Tellus (1996), 48B, 474–486
Printed in UK – all rights reserved

TELLUS
ISSN 0280–6509

The selection of model-generated CO_2 data: a case study with seasonal biospheric sources

By RACHEL LAW, *Cooperative Research Centre for Southern Hemisphere Meteorology, Monash University, Clayton 3168, Australia*

(Manuscript received 23 October 1995; in final form 1 April 1996)

ABSTRACT

We assess whether data selection is required when comparing modelled concentrations with observed data that have been selected for 'background' conditions. To do this, the Melbourne University Tracer Model was forced with a set of seasonally varying biospheric CO_2 sources. The variation in predicted seasonal cycles of CO_2 concentrations up to 500 km from CO_2 monitoring sites was used as an indication of how regionally representative the model data is for each location. Where the variation is due to land influences this was reduced by removing air with recent land contact from the model output. Modelled radon-222 concentration from a simulation using the same wind fields as the CO_2 case was found to provide an effective selection method. Selection generally reduces regional variability. The impact on the seasonal cycle, however, is relatively small and there are only a small number of sites where selection clearly improves the comparison with the observed seasonal cycle.

1. Introduction

The potential climate impact of increasing CO_2 in the atmosphere has stimulated much work to better understand the carbon cycle. One aspect of this work has been the simulation of the atmospheric transport of CO_2, initially with 2-dimensional models (e.g. Pearman et al., 1983) and more recently with 3-dimensional models (e.g. Fung et al., 1983, 1987; Heimann and Keeling, 1989; Tans et al., 1990; Enting et al., 1995; Denning et al., 1995). These models provide the link between sources and sinks of CO_2 to the atmosphere and observed atmospheric concentrations. The transport models have typically been used in 2 ways. In the first, proposed sources and sinks of CO_2 are input to the model and plausible combinations of these sources are determined by matching the resulting atmospheric concentrations to observations (the references listed above tend to be of this type). Alternatively the models are constrained with the observed concentrations to estimate the sources and sinks (Tans et al., 1989; Enting and Mansbridge, 1989, 1991; Conway et al., 1994a; Ciais et al., 1995; Law and Simmonds, 1996). In either case, it is necessary to equate model values to observations despite some fundamental differences in the characteristics of these two measures.

For CO_2, most observations are taken at surface monitoring sites, either measuring continually or by flask samples taken at regular intervals. In both cases the data are selected for 'background' or 'baseline' conditions. The aim of this selection is to provide measurements that are representative of large-scale air masses. This requirement is most easily satisfied for marine or coastal sites where local surface sources are small and atmospheric stability near the surface is less variable than over land. Enting and Pearman (1993) discuss the reasons for observational data selection and also the typical methods used. Selection is designed to eliminate periods in the concentration record that are contaminated by local sources. For coastal sites, this is usually when the air has had recent contact with land surfaces. The selection method used is site specific, requiring a good knowledge of meteorological conditions and the variability of CO_2 at the site. Typical criteria in use include wind direction, wind speed, the temporal consistency of the observations and the concurrent concentrations of other species. Where monitoring uses flask samples, most selection occurs through the timing of the sample. By contrast, continuous monitoring allows different selection criteria to be tested (Beardsmore and Pearman, 1987).

Compared to observations, modelled concentrations automatically represent relatively large areas; modelled values are grid box averages where a typical grid size might be $4 \times 5°$ horizontally with a surface layer depth of a few hundred metres. However, the relatively low model resolution also means that geographical structure may be misrepresented. Examples include simplified land/ocean distributions in the vicinity of a monitoring site and problems with the altitudes of mountain sites. This could introduce discrepancies between the modelled concentrations and the observations.

This fundamental difference between observed and modelled concentrations raises the question of how they should be compared. For example, should model data be selected in a similar fashion to the observed data or does the low resolution of typical tracer models adequately mask contamination by local sources? This issue has been recognised by a number of authors (Fung et al., 1987; Heimann et al., 1989; Taylor, 1989) but in these studies the modelled concentrations were used without modification. The major problem is recognised to be with coastal stations. Fung et al. (1991) and Denning et al. (1995) tackled this by comparing the observations with the ocean grid point nearest to the station in the direction from which the observations are selected. This reduces the impact of the local land source. Enting et al. (1993) considered this option but found the impact to be small compared to other uncertainties in the inversion.

Studies of model data selection have been done for a small number of sites. Enting and Trudinger (1993) looked primarily at Cape Grim and found that their modelled short timescale variability was small compared to that observed and was not reduced by selection based on wind direction. One reason for this result was that the model used zonally averaged sources. Thus the extra variability resulting from regional sources was not present and hence could not be removed by selection. By contrast, Ramonet and Monfray (1996) found a significant impact from selection at Cape Grim and Cape Meares with less impact at Amsterdam Island. They attribute the low variability seen by Enting and Trudinger to the low resolution of the model used. Ramonet and Monfray use a high (2.5°) resolution window in their model. This allows better resolution of both the source and concentration fields and would account for the larger variability they found and the larger impact due to selection. They tested selection methods analogous to those used for the observations as well as trajectory and radon-based methods.

Most of the emphasis on selection in the previous studies has been to reduce the *temporal* variability of the CO_2 timeseries. In the case of observed timeseries, low temporal variability is the best indication available of concentrations that are representative of large-scale air masses. With modelled concentrations however, it is possible to check directly whether the timeseries are representative of that region. Thus in this paper, we focus on the *spatial* variability of CO_2 concentrations. We examine how one feature of the CO_2 data (the seasonal cycle) varies over distances of 500–1000 km. We then consider how much of the spatial variability found may be attributed to local land influences and whether this can be removed. The impact of selection on spatial variability is then presented and lastly comparisons are made with observed seasonal cycles.

2. Model simulations

For this study we focus on one component of the sources and sinks of CO_2 to the atmosphere: the seasonal land biosphere exchange. This was chosen because in the northern mid and high latitudes the seasonal cycle of atmospheric CO_2 concentrations is almost completely determined by the biospheric exchange. Thus it is possible to compare this feature of the model simulations with observations without having to model the other exchanges of CO_2 with the atmosphere. In the tropics and southern hemisphere, where other sources make a significant contribution to the seasonal cycle, the comparison with observations is not valid but the impact of selection can still be assessed.

