Source regions for CO$_2$ at Cape Point assessed by modelling, $^{222}$Rn and meteorological data

S. Whittlestone, University of Wollongong, Wollongong, Australia
E. Kowalczyk CSIRO Atmospheric Research, Mordialloc, VIC, Australia
E.-G. Brunke and C. Labuschagne, South African Weather Service
Source regions for CO2 at Cape Point assessed by modelling, radon-222 and meteorological data

S. Whittlestone
University of Wollongong, Wollongong, Australia

E. Kowalczyk
CSIRO Atmospheric Research, Mordialloc, VIC, Australia

E.-G Brunke and C. Labuschagne
South African Weather Service, Global Atmosphere Watch Station Cape Point
Stellenbosch, South Africa


June 2009
Executive Summary

This study explores methods of characterising experimental and modelling data to see if trace gas measurements at the Cape Point GAW station could improve our understanding of sources from the continent. Selection criteria have been defined which make it possible to divide the samples into categories with predominant transport to Cape Point from one of four source regions: City (greater Cape Town); close rural (within about 150 km of Cape Point); distant rural (from 150 to about 500 km) and very distant rural covering the rest of southern Africa. For species with fairly uniform source such as radon or CO, this regional selection worked well, but for CO₂ the respiration and photosynthesis cycle from very close vegetation usually overwhelmed the input from other sources. It was only in winter that the draw-down by wheat growing in the close rural zone was clearly evident in the Cape Point CO₂ data.

An invaluable tool for this study was a simulation of trace gas concentrations by a model of the global atmosphere, the Conformal Cubic Atmospheric Model (CCAM) developed by McGregor. The transport component of the model was sufficiently accurate to make meaningful estimates of the origin of CO and Rn, but the sources of CO₂ available for the model were inadequate for the purpose of the present study, which resulted in the simulated CO₂ being uncorrelated with the measurements.

Two superficially attractive selection criteria proved to be ineffective. One, wind direction, was so perturbed by local topography that there was no correlation of the measured wind direction with the bearing of the source of more distant trace gases from the critical north to east sector. The other, back trajectories, were effective in determining if contact with the southern African continent had occurred, but the indicated time and location of land contact was highly inaccurate.

Selection criteria which proved to be successful were:

i). Radon. A threshold of 250 mBq m⁻³ proved useful to select samples strongly influenced by land without excluding samples from coastal towns which have large CO and CO₂ sources but a short land fetch;

ii). Age. The simulated age of samples was not used directly as a selection criterion, but was useful in evaluation of data sets selected by other criteria.;

iii). Persistent Percentile Filter. Based on long-term variability of the data, this filter was able to exclude samples which had essentially marine baseline levels of the other species even though their passage over land yielded 250 mBq Rn;

iv). Spike Filter. Derived from the change of CO₂ mixing ratio from one hour to the next, this filter proved highly effective in selecting samples from greater Cape Town;

v). Predominant Zone. The CCAM was able to determine the region that contributed the greatest amount of direct transport to Cape Point.

The CCAM has shown the potential of CO₂ measurements at Cape point to provide estimates of the larger continental sources and sinks within a few hundred km of the station. However, as configured for this simulation, it is possible to extract only very limited information about these sources. To make full use of the CCAM, it would need to be configured to trace the transport of the simulated species from smaller regions than used for the present work, and output more detailed information about CO₂ fluxes from these regions.

It might then be possible to use inversion modelling of carefully selected data to estimate regional fluxes of the trace gases.
<table>
<thead>
<tr>
<th>Contents</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2. Experimental data</td>
<td>1</td>
</tr>
<tr>
<td>3. The conformal cubic atmospheric model (CCAM)</td>
<td>1</td>
</tr>
<tr>
<td>3.1. Model grid and flux zones</td>
<td>2</td>
</tr>
<tr>
<td>3.2. Model output processing</td>
<td>3</td>
</tr>
<tr>
<td>3.3. Sample age</td>
<td>4</td>
</tr>
<tr>
<td>4. Comparison of CCAM with measurements</td>
<td>6</td>
</tr>
<tr>
<td>4.1. Comparison of simulated wind direction with experiment</td>
<td>6</td>
</tr>
<tr>
<td>4.2. Comparison of simulated radon with experiment</td>
<td>8</td>
</tr>
<tr>
<td>4.2.1. CCAM transport timing</td>
<td>8</td>
</tr>
<tr>
<td>4.2.2. Model accuracy for close zones</td>
<td>9</td>
</tr>
<tr>
<td>4.3. Comparison of simulated CO₂ with experiment</td>
<td>10</td>
</tr>
<tr>
<td>5. Air mass selection criteria</td>
<td>12</td>
</tr>
<tr>
<td>5.1. Wind direction</td>
<td>13</td>
</tr>
<tr>
<td>5.2. Radon</td>
<td>13</td>
</tr>
<tr>
<td>5.3. Age</td>
<td>13</td>
</tr>
<tr>
<td>5.4. Persistent percentile filter for CO₂</td>
<td>13</td>
</tr>
<tr>
<td>5.5. Spike filter</td>
<td>14</td>
</tr>
<tr>
<td>5.6. Predominant zone</td>
<td>14</td>
</tr>
<tr>
<td>5.7. Back trajectories</td>
<td>15</td>
</tr>
<tr>
<td>6. Results</td>
<td>17</td>
</tr>
<tr>
<td>6.1. Diurnal variation of sample number</td>
<td>17</td>
</tr>
<tr>
<td>6.2. Diurnal variation of trace gas median concentrations</td>
<td>18</td>
</tr>
<tr>
<td>6.3. Seasonal variation</td>
<td>19</td>
</tr>
<tr>
<td>6.4. Annual variation</td>
<td>20</td>
</tr>
<tr>
<td>7. Conclusions</td>
<td>24</td>
</tr>
<tr>
<td>8. Endorsements</td>
<td>25</td>
</tr>
<tr>
<td>9. Acknowledgements</td>
<td>26</td>
</tr>
<tr>
<td>10. References</td>
<td>26</td>
</tr>
</tbody>
</table>
1. Introduction

The Cape Point GAW station was established in 1977 (Brunke et al 1990) primarily to study trace gas concentrations from the southern hemispheric mid latitudes in baseline conditions. A study of selection criteria for marine baseline samples (Brunke et al 2004) showed that a substantial proportion of the data were influenced by the continental land mass of southern Africa. This study explores methods of characterising data to see if these non-baseline data could improve our understanding of CO₂ sources from the continent. Selection criteria are developed which separate data into sets with trace gases predominantly from large land areas, from close towns and from a mix of land and towns.

A major challenge was to separate out samples for which the predominant influence was local. Certainly some of the criteria used for baseline marine samples could be applied: CO₂, CO and CH₄ variability on a time scale of a few hours should be low for distant samples. The concentration of Rn gas should be high for a long continental path. However, it was not possible using only these criteria to discriminate between samples which had spend a long time in low wind speed over a small part of Africa close to Cape Point, and air masses which had passed over a thousand or so km of Africa and were representative of a much larger region.

It was decided to investigate whether a better understanding of the data could be gained by simulating the concentrations with a model of the global atmosphere. The model used was the Conformal Cubic Atmospheric Model (CCAM) developed by McGregor (McGregor et al 2001).

