GEOPHYSICAL RESEARCH LETTERS, VOL. 28, NO. 1, PAGES 175-178, JANUARY 1, 2001

The utility of remotely sensed CO_2 concentration data in surface source inversions

P. J. Rayner

Cooperative Research Centre for Southern Hemisphere Meteorology and CSIRO Atmospheric Research, Aspendale, Victoria, Australia

D. M. O'Brien

CSIRO Atmospheric Research, Aspendale, Victoria, Australia

Abstract. This paper aims to establish the required precision for column-integrated CO_2 concentration data to be useful in constraining surface sources. We use the method of synthesis inversion and compare the uncertainties in regional sources calculated from a moderate-sized surface network and either global or oceanic coverage of column-integrated pseudodata. With a simple measure of total uncertainty, we require precision of monthly averaged column data better than 2.5 ppmv on a $8^\circ \times 10^\circ$ footprint for comparable performance with the existing surface network. If coverage is only oceanic we require 1.5 ppmv precision. We recommend more detailed studies on the feasibility of obtaining such observations from current and future satellite instruments.

1. Introduction

There is a growing need, both scientifically and policy driven, to monitor the global carbon cycle. While current international agreements such as the Kyoto Protocol seek to control aspects of the carbon cycle most directly under human control, the over-arching aim is to stabilize trace gas (particularly CO_2) concentrations. Such an aim requires monitoring of the total source of CO_2 to the atmosphere. For this reason, as well as for intrinsic scientific interest, the attempt to understand quantitatively the spatial structure of carbon sources has gathered pace over the last decade.

The most common technique, called synthesis inversion, uses the integrating power of the atmosphere to estimate large-scale sources from their atmospheric signature in space and time. This requires assumptions about the smoothness of sources, in general that they vary according to certain large-scale patterns. It is the magnitude or scaling factors for these patterns that are estimated in this technique. The approach is limited both by the quality of the relationship between the unknown surface source and the observations as well as the simple limitation of data. The level of spatial detail is limited by the spatial coverage of observations. The limitation is manifested as large uncertainties on local sources or extreme sensitivity to additional data. Obviously we need additional data but these also are expensive and difficult. In general it seems wise to assess the utility of potential new data sources in providing useful constraints.

Paper number 2000GL011912.
0094-8276/01/2000GL011912$05.00

In this paper we aim to set thresholds at which column-integrated data may be useful for such inversions. The motivation is the suggestion that such data will be available, although imprecisely, at high spatial coverage from upcoming satellite missions [e.g. *Chedin* 1999].

There are several potential techniques for retrieving information on atmospheric CO_2 distributions from satellites. For example we can measure radiances in thermal infrared bands of CO_2. Similar techniques have been studied for CO, CH_4 and O_3 [see for example *Clerbaux et al.* 1998, *Wetzel et al.* 1995]. The technique requires that the surface properties, temperature profile and emissions by other gases (notably H_2O) are known. Coverage would be limited by cloud, but the instrument would function by both day and night. An instrument such as AIRS on the EOS PM platform would provide twice daily global coverage at 15 km horizontal resolution. After allowing for cloud, one could expect global coverage every few days.

Alternatively, we could measure sunlight reflected from the surface at frequencies in near infra-red absorption bands of CO_2. The technique is relatively insensitive to the temperature profile but functions only in sunlight and cloud free conditions. The water vapour profile need be known only approximately. Because Rayleigh scattering is negligible at these frequencies, the principal scattering sources are undetected cloud and aerosol. Experiments carried out by *O'Brien et al.* (1998) in the context of measuring the column amount of oxygen suggest that high precision (0.1%) is possible. *Aoki et al.* (1993) have studied the even more ambitious task of determining vertical profiles of trace gases using tunable etalons. In both the preceding studies, it was assumed that the sensor would observe sunglint over the oceans in order to increase the signal-to-noise ratio at the expense of spatial coverage. Nevertheless, extensive regions of the globe still would be sampled on every orbit of the satellite [*O'Brien* 1990].