The simulations have been performed using the Melbourne University tracer model (Law et al., 1992). This is an 'offline' tracer model derived from the Melbourne University General Circulation Model (MUGCM) (Simmonds, 1985; Simmonds et al., 1988). The tracer concentration is described by a spectral representation with rhomboidal, 21 wave resolution at 9 vertical levels. When gridded fields are required, a 64×54 grid is used. The 64 longitude points have a spacing of 5.625° while the latitudes are gaussian values with an approximate 3.3° spacing. Advection is driven by daily wind data from a MUGCM control run which is interpolated to the tracer model timestep. In general one year of wind data is used repeatedly. There is no diurnal cycle. The trace gas is also subject to diffusion, both horizontal and vertical and to sub-grid scale vertical transport by convection, which is parameterised using convection statistics from the MUGCM. Daily statistics are used rather than the monthly averages used in Law et al. (1992).

The biospheric sources and sinks used are from Fung et al. (1987); these sources have been used for simulations with a number of different tracer models as part of a tracer model intercomparison (Rayner and Law, 1995). The sources were input directly into the surface layer of the model and were kept fixed over the month to which they apply. Thus all variability in modelled concentrations on timescales less than a month is due to transport rather than variability in the sources. The model was integrated from an initial atmosphere containing no CO_2. Analysis is of the third model year which is sufficient for the troposphere to reach a 'steady-state'.

In addition to the CO_2 simulation, a radon-222 experiment has been performed for use in the identification of land influenced air. ^{222}Rn is a radioactive decay product of Uranium-238 and is released from land surfaces at a rate 100–1000 times larger than from oceans. It has a relatively short half-life of 3.8 days and is therefore a good tracer for indicating air that has had recent contact with land surfaces (Whittlestone, 1985). We used a constant source of 1 atom.cm^{-2}.s^{-1} (as Heimann et al., 1990) for all land grid points except Antarctica. The model was run from 1 December through to the end of the following year; one month of initialisation is sufficient given the short lifetime of ^{222}Rn.

3. Spatial variability over 500–1000 km

Given that the major purpose of selecting CO_2 concentration measurements is to provide data that are representative of large regions, it is of interest to examine how much spatial variability is present in the modelled concentrations in the regions surrounding the locations at which CO_2 is measured. To do this we construct timeseries of daily (instantaneous not mean) concentrations at the nearest grid point to a number of monitoring locations and at the eight surrounding grid points to these locations. The list of locations used along with the latitude and longitude of the nearest grid point are given in Table 1.

Since we are interested principally in the simulated seasonal cycle, the daily timeseries have been fitted with smoothing splines (Enting, 1987) with a 50% attenuation period of 90 days. A variety of attenuation periods was investigated. 90 days was chosen to provide a similar scale of seasonality as that produced by a definition of the seasonal cycle based on two harmonics as used by Enting et al. (1993), for example. Smoothing splines were used in this study because of their ability to fit irregular data such as occurs after selection. Examples of the seasonal cycles obtained are shown in Fig. 1 for 4 sites.

At Barrow (Fig. 1a), the amplitude of the seasonal cycle is larger for points to the south and south east of the Barrow grid point. The maximum and minimum also occur earlier in the year at these locations. During July and August the 9 grid points can span up to 8 ppmv in concentration, the lower values being those with more continental influence as photosynthesis (a CO_2 sink) dominates at this time of year.

The variability at Kumakahi (Fig. 1b) is smaller as expected for this remote ocean location (the model resolution is too low for any representation of the Hawaiian islands). The largest differences occur at the maximum and minimum concentrations where the 9 points span about 1 ppmv. The amplitude of the seasonal cycle is smaller to the south and larger to the north. This reflects the general decrease in amplitude away from the

Table 1. *Latitude and longitude of a selection of CO_2 monitoring sites and the nearest model grid point latitude and longitude*

Station	Code	Location (lat, long)	Nearest model (lat, long)	Radon	Days
Alert	ALT	82.5, 297.7	80.9, 298.1	141	255–363
Mould Bay	MBC	76.2, 240.7	77.6, 241.9	92	288–363
Barrow	BRW	71.3, 203.4	71.0, 202.5	73	70–363
Station M	STM	66.0, 2.0	64.4, 0.0	43	271–342
Cold Bay	CBA	55.2, 197.3	54.5, 196.9	39	119–350
Shemya	SHM	52.8, 174.1	51.2, 174.4	53	289–359
Cape Meares	CMO	45.0, 236.0	44.6, 236.3	56	99–340
Azores	AZR	38.8, 332.9	38.0, 331.9	32	337–363
Bermuda E	BME	32.4, 295.4	31.4, 298.1	9	102–239
Bermuda W	BMW	32.3, 294.1	31.4, 292.5	13	90–254
Sand I	MID	28.2, 182.6	28.1, 180.0	6	124–310
Key Biscayne	KEY	25.7, 279.8	24.8, 281.3	22	140–350
Kumakahi	KUM	19.5, 205.2	18.2, 202.5	6	269–318
St Croix	AVI	17.8, 295.3	18.2, 292.5	12	284–352
Guam	GMI	13.4, 144.8	14.9, 146.3	–	–
Ragged Point	RPB	13.2, 300.6	11.6, 298.1	31	121–365
Christmas I	CHR	2.0, 202.7	1.7, 202.5	3	135–329
Seychelles	SEY	−4.7, 55.2	−5.0, 56.3	3	75–259
Ascension	ASC	−7.9, 345.6	−8.3, 343.1	6	114–349
Samoa	SMO	−14.3, 189.4	−14.9, 191.3	–	–
Amsterdam	AMS	−38.0, 77.5	−38.0, 78.8	6	292–355
Cape Grim	CGO	−40.7, 144.7	−41.3, 146.3	19	60–338
Palmer	PSA	−64.9, 296.0	−64.4, 298.1	6	283–345
Syowa	SYO	−69.0, 39.6	−67.7, 39.4	7	358–364
Halley Bay	HBA	−75.7, 279.8	−74.3, 331.9	6	319–356
South Pole	SPO	−90.0, 335.2	−87.5, 337.5	5	246–298

The station codes are used in Figs. 2, 5, 7. The final two columns give the radon threshold concentration (in pCi/SCM) used for the data selection described in Section 5 and the number of days remaining after selection. A range is given because a different number of days remains for each of the nine locations in the region around the monitoring site.

region around 60°N where the largest seasonal cycle of sources occurs.

