# Synoptic variations in atmospheric $CO_2$ at Cape Grim: a model intercomparison

*By* RACHEL M. LAW[1]*, L. PAUL STEELE[1], PAUL B. KRUMMEL[1] and WLODEK ZAHOROWSKI[2], [1]*Centre for Australian Weather and Climate Research, CSIRO Marine and Atmospheric Research, PMB 1, Aspendale, Vic 3195, Australia;* [2]*Australian Nuclear Science and Technology Organisation, Kirrawee, NSW 2232, Australia*

(Manuscript received 14 December 2009; in final form 10 June 2010)

### ABSTRACT

A 'TransCom' model intercomparison is used to assess how well synoptic and diurnal variations of carbon dioxide ($CO_2$) and $^{222}$Rn (radon) can be modelled at the coastal site, Cape Grim, Australia. Each model was run with prescribed fluxes and forced with analysed meteorology for 2000–2003. Twelve models were chosen for analysis based on each model's ability to differentiate baseline $CO_2$ concentrations from non-baseline $CO_2$ (influenced by regional land fluxes). Analysis focused on non-baseline events during 2002–2003. Radon was better simulated than $CO_2$, indicating that a spatially uniform radon land flux is a reasonable assumption and that regional-scale transport was adequately captured by the models. For both radon and $CO_2$, the ensemble model mean generally performed better than any individual model. Two case studies highlight common problems with the simulations. First, in summer and autumn the Cape Grim observations are sometimes influenced by Tasmanian rather than mainland Australian fluxes. These periods are poorly simulated. Secondly, an event with an urban plume demonstrates how the relatively low spatial resolution of the input $CO_2$ fluxes limits the quality of the simulations. Analysis of periods with below baseline concentration indicates the possible influence of carbon uptake by winter crops in southern mainland Australia.

## 1. Introduction

In situ measurements of atmospheric $CO_2$ show large diurnal and synoptic variations, particularly for coastal and continental sites. The variations result from temporally varying fluxes of carbon to or from the atmosphere combined with atmospheric transport from a source region to an observing site. We would like to be able to use the atmospheric $CO_2$ measurements to estimate carbon fluxes, but this requires a good simulation of the atmospheric transport. Until recently, atmospheric inversions (e.g. Baker et al., 2006) have been run at monthly temporal resolution, with a range of atmospheric transport models used to provide an estimate of flux uncertainty due to transport uncertainty. However, synthetic data experiments (e.g. Law et al., 2002) have shown that fluxes can be estimated with smaller uncertainties as the frequency of observations is increased. Hence current inversions (e.g. Peters et al., 2007; Lauvaux et al., 2009; Rödenbeck et al., 2009) are attempting to use more frequent atmospheric $CO_2$ data. This places greater demands on the quality

of atmospheric transport models and so the $CO_2$ data are often selected for times when the $CO_2$ is expected to be more easily modelled.

For low-altitude, continental sites, it is common to only use day-time data (Peters et al., 2007) since this is when atmospheric mixing is more vigorous and the $CO_2$ measurements are likely to be regionally representative. At mountain sites night-time data is often selected because this is representative of regional fluxes while day-time data may be influenced by local fluxes due to upslope transport.

Coastal sites present different problems. Here, the sites were originally chosen to sample marine air, in order to provide a 'baseline' trend and seasonal cycle that was representative of the latitude of the site and not contaminated by local land biosphere or anthropogenic sources. It is baseline-selected data that have generally been used in monthly mean inversions, with the baseline selection being site dependent. However, the non-baseline data contain equally useful flux information from surrounding land regions. To use these data in an inversion, we cannot simply select by time of day; we need to know under what circumstances our transport models are able to reliably distinguish between baseline and non-baseline air. This has been investigated for Cape Point, South Africa by Whittlestone et al. (2009). Here,
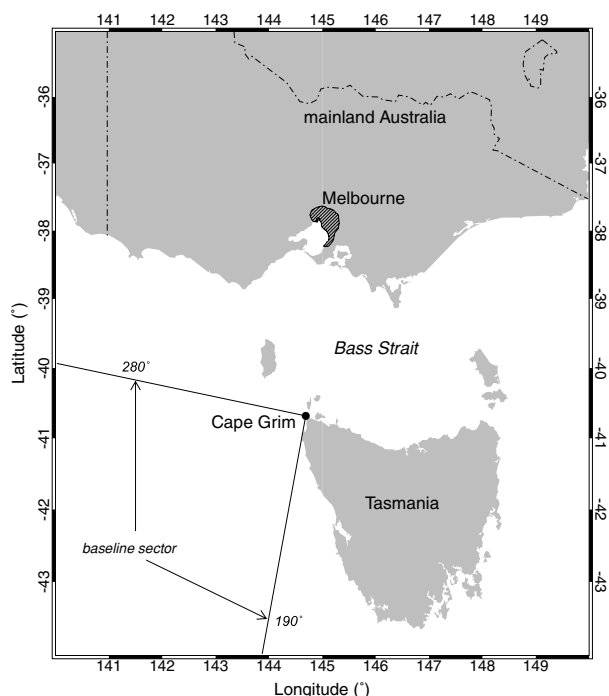
*Fig. 1.* Location of Cape Grim, Tasmania and the Melbourne urban area. The wind direction sector for baseline selection is also shown.

we will investigate this question by assessing the performance of a range of global atmospheric models at another coastal site, Cape Grim, Australia.

Cape Grim (Fig. 1 ) is a key Southern Hemisphere site, measuring an extensive range of greenhouse gases, ozone-depleting chemicals and aerosols, with some records dating back to the late 1970s. The baseline sector at Cape Grim samples southern ocean air, providing very clean records for detecting long-term trends in trace gas concentrations (e.g. Francey et al., 2010). Non-baseline air is influenced by southern Australia, including the city of Melbourne. $CO_2$ observations from the non-baseline sector have been used to constrain biosphere flux estimates from southeastern Australia (Wang and McGregor, 2003). There have been a number of limited area modelling studies for various trace gas species for Cape Grim. Kowalczyk and McGregor (2000) modelled $CO_2$ and radon at 75 km resolution, while other studies have used higher resolution models (down to 5 km) to focus on urban tracers, for example, carbon monoxide (CO; McGibbony et al., 2005) and dichloromethane (Cox et al., 2003). Kowalczyk and McGregor (2000) found that radon was better simulated than $CO_2$. McGibbony et al. (2005) focuses on the impact of seasonal differences in frontal dynamics on the transport of polluted air to Cape Grim.

## 2. Methods

During 2006–2007 a model intercomparison project ('TransCom-continuous', http://www.purdue.edu/transcom/

T4_continuousSim.php) was conducted to assess the ability of atmospheric models to simulate diurnal and synoptic variations in atmospheric $CO_2$. An overview of the results has been presented by Law et al. (2008) and Patra et al. (2008). Here we focus on the model simulations for a single site.

### 2.1. TransCom experiment

The experiment is described in Law et al. (2008) with more technical details available in the experiment protocol (Law et al., 2006b). In brief, each atmospheric model was run for the years 2000–2003; off-line (trace gas transport only) models were forced with analysed meteorology while for on-line (full climate) models, the horizontal wind fields (and sometimes temperature) were nudged to analyses. The first 2 yr are discarded as spin-up. The years 2002 and 2003 are used for analysis. At the time the TransCom protocol was written, these were the most recent years with $CO_2$ observations available.

