JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 100, NO. D6, PAGES 11,593–11,610, JUNE 20, 1995

Extension and integration of atmospheric carbon dioxide data into a globally consistent measurement record

Kenneth A. Masarie
Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder

Pieter P. Tans
Climate Monitoring and Diagnostics Laboratory, National Oceanic and Atmospheric Administration, Boulder, Colorado

Abstract. Atmospheric transport models are used to constrain sources and sinks of carbon dioxide by requiring that the modeled spatial and temporal concentration patterns are consistent with the observations. Serious obstacles to this approach are the sparsity of sampling sites and the lack of temporal continuity among observations at different locations. A procedure is presented that attempts to extend the knowledge gained during a limited period of measurements beyond the period itself resulting in records containing measurement data and extrapolated and interpolated values. From limited measurements we can define trace gas climatologies that describe average seasonal cycles, trends, and changes in trends at individual sampling sites. A comparison of the site climatologies with a reference defined over a much longer period of time constitutes the framework used in the development of the data extension procedure. Two extension methods are described. The benchmark trend method uses a deseasonalized long-term trend from a single site as a reference to individual site climatologies. The latitude reference method utilizes measurements from many sites in constructing a reference to the climatologies. Both methods are evaluated and the advantages and limitations of each are discussed. Data extension is not based on any atmospheric models but entirely on the data themselves. The methods described here are relatively straightforward and reproducible and result in extended records that are model independent. The cooperative air sampling network maintained by the National Oceanic and Atmospheric Administration Climate Monitoring and Diagnostics Laboratory in Boulder, Colorado, provides a test bed for the development of the data extension method; we intend to integrate and extend CO_2 measurement records from other laboratories providing a globally consistent atmospheric CO_2 database to the modeling community.

1. Introduction

The monitoring of carbon dioxide (CO_2) in the atmosphere serves a twofold purpose. First, it documents the gradual global buildup of this gas due to man's activities. Toward this end, one or two remote measurement sites would be sufficient. Second, the measurements are used to try to better elucidate the operation of today's global carbon cycle. Success in the second endeavor will allow better predictions of future atmospheric levels of this greenhouse gas for given scenarios of fossil fuel burning. The ultimate goal of CO_2 monitoring therefore is to provide an assessment of some important large-scale and long-term environmental consequences of the burning of fossil fuels and deforestation. A major tool in the study of the contemporary carbon cycle is provided by the detailed temporal and large-scale spatial patterns of the mixing ratio of CO_2 and its stable isotopic composition in the atmosphere. These patterns reflect the location and strength of major sources and sinks as modified by the action of transport and mixing in the atmosphere. An essential component of this study is the use of numerical models of atmospheric transport and mixing to translate the CO_2 mixing ratio patterns into sources and sinks, invoking conservation of mass.

Two-dimensional (latitude, height) atmospheric transport models have been used [*Bolin and Keeling*, 1963; *Pearman and Hyson*, 1980; *Tans et al.*, 1989; *Enting and Mansbridge*, 1989, 1991] to deduce time-dependent surface sources and sinks of CO_2 from the observations. A major drawback is that in the real three-dimensional world there is considerable variation as a function of longitude that is entirely ignored in the two-dimensional approach. This leads to some misallocation of sources as a function of latitude by the two-dimensional model because the observations have thus far tended to be biased toward the marine boundary layer [*Tans et al.*, 1989].

Three-dimensional models have been used to explain the observed seasonal cycle of CO_2 at various places around the globe in terms of the patterns of photosynthesis and respiration of land plants [*Fung et al.*, 1983; *Heimann et al.*, 1989] and to deduce from the observed annual mean mixing ratios the annual balance of sources and sinks, i.e., net uptake (positive or negative) by large areas of the oceans, net uptake by ecosystems on land, and their locations [*Keeling et al.*; 1989, *Tans et al.*; 1990, *Law et al.*; 1992; *Yamazaki and Chiba*; 1993]. When using three-dimensional models it is very apparent that the observations are extremely sparse, even if we limit our questions

Paper number 95JD00859.
0148-0227/95/95JD-00859$05.00

to spatial scales as large as the entire expanse of boreal forest ecosystems.

A global CO_2 monitoring system with unprecedented spatial resolution and temporal continuity is needed. Since both the costs and the logistics of operating such a global system prove too formidable a task for a single organization, an international effort is required. Carbon dioxide is being monitored by an increasing number of laboratories around the world, with careful attention to the calibration and the methodology of the measurements. The World Meteorological Organization (WMO) has played for many years and continues to play a coordinating role in this activity through its Global Atmosphere Watch Programme (GAW) and its forerunner Baseline Air Pollution Monitoring Network (BAPMoN) program. Since 1991, the international monitoring of CO_2 is also one of the activities of International Global Atmospheric Chemistry (IGAC), a core project of the International Geosphere-Biosphere Program (IGBP). We are indeed witnessing the gradual emergence of a global carbon cycle observing system with the explicit purpose of using the integrating action of atmospheric mixing to detect the sources and sinks of CO_2 on very large regional scales.

Considered alone, high-precision measurements made by individual laboratories will continue to be sparse because of operational costs and logistical constraints. Considered together, an international measurement database will expand spatial and temporal coverage. Although the issue of sparse data is not eliminated, it is improved, and this is an immediate benefit to the modeling community. It is our aim to integrate as many measurements as possible from different laboratories into one global database with careful attention to direct compatibility with respect to the calibration and methodology. The observations from most of the sites are in the form of flask sample data where the measurements have been performed on discrete samples of whole air obtained at the site on a regular (e.g., weekly) basis. At a smaller number of sites, continuous measurements have been made. We would also like to incorporate air measurements of a more purely episodic nature, such as those performed on oceanographic research ships and during aircraft measurement campaigns.

The introduction of many measurement records into a transport model poses an additional problem. Data records tend to begin and end at different times. Nonsimultaneous records that "appear" or "disappear" to a model may be misinterpreted by the model. As an example, weekly measurements from the middle of the Baltic Sea were added fairly recently to the National Oceanic and Atmospheric Organization (NOAA), Climate Monitoring and Diagnostics Laboratory (CMDL), cooperative air sampling network. Especially during the winter season, CO_2 mixing ratios over the Baltic tend to be high compared to mid-oceanic sites at the same latitude. This is to be expected because of the relative proximity of the Baltic Sea to major regions of fossil fuel burning and terrestrial respiration. If in a model the measurements from the Baltic begin suddenly, the model would be forced to infer that fossil fuel burning or terrestrial respiration abruptly increased at that same moment in Europe, or equivalently, that some CO_2 uptake process suddenly diminished. We would like to have a way to extend the knowledge gained during the period when the Baltic measurements were active to other times when there were no data from the Baltic.

The problem of deducing sources and sinks of CO_2 from sparse observations, which include a component of noise as well, has been characterized as a mathematically ill-posed inverse problem. The potential for serious instability of the solutions increases as the time resolution and the spatial scale become smaller [*Newsam and Enting*, 1988]. To at least partially overcome that problem, we remove some of the high-frequency variability in the data by using smoothing filters in the time domain. Clearly, the transport model, with its sources and sinks, should not be required to reproduce every measured point exactly because then certainly measurement errors and "random" weather events at the site would be translated into enormous apparent significance in terms of sources and sinks. Likewise, the model should not be required to exactly reproduce the spatial differences of time-smoothed records of sites in relative proximity to each other, as this would certainly lead to the appearance of strong but unstable source-sink pairs. We think that the best strategy for the community to proceed is to try different versions of the inverse approach in practice, add other constraints and/or prior knowledge, and see how far it can be pushed, while being aware of the problems that have been pointed out by the more theoretical work [*Newsam and Enting*, 1988]. After all, this approach has already produced, with fewer data and on the largest spatial scales (hemispheric), one of the more significant findings of recent carbon cycle research, namely, the existence of a large terrestrial sink of CO_2 in the northern hemisphere [*Keeling et al.*, 1989; *Tans et al.*, 1990; *Ciais et al.*, 1995].

In this paper we propose a data integration and extension scheme designed to incorporate data from various measurement programs and laboratories and to extrapolate in time the knowledge gained during a limited period of measurements. At this stage we have used only the data from the CMDL cooperative air sampling network and from the continuous CO_2 measurement programs operating at the four CMDL observatories as our platform to develop the methods presented here. Use of the extended records (records containing measurement data and interpolated and extrapolated values) may already have an influence on possible biases introduced by earlier work into the derived trends because the locations of sites in the CMDL network have changed over time. We plan to incorporate data from other measurement programs as the next step, and we see this as a possible starting point for the integration and extension of other atmospheric data as well, such as CH_4, CO, and isotopic ratios. The technique proposed here is not based on any atmospheric models but entirely on the data themselves, so as not to bias the use of these data in different model applications. The extended CO_2 records will be made available as smoothed synchronized time series, with an indication of whether the values are based directly on measurements, or are interpolated or extrapolated from a period when no original data were available. The original data themselves remain available from data centers such as the WMO World Data Center for Greenhouse Gases in Tokyo and the Carbon Dioxide Information Analysis Center in Oak Ridge, Tennessee.

2. Cooperative Air Sampling Network

In 1968 the CMDL (formerly GMCC) Carbon Cycle Group began making measurements of atmospheric CO_2 from air collected in glass bottles [*Komhyr et al.*, 1985]. The group established what is now called the CMDL cooperative air

sampling network with the aim of documenting trends, changes in trends, and the global distribution of CO_2. The network is a globally distributed set of sampling sites that are established in cooperation with national and international organizations whose facilities and personnel are already present in regions of interest. Historically, sampling sites were established where air samples were representative of very large, well-mixed tropospheric air masses, without the influence of local CO_2 sources and sinks. This constraint placed the majority of sampling locations in remote marine locations. More recently, network expansion has included oceanic and continental locations in closer proximity to strong regional sources or sinks [*Conway et al.*, 1994]. In continental desert regions we employ the traditional weekly ground-based flask sampling method. In vegetated areas, other measurement techniques are being developed that employ very tall towers and the use of small aircraft. The addition of these sites to the network will help improve our understanding of regional source and sink strengths.

The CMDL network has experienced periods of rapid expansion and occasional attrition (Table 1). In late 1986, for example, sampling began in cooperation with Blue Star Line, Limited, aboard the container ship *Southland Star*, that sailed regularly between the U.S. west coast and Auckland, New Zealand. Air samples were collected at approximately 5°

Table 1. NOAA/CMDL Cooperative Air Sampling Network Sites Used in Data Extension Procedure

Site	Location	Latitude	Longitude	Elev., m[a]	Dates[b]	MBL[c]
ALT	Alert, Northwest Territories	82°27'N	62°31'W	210	1985-	y
AMS	Amsterdam Island	37°57'S	77°32'E	150	1979-1990	y
ASC	Ascension Island	07°55'S	14°25'W	54	1979-	y
AVI	St. Croix, Virgin Islands	17°45'N	64°45'W	3	1979-1990	y
AZR	Terceira Island, Azores	38°45'N	27°05'W	30	1979-	y
BAL	Baltic Sea (*Balanga Sister*)	55°30'N	16°40'E	7	1992-	n
BME	St. David's Head, Bermuda	32°22'N	64°39'W	30	1989-	y
BMW	Southhampton, Bermuda	32°16'N	64°53'W	30	1989-	y
BRW[d]	Barrow, Alaska	71°19'N	156°36'W	11	1971-	y
CBA	Cold Bay, Alaska	55°12'N	162°43'W	25	1978-	y
CGO	Cape Grim, Tasmania	40°41'S	144°41'E	94	1984-	y
CHR	Christmas Island	1°42'N	157°10'W	3	1984-	y
CMO	Cape Meares, Oregon	45°29'N	123°58'W	30	1982-	n
CRZ	Crozet, Indian Ocean	46°27'S	51°51'E	120	1991-	n
GMI	Guam, Mariana Islands	13°26'N	144°47'E	2	1978-	y
HBA	Halley Bay, Antarctica	75°40'S	25°30'W	10	1983-	y
ICE	Vestmannaeyjar, Iceland	63°15'N	20°09'W	100	1992-	y
IZO	Tenerife, Canary Islands	28°18'N	16°29'W	2300	1991-	n
KEY	Key Biscayne, Florida	25°40'N	80°12'W	3	1972-	y
KUM	Cape Kumukahi, Hawaii	19°31'N	154°49'W	3	1971-	y
MBC	Mould Bay, Northwest Territories	76°15'N	119°21'W	58	1980-	y
MHT	Mace Head, Ireland	53°20'N	9°54'W	25	1991-	y
MID	Sand Island, Midway	28°13'N	177°22'W	4	1985-	y
MLO[d]	Mauna Loa, Hawaii	19°32'N	155°35'W	3397	1969-	n
NWR	Niwot Ridge, Colorado	40°03'N	105°35'W	3475	1967-	n
OPC[e]	Pacific Ocean (*California Star*)			8	1993-	y
OPW	Olympic Peninsula, Washington	48°15'N	124°25'W	488	1984-1990	n
PAC[e]	Pacific Ocean (*Southland Star*)			15	1986-1993	y
PAW[e]	Pacific Ocean (*Wellington Star*)			7	1990-1993	y
PSA	Palmer Station, Antarctica	64°55'S	64°00'W	10	1978-	n
QPC	Qinghai Province, China	36°16'N	100°55'	3810	1990-	n
RPB	Ragged Point, Barbados	13°10'	59°26'W	3	1987-	y
SCS[f]	South China Sea (*Carla A Hills*)			15	1991-1993	n
SCS[f]	South China Sea (*Great Promise*)			15	1993-	n
SEY	Seychelles, Mahe Island	4°40'S	55°10'E	3	1980-	y
SHM	Shemya Island, Alaska	52°43'N	174°06'E	40	1985-	y
SMO[d]	Tutuila, American Samoa	14°15'S	170°34'W	42	1972-	y
SPO[d]	South Pole, Antarctica	89°59'S	24°48'W	2810	1975-	y
STM	Ocean Station "M"	66°00'N	2°00'E	7	1981-	y
SYO	Syowa, Antarctica	69°00'S	39°35'E	11	1986-	y
TAP	Tae-ahn Peninsula, Korea	36°44'N	126°08'E	20	1990-	n
UUM	Ulaan Uul, Mongolia	44°27'N	111°06'E	914	1992-	n

[a]Elevation in meters above mean sea level.

[b]Beginning and ending year of sampling.

[c]Marine boundary layer sites used in weekly latitude distributions.

[d]Flask and continuous CO_2 measurement programs treated as two sampling locations.

[e]Combined OPC, PAC, and PAW sites generates 17 effective shipboard sites.

[f]SCS generates seven effective shipboard sites.

latitude intervals from 45°N to 35°S with an average time interval of ~3 weeks for every 5° latitude position. This effort effectively added 17 new sampling locations in the Pacific Ocean and doubled the size of the network. During the 3-year period 1991-1993, 11 new sites, listed in Table 1, were added to the network. One of these, South China Sea (SCS), was a cooperative effort with Chevron Shipping Company aboard the refined product carrier *Carla A. Hills* that sailed regularly between Hong Kong and Singapore across the South China Sea (currently, the same route is being sampled in collaboration with the Nippon Yusen Kaisha Shipping Company). Air samples aboard these vessels were collected at ~3° latitude intervals from 3°N to 21°N with an average frequency of approximately one sample per week per 3° latitude position effectively adding seven new sampling locations. Sampling at St. Croix, Virgin Islands (AVI) was discontinued in 1990 concluding 10 years of successful cooperation with Fairleigh Dickinson University. Twelve years of sampling at Amsterdam Island (AMS), in collaboration with the French Centre des Faibles Radioactivités, also ended in 1990. Figure 1 shows the evolution of network sampling density. Periods of expansion and attrition are evident as well as occasional gaps within individual sampling site records.

3. Data Preparation

Carbon dioxide data from 37 land-based flask sampling sites and 24 shipboard positions of the CMDL cooperative air sampling network and from the CO_2 continuous sampling programs at the four CMDL baseline observatories were used in the extension procedure (Table 1). To characterize an average seasonal cycle, we required sampling records be at least 2 years in length; shorter records were excluded from the extension procedure. Techniques for sample collection and analysis and data editing and selection have been described in detail for the flask program [*Conway et al.*, 1988, 1994] and the continuous programs [*Peterson et al.*, 1986; *Gillette et al.*, 1987; *Komhyr et al.*, 1989; *Thoning et al.*, 1989; *Waterman et al.*, 1989]. Briefly, both measurement programs use nondispersive infrared gas analyzers. All CO_2 mixing ratios are determined relative to standards traceable to the WMO X85 mole fraction scale and maintained at the WMO Central CO_2 Laboratory at the Scripps Institution of Oceanography. Both flask and continuous data have been selected to ensure that samples are very likely to be representative of large well-mixed air masses. Flask samples are collected in pairs approximately weekly, and pairs retained by the selection process are averaged. Figure 2 shows retained flask averages at Shemya Island, Alaska (SHM) and Cape Grim, Tasmania (CGO). At the CMDL observatories, measurements are made every minute and reported as hourly averages; for this work, selected daily averages were used. Figure 3 shows daily averages from the Barrow, Alaska (BRW) and South Pole, Antarctica (SPO) continuous measurement programs.

It is clear from Figure 1 that flask sampling does not occur at evenly spaced intervals. Designated sampling days may be shifted by one or more days because of inclement weather or unacceptable sampling conditions. Sampling gaps greater than 2 weeks occasionally occur because of sampling problems, site personnel changes, or insufficient flask supplies. Because of the sampling variability at individual sites, sampling among CMDL sites is not synchronized. To ensure that the global database can be used easily and unambiguously by the modeling community, we require that integrated and extended records have evenly spaced, synchronized time steps. To achieve this, we first fit a smooth curve to the measurements at each site and then extract values from the curve at evenly spaced synchronized steps. The

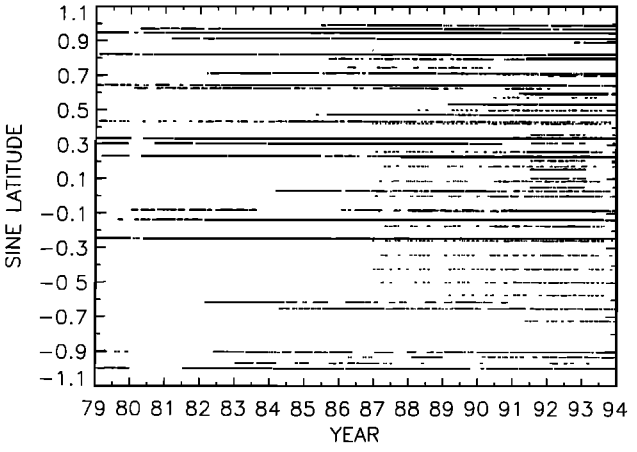

Figure 1. The density and distribution of air samples from the Climate Monitoring and Diagnostics Laboratory (CMDL) cooperative air sampling network plotted as sine of latitude versus time. Each solid circle indicates an air sample resulting in a CO_2 measurement believed to be representative of a large, well-mixed air mass.

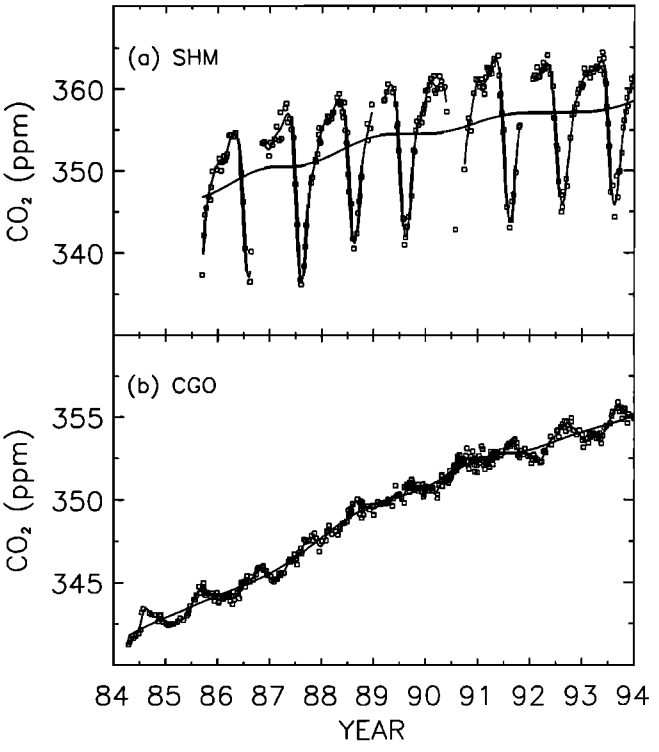

Figure 2. Carbon dioxide mixing ratios determined from flask samples at (a) Shemya Island, Alaska (SHM) and (b) Cape Grim, Tasmania (CGO). Squares are values believed to be representative of large, well-mixed air masses. The smooth curve, $S_{STA}(t)$, and deseasonalized long-term trend, $T_{STA}(t)$, are shown as solid curves (see section 3). Internal gaps greater than 8 weeks are denoted by breaks in the smooth curve.