Around the equator and in the southern hemisphere, the seasonal CO_2 cycles become much smaller and local variability becomes more noticeable. Around Christmas Island (2.0°N) the concentrations are clearly dependent on latitude (Fig. 1c). As at Kumakahi the amplitude decreases going south. There is also a shift in phase with the maximum at the southern points occurring almost 3 months later than at the northern points. The minimum concentrations occur closer together in time. Since Christmas Island is situated in the Pacific Ocean the differences seen here are clearly not due to local land influences. Rather each point experiences air from different hemispheres at different times of year due to the latitudinal shift of the inter-tropical convergence zone (ITCZ). This sensitivity to the location of the ITCZ means that Christmas Island could provide a good test of a model's treatment of tropical transport if all components of the CO_2 budget were modelled.

Fig. 1d shows the 9 points surrounding Cape Grim. In this case, a wide range of seasonal cycles is found with maximum values occurring between late June and September. The minimum values

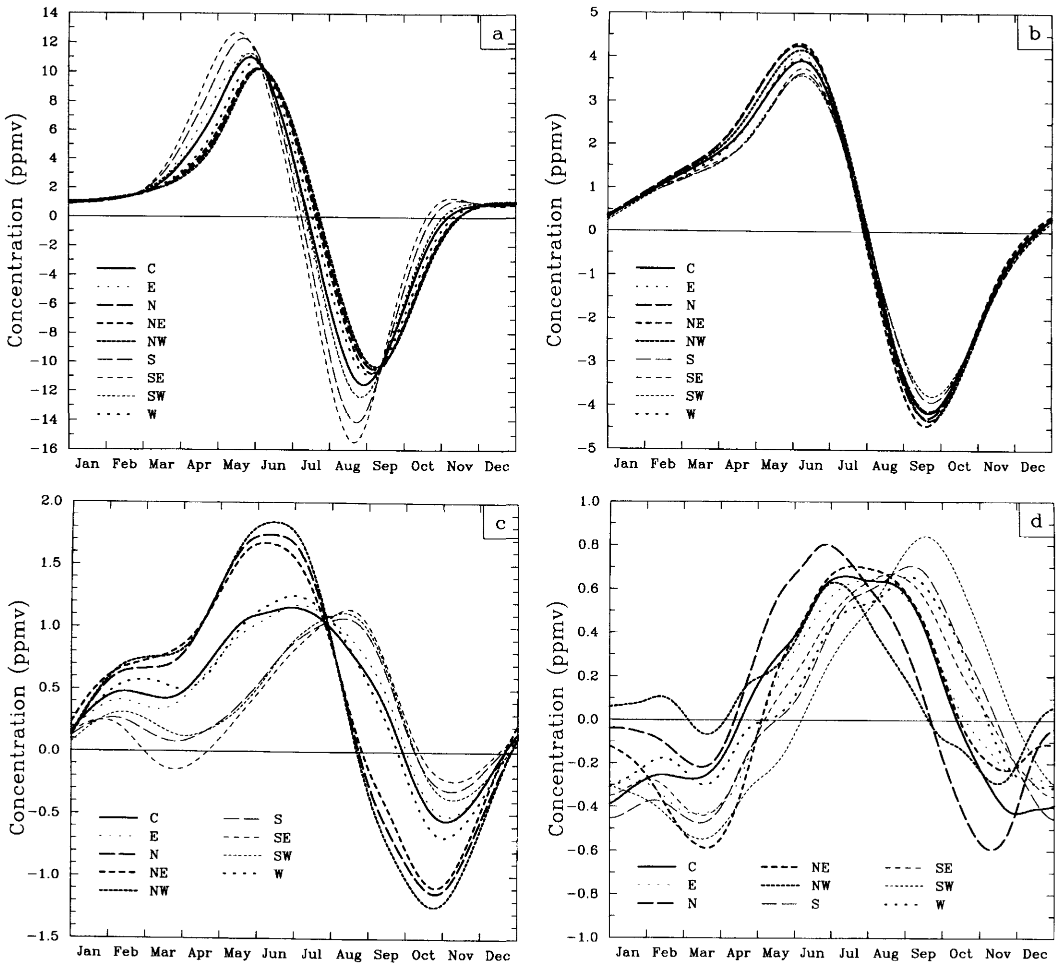

Fig. 1. Modelled seasonal cycles of CO_2 concentration (in ppmv) for the nearest grid point to the site (C) and for the eight surrounding grid points (indicated by compass point in the key). The sites shown are (a) Barrow, (b) Kumakahi, (c) Christmas Island and (d) Cape Grim.

tend to occur in either March or November. Peak to peak amplitudes range from about 0.8 to 1.4 ppmv. During October the range in concentrations is around 1 ppmv. While this is small compared to the differences seen at Barrow, it is large relative to the amplitude of the seasonal cycle at Cape Grim and hence important if we are to draw any conclusions about the seasonality of CO_2 concentrations at this location.

To provide a measure of the spatial variability, we have calculated the time mean of the spatial standard deviation of the value of the spline (x) at each of the nine points (i) evaluated for each day of the year (n), i.e

$$\text{Mean sd} = \frac{1}{365}\sum_{n=1}^{365}\left(\frac{1}{8}\sum_{i=1}^{9}(x_i(n)-\bar{x}(n))^2\right)^{1/2} \quad (1)$$

The results are plotted as a function of the station grid point latitude in Fig. 2. It is clear that the mean standard deviation increases from south to north: southern hemisphere values are less than 0.2 ppmv and often less than 0.1 ppmv whereas the northern hemisphere values range from about 0.2–1.0 ppmv. The largest values occur between 25–70°N. In general, spatial variability is larger for those regions where the amplitude of the

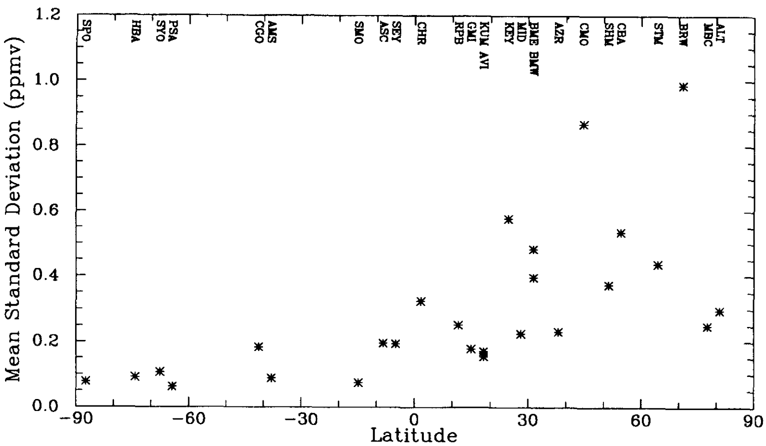

Fig. 2. Annual mean of standard deviation of the splines representing the seasonal CO_2 cycle at the nine grid points surrounding each monitoring site listed in Table 1.

seasonal cycle is larger and for grid points near to land where local sources can have a large impact. A few exceptions are Alert and Mould Bay where the variabilities are relatively small presumably because the local sources are very small and the equatorial stations such as Christmas Island where the variability is larger than might be expected. As mentioned earlier this is due to the seasonal movement of the ITCZ.