Any retrieval method is likely to be fraught with difficulty. Even if the observational problems can be overcome the potential utility of such data is not clear. In general the surface sources of CO_2 we wish to constrain are input into a preexisting column of CO_2, which makes their signature hard to detect. The potential pay-off is large though. There is a chance of complete spatial coverage at weekly frequencies or better on a footprint of tens of kilometres. This is perhaps five orders of magnitude more concentration data than have been available until now. This sheer volume of low precision data may outweigh the small amounts of high pre-

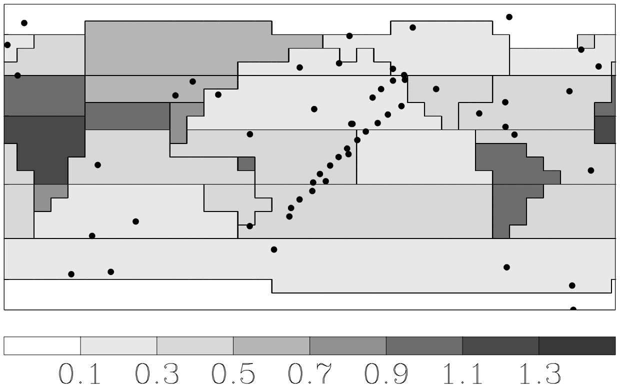

Figure 1. Standard deviation of sources (GtC yr^{-1}) for 26 regions in an inversion using 56 mainly surface stations. Stations are marked by dots.

cision data currently available. Here we will try to establish benchmarks of precision for this to occur.

2. Methods and Data

The method we use is Bayesian synthesis inversion previously used for atmospheric trace gas studies by *Enting et al.* (1995) and *Rayner et al.* (1999). We choose a series of source regions in advance and separately use each as a flux boundary condition for an atmospheric transport model. We sample the resulting concentrations at a series of actual or proposed observing sites. We adjust a scaling parameter for each source region until we obtain an optimal match for concentrations at those sites. If we use a quadratic cost function to express optimality we can use linear regression to solve for the magnitudes. An advantage of such regression is that we find the uncertainties for the source magnitudes independently from the optimal estimate. This means we can find the uncertainty of sources returned by a particular data network before that network is built, provided we can specify the error statistics of the data. This property has been previously used by *Rayner et al.* (1996) and *Gloor et al.* (2000) to suggest optimal extensions to existing networks for monthly mean observations.

The only requirement for the so-called pseudodata from a proposed network is that it is a linear function of the sources. In previous studies the pseudodata have been monthly mean surface observations at a point. In this study we extend the idea to consider vertical integrals of concentration. Such integrals are, in general, weighted according to the radiative properties of the atmosphere and instrumental characteristics. In this paper we assume the simplest case of uniform weighting i.e. a simple vertical integral.

The other choice of input to our procedure is prior knowledge of fluxes and their uncertainties. For prior uncertainties we choose 1.0 GtC yr^{-1} for most ocean regions except the well-sampled far North Atlantic and North Pacific for which we use 0.5 GtC yr^{-1} and the large area of the midlatitude South Pacific for which we use 1.5 GtC yr^{-1}. We use a uniform 1.2 GtC yr^{-1} for all land regions and 0.3 GtC yr^{-1} for the fossil fuel source.

We also need to choose a set of regions for which we seek flux estimates and an atmospheric transport model to relate fluxes to concentrations. In both of these we follow *Rayner et al.* (1999). This means we solve for 26 regions, 12 over ocean and 14 over land as well as one pattern for the global fossil fuel source. The transport model is the Goddard Institute for Space Studies (GISS) model [*Fung et al.* 1983].

3. Results

As a baseline, Fig. 1 shows the predicted uncertainty of annual mean sources in a calculation using 56 stations (shown as dots) from *GLOBALVIEW-CO2* (1999). Data uncertainties are also taken from *GLOBALVIEW-CO2* (1999). Data uncertainties reflect variability of actual flask data around a fit to the monthly mean, this variability being an indication of the stability of a monthly mean estimate. Unsurprisingly, source uncertainties are lowest for those regions that contain observing sites. This means that large regions will, by construction, have a better chance of being well-constrained. Also the bias of the network towards marine sites is manifest in large uncertainties over many land regions, particularly in the tropics and south. Although we have not shown it here, the surface network also is less effective in regions where signals are diluted by rapid vertical transport so that a given observation in the tropics will produce less impact than one of the same precision at high latitudes.