Twenty-five models or model variants participated in the intercomparison, four of which were regional models run for portions of the Northern Hemisphere only. The complete list of models is given in Law et al. (2008) and the complete model data set is available by anonymous ftp (ftp-kg01.eas.purdue.edu). Nine separate trace gases were simulated; fossil $CO_2$ (Olivier and Berdowski, 2001; Marland et al., 2003), ocean $CO_2$ (Takahashi et al., 2002), sulphur hexafluoride ($SF_6$) (http://www.rivm.nl/edgar), $^{222}$Rn (radon) and five biospheric $CO_2$ tracers. The biospheric $CO_2$ tracers were from two biosphere models, SiB3 (Baker et al., 2009) and CASA (Randerson et al., 1997; Olsen and Randerson, 2004). The SiB3 fluxes were run with hourly, daily and monthly temporal resolution. The CASA fluxes were run at three-hourly and monthly resolution. The ocean $CO_2$ fluxes had monthly temporal resolution while the fossil emissions were constant in time. For radon, we used a constant flux from land surfaces between 60°S and 60°N of $1.66 \times 10^{-20}$ mol m$^{-2}$ s$^{-1}$ (1 atom cm$^{-2}$ s$^{-1}$), a flux from ocean surfaces between 70°S and 70°N of $8.30 \times 10^{-23}$ mol m$^{-2}$ s$^{-1}$ and applied radioactive decay with a half-life of 3.8 d.

### 2.2. Time series processing

For the analysis here, we compare modelled and observed $CO_2$, using radon to interpret the $CO_2$ time series. Radon is a valuable tracer of non-baseline air since its source from land surfaces is much larger than from ocean surfaces and it has a relatively short atmospheric lifetime. For modelled atmospheric $CO_2$, we use the sum of the fossil and ocean $CO_2$ tracers plus either the CASA or SiB diurnally varying biosphere tracer. When not explicitly stated, the $CO_2$ sum with CASA is the case used for analysis. We should note that the modelled $CO_2$ is expected to capture most diurnal and synoptic variations of $CO_2$ at a site but will not match the observed trend since not all components of the $CO_2$ budget are modelled (e.g. any net biospheric sink). We therefore

limit our comparison between modelled and observed $CO_2$ to residuals from baseline.

We fit the baseline trend and seasonal cycle to the modelled concentrations for 2002–2003 using an iterative procedure, as follows. A quadratic plus harmonics, $C_{fit}$, is fitted

$$
\begin{aligned}
C_{fit} = & a_1 + a_2 t + a_3 t^2 + a_4 \cos(2\pi t) + a_5 \sin(2\pi t) \\
& + a_6 \cos(4\pi t) + a_7 \sin(4\pi t) \\
& + a_8 \sin(6\pi t) + a_9 \cos(6\pi t),
\end{aligned}
\tag{1}
$$

where $t$ is time in years (i.e. from 0 to 2) and $a_n$ are the constants determined by the fit. The standard deviation of the residuals from the fit is calculated. Any data points that are more than two standard deviations from the fit are discarded and the reduced data set is again fitted and a new standard deviation calculated. The process is repeated until no further data points are excluded from the fit. In practice, we limit the number of iterations to 40, at which point the differences in the fit from one iteration to the next are less than 0.01 ppm at any hour in the time series. The final fitted concentrations are subtracted from the modelled $CO_2$ to give the $CO_2$ residuals from baseline.

### 2.3. Site and observations

Cape Grim is on the northwest coast of Tasmania (40.7°S, 144.7°E), south of mainland Australia (Fig. 1). The site is at the top of a 90 m cliff and the air inlet for the in situ $CO_2$ measurements is from 164 m asl. During 2002–2003, in situ measurements of atmospheric $CO_2$ were being made at Cape Grim by two separate $CO_2$ analyser systems. One system (known as BASGAM) was based upon a Siemens Ultramat 5E non-dispersive infrared gas analyser. The other (known as LoFlo) is based upon a Li-Cor 6261 non-dispersive infrared gas analyser (Steele et al., 2004, 2006). These two systems were being operated side-by-side over an extended period to compare their performance.

The differences in the instrumental records are very small (±0.2 ppm) compared to the differences between the modelled and observed $CO_2$. For this analysis we have used the LoFlo record, filling any gaps with BASGAM measurements if these were available. Hourly average data from both instruments are used in this study. The hourly values are computed from 1 min average values. Only those data which have been accepted by a test of optimal instrument performance are used. Using the same procedure as for the modelled $CO_2$, the baseline trend and seasonal cycle is removed from the observations to produce a time series of $CO_2$ residuals from baseline (Fig. 2). The merging of the LoFlo and BASGAM time series was performed after removal of the baseline fit since this also removes small (±0.1 ppm) slowly varying differences between the two analysers. These differences are linked to allocation of concentration values to the working gas standards required by the analysers. The time series of residuals shows many positive deviations from baseline, with maximum deviations around 10–20 ppm. Nega-
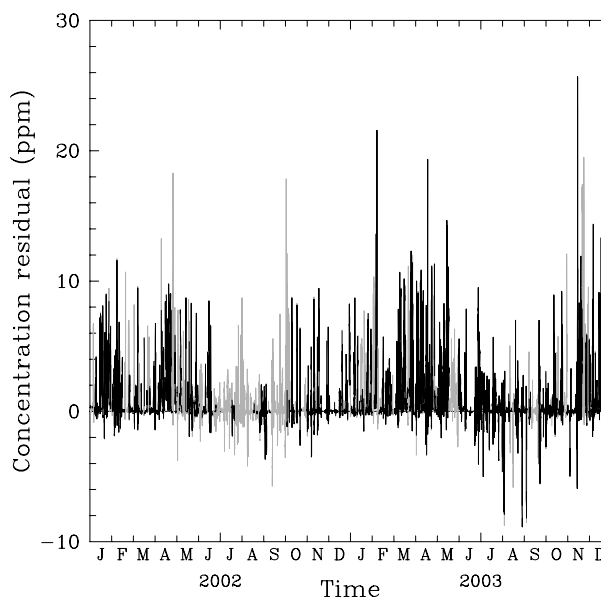


*Fig. 2.* Observed $CO_2$ residuals from baseline at Cape Grim for the LoFlo (black) and BASGAM (grey) analysers. The *x*-axis is marked with the initial letters of each month of 2002 and 2003.

tive residuals are less common, occur predominantly in winter and spring (June–November) and are more frequent and more negative in 2003 than 2002.

Hourly radon observations were made at Cape Grim using a 5000 L dual flow loop, two-filter radon detector (Whittlestone and Zahorowski, 1998), equipped with two sensing heads. The air intake, at a rate of approximately 288 L min$^{-1}$ was from an inlet at 70 m agl (164 m asl). The sensitivity of the detector was 0.66 counts per second per Bq m$^{-3}$. Calibrations were performed using a Pylon flow-through radon source (20.9 ± 0.8 kBq radium-226), traceable to NIST standards.

There are a number of ways to define whether any given air sample is considered baseline or not. While the fitting procedure described above produces a baseline trend and seasonal cycle, it does not categorise individual hours as baseline or not. For this categorisation we use one of the baseline definitions employed at Cape Grim for $CO_2$. Two conditions (wind direction and $CO_2$ stability) must be met: (1) the hourly mean wind direction is between 190 and 280 degrees and (2) the hourly mean $CO_2$ concentration is one of a group of at least five consecutive hours with variations in hourly mean $CO_2$ concentrations of less than 0.3 ppm. More rigorous definitions of baseline may also include a minimum wind speed requirement and iterative removal of outliers (Steele et al., 2003) but these were not included here.