The large spatial variability at some sites (despite the relatively low model resolution) makes comparisons with observations difficult. It is not clear which model result, for the region around the given site, can most appropriately be compared to the observations. While the spline fit smooths the extreme values in the daily timeseries, these are nevertheless averaged into the resulting seasonal cycle rather than being removed as data selection would tend to do. This can bias the model concentrations compared to the observations. Thus we examine the potential for reducing the spatial variability by selecting the model data.

4. Selection method

There are a number of ways of selecting data to eliminate air contaminated by land contact. For observed data, the selection criteria are often based on wind direction or a combination of wind direction and other measures such as wind speed, consistency of the measurements in time, the concentrations of other trace species etc. When selecting data from a model timeseries, it is not always appropriate to use the same criteria as were used for the observations. For example, a wind direction criterion is determined by the land/ocean distribution in the vicinity of the monitoring site which may not be well represented by the model.

For this study of the impact of selection on spatial variability, three selection methods were considered: wind direction, back-trajectories and modelled ^{222}Rn concentration. Modelled ^{222}Rn concentration was chosen for a number of reasons. A wind direction criteria is not well suited to this situation because it is primarily a local method. The 'clean-air' sector is specific to a given site and may not be applicable to each of the 9 points we are using to examine the spatial variability around a site. Ramonet and Monfray (1996) found that back-trajectories and ^{222}Rn concentration were equally effective selection methods (and better than local selection based on wind direction and speed). In this case, the advantages of the ^{222}Rn method over a trajectory method are (i) the ^{222}Rn is transported by the same resolved and sub-grid scale processes as used for the CO_2 simulation whereas the available trajectories were calculated on isosigma surfaces with no sub-grid scale component, (ii) the ^{222}Rn method allows different datasets to be selected easily, by varying the ^{222}Rn threshold value used (a similar capability with the trajectory method would require estimating the length of time a given trajectory was over land) and (iii) the ^{222}Rn method is easier to implement over a region of grid-points (one ^{222}Rn simulation provides the required data at all points).

While the potential use of ^{222}Rn in the selection of observed data (Whittlestone, 1985) has prompted its use here, there is no necessity to use a real tracer. The tracer model gives us the ability to simulate an artificial tracer which could be tuned to enhance certain selection conditions. We have tested the sensitivity to changing the tracer decay rate and the results are discussed at the end of the next section.

5. Reduction of spatial variability by selection

One decision required for ^{222}Rn-based selection is the threshold ^{222}Rn value which will determine whether data is accepted or rejected. We use Barrow to illustrate some of the issues involved in choosing the threshold value and the type of result produced. The spline fits to the unselected data from the set of nine points around Barrow were shown in Fig. 1a. Fig. 3 shows spline fits to various sets of selected Barrow data. Selection was based on ^{222}Rn concentration and the results for three ^{222}Rn threshold values are shown. In each case, the daily CO_2 values were accepted or rejected from the timeseries depending on whether the ^{222}Rn concentration at the same time was less than or greater than the chosen threshold value. A spline was then fitted to the selected data in the same manner as for the unselected data.

Fig. 3a shows the nine spline fits for the points around Barrow with an acceptable ^{222}Rn level of 90 pCi/SCM (standard cubic metre). It can be clearly seen that the nine curves lie closer together than for the unselected data particularly at the times of maximum and minimum concentrations. In August the unselected CO_2 values to the south and south east of Barrow were noticeably lower than the other CO_2 values due to the continental sink south of Barrow at this time. These lower values are no longer present when the data for high ^{222}Rn values is removed indicating that these low CO_2 values were indeed associated with continental air. When the acceptable ^{222}Rn value is lowered to 73 (Fig. 3b) and 50 pCi/SCM (Fig. 3c) most of the curves continue to move together. However in the 73 pCi/SCM case, the SE point is beginning to drift away from the other curves and by 50 pCi/SCM the S and E points are showing similar behaviour. This occurs because the selection has reduced the timeseries to such an extent that a realistic seasonal cycle can no longer be obtained. For example, when the selection criteria is 50 pCi/SCM, the timeseries to the SE and S are reduced to only 8 and 19 points respectively. For the same selection criteria, the point to the NW still has 342 acceptable days in its timeseries.

This raises a number of options: (i) different selection criteria could be used for each location, (ii) locations which dropped below a certain number of selected days of data could be ignored altogether or (iii) we could adopt a compromise between maintaining sufficiently large datasets and removing land influence. We have chosen the third option based on the assumption that for any station close to land, a regionally representative value would be expected to maintain some land influence; it is only the local effects that should be eliminated. This choice also minimises the dependence of the selection on the half-life of the tracer as we will discuss later.

Since our aim is to reduce the spatial variability, we seek the threshold ^{222}Rn value (to the nearest 1 pCi/SCM) which gives the smallest mean standard deviation between the nine curves (as eq. (1)). For Barrow, this is the 73 pCi/SCM case with a mean sd of 0.69 ppmv (compared to 0.99 ppmv for the unselected data, 0.79 ppmv for 90 pCi/SCM and 1.01 ppmv for 50 pCi/SCM). Note that 73 pCi/SCM is a relatively high ^{222}Rn mixing ratio so that, for example, air coming from the Siberian continent would be accepted because its ^{222}Rn concentration would typically be below this threshold by the time it reached Barrow.

