Tellus (1996), 48B, 433–444
Printed in UK – all rights reserved

TELLUS
ISSN 0280–6495

Optimizing the CO_2 observing network for constraining sources and sinks

By P. J. RAYNER*, I. G. ENTING and C. M. TRUDINGER
Cooperative Research Centre for Southern Hemisphere Meteorology, c/o CSIRO DAR, PMB 1, Aspendale 3195 Australia

(Manuscript received 12 October 1995; in final form 11 March 1996)

ABSTRACT

This paper presents a combination of synthesis inversion and simulated annealing as a method for suggesting optimal networks for constraining the global carbon budget. Synthesis inversion uses an atmospheric tracer transport model and, in our applications, prior estimates of sources to return estimates of sources' strengths along with their uncertainties. Simulated annealing is a commonly used technique for optimizing multi-variate functionals. By treating the predicted uncertainty for a source as this functional and station locations as parameters, we are able to suggest improved observing networks. The results suggest that considerable improvement is possible with current station densities. Further, they suggest surprising constraints are possible because of the combination of apparently disparate data by the inversion technique. For example, the best improvement available in estimates of total ocean uptake from one extra station is obtained when that station is sited over tropical South America. This follows from the global balances of the CO_2 budget. Finally, considerable specificity is required in defining the objectives for such an optimization, since apparently similar quantities might require very different strategies.

1. Introduction

It is clear that a significant source of uncertainty in prediction of climate change is the future concentrations of the radiatively active trace gases themselves (Schimel et al., 1995). Furthermore, a major source of uncertainty in future trends of the main important trace gases arises from uncertainties in the current budget. Thus we should, where possible, try to reduce uncertainty in current budgets of trace gases. We should also attain the capability to monitor changes in these budgets, both for assessing the effectiveness of any emission control actions and for detecting any CO_2-climate feedback processes. In the case of the single main radiatively active trace gas, CO_2, this problem is complicated by the number of processes controlling the atmospheric budget.

The primary measurements we use to constrain the atmospheric carbon budget are either of fluxes or concentrations. Flux measurements, like those of Takahashi et al. (1993) over the oceans or Wofsy et al. (1993) over the land, have the advantage that they contribute directly to the budget. The disadvantage is that they are highly local in space and time. Given the high degree of spatial and temporal variability in such fluxes, gathering enough measurements to produce a stable estimate of the budget over a large region over several years is problematic. Also many of the locations likely to exercise strong control over the global carbon cycle (e.g., the high latitudes of the southern ocean) are remote.

An alternative way of determining source strengths is to measure concentrations and derive

* Corresponding author.

fluxes from them, i.e. the inverse approach. This is a classical inverse problem with all its attendant difficulties.

There is a growing set of concentration measurements performed by several groups around the world, e.g., Conway et al. (1994), Keeling et al. (1995) and Francey et al. (1996). In the past, there has been one overwhelming criterion for siting of measurements of CO_2, namely the avoidance of local contamination. This choice was guided by the real concern that, with few measurements ever likely to be available, it was important that each observation should characterize as large a region as possible. Even with the current network, the measurement community takes care to construct so-called baseline conditions in which the effects of local sources are screened out. The strong local sources and sinks which are being avoided under the siting policy are usually associated with the terrestrial biosphere. Consequently, measurements have usually been taken either in the middle of ocean basins or on continental margins downwind of ocean basins. This leaves considerable gaps in our knowledge, only partly addressed by the detailed records available from individual continental sites, e.g., Bakwin et al. (1995).

The measurement network was evolved considerably over time and will continue to do so. Not only new stations but also new types of data are being added constantly, e.g., Keeling et al. (1993) for oxygen-nitrogen ratios. The measurements are difficult and expensive. It seems reasonable then to use whatever guidance we can in selecting optimal locations for making such measurements.

In this paper we outline one technique for suggesting optimal extensions to the current observing network for constraining CO_2 sources and sinks. The next section outlines the method involved after which a series of case studies and analyses are presented. The first case is the complete reorganization of the current observing network. This is both a feasibility check on the method and a measure of the limit of detection with the current number of stations. As a more potentially useful example the following section considers the addition of a few extra stations to the current network. In particular we see the coupling of various data sources through the inversion procedure leads to some surprising results. Finally we apply the technique to different measures of uncertainty to see how suggested networks might change if we target different quantities.

2. Method

The approach used in this paper is a combination of two pre-existing methods, namely Bayesian synthesis inversion and simulated annealing. We will not describe either method in detail but hope to provide enough information to understand the later applications. The application of simulated annealing to the design of sampling networks for an inversion problem derives from the seismological studies of Hardt and Scherbaum (1992, 1994).

2.1. *Synthesis inversion*

Synthesis inversion is a technique for recovering source estimates from concentration data by the use of an atmospheric tracer transport model. In this paper we use the term "source" to refer to a net source which might in fact be a sink. The Bayesian synthesis method was described and used by Enting et al. (1993, 1995). In synthesis inversion, sources of predetermined or assumed structure and unit magnitude are used as inputs to a transport model. Each source produces a pattern of responses over the globe. In particular each source produces a unique set of concentrations at observing stations. The final step is to find the linear combination of source magnitudes which best matches the observed structure. This is treated as a linear least squares problem. We calculate a matrix $\mathbf{A}$ by inserting each proposed source (at unit magnitude) into an atmospheric tracer transport model and sampling the resulting concentrations at observing locations. For this work, the model is that of the NASA Goddard Institute for Space Sciences (GISS) and is described by Fung et al. (1983). If we further define the vector of unknown source magnitudes by $\boldsymbol{S}$ and the vector of observed concentrations by $\boldsymbol{D}$ we are left solving the least squares problem

$$\mathbf{A}\boldsymbol{S} = \boldsymbol{D} \tag{1}$$

i.e. we are to minimize the length of the vector given by $\mathbf{A}\boldsymbol{S} - \boldsymbol{D}$. This assumes homogeneous variance of the data whereas in practice we weight the residual by the inverse of the data covariance.