To demonstrate the potential impact of column data, Fig. 2 shows the same predicted source uncertainty in the case of a column-integrated observation of monthly mean CO_2 concentration over each $8^\circ \times 10^\circ$gridbox of the atmospheric transport model. Each pseudodatum is assigned a precision of 1 ppmv. As might be expected from the coverage of observations, the reduction in uncertainty is much more uniform across the globe. Only in the southern Indian ocean does the surface network perform much better than the column data while the column data do better over almost all land regions. The dilution of signal (and hence data impact) by vertical transport does not appear in the column data experiment.

To summarize the sensitivity of source uncertainty to data precision, Fig. 3 shows a measure of global source uncertainty Σ as a function of the precision of the column-

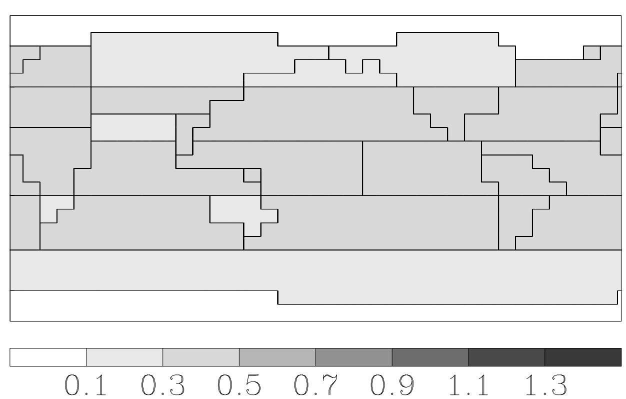

Figure 2. Standard deviation of sources (GtC yr^{-1}) for 26 regions in an inversion using global coverage of column-integrated concentration pseudodata. Precision on the pseudodata is set at 1 ppmv.

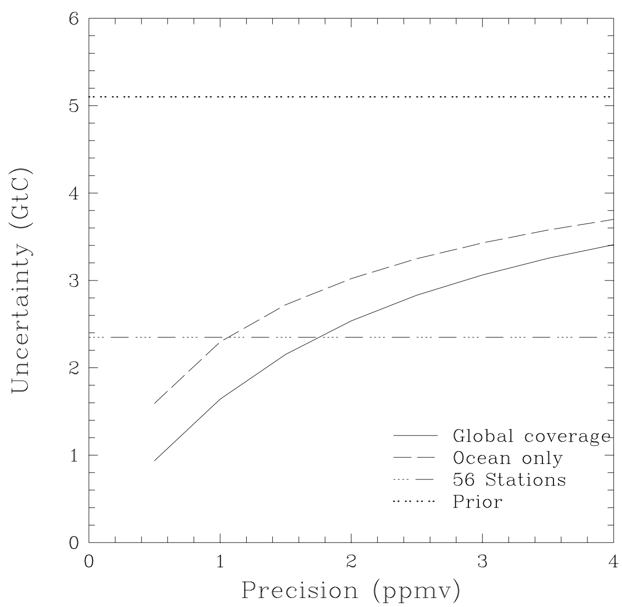

Figure 3. Plot of "total uncertainty" (see Eq. 1) in GtC yr^{-1} against the precision (ppmv) of column-integrated data. The dotted horizontal line shows the prior uncertainty while the dash-dot horizontal line shows the case for the surface network of Fig. 1.

integrated data δ. We calculate Σ as

$$\Sigma = (\sum_n \sigma_n^2)^{1/2} \quad (1)$$

where σ_n is the uncertainty for region n and we sum over all 26 regions. The solid line shows the case of a global column data coverage (Fig. 2 was the $\delta = 1.0$ point on this line) while the dashed line shows the case where retrievals are only possible over those gridboxes that contain only ocean. The dotted horizontal line shows the prior uncertainty and the dash-dot horizontal line the existing surface network case from Fig. 1. Under this gross measure, the column data require a precision of 2.5 ppmv for an equivalent performance to the surface network and the ocean-only case requires $\delta < 1.5$ ppmv.