Another station that provides a good example of the impact of selection is Cape Grim (Fig. 4). The considerable variability in the seasonal cycles in the unselected data (Fig. 1d) has been reduced markedly by selecting the data using a ^{222}Rn value of 19 pCi/SCM (sd = 0.10 ppmv compared to 0.18 ppmv for the unselected data). The seasonal cycles for the points to the north and north west of Cape Grim show the largest changes. The final result is a seasonality which resembles that from the unselected data to the south west of Cape Grim. This agrees with the direction from which clean air is sampled at Cape Grim and would tend to justify those who use an offshore point to

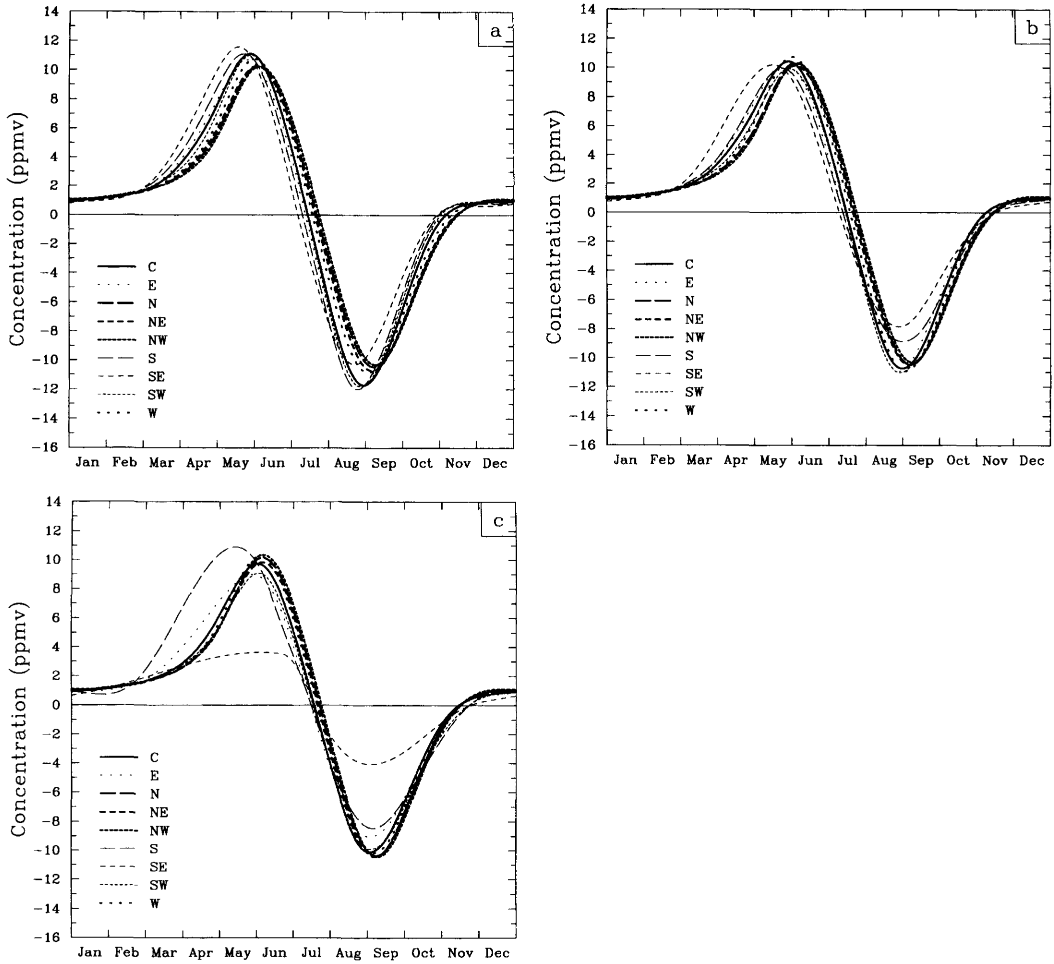

Fig. 3. Seasonal cycles fitted to selected CO_2 concentration for the nearest grid point to Barrow (C) and for the eight surrounding grid points (indicated by compass point in the key). Data were selected for radon concentrations less than (a) 90 pCi/SCM, (b) 73 pCi/SCM and (c) 50 pCi/SCM.

represent a coastal station. Ramonet and Monfray (1996) found a similar shift in phase when they selected model data for Cape Grim using either back-trajectories or ^{222}Rn concentration. They also found an approximate 30% reduction in amplitude whereas we find a slight increase in the amplitude. This difference may be because they simulate the full set of CO_2 sources and sinks rather than just the biosphere component.

The process of selecting and fitting the data and finding the minimum mean sd has been repeated for all the sites listed in Table 1. It is important to check that selection has little effect on those sites far from land as well as testing the impact at those stations more likely to show significant land influence. Also given in Table 1 is the ^{222}Rn concentration which was found to give the minimum standard deviation and the number of days remaining after selection for the timeseries with the least and most acceptable data (out of the 9 timeseries for the points around the site). As would be expected the threshold ^{222}Rn values generally increase from south to north and are larger for coastal sites. Coastal sites also tend to have larger ranges of acceptable days indicative of inland points where much of the timeseries is rejected

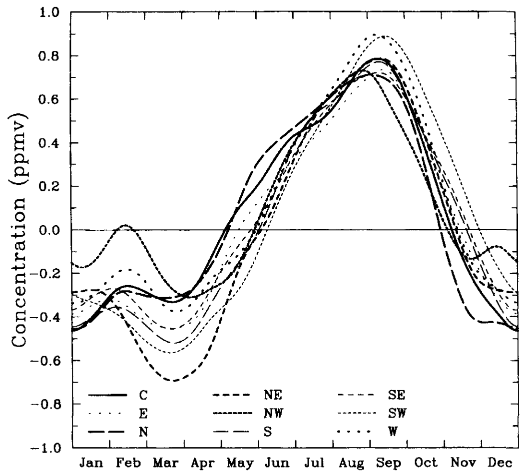

Fig. 4. Seasonal cycles fitted to selected CO_2 concentration for the nearest grid point to Cape Grim (C) and for the eight surrounding grid points (indicated by compass point in the key). Data were selected for radon concentrations less than 19 pCi/SCM.

and offshore points where most data is acceptable. The table indicates that even for ocean sites the selection can substantially reduce the number of days in the timeseries. This is typically due to the use of a very low ^{222}Rn threshold based on minimising the mean sd. However the minimum sd found is usually very close to that obtained for the unselected data indicating that selection is not necessary in these cases.

The minimum mean sd after selection is shown in Fig. 5. All stations except Guam and Samoa showed some reduction in the mean standard deviation although in many cases the changes were very small. Eleven stations produced standard deviations after selection that were less than 80% of their unselected values, the largest reductions being at Bermuda and Cold Bay (52–53% of original). It is clear (by comparison with Fig. 2) that the largest reductions occur in the northern mid-latitudes. Only two stations remain with sd greater than 0.5 ppmv (Barrow and Cape Meares). It is worth noting that after selection the mean sd at Cape Grim is much closer to the values obtained for the other southern hemisphere sites.