### 2.4. Model selection

We made a preliminary assessment of each model submission by checking how well the models differentiated between baseline and non-baseline periods. We chose to test this differentiation by comparing the mean diurnal cycle of $CO_2$ concentration
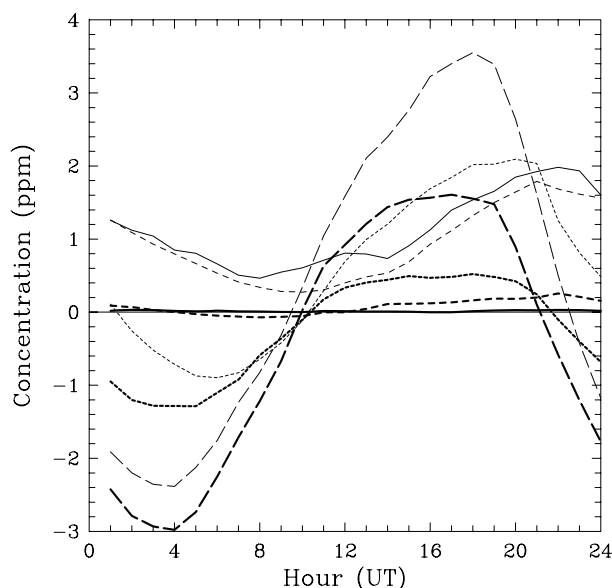
*Fig. 3.* Mean diurnal cycle of $CO_2$ at Cape Grim for all baseline hours (bold) and non-baseline hours (standard) for BASGAM observations (solid), and example model submissions for a land grid point (long-dash), an ocean grid point (short-dash) and for interpolation to the location of Cape Grim (dot).

calculated separately for baseline and non-baseline hours (with baseline hours defined using the observed wind direction and $CO_2$ stability criteria above). Since any diurnal cycle is driven primarily by continental $CO_2$ fluxes, we expect little or no diurnal cycle for baseline hours and a small diurnal cycle (after transport from mainland Australia) for non-baseline hours. Fig. 3 shows that this is true for observed $CO_2$ with a diurnal peak-to-peak amplitude of 0.03 ppm for baseline hours and 1.5 ppm for non-baseline hours.

The experiment protocol asked modellers to submit both an ocean location and a land location to represent Cape Grim, since it was not clear which would best represent the site for any given model. Most modellers chose to submit their nearest ocean and land grid points, whereas other modellers interpolated their model output to the location of Cape Grim and to another offshore location. This choice makes a large difference to the simulated diurnal cycles, illustrated for some example models in Fig. 3. Land grid points tend to give much larger amplitude diurnal cycles than observed, for both baseline (4–11 ppm) and non-baseline hours (4–16 ppm). This is also true for models that interpolated to the Cape Grim location; the inclusion of land grid points in the interpolation causes overestimated diurnal cycles, particularly for baseline hours (1–3 ppm). For most models the ocean grid point gives the best match to the observed diurnal cycles.

Based on these results, models were selected using a criterion that the diurnal amplitude for non-baseline hours should be at least three times larger than for baseline hours. While different criteria could have been chosen, this one focuses on the differentiation between baseline and non-baseline periods. This selection gave twelve model submissions for further analysis (Table 1). Of these 12, nine were for the submitted ocean grid point. The remaining three (AM2, AM2t and PCTM.CSU) were for their land submissions, since these gave diurnal cycles closer to those observed than their ocean submissions. In these three models there were no emissions (biosphere or fossil) for the land location chosen and diurnal cycles were consequently small for baseline hours, similar to other models' ocean grid points.

### 2.5. Baseline events

One way to consider the observed $CO_2$ record is as a series of non-baseline events separated by periods of baseline sampling.

*Table 1.* Selected models

| Model | Horizontal resolution[a] | Reference |
|---|---|---|
| AM2.GFDL | $2.5° \times 2.0°$ | GFDL GAMDT (2004) |
| AM2t.GFDL | $2.5° \times 2.0°$ | GFDL GAMDT (2004) |
| CCAM.CSIRO | $\sim$220 km | Law et al. (2006a) |
| CCSR_NIES1.FRCGC | T42 ($\sim$2.8° $\times$ 2.8°) | Takigawa et al. (2005) |
| CCSR_NIES2.FRCGC | T106 ($\sim$1.1° $\times$ 1.1°) | Takigawa et al. (2005) |
| IMPACT.LLNL | $2.5° \times 2.0°$ | Rotman et al. (2004) |
| NICAM.CCSR | $\sim$240 km | Satoh et al. (2008) |
| PCTM.CSU | $1.25° \times 1.0°$ | Kawa et al. (2004) |
| PCTM.GSFC | $2.5° \times 2.0°$ | Kawa et al. (2004) |
| STAG.AIST | $1.125° \times 1.125°$ | Wada et al. (2007) |
| STAGN.AIST | T63 ($\sim$1.88° $\times$ 1.90°) | Sawa et al. (2007) |
| TM3_vfg.MPIBGC | $1.875° \times 1.875°$ | Heimann (1996) |

[a]Longitude $\times$ latitude or distance or spectral resolution indicated by T (triangular) maximum wave number.
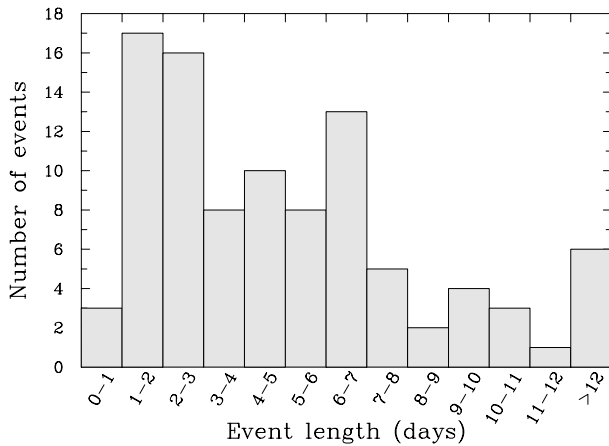
*Fig. 4.* Histogram of the length of events in days.

Typically an event might last for a few days and may encompass concentrations that are both above and below baseline. The definition of an event is somewhat arbitrary. Here we take a two-step approach. Since we are looking for non-baseline events that occur between sustained periods of baseline sampling, we first search for all periods of at least 12 consecutive baseline hours. The midpoints of these baseline periods are used to define the start and end of a period that contains a potential non-baseline event. We are only interested here in those events that are of sufficient length and amplitude to be captured by the global-scale models used in this analysis. Therefore, from these potential events we select only those with at least 12 h of non-baseline data and non-baseline concentrations that span at least 2 ppm. We also exclude potential events with more than 50% missing data during non-baseline hours. Using the LoFlo-BASGAM merged observations we find 167 potential events, which reduces to 96 events when the selection criteria are applied. Most of the excluded events fail both the event length and amplitude checks. The selected events are generally between 1 and 7 d long (Fig. 4 ), with four events longer than 20 d.

The modelled $CO_2$ residual time series are divided into the same events as the observations. An ensemble model mean is also created for each event by averaging the $CO_2$ residuals for the 12 models. The time series of modelled and observed radon are also divided into the same events. Six of the $CO_2$ events have no observed radon data available. Again an ensemble model mean is created by averaging the 12 modelled radon time series.