One novel feature of the two studies cited above is the Bayesian approach which uses prior estimates for source magnitudes. These initial guesses act both to add other information, such as flux observations or other modelling studies, as well as to stabilize the inversion of an underdetermined problem. We denote the initial guess for the sources by $\boldsymbol{S}_0$ with covariance $\mathbf{C}(\boldsymbol{S}_0)$. Now we are to minimize the mismatch between initial and predicted sources as well as the residuals of the predicted responses from the observations. This can be written as

$$|(\boldsymbol{S}-\boldsymbol{S}_0)\mathbf{C}(\boldsymbol{S}_0)^{-1}(\boldsymbol{S}-\boldsymbol{S}_0)^{\mathrm{T}}| + |(\mathbf{A}\boldsymbol{S}-\boldsymbol{D})\mathbf{C}(\boldsymbol{D})^{-1}(\mathbf{A}\boldsymbol{S}-\boldsymbol{D})^{\mathrm{T}}|. \tag{2}$$

The products of such a minimization include not only an optimal estimate for $\boldsymbol{S}$ but also the estimated covariance matrix $\mathbf{C}(\boldsymbol{S})$ which represents the confidence we should have in that estimate. It is a general property of such linear least squares calculations that $\mathbf{C}(\boldsymbol{S})$ does not depend on $\boldsymbol{S}_0$ or $\boldsymbol{D}$ but only on $\mathbf{C}(\boldsymbol{S}_0)$, $\mathbf{C}(\boldsymbol{D})$ and $\mathbf{A}$. The expression for the estimated source covariance, taken from Tarantola (1987, equation 1.90) is

$$\mathbf{C}(\boldsymbol{S}) = [\mathbf{A}^{\mathrm{T}}\mathbf{C}(\boldsymbol{D})^{-1}\mathbf{A} + \mathbf{C}(\boldsymbol{S}_0)^{-1}]^{-1} \tag{3}$$

An important consequence of the independence of $\mathbf{C}(\boldsymbol{S})$, the predicted error estimates of sources, from actual data is that we can predict the impact on variance of changes to the station network even though we have no data at the new fictitious stations. This is the approach we will use: manipulate the observing network and minimize the estimated variance of one or more sources. The choice of exactly what quantity we choose to minimize will turn out to be of pivotal importance in our analysis and we postpone discussion of this point until later.

An advantage of synthesis inversion techniques is that the data against which we match source combinations can take many forms. Enting et al. (1995) used the detrended annual mean concentrations and seasonal cycles of CO_2 at 20 stations as well as the $\delta^{13}C$ ratios from 5 of these stations. The seasonal cycles were represented by their first two Fourier components. They also considered global trends in CO_2 and $\delta^{13}C$. These two global trends actually control the global budgets. The additional information provided by the spatial distributions does little to help reduce the uncertainties in these budgets. The use of Fourier components from the data means that Fourier components must be calculated for each response. The seasonality in the response may arise either from the seasonality of the source itself or the seasonality of transport.

While the actual data values used have no impact on the analyses presented here, the types of data used and the variance structures assigned to them well might. For our initial cases we use the same variance for data as Enting et al. (1995) namely: there is no correlation among data errors, the errors are homogeneous so that no station is considered better or worse than another, the standard deviation of annual mean CO_2 concentrations is 0.3 ppmv and that for the amplitude of seasonal responses is 0.5 ppmv. The initial uncertainty of the $\delta^{13}C$ values varies spatially as described in Enting et al. (1993). For the annual trends, the initial errors are 0.15 ppmv yr^{-1} for the CO_2 concentration and 0.005‰ yr^{-1} for the $\delta^{13}C$. The possibility of better "error models" for the observations is discussed by Enting and Pearman (1993) and Ciais et al. (1995). While we use the network from Enting et al. (1995) in this study we realize that concrete suggestions for extensions to the network must start with the current network as a base. In particular, $\delta^{13}C$ observations are now available from the same stations as CO_2.

2.2. Simulated annealing

Simulated annealing is a technique for optimizing highly structured nonlinear functionals of many variables. It proceeds by analogy with the annealing of a crystal lattice in which quasi-steady state conditions while cooling allow the attainment of a minimum energy state. The general procedure for minimizing $f(\boldsymbol{x})$ proceeds as follows: At any time we have some state $\boldsymbol{x}_i$. We perturb this by an amount $\delta\boldsymbol{x}$ proportional to pseudo-temperature, T. If $f(\boldsymbol{x}_i+\delta\boldsymbol{x}) < f(\boldsymbol{x}_i)$ then we accept $\boldsymbol{x}_i+\delta\boldsymbol{x}$ as $\boldsymbol{x}_i+1$. If $f(\boldsymbol{x}_i+\delta\boldsymbol{x}) \geqslant f(\boldsymbol{x}_i)$, then we may still accept the new state based on a Boltzmann probability distribution. We calculate the probability of the new state as

$$\exp\left(-\frac{f(\boldsymbol{x}_i+\delta\boldsymbol{x}) - f(\boldsymbol{x}_i)}{kT}\right), \tag{4}$$

where T is the pseudo-temperature and k is the scaling factor between temperature and energy, i.e., Boltzmann's constant. We accept the new state if the above Boltzmann expression is greater than a pseudo-random number between 0 and 1.