The sensitivity of these results to the tracer used for selection has been tested. A model simulation was performed for a tracer with the same source as ^{222}Rn but with a half-life of 1.9 rather than 3.8 days. Selection of the CO_2 data was performed with this tracer, in the same manner as for ^{222}Rn. Again the threshold value was determined by minimising the spatial variability as measured by the mean sd. In general the threshold values for the 1.9 day half-life tracer are 30–70% of those determined for ^{222}Rn. This reflects the generally lower concentrations of this tracer in the atmosphere because of its shorter lifetime. The minimum mean standard deviations obtained are close to those obtained from the ^{222}Rn selection (typically within 10% and with a mean increase over the 26 sites of 4%). Thus the change in the half-life of the tracer has had little impact on the selection. This is a consequence of performing the selection across a region using the same threshold value for all nine locations. It appears that this regional implementation is a stronger determinant of the spatial scales associated with the rejection of data than the half-life of the radioactive tracer.

6. Impact of selection on the regional mean seasonal cycle

The results from selecting the data have shown that spatial variability in the vicinity of specific locations can be reduced for those regions close to land. The question remains as to whether the mean seasonal cycle for that region is significantly altered by this process and how this might impact on any comparisons with observations. We compare three approximations to the seasonal cycle at a given station: the fit to the unselected data for the nearest grid point to the site, the mean of the nine fits to the unselected data surrounding the site and the mean of the nine fits to the selected data around the site. The difference between the first two gives an indication of whether there is value in considering the region around a site rather than just a single location while the difference between the second and third cases indicates the possible value of data selection. The results for three sites (Barrow, Cape Meares and Cape Grim) are shown in Fig. 6.

In each case, seasonal cycles are plotted for the nearest grid point and regional means for selected and unselected data. The fourth curve is the observed seasonal cycle represented by two Fourier components of detrended data. The fit is to the data of Conway et al. (1994b). At Barrow

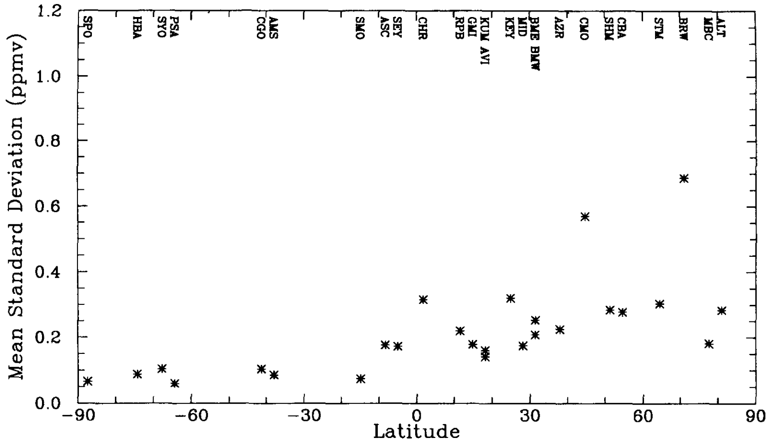

Fig. 5. Annual mean of standard deviation of the splines representing the seasonal CO_2 cycle from selected data at the nine grid points surrounding each monitoring site listed in Table 1.

(Fig. 6a) the seasonal cycles at the nearest grid point and for the mean unselected data are almost identical indicating that the region around the site is well represented by the central grid point. Selection produces some change in the seasonal cycle, primarily in the value of the minimum and maximum values. The minimum value from the selected data is reasonably close to the observations but the maximum value is substantially larger than observed.

At Cape Meares (Fig. 6b) the central point has a lower minimum than the regional mean but for the remainder of the year the values are similar. Selection increases the minimum value further, bringing it closer to the observations. The changes due to selection are smaller during the rest of the year but tend to move the model result away from the observed values. For both these northern mid and high latitude sites (and those not shown) the time of maximum concentration is clearly delayed in the model compared to the observations and often of excessive magnitude. This appears to be a problem with the sources as a similar result was found for all models in a recent intercomparison (Rayner and Law, 1995).

The model results for Cape Grim (Fig. 6c) for the nearest grid point show maximum values through July and August and minimum values between December and March. The regional mean curve is similar but with slightly reduced extremes (except during March). The times of zero concentration are also slightly delayed. Selection produces a further shift in phase with the maximum concentration now occurring in September. This is similar to the timing of the maximum in the observations although the model magnitude is larger than observed. The times of zero concentration indicate that even after selection the model seasonal cycle still lags the observed values by around a month but this is a significant improvement over the unselected data. We do not expect to see complete agreement with the observations because other CO_2 sources may contribute significantly to the seasonality at sites such as Cape Grim, which have a small amplitude seasonal cycle.

These examples have shown that the difference between seasonal cycles calculated from data with or without selection is small compared with differences between the modelled seasonal cycle and that observed. To assess the impact of selection for all sites we calculate the mean difference between the various estimates of the seasonal cycle and the observed seasonal cycle using

$$\text{Mean diff} = \frac{1}{365} \sum_{n=1}^{365} |x_s(n) - x_0(n)| \qquad (2)$$

where n is the day number and x_0 is the observed seasonal cycle calculated from the Fourier components. The mean difference is evaluated for 3 cases of x_s: the spline fit to the nearest gridpoint, the mean of the splines fitted to the unselected data and the mean of the splines fitted to the selected data.