4. Discussion

The results above suggest that column-integrated data can provide a comparable constraint to current surface networks. This is perhaps surprising given that surface sources induce weaker signatures in the total column amounts than in surface concentrations. There appear to be three major reasons for the relative power of column observations. The first, as we mentioned in the introduction, is the global coverage. The biased constraint of the surface network is clear from Fig. 1 and Fig. 2. In general, uncertainties are lower for the surface network wherever a station lies within a region. This is true even though the precision on the surface concentration observation is, in some cases, worse than for the column data. Much of the contribution to the total uncertainty comes from the poorly observed tropical and southern continents all of which fare better with the column observations. The importance of global coverage is also reflected in the more stringent precision requirement in the ocean-only case.

Second is the sheer quantity of data. We can, roughly speaking, calculate an analogous measure of total data uncertainty for the two networks as

$$\Delta = 1/N(\sum_n \delta_n^2)^{1/2} \quad (2)$$

where δ_n represents the uncertainty of the nth observation and N is the total number of observations. For the column data, δ is uniform and $N = 792$. For the surface network, δ is highly variable and $N = 56$ yielding $\Delta = 0.12$. If data precision alone were the determinant of data impact, a column network with $\delta = 3.4$ would be equivalent to the surface network. As we can see from Fig. 3 δ needs to be lower than this for the two networks to perform equivalently, a symptom of the dilution of surface sources in the column observations.

A third reason for the relative performance of the column observations lies in the nature of atmospheric transport. While surface sources are relatively diluted in the large background concentration of the total column, this dilution is relatively uniform across the globe. To see this directly involves comparing the response functions from different regions, which is beyond our scope here. However the behaviour is manifest in the relatively uniform reduction in uncertainty across the globe. For surface observations the pattern is different. Rapid vertical mixing in the tropics dilutes surface signals of surface sources so that a given observation has less impact than in the relatively stable atmosphere of the extratropics. This suggests that even without noise and feasibility problems for increased tropical sampling, surface observations will be challenged in this region.

The other question one can ask of such a data impact comparison is whether it is fair. This question turns on the assignment of prior uncertainties to sources and of uncertainty to the data and pseudodata. We have tested the sensitivity to prior source uncertainty by a parallel series of experiments in which prior uncertainty for a region is set proportional to the area of the region. This is equivalent to assuming that fluxes are equally uncertain everywhere. The column data perform equally well in both cases. In fact, the linear region of the two curves in Fig. 3 covers a regime in which source uncertainty is determined by data uncertainty not prior source uncertainty. Thus, results in this regime are insensitive to the assignment of prior uncertainty, including not using prior estimates at all. The surface network performs better in the second case since the poorly sampled continental regions are smaller, and hence have smaller prior uncertainties. The precision required for the column data to outperform the surface network is now 1.5 ppmv for global coverage.

The data uncertainties for the surface network (ranging from 0.3 ppmv to 2.4 ppmv) are much larger than suggested by instrumental precision. The overwhelming contribution to this error comes from our inability to fit a series of point observations at one point in space using a statistical model based on sources over nearly continental scale and constant throughout a month. Further, the transport model resolution of several hundred km is also inadequate to model observations at a point, notwithstanding efforts to maximize the atmospheric volume sampled by observations. It is hard to quantify the various contributions to this error but the empirically determined errors here have the right structure, being large in regions with heterogeneous sources nearby. In this respect the column data are closer to the simulated quantity so that this error will make a smaller contribution.

It is also likely that the column-integrated data are less susceptible to uncertainties in model transport than the surface network. In their model comparison, *Law et al.* (1996) showed that, on a hemispheric scale, model-model differences in vertically integrated concentration were about half those at the surface for the fossil-fuel and terrestrial biosphere sources. If we can regard the 500 hPa level as a proxy for the vertical integral, this was also shown by *Rayner and Law* (1995) (Figs. 3.5 and 4.9) at smaller scales. These uncertainties in forward models would also affect inversions.

Finally we should ask whether the data uncertainties used for the column data are possible from satellite retrievals. The precision is probably greater than that from a single retrieval. However the pseudodata used here are for concentration averaged over a $8^\circ \times 10^\circ$ model grid cell. For a 100 km footprint, this represents an average of approximately 80 retrievals. While uncertainties on these retrievals are unlikely to be completely independent, nine independent retrievals with an uncertainty of 6 ppmv would yield an uncertainty in the mean concentration of 2 ppmv, within our required precision provided the coverage is global. Also, the more detailed sampling of these individual retrievals should allow estimates on smaller spatial scales than we have attempted here.