We carry out a series of such trials for a given temperature then reduce the temperature. Reducing the temperature has two effects. It reduces the amplitude of perturbations. Secondly a lower temperature makes the acceptance of an increase in f less likely.

The reasoning for the method is quite simple. In general we accept lower values for the functional most of the time. This alone would get the optimization stuck in any local minimum so the Boltzmann criterion above allows the possibility of escape. As we reduce the temperature the perturbations decrease as does the chance of escape from any potential well. The hope is that by this time we are in the global minimum and will gradually settle towards the bottom. For our problem it is not critical to find the single global minimum corresponding to the very best network. Significant improvement will suffice even if, as is quite possible in this type of problem, the chosen station distribution is quite unlike the true optimum.

There are three parameters one can tune for the simulated annealing algorithm. The first is the cooling rate. In general this is simply a computational trade-off, the slower the cooling the less likely we are to inadvertently freeze the system in a local minimum. The second parameter is the relationship between the temperature and the size of perturbations to the function parameters. The last parameter is the Boltzmann constant for the particular problem. In our case there is little guidance for the choice of these parameters. There are a few cases we can check directly against a complete map of the functional. One of these is the case for adding one extra station considered in the next section. We chose parameters which optimized this case correctly, i.e., which placed the new station in the minimum of the objective function.

For our specific problem, the functional we optimize is a variance estimate for a source or group of sources. The parameters are the station locations. Changing the station locations will change the matrix $\mathbf{A}$ since it varies the places at which the global responses to a given source are sampled. In the more realistic cases considered later it may also vary the data covariance $\mathbf{C}(\boldsymbol{D})$. This will occur if we make some assumption about the relative data quality in various regions.

We make the number of gridpoints in the zonal direction equal to the temperature for each perturbation. Because of the aspect ratio of the GISS tracer transport model grid (24×36) we make meridional perturbations $\frac{2}{3}$ as large. We always choose an initial temperature of 20 so that initial perturbations are of hemispheric scale. This should guarantee that the details of the initial network are unimportant. For each perturbation we multiply the temperature by a random number between -1 and 1 for each of the zonal and meridional increments.

We cool the system in steps of 1 with a minimum temperature of 1 since anything lower would not alter the locations. The usual case was 50 trials at each temperature above 10 and 100 trials at all lower temperatures. Finally the Boltzmann constant for the optimization (a parameter for which we could see little if any physical guidance) was set at 400.

The sources whose estimates we wish to improve are also a matter of choice. We use the set from the inversion of Enting et al. (1995). See Section 3 of that paper for a full description. There are 54 sources in all and we break them into groups as follows.

- 2 to represent the input from fossil fuel and the oxidation of CO.
- 12 to represent the uptake by twelve ocean regions of CO_2. These may also have a $\delta^{13}C$ signature depending on the different solubilities of $^{13}CO_2$ and $^{12}CO_2$ in seawater.
- 12 representing the gross flux between ocean and atmosphere. These fluxes do not affect the CO_2 concentration but do have a $\delta^{13}C$ signature due to the isotopic disequilibrium between ocean and atmosphere.
- 8 representing unbalanced CO_2 uptake by the terrestrial biota. They will include effects of CO_2-enhanced growth, climatic effects on the size of the biota and also fluxes where CO_2 is taken up by the biota and is returned to the atmosphere as CO or its precursors or transported to the ocean by rivers.
- 8 representing the difference between seasonal uptake and release of CO_2 by the terrestrial biota. These sources have zero annual mean. These fluxes have a biotic $\delta^{13}C$ signature.

• 4 representing the source of carbon due to land-use change. These also have their associated $\delta^{13}C$ signature due to the isotopic composition of the terrestrial biota.

• 8 representing the flux of ^{13}C due to the isotopic disequilibrium between the terrestrial biota and atmosphere.

The initial uncertainties we assign to these quantities are those of Enting et al. (1995). These are between 0.5 Gt C yr^{-1} and 1.5 Gt C yr^{-1} on the ocean net fluxes, 1.5 Gt C yr^{-1} on the unbalanced terrestrial uptake terms, large uncertainties proportional to the fluxes on the gross ocean fluxes, and very large uncertainties equal to the fluxes on the seasonal biotic exchange. The land-use flux is assigned a relatively small uncertainty of 0.5 Gt C yr^{-1} for each tropical region and 0.3 Gt C yr^{-1} for the temperate.

We will usually wish to minimize the variance of a group of these sources, such as the total ocean uptake. This can be calculated by summing over the terms in the total covariance matrix, i.e., the variance of a sum is the sum over the whole covariance matrix for that quantity.