2. Experimental data

CO₂, CO, CH₄ and ²²²Rn concentrations as well as wind speed and direction have been taken from the approximately 6 year period, March 1999 to June 2005. Since the only Rn isotope used in this report is ²²²Rn, the “²²²” will be omitted subsequently. The instrumentation is described by Brunke et al (1990) and Brunke et al (2002). Of special interest is the effect of the most recent transit of the air parcel over southern Africa. Therefore the concentrations of the longer life-time species have been de-trended to eliminate seasonal and longer term trends, as described by Brunke and Scheel (2002). The long-term trend value has been subtracted from all CO₂ mixing ratios presented here, and an arbitrary 7 ppm added back in so that most of the values are positive, which improves readability.

3. The conformal cubic atmospheric model (CCAM)

The CSIRO conformal-cubic atmospheric model (CCAM) is a two time-level semi-implicit hydrostatic primitive equation model originally described by McGregor (1987) and more recently by McGregor and Dix (2001). The model is formulated upon a quasi-uniform grid derived by projecting the panels of a cube onto the surface of the Earth. Distinctive features of the CCAM dynamics include semi-Lagrangian horizontal advection with bi-cubic spatial interpolation, total-variation-diminishing vertical advection and an unstaggered grid (McGregor (1993), McGregor (1996)).

CCAM has an extensive set of physical parameterizations: interactive diagnosed cloud distribution, cumulus convection, Tiedke shallow convection, gravity wave drag, diurnally-varying skin temperature for soil surface temperatures and evaporation of rainfall. Longwave and shortwave radiative contributions are provided by GFDL parameterization (McGregor (1996)), CCAM has 18 levels in the vertical, with four levels within the first
1000 m and the lowest two levels at 38 m and 184 m. We have found this vertical resolution adequate to capture the main features of the surface/atmosphere gas exchange.

Coupled to CCAM is the CSIRO Atmosphere Biosphere Land Exchange (CABLE) land surface model. CABLE calculates carbon, water and heat exchanges between the land surface and atmosphere as described in Kowalczyk et al. (2006). The net biospheric CO₂ flux to the atmosphere is the sum of photosynthesis and respiration (flux units are gC m⁻² s⁻¹).

Surface fluxes for all the tracer species in this experiment are input to the lowest model level and converted to a change in concentration. After input to the atmosphere, the tracer species are transported as passive tracers with provision for a loss process such as radioactive decay, and are subject to advection by the resolved winds and sub-grid scale transport. Simulation of transport of trace gases by a CCAM predecessor DARLAM (CSIRO Div. of Atmospheric Research Limited Area Model) are described in Kowalczyk and McGregor (1998) and with the current model by Law et al. (2006).

3.1. Model grid and flux zones

The model was configured with fine grids over South Africa and progressively coarser grids as the distance from South Africa increased (Fig 1). In the vicinity of Cape Point, the grids represented almost square areas of about 62 x 62 km. The main species modelled was CO₂, but it was also possible to follow 3 additional tracer species, each with its own transport characteristics. Their source strengths (fluxes), but not that of CO₂, could be set at 8 land areas called tracer zones, which are the areas numbered 2 to 9 in Fig. 1.

Fig. 1. CCAM grid and tracer zone boundaries
The tracer species chosen were CO and Rn and a continental background tracer (CB). The latter was given a very long decay time (1 year) compared to Rn (3.84 days) in the atmosphere. This tracer could be used as a reference since it was conserved during transport. By comparing the concentrations of Rn and this tracer, it was possible to estimate the average time the contribution from each zone had spent in moving from the zone to Cape Point. Because the CB tracer is not a real gas it has no conventional unit, and has been ascribed an arbitrary scale factor.

Tracer 1 for each species was the contribution from a uniform flux from all land areas of the world. The model was run with unit source strength for each zone, and the contribution of each zone saved. After the run it was possible to vary the flux for the three tracer species for each zone to help interpret differences between the simulation and the observation.

Tracer zones 8 and 9 are north of the equator, and serve as a test of the model transport. As expected, transport from these zones to Cape Point proved to be negligible.

Fig. 1 has an expanded view of the area near Cape Point. This shows 4 grids labelled, S1, S2, S3 and L. They surround Cape Point, and when comparing the simulation to measurements, it was necessary to combine them to get the best match to the measured concentrations. For each grid, simulations were done for two heights, 38 m (level 1) and 184 m (level 2).

3.2 Model output processing

Rn was chosen as the test species because its sources were considered to be more uniform with time and location than the other species. The decision of how to combine the model output for the different grids and levels was based on the combination that gave best agreement with radon concentrations.

Fig. 2 shows the concentration of Rn for 5 days in March 1999. It is immediately clear that the level 1 simulation is frequently grossly higher than the observations. Thus only level 2 data were used for Cape Point simulations.

With Cape Point on the boundary between grids L and S3, it seemed logical to use an average of the two. Because the model does not allow for the discontinuity between land and sea or for locally generated winds, a procedure was devised to test whether improvements could be made by using different proportions of the grids when the local wind was from the sea rather than the land. Two parameters were selected to check the quality of the match for a full year’s data: standard deviation (STD), which emphasises the higher concentrations; and fraction of simulations within 50% of the observation (F50%), which gives all concentrations equal weight. Fortunately both parameters gave the same result.

A simple average of S3 and L grids yielded F50% of 0.507. The best result was for 30% S3 and 70%L when the experimental winds were in the land sector (0 to 120°), and 70% S3 and 30% L for other wind directions. This gave a slightly improved F50% of 0.531, and is represented in Fig. 2 by the solid line. Incorporating S1 and S2 did not improve F50%. Since the wind sector and relative proportions of S3 and L proved to have only a small effect on F50%, the use of the procedure optimum for the test period was likely to be close to optimum for other years. All simulations presented subsequently have been obtained...
using the wind direction dependent combination of S3 and L concentrations as described above.

Fig. 2: Experimental Rn concentration and simulation at two levels.

Fig. 2 illustrates the substantial improvement gained by combining S3 and L grids. It also shows that the simulation is not perfect, sometimes too high (day 63) and sometimes too low (day 66). This represents partly the limitations of the model and partly the inaccuracy of the estimate of the source flux. Model errors generally reduce when averages are made over longer periods. When there are long periods with a substantial proportion of a trace gas coming from a particular zone, it is possible to derive an estimate of the flux from that zone.

3.3 Sample age

In air samples there will be a mix of air which has contacted land over a range of distances from meters to thousands of km. It is of interest to know the approximate distance of the region which most influences the sample. The CCAM provides some estimate of distance because the contribution of each zone is known. However, with the large flux zones, there is a wide range of possible distances. An alternative is to estimate the transit time of the samples.

A good proxy for transit time is the “age” of a sample. This is defined in terms of the reduction in concentration of a species which has a steady decay rate, such as a radioactive isotope. In the case of Rn, if a sample has half the concentration it would have had if it had not decayed, its age is 3.84 days. This applies regardless of whether the sample is, say, half from a very close source and half from a very distant source, or all from a point which 3.84 days upwind.
Fig. 3: Continental background tracer for each flux zone.