# 3. Results

We focus our analysis on the correlation between modelled and observed $CO_2$ for each of the 96 selected events in 2002 and 2003. The correlation coefficient, $R$, is defined as

$$R = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}}, \qquad (2)$$

where the sum is over the number of non-missing hours in the event, $x$ and $y$ are the modelled and observed $CO_2$ concentrations and $\bar{x}$ and $\bar{y}$ are the mean $CO_2$ concentrations. The correlation gives a measure of how well we model the temporal evolution of an event. Any errors in the temporal evolution are likely to be hard to correct by a $CO_2$ inversion since a transport error may be involved. By contrast, a poor simulation of the magnitude of an event could more easily be corrected since the inversion is designed to adjust the flux magnitude to best fit the observed concentrations. Events with poor correlation may need to be excluded from an inversion or at least given less weight.

## 3.1. Correlation of individual models and ensemble

Figure 5 shows the correlation between modelled (CASA+ fossil+ocean) and observed $CO_2$ for each event and each model. The events (*x*-axis) are sorted for each model so that the correlations are plotted from most positive to most negative, in order to more easily make comparisons. For most events and most models, the correlations are low; typically less than 20% of events give correlations greater than 0.6. The ensemble model mean generally gives correlations at the upper end of the model spread. All models give a small number of events with negative correlation. These indicate that the transport is reasonably modelled but there is a problem with the prescribed fluxes. This is most likely when there are competing fluxes, i.e. a fossil source and a biospheric sink.

One model, CCAM.CSIRO, performs poorly relative to the other models. At least for the events with higher correlation, the problem with the CCAM.CSIRO model often appears to be a lag, behind the other models and observations, particularly as concentrations return to baseline following a peak or trough. Subsequent tests show that the CCAM.CSIRO results can be
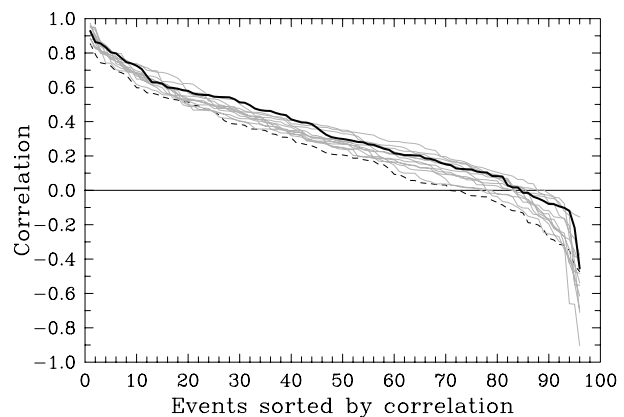


*Fig. 5.* Correlation between observed and modelled $CO_2$ for each event. The events are sorted from highest to lowest correlation, for individual models (grey, thin lines), the CCAM.CSIRO model (black, dashed line) and the ensemble model mean (black, thick line). The modelled $CO_2$ uses the CASA, fossil and ocean fluxes.

improved by changing the way the model is nudged to the analyses.

The correlation between the modelled CO$_2$ and observations when the SiB fluxes are used (not shown) are broadly similar to those that included the CASA fluxes. The performance of the ensemble mean across events seems to be slightly better. For example, 21 events give a correlation greater than 0.6 compared to 15 events when the CASA fluxes are used. The better performance may be because the SiB fluxes are hourly whereas the CASA fluxes are three-hourly. For individual models, the results are mixed. CCAM.CSIRO performs better with the SiB fluxes and, while still amongst the poorer performing models, is no longer an outlier. TM3_vfg.MPIBGC performs worse with the SiB fluxes than the CASA fluxes.

### 3.2. Paired comparisons

In our 12 selected models there are two versions of a number of models. These cases are useful for testing whether any particular element of a model set-up clearly improves the simulation. We focus our analysis on the 21 events that are best modelled across all 12 models, defined as those events that give a correlation of at least 0.3 for all models. Four paired comparisons are available. AM2 is an online model, nudged to NCEP analyses. For AM2 only the zonal (u) and meridional (v) wind components are nudged while for AM2t, temperature is also nudged. Figure 6(a) shows generally small differences (less than 0.05 change in correlation) between the two AM2 simulations. Of the six events with correlation differences greater than 0.05, five give a lower correlation when the temperature nudging is included.

Two pairs of models differ primarily in their horizontal resolution: CCSR_NIES1 ($\sim$2.8°) compared to CCSR_NIES2 ($\sim$1.1°), and PCTM.CSU (1.25° $\times$ 1.0°) compared to PCTM.GSFC (2.5° $\times$ 2.0°). The PCTM models also differ in the frequency of their meteorological forcing with the GSFC version using 3 hourly forcing while the CSU version uses 6 hourly forcing. Figs. 6b and c show no obvious improvement due to higher resolution in the CCSR_NIES model but in the PCTM case, the correlations are generally higher when higher spatial resolution is used. For some events (6, 13, 14) the two PCTM cases span close to the full range of correlations across models with the high resolution PCTM at the top of the range.

The two STAG models differ in their forcing meteorology with STAG using ECMWF forcing and STAGN using NCEP forcing. Along with the change in forcing is a change in resolution, with STAG having higher horizontal and vertical resolution. As with the CCSR_NIES pair, there is little evidence to indicate that one STAG simulation performs consistently better than the other (Fig. 6d).

These paired comparisons suggest that the model simulations for Cape Grim may be more limited by the spatial resolution of the input flux fields than by the resolution of the simulation. We explore this idea further in Section 3.4. This finding is likely to be
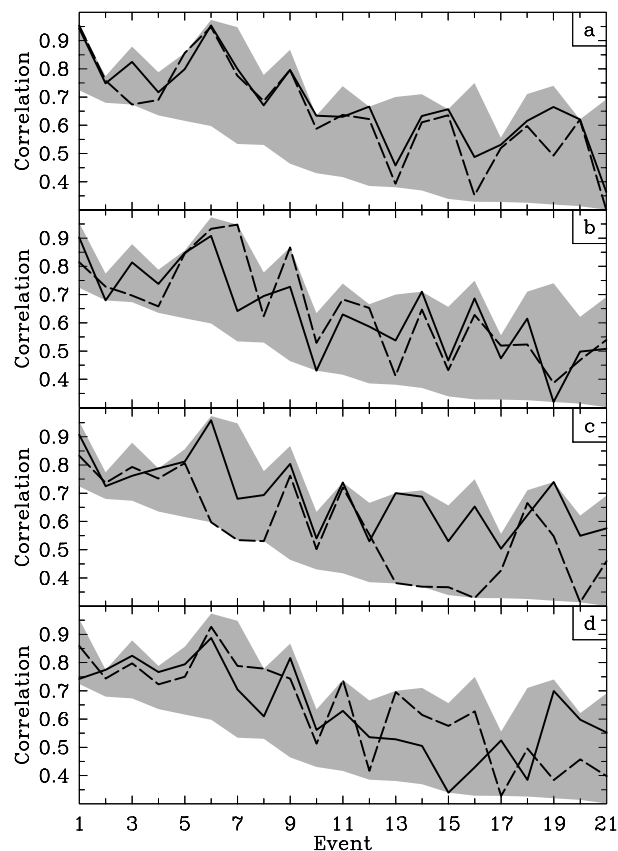


*Fig. 6.* Correlation between modelled and observed CO$_2$ for the 21 events with correlation greater than 0.3 for all 12 models for (a) AM2 (solid) and AM2t (dashed), (b) CCSR_NIES1 (solid) and CCSR_NIES2 (dashed), (c) PCTM.CSU (solid) and PCTM.GSFC (dashed) and (d) STAG (solid) and STAGN (dashed). The shaded region shows the range of correlation across all 12 models for each event. The events are sorted by the minimum correlation across all models.

site dependent. Patra et al. (2008) found an improved simulation at most sites with higher spatial resolution but concluded that this was mostly due to a higher resolution model being able to sample nearer to the true observing location. This is probably not a factor in this analysis of Cape Grim since the model selection criteria removed models that sampled too far from the site to give plausible simulations.