The implementation of the optimization is very simple indeed. The simulated annealing code runs as a controlling program. For each trial it assembles a station list in the form required by the synthesis inversion then calls the inversion. Trivial post-processing of the synthesis inversion output returns the required variance to the simulated annealing program. Whenever the process yields a new best estimate, the appropriate station list is recorded. It is common that the best list is not the final one, with the optimum location in phase-space having been visited during the process then jumped away from by the random acceptance stage in the algorithm. The main reason for this is that the minima are quite flat so that optimum positions are only defined to ± 1 gridpoint. The process with the temperature profile discussed above takes about 3 h of computation time on a moderately powerful workstation.

3. Results

As an initial test of the method we consider the optimization of the current network, i.e., shifting the current stations. This is firstly to test the optimization procedure. Secondly, while the new network is not any kind of practical proposition and would not have any climatology established, it is worth knowing the limit of detection available from any given number of stations with given instrumental properties. Simulated annealing should produce a near-optimal network with which real proposed networks can be compared. It may be that while the proposed network is impractical, some of its more desirable properties might be achievable.

Fig. 1 shows the case where all stations are allowed to move. "+" represents the current sites and "*" the optimal locations. We optimize for the estimated error of the global sum of net ocean flux. The error (expressed as a standard deviation) changes from 1.2 Gt C yr^{-1} to 0.7 Gt C yr^{-1}, a

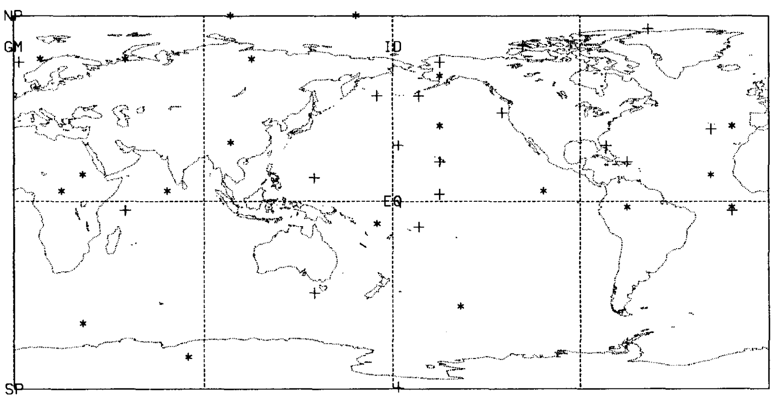

Fig. 1. Map showing current and optimized station network. The network is optimized for the variance in global ocean uptake. The current network is denoted with a "+", the network optimized with simulated annealing with a "*".

considerable improvement in the uncertainty. The most striking fact about the optimal locations is that there are twelve of the twenty stations located over the oceans, about what one would expect based on area. Thus they show no discernible preference for ocean locations. We shall say more about this in a simpler, more systematic study in Section 4.

By maximizing the impact of station data on the inversion procedure, we are likely to have maximized the impact of improvement in data quality. To investigate this we repeat one of the experiments of Enting et al. (1995). Recall that, in the synthesis inversion as it is used here, the station data only contribute information regarding spatial structure, global budget information is principally contained in the global trend terms. To see what effect increasing precision of station data (and hence increased knowledge of spatial distributions) had on the synthesis inversion, Enting et al. tightened the initial estimate of annual mean CO_2 concentration from a standard deviation of 0.3 ppmv to 0.1 ppmv at each station. They found that the estimated error in global net ocean uptake only decreased from 1.16 to 1.10 Gt C yr^{-1}. This suggested that estimates of global fluxes were derived mainly from global-scale information and that the spatial structure was adding little information globally. Our best network allows us to test whether or not this is a general property of the system or a function of the current network. It is, of course, a best case test. When we repeat the experiment with the new network, the estimated error drops from 0.73 to 0.50 Gt C yr^{-1}, a larger improvement in both relative and absolute terms. This suggests that the limited role played by the spatial structure of CO_2 in constraining global sources in the work of Enting et al. (1993, 1995) was due partly to the limitations of the current observing network.

The other major investigation presented here involves the addition of a few more stations to the current network. The general procedure is to clone one existing station one or more times then allow these new stations to shift to optimize the desired estimate. For the simplest case, however, we have a more direct approach available. The relatively coarse resolution of the tracer transport model used by this study means it is inexpensive to do a global test with one station. This process doesn't involve simulated annealing at all, we simply carry out the synthesis inversion for the current network with an additional station successively placed at every model gridpoint. This is very useful as a test of the optimization procedure (we now can identify global minima for the one extra station case) as well as providing a complete description of the system for this simple case.

Fig. 2 shows the complete portrait of the one extra station case for the estimated error in global net ocean flux. The station added has the measurement properties of the station at Cape Grim,

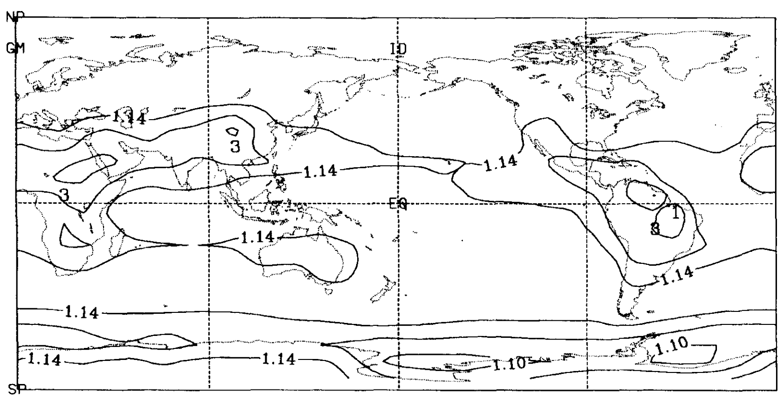

Fig. 2. Standard deviation (Gt C yr^{-1}) of global ocean uptake as a function of the position of one additional station to the current network. The station measures the same species with the same uncertainties as the existing station at Cape Grim, Tasmania. The contouring is curtailed at 80° to avoid problems arising from multi-valued responses at the poles. The contour interval is 0.02. The map also shows the positions of one and three additional stations calculated (with simulated annealing) to minimize the variance of net ocean uptake.