For the CCAM run, the CB tracer had the same transport conditions as Rn, but had a 1 year half-life. For relatively short transport times, its concentration for each flux zone was therefore the same as Rn which had not decayed. Because of the slow decay, CB built up in the global atmosphere, particularly from large flux zones. Fig 3 shows CB from each zone for periods of 10 days from the start of the run and at the end. In order to use the CB for age calculation for recent transport, the CB was de-trended by subtracting a background in the same way as the other long-lived trace gases (Brunke and Scheel (2002)). The background concentration is shown as the dotted lines in Fig 3B.

The age is calculated from the formula:

\[ t = \frac{\ln(CB/Rn)}{\ln(2)} \]

Where \( L \) is the half-life of Rn (3.835 days)

It was found that precision limits within the CCAM resulted in noise being added to the age estimates. For very low Rn concentrations the age estimates were essentially random numbers which were not useful. In the present work, our interest is in the set of samples with levels of Rn above 250 mBq m\(^{-3}\), which is high enough to avoid precision problems except when the contribution of an individual tracer zone was only a small fraction of the total. Accordingly, age estimates are considered valid for an individual tracer zone when it contributes more than 5% of the total CB concentration. The number distributions of ages for this set are shown in Fig. 4. which also shows a table of the median tracer ages.

A reasonable attribute of the distributions is that the more distant flux zones have greater ages. Somewhat surprising are the very large ages up to 70 days for samples from zones 5 to 7. These imply that air parcels can retain concentrations well above background levels for many weeks. There is experimental evidence for such long atmospheric “storage” times. Piketh et al (1999) show that there can be absolutely stable layers from 850 to 700 mb which persist for tens of days, trapping trace gases from southern Africa. Cosijn et al (1996) report stable layers over southern Africa persisting for as much as 40 days. The large ages simulated by the CCAM are therefore consistent with observations.
More rigorous evidence of transport from zone 7 is not necessary for the present work because the age of the sum of all zones (Fig 4b) is never more than 15 days. This means that the amount of gas transported via stable layers to Cape Point is never a significant part of the total gas concentration. During periods when the distant zones do contribute substantially to the sum, such as 20 to 23 May 2005 (Fig 3b), transport is direct and the ages of individual tracers from zones 5, 6 and 7 are 3, 7 and 14 days respectively.

4. Comparison of the CCAM with measurements
4.1. Comparison of simulated wind direction with experiment

It was expected that the local topology would lead to differences between the modelled wind directions (CCWD) and experimental wind directions (Expt WD), particularly to the
north where there is a small mountain range which runs north to south for 30 km and reaches 400 m. Fig. 5 shows a time series with a range of conditions. In addition to the experimental and simulated wind directions, the figure shows a running 5 point standard deviation of the difference between these directions. This is sensitive to times when the wind direction is varying substantially on time scales of an hour and one would not expect the simulation to match the measurement. The 5 point std is not sensitive to steady systematic deviations such as those caused by local topology. The presence of such deviations is confirmed by events where the running std is low, but the difference between observation and simulation is large (18 and 28 March, Fig. 5). For the whole period of the study, 75% of the hourly data had the 5Pt Std below 27 degrees.

To make a more detailed comparison between the CCWD and Expt WD, the data were first separated into a mainly maritime set with Rn < 250 mBq m$^{-3}$, and a continental set with higher Rn. A "smooth" set was generated by including only those data points for which the 5Pt Std was < 27°. The CCWD was adjusted in increments of 360° so that the difference from the Expt WD was less than 180°. These sets were grouped into 5 degree Expt WD bins, and plotted in Fig. 6.

From this we can observe:

i). CCWD is strongly deviant from Expt WD from 330° through north to 90°.

ii). CCWD is mostly close to Expt WD for WD between 90° and 330°.

iii). The exception to ii) is at 180° where for the high Rn complete data set the median CCWD is about 80° higher than Expt WD.

iv). This discrepancy is not present in the smooth set.

v). The discrepancies other than that in iii) are much the same for all data as for smooth data.

vi). Agreement is best for the low Rn set, which also has a consistently high percentage of smooth data.

vii) The number of data points in the WD range for which discrepancies are large is small for the low Rn group, but large for the high Rn group.

It was expected that the CCWD would simulate Expt WD for the low Rn set more accurately than it does for the high Rn set. A quantitative indication of the accuracy is the median deviation of the CCWD values from the Expt WD. This is 21° for low Rn and 32° for high Rn. The discrepancy is present even when the WD is smooth in time.
Particularly relevant to the present work is the deviation in the continental sector, i). With a median absolute deviation of 51° from the Expt WD, the CCWD has no useful relationship to the Expt WD in this sector. This has two implications. Firstly, the effect of local trace gas sources will not be reflected in the model results. Secondly, the Expt WD will be a very poor indicator of the location of more distant sources.

4.2. Comparison of simulated radon with experiment

Radon is a good species for large scale transport model verification because it is inert and has a well defined source, the land. While there are wide source variations over kilometre scales, most dry land without snow cover emits close to 20 mBq m\(^{-2}\) of Rn. A survey of Australia soils by Schery et al, (1989) resulted in a mean flux of 19.2 ± 2.4 mBq m\(^{-2}\) with a standard deviation of the individual measurements of 21 mBq m\(^{-2}\). Because of its 3.82 day half life, radon tests transport on a continental rather than global scale.

4.2.1. CCAM transport timing

Fig. 2 shows that there is a measure of agreement between CCRn and ExptRn. Over the full 5 year period, the correlation between them was good, at 78%. To investigate the possibility that the CCAM output had a systematic time shift, the correlation was performed with a range of time lags between CC and Expt Rn. As Fig. 7 shows, the correlation vs lag curve peaks when the CCAM values 2 hours before the expt values have a slightly better correlation (79.5%).

On its own this was not considered enough to justify time-shifting the CCAM data, so an alternative way of checking the effect of time shift was devised. Correlation only requires that high values of one variable are associated with high values of the other. Given that most of the data change less than 10% per hour, correlation using the whole data set is not sensitive to the times when change is occurring. Yet these are the times when model timing is most important. The new test looked at times when the data were increasing (or decreasing) by more than 10% from one hour to the next. Because of our focus on continental data, and to avoid the effects of instrumental error at low concentrations, data were restricted to times when the expt Rn was greater than 250 mBq m\(^{-3}\).
Ideally, both the CCAM and experimental data would increase and decrease together. The test parameter was the number of samples for which the CCRn and expt Rn were increasing or decreasing together. This was calculated for a range of time lags between CCAM and expt. Fig. 8 shows that the two data sets move together most often when the time lag is zero, and markedly less often as the time lag moves away from zero. Clearly there is no systematic error in the CCAM transport timing.

4.2.2. Model accuracy for close zones

Radon also helps us investigate in more detail the indication from wind simulations (4.1) that transport from close sources was unlikely to be well modelled. It also helps to evaluate the effectiveness of the wind direction roughness filter to test for local influence. From CCAM it was possible to find the measurements for which particular zones yielded most of the radon.