### 3.3. Comparison with radon

The modelled radon simulations allow the modelled transport to be tested for a simpler trace gas than CO$_2$. Fig. 7 shows that the correlations between the ensemble model mean and observed radon are larger than for CO$_2$ for more events; around 70% of events give correlations greater than 0.6. This is consistent with Kowalczyk and McGregor (2000) who found monthly correlations between observed and modelled radon ranging from 0.60–0.80 compared to 0.21–0.61 for CO$_2$. The CO$_2$ correlations
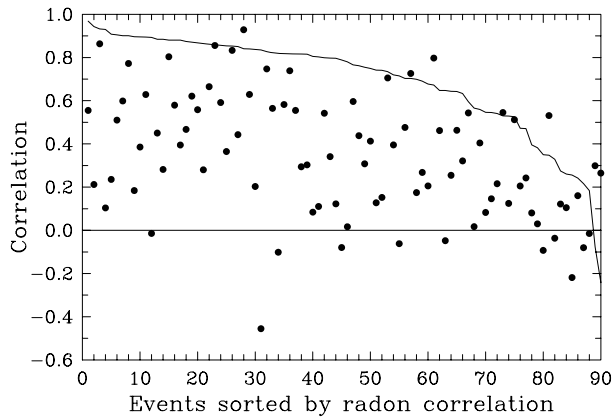
*Fig. 7.* Correlation between observed and modelled radon (line) and observed and modelled $CO_2$ (CASA+fossil+ocean) (dots) for each event. The events are sorted from highest to lowest radon correlation.

from the ensemble mean model case are also shown in Fig. 7 sorted as for the radon events. This shows that there are very few events for which the $CO_2$ correlation is larger than the radon one. It is also clear that a well-modelled radon event does not ensure that a $CO_2$ event is also well-modelled. The difference is primarily due to the radon flux being a reasonably uniform source to the atmosphere from all land surfaces while the $CO_2$ flux is from multiple processes, has higher spatial variability and changes between a source and a sink, often on diurnal timescales. This is illustrated in the second case study below.

### 3.4. Case studies

We use two case studies to illustrate an event where $CO_2$ and radon show similar problems in the simulation and an event where radon is well simulated but $CO_2$ is poorly simulated. We use the ensemble model mean for this analysis but the results would be similar for any of the individual models.

The first case occurs in March 2003 (Figs. 8a and b). The event consists of three smaller peaks followed by a longer non-baseline period with higher concentrations. In both the radon and $CO_2$

case the main non-baseline period is reasonably well modelled, leading to a model-observations correlation of 0.90 for radon and 0.77 for $CO_2$. However the early peaks in the event are almost completely missed by the model simulations. This type of behaviour appears to be relatively common in summer and autumn. The smaller peaks are often associated with air that has passed across or close to Tasmania, while the period of highest concentrations is for air that has passed over mainland Australia. It is clear that the models, with typically 100–200 km horizontal resolution, are finding it difficult to resolve Tasmania.

To test this, we have used the wind direction at each observation time to exclude any observations when the wind direction was between $70°$ and $190°$ (around 26% of the observations). We have then recalculated the event correlations between the ensemble model mean and observations without the Tasmanian sector data. We find that 31% of events show higher correlation by at least 0.1, 80% of which occur in summer and autumn. The number of events for which the correlation between the ensemble mean modelled $CO_2$ and observed $CO_2$ is greater than 0.6 increased from 16 to 30%. Clearly many summer and autumn events are currently impacted by the inability to correctly simulate the contribution of Tasmanian fluxes to the $CO_2$ and radon concentration at Cape Grim.

The second case (Figs. 8c and d) occurs in August 2003. For this event, radon was well simulated ($R^2 = 0.93$) while $CO_2$ was poorly simulated ($R^2 = 0.21$ including the CASA tracer and $R^2 = 0.29$ including the SiB tracer). The $CO_2$ observations show rapid transitions from positive to negative residuals and vice versa. The model ensemble mean fails to capture this; in the CASA case, the positive residual on 12 August is completely missed while the SiB case gives a small positive concentration that is earlier than observed.

The good radon simulation suggests that the transport is modelled reasonably through this event so it is likely that the problem is with the $CO_2$ fluxes. The hourly SiB fluxes give a slightly better result than the three-hourly CASA ones, so the temporal resolution of the fluxes is likely to be one problem. The main feature of the observations that is missed in the model simulations is the large positive residual on August 12. The timing of this peak at
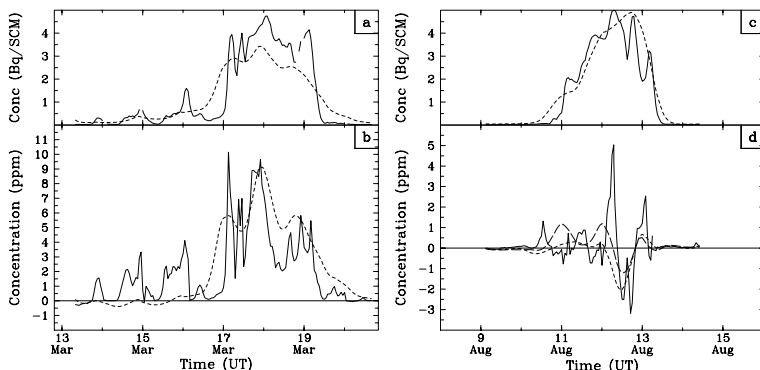


*Fig. 8.* Observed (solid) and ensemble mean modelled (dashed) radon (a, c) and $CO_2$ (b, d) for 13–20 March 2003 (a, b) and 9–15 August 2003 (c, d). The modelled $CO_2$ is the CASA case (short dash) and the SiB case (long dash, shown in panel d only).
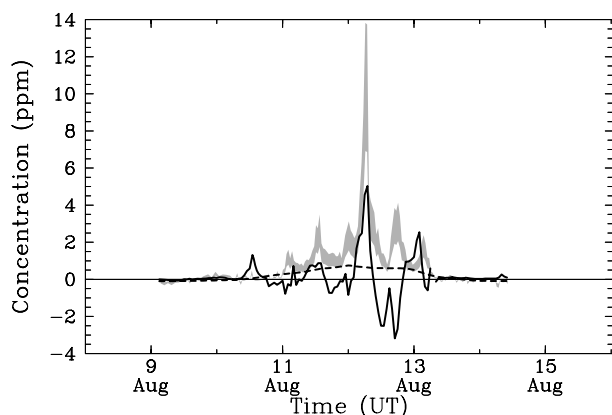
*Fig. 9.* Observed CO$_2$ (solid) and ensemble mean modelled fossil CO$_2$ (dashed) for 9–15 August 2003. The shaded region is observed CO, scaled to CO$_2$ using a ratio of 10 ppb.ppm$^{-1}$ (upper bound) or 20 ppb.ppm$^{-1}$ (lower bound).

1600 local time suggests that it is unlikely to be biospheric in origin, even accounting for around 6 h of transport time from mainland Australia. Back-trajectories (Draxler and Hess, 1998) suggest that the air parcel has travelled across Melbourne, and consequently would have a significant fossil fuel component. This is confirmed by CO measurements from Cape Grim that show an enhancement of around 140 ppb above baseline at this time. If a CO to fossil fuel CO$_2$ ratio of 10–20 ppb.ppm$^{-1}$ is assumed (Gamnitzer et al., 2006), this corresponds to a CO$_2$ enhancement of 7–14 ppm. The bottom end of this range is more consistent with the CO$_2$ observations assuming a modest offset by biospheric uptake.

