GLOBAL BIOGEOCHEMICAL CYCLES, VOL. 14, NO. 1, PAGES 407-428, MARCH 2000

Optimal sampling of the atmosphere for purpose of inverse modeling: A model study

Manuel Gloor, [1]Song-Miao Fan, Stephen Pacala, and Jorge Sarmiento
Carbon Modeling Consortium, Atmospheric and Oceanic Sciences Program and Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey

Abstract. The 66 stations of the GLOBALVIEW-CO2 sampling network (GLOBALVIEW-CO2: Cooperative Atmospheric Data Integration Project - Carbon Dioxide, (1997)) are located primarily remotely from continents where signals of fossil fuel consumption and biospheric exchange are diluted. It is thus not surprising that inversion studies are able to estimate terrestrial sources and sinks only to a very limited extent. The poor constraint on terrestrial fluxes propagates to the oceans and strongly limits estimates of oceanic fluxes as well, at least if no use is made of other information such as isotopic ratios. We analyze here the resolving power of the GLOBALVIEW-CO2 network, compare the efficiency of different measurement strategies, and determine optimal extensions to the present network. We find the following: (1) GLOBALVIEW-CO2 is well suited to characterize the meridional distribution of sources and sinks but is poorly suited to separate terrestrial from oceanic sinks at the same latitude. The most poorly constrained regions are South America, Africa, and southern hemispheric oceans. (2) To improve the network, observing stations need to be positioned on the continents near to the largest biospheric signals despite the large diurnal and seasonal fluctuations associated with biological activity and the dynamics of the PBL. The mixing in the atmosphere is too strong to allow positioning of stations remote from large fluxes. Our optimization results prove to be fairly insensitive to the details of model transport and the inversion model with the addition of $\sim$ 10 optimally positioned stations. (3) The best measurement strategy among surface observations, N-S airplane transects, and vertical profiles proves to be vertical profiles. (4) Approximately 12 optimally positioned vertical profiles or 30 surface stations in addition to GLOBALVIEW-CO2 would reduce estimate uncertainties caused by insufficient data coverage from $\sim$ 1 Pg C yr^{-1} per region to $\sim$ 0.2 Pg C yr^{-1} per region.

1. Introduction

Atmospheric inverse modeling provides one of the most powerful tools to estimate the uptake and release of CO_2 and other gases by the planet's surface. Atmospheric transport models can be used to infer the pattern of atmospheric concentration that corresponds to a distribution of surface fluxes. In an inverse problem, one simply inverts this relationship to infer the surface fluxes most compatible with concentration data. Such estimates are critical for verification of fossil fuel emissions and assessment of ecosystem carbon storage and ocean uptake of CO_2.

Previous studies that have attempted to estimate the spatial distribution (both latitude and longitude) of CO_2 sources and sinks with an inversion of atmospheric transport faced large technical problems [*Keeling et al.*, 1989; *Tans et al.*, 1990]. *Keeling et al.* [1989] and *Tans et al.* [1990] were both forced to prescribe fluxes from most regions to permit the estimation of fluxes from the remaining ones. *Keeling et al.* [1989] estimated two parameters, the strength of the North Atlantic uptake and equatorial outgassing, whereas *Tans et al.* [1990] estimated the strength of the Northern Hemisphere terrestrial and Southern Hemisphere oceanic fluxes. The results of these studies disagree substantially; one concluded that the Northern Hemisphere sink is oceanic; the other concluded that it is terrestrial. Similar problems have been encountered more recently by *Fan et al.* [1998] where systematic and estimation errors prevented estimates of fluxes from more than four regions globally.

[1]Now at Max-Planck-Institut fur Biogeochemie, Jena, Germany.

Paper number 1999GB900052.
0886-6236/00/1999GB900052$12.00

This summary points to several concerns about inverse methods. First, the spatiotemporal resolution of the method is currently very coarse (i.e., 30° of latitude with no zonal resolution or northern hemispheric continents with no meridional resolution). What are the fundamental limitations on resolution? How much more data would we need to improve it and by how much? The amount of data required to achieve any given level of resolution depends critically on the natural level of mixing in the atmosphere. If mixing is too fast, then the footprint of each region will be very weak and estimates will be very uncertain (amplification of high-frequency data variability). Second, published studies employ different representations of atmospheric transport. Errors in atmospheric transport are particularly troublesome because they are likely to create systematic biases that cannot be overcome simply by adding data. Systematic errors may also be introduced through the assumptions about the spatiotemporal structure of surface fluxes assumed in the inversion.

The issues of data coverage and systematic error were investigated in a modeling study by *Gloor et al.* [1999]. They employed three different atmospheric general circulation models (AGCMs) (one based on analyzed winds), several formulations for terrestrial fluxes (e.g., based on satellite imagery of vegetation or spatially uniform), two different schemes for oceanic fluxes (*Takahashi et al.* [1997] and uniform), and a wide range of resolutions for the inversions. They explicitly separated errors caused by insufficient data coverage, given the natural level of information loss by mixing, from systematic errors in transport or surface fluxes.

Their results suggest that the first source of error, insufficient data coverage, is indeed considerably reducable with additional observing stations. By comparing randomly constructed networks with different numbers of stations, *Gloor et al.* [1999] showed that $\sim$ 150 surface stations were necessary to achieve a precision of $\sim$ 0.2 Pg C yr^{-1} per region for 10-20 regions globally. The benefit of adding stations showed a pattern of diminishing returns. Estimates for 10-20 regions improved only marginally if the networks were increased in size beyond 150 stations, whereas estimates deteriorated quickly if the networks were reduced in size. For 150 observation stations less than half of the uncertainty was due to information loss caused by atmospheric mixing.

In this paper, we build on the general study of *Gloor et al.* [1999] to diagnose weaknesses of the current GLOBALVIEW-CO2 network (GLOBALVIEW-CO2: Cooperative Atmospheric Data Integration Project - Carbon Dioxide (1997)) and to derive optimal extensions of the network. Similar to the previous study, this study employs two AGCMs, two models of terrestrial fluxes, two models of oceanic fluxes, and estimates of emissions from fossil fuel burning. GLOBALVIEW-CO2 is the result of an international effort to unite measurements from different groups and encompasses observations from 66 stations run by 16 different institutions.

A potential problem with the terrestrial models used in this study is that they lack a diurnal cycle and thus largely lack diurnal fluctuations present at the surface. We attempt to ameliorate this problem in two ways. First in some of the work we assume ten-fold greater high-frequency data variability at continental stations than at marine stations and obtain qualitatively similar conclusions despite this change. We show that this ten-fold increase in variability is larger than the variability induced by a diurnal cycle. Second, one of the conclusions of the study is that vertical profiles are the best way to extend the network. This conclusion would be strengthened by the presence of a diurnal cycle because the column integral of concentrations from a vertical profile does not depend on the diurnal cycle. Such a sampling strategy has in fact been proposed by *Tans* [1993].

The characteristics of an optimal sampling network for atmospheric CO_2 depend critically on the signal-to-noise ratio associated with the measurements. On the one hand, measurements over the oceans exhibit comparably little variability, either day-to-day or with altitude because oceanic stations are generally remote from large fluxes; however, the signals associated with fossil fuel emissions and biospheric exchange are strongly diluted and hence small. On the other hand, measurements over the land exhibit large signals because of their close proximity to large sources and sinks (fossil fuel and biosphere) but comparatively large variance. An optimal network and measurement strategy for inverse estimates of surface fluxes must account for amplification of errors associated with the dilution of signals and the magnitude of the noise.

To determine error amplification for a network, we use a simple inversion scheme that solves for 7 continental and 10 oceanic regions and the analytical expression for this quantity in a linear inversion problem. We do not consider in this paper cases with a larger or a smaller number of regions. We model high-frequency variability in the data (the noise) in a variety of ways that bracket the observed patterns in existing data. Optimization of the network is accomplished using a simulated annealing procedure [*Kirkpatrick et al.*, 1983].

The paper is structured as follows. In a first technical part we explain the inversion scheme and the analytical expression for amplification of high-frequency data variability as well as the optimization routine. We then examine the GLOBALVIEW-CO2 network, discuss the properties of optimal networks in general, and identify optimal extensions of GLOBALVIEW-CO2 for various measurement strategies.

2. Methods

2.1. Inversion Scheme

The inversion method used in this paper relies on the linearity of the tracer transport equation which allows one to decompose an observed mixing ratio distribution into the contributions from sources and sinks in different regions. Consider a flux of CO_2 from a region r, at rate ϕ_r (Pg C yr^{-1}) and suppose there are R such regions. After a few years the mixing ratio distribution increases at the same rate throughout the entire atmosphere with a spatial pattern characteristic for a flux from region r. To reveal the spatial pattern of the annual mean mixing ratio distributions independently from the specific model year, we refer the mixing ratio distributions to an arbitrary reference station (here the South-Pole): $\Delta\chi_r(\vec{x}) \equiv \chi_r(\vec{x}) - \chi_r(\vec{x}_{\text{ref}})$, ($\vec{x}$ spatial location). These referenced annual mean mixing ratio distributions are then sampled at the observation stations and arranged in column vectors: $\Delta\vec{\chi}_r$. To obtain flux estimates we minimize the sum of squares of $|\Delta\vec{\chi}^{\text{obs}} - \sum_{r=1}^{R} \lambda_r \Delta\vec{\chi}_r|$ where $\vec{\chi}^{\text{obs}}$ is the vector of observed mixing ratios at the observation stations and the λ_r is a multiplier that gives the flux from region r in units of ϕ_r. The estimate of the flux from region r is then $\lambda_r\phi_r$.

To set up such an inversion, a choice has to be made as to which flux patterns to use for the forward predictions (for calculating $\Delta\vec{\chi}$). Here we use two sets: (1) spatially and temporally structureless (uniform) fluxes over a region and (2) fluxes with the same spatiotemporal distribution as net primary productivity (NPP) estimated from satellite data for continental regions (from the Carnegie-Ames-Stanford approach (CASA) model of *Potter et al.* [1993] and air-sea partial pressure difference measurements for oceanic regions (from *Takahashi et al.* [1997]).

The inversion based on the spatially uniform flux set permits us to isolate the role played by atmospheric transport and mixing in the selection of optimal stations from other factors, such as the spatiotemporal pattern of the high-frequency data variability and the spatiotemporal pattern of fluxes.

The second set of fluxes uses the same flux patterns as one of the sets used by *Fan et al.* [1998], which builds on existing knowledge of fluxes. In particular, they subtract the comparatively well-known fossil fuel signal from the observations before the inversion and solve for oceanic and terrestrial biospheric fluxes. They assume proportionality between terrestrial net ecosystem productivity (NEP) and NPP (CASA) and use the spatial pattern of oceanic fluxes identical with the spatial pattern of measured air-sea partial pressure differences [*Takahashi et al.*, 1997]. For our main optimizations we use their spatial flux patterns of CASA NPP for continental regions and partial pressure differences of *Takahashi et al.* [1997] for oceanic regions. We hence assume that in practice the simulated mixing ratio signal from fossil fuel emissions will be presubtracted from observations before an inversion. Note that this assumption does not imply the exclusion of the high-frequency data variability caused by fossil fuel emissions from our models of high-frequency data variability.

The inverse calculation is simplified if one first assembles the mixing ratio distributions in a matrix $\mathbf{A} = \{\Delta\vec{\chi}_1, ..., \Delta\vec{\chi}_R\}$, which maps the size of the fluxes $\vec{\lambda}$ to the expected mixing ratio distribution $\Delta\vec{\chi} = \mathbf{A}\vec{\lambda}$. The solution of the minimization problem is then given by the product of the "pseudo-inverse" of $\mathbf{A}$ with the vector of observations. The pseudo-inverse of a matrix with more rows than columns may be determined from its singular value decomposition $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ as $\mathbf{A}^{-1} = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^T$, where the columns of $\mathbf{U}$ as well as $\mathbf{V}$ are orthogonal and the matrix $\mathbf{\Sigma}$ is diagonal. An advantage of the singular value decomposition (SVD) is that it allows one to determine which linear combination of regions is most robustly determined by the data. For example, the GLOBALVIEW-CO2 network turns out to be well suited for meridional resolution of sinks and sources and poorly suited for separating South America and Africa. Specifically, the linear combinations of regions are given by the columns of $\mathbf{V}$ because these are the principal axes of the error ellipsoid of the estimates, if residual errors (high-frequency data variability) are independent and identically distributed Gaussian random variables, and the standard deviation (STD) of the estimate of the ith such combination of regions is the inverse of the ith diagonal element of $\mathbf{\Sigma}$, the matrix of singular values [e.g., *Press et al.*, 1992].