<table>
<thead>
<tr>
<th>Wind direction variability</th>
<th>Predominant Zone (&gt;=50%)</th>
<th>Number of samples</th>
<th>Number of samples as % of total</th>
<th>Correlation of Expt Rn with CCRn</th>
<th>median Expt Rn 250 (mBq m^-3)</th>
<th>median CCRn 250 (mBq m^-3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rough</td>
<td>2</td>
<td>1652</td>
<td>21%</td>
<td>0.22</td>
<td>359</td>
<td>172</td>
</tr>
<tr>
<td></td>
<td>3+4</td>
<td>91</td>
<td>1%</td>
<td>-0.12</td>
<td>403</td>
<td>656</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2874</td>
<td>37%</td>
<td>0.45</td>
<td>1978</td>
<td>2231</td>
</tr>
<tr>
<td></td>
<td>6+7</td>
<td>57</td>
<td>1%</td>
<td>0.29</td>
<td>961</td>
<td>694</td>
</tr>
<tr>
<td></td>
<td>none&gt;50%</td>
<td>3174</td>
<td>40%</td>
<td>0.58</td>
<td>659</td>
<td>525</td>
</tr>
<tr>
<td></td>
<td>tr2&lt;50%</td>
<td>6196</td>
<td>79%</td>
<td>0.65</td>
<td>1166</td>
<td>1164</td>
</tr>
<tr>
<td>smooth</td>
<td>2</td>
<td>2234</td>
<td>12%</td>
<td>0.19</td>
<td>367</td>
<td>157</td>
</tr>
<tr>
<td></td>
<td>3+4</td>
<td>212</td>
<td>1%</td>
<td>-0.01</td>
<td>436</td>
<td>541</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>5706</td>
<td>30%</td>
<td>0.64</td>
<td>1460</td>
<td>1493</td>
</tr>
<tr>
<td></td>
<td>6+7</td>
<td>194</td>
<td>1%</td>
<td>0.20</td>
<td>703</td>
<td>533</td>
</tr>
<tr>
<td></td>
<td>none&gt;50%</td>
<td>10978</td>
<td>57%</td>
<td>0.66</td>
<td>450</td>
<td>339</td>
</tr>
<tr>
<td></td>
<td>tr2&lt;50%</td>
<td>17090</td>
<td>88%</td>
<td>0.76</td>
<td>628</td>
<td>508</td>
</tr>
<tr>
<td>rough + smooth</td>
<td>Close&gt;50%</td>
<td>13019</td>
<td>48</td>
<td>0.37</td>
<td>399</td>
<td>255</td>
</tr>
<tr>
<td></td>
<td>distant&gt;50%</td>
<td>14153</td>
<td>52</td>
<td>0.68</td>
<td>1222</td>
<td>1113</td>
</tr>
</tbody>
</table>

Table 1: Simulation performance for radon when a zone contributes more than half the total concentration. Conditions: Rn > 250 mBq m-3.
Table 1 shows some properties of groups of data selected according to which zone contributed most radon. Thus values in the top row are for samples where zone 2 contributed more than half the total. A further subdivision is made on the basis of the “smoothness” of the wind direction as defined in section 4.1.

The points in Table 1 most relevant to the accuracy of close zone simulations are:

i). Agreement between the median experiment and simulated Rn groups is remarkably good, even where the correlation is poor. This validates the use of CCAM to identify the source zones.

ii). The correlation between experiment and simulation is much worse when the zones closest to the observatory (2, 3 and 4) are predominant.

iii). The radon concentrations are very much lower when the close zones contribute most of the radon.

iv). The correlations are better for smooth compared to rough data in the vast majority of samples for which zone 2 contributes less than 50%

v). The differences between the rough and smooth results for each zone group are much the same for close and distant zones.

vi). There are too few points for the zone 3+4 and 6+7 groups to draw strong conclusions about these zones.

Points i) and ii) show that CCAM can be used to select data from either close or distant sources, but the simulated trace gas concentrations will be poor for the close zones. iv) and v) show on one hand that the roughness filter does select slightly better quality data, but on the other it does not discriminate between close and distant sources. It seems that CCAM does not simulate close zones well, but this is not because of wind direction variability. For the continental background it has been decided that it would be best to include the data with high wind direction variability because this set contains a lot of data which is not obviously more influenced by close sources than the smooth set.

We conclude that the CCAM can be used to divide radon data into local (zones 2 to 4) and distant (zones 5 to 7) groups, but the data do not support finer spatial subdivision. The last two lines of Table 1 show properties of the close and distant radon groups.

4.3. Comparison of simulated CO2 with experiment

The sources and sinks of CO2 are much more complex than those of Rn, so it is expected that the simulation will not be as accurate for CO2 as for Rn. The time series of expt CO2 and CC-CO2 for May 1999, shown in Fig. 9, illustrates some aspects of the simulation. There are very obvious brief periods when the expt-CO2 is as much as 10 ppm higher than the CC-CO2. These are likely to be times when the sample has been strongly influenced by the nearby Cape Town, an effect for which the model does not allow. Also, there are times such as May 7 when the CC_CO2 is drawn down sharply and the expt-CO2 follows some hours later.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Correlation for zero lag</th>
<th>Lag (hours expt after CC) for max correlation</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>all</td>
<td>0.12</td>
<td>5</td>
<td>0.34</td>
</tr>
<tr>
<td>Draw-down event, June 1999</td>
<td>-0.13</td>
<td>8</td>
<td>0.63</td>
</tr>
</tbody>
</table>

Table 2: Lagged correlations between expt- CO2 and CC- CO2
For the whole data set the correlation of CC-CO\(_2\) and expt-CO\(_2\) was 0.12 (Table 2), suggesting that the simulated CO\(_2\) was unrelated to the real world. However, we have an impression from Fig. 9 that there are times when the simulation worked, and that times free from local influence together with some time shifts could lead to some improvement.

First we studied draw-down events. These were times when the land contact with dense vegetation was short enough to prevent smearing out in the arrival time of the photosynthetic CO\(_2\) draw-down. There were only 10 events over 5 years when these conditions prevailed for a few days, the best of which was in June 1999. Expt-CO\(_2\) and CC-CO\(_2\) for this event are shown in Fig. 10, together with a summary of back trajectories.
The latter, shown in detail in Fig. 11, are consistent with the model: first land contact on 26th was fairly brief, followed by progressively longer land paths which would smooth out diurnal variation of mixing ratio. The best correlation of CC-CO₂ with expt-CO₂ was 0.6 when the CCAM was delayed in time by 8 hours. This was typical of the other draw-down events, and demonstrates that the model does allow for draw-down during passage over vegetation, but the timing is wrong. There is therefore an inconsistency within CCAM simulations of CO₂ for which the transport timing is correct, (4.2.1. CCAM transport timing, above) but the source/sink timing is wrong. This concept of a mix of right and wrong is supported by the time lag that gave the best correlation, 0.34 (Table 2), for the whole set. At 5 hours it was about half the lag of the draw-down events. We conclude that the CC-CO₂ hourly data have no useful relationship with hourly expt-CO₂.