Tasmania. Most importantly it measures both CO_2 and $\delta^{13}C$. Recall that the current network returns 1.16 Gt C yr^{-1} for this estimate. This must be an upper bound for the global map since any new station will add some information no matter how badly it is placed.

As one might expect, with only one new station, the improvement is not large, with a global minimum of 1.08 Gt C yr^{-1}. What was quite unexpected is the position of the various minima on this map. There are three minima. The deepest is over northern Brazil, the second and much more diffuse one is over tropical Africa, the Middle East and the mid and low latitudes of Asia. The third, really a double minimum, is off the Antarctic coast.

Given that we are optimizing for knowledge of an oceanic flux, the suggestion that the best location on the globe for a new station is the Amazon basin seems mystifying. The explanation lies, as it must, in the behaviour of the synthesis inversion procedure. The global net ocean flux is one of few terms in a closed global budget which, for CO_2 at least, can be written schematically

$$\text{fossil} + \text{land-use} + \text{ocean} + \text{terrestrial biota} = \text{trend} \qquad (5)$$

We can regard this sum as a very simple Bayesian inversion problem. The prior estimates for each global flux are formed from the prior estimates for each regional component and the requirement to match the CO_2 structure. The new station acts to better constrain a regional source and hence reduce the prior uncertainty for that global flux. In a Bayesian inversion this will be reflected in reduced uncertainties for all the global terms.

Table 1 lists the initial and final uncertainties for some groups of sources for a synthesis inversion with the current network. The chosen sources are those which contribute to the CO_2 budget.

Table 1. *Initial and final uncertainties for various groups of sources with the current station network*

Source group	Prior s.d.	Final s.d.
fossil	0.36	0.35
biotic uptake	4.24	1.39
land-use	0.92	0.88
ocean uptake	3.43	1.16

The sources shown are those that contribute to the CO_2 budget. The fossil group also contains the source due to CO oxidation.

Partly by design of the current network and partly by the initial uncertainties we give to various source estimates, the most uncertain term is currently terrestrial biotic uptake. It is reasonable therefore that this terrestrial flux is amenable to the greatest improvement. Locally it is the tropical American source which is the most uncertain component of terrestrial biotic uptake. The large minimum over Asia and Africa is explained by the same argument. To show that direct constraints can be of some use, the other minimum near Antarctica is a direct constraint on ocean sources themselves.

To demonstrate further the links between various terms in global budgets, we can calculate the same global map for the estimated error in terrestrial biotic uptake. This is shown in Fig. 3. The main point to notice is the similarity between Figs. 2 and 3. In particular, the two major minima for the oceanic case correspond with minima for the terrestrial case. This example illustrates a general property of the synthesis inversion method. It indicates that estimates of many quantities are greatly improved by consideration of apparently unrelated quantities, the overall budget providing the nexus.

We have tested how good a job the simulated annealing is doing of finding optimal states. Fig. 2 shows the location of one and three extra stations added to the current network with their positions optimized by simulated annealing. Each trial is optimized for net ocean uptake. We can see that the stations generally successively fill in the minima on the global map, indicating that the optimization is, indeed, finding the relevant minima. For five new stations (not shown) two of the minima have two stations nearby. The figure suggests that the new stations don't interact strongly and that the global maps are a good predictor for the locations of more than 1 new station. If we added enough new stations, all the clear minima in Fig. 3 would be taken up. Repeating the experiment of adding one new station to this enhanced network would then produce quite a different map. Also, it is only the low resolution of the GISS model, ($8 \times 10°$) which makes such a comprehensive test computationally feasible. Fig. 4 shows the improvement offered by each set of stations. We see that by the time 5 new

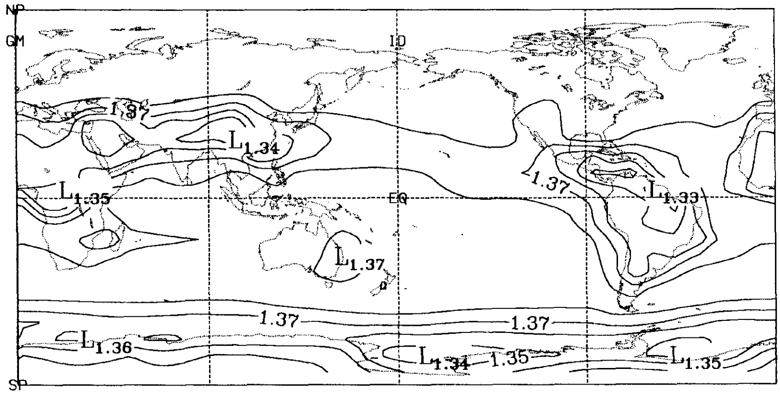

Fig. 3. Standard deviation (in Gt C yr^{-1}) of global terrestrial biotic uptake as a function of the position of one additional station to the current network. The additional station has the properties of the station at Cape Grim. The contour interval is 0.01.