The two AGCMs used in this study are from GFDL/-NOAA (Geophysical Fluid Dynamics Laboratory / National Oceanic and Atmospheric Administration): SKYHI and GCTM (Global Chemical Transport Model). Detailed discussions of the properties of the two models are given, for example by *Levy II and Moxim* [1989] for GCTM and *Hamilton et al.* [1995] for SKYHI. To set up the inversion scheme, we partitioned the Earth's surface into 17 regions, 7 continental and 10 oceanic (Figure 1). This selection was guided by the position of oceans and continents and the main ocean circulation features and large-scale terrestrial climate regions. It is important to keep in mind throughout this paper that the results presented here are based on this specific spatial resolution of the globe.

To understand our results for optimal networks, it is useful to consider an explicit example of a mixing ratio distribution that results from a regional CO_2 flux. Consider the annual mean $\Delta\chi_r(\vec{r})$ that results from monthly varying North American NPP as modeled in CASA

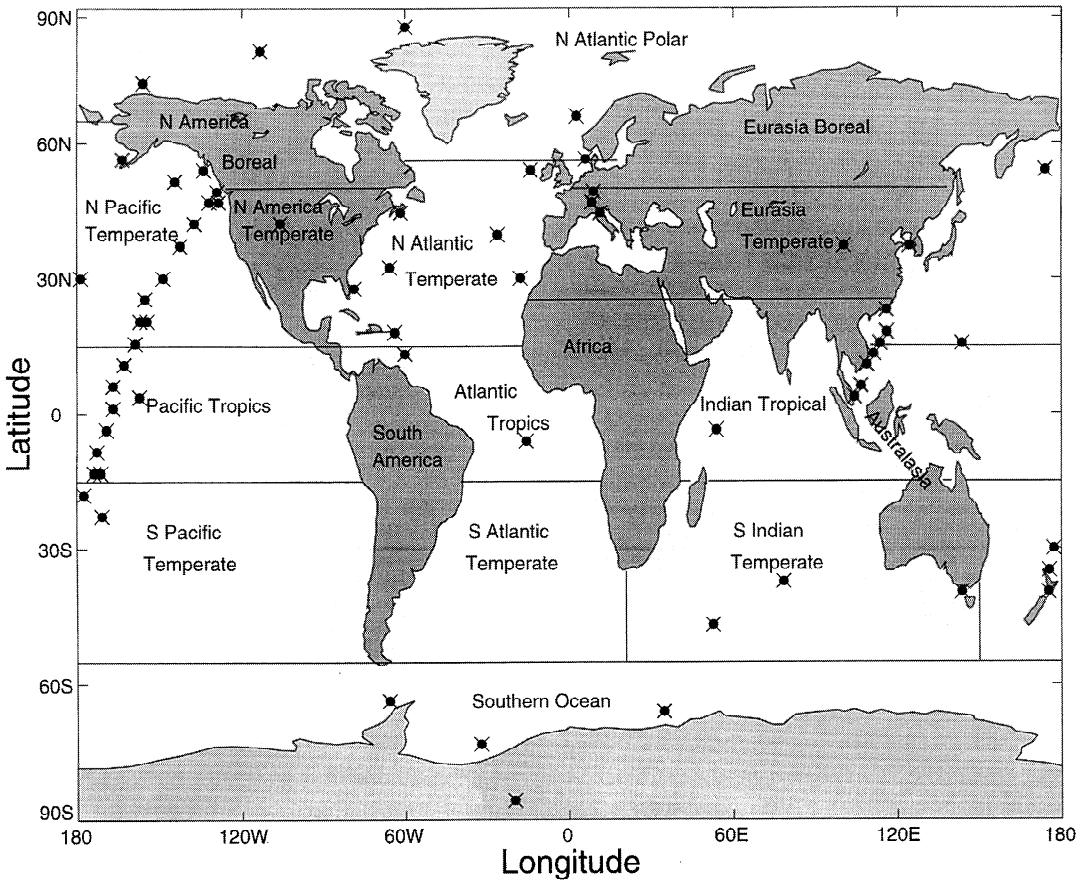

Figure 1. Partitioning of the Earth's surface in flux regions and observation locations of the GLOBALVIEW-CO2 network.

(Figure 2, *Potter et al.* [1993]). This mixing ratio distribution reflects the magnitude of various atmospheric mixing timescales: (1) the highest concentrations in the planetary boundary layer (PBL) of the region are determined by the PBL ventilation rate (approximately days), (2) zonally well-mixed near surface concentrations are determined by the time scale for zonal mixing in the free troposphere ($\sim$ 1 month), and (3) decreasing concentration with increasing meridional distance from the flux region is determined by the interhemispheric exchange time ($\sim$ 1 year). Typically, the largest signals within a flux region differ from the rest of the same zonal band by $\sim 1-2$ ppm Pg C^{-1} yr^{-1} and from the farthest removed latitudes by $\sim$ 1-3 ppm Pg C^{-1} yr^{-1}. The largest local within-zone signals occur in high latitudes, and the smallest ones in the equatorial continental regions where PBL ventilation caused by strong convection is largest. For example, Figure 3 shows the annual average mixing ratios at the surface resulting from 1 Pg C yr^{-1} emitted from each terrestrial region with within-region spatio-temporal distribution given by NPP from the CASA model (using GCTM). Note that the signal is larger in the Northern Hemisphere because strong convection dilutes the signals from South America, Africa, and Australasia. These general characteristics of mixing ratio distributions let one expect that the meridional resolution of sources and sinks with an inversion is much easier than the resolution of sources and sinks within a zonal band and that equatorial sources and sinks are particularly difficult to resolve.

Figure 4 similarly shows the spatial pattern of oceanic fluxes estimated from partial pressure differences by *Takahashi et al.* [1997] that we used for the forward predictions. The structure of the partial pressure distribution in the oceans coincides with major upwelling and downwelling regions as well as regions with strong convection. This spatial pattern like the pattern for fluxes of the terrestrial biosphere will be helpful to interpret the results of network optimizations.

2.2. High-Frequency Data Variability and Estimate Uncertainty (Error Amplification)

The variance-covariance matrix for the vector of estimates for a linear inversion is given by $\mathbf{C}_\varphi = \mathbf{A}^{-1}\mathbf{C}_{\Delta\chi}(\mathbf{A}^{-1})^T$ where $\mathbf{C}_{\Delta\chi}$ is the assumed variance-covariance matrix of high-frequency data variability. As high frequencies, we consider fluctuations with periods of the order of the synoptic timescale and

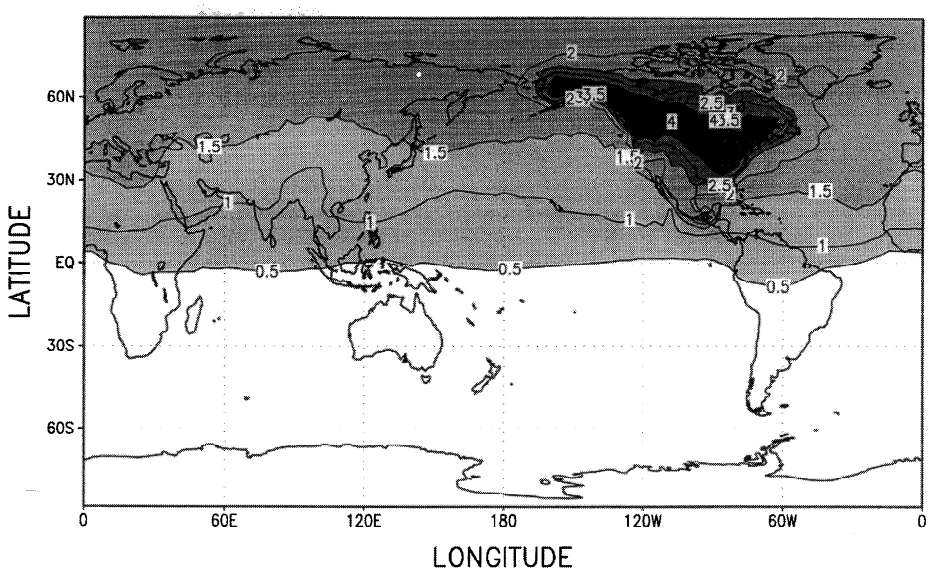

Figure 2. Annual mean mixing ratio distribution (ppm) in the atmosphere simulated with GCTM that results from net primary productivity fluxes in North America as estimated from satellite data and the CASA biosphere model [*Potter et al.*, 1993].

less. Note that this formula contains two parts, one summarizing high-frequency data variability ($\mathbf{C}_{\Delta\chi}$) and the other the amplification of this high-frequency data variability (the premultiplication and postmultiplication of the data variance-covariance matrix). A measure for the error amplification is: $\mathbf{ErrAmp} \equiv \mathbf{A}^{-1}(\mathbf{A}^{-1})^T$ with dimensions $[(\text{Pg C yr}^{-1}\ \text{region}^{-1}\ \text{ppm}^{-1})^2]$ [*Menke*, 1989]. The magnitude of the entries of $\mathbf{A}$ is a reflection of the magnitude of the information loss caused by mixing in the atmosphere, given a specific observation network and partitioning of the Earth's surface into regions. A natural average measure of error amplification is the mean of the diagonal elements of the error amplification matrix $\overline{\mathbf{ErrAmp}} \equiv 1/R \sum_{r=1}^{R} (\mathbf{A}^{-1}(\mathbf{A}^{-1})^T)_{rr}$.

What processes contribute to the high-frequency variability of measurements (to $\mathbf{C}_{\Delta\chi}$) ? What are their magnitudes and spatial patterns? The high-frequency variability of data is due to an interplay of the spatiotemporal variability of fluxes like the daily and seasonal cycles of the biosphere and variability of transport processes such as turbulence and convection in the PBL or weather systems. GLOBALVIEW-CO2 pub-

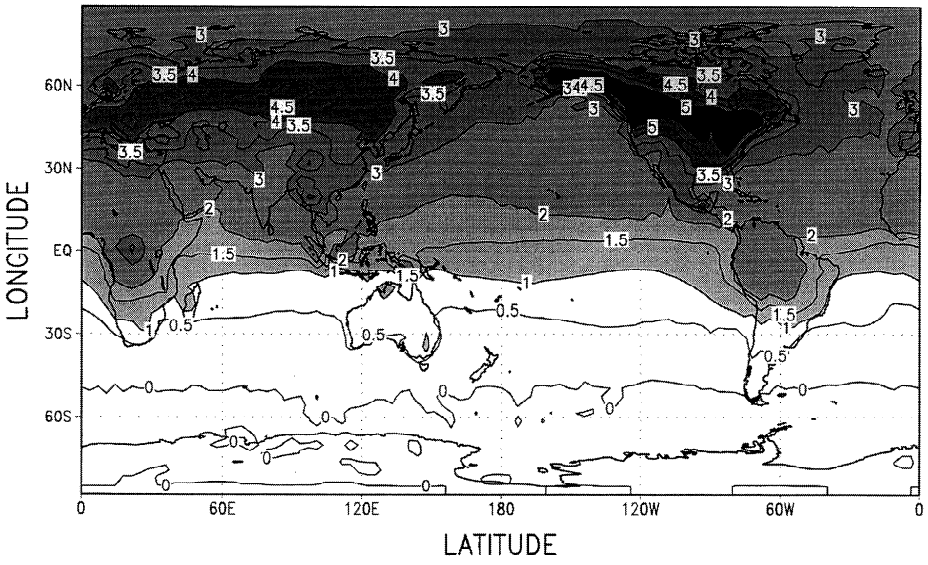

Figure 3. Same as Figure 2 for a flux of 1 Pg C yr^{-1} region^{-1} from all continental regions in Figure 1 with spatial flux pattern of net primary productivity (CASA biosphere model).