![Fig. 11: Back trajectories for CO₂ draw-down event in June 1999. Legend format is yymmdd hh:mm. The trajectories were obtained from the NOAA ESRL website, http://www.esrl.noaa.gov/gmd](image)

**5. Air mass selection criteria**

Section 4 has established the properties of parameters that might be used to select air masses typical of different regions of the African continent. Here we examine how to use these parameters, and develop criteria based on the temporal behaviour of trace gas concentrations. The aim was to use experimental data for the selection as far as possible.
5.1. Wind direction

Wind direction at Cape Point is a rough indicator of air mass origin. Winds from the north to east sector bring high concentrations of trace gases. But, as shown in section 4.1, local wind direction has very little correlation with the synoptic scale winds which govern transport from land the order of hundreds of km away. Local topology strongly distorts the wind field in the sector of interest.

Even the synoptic wind directions used by the trajectory model give very little idea of the origin of air samples, as can be seen in Fig. 11. For example, trajectory 990626 12:00 arrived from the north east after crossing the west coast and spending less than a day over land. Trajectory 990628 12:00 arrived from the north west after spending many days over land to the north east of Cape point.

A further difficulty with wind direction is indicated by CCAM, which shows that there is a large difference between surface concentrations and those at the elevation of Cape Point (Fig. 2). Sometimes air that has passed above Cape Town would not be substantially affected by the city. It was therefore not possible to use either experimental or modelled wind directions as selection criteria.

5.2. Radon

Radon has been established as the most effective single criterion for determining if air is of maritime origin at Cape Point (Brunke et al 2004): 98.5% of samples with Rn concentrations below 100 mBq m\(^{-3}\) have marine baseline levels of trace gases. Brunke et al 2004 also show that samples with Rn between 100 and 250 mBq m\(^{-3}\) are slightly affected by land contact. We consider these to be of intermediate status, and so should be rejected from a study of continental air. The data used for this study therefore are required to have Rn>=250 mBq m\(^{-3}\).

The decay of Rn could be a problem because in principle a sample could have made land contact at a distant part of the continent then been transported without further land contact over a period of 10 days or so. Such samples could be strongly affected by the continent yet have low Rn. However, CCAM has shown (fig. 4) that even though some air from distant zones can be present with very long transport times, mixing with closer sources results in a median age of 2.3 days and 95% of the time an age of less than 6.2 days. The reduction in Rn with its half life of 3.82 days would therefore be about 30%. For most samples then, Rn can be taken as an indication of duration of land contact.

5.3. Age

Age has shown its value for interpreting Rn measurements. However, it is generated entirely from the CCAM simulation, which does not allow for the effect of close towns on trace gases. Given the imprecision of the simulation, it was considered best to rely on experimental Rn for estimating land contact time for individual measurements. Age will be used for evaluating data sets rather than as a selection criterion.

5.4. Persistent percentile filter for CO\(_2\)

The percentile filter for CO\(_2\) (CO\(_2\) (PF)) used for baseline air mass selection at Cape Point is described by Brunke and Scheel, (2002). The upper and lower cut-off limits are set by 11 day moving percentiles. It took the value 1 for baseline conditions and 0 otherwise. This
filter was very effective as a necessary condition for baseline, but it was not sufficient and let through 77% of the data, of which only one fifth was marine baseline. This marine component was removed automatically from our data set by the requirement that Rn be > 250 mBq m\(^{-3}\). The remaining data were influenced by land to varying degrees.

To pick out the set with minimal land influence, the CO\(_2\)(PF) filter was extended by adding the requirement that the filter value be 1 for at least 5 consecutive hours. This filter, CO\(_2\) (Ps5), or just Ps5, was tested by comparing the data sets obtained by applying it to marine baseline (Rn<100 mBq m\(^{-3}\)) and Rn>250 mBq m\(^{-3}\) data. The results are shown in Table 3. For marine baseline, Ps5 =1 for 80% of the data and the Ps5=0 set is almost identical to the Ps5=1 set. For the continental set with Rn>250, Ps5=1 for only 32% of the data and there is a marked difference between the Ps5=1 and Ps5=0 sets. The former has CO and CO\(_2\) mixing ratios very close to marine baseline and low Rn. The latter is clearly strongly influenced by land.

The Ps5 filter has strong selective power for marginally land-influenced data. Thus we have chosen our continental data set to be that with Rn>250 and Ps5=0.

<table>
<thead>
<tr>
<th>Data set</th>
<th>% of data with Ps5=1</th>
<th>median Rn (mBq m(^{-3}))</th>
<th>median CO (ppb)</th>
<th>median CO(_2) (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>marine baseline, Ps5=1</td>
<td>80</td>
<td>54</td>
<td>6.7</td>
<td>7.06</td>
</tr>
<tr>
<td>Ps5=0</td>
<td></td>
<td>55</td>
<td>6.6</td>
<td>7.06</td>
</tr>
<tr>
<td>Rn&gt;250, Ps5=1</td>
<td>32</td>
<td>379</td>
<td>7.7</td>
<td>7.13</td>
</tr>
<tr>
<td>Ps5=0</td>
<td></td>
<td>985</td>
<td>17.8</td>
<td>8.05</td>
</tr>
</tbody>
</table>

Table 3: Effect of the persistent percentile filter, Ps5.

5.5. Spike filter

Within the continental data is a mix of trace gases from greater Cape Town, nearby rural areas and more distant sources. We explored filters to exploit the higher variability from one hour to the next of local sources. A simple “spike” filter proved effective for CO\(_2\). The value Sp1(t) at a given time, t in hours is the sum:

\[
Sp1(t) = \text{abs}(C(t)-C(t-1)) + \text{abs}(C(t)-C(t+1))
\]

This is sensitive to the magnitude of changes in CO\(_2\), positive or negative, which is important because CO\(_2\) has both sources and sinks on land. This filter can be evaluated best in conjunction with the data divided according to the predominant zone, which is discussed below (5.6).

5.6. Predominant zone

The concept of predominant zone has been covered in section 4.2 for Rn. There it was shown that the sum from zones 6 and 7 of Rn was rarely more than 50% of the total Rn at Cape Point (Table 1). However, longer lived species can be transported over longer distances, so, for example, the sum of CO from zones 6 and 7 might well be more than half the total more often than Rn. For these species the CB tracer (3.1) is a better indicator of transport. Table 4 shows some results when the data are divided according to which zone group contributed most CB to the simulated value. Indeed zone group 6-7 is predominant for CB in 6.9% if samples, compared to 1% for Rn. The age of this group is also the longest, at 4.2 days, which would lead to a factor of 2 loss of Rn in transit.
For a zone group to be useful in the subsequent analysis, it needed to have enough samples to permit a study of seasonal and diurnal variations, and to have a distinct air mass origin. Zone group 2-4 clearly meets these requirements, as does zone 5. The distant zone group, 5-7, has almost a one day larger age and markedly lower CB and Rn concentrations compared to the zone 5 >50% zone group. It was not as easy to decide whether 6-7, the very distant group, was useful. Although it has only a marginal number of samples and the median CO and CO$_2$ are not very different from the other groups, it represents more distant land than the others. A final decision not to pursue study of this group further was based on various analyses which showed that any difference in trace gas concentrations between zone 6-7 and the others was small, and interpretation difficult because of the small numbers.

