periods. This provides a reference against which the sensitivity tests can be assessed. Figure 2 shows Australian region (total of large region plus 45 grid cells) source estimates for six inversions with uncertainties for three cases. The 4-h case is the same as that presented in L02. The inversion with daily data is very close to that from four-hour data. As the averaging period is further increased there is some degradation of the estimates, particularly in February and May. The uncertainty also increases as the averaging period is lengthened. The reason for this must be the response functions, since the only other terms in eq. (2) (prior source and data uncertainties) are unchanged. This can be shown experimentally by performing inversions that preserve the high-frequency data but use time-averaged responses. The uncertainty retrieved is equivalent to that obtained using time-averaged data. L02 conjectured a reason for this, namely that the fluctuations in the transport at high frequencies would help locate sources in space. This is supported by the variation of RMSU as a function of the averaging period

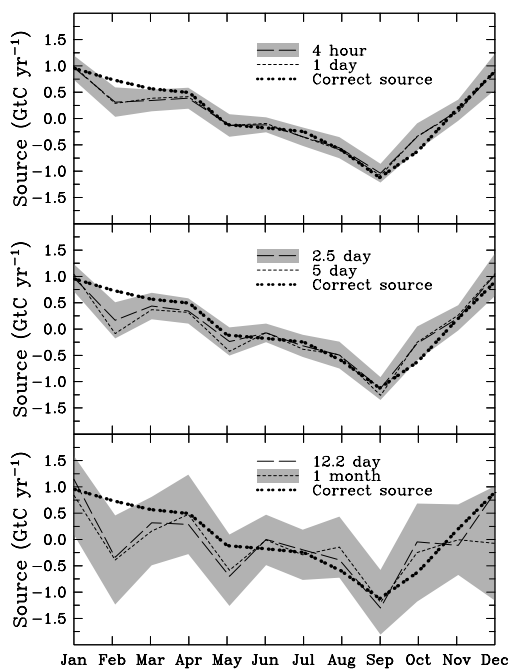


Fig. 2. Correct and estimated Australian region sources. Two inversions are shown in each panel as indicated in the key. The shaded region shows the estimated uncertainties for one inversion in each panel.

(Table 2). Dominant variations of transport occur on synoptic time-scales (2–5 d) and provided these are preserved in the response functions the RMSU is relatively low, but it increases sharply once these signals are averaged away.

RMSB for the six cases is also given in Table 2. In each case RMSB is less than RMSU, which indicates that the estimated uncertainty is providing a realistic measure of the inversion capability. Given the argument in section 2.1, this also suggests that our choice

Table 2. Australian source RMSB and RMSU for inversions using clean data

| Time period | RMSB  | RMSU  |
|-------------|-------|-------|
| 4-h         | 0.168 | 0.225 |
| 1-d         | 0.169 | 0.238 |
| 2.5-d       | 0.214 | 0.273 |
| 5-d         | 0.301 | 0.325 |
| 12.2-d      | 0.424 | 0.459 |
| 1-month     | 0.499 | 0.730 |

of data uncertainty is reasonable. It is worth noting here that the monthly case is not equivalent to current inversions using flask data because of our use of smaller data uncertainties.

### 3.1. Test 1

Figure 3 shows examples of two of the inversion sets with offsets up to  $\pm 0.5$  ppm. In the 4-h case the estimated Australian sources are still reasonably representative of the correct sources. In the monthly case, the estimated sources are very scattered and bear little resemblance to the correct source. RMSB statistics for all the 1a and 1b tests are shown in Fig. 4 along with the clean data RMSB and RMSU. RMSU is not altered by changing data values, so the clean data RMSU is applicable to all inversions at the relevant averaging period. For all except the monthly,  $\pm 0.5$  ppm case, the minimum RMSB is smaller than the clean data RMSB. It appears that some combinations of offsets can improve the inversion by compensating for other errors in the inversion such as the use of large regions outside of Australia. Overall the impact of the offsets on