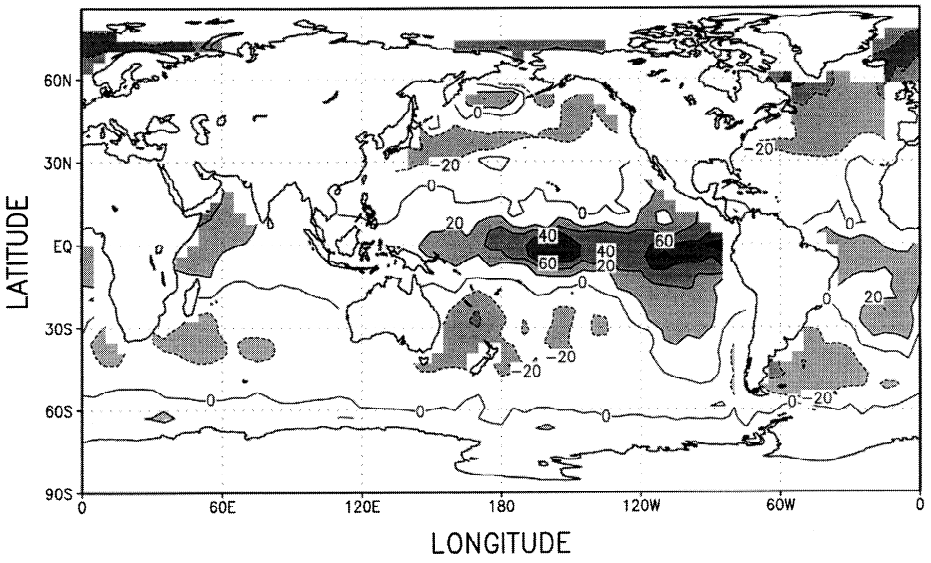

Figure 4. Oceanic CO_2 exchange fluxes (gC m^{-2} $month^{-1}$) estimated from measurements of air-sea partial pressure differences by *Takahashi et al.* [1997].

lishes residual standard deviations of their measurements from which the seasonal and long-term trends are removed. These residual standard deviations (Figure 5, top) are largest for continental stations and generally decrease with decreasing latitude. In the Northern Hemisphere the ratio of the thus determined high-frequency data variability for stations within the continents to remote stations varies between 2 and 6.

Our first model of high-frequency data variability (for $\mathbf{C}_{\Delta\chi}$) uses the combined signal of fossil fuel emissions plus biospheric exchange fluxes plus oceanic exchange fluxes [*Andres et al.*, 1996; *Potter et al.*, 1993; *Takahashi et al.*, 1997] simulated with GCTM. From the thus modeled mixing ratio time-series we determine the residual standard deviation from the low-pass filtered version in the same way as GLOBALVIEW-CO2 [e.g., *Conway*

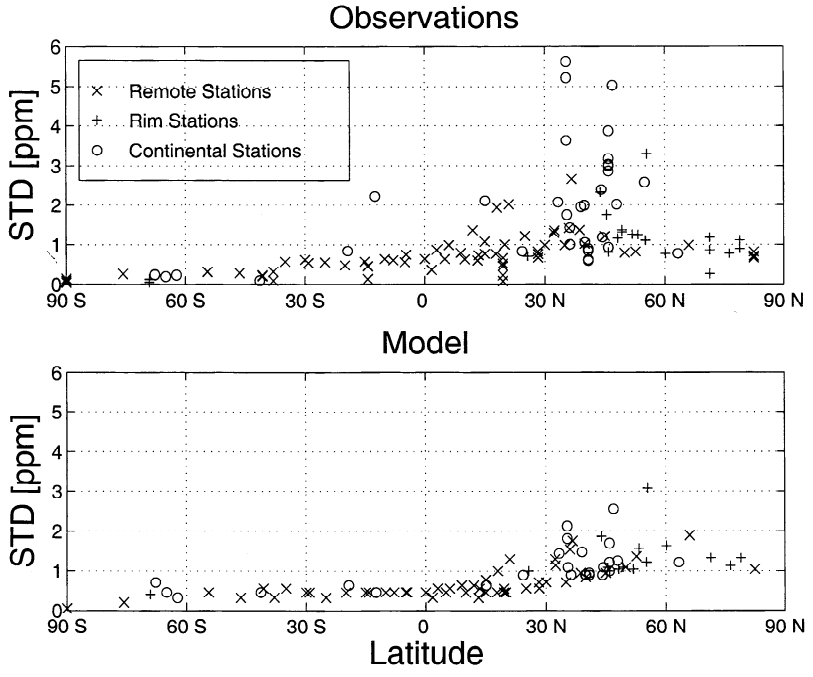

Figure 5. Residual standard deviation of (top) the data of GLOBALVIEW-CO2 (1997) and of (bottom) the combined signal of fossil fuel emissions plus biospheric fluxes plus oceanic fluxes[6] [*Andres et al.*, 1996; *Potter et al.*, 1993; *Takahashi et al.*, 1997] simulated with GCTM and sampled at the GLOBALVIEW-CO2 observation stations.

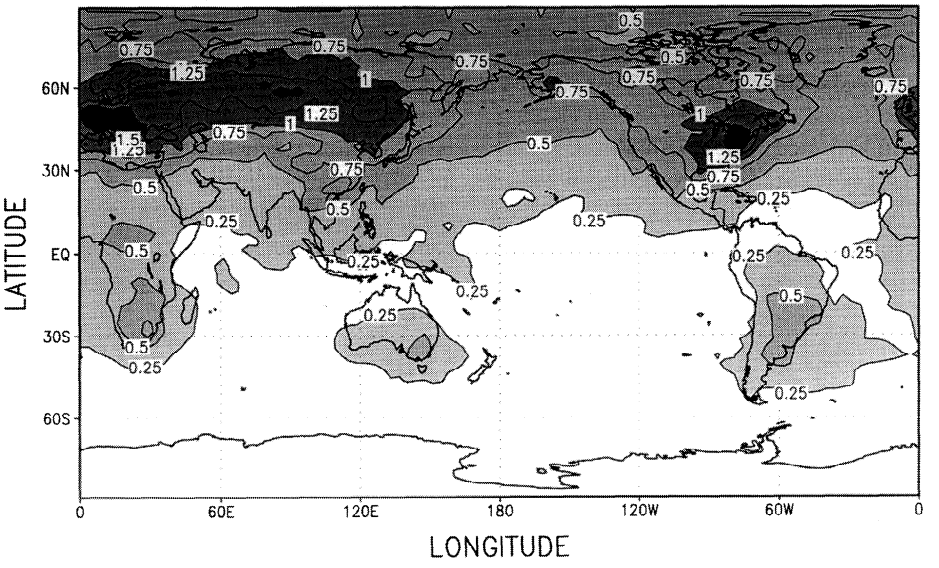

Figure 6. Residual standard deviation of the combined CO_2 signal of the biosphere (CASA), the oceans [*Takahashi et al.*, 1997], and of fossil fuel burning caused by natural variability of atmospheric transport as simulated by GCTM.

et al., 1994]. The variability of annual mean mixing ratios though is not given by the residual standard deviation but the residual standard deviation divided by the square root of the number of observations. We have to account for that in the optimizations. We do so by normalizing the spatial distribution of the residual standard deviation such as to match the reported values of *Conway et al.* [1994] most closely.

The spatial pattern of the residual standard deviation reflects the positions of largest fossil fuel emissions and NPP (Figure 6) as it should, and the overall meridional structure compares reasonably well with GLOBALVIEW-CO2 (Figure 5, bottom) even though the value for continental stations is underestimated (approximately by a factor 2-3). This error model is realistic insofar as it considers all fluxes which contribute to high-frequency data variability (fossil fuel emissions and biospheric and oceanic exchange fluxes) but unrealistic in discarding the daily cycle and short-term fluctuations of the biosphere (the CASA biosphere varies on a monthly basis) and in misrepresenting the daily dynamics of the PBL. The spatial structure of this error model is exclusively caused by the interplay between synoptic-scale transport variability and the relative magnitude of fluxes.

Because our first model for high-frequency data variability underestimates the contrast between continental and remote stations and is based on oversimplified physics, we use a second model to test our conclusions on the partitioning of stations between continents and the oceans. This model assumes 10 times higher high-frequency data variability (standard deviation) for continental stations compared to oceanic stations. Note that for the distribution of stations between continents and oceans only the relative magnitude of high-frequency data variability matters, not the absolute magnitude. Note also that we neglect measurement error in both models because it is comparably small.

From the models of high-frequency data variability, we determine the diagonal elements of the data covariance matrix needed for the network optimization and assume that these are uncorrelated, i.e., that the off-diagonal elements are zero. If observation stations do not lie too close to each other as is the case for "small" networks (≤ 150 stations), this is probably a reasonable assumption.

2.3. Optimization

Optimizations of network spatial structure are performed with simulated annealing [*Kirkpatrick et al.*, 1983; *Rayner et al.*, 1996]. This algorithm mimics cooling of a material to 0 K, which eventually attains its minimal energy configuration because thermal fluctuations allow it to escape from locally minimal configurations. Similarly, we look in our application for networks that permit us to estimate surface fluxes with minimal uncertainty [*Rayner et al.*, 1996]. The analog of the arrangement of atoms within the cooling material accordingly is the configuration of the observation stations, and the analog of the energy of the material is the uncertainty of the estimates if based on this configuration. To perform optimizations, a network of N observation stations is specified by the stations' longitudes and latitudes φ_i and ϑ_i, $i = 1, ..., N$. Cooling is modeled by a geometric progression $T_n = \gamma^n T_0$ with $0 < \gamma < 1$, where T mimics temperature and γ is the cooling rate. The initial temperature T_0 is of the same order as half the parameter range, 90° for latitude and 180° for lon-

gitude. At each time step, the network is perturbed by the temperature dependent amount: $\delta\varphi_i = \mathrm{sgn}\,\alpha T_n$, where α is a random number between 0 and 1, sgn = -1 if $\gamma <= 0.5$ and sgn = $+1$ if $\gamma > 0.5$. After each perturbation the quantity to be minimized is calculated for the new configuration and compared to the previous value. If the new value is lower than the old value, then the new configuration is accepted (replaces the old value). If the new value is larger than the old value, then it is still accepted with probability $\exp(-(\text{New-Old})/kT_n)$ where k is a tunable constant. This permits us to abandon local minima.

Although this method is robust, it does require considerable time to converge, and so one cannot be certain that one has found the global minimum (we ran the method for up to $5\text{x}10^5$ iterations, depending on the size of the network). Nevertheless, we checked the convergence directly in simple cases, and in all cases we performed repeated optimizations using different initial configurations to ensure convergence to a similar solution.

3. Analysis of GLOBALVIEW-CO2

The stations of GLOBALVIEW-CO2 (Figure 1) are positioned remotely from continents on islands in the ocean with a few exceptions. The rationale behind this arrangement is to sample only air that is remote from large fluxes and thus relatively homogeneous with correspondingly low CO_2 variability. Nevertheless, there is a price to pay for the reduced variability: Flux signals are strongly diluted at these stations. Remote stations are well suited to measure a mean mixing ratio of the atmosphere but not to separate local sources and sinks in an inversion.

To illustrate the problem which arises in an inversion based on a limited number of observations at remote stations like the ones of GLOBALVIEW-CO2, suppose that we wished to estimate, by inversion, the current pattern of fossil fuel emissions (Figure 7). We use the same transport model both in the inversion and to generate the mixing ratio distribution for fossil fuel emissions. Thus there can be no error due to transport. However, supposing that we do not know the true pattern of fluxes, we assume spatially uniform fluxes for each region in calculating the footprints $\Delta\chi_r(\vec{x})$. This systematic error is compounded by the relatively large estimation errors associated with insufficient data coverage, which generally results in strong amplification of the inconsistencies between the "true" and modeled fluxes (compare section 4.4).

Differences between the true and the spatially structureless mixing ratio distributions are concentrated in the main industrial regions. Although one might expect that this discrepancy would cause large errors in the Northern Hemisphere, the largest errors instead occur for the equatorial continents: Africa, South America, and Australasia (1.4, 0.6, and 1.6 Pg C yr^{-1}). There are two reasons for this pattern of error. First, estimates of fluxes from South America, Africa and Atlantic Tropics are extremely sensitive to small changes in the observed concentrations (up to 3 Pg C yr^{-1} region^{-1} ppm^{-1}). We measure the sensitivity of an estimate to high-frequency variability in data by $\delta\phi_i/\delta\Delta\chi_j = (A^{-1})_{ij}$. These sensitivities are displayed in Figure 8 for all stations of GLOBALVIEW-CO2. The station's locations are displayed in Figure 9 with the station number equal to the station number in Figure 8. Most sensitive stations to estimate fluxes from Africa, South America, and equatorial Atlantic are marked with black dots, inner dark grey circles and outer bright grey circles respectively in Figure 9. Because natural variation in atmospheric transport causes inevitable variation in CO_2 concentrations of the order of 0.2-0.4 ppm [*Conway et al.*, 1994], sensitivities as large as 3 Pg C yr^{-1} region^{-1} ppm^{-1} cause large estimation errors. The stations yielding the largest sensitivities for South America and Africa are remote from both regions. Moreover, South American and African sensitivities are mostly anticorrelated (Figure 8).

The error amplification matrix $\mathbf{A}^{-1}(\mathbf{A}^{-1})^T$ for this inversion is a square matrix with dimensions equal to the number of flux regions (Figure 10). The ith diagonal element is the estimate error per high-frequency data variability of the estimate of fluxes from region i. The ijth off-diagonal element of the error amplification matrix is a measure for the correlation between the estimates of fluxes from regions i and j per high-frequency data variability (compare section 2.2). This error amplification matrix (Figure 10) shows generally that GLOBALVIEW-CO2 is not well suited to estimate fluxes from the equatorial regions. The structure of this matrix is very similar if determined either with GCTM or SKYHI indicating that it is not dependent on the details of the transport model. Although temperate and boreal regions are better constrained than tropical regions, the error amplification matrix shows that several regions in the Northern Hemisphere are correlated with Africa and South America (i.e., boreal Eurasia and Africa). Thus uncertainty in the tropics may cause uncertainty in regions far from the tropics. These correlations are to some extent a result of the design strategy of GLOBALVIEW-CO2. Stations creeping around the perimeter of continents contribute substantially to estimates from several different regions, thereby strenghtening the interdependence of estimates. Locating stations within the continents would make estimates more independent.

Why are estimates from South America and Africa so poorly constrained by GLOBALVIEW-CO2? To some extent, the problem is not only with the network design but also with atmospheric transport, which accentuates

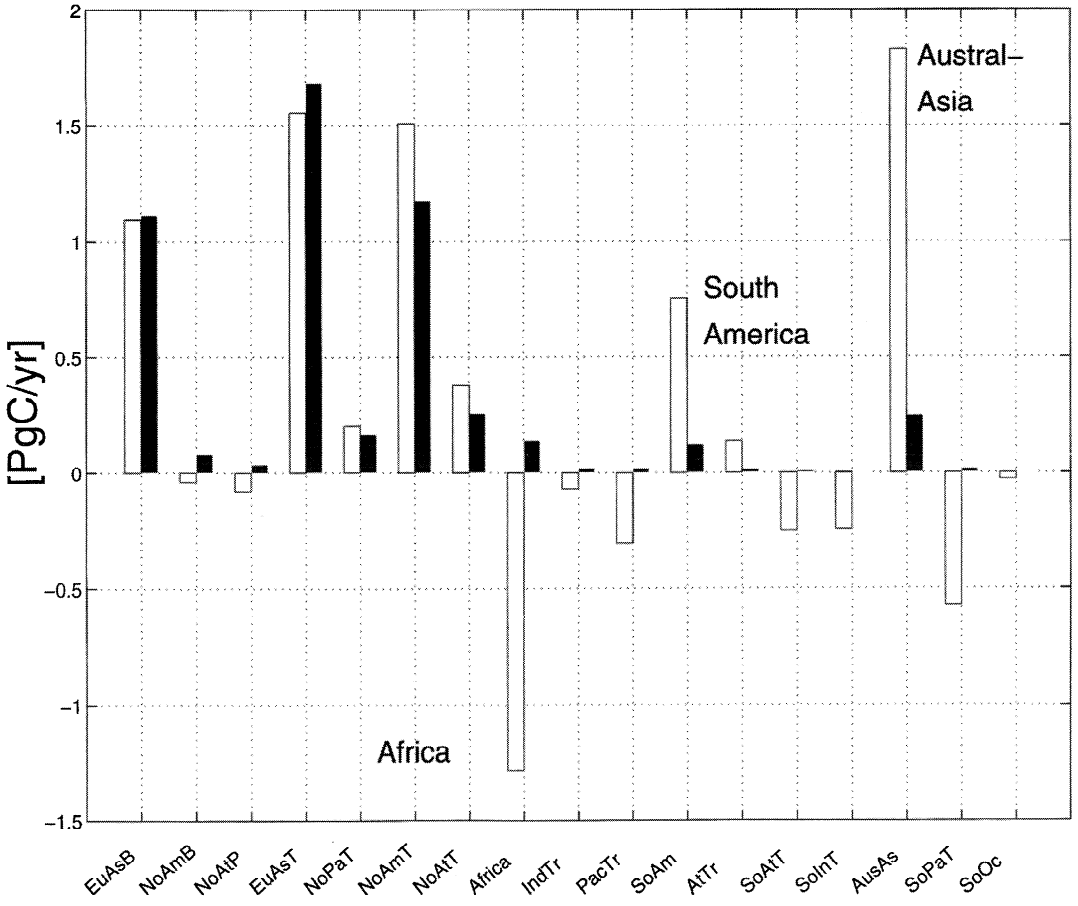

Figure 7. Estimate of fossil fuel emissions using structureless footprints for the inversion and the fossil fuel footprint as data, sampled at the stations of GLOBALVIEW-CO2 (left, light bars), compared with the true emissions used to simulate the fossil fuel footprint (right, dark bars). Abbreviations for the names of the regions are as follows: EuAsB, Eurasia Boreal; NoAmB, North America Boreal; NoAtP, Polar North Atlantic; EuAsT, Temperate Eurasia; NoPaT, Temperate North Pacific; NoAmT, Temperate North America; NoAtT, Temperate North Atlantic; Africa, Africa; IndTr, Tropical Indian Ocean; PacTr, Tropical Pacific; SoAm, South America; AtTr, Tropical Atlantic; SoAtT, Temperate South Atlantic; SoInT, Temperate South Indian Ocean; AusAs, Australasia; SoPaT, Temperate South Pacific; SoOc, Southern Ocean.

the deficiencies of the network. *Gloor et al.* [1999] show that estimation errors for tropical continents are 2-5 times larger than errors for temperate continents with equal coverage by observation stations because the rapid ventilation of the PBL in the tropics strongly reduces the size of the signals there (e.g., Figure 3).

The elements of the matrix $\mathbf{\Sigma}$ from the singular value decomposition (the singular values) confirm that the worst estimated linear combination of regions, the first column of the matrix $\mathbf{V}$ (compare section 2.1) - is a source in Africa and a balancing sink in South America, or vice versa, because of strong anticorrelation between these regions (STD=17 Pg C yr^{-1} ppm^{-1}). The reason for the strong anticorrelation is both the spare coverage of the tropical region by observation stations and the quick zonal mixing that makes it difficult to distinguish fluxes from regions within the same zonal band. Not surprisingly, the best constrained combination of regions, the last column of the matrix $\mathbf{V}$, is the interhemispheric gradient (Figure 11).

An analysis of sensitivities of estimates for the northern hemispheric continental regions (Eurasia Boreal, Eurasia Temperate, North America Boreal, and North America Temperate (Table 1)) identifies Temperate Eurasia as the most weakly constrained northern terrestrial region. In particular, the natural high-frequency variability in mixing ratios at the stations in the South China Sea may easily change the estimate for the flux from Eurasia by 0.5 Pg C yr^{-1}.

In summary, although the GLOBALVIEW-CO2 network is well designed to characterize the meridional gradient in CO_2, it is not well suited to characterize several regions of critical importance to the global carbon cycle. The mean standard error (1 σ) of flux estimates

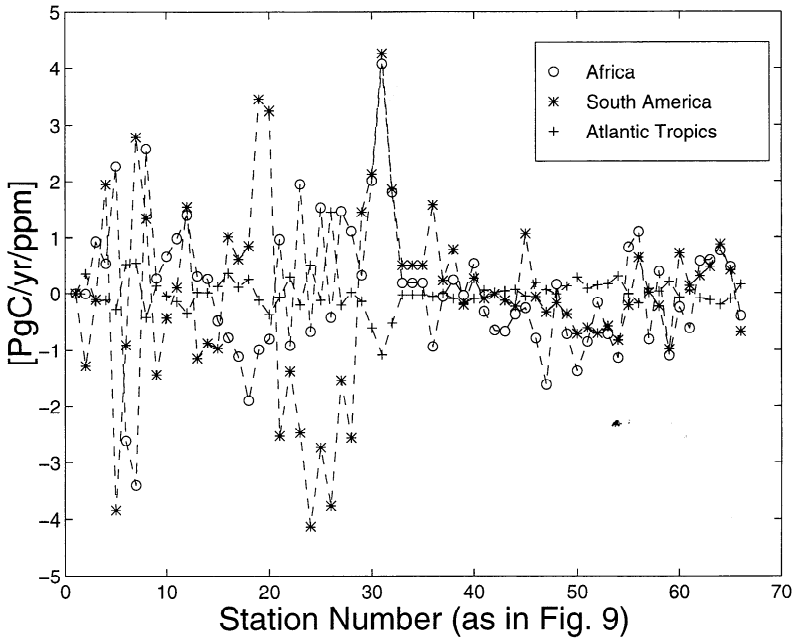

Figure 8. Sensitivities of estimates of fluxes from South America, Africa, and Atlantic Tropics to changes in observations ($\delta\phi_r^{est}/\delta\Delta\chi_i$, (Pg C yr^{-1} region^{-1} ppm^{-1})) if based on GLOBALVIEW-CO2. The station numbers increase with increasing latitude and agree with the station numbers in Figure 9.

for mean high-frequency data variability of 0.3 ppm is 0.7 Pg C yr^{-1} region^{-1} in Figure 10 and the maximum standard error is 2.9 Pg C yr^{-1} region^{-1} (for Africa). There are simply not enough stations, those which are available are biased to the Northern Hemisphere and to North America in particular, and most stations are positioned remote from the regions with largest carbon fluxes (fossil fuel emissions and net primary productivity and respiration of the terrestrial biosphere). Particularly sparsely covered regions are Africa and South America and southern hemispheric oceanic regions.

4. Optimal networks

What is the best way to improve the GLOBALVIEW network? To answer this question, we first consider some characteristics of optimal surface networks. The value of optimizing the network may be shown by comparing uncertainties of estimates from random (where stations have random coordinates) and optimized networks as a function of the number of observation stations. To optimize the networks in Figure 12, we minimized the mean of the estimate uncertainties $1/R\sum_{r=1}^{R}\mathbf{C}_{\varphi,rr}$ following *Rayner et al.* [1996]. For the inversion scheme, we followed *Fan et al.* [1998] with 7 land regions with CASA-biosphere flux patterns and 10 ocean regions with observed partial pressure flux patterns. The error model is based on the CASA biosphere, fossil fuel emissions, and observed oceanic air-sea partial pressure differences. Note that small random networks are substantially inferior to optimized ones but that the value of optimization decreases with network size. For larger networks of the order of 150 stations, estimate uncertainties are smaller than currently unavoidable systematic errors [*Gloor et al.*, 1999]. Thus optimization gives a substantial benefit for networks as small as GLOBALVIEW-CO2. In what follows, we focus on small networks; we first consider their general properties and then examine extensions of GLOBALVIEW-CO2.

4.1. Properties of Optimal Networks

As explained in section 2.2, there are two contributors to the uncertainty of estimates ($\mathbf{C}_\varphi = \mathbf{A}^{-1}\mathbf{C}_{\Delta\chi}(\mathbf{A}^{-1})^T$): high-frequency data variability, $\mathbf{C}_{\Delta\chi}$, and the amplification of the variability, $\mathbf{A}^{-1}(\mathbf{A}^{-1})^T$. To minimize $\mathbf{C}_{\Delta\chi}$, stations have to avoid the regions with largest fluxes, those with strong biospheric exchange fluxes and large fossil fuel emissions (Figure 6). This is the strategy behind GLOBALVIEW-CO2. The structure that minimizes the error amplification, $\mathbf{A}^{-1}(\mathbf{A}^{-1})^T$, might be a little less obvious. To determine it, we build networks from scratch that minimize the mean error amplification, $\overline{\mathbf{ErrAmp}}$.

Figure 13 shows the result of minimizing mean error amplification for optimal networks of 20 and 40 stations. For simplicity, fluxes were spatiotemporally uniform in each of the 17 regions, but the conclusions are not changed qualitatively if we use fluxes structured like the pattern of CASA NPP, oceanic uptake from *Taka-*

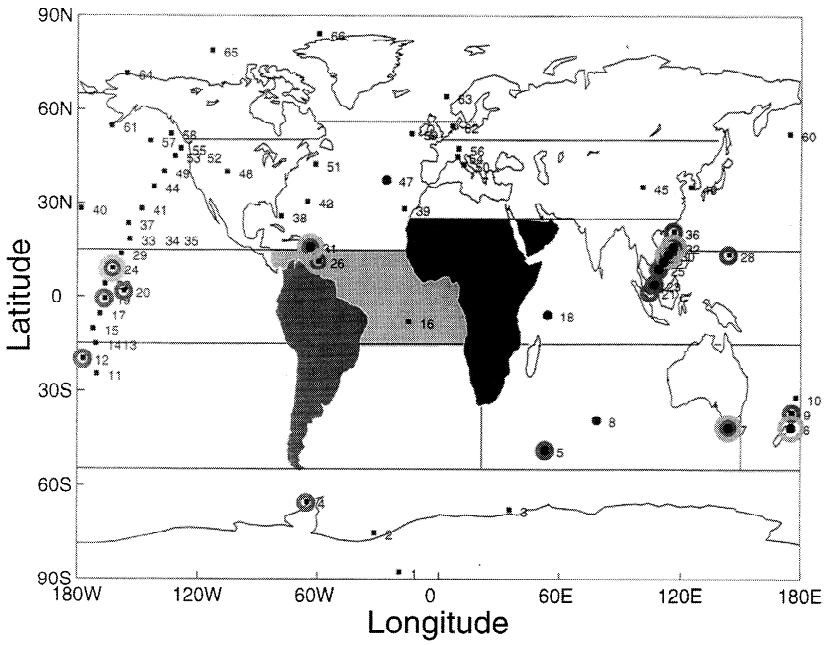

Figure 9. Locations of most sensitive GLOBALVIEW-CO2 stations for the estimation of fluxes from South America (intermediate dark grey circles), Africa (black dots), and Atlantic Tropical (outer bright grey circles). Station numbers are identical with the station numbers in Figure 8.

hashi et al. [1997], or fossil fuel consumption. Note that the networks that minimize error amplification have (1) approximately equal numbers of observing stations in each region and (2) observing stations located where the signals within each region are largest. The locations of largest signals are determined entirely by atmospheric circulation in these examples because fluxes are structureless within each region. Regional maxima for subtropical oceanic regions are at the center of high-pressure cells where air is descending; regional maxima

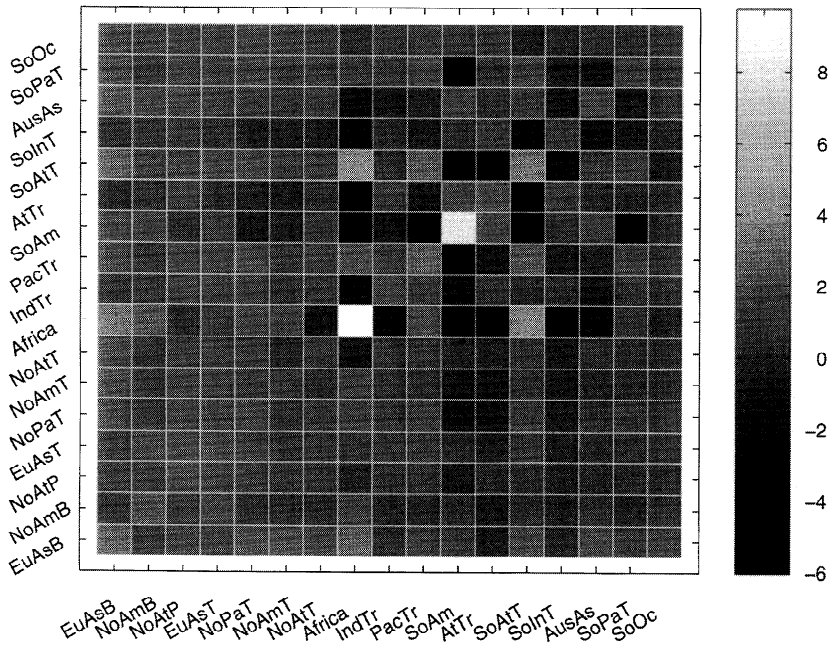

Figure 10. Error amplification matrix (cf. section 2.2) for the estimation of fluxes from the 17 regions in Figure 1 if based on annual mean mixing ratios measured at the GLOBALVIEW-CO2 stations. The units of the matrix elements are (Pg C yr^{-1} region $^{-1}$ ppm^{-1})^2.

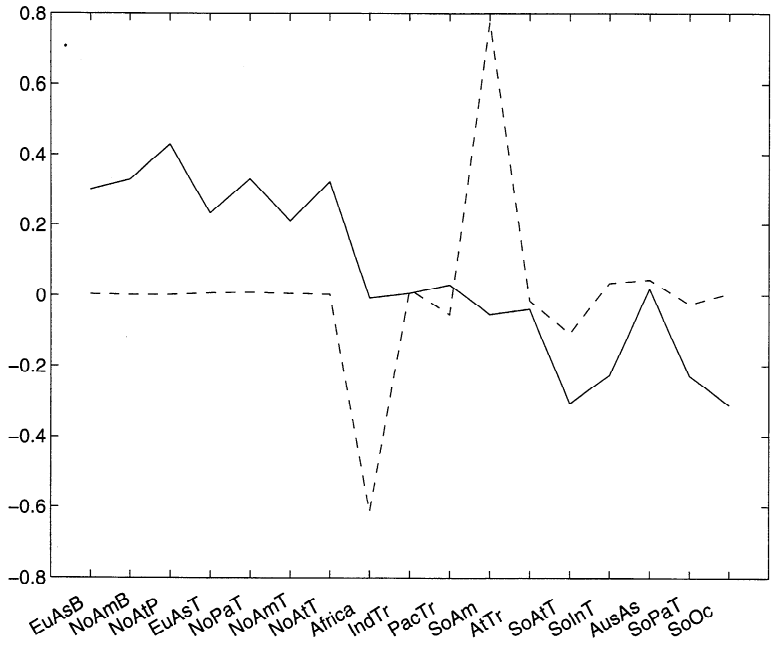

Figure 11. Best (black line) and worst (dashed line) estimable combinations of flux regions if based on GLOBALVIEW-CO2 observation stations. These combinations are a byproduct of the singular value decomposition $\mathbf{A} = \mathbf{U\Sigma V}$ of the matrix $\mathbf{A}$ which maps fluxes to the expected mixing ratio distribution. They are given by the columns of $\mathbf{V}$; the higher the column number, the larger the uncertainty of the estimate of a linear combination of regions combined according to the elements of the column of $\mathbf{V}$ (compare section 2.1).

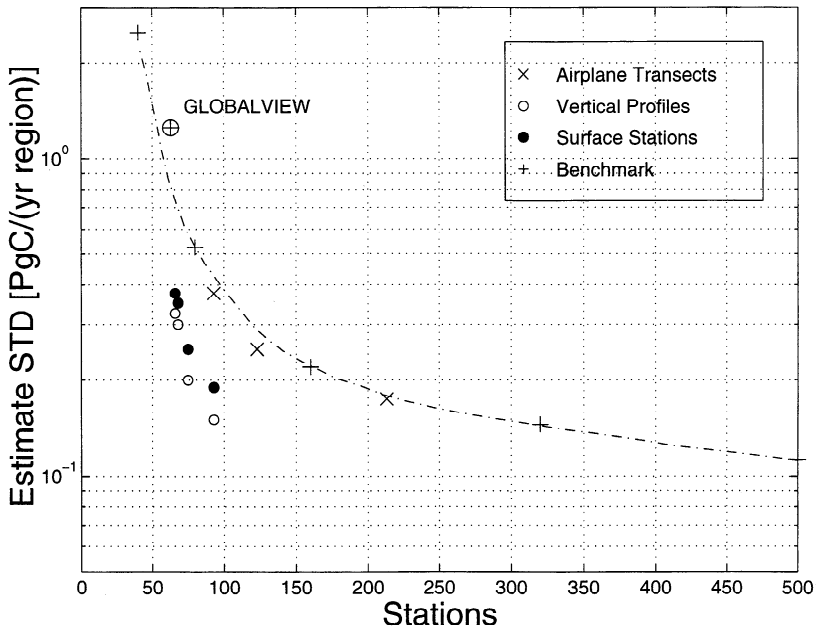

Figure 12. Comparison of the mean estimate uncertainty as a function of the number of observation stations for randomly positioned networks (benchmark) with the minimally achievable estimate uncertainty if using the following different measurement strategies: (1) optimally located N-S airplane transects, (2) 3,5,12, and 30 optimally positioned vertical profiles in addition to GLOBALVIEW-CO2, and (3) 3,5,12, and 30 optimally positioned surface stations in addition to GLOBALVIEW-CO2.

for equatorial oceans are on the equator because of the convergence of the doldrums and maxima for continental regions are located near the center of each region. (The choices of stations next to each other may look suspicious. Because the data covariance structure implicit in minimizing mean error amplification is the identity matrix though, there is no correlation structure which would penalize nearby stations. Such choices hence underline the need to reduce estimate uncertainty for this region which might for example be accomplished by sampling air more often at this specific location.)

Thus strategies minimizing $\mathbf{C}_{\Delta\chi}$ demand placement away from continents, whereas strategies minimizing error amplification demand placement within the most variable regions of continents. We now balance these conflicting demands by minimizing mean estimate uncertainty: $1/R\sum_{r=1}^{R}\mathbf{C}_{\varphi,rr}$ where the inversion model is the one of *Fan et al.* [1998].

The results for the first of our two scenarios for high-frequency data variability (the annual variance of the superposed signals of CASA terrestrial biosphere, fossil fuel emissions, and air-sea partial pressure differences (section 2.2)) are shown for networks of 20 and 40 stations in Figure 14. Note that the even distribution of stations per regions persists but that locations within the largest signals are generally avoided, especially over the continents, because of large variability caused by the fluxes of the terrestrial biosphere and fossil fuel burning. Nevertheless, because stations in the continents tend to be located near the hottest spots, the need to capture regional signals still dominates over the disadvantage of large high-frequency data variability. The twin conclusions of even global coverage and placement of stations close to (but not in) hot spots are also obtained for the scenario with 10 times higher high-frequency data variability in continental regions than in oceanic regions. Networks that minimize uncertainty of the estimates do not support the design strategy behind the current GLOBALVIEW network.

4.2. Optimal Extension of GLOBALVIEW-CO2

In this section we examine optimal extensions of GLOBALVIEW-CO2. Specifically we add 3, 5, or 12 new surface stations that minimize the uncertainty of the mean of flux estimates from all regions (Plate 1). We also consider alternate strategies: weekly vertical profiles in 3, 5, and 12 optimal locations and elevations (0-3 km elevation) and weekly 1, 2, or 5 N-S airplane transects with optimal locations and elevations from the ground to the free troposphere at 8.7 km. As expected from the analysis presented in section 3, optimal surface stations or vertical profiles are positioned in South America, Africa, and the temperate South Atlantic if either 3, 5, or 12 stations or profiles are added because these are the most poorly constrained regions at present. With five added optimal stations or profiles, we also see a station in the temperate South Pacific and in Siberia. Note that incremental implementation of a 5-station or 12-station optimal design carries little or no penalty because the larger designs also contain the loca-

Table 1. Largest Sensitivities of Flux Estimates to Perturbations of Observations at a Single Station Holding the Others Constant for Estimation of Fluxes From Eurasia Boreal and Temperate and North America Boreal and Temperate Using GLOBALVIEW-CO2

Eurasia Boreal		
Shemya	(60)	0.75
Mould Bay	(65)	-0.60
Station M	(63)	0.43
Cold Bay	(61)	0.38
Barbados	(26)	-0.28
Virgin Island	(31)	0.23
North America Boreal		
Station M	(63)	-0.48
Barbados	(26)	0.33
Alert	(66)	0.22
Sable Island	(51)	0.18
Cape St. James	(58)	0.18
Olympic Peninsula	(55)	-0.10
Niwot Ridge	(48)	-0.10
Eurasia Temperate		
S China Sea	(36)	1.8
Virgin Island	(31)	-1.4
Izana	(39)	1.3
Christmas Island	(20)	0.8
Mauna Loa	(34)	-0.8
Qinghai	(45)	0.8
Syowa	(3)	-0.8
Monte Cimone (*)		0.2
China, (Hunan Province) (*)		0.3
Schauinsland (*)		0.2
North America Temperate		
Sable Island	(51)	0.9
Biscayne	(38)	0.9
Izana	(39)	-0.6
Virgin Island	(31)	-0.5
Bermuda	(42)	0.5
Mace Head	(59)	-0.5
Mauna Loa	(34)	-0.5
Niwot Ridge	(48)	0.5
Syowa	(3)	-0.5

The units of the sensitivities are Pg C yr^{-1} ppm^{-1}. Numbers in parentheses refer to the position of the stations in Figure 8. The Eurasian temperate stations marked by an asterisk are the three largest sensitivities after inclusion of two optimal stations.

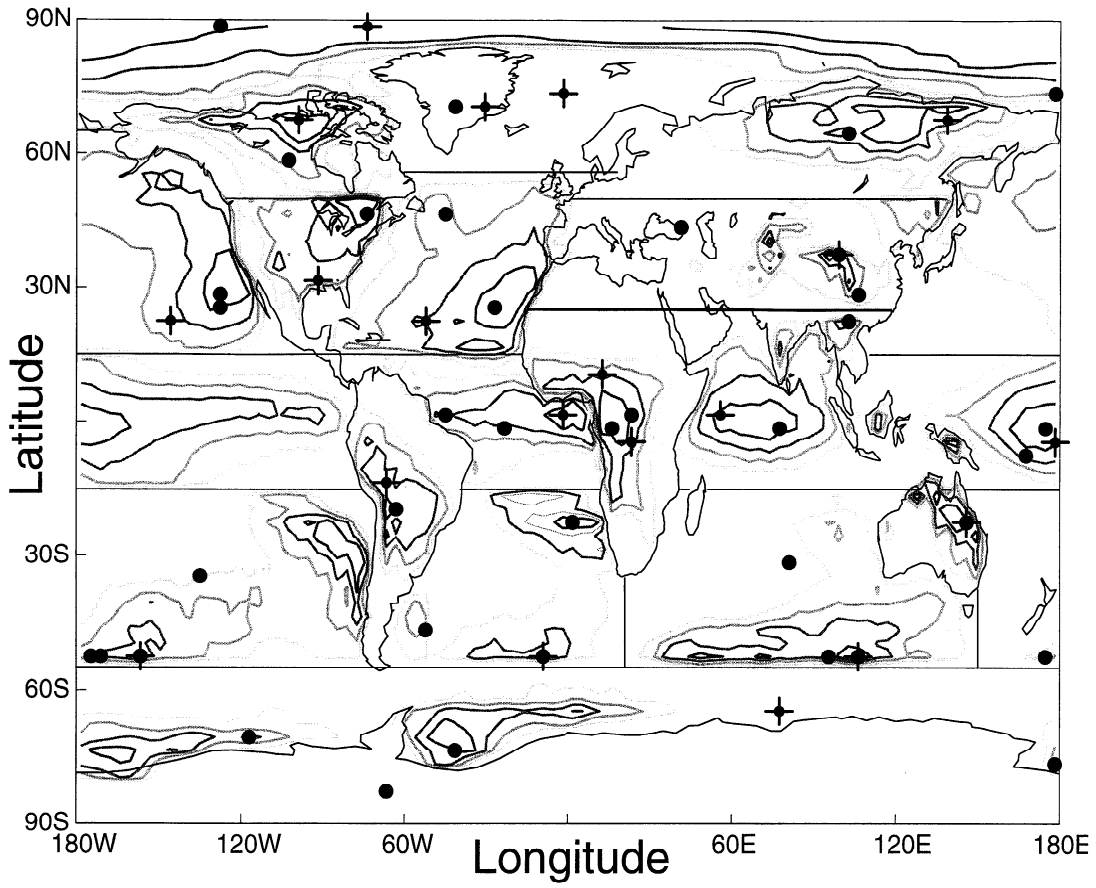

Figure 13. Positions of networks encompassing 20 (crosses) and 40 (dots) observation stations that minimize mean error amplification. The inversion scheme uses structureless flux patterns. Also shown are the isolines of the mixing ratio distributions in the model suface layer for a flux from each of the 17 source regions separately in percentage of the maximal signal (90, 80, 70, and 60%).

tions of stations or profiles in the smaller designs. The locations of the added stations and vertical profiles also confirm the general properties of optimal networks discusseed in section 4.1. The new locations make global coverage more even and usually occur near the largest local signals (compare Figures 3, 4 and 14).

The analysis suggests the most needed locations for the extension of GLOBALVIEW-CO2: (1) Central Africa, Zaire, Uganda or Tanzania; (2) Equatorial South America, Brazil, Amazon; (3) South Atlantic Temperate, below southern tip of Africa (approximately near Bovet island) or South Georgia island (U. K.); (4) Eurasia, Central Siberia, Krasnoyarsk or Magadan, and China, Hunan Province; (5) South Indian Temperate, Macquarie Island (Australia) and Kerguelen (F); (6) Tropical and Southern Temperate Pacific, off-shore South America 15 S or Galapagos islands and 40 S, Juan Fernandez (Chile); (7) Indonesia, Papua New Guinea; (8) North America, Edmonton(Can) and Southern Mexico.

The optimal airplane transects are generally located next to the Earth's surface. The only exception is a single transect through Indonesia for the case of five transects, and this is in the second atmospheric layer. Higher elevation transects are not chosen because of the strong decrease of signals with altitude caused by mixing in the atmosphere: Estimate errors increased from 0.5 Pg C yr^{-1} $region^{-1}$ if observed at 0.08 km height to 1.1 Pg C yr^{-1} $region^{-1}$ at 3.1 km height on the average for 80 randomly located measurements despite the decrease of high-frequency data variability with height.

The positions of the optimal airplane transects confirm some of the measurement priorities determined for surface stations and profiles: For 1, 2, and 5 transects, at least one crosses right through Africa to reduce the estimate uncertainty of the most poorly constrained, equatorial regions. In addition, transects through Central Siberia are chosen repeatedly, and this has already been identified as an insufficiently covered region.

We now turn to the optimization results for specific regions and combinations of regions: (1) North America, (2) Eurasia, (3) North America and Eurasia combined, (4) the northern hemisphere (7 northernmost re-

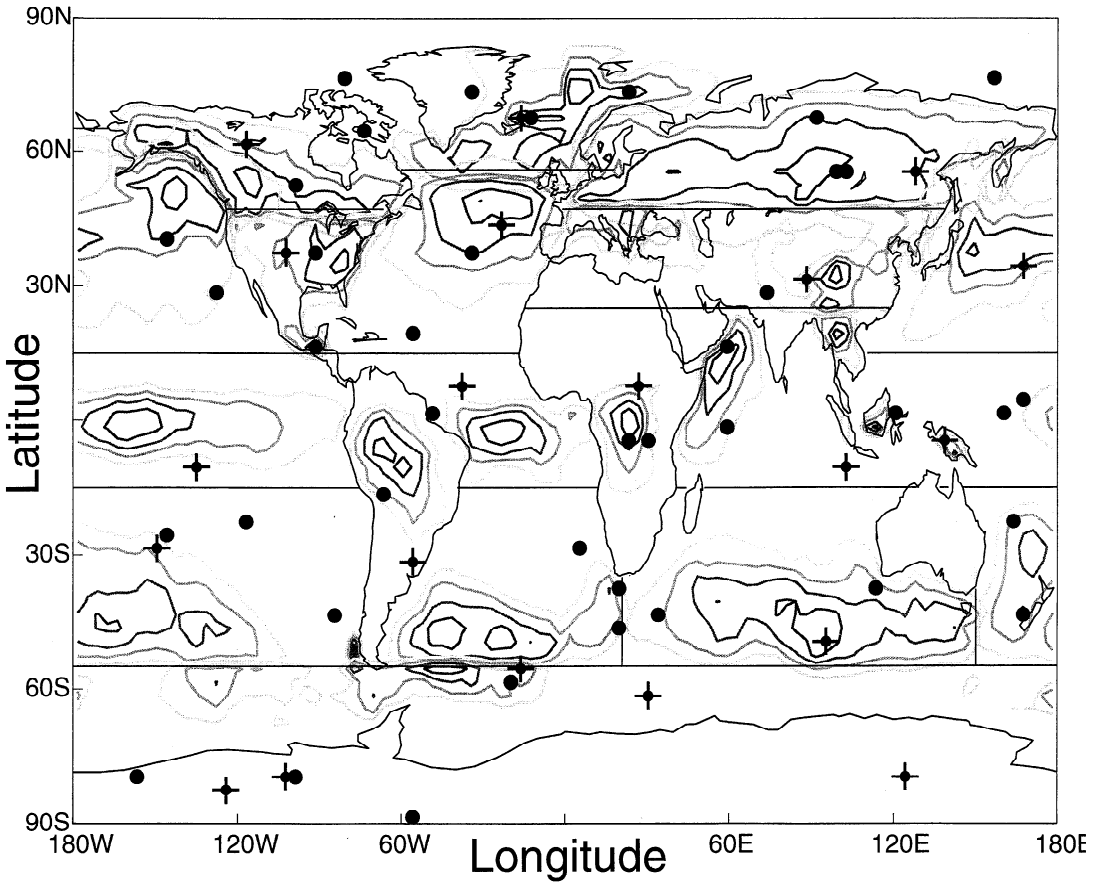

Figure 14. Same as in Figure 13 for networks which minimize the mean estimate uncertainty for an inversion scheme based on net primary productivity flux patterns for continental regions and flux patterns estimated from measurements of air-sea partial pressure differences for oceanic regions.

gions), and (5) all continental regions combined. The quantity we minimize for these cases is the sum over the diagonal elements of the covariance matrix of the estimates corresponding to the regions that we consider; in the case of North America, for example, this is one diagonal element of the covaraince matrix of the estimates. In each case we considered additions of 3, 5, and 12 surface stations.

The results for most of these combinations are shown in Plate 2 and Plate 3 (omitted cases are North America with 12 new stations and Eurasia with 12 new stations because these cluttered the figures while providing little additional insight). By comparing the locations in Plate 2 and Plate 3 with the signals (isolines) in Figure 14, one can again explain the placement of stations in each case as an intersection of three conflicting factors. Within the region whose estimation uncertainty is minimized, stations are placed near to the biggest signals to reduce error amplification but not always within them to reduce high-frequency data variability. For example, note the North American (green) circles near the maxima in western Canada and southern Mexico. Note also the enormous decrease (six-fold) of the largest sensitivities of estimates to high-frequency data variability for the case of Eurasia Temperate in Table 1 if three optimal stations are added.

In addition, the patterns show a surprisingly large influence of the need for globally even coverage even though the networks are now optimized to reduce subglobal uncertainty. In most cases, half or more of the new stations are placed outside of the region that is optimized. Examples include the six Southern Hemisphere stations chosen to reduce uncertainty in the combination of Eurasian and North American (dark blue circles with crosses in Plate 2), and the Siberian and near-Siberian stations (green circles in Plate 2) chosen to minimize North American uncertainty. Because the atmosphere couples all regions together, an understanding of the globe is necessary to pin down any one region (even so we did not impose a constraint here that ensures mass conservation in the inverse calculations). Thus already heavily sampled regions such as the oceanic regions in the Northern Hemisphere are almost never chosen in Plate 2 and 3, while the poorly

covered oceanic regions of the southern hemisphere are frequently chosen. One other interesting detail is that locations outside of the optimized region are not generally near the largest local signals, in contrast to stations within the optimized regions. Evidently, stations outside of an optimized region are chosen to constrain large-scale features, such as the interhemispheric gradient.

4.3. Comparison of Sampling Strategies

Figure 12 compares the GLOBALVIEW-CO2 network against optimally added surface stations, vertical profiles, and N-S transects (note that we have also 30 added optimal surface stations and profiles to supplement the cases of 3, 5, and 12 in Plates 1,2, and 3). As a benchmark, we also include the mean over 1000 randomly constructed networks of 40, 80, 160, 320, and 500 surface stations. Note that the average estimate uncertainty for the 66 station GLOBALVIEW-CO2 network is just over 1 Pg C yr^{-1} $region^{-1}$ and that this is close to what one would expect for a random network (actually slightly worse). The good news is that 3 optimal stations reduce this error by a factor of 3 while 30 reduce it by an order of magnitude. Optimizations with very large networks (i.e., 500 stations) show negligible advantage of optimizing with sampling this dense. Because GLOBALVIEW-CO2 with 30 optimal stations already has errors nearly as small as the 500-station network shown, the return on new stations diminishes sharply after 30, at least by the metric of average estimate uncertainty. Note that this analysis is restricted to the estimation of fluxes from 17 regions.

Vertical profiles appear to be marginally better than surface stations: 3 optimal profiles reduce uncertainty by a factor of 4, while 12 reduce it by an order of magnitude. It is important to understand that vertical profiles provide benefits beyond the relatively modest advantage over surface stations shown in Figure 12. Specifically, vertical profiles would provide information critical to improving and/or verifying the vertical transport in an AGCM, and this is essential to reduce systematic error in inversions. Also, the column integral of CO_2 concentration from a vertical profile should be largely unaffected by diurnal fluctuations in the boundary layer and thus should exhibit much less variability than surface data, while retaining information about surface fluxes (compare also *Tans et al.* [1996]).

Finally, a single, optimal N-S transect (which is chosen in the lowest model layer) has benefits comparable to 3 optimal surface stations, 2 transects are similar to 5 surface stations, and 5 transects are similar to 30 surface stations. Obviously, each flight from the Arctic to the Antarctic would entail considerable expense.

4.4. Robustness of Optimal Design to Structural Uncertainty in Transport, Terrestrial Biosphere, and Oceanic Uptake

Our optimizations use particular models for atmospheric transport (GCTM) and the spatiotemporal flux patterns (CASA biosphere and measurements of oceanic partial pressure). Are our results mainly driven by these choices, or are they valid more generally? Let us assume that true atmospheric transport is given by SKYHI as opposed to the transport used in the network optimizations (GCTM) and that continental and oceanic fluxes are structureless in space and time rather than given by CASA and *Takahashi et al.* [1997]. How well do our optimized networks perform in recovering these fluxes compared to GLOBALVIEW-CO2? In Figure 15 we list estimates for each of the 17 footprints from structureless fluxes in SKYHI (i.e., an estimate with 17 components for each of the 17 footprints) for both GLOBALVIEW-CO2 (left) and 12 optimal additional stations (right). The inversion itself is based on footprints resulting from the spatial patterns of the CASA biosphere and ocean partial pressure differences [*Takahashi et al.*, 1997] and simulated with GCTM. If the inversion were perfect, elements along the diagonal would equal one and all remaining elements would be zero. This is indeed approximately the case for the optimally extended network in contrast to GLOBALVIEW-CO2, which performs quite poorly. This result gives us confidence that our optimizations are fairly robust to the details of transport and spatiotemporal flux pattern at least as long as we add a few stations only.

We may identify the level of additional stations at which systematic errors from mismatch of transport and spatiotemporal flux patterns begin to dominate over the gain of additional information by comparing

$$\sigma_{\text{Bias}} \equiv \Big(\sum_{\text{regions}} \frac{1}{17} |\vec{\phi}^{\text{truth}}_{\text{region}} - \vec{\phi}^{\text{est}}_{\text{region}}|^2 \Big)^{1/2}$$

with the uncertainty as determined by error propagation (Figure 16). The parallel strong decrease with a few additional stations of both the estimate uncertainty and the bias (σ_{Bias}) indicates that the decrease of the bias is caused by the reduction of error amplification (of the mismatches between models). However, biases do not decline appreciably for addition of more than 12 stations because systematic errors dominate once error amplification is sufficiently reduced.

5. Summary and Conclusions

Our analysis reveals serious deficiencies of the GLOBALVIEW-CO2 network. Although well designed to

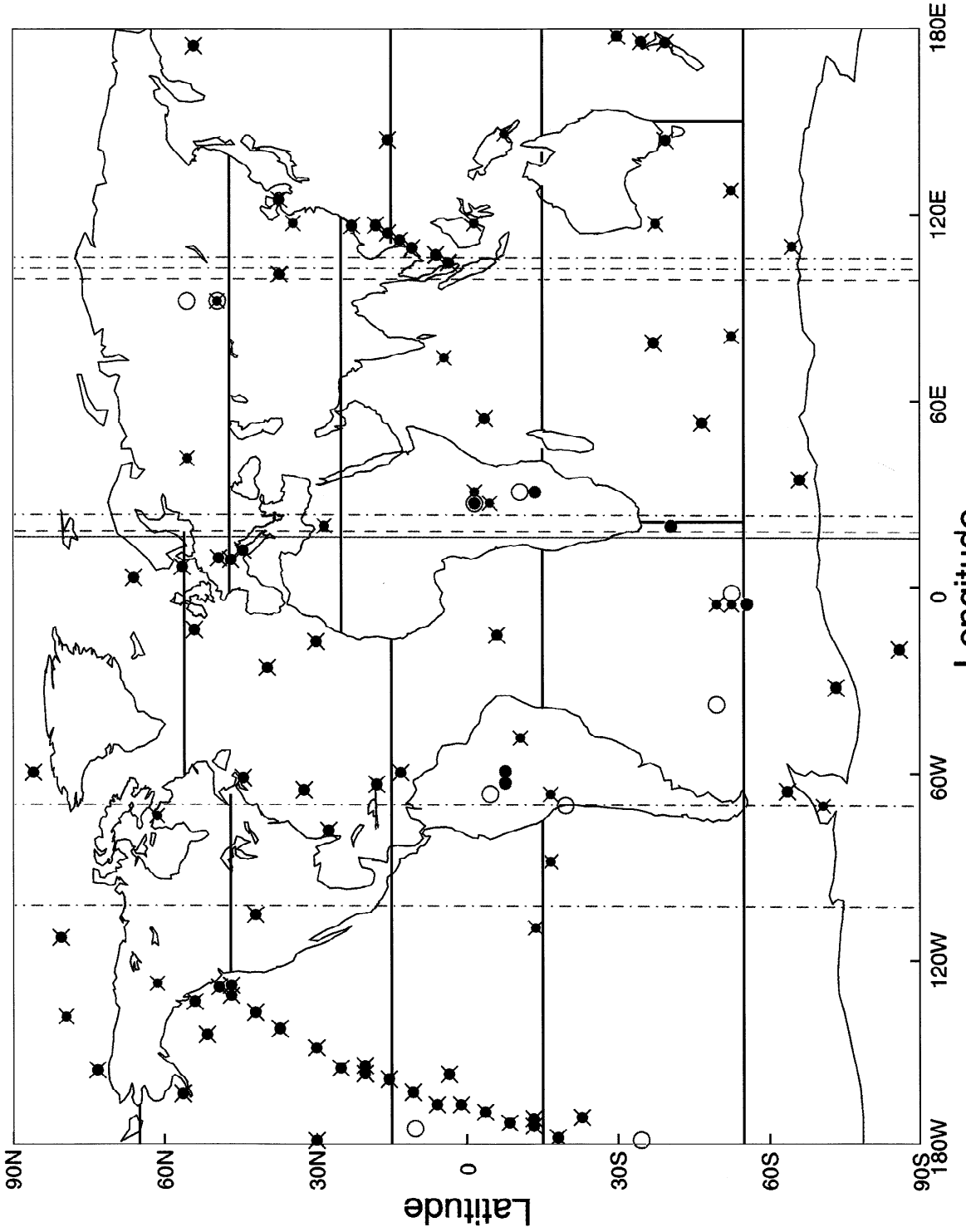

Plate 1. Locations of GLOBALVIEW-CO2 observation stations and (1) optimal surface stations in red (3 (dots), 5 (circles), and 12 (crossed dots) additional stations), (2) optimal profiles in dark blue (3 (dots), 5 (circles), and 12 (crossed dots) additional profiles), and (3) optimal N-S transects in green (1 (line), 2 (dashed lines), and 5 (dash-dotted lines) additonal transects) to minimize estimate uncertainty. Crossed black dots mark GLOBALVIEW-CO2 sites.

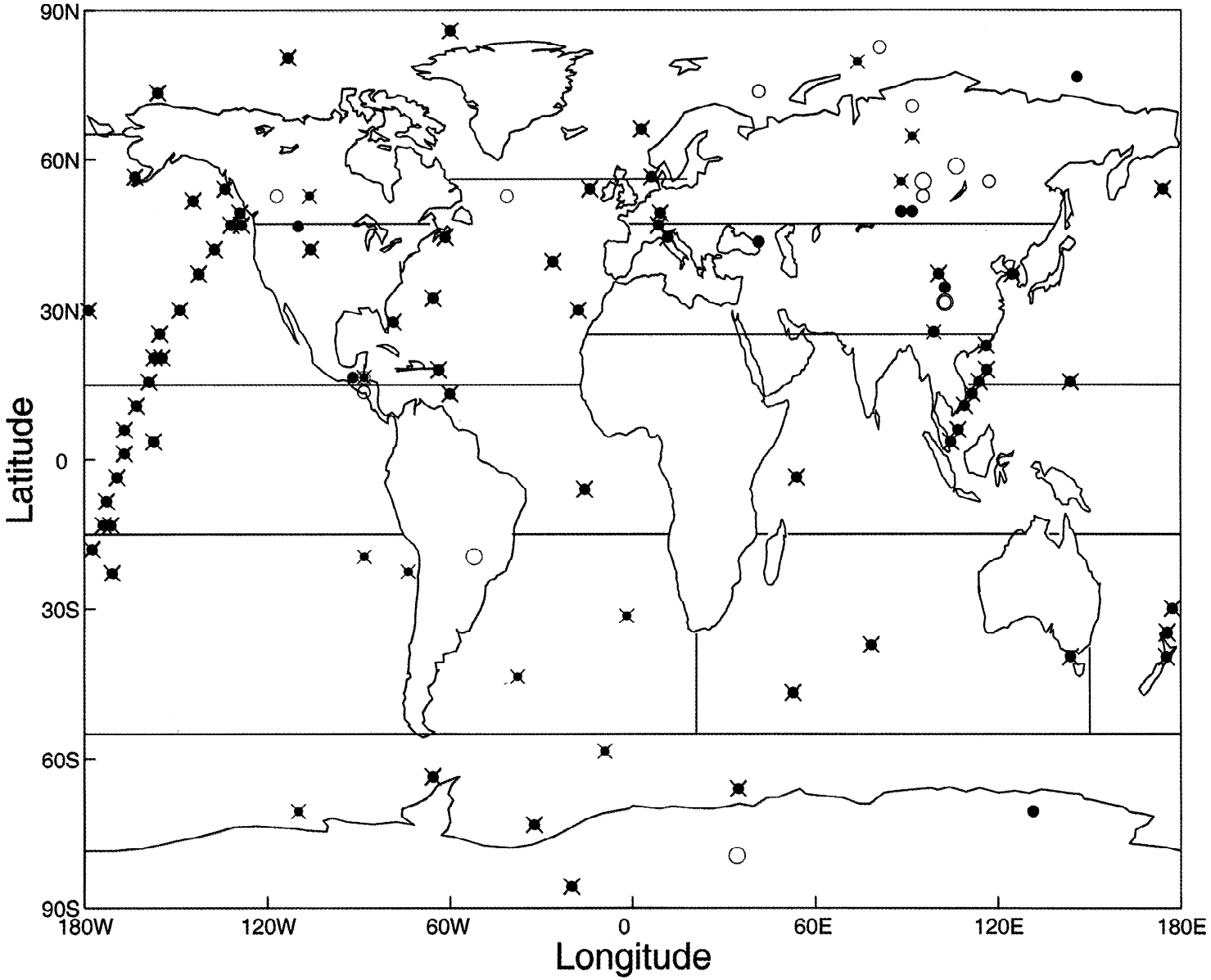

Plate 2. Locations of GLOBALVIEW-CO2 observation stations and optimal surface stations for the estimation of fluxes from (1) North America (green), (2) Eurasia (red), and (3) North America and Eurasia combined (dark blue) with minimal uncertainty. For North America and Eurasia 3 (dots) and 5 (circles) additional stations and for North America and Eurasia combined in addition 12 (crossed dots) stations are displayed. Crossed black dots mark GLOBALVIEW-CO2 sites.

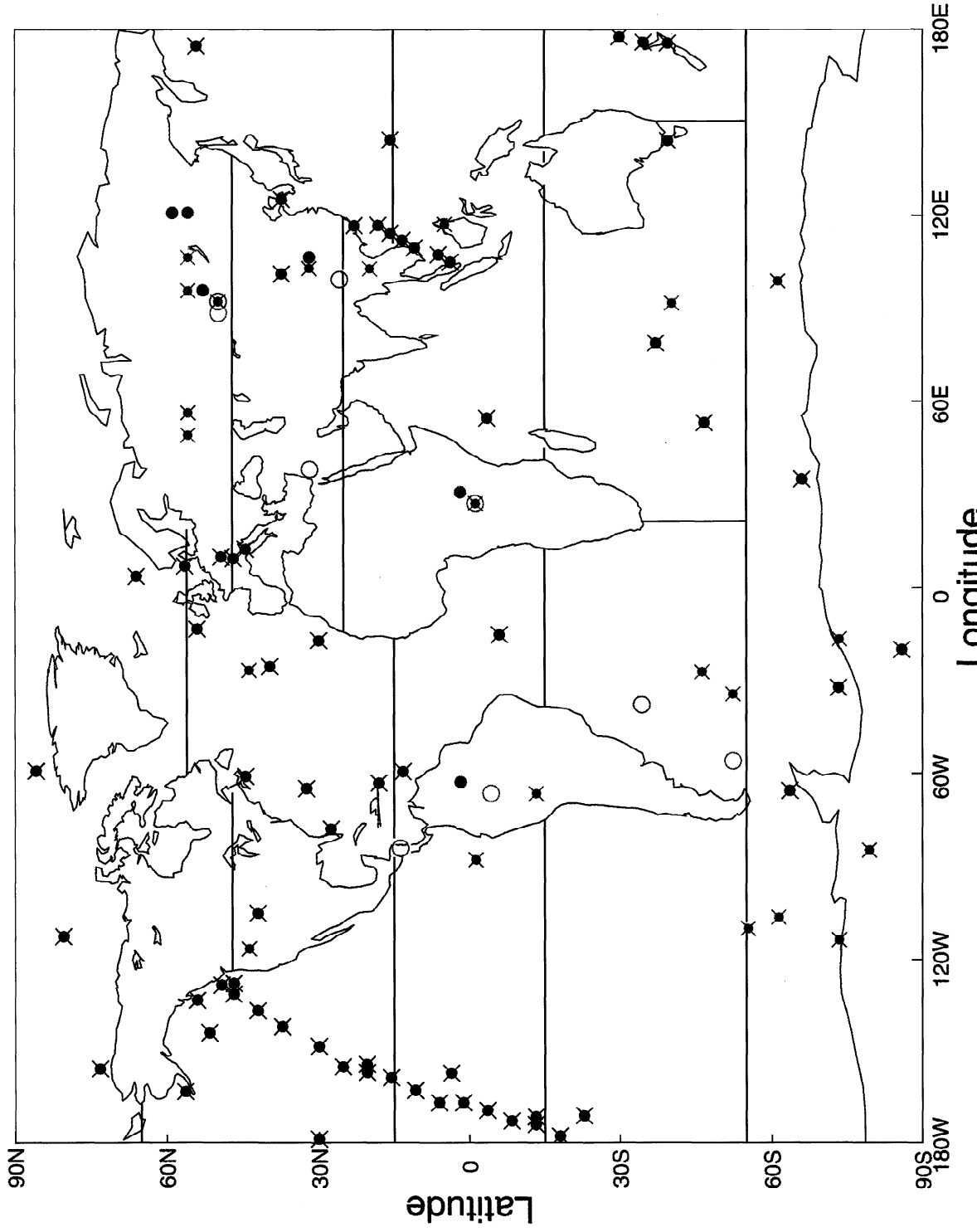

Plate 3. Same as Plate 2 for the estimation of fluxes from (1) all continents combined in green and (2) the Northern Hemisphere in red (in each case 3 (dots), 5 (circles), and 12 (crossed dots) additional stations) with minimal uncertainty. Crossed black dots mark GLOBALVIEW-CO2 sites.

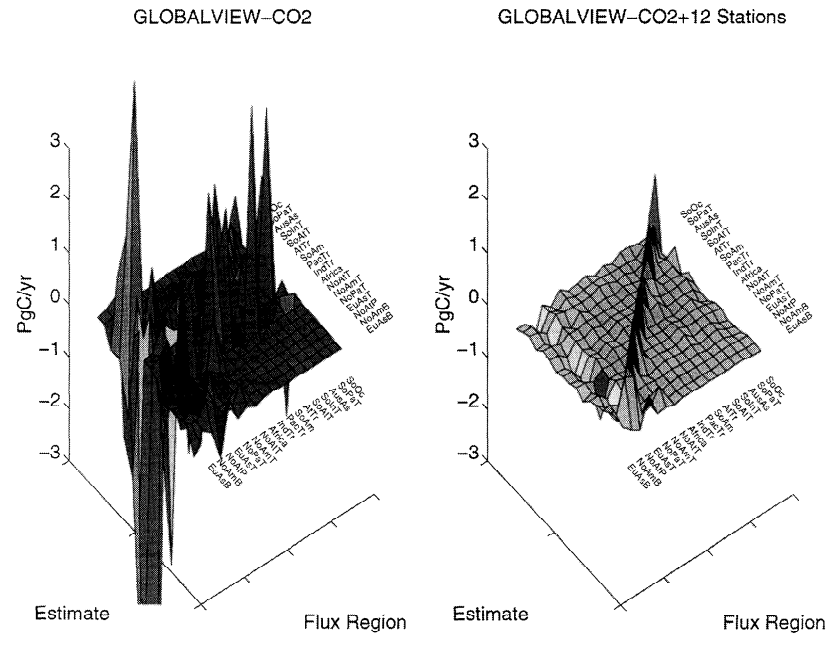

Figure 15. Flux estimates from the inversion of 17 different footprints simulated with SKYHI resulting from spatially uniform fluxes from each of the regions in Figure 1 separately. The footprints for the inversion are simulated with GCTM and use CASA biosphere flux patterns for the continents and air-sea partial pressure difference flux patterns for the oceans. Estimates are based on the observation stations of GLOBALVIEW-CO2 (left) and its optimal extension by 12 additional stations for minimal uncertainty of estimates.

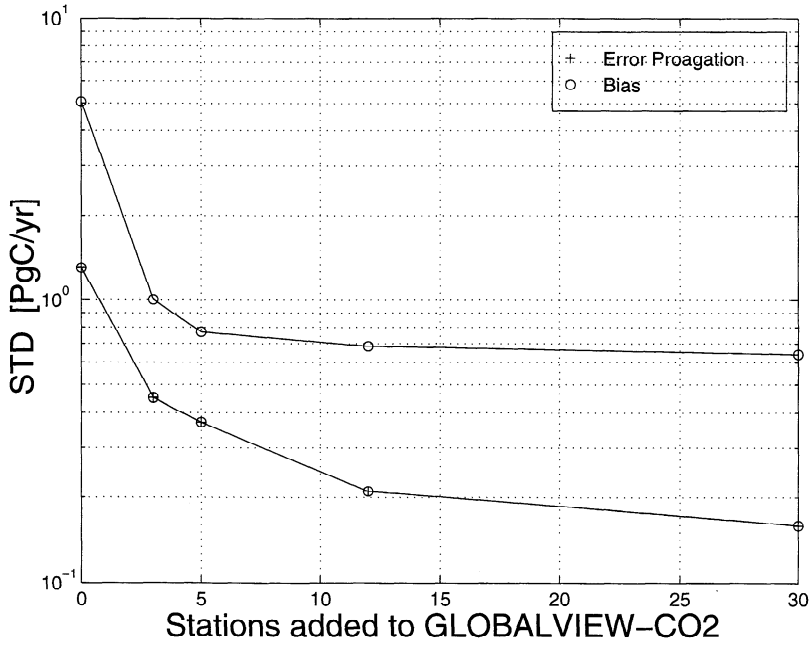

Figure 16. Comparison of the decrease of the standard deviation of estimates in Figure 15 from correct fluxes (σ_{Bias}, section 4.4) with additional optimal stations to GLOBALVIEW-CO2 with the decrease estimated by error propagation.

characterize the meridional distribution of CO_2, the sparse coverage of several biogeochemically critical areas of the globe leads to prohibitively large uncertainty in tracer-transport inversions (on average over 1 Pg C yr^{-1} $region^{-1}$ for the 17 regions in Figure 1). The most poorly constrained areas are (in order) (1) South America and Africa, (2) the Southern Hemisphere Oceans (especially the South Temperate Atlantic), and (3) Eurasia (especially outside of Europe). When optimizing the CMDL (Climate Monitoring and Diagnostics Laboratory, Boulder) network, a subset of the GLOBALVIEW-CO2 network, for the estimation of fluxes from and to the oceans, *Rayner et al.* [1996] similarly identified South America and Africa as regions that most urgently need coverage with observations. Note that their method differs from ours. They base their analysis on a Bayesian inversion approach that somewhat complicates the interpretation of their results.

Models of networks identify three partially antagonisitic principles that should guide the placement of new stations: (1) even coverage over all regions (2) proximity to the largest local signals, and (3) minimal high-frequency data variability. The optimization results indicate that (1) and (2) currently dominate significantly over (3) and strongly encourage to dare the "step into the continents" with profile measurements in general and for South America and Africa in particular. We find that three additional stations (profiles) to GLOBALVIEW-CO2 in these regions would decrease the mean error of estimates strongly (by a factor 3) and 12 additional vertical profiles or 30 optimal surface stations would reduce the mean error caused by error amplification to $\sim$ 0.2 Pg C yr^{-1}. Regions in greatest need of additional stations to those mentioned are Temperate South Atlantic, Central Eurasia, and the South Temperate Indian and Pacific Oceans.

Finally, we do not recommend airplane transects aloft ($\geq$ 3 km) as a strategy to complement GLOBALVIEW-CO2. Such signals are too diluted to be useful for inversions.

Acknowledgments. We would like to thank Bud Moxim for his help to run the Atmospheric Transport Model GCTM, Jerry Mahlman for helpful advice and Pieter Tans, Peter Rayner, and Lori Perliski for carefully reviewing the manuscript. This work has been supported by NASA (NAG5-3510) and the NOAA Office of Global Programs (NA56GP04-39).

References

Andres, R. J., G. Marland, I. Fung, and E. Matthews, A 1°x1° distribution of carbon dioxide emissions from fossil fuel consumption and cement manufacture, 1950-1990, *Global Biogeochem. Cycles, 10*, 419–429, 1996.

Conway, T. J., P. P. Tans, L. S. Waterman, K. W. Thoning, D. R. Kitzis, K. A. Masarie, and N. Zhang, 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*(D11), 22831–22855, 1994.

Fan, S., M. Gloor, J. Mahlman, S. Pacala, J. Sarmiento, T. Takahashi, and P. Tans, A large terrestrial carbon sink in North America implied by atmospheric and oceanic carbon dioxide data and models, *Science, 282*, 442–446, 1998.

Gloor, M., S.-M. Fan, S. W. Pacala, J. L. Sarmiento, and M. Ramonet, A model-based evaluation of inversions of atmospheric transport, using annual mean mixing ratios, as a tool to monitor fluxes of nonreactive trace substances like CO_2 on a continental scale, *J. Geophys. Res., 104*, 14245–14260, 1999.

Hamilton, K. P., J. Wilson, J. D. Mahlman, and L. M. Umscheid, Climatology of the SKYHI troposhere-stratosphere-mesosphere model, *J. Atmos. Sci., 48*, 651–678, 1995.

Keeling, C. D., S. C. Piper, and M. Heimann, A three dimensional model of atmospheric CO_2 transport based on observed winds, 4, Mean annual gradients and interannual variations, in *Aspects of Climate Variability in the Pacific and the Western Americas*, edited by D. H. Peterson, Geophys. Monogr. Ser., 55, pp. 305 363, AGU, Washington, D.C., 1989.

Kirkpatrick, S., J. C. D. Gelatt, and M. P. Vecchi, Optimization by simulated annealing, *Science, 220*, 671–680, 1983.

Levy II, H., and W. J. Moxim, Simulated global distribution and deposition of reactive nitrogen emitted by fossil fuel combustion, *Tellus, 41B*, 256–271, 1989.

Menke, W., *Geophysical Data Analysis: Discrete Inverse Theory*, Academic Press, San Diego, 1989.

Potter, C. S., J. T. Randerson, C. B. Field, P. A. Mason, P. M. Vitousek, H. A. Mooney, and S. A. Klooster, Terrestrial ecosystem production: A process model based on global satellite and surface data, *Global Biogeochem. Cycles, 7*, 811–841, 1993.

Press, W., S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes* (2nd ed.)., Cambridge University Press, Cambridge, 1992.

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.

Takahashi, T., R. A. Feely, R. Weiss, R. H. Wanninkhof, D. W. Chipman, S. C. Sutherland, and T. T. Takahashi, Global air-sea flux of CO_2: An estimate based on measurements of sea-air pCO_2 difference, in *Revelle Symposium: Proceedings of the Natl. Acad. Sci.*, edited by C. D. Keeling, vol. 94, pp. 8292 – 8299, Natl. Acad. Sci., Washington, D. C., 1997.

Tans, P., Observational Strategy for Assessing the Role of Terrestrial Ecosystems in the Global Carbon Cycle: Scaling Down to Regional Scales., in *Scaling Physiological Processes*, edited by J.P.Ehleringer and C.B.Field, Physiological Ecology, pp. 179–190, Academic Press, San Diego, 1993.

Tans, P. P., P. S. Bakwin, and D. W. Guenther, A feasible global carbon cycle observing system: a plan to decipher today's carbon cycle based on observations, *Global Ch. Biology, 2*, 309–318, 1996.

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

S. Fan, J. Sarmiento, Atmospheric and Oceanic Sciences Program, Princeton University, Sayre Hall, Forrestal Campus, P.O. Box CN710, Princeton, NJ 08544-0710. (fan@splash.princeton.edu, jls@splash.princeton.edu)

M. Gloor, Max-Planck-Institut für Biogeochemie, Postfach 100164, D-07701 Jena, Germany, (manuel.gloor@bgc-jena.mpg.de)

S. Pacala, Ecology and Evolutionary Biology, Princeton University, 101 Eno Hall, Princeton, NJ 08544, (steve@eno.princeton.edu)

(Received September 8, 1998; revised July 19, 1999; accepted July 27, 1999.)