<table>
<thead>
<tr>
<th>zone group</th>
<th>Description</th>
<th>Number of samples (5-7 includes 6-7)</th>
<th>% of total</th>
<th>CB</th>
<th>Age</th>
<th>Rn</th>
<th>CO (ppb)</th>
<th>CO$_2$ (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-4</td>
<td>close</td>
<td>3521</td>
<td>20.1</td>
<td>27</td>
<td>1.6</td>
<td>457</td>
<td>9.7</td>
<td>7.9</td>
</tr>
<tr>
<td>5</td>
<td>medium</td>
<td>7119</td>
<td>40.6</td>
<td>147</td>
<td>2.1</td>
<td>1675</td>
<td>22.5</td>
<td>8.1</td>
</tr>
<tr>
<td>5-7</td>
<td>distant</td>
<td>6878</td>
<td>39.3</td>
<td>49</td>
<td>2.9</td>
<td>829</td>
<td>18.5</td>
<td>8.1</td>
</tr>
<tr>
<td>6-7</td>
<td>very distant</td>
<td>1208</td>
<td>6.9</td>
<td>47</td>
<td>4.2</td>
<td>779</td>
<td>22.3</td>
<td>7.9</td>
</tr>
</tbody>
</table>

Table 4: Number of samples for which the sum of the CB tracer in each zone group is more than 50% of total CB for the continental data set (Rn>250, Ps5=0). CB and age are medians of CCAM generated values, and Rn, CO and CO$_2$ are experimental medians for each zone group.

5.7. Back trajectories

Back trajectories have an intuitive appeal because they appear to show the likely source of land contact. Indeed, the experimental record and CCAM simulation of CO$_2$ shown in Fig. 10 are superficially in agreement with the back trajectories shown in Fig. 11. Until half-way through 26 June, the CO$_2$ was steady apart from what is probably a spike from nearby Cape Town, and the trajectories are all maritime. In the following days the CO$_2$ swings above and below the baseline level of 7 ppm, and the trajectories show progressively longer paths over land. It would therefore seem conceivable that one could use trajectories as a part of a set of selection criteria. For example, if one was interested in some region of the continent, one could select data for which the trajectories all passed over that region. This would involve a large investment in calculation and analysis of trajectories, so we have examined the set of trajectories in Fig. 11 in more detail.

For a selection criterion to be useful it should include most of the data for which the region of interest did contribute significantly to the trace gas signal, and exclude other data. It is well known that the line used to represent a trajectory can at best be regarded as an average of a very broad distribution. A brush which widened with distance from the end point would be more accurate. The question pertinent to this project was how broad the brush should be for continental trajectories.

The marine trajectories, before and including trajectory 990626 0.00 (Fig. 11), agreed well with both experiment and the CCAM, which simulated very low trace gas transport from all zones. At 990626 12:00 there was agreement to the extent that there was land contact, but the CCAM estimate that 2/3 of the contact was with zone 5 is very much higher than one would expect from the trajectory, which did not cross zone 5.
The subsequent trajectories again show superficial agreement with the CCAM in that there is strong land contact, with progressively less contribution from the close zones, 2-4. However, a detailed comparison reveals some serious differences, which we illustrate for the trajectory 990628 12:00 in Fig. 12. 32% transport from zone 6 and 15% from zone 7 indicates a much broader fetch region than the trajectory implies. Even more striking is the difference in the period over which the main contact with zone 5 occurred. With a simulated age of 1.7 days for zone 5, the CCAM indicates contact during the period when the trajectory was some 100 km out to sea. The long path over land implied by the trajectory appears to have led to a minimal contribution from the land over which it passed. A short period of land contact as indicated by the CCAM is supported by the experimental CO₂ trace with its strong draw-down. Over a long land path these photosynthetic draw-downs are not apparent because of mixing with CO₂ from nocturnal respiration.

Given the greater sophistication of the CCAM and supporting experimental evidence of our own and Piketh et al (1999), we conclude that the complexities of transport over the African continent are such that back trajectories provide only very limited indication of the predominant source region for trace gases at Cape Point. Use of trajectories as a basis for selection of data from a particular source region is thus unlikely to be valid.

Fig. 12: The contribution of each zone to total transport from land to Cape Point and age of each contribution at the arrival time of the trajectory.
6. Results

From the various selection criteria discussed in section 5 we have taken Rn and the persistent percentile filter for CO₂ to remove the mainly maritime data. The CCAM is used to select the data for which a particular group of zones contributes most land contact. Finally, a CO₂ spike filter helps to filter out data from nearby towns. Here we examine the properties of the data groups so selected.

6.1. Diurnal variation of sample number

Diurnal variations provide strong evidence of the transit time of a sample over the source region. In the artificial case where the maximum transit time is 24 hours, the source is uniform and the transport per m² the same from all distances, any diurnal variation in the source will be averaged out to zero at the observation point. As the duration of the contact with the source region increases, the observed diurnal variation decreases. The presence of strong diurnal variation therefore indicates that the contact time with the source region is less than 24 hours. There remains the question of whether the diurnal change is the result of transport or source variation. This where Rn is valuable. Its diurnal source variation over a hundred km or so is usually small compared to that of other trace gases.

Before considering properties of any data groups it is important to put them in the context of how many experimental points they represent. Fig. 13 shows the number of points in 3 hour time bins for each of the main zone groups with the data further separated according to the season (defined below in 6.3) and spike filter value. The choice of bin widths for time and CO₂ spike filter were a compromise between fine bins to pick up detail and having as many data points as possible. Because the groups with medium spike filter values usually had properties closer to the smooth than the spiky groups, they add little to the discussion and will not be presented below.

![Fig. 13: Number of samples in 3 hour time bins for each zone group, subdivided into seasons for smooth and spiky CO₂ filter groups. The horizontal axis shows time of day of the lower bin boundary for alternate bins.](image-url)
Most importantly, Fig. 13 shows that there is good representation of smooth data for most seasons, times of day and zone groups. Spikey data on the other hand are poorly represented after 3 pm, particularly in summer.

![Fig 14: Trace gas medians in 3 hour time bins for each zone group.](image)

### 6.2. Diurnal variation of trace gas median concentrations

Fig. 14 shows the trace gas medians in 3 hour time bins for each zone group. CO and Rn confirm the success of the CCAM based zone selection, with much higher CO and Rn from smooth zone 5 samples. The CO$_2$ spike filter is also very effective, selecting CO samples from zones 2-4 with smooth data only marginally higher than baseline and spiky data with very high concentrations showing strong diurnal variation. The very strong daytime peak for CO is likely to be from Cape Town, only 50 km to the north. The CO graphs strongly support the validity of the spike filter in selecting samples strongly affected by the city. CO from the smooth sets correlate well with Rn, which shows that on the larger distance scale the CO and Rn sources overlap. The more land you cross, the more towns you pass.

The spiky CO$_2$ trace with its markedly elevated concentrations and strong diurnal variation is adequately explained by the influence of the city on this group. However the smooth CO$_2$ agrees with neither Rn nor CO, and has diurnal variation of the same phase and amplitude for all three zone groups. Almost hidden by the 1.5 ppm diurnal variation, the increase in zone 5 daily average CO$_2$ is 0.4 ppm compared to zones 2-4. The major influence on smooth CO$_2$ is therefore whatever causes the diurnal variation, which has a phase consistent with the daytime draw-down and nightly respiration of photosynthesis. We suggest that this is vegetation within a few hours transport of Cape Point.