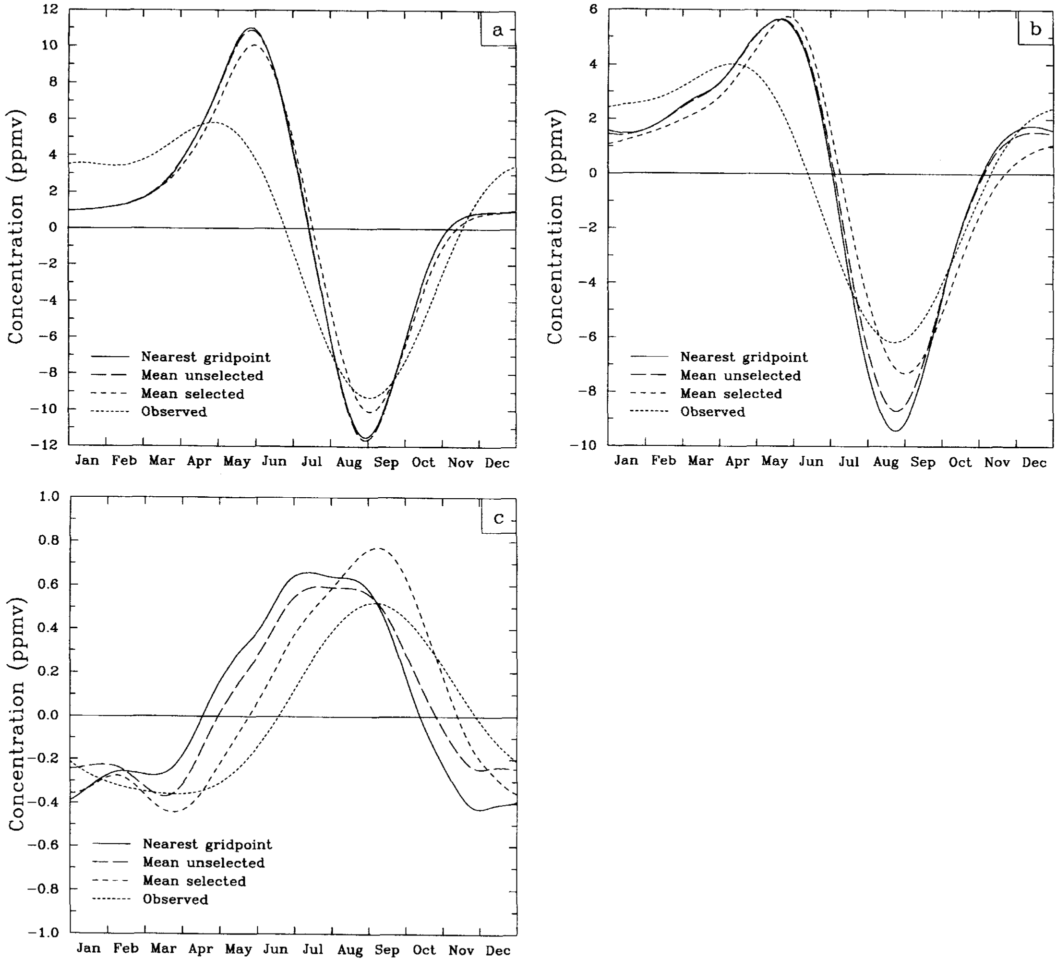

Fig. 6. Seasonal cycles of CO_2 concentration for the nearest grid-point to the site (solid), the mean of the splines fitted to the unselected data for the 9 nearest grid-points to the site (long dash), the mean of the splines fitted to the selected data for the 9 nearest grid-points (medium dash) and the observed seasonal cycle (short dash) for (a) Barrow, (b) Cape Meares and (c) Cape Grim.

The mean differences (from the observed seasonal cycle) are plotted in Fig. 7. The spread of the three points indicates the impact of selection relative to the typical difference from the observations. In almost all cases the impact is small. For most sites the regional mean from the unselected data appears to be a slight improvement on the nearest gridpoint data for modelling the observed seasonal cycle. By contrast, selection produces larger differences from the observations (e.g., Cold Bay, Bermuda, Cape Meares) more often than it improves the comparison with the observations (e.g. Barrow, Cape Grim). Clearly we need to improve our simulation of the CO_2 seasonal cycle before the issues associated with selection will become important.

7. Conclusions

The spatial variability of simulated CO_2 seasonal cycles driven by the terrestrial biotic flux has been examined for regions around a number of CO_2 monitoring sites. Significant variability is

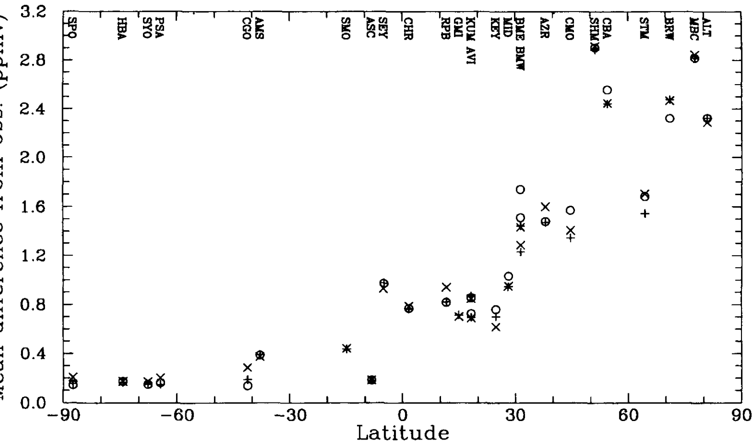

Fig. 7. Mean difference (as eq. 2) from the observed seasonal cycle of CO_2 concentration for the nearest grid-point (x), the mean unselected spline (+) and the mean selected spline (o) for each of the sites listed in Table 1.

found, particularly for northern mid and high latitude coastal sites. This variability is reduced when the simulated timeseries are selected to remove air that has had recent contact with land. The concurrent modelled concentration of ^{222}Rn provides a reliable selection method. Identifying a regionally representative timeseries by minimising a measure of the local spatial variability gives one method of choosing an appropriate radon-222 concentration for use in the selection.

The comparison with the observed seasonal cycle was more problematic. The impact of selection was largely masked by the large differences between the simulated and observed seasonal cycles. It appears that until we improve our simulation of the seasonal cycle, the relatively small effects of selection may be neglected without introducing significant errors. It is important to remember, however, that this conclusion is based on the use of a moderate resolution model with no diurnal cycle and no sub-monthly temporal variability in the sources. Any move to higher spatial or temporal resolution would presumably increase the modelled CO_2 variability and selection may become more important (as suggested by Ramonet and Monfray (1996)). This study has indicated at which locations the impact of selection will be largest.

8. Acknowledgements

I would like to thank Dr Ian Enting and Dr Peter Rayner for helpful comments on this manuscript. This study was supported through the Australian Government Cooperative Research Centres Program.

REFERENCES

Beardsmore, D. J. and Pearman, G. I. 1987. Atmospheric carbon dioxide measurements in the Australian region: Data From Surface Observations. *Tellus* **39B**, 42–66.

Ciais, P., Tans, P. P., White, J. W. C., Trolier, M., Francey, R. J., Berry, J. A., Randall, D. R., Sellers, P. J., Collatz, J. G., and Schimel, D. S. 1995. Partitioning of ocean and land uptake of CO_2 as inferred by $\delta^{13}C$ measurements from the NOAA Climate Monitoring and Diagnostics Laboratory Global Air Sampling Network. *J. Geophys. Res.* **100**, 5051–5070.