Since separate simulations were performed for each CO$_2$ tracer, we can check the fossil simulation for this period. Fig. 9 shows that the ensemble model mean shows a broad, small concentration maximum for fossil CO$_2$ of only around 0.7 ppm. It is clear that the fossil flux fields, which were provided at 1° × 1° resolution and then interpolated to each model grid, do not have sufficient spatial resolution to simulate this type of event at Cape Grim.

We have investigated whether this is a common problem through the simulation by testing whether the correlation for each event is improved if periods of high CO are excluded. We remove all hours for which the CO is at least 50 ppb above baseline. For the event shown here, this improves the correlation in the CASA case from 0.21 to 0.51 and in the SiB case from 0.29 to 0.37. However this magnitude of improvement is not typical. Many correlations show little change and some are made worse when high CO hours are removed. Events with longer periods of high CO are often modelled reasonably for CO$_2$, suggesting that the flux resolution only becomes critical when an air parcel has brief contact with urban areas.

The CO record also shows some periods of very high CO, which are often coincident with high CO$_2$, methane (CH$_4$) and hydrogen (H$_2$). It is likely that these periods are affected by biomass burning. Since burning is not included in the model simulations of CO$_2$, these periods are not well simulated. For example, a long event (22 d) in November 2003 has multiple CO$_2$ peaks; some are reasonably modelled while others show evidence of biomass burning and are missed by the model. The CO$_2$ correlation between model mean and observations is 0.31 (CASA) and 0.23 (SiB) compared to 0.75 for radon. Removing periods with high H$_2$ or CO increases the correlation for CO$_2$ by 0.1–0.2. Removing the Tasmanian wind sector has a similar impact, indicating that the fires were local rather than from mainland Australia.

### 3.5. Drawdown periods

The CO$_2$ observations at Cape Grim show a number of occasions when the concentration is below baseline. These periods of drawdown are due to CO$_2$ uptake by the biosphere, predominantly over continental Australia. It is useful to understand how well we can simulate these drawdown periods to determine whether we can use the Cape Grim record to estimate biospheric uptake over southern Australia. We first isolate drawdown periods by identifying all occasions when the CO$_2$ concentration drops more than 2 ppm below baseline. We find 52 periods (Fig. 10a); the majority are at least 2 ppm below baseline for 3 h or less



*Fig. 10.* Time of day and time of year for the minimum concentration of each period when the (a) observed and (b) ensemble model mean CO$_2$ concentration is at least 2 ppm below baseline. The length of the period below 2 ppm is less than 3 h (+), 3–9 h (o), greater than 9 h (x).
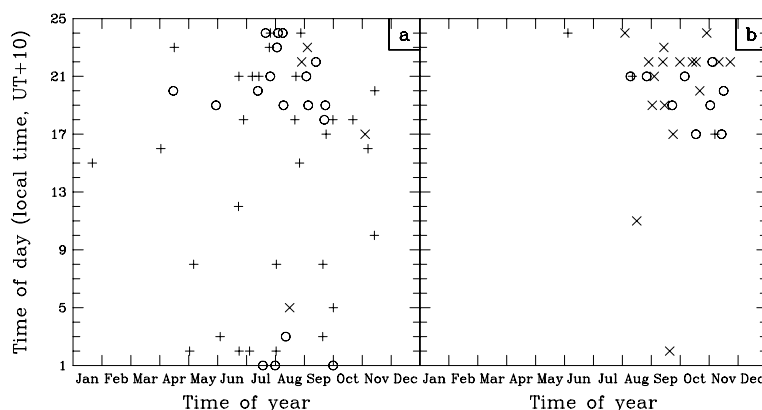
(58%) with only 8% where the concentration is below baseline for at least 9 h. Most cases occur from July to September (62%) and have their minimum concentration between 1800 and 300 local time (69%). This time of day is consistent with the low concentration air coming from continental Australia in the afternoon, assuming a 6–8 h transit time for the ∼200 km crossing of Bass Strait at a typical wind speed of 10 ms$^{-1}$.

The characteristics of the drawdown periods for the modelled $CO_2$ are somewhat different than for the observations (Fig. 10b). For the case using CASA fluxes, the ensemble mean model gives only 29 periods in which the concentration drops more than 2 ppm below baseline. The modelled low concentration periods tend to be longer than those observed, with 59% having concentrations more than 2 ppm below baseline for 9 h or more. The time of day of the minimum concentration is consistent with the observations but is more tightly constrained, with all but two cases lying between 1700 and 2400 local time. This suggests that the diurnal cycle of carbon uptake in the CASA fluxes and the subsequent transport to Cape Grim is modelled reasonably.

The seasonal difference in modelled and observed drawdown periods is worth noting. The modelled drawdown periods tend to occur slightly later in the year (97% between August–November) than observed. This suggests that the CASA fluxes do not properly represent the seasonality of carbon uptake for southern Australia, with uptake beginning later in winter and extending further into spring than the observations suggest. One possibility for the difference in seasonality is that CASA models natural ecosystems while the observed seasonality may be dominated by winter crops, such as wheat, barley and canola. There are large areas of crops to the west and northwest of Melbourne which are likely to be traversed by continental air reaching Cape Grim. Gervois et al. (2004) has shown, for two Northern Hemisphere sites, that modelling the phenology of winter wheat shifts carbon uptake earlier in the year compared to natural grasslands, consistent with the shift in timing that we would need to match the Cape Grim observations. Whittlestone et al. (2009) also found that the impact of winter wheat crops was detectable in the $CO_2$ observations at Cape Point, South Africa.

The ensemble model mean using the SiB fluxes gives a poorer simulation of drawdown periods with only four periods meeting the 2 ppm below baseline criterion. It is apparent that the SiB model does not simulate sufficient photosynthetic uptake over southern Australia. This is confirmed by comparing the diurnal amplitude of monthly mean SiB and CASA fluxes averaged over all land points between 135–147°E and 35–39°S. The CASA fluxes give diurnal amplitudes ranging from 4.3 to 22 $\mu$mol m$^{-2}$ s$^{-1}$ compared to only 2.5–8.6 $\mu$mol m$^{-2}$ s$^{-1}$ for the SiB fluxes.

One drawdown period is noteworthy due to its unusual timing. The episode occurs on 16–17 August 2003 and is shown in Fig. 11. Concentrations are 2 ppm below baseline from about 1700 16 August to 700 17 August UT (300–1700 17 August local time), much later in the day than is typical. Back-trajectories
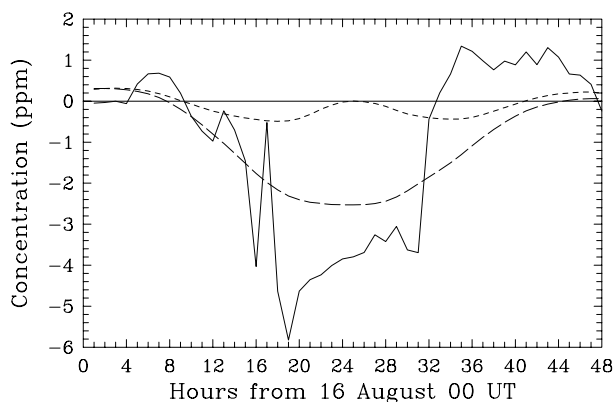


*Fig. 11.* Observed $CO_2$ (solid) and ensemble mean modelled $CO_2$ including CASA fluxes (long dash) and SiB fluxes (short dash) for 16–17 August 2003.

suggest that transport is consistently from the region west of Melbourne for up to 24 h around this time, with some indication that the air parcels have come from a few hundred metres above the ground over the continent. This elevation of the trajectories may explain why the concentration residuals remain negative for longer than expected; positive nighttime concentrations may be trapped closer to the ground and not transported to Cape Grim. The episode is reasonably well simulated by the ensemble mean model using CASA fluxes but not with the SiB fluxes (individual models show a similar contrast between the CASA and SiB cases). The CASA fluxes for 16 August show more uptake than is typical for August; the diurnal amplitude of the flux for the 135–147°E, 35–39°S region is 12.5 $\mu$mol m$^{-2}$ s$^{-1}$ for 16 August compared to 9.2 $\mu$mol m$^{-2}$ s$^{-1}$ for the August mean. It appears that above-average uptake on 16 August was important for simulating the unusual drawdown episode that was observed. The diurnal amplitude of the SiB flux for 16 August was 6.7 $\mu$mol m$^{-2}$ s$^{-1}$, which was clearly insufficient to simulate the drawdown episode at Cape Grim.