Another implication of the diurnal variation of CO$_2$ is that medians or averages over times longer than a few hours will be subject to sampling bias. The peak in the number of
observations coincides with the maximum CO₂ mixing ratio. Further bias will occur because of the wide variation in sample number with season. These biases will seriously distort analysis of the relationships between the groups, so we have examined the and season bins.

6.3. Seasonal variation

The seasons have been defined in the usual way with summer being the 3 months December, January and February and so on. All the trace gases had seasonal variation, but the strongest changes were for CO₂, shown in Fig 15. Medians are only displayed when there are at least 6 points.

The diurnal signal is strong for the data selected by the smooth CO₂ filter, and varies slightly with season, being slightly stronger in spring, which is consistent with the temperate annual vegetation growth cycle. A greater seasonal variation is evident in the “city” data selected by the spiky CO₂ filter. With less land contact, the zone 2-4 zone group has its highest peak in the 9 am bin in winter. Unfortunately there are not sufficient data to determine the full diurnal variation of this group. The other zone groups are consistent with rural CO₂ plus a city signal which moves the peak later in the day than it occurs in the smooth set, and raises the non-spring season CO₂ well above the spring levels. In some cases the diurnal variation is greater for the spiky than the smooth set. This is likely to be because of greater land contact, which seems likely given the higher Rn in the spiky set (Fig. 14) even with the reservation that most of the Rn comes from further afield than the CO₂.

One way to examine the relationships between the trace gases without sampling bias is to compare the medians for each time/season/zone/spike filter bin, which are plotted for CO₂.
in Fig 15. There are enough data to represent 258 of the 288 bins, and only summer night times with spiky data are poorly covered (Fig 13). Correlations of the trace gas medians yield a striking vindication of the filtering process (Fig. 16).

For the smooth set the strongest correlation is between Rn and CO, reflecting the blending of Rn and CO sources over long land paths. Next down is a lower but still significant correlation between Rn and CO₂. It is lower because the predominant influence on CO₂ is close rural land, which supplies only part of the Rn. This close rural land produces a negligible part of the CO, so there is no significant correlation between CO and CO₂.

When data are selected by the spiky CO₂ filter, the close city is the predominant source of CO and CO₂. Thus there is no significant correlation between Rn and CO or CO₂, whereas CO does correlate moderately well with CO₂.

6.4. Annual variation

A major goal of this work is to see if Cape Point data are useful in monitoring long term trends in trace gas emissions. Ideally, we should be able to follow annual and seasonal trends from as many source categories as can be derived from the selection criteria.

The spike filter allows us to divide the data from the close zones into a spiky set affected mainly by the close city and a smooth set for which any land contact is essentially rural. When the more distant zones are predominant in terms of transport, spiky data have a mix of city and rural or distant town influences, and are not useful for assessing annual trends. It is the smooth data from these zone groups that let us see the trends in trace gas concentrations attributable to rural areas. As a result, we can divide the data for annual analysis into four source categories: city, rural close, rural distant and rural very distant as set out in Table 5. The number of samples for each category (Table 5) is sufficient to divide the data into 4 seasons for each of the 7 years.

<table>
<thead>
<tr>
<th>source</th>
<th>spike filter</th>
<th>CCAM tracer zone group</th>
<th>total number of samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>city</td>
<td>&gt;2 (spiky)</td>
<td>2-4</td>
<td>700</td>
</tr>
<tr>
<td>rural, close</td>
<td>&lt;=1 (smooth)</td>
<td>2-4</td>
<td>2085</td>
</tr>
<tr>
<td>rural, distant</td>
<td>&quot;</td>
<td>5</td>
<td>3889</td>
</tr>
<tr>
<td>rural, very distant</td>
<td>&quot;</td>
<td>5-7</td>
<td>4276</td>
</tr>
</tbody>
</table>

Table 5: Data sets for annual analysis
The annual variation of each of the major parameters and trace gas concentrations for each season is shown in Fig. 17. Low sample numbers for the city group mean that those results may not be fully representative of the true influence of the city. Because of the strong diurnal variation of CO₂ concentrations, the median hour of day is shown in Fig 17.
Most of the samples were taken in the 3:00 to 12:00 period when CO₂ is at its maximum (Fig. 15). A true daily average would therefore be lower than the medians presented, but such an average would require more samples for each time of day group than was available.

Another diagnostic shown is the median sample age. This indicates that transport times over land have not varied much from year to year. The larger variation is with season, for which the distant and very distant groups have about one day shorter transport times in spring and summer. Radon concentrations are lower in these seasons, indicating that time spent in the zone is strongly related to land contact. These are the seasons for which the age of the close rural group is the same as that of the distant rural group. Transport from the latter group is faster, and to judge from the higher radon, much more efficient.
City CO₂ is substantially higher than rural CO₂, and at its highest in winter. It closely follows the city CO, as expected given the strong correlation between spiky CO and CO₂ shown in Fig 16. There is no clear annual trend.

Rural CO₂ has both seasonal and annual variations which can be understood in terms of the major biomes and crop distributions within about 400 km of Cape Point (Fig. 18). In all seasons except winter all three rural zones have very similar CO₂. In other words, the CO₂ is largely independent of the distance of the predominant area of land contact from Cape Point. Zone 5 grades from the mid- to low-productivity fynbos biome through succulent-karoo to the extremely low productivity nama-karoo, so this zone is not expected to contribute much CO₂. When the close zones predominate, there is no change, which indicates that the natural fynbos in these zones does not have a large effect on CO₂ at Cape Point. The clue to the origin of rural CO₂ is the strong diurnal variation (6.2), which proves that the source is very close vegetation within zone 2, and perhaps just the land to the north of the Cape.

For winter the discussion above does not apply. The close rural CO₂ is markedly lower than that of the other zone groups (Fig. 17). As shown in Fig. 18, there are extensive winter wheat areas in zones 3 and 4, which have very high productivity. Evidently they draw down the CO₂ sufficiently to have a bigger influence on CO₂ than the local vegetation which predominates in the other seasons. It would be expected that such a draw-down of CO₂ over a large area would lead to a reduced diurnal variation as samples with low CO₂ arrive at Cape Point with different transit times. Fig. 15 shows that this is reflected in the 3 hourly medians. The winter diurnal variation of close rural CO₂ (smooth, zones 2-4) is less than that for other seasons.

A feature of the rural spring CO₂ is the markedly lower median mixing ratio for years 2001 and 2002. This suggests higher growth rates of the local fynbos in these years, which were probably driven by rainfall. To avoid the complexity of analysing the time lags between rainfall and natural growth rate, we have taken the total annual Cape Point rainfall for each year and compared this with the close rural spring CO₂ median mixing ratios. As shown in Fig. 19, there is a strong anti-correlation between rainfall and CO₂ mixing ratio. This confirms the validity of the selection process. However, the detailed study of the rainfall distribution needed to assess the full significance of this relationship, is beyond the scope of this report.

![Fig. 19: Annual rainfall at Cape Point and close rural spring median CO₂ mixing ratio.](image-url)
CO has a stronger seasonal variation than CO₂, but unlike rural CO₂, rural CO has a similar seasonal variation to the city CO. In the case of rural CO, the variation is driven by transport rather than source strength. Therefore we find a very strong correlation between rural CO and Rn within each tracer zone group.