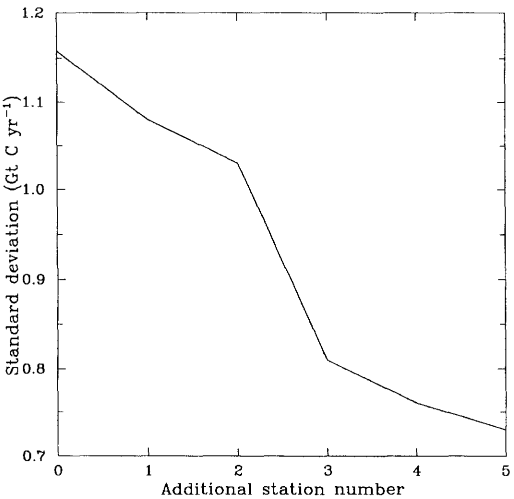

Fig. 4. Graph of the standard deviation (in Gt C yr^{-1}) of global ocean uptake for 1–5 additional stations added to the current network. In all cases the standard deviation shown is that from the best network found with simulated annealing. The current network is equivalent to adding 0 extra stations.

stations have been added, the case is about as good as the re-siting of the entire current network. This last result needs to be treated with caution as the number of sites, 25, is close to the number of poorly-known CO_2 source components in the inversion. This may mean the results are sensitive to the assumptions about source discretisation. This requires further study.

The above calculations suggest that the combination of synthesis inversion and simulated annealing can yield significant improvement in source estimates. The examples are obviously only illustrative and it is doubtful that any real decisions on new sites should be based on them. To move a little more in this direction we have attempted to take some aspects of site quality into account.

One of the reasons why the optimization has picked out continental locations and continental sources is the lack of observations there at present. This in turn is based on well-founded concerns about the representativeness of continental observations, both spatially (there are strong local sources) and temporally (the variability is much higher so it is harder to produce a stable record). In an attempt to deal with the second of these we have weighted potential station locations by some measure of quality. The quality measure we use is the standard deviation of the daily data around the monthly mean. The argument is that a flask sampling site will only take some points on a time series. The more highly variable this time series, the poorer the estimate of the mean value for a given month. This approach is similar to the measure adopted by Enting et al. (1993, 1995) when they based their estimates of observational uncertainty on the results of the study of sampling statistics by Tans et al. (1990). To find the submonthly variability we use daily concentrations from a tracer model experiment forced by the terrestrial biotic fluxes from Fung et al. (1987). There is no reason why the tracer model used for the data quality calculation should be at all related to the model which generated the response functions we use for the synthesis inversion; we only want some estimate of quality, the more realistic the better. We actually used the model of Law et al. (1992). The annually averaged monthly standard deviation of concentration from the Fung et al. terrestrial fluxes is shown in Fig. 5. The results are largely what one would expect, with high values in regions where the fluxes have large annual cycles. This will particularly disadvantage high northern latitude sites.

The relationship between the variability at a particular site and the uncertainty of monthly mean measurements at that site is arbitrary, presumably depending on sampling protocols, data selection and the like. Provided we assume the relationship is constant, Fig. 5 does show how much better or worse one station is than another. In order to calculate the real station uncertainty, we assume that the error contributed to each datum by this new sampling uncertainty is the same as the original uncertainty at Cape Grim when the new station is placed at Cape Grim. Thus the additional error E is given by

$$E = E_{CG} \times \frac{\sigma}{\sigma_{CG}},$$

where the σ values are the standard deviations from Fig. 5 and the subscript CG refers to Cape Grim. This is absolutely arbitrary of course but it was really the structure of the data quality penalty we were interested in. Note that this is a case where the actual form of the matrix $\mathbf{C}(\boldsymbol{D})$ changes with the station locations not just $\mathbf{A}$.

The map of uncertainty in net ocean uptake with the position of one new station with this extra penalty on certain sites is shown in Fig. 6. This should be compared with Fig. 2 which is the same map without the station quality condition. The main impact is the downgrading of the available region over Asia and Africa. Here, the impact of the seasonal cycle increases the sampling error enough to defeat the site advantages. The most striking feature of Fig. 2, the region over the Amazon, remains however. This calculation is just one step along the way to a concrete suggestion for new measurement sites. For example variability of sources and, particularly, boundary layer behaviour (Denning et al., 1995) also act against our flask sampling measurements. As a limiting case, if we penalized continental measurements enough, we would be forced back to the mainly marine measurements of the current network.