## 4. Conclusions

Continuous measurements of $CO_2$ at coastal sites provide valuable records of both baseline $CO_2$ concentrations and concentrations that are influenced by local and regional land fluxes. Extracting this flux information from the non-baseline concentrations remains challenging. Here we have assessed the ability of global transport models to simulate $CO_2$ and radon concentrations at Cape Grim. We have found that the choice of sampling location in the model is critical; indiscriminately interpolating to the site location is not appropriate for a coastal site, rather an ocean location is required. Both Kowalczyk and McGregor (2000) and Whittlestone et al. (2009) used a weighted interpolation based on wind direction which they both found to be suitable. Such a choice might be useful for models which have a fractional land–sea mask, for which the nearest pure ocean grid point may be some distance from the true site location.

The ensemble model mean performs as well or better than any individual model; there was no clear evidence that the models with ∼1° resolution performed better than those with ∼2° resolution. The radon simulation indicated that atmospheric transport is reasonably well simulated, although we should note that some transport errors may be missed because the radon land source is approximately uniform. The quality of the CO$_2$ simulations was limited by the spatial and temporal resolution of the input CO$_2$ fluxes. Higher spatial resolution CO$_2$ fluxes would assist with resolving contributions from Tasmania to the CO$_2$ record at Cape Grim and would justify using a higher resolution transport model.

In situ records of other trace gases are very valuable for interpreting the CO$_2$ record. At Cape Grim, high CO often indicated urban air, while very high CO and high H$_2$ indicated biomass burning, a CO$_2$ flux that was missing from these simulations. Case studies of individual non-baseline events have been useful in determining why the models failed in certain circumstances, but generalising these cases is not always easy. An analysis of drawdown periods highlighted problems with the seasonality and magnitude of carbon uptake in the prescribed biospheric fluxes for southern Australia, most likely due to carbon uptake by winter crops which is not modelled.

While using non-baseline CO$_2$ concentrations from coastal sites in CO$_2$ flux inversions remains difficult, this study has provided some insights into how a coastal record might be selected to only use the CO$_2$ observations that can be most reliably modelled. Poorly modelled radon is indicative of poorly modelled CO$_2$. Local, rather than regional, land influence might be excluded through a wind direction selection. Other trace gases help to identify when CO$_2$ observations are influenced by different CO$_2$ fluxes (e.g. urban or biomass burning); these observations can then be excluded if the associated fluxes would not be resolved by the inversion. Hourly CO$_2$ observations will be increasingly valuable as interest grows in regional, higher resolution CO$_2$ flux inversions; consequently, this type of study will be required for each observing site, in order to understand any modelling limitations and to ensure that reliable flux estimates are obtained.

## 5. Acknowledgments

The TransCom-continuous model simulations were provided by AM2/AM2t (S. Fan, S.-J. Lin), CCAM (R. M. Law), CCSR_NIES (P. Patra, M. Takigawa), CDTM (T. Maki), CHIMERE (C. Aulagnier, L. Rivier, R. Vautard), COMET (G. Pieterse, A. T. Vermeulen), DEHM (C. Geels, J. Brandt, J. H. Christensen), IFS (S. Serrar), IMPACT (D. J. Bergmann, P. J. Cameron-Smith), LMDZ/LMDZ_THERM (P. Bousquet, F. Delage), NICAM (Y. Niwa, R. Imasu, M. Satoh), NIES05 (S. Maksyutov, R. Onishi), PCTM.CSU (R. Lokupitiya, A. S. Denning, N. Parazoo, J. Kleist), PCTM.GSFC (S. R. Kawa, Z. Zhu), REMO (U. Karstens), STAG/STAGN (S. Taguchi), TM3_fg/TM3_vfg (C. Rödenbeck), TM5_glb3x2/TM5_nam1x1 (W. Peters, L. Bruhwiler), TM5_eur1x1 (M. C. Krol, S. Houweling).

## References

Baker, D. F., Law, R. M., Gurney, K. R., Rayner, P., Peylin, P. and co-authors. 2006. TransCom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional CO$_2$ fluxes, 1988–2003. *Global Biogeochem. Cycles* **20**(1), GB1002, doi:10.1029/2004GB002439.

Baker, I. T., Denning, A. S., Prihodko, L., Schaefer, K., Berry, J. A. and co-authors. 2009. SiB3 Modeled Global 1-degree Hourly Biosphere-Atmosphere Carbon Flux, 1998-2006. Data set., Available on-line [http://www.daac.ornl.gov] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA, doi:10.3334/ORNLDAAC/909.

Cox, M., Siems, S., Fraser, P., Hurley, P. and Sturrock, G. 2003. TAPM modelling studies of AGAGE dichloromethane observations at Cape Grim. In: *Baseline Atmospheric Program (Australia) 1999-2000* (eds N. W. Tindale, N. Derek and P. J. Fraser. Bureau of Meteorology and CSIRO Atmospheric Research), Melbourne, 25–30.

Draxler, R. R. and Hess, G. D. 1998. An overview of the HYSPLIT_4 modeling system for trajectories, dispersion, and deposition. *Aust. Met. Mag.* **47**, 295–308.

Francey, R. J., Trudinger, C. M., van der Schoot, M., Krummel, P. B., Steele, L. P. and co-authors. 2010. Differences between trends in atmospheric CO$_2$ and the reported trends in anthropogenic CO$_2$ emissions. *Tellus* this issue, doi:10.1111/j.1600-0889.2010.00472.x.

Gamnitzer, U., Karstens, U., Kromer, B., Neubert, R. E. M., Meijer, H. A. J. and co-authors. 2006. Carbon monoxide: a quantitative tracer for fossil fuel CO$_2$?. *J. Geophys. Res.* **111**, D22302. doi:10.1029/2005JD006966.

Gervois, S., de Noblet-Ducoudré, N., Viovy, N. and Ciais, P. 2004. Including croplands in a global biosphere model: methodology and evaluation at specific sites. *Earth Interact.* **8**(16), 1–25.

GFDL Global Atmospheric Model Development Team 2004. The new GFDL global atmosphere and land model AM2-LM2: Evaluation with prescribed SST simulations. *J. Climate* **17**, 4641–4673.

Heimann, M. 1996. The global atmospheric tracer model TM2, *Technical report no. 10, Deutsches Klimarechenzentrum, Hamburg, Germany (ISSN 0940-9327)*.

Kawa, S. R., Erickson D. J. III, Pawson, S. and Zhu, Z. 2004. Global CO$_2$ transport simulations using meteorological data from the NASA data assimilation system. *J. Geophys. Res.* **109**, D18312. doi:10.1029/2004JD004554.