Conway, T. J., Tans, P. P., Waterman, L. S., Thoning, K. W., Kitzis, D. R., Masarie, K. A., and Zhang, N. 1994a. Evidence for interannual variability of the carbon cycle from the National Oceanic and Atmospheric Administration/Climate Monitoring and Diagnostics Laboratory Global Air Sampling Network. *J. Geophys. Res.* **99**, 22831–22855.

Conway, T. J., Tans, P. P., and Waterman, L. S. 1994b. Atmospheric CO_2 records from sites in the NOAA/CMDL air sampling network. *Trends '93: A Compendium of data on global change*, ed. T.A. Boden, D.P. Kaiser, R.J. Sepanski, F.W. Stoss, Carbon Dioxide Information Analysis Center, Oak Ridge, U.S.A, 41–119.

Denning, A. S., Fung, I. Y., and Randall, D. A. 1995. Gradient of Atmospheric CO_2 due to Seasonal Exchange with Land Biota. *Nature* **376**, 220–223.

Enting, I. G. 1987. On the use of smoothing splines to filter CO_2 data. *J. Geophys. Res.* **92D**, 10977–10984.

Enting, I. G. and Mansbridge, J. V. 1989. Seasonal sources and sinks of atmospheric CO_2: direct inversion of filtered data. *Tellus* **41B**, 111–126.

Enting, I. G. and Mansbridge, J. V. 1991. Latitudinal distribution of sources and sinks of CO_2: results of an inversion study. *Tellus* **43B**, 156–170.

Enting, I. G. and Pearman, G. P. 1993. Average global distributions of CO_2. The global carbon cycle: *Proceedings of the NATO ASI*, at Il Ciocco, Italy, September 1991, ed. M. Heimann, Springer-Verlag, 31–64.

Enting, I. G., Trudinger, C. M., Francey, R. J., and Granek, H. 1993. Synthesis inversion of atmospheric CO_2 using the GISS tracer transport model. *CSIRO Div. Atmos. Res. Tech. Paper no. 29*, 44 pp.

Enting, I. G. and Trudinger, C. M. 1993. Modelling studies of small-scale variability in atmospheric CO_2. *Baseline atmospheric program (Australia) 1990*, ed. S.R. Wilson and J. Gras, Dept. of Admin. Services and CSIRO: Australia, 10–16.

Enting, I. G., Trudinger, C. M., and Francey, R. J. 1995. A synthesis inversion of the concentration and $\delta^{13}C$ of atmospheric CO_2. *Tellus* **47B**, 35–52.

Fung, I., Prentice, K., Matthews, E., Lerner, J., and Russell, G. 1983. Three-dimensional tracer model study of atmospheric CO_2: response to seasonal exchanges with the terrestrial biosphere. *J. Geophys. Res.* **88**, 1281–1294.

Fung, I. Y., Tucker, C. J., and Prentice, K. C. 1987. Application of advanced very high resolution radiometer vegetation index to study atmosphere-biosphere exchange of CO_2. *J. Geophys. Res.* **92**, 2999–3015.

Fung, I., John, J., Lerner, J., Matthews, E., Prather, M., Steele, L. P. and Fraser, P. J. 1991. Three-dimensional model synthesis of the global methane cycle. *J. Geophys. Res.* **96**, 13033–13065.

Heimann, M. and Keeling, C. D. 1989. A three-dimensional model of atmospheric CO_2 transport based on observed winds (2). Model description and simulated tracer experiments. *Aspects of climate variability in the Pacific and the Western Americas, Geophysical Monograph* **55**, ed. D.H. Peterson, AGU, Washington (USA), 237–275.

Heimann, M., Keeling, C. D., and Tucker, C. J. 1989. A three-dimensional model of atmospheric CO_2 transport based on observed winds (3). Seasonal cycle and synoptic time scale variations. *Aspects of climate variability in the Pacific and the Western Americas, Geophysical Monograph* **55**, ed. D.H. Peterson, AGU, Washington (USA), 277–303.

Heimann, M., Monfray, P., and Polian, G. 1990. Modeling the long-range transport of ^{222}Rn to subantarctic and antarctic areas. *Tellus* **42B**, 83–99.

Law, R., Simmonds, I., and Budd, W. F. 1992. Application of an atmospheric tracer model to the high southern latitudes. *Tellus* **44B**, 358–370.

Law, R. and Simmonds, I. 1996. The sensitivity of deduced CO_2 sources and sinks to variations in transport and imposed surface concentrations. *Tellus* **48B**, in press.

Pearman, G. I., Hyson, P., and Fraser, P. J. 1983. The global distribution of atmospheric carbon dioxide (1). Aspects of observations and modeling. *J. Geophys. Res.* **88**, 3581–3590.

Ramonet, M. and Monfray, P. 1996. CO_2 baseline concept in 3-D atmospheric transport models. *Tellus* **48B**, this issue.

Rayner, P. J. and Law, R. M. 1995. A comparison of modelled responses to prescribed CO_2 sources. *CSIRO Div. of Atmos. Res. Tech. Paper no* **36**/CRC Sthn. Hem. Met. Tech. Rep. No 1, 82 pp.

Simmonds, I. 1985. Analysis of the 'spinup' of a general circulation model. *J. Geophys. Res.* **90**, 5637–5660.

Simmonds, I., Trigg, G., and Law, R. 1988. *The climatology of the Melbourne University general circulation model*, pub. no. **31**, Department of Meteorology, University of Melbourne, 67pp. [NTIS PB 88 227491.]

Tans, P. P., Conway, T. J., and Nakazawa, T. 1989. Latitudinal distribution of the sources and sinks of atmospheric carbon dioxide derived from surface observations and an atmospheric transport model. *J. Geophys. Res.* **94**, 5151–5172.

Tans, P. P., Fung, I. Y. and Takahashi, T. 1990. Observational constraints on the global atmospheric CO_2 budget. *Science* **247**, 1431–1438.

Taylor, J. A. 1989. A stochastic Lagrangian atmospheric transport model to determine global CO_2 sources and sinks—a preliminary discussion. *Tellus* **41B**, 272–285.

Whittlestone, S. 1985. Radon Measurements as an aid to the interpretation of atmospheric monitoring. *J. Atmos. Chem.* **3**, 187–201.