Between these zone groups there is a marked difference driven by the population densities of the source regions. Thus city CO/Rn is much higher (about 0.06) than close rural CO/Rn (about 0.02). The ratio for the very distant zone group which is predominantly savannah and grassland with a similar population density to the close rural group, is also about 0.02. With its main fetch region including the sparsely populated Karoo biome, the distant rural group has the lowest CO/Rn ratio of about 0.009. The experimental CO/Rn ratios therefore strongly support the sample fetch regions defined by the CCAM.

### 7. Conclusions

Selection criteria have been defined which make it possible to divide the samples with predominant transport to Cape Point from one of four source regions: City (greater Cape Town); close rural (within about 150 km of Cape Point); distant rural (from 150 to about 500 km) and very distant rural covering the rest of southern Africa. For species with fairly uniform source such as radon or CO, this regional selection worked well, but for CO₂ the respiration and photosynthesis cycle from very close vegetation usually overwhelmed the input from other sources. It was only in winter that the draw-down by wheat growing in the close rural zone was clearly evident in the Cape Point CO₂ data.

To achieve this degree of selection, a number of criteria were evaluated, of which some superficially attractive ones proved to be ineffective. These included:

- **a. Wind direction.** A comparison of measured and simulated wind directions showed that local topography destroyed the correlation of the measured wind direction with the bearing of the source of more distant trace gases from the critical north to east sector. The simulated wind direction was also not useful in this sector because of the unpredictable mixing with of air from local sources;

- **b. Back trajectories.** Although the back trajectories were effective in determining if contact with the southern African continent had occurred, the indicated time and location of land contact was highly inaccurate;

- **c. Simulated CO₂ mixing ratios.** The CCAM transport was good, which lead to good correlations with experimental and simulated Rn. However, the sources of CO₂ available for the model were inadequate for the purpose of the present study, which resulted in the simulated CO₂ being uncorrelated with the measurements.

Selection criteria which proved to be successful included transport information from the CCAM and experimental parameters based on the previous work on selection of marine baseline samples. These were:

- **d. Radon.** A threshold of 250 mBq m⁻³ proved useful to select samples strongly influenced by land without excluding samples from coastal towns which have large CO and CO₂ sources but a short land fetch;

- **e. Age.** The simulated age of samples was not used directly as a selection criterion, but was useful in evaluation of data sets selected by other criteria. For example, it was shown
that most transport was from land within a few days of Cape Point, which means that Rn decay did not detract from Rn as a selection criterion;

f. Persistent percentile filter. Based on long-term variability of the data, this filter was able to exclude samples which had essentially marine baseline levels of the other species even though their passage over land yielded 250 mBq Rn;

g. Spike filter. Derived from the change of CO₂ mixing ratio from one hour to the next, this filter proved highly effective in selecting samples from greater Cape Town;

h. Predominant zone. The direct contribution of the CCAM to the selection process was its selection of sample origin in terms of the zone group that contributed the greatest amount of direct transport. Although we have used a 50% cut-off for each zone group, different thresholds would permit focussing the selection to the needs of a particular study. Thus data availability could be improved by accepting more data from zones with strong sources.

Some improvement to the CCAM regional selection could be made by dividing zone 5 into several smaller zones within about 1000 km of Cape Point. Smaller zones within the close zone group would also be advantageous. If CO₂ remained the focus of the simulation, the zones would be aligned to major biomes or regions containing large sources.

In terms of the original objective of the project, the CCAM has shown the potential of CO₂ measurements at Cape point to provide estimates of the larger continental sources and sinks within a few hundred km of the station. However, as configured for this simulation, it is possible to extract only very limited information about these sources. To make full use of the CCAM, it would need to have the facility that is available for the other tracers: it should be possible to adjust the CO₂ flux for each zone after the run, thus making it possible to carry out inverse modelling to obtain flux estimates.

In the case of Rn and CO the transport component of the CCAM is probably good enough to permit inversion of carefully selected data. However, this would be better if the zone structure were revised for this purpose.

8. Endorsements

8.1
This report assesses a number of selection criteria to distinguish measurements of atmospheric composition at Cape Point that were heavily influenced by interaction with the southern African landmass, from samples that were predominantly of marine origin. The report is quite complex as it details a very thorough exercise and includes a number of criteria that in the end proved unsatisfactory for one reason or another.

In interpreting the final dataset of CO₂, CO and Rn detailed knowledge of the selection criteria used is required in order to understand the potential biases present in the data. Overall this represents an impressively thorough study of tools and techniques available to extract the most useful information from the measurements from Cape Point. The final objective of understanding the influences of land-atmosphere interactions over southern Africa in terms of the sources and sinks of CO₂ is of great importance in our efforts to better understand the global budget of this primary greenhouse gas.

Clare Murphy
Lecturer in Physical Chemistry, School of Chemistry
University of Wollongong
Wollongong, N.S.W. 2522, AUSTRALIA.

8.2
I have been through the report by Whittlestone et al. and am impressed by this very thorough study. I should say up-front that I am neither a climatologist nor do I have experience of modelling at this level, but I think I
understood the general approach and findings of the model. However, what was of great interest to me as a botanist/agriculturist, was the interpretation of the CO2 data, especially the seasonal and annual trends. The link made between the substantial draw-down in winter and the high productivity (ie. photosynthetic CO2 use) of winter grains seems to me to be a valid explanation, since productivity of natural plants and other crops in the region (e.g. vines, fruit) is very low during this season. The only other possibility is the higher productivity of natural and planted pastures during the wet winter season, but this comprises a much smaller area than winter wheat and I do not believe that it is a major contributor. The draw-down during 2001 and 2002 could well be linked to rainfall, although as the authors point out a more thorough investigation is beyond the scope of this study. However, it would be fascinating to examine the relationships between the Cape Point CO2 data and rainfall patterns (ie. potential productivity of natural ecosystems and rain-fed crops) using, for example, remote sensing (such as NDVI) or eddy correlation – perhaps someone with this expertise could be encouraged to participate in such an analysis.

Prof Stephanie Midgley (nee Wand)
Associate Professor
Dept Horticultural Science
Stellenbosch University, South Africa

9. Acknowledgements

The authors wish to acknowledge the contribution of Hans-Eckhart Scheel of (FZK IMK-IFU) . We thank him for both his valuable scientific insights and editorial assistance in preparing this manuscript.

10. References

Brunke, E.-G., Scheel, H.E., Seiler W., Trends of Tropospheric CO, N2O and CH4 as observed at Cape Point, South Africa. Atmospheric Environment 24A (1990), 585-595.


Brunke, E.-G, C. Labuschagne, B. A. Parker, H. E. Scheel and S. Whittlestone, Baseline air mass selection at Cape Point, South Africa: application of 222Rn and other filter criteria to CO2, Atmospheric Environment 38 (2004), 5693-5702


McGregor, J.L., Accuracy and initialization of a two-time-level split semi-Lagrangian model. Collection of papers presented at WMO/IUGG NWP symposium, Tokyo, 4-8, August 1987 , 233-246.