5. Conclusions

We have estimated the required precision for column-integrated concentration data for use in synthesis inversions. We have found that we require a precision of 2.5 ppmv for monthly mean column-integrated concentration on an $8^\circ \times 10^\circ$ footprint for equivalent performance to a moderate global surface network. A major advantage of the column data as used here is their global coverage; improvement is not as good if the coverage is only oceanic. Given these findings, we recommend further study into the feasibility of physical retrievals of such data. Once characteristics of such retrievals are established (e.g. vertical weighting) the calculations presented here will need to be repeated using such characteristics.

Acknowledgments. This study was carried out with the support of the Australian Government through its Cooperative Research Centres Programme. The authors wish to thank Ian Enting, Rachel Law and Fred Prata for their comments on the manuscript.

References

Aoki, T., M. Fukabori, and T. Aoki, Trace gas remote sounding from near IR sun glint observations with tunable etalons, in *High spectral resolution infrared remote sensing for Earth's weather and climate studies*, edited by A. Chedin, M. T. Chahine, and N. A. Scott, vol. 19 of *ASI*, pp. 309–322, Springer-Verlag, Berlin, 1993.

Chedin, A., N. A. Scott, R. Saunders M. Matricardi, J. Etcheto and C.Clerbaux, IASI Capability to Monitor CO_2 , *Final Report, EUMETSAT Contract No 98/310*, 1999.

Clerbaux, C., P. Chazette, J. Hadji-Lazaro, G. Mégie, J.-F. Müller, and S. A. Clough, Remote sensing of CO, CH_4, and O_3 using a spaceborne nadir-viewing interferometer, *J. Geophys. Res.*, *103*, 18,099–19,013, 1998.

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

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

GLOBALVIEW-CO2, Cooperative AtmosphericData Integration Project - Carbon Dioxide, CD-ROM, NOAA/CMDL, Boulder, Colorado, 1999, [Also available on Internet via anonymous FTP to ftp.cmdl.noaa.gov, Path: ccg/co2/GLOBALVIEW].

Gloor, M., S. M. Fan, S. Pacala, and J. L. Sarmiento, Optimal sampling of the atmosphere for purpose of inverse modelling - a model study, *Global Biogeochem. Cycles*, *14*, 407–428, 2000.

Law, R. M., et al., Variations in modelled atmospheric transport of carbon dioxide and the consequences for CO_2 inversions, *Global Biogeochem. Cycles*, *10*, 783–796, 1996.

O'Brien, D. M., and R. M. Mitchell, Zones of feasibility for retrieval of surface pressure from observations of absorption in the A-band of oxygen, *Tech. Rep. 19*, CSIRO Div. Atmos. Res., Aspendale, 1990.

O'Brien, D. M., R. M. Mitchell, S. A. English, and G. A. Da Costa, Airborne measurements of air mass from O_2 A-band absorption spectra, *J. Atmos. Oceanic Technol*, *15*, 1272–1286, 1998.

Rayner, P. J., and R. M. Law, A comparison of modelled responses to prescribed CO_2 sources, *Tech. paper no 36*, CSIRO Div. of Atmos. Res., 1995.

Rayner, P. J., I. G. Enting, and C. M. Trudinger, Optimizing the CO_2 observing network for constraining sources and sinks, *Tellus*, *48B*, 433–444, 1996.

Rayner, P. J., I. G. Enting, R. J. Francey, and R. L. Langenfelds, Reconstructing the recent carbon cycle from atmospheric CO_2, $\delta^{13}C$ and O_2/N_2 observations, *Tellus*, *51B*, 213–232, 1999.

Wetzel, G. H., G. H. Fischer, and H. Oelhaf, Remote sensing of trace gases in the midinfrared spectral region from a nadir view, *Appl. Opt.*, *34*, 467–479, 1995.

D. M. O'Brien, P. J. Rayner CSIRO-DAR, PMB 1 Aspendale, Victoria 3195, Australia. (e-mail: denis.obrien@dar.csiro.au, peter.rayner@dar.csiro.au)

(Received June 19, 2000; revised October 11, 2000; accepted October 16, 2000.)