In the cases considered so far, we have dealt with optimizations involving summed groups of sources. We have also discovered a propensity of the synthesis inversion to couple various sources so that improvements in one lead to improvements in another via covariance effects. It is worthwhile asking now whether the optimal networks for sums of sources are also good for estimates of individual source components. This is not necessarily true; sets of locations which best constrain

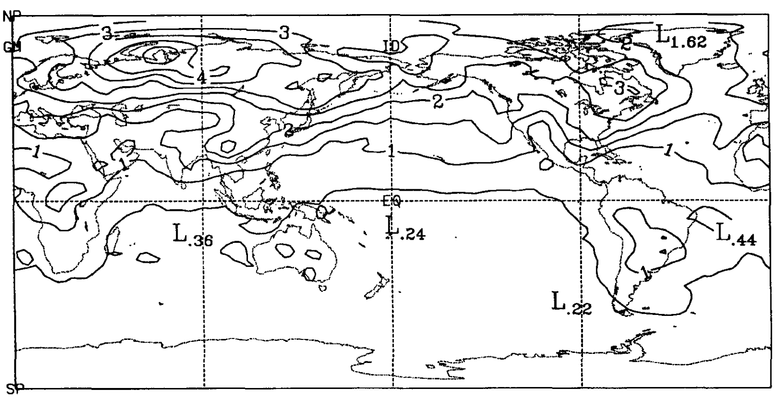

Fig. 5. Map of the annually averaged standard deviation (in ppmv) for each month of the CO_2 concentration resulting from a representation of the terrestrial biotic source. The values were calculated by first calculating the standard deviation for each month of the daily values then averaging this annually. The source used was that of Fung et al. (1987) and the tracer model that of Law et al. (1992). The contour interval is 0.5.

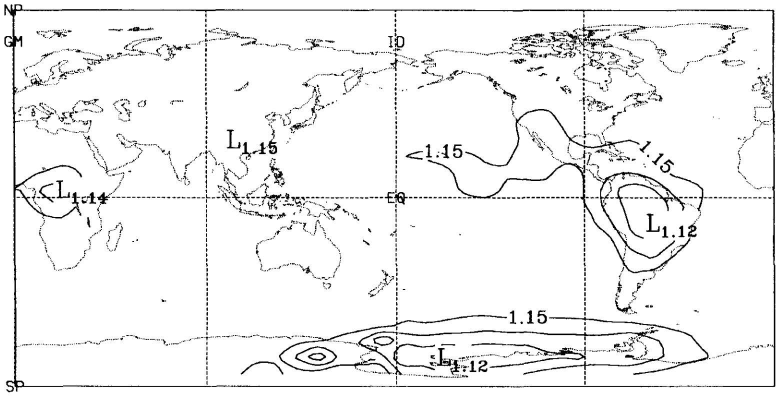

Fig. 6. Standard deviation (Gt C yr^{-1}) of global ocean uptake as a function of the position of one additional station to the current network. The submonthly variability resulting from the simulated response to the seasonal biospheric source is allowed to impact the uncertainty of both seasonal and annual mean data for the additional station. See text for details. The contouring is curtailed at 80° to avoid problems arising from multi-valued responses at the poles. The contour interval is 0.01.

global flux estimates may maximize correlation among a group of sources. This could reduce the variance of the sum but not those of the individual components.

To investigate this we repeat the "one extra station" calculation but this time minimizing the trace of the covariance submatrix for the ocean uptake rather than the sum over all its elements. This means we are minimizing the sum of the variances of the individual source components without considering the effect covariances among them may have on the variance of the summed flux. This is some crude measure of the average constraint on individual sources in the group rather than on the sum. We do not include the station quality condition. The resulting map is shown in Fig. 7. This result looks radically different from any of the three comparable diagrams earlier. Figs. 2 and 3, in particular, look quite similar although they are optimizing different flux estimates. Also note that the numbers are much higher than for the sum case. The initial value for

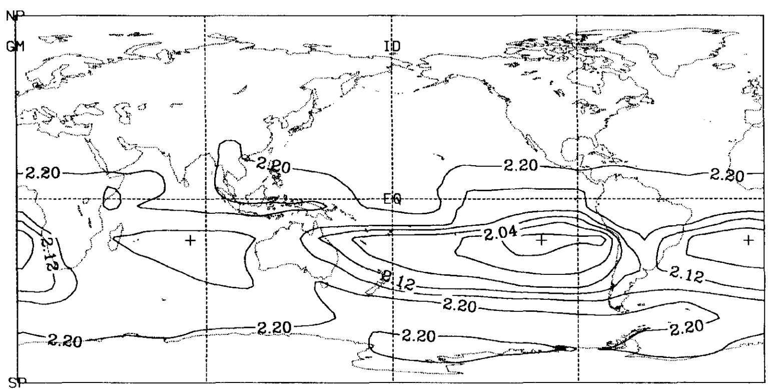

Fig. 7. Map of the square root of the sum of the variance of individual oceanic source components (Gt C yr^{-1}) as a function of the position of one extra station added to the current network. The station measures the same species with the same uncertainties as the station at Cape Grim. This is intended as a measure of the average constraint on each source component separately. The positions of three additional stations optimized for the contoured quantity are denoted by a "+". The contouring is curtailed at 80° to avoid problems arising from multi-valued responses at the poles. The contour interval is 0.04.

the standard deviation of the ocean group is listed in Table 1 as 3.43 Gt C yr^{-1}. Since we assume no correlation among the prior estimates this initial error is the same for the sum and the trace case. The error of the sum, however, is reduced by the synthesis inversion much more than the trace indicating that the inversion technique introduces strong correlations between estimated sources of the same type at different locations.

The second thing to notice about Fig. 7 is the placement of the minima. The three strongest correspond to the southern hemisphere subtropical gyres. These locations are also selected by an optimization of three additional stations using this summed variance as an objective function. These locations are shown by the "+" in Fig. 7.

The reasons for the preference for these locations can again be seen in the behaviour of the synthesis inversion. These sources have the same initial error as for all the other net ocean sources. However, the lack of convenient sites means the current network is unable to constrain them very much. Thus these three sources have the highest estimated uncertainty using an inversion with the current network.