Kowalczyk, E. A. and McGregor, J. L. 2000. Modeling trace gas concentrations at Cape Grim using the CSIRO Division of Atmospheric Research Limited Area Model (DARLAM). *J. Geophys. Res.* **105**(D17), 22 167–22 183.

Lauvaux, T., Gioli, B., Sarrat, C., Rayner, P. J., Ciais, P. and co-authors. 2009. Bridging the gap between atmospheric concentrations and local ecosystem measurements. *Geophys. Res. Lett.* **36**, L19809. doi:10.1029/2009GL039574.

Law, R. M., Rayner, P. J., Steele, L. P. and Enting, I. G. 2002. Using high temporal frequency data for $CO_2$ inversions. *Global Biogeochem. Cycles* **16**, 1053. doi:10.1029/2001GB001593.

Law, R. M., Kowalczyk, E. A. and Wang, Y. P. 2006a. Using atmospheric $CO_2$ data to assess a simplified carbon-climate simulation for the 20th century. *Tellus* **53B**, 427–437.

Law, R. M., Peters, W., Rödenbeck, C., Aulagnier, C., Baker, I. and co-authors. 2008. TransCom model simulations of hourly atmospheric $CO_2$: Experimental overview and diurnal cycle results for 2002. *Global Biogeochem. Cycles* **22**, GB3009. doi:10.1029/2007GB003050.

Law, R., Peters, W. and Rödenbeck, C. 2006b. Protocol for the TransCom continuous data experiment. *Purdue climate change research center tech. rep.* http://www.purdue.edu/climate/publications/T3%5B1%5D.continuous.protocol_v6.pdf.

Marland, G., Boden, T. A. and Andres, R. J. 2003. Global, regional, and national annual $CO_2$-emissions from fossil-fuel burning, cement production, and gas flaring: 1751-2002, *NDP-030*, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory. http://cdiac.esd.ornl.gov/ndps/ndp030.html.

McGibbony, S. C., Siems, S. T., Physick, W. L. and Hurley, P. 2005. The sensitivity of simulations of air pollution events at Cape Grim to the modelled meteorology. *Aust. Met. Mag.* **54**, 321–331.

Olivier, J. G. J. and Berdowski, J. J. M. 2001. Global emissions sources and sinks. In: *The Climate System*, (eds J. J. M. Berdowski, R. Guicherit and B. J. Heij), A.A. Balkema Publishers/Swets & Zeitlinger Publishers, Lisse, The Netherlands, 33–78.

Olsen, S. C. and Randerson, J. T. 2004. Differences between surface and column atmospheric $CO_2$ and implications for carbon cycle research. *J. Geophys. Res.* **109**, D02301. doi:10.1029/2003JD003968.

Patra, P. K., Law, R. M., Peters, W., Rödenbeck, C., Takigawa, M. and co-authors. 2008. Transcom model simulations of hourly atmospheric $CO_2$: analysis of synoptic scale variations for the period 2002-2003. *Global Biogeochem. Cycles* **22**, GB4013. doi:10.1029/2007GB003081.

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J. and co-authors. 2007. An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker. *Proc. Natl. Acad. Sci. USA* **104**, 18 925–18 930.

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.

Rödenbeck, C., Gerbig, C., Trusilova, K. and Heimann, M. 2009. A two-step scheme for high-resolution regional atmospheric trace gas inversions based on independent models. *Atmos. Chem. Phys.* **9**, 5331–5342. www.atmos-chem-phys.net/9/5331/2009/.

Rotman, D. A., Atherton, C. S., Bergmann, D. J., Cameron-Smith, P. J., Chuang, C. C. and co-authors. 2004. IMPACT, the LLNL 3-D global atmospheric chemical transport model for the combined troposphere and stratosphere: Model description and analysis of ozone and other trace gases. *J. Geophys. Res.* **109**, D04303. doi:10.1029/2002JD003155.

Satoh, M., Matsuno, T., Tomita, H., Miura, H., Nasuno, T. and co-authors. 2008. Nonhydrostatic Icosahedral Atmospheric Model (NICAM) for global cloud resolving simulations. *J. Computational Phys.* **227**, 3486–3514. doi:10.1016/j.jcp.2007.02.006.

Sawa, Y., Tanimoto, H., Yonemura, S., Matsueda, H., Wada, A. and co-authors. 2007. Widespread pollution events of carbon monoxide observed over the western North Pacific during the EAREX 2005 campaign. *J. Geophys. Res.* **112**, D22S26. doi:10.1029/2006JD008055.

Steele, L. P., Krummel, P. B., da Costa, G. A., Spencer, D. A., Porter, L. W. and co-authors. 2003. Baseline carbon dioxide monitoring. In: *Baseline Atmospheric Program Australia 1999-2000*, (eds. N. W. Tindale, N. Derek and P. J. Fraser), Bureau of Meteorology and CSIRO Atmospheric Research, Melbourne, Australia, 80–84.

Steele, L. P., Krummel, P. B., Spencer, D. A., Porter, L. W., Baly, S. B. and co-authors. 2004. Baseline carbon dioxide monitoring. In: *Baseline Atmospheric Program Australia 2001-2002*, (eds. J. M. Cainey, N. Derek and P. B. Krummel), Bureau of Meteorology and CSIRO Atmospheric Research, Melbourne, Australia, 36–40.

Steele, L. P., Krummel, P. B., Spencer, D. A., Porter, L. W., Baly, S. B. and co-authors. 2006. Baseline carbon dioxide monitoring. In: *Baseline Atmospheric Program Australia 2003-2004*, (eds. J. M. Cainey, N. Derek and P. B. Krummel), Australian Bureau of Meteorology and CSIRO Atmospheric Research, Melbourne, Australia, 39–42.

Takahashi, T., Sutherland, S. C., Sweeney, C., Poisson, A., Metzl, N. and co-authors. 2002. Global sea-air $CO_2$ flux based on climatological surface ocean $pCO_2$, and seasonal biological and temperature effects. *Deep-Sea Research, Part II, Topical Studies in Oceanography* **49**, 1601–1622.

Takigawa, M., Sudo, K., Akimoto, H., Kita, K., Takegawa, N. and co-authors. 2005. Estimation of the contribution of intercontinental transport during the PEACE campaign by using a global model. *J. Geophys. Res.* **110**, D21313. doi:10.1029/2005JD006226.

Wada, A., Sawa, Y., Matsueda, H., Taguchi, S., Murayama, S. and co-authors. 2007. Influence of continental air mass transport on atmospheric $CO_2$ in the western North Pacific. *J. Geophys. Res.* **112**, D07311. doi:10.1029/2006JD007552.

Wang, Y.-P. and McGregor, J. L. 2003. Estimating regional terrestrial carbon fluxes for the Australian continent using a multiple-constraint approach. II. The atmospheric constraint. *Tellus* **55B**(2), 290–304. doi:10.1034/j.1600-0560.2003.00030.x.

Whittlestone, S., Kowalczyk, E., Brunke, E.-G. and Labuschagne, C. 2009. Source regions for $CO_2$ at Cape Point assessed by modelling, [222]Rn and meteorological data, *South African Weather Service, Pretoria, South Africa*. ISBN 978-0-9584463-9-6, http://ro.uow.edu.au/scipapers/170/.

Whittlestone, S. and Zahorowski, W. 1998. Baseline radon detectors for shipboard use: development and deployment in the first aerosol characterisation experiment (ACE 1). *J. Geophys. Res.* **103**, 16 743–16 751.