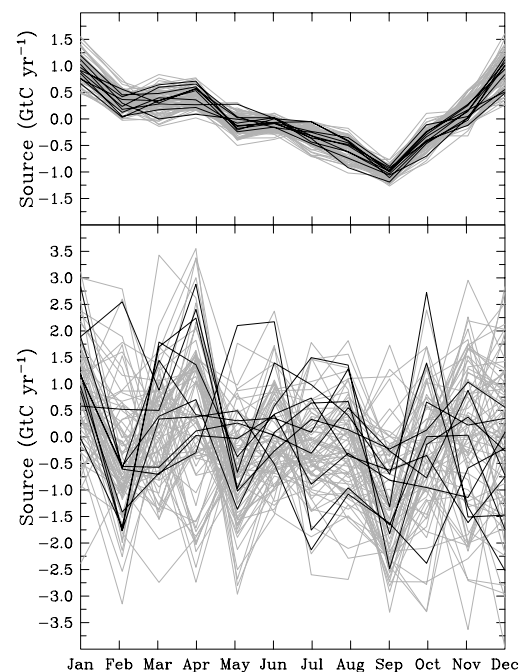


Fig. 3. Estimated Australian region sources for the set of inversions of 4-h data (top) and 1-month data (bottom) both with up to  $\pm 0.5$  ppm offsets. For clarity most cases are plotted in grey with ten example cases shown in black.

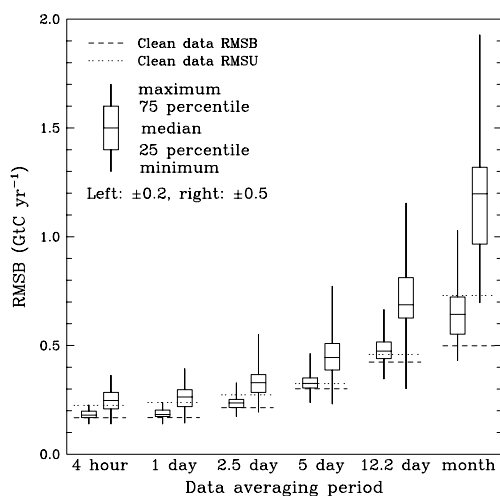


Fig. 4. Root mean square bias (RMSB) for the Australian region source from the 1a (left) and 1b (right) sensitivity tests. The horizontal lines are the RMSB and RMSU for inversions with clean data. The boxes and whiskers represent the minimum, 25th percentile, median, 75th percentile and maximum RMSB for the sets of inversions at each time period.

the inversion is greater at longer averaging periods. At 4-h and 1-d periods all the  $\pm 0.2$  ppm cases give RMSB less than RMSU, which suggests that intercalibration differences of up to 0.2 ppm could be accommodated in inversions using very frequent data. Only one  $\pm 0.2$  ppm case, the 12.2-d inversion, gives the median RMSB greater than RMSU, whereas all the  $\pm 0.5$  ppm cases have larger median RMSB than RMSU. At high data frequencies (4-hourly and daily), the estimated sources with  $\pm 0.5$  ppm offsets are still preferable to those estimated from longer time periods (over 5 d) and smaller offsets ( $\pm 0.2$  ppm).

### 3.2. Test 2

The RMSB and RMSU for the Australian region are shown in Fig. 5 for the addition of a perfectly calibrated continental site (case 2a). As expected the RMSU is decreased at all time periods. RMSB is also reduced, by a proportionally slightly greater amount than the reduction in RMSU. The reductions also tend to be proportionally larger at the longer time periods. The results from case 2b, in which the extra site is no longer assumed to be well calibrated, are also shown in Fig. 5. There is no significant increase in RMSB for case 2b relative to 2a for time periods shorter

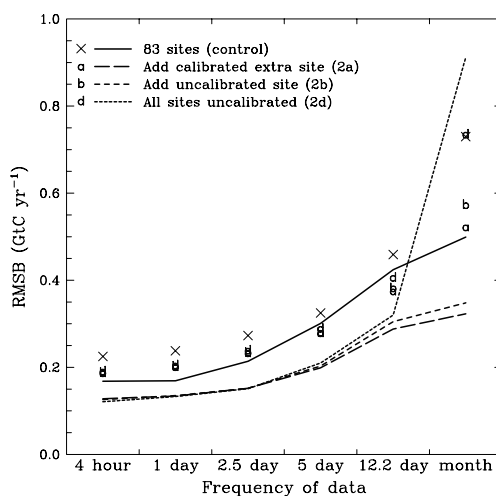


Fig. 5. RMSB (lines) and RMSU (symbols) for the Australian region source for the 2a, 2b and 2d sensitivity tests. The line and symbol identification is given by the key.

than 5 d, while at the 12.2-d and one-month periods there is a small increase in RMSB. RMSU shows similar behaviour. The inversion also provides an estimate of the unknown calibration for this site. In this case, we did not intentionally offset the data, so the inversion was attempting to retrieve a value of zero. The estimated offsets were  $-0.10 \pm 0.09$ ,  $0.02 \pm 0.13$ ,  $0.03 \pm 0.14$ ,  $0.06 \pm 0.18$ ,  $0.17 \pm 0.25$  and  $0.15 \pm 0.33$  ppm going from shortest (4-h) to longest (one-month) time periods, respectively. The uncertainty on the estimated offset increases with the increasing averaging period of the data. In all but the 4-h case, the correct offset (zero) lies within the estimated range of offsets. Some additional cases (2c) were run in which a non-zero offset was added to the extra continental site. Since the non-zero offsets tested were no larger than the prior uncertainty chosen for the unknown calibration ( $\pm 2$  ppm), there was essentially no change to the estimated sources. Estimated offsets were also as expected, e.g. inversions using 4-h and 1-d data with an imposed offset of 2.0 ppm gave estimates of  $1.90 \pm 0.09$  and  $2.01 \pm 0.13$  ppm, respectively.

A more extreme case (2d), in which we estimated the relative calibration of all sites, is also shown in Fig. 5. Perhaps surprisingly there is little impact on the RMSB for Australia for all averaging periods except one-month. The one-month result is poor with RMSB greater than RMSU. The large RMSB is due to a shift in source estimates in all months of about  $1 \text{ GtC yr}^{-1}$ ; the

seasonal cycle of sources is still estimated reasonably. In almost all the inversions the estimated offsets are within the estimated uncertainty of the correct (zero) offsets for the sites around Australia. One exception is Darwin at all time periods greater than 4-h, e.g. the 5-d inversion gives  $0.32 \pm 0.25$  ppm. Whereas in the 2b case the uncertainty on the estimated offset decreased with decreasing averaging period, here the uncertainty is reasonably constant at 0.23–0.24 ppm for all periods less than 5 d at many sites.

### 3.3. Test 3

The results from the tests to weaken the ‘perfect transport’ assumption are shown in Fig. 6. Case 3a, with temporally shifted time series, is shown by the left box of each pair while case 3b, the amplified or reduced ‘events’ case, is shown by the right-hand boxes. Note that the y-axis range is half that of Fig. 4. For case 3a, there is less impact on RMSB as the averaging period of the data is increased in the inversion. This is reasonable as temporal shifts of up to one day will be more significant in 4-h or 1-d data than when averaged over longer periods. The smallest median RMSB is for the 2.5-d case due to the compromise between lower RMSB with more frequent data and less degradation as

averaging period increases. The distribution of RMSB also tends to be wider for the shorter averaging periods.

The impact on RMSB for case 3b is qualitatively similar with smaller impact at the longer time periods. Unlike case 3a, the median RMSB remains smaller than RMSU for all time periods. The spread in RMSB increases for the shorter time periods. In the 4-h case, the larger RMSBs are associated with a weakening of the seasonal cycle of the estimated sources. In particular the February to April sources and August to October sinks are underestimated. These larger RMSB cases are also those cases where the random factor applied to non-baseline events resulted in most sites around Australia having their non-baseline events reduced. This seems reasonable since, particularly for coastal sites, non-baseline samples are usually those that have had recent land contact, while baseline data are sampling marine air. Thus, reducing the magnitude of non-baseline events is effectively reducing the impact of Australian sources on the concentrations at sites and so smaller Australian sources and sinks result. The opposite also occurs. When the non-baseline events are increased at most Australian sites, the seasonality of estimated sources is increased. However, this reduces rather than increases RMSB because the sources estimated with clean data slightly underestimated source seasonality. Thus an increase in seasonality gives estimates that are closer to the correct source. This result is contingent on the particular biases in the Australian region inversions and may not be globally applicable.

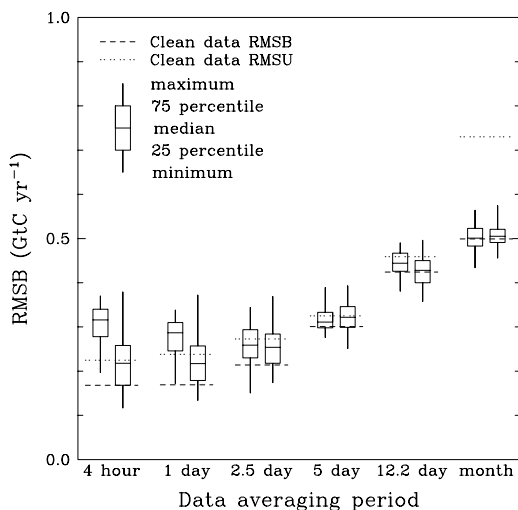


Fig. 6. RMSB for the Australian region source from the 3a (left) and 3b (right) sensitivity tests. The horizontal lines are the RMSB and RMSU for inversions with clean data. The boxes and whiskers represent the minimum, 25th percentile, median, 75th percentile and maximum RMSB for the sets of inversions at each time period.

## 4. Discussion and conclusions

Overall the tests focussed around calibration issues showed that inversions using short averaging period data are less susceptible to problems than those using longer averaging periods. The tests focussed around transport errors showed that some data averaging may be necessary to minimise source biases from inversions. If these results for the Australian region are representative of behaviour globally then it suggests that the initial focus for inverting continuous measurements should be on the synoptic timescale (2–5 d). It is worth noting that, although flask data are available bi-weekly at some sites, these data are unlikely to contribute to the source estimation to the same extent as continuous measurements since they are usually collected to avoid non-baseline signals. Ideally any continuous measurements used by inversions should be part of a well calibrated network, but our tests have

shown that less well calibrated data (such as that collected at flux tower sites) may still be useful in this inversion application, providing the calibration does not significantly drift in time. In fact, the inversion may be able to provide some estimate of any calibration offset.

While these results all seem positive, a number of issues still need to be tackled before we can begin to test inversions with real data. These synthetic data tests have used a 'cyclo-stationary' inversion in which the winds are assumed to be the same from year to year. Clearly this is not the case in reality, and so a time-dependent inversion is required. Using a synthesis inversion method, this is computationally expensive if winds are allowed to vary from year to year, especially with the higher spatial resolution that we know is required to correctly use the high frequency data. Thus we need to explore in what ways we can reduce the computational cost of this approach without losing important information, much as we have been able to successfully make use of the short 3-month responses here for the Australian grid-cells.

A second major question relates to the impact of sub-monthly variations in source, including diurnal variations. It is likely that some of the power of our approach relies on variations of sources which are slow relative to variations in transport. The current tests that we have performed use 'square' monthly pulses to cal-

culate the response functions (i.e. constant emissions through the month are assumed), while the synthetic data are created from mid-month sources that are linearly interpolated between mid-months. In reality we are likely to have both synoptic scale and diurnal variations in sources. We need to know whether we can ignore these source variations in the inversion or whether some of the variation needs to be estimated by the inversion. Finally, consideration needs to be given to other types of transport model error not included in this study.

The future for CO<sub>2</sub> inversions would appear to lie in using high frequency data, either from the surface or from satellites (Rayner et al., 2002) or both. While high-frequency observations remain relatively limited, synthetic data tests are a useful way forward for exploring both the potential of such data and the challenges to our inversion methodology that we are most likely to encounter. Here we have shown that high data frequency inversions are potentially more robust to some measurement limitations than inversions of monthly data.

## 5. Acknowledgments

We thank Roger Francey for helpful comments on this manuscript.

## REFERENCES

- Andres, R. J., Marland, G., Fung, I. and Matthews, E. 1996. A  $1^\circ \times 1^\circ$  distribution of carbon dioxide emissions from fossil fuel consumption and cement manufacture, 1950–1990. *Global Biogeochem. Cycles* **10**, 419–429.
- Bolin, B. and Keeling, C. D. 1963. Large-scale atmospheric mixing as deduced from the seasonal and meridional variations of carbon dioxide. *J. Geophys. Res.* **68**, 3899–3920.
- Bousquet, P., Peylin, P., Ciais, P., Le Quééré, C., Friedlingstein, P. and Tans, P. P. 2000. Regional changes in carbon dioxide fluxes of land and oceans since 1980. *Science* **290**, 1342–1346.
- Enting, I. G. 2000. Green's function methods of tracer inversion. In: *Inverse methods in global biogeochemical cycles* (eds. P. Kasibhatla, M. Heimann, P. Rayner, N. Mahowald, R. Prinn and D. Hartley). *Geophysical Monograph* 114, American Geophysical Union, Washington, DC, 19–31.
- Francey, R. J., Rayner, P. J. and Allison, C. E. 2001. Constraining the global carbon budget from global to regional scales – the measurement challenge. In: *Global biogeochemical cycles in the climate system* (eds. E.-D. Schulze, M. Heinemann, S. Harrison, E. Holland, J. Lloyd, I. C. Prentice and D. Schimel). Academic Press, San Diego, CA, 245–252.
- GLOBALVIEW-CO<sub>2</sub>. Cooperative Atmospheric Data Integration Project – Carbon Dioxide. CD-ROM, NOAA CMDL, Boulder, Colorado 2001. [Also available on Internet via anonymous FTP to ftp.cmdl.noaa.gov, Path: ccg/co2/GLOBALVIEW].
- 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. 2002. Towards robust regional estimates of CO<sub>2</sub> sources and sinks using atmospheric transport models. *Nature* **415**, 626–630.
- Kaminski, T., Rayner, P. J., Heimann, M. and Enting, I. G. 2001. On aggregation errors in atmospheric transport inversions. *J. Geophys. Res.* **106**, 4703–4715.
- Law, R. M., Rayner, P. J., Steele, L. P. and Enting, I. G. 2002. Using high temporal frequency data for

- CO<sub>2</sub> inversions. *Global Biogeochem. Cycles* **16**(4), 1053, doi:10.1029/2001GB001593/.
- Masarie, K. A., Langenfelds, R. L., Allison, C. E., Conway, T. J., Dlugokencky, E. J., Francey, R. J., Novelli, P. C., Steele, L. P., Tans, P. P., Vaughn, B. and White, J. W. C. 2001. NOAA/CSIRO Flask air intercomparison experiment: A strategy for directly assessing consistency among atmospheric measurements made by independent laboratories. *J. Geophys. Res.* **106**, 20445–20464.
- Pearman, G. 1993. Survey of results of the 1991–1992 round-robin inter-laboratory intercomparison. In: *Report of the seventh WMO meeting of experts on carbon dioxide concentration and isotopic measurement techniques* (eds G. I. Pearman and J. T. Peterson). Rep. No. 88, Global Atmos. Watch, World Meteorological Organization, Geneva, 104–107.
- Peterson, J., Tans, P. and Kitzis, D. 1997. CO<sub>2</sub> round-robin reference gas intercomparison. In: *Report of the ninth WMO meeting of experts on carbon dioxide concentration and related tracer measurement techniques* (ed R. Francey). Rep. No. 132, Global Atmos. Watch, World Meteorological Organization, Geneva, 30–33.
- Potosnak, M. J., Wofsy, S. C., Denning, A. S., Conway, T. J., Munger, J. W. and Barnes, D. H. 1999. Influence of biotic exchange and combustion sources on atmospheric CO<sub>2</sub> concentrations in New England from observations at a forest flux tower. *J. Geophys. Res.* **104**, 9561–9569.
- Randerson, J. T., Thompson, M. V., Conway, T. J., Fung, I. Y. and Field, C. B. 1997. The contribution of terrestrial sources and sinks to trends in the seasonal cycle of atmospheric carbon dioxide. *Global Biogeochem. Cycles* **11**, 535–560.
- Rayner, P. J., Law, R. M., O'Brien, D. M., Butler, T. M. and Dille, A. C. 2002. Global observations of the carbon budget: 3. Initial assessment of the impact of satellite orbit, scan geometry and cloud on measuring CO<sub>2</sub> from space. *J. Geophys. Res.* **107**(021), 4557, doi:10.1029/2001JD000618
- Running, S. W., Baldocchi, D. D., Turner, D. P., Gower, S. T., Bakwin, P. S. and Hibbard, K. A. 1999. A global terrestrial monitoring network integrating tower fluxes, flask sampling, ecosystem modeling and EOS satellite data. *Remote Sens. Environ.* **70**, 108–127.
- Takahashi, T., Sutherland, S. C., Sweeney, C., Poisson, A., Metzl, N., Tilbrook, B., Bates, N., Wanninkhof, R., Feely, R. A., Sabine, C., Olafsson, J. and Nojiri, Y. 2002. Global sea–air CO<sub>2</sub> flux based on climatological pCO<sub>2</sub>, and seasonal biological and temperature effects. *Deep-Sea Research, Part II, Topical Studies in Oceanography* **49**, 1601–1622.
- World Meteorological Organization (WMO) 1995. Annex 5.10. The CO<sub>2</sub> measurement programme of NOAA CMDL, U.S.A. In: *Report of the eighth WMO meeting of experts on carbon dioxide concentration and isotopic measurement techniques* (ed T. J. Conway). Rep. No. 121, Global Atmos. Watch, World Meteorological Organization, Geneva, 97–102.