The results of the above case have little practical outcome since a measurement programme for such locations would be difficult. It does however point out an important and general point. We should be careful in this and any other experimental design problem to define precisely what we wish to constrain. The two quantities, here called total and global variance, appear quite similar but demand radically different measurement strategies. The difference highlights a dilemma between the requirement of better understanding particular processes (which would suggest knowledge of individual sources hence the total variance) versus the desire to constrain the global carbon cycle (for which we require knowledge of global fluxes).

4. Conclusions

In this paper, we have demonstrated a combination of an inversion technique for obtaining flux estimates from concentration data and an optimization technique for designing networks for best constraining those sources. The conclusions can be summarized as follows.

(1) A combination of synthesis inversion and simulated annealing can suggest considerable improvements in observational network design for constraining terms in the carbon budget. Application of the technique will require refinement of the inversion formalism with particular attention to the "error model".

(2) We can understand much of the behaviour of such an optimization by looking at the impact of adding one station at all available gridpoints.

(3) The ability of synthesis inversion to couple apparently disparate information means that highly indirect constraints might be preferable. The example shown is that the best single site for improving the constraint of global ocean uptake is the Amazon basin.

(4) We must be quite precise in specifying the requirements of an observational network in such problems since apparently similar quantities can demand different strategies.

5. Acknowledgements

The authors wish to thank Dr. Rachel Law and Martin Dix for valuable comments on the manuscript. This study was carried out with the support of the Australian Government through its Cooperative Research Centres Programme. The CSIRO transport modelling project was funded by the State Electricity Commission of Victoria. The GISS tracer transport model was kindly provided by Prof. Inez Fung. We wish to thank M. Hardt and F. Scherbaum for supplying a preprint prior to publication.

REFERENCES

Bakwin, P. S., Tans, P. P., Zhao, C., Ussler, W. and Quesnell, E. 1995. Measurements of carbon dioxide on a very tall tower. *Tellus* **47B**, 535–549.

Ciais, P., Francey, R. J., Tans, P. P., White, J. W. C. and Trolier, M. 1995. An analytical error estimate for the ocean and land uptake of CO_2 using $\delta^{13}C$ observations in the atmosphere. *NOAA Technical Memorandum ERLCMDL*-**8**, 13 pp.

Conway, T. J., Tans, P. P., Waterman, L. S., Thoning, K. W., Kitzis, D. R., Masarie, K. A. and Zhang, N. 1994. 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.

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**, 240–243.

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, M. Heimann ed. Springer-Verlag, pp. 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., 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.

Francey, R., Steele L. P., Langenfelds, R., Lucarelli, M., Allison, C., Beardsmore, D., Coram, S., Derek, N., de Silva, F., Etheridge, D., Fraser, P., Henry, R., Turner, B., Welch, E., Spencer, D., and Cooper, L. 1996. Global Atmospheric Sampling Laboratory (GASLAB): Supporting and Extending the Cape Grim Trace Gas Program. *Baseline atmospheric program (Australia) 1993*, R. J. Francey, A. L. Dick and N. A. Derek, eds. Bureau of Meteorology and CSIRO, 8–29.

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.

Hardt, M. and Scherbaum, F. 1992. Optimizing the station distributions of seismic networks for aftershock recordings by simulated annealing. EOS Supplement. *Transactions of the Fall Meeting, 1992*. AGU, 351.

Hardt, M. and Scherbaum, F. 1994. The design of optimum networks for aftershock recordings. *Geophys. J. Int.* **117**, 716–726.

Keeling, R. F., Najjar, R. P., Bender M. L. and Tans, P. P. 1993. What atmospheric oxygen measurements can tell us about the global carbon cycle. *Glob. Biogeochem. Cyc.* **7**, 37–68.

Keeling, C. D., Whorf, T. P., Wahlen, M. and van der Plicht, J. 1995. Interannual extremes in the rate of rise of atmospheric carbon dioxide since 1980. *Nature* **375**, 666–670.

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

Schimel, D., Enting, I., Heimann, M., Wigley, T., Raynaud, D., Alves, D. and Siegenthaler, U. 1995. CO_2 and the carbon cycle. *Climate Change* 1994: radiative forcing of climate change and an evaluation of the IPCC IS92 emission scenarios, J. Houghton, L. M. Filho, J. Bruce, H. Lee, B. Callander, E. Haites, N. Harris and K. Maskell, eds. Cambridge University Press, 35–71.

Takahashi, T., Olafsson, J., Goddard, J. G., Chipman, D. W. and Sutherland, S. C. 1993: Seasonal variations of CO_2 and nutrients in the high-latitude surface oceans: a comparative study. *Glob. Biogeochem. Cyc.* **7**, 843–878.

Tans, P. P., Thoning, K. W., Elliott, W. P. and Conway, T. J. 1990. Error estimates of background atmospheric CO_2 patterns from weekly flask samples. *J. Geophys. Res.* **95**, 14063–14070.

Tarantola, A. 1987: *Inverse problem theory: methods for data fitting and parameter estimation.* Elsevier, Amsterdam.

Wofsy, S. C., Goulden, M. L., Munger, J. W., Fan, S. M., Bakwin, P. S., Daube, B. C., Bassow, S. L. and Bazzaz, F. A. 1993. Net exchange of CO_2 in a mid-latitude forest. *Science* **260**, 1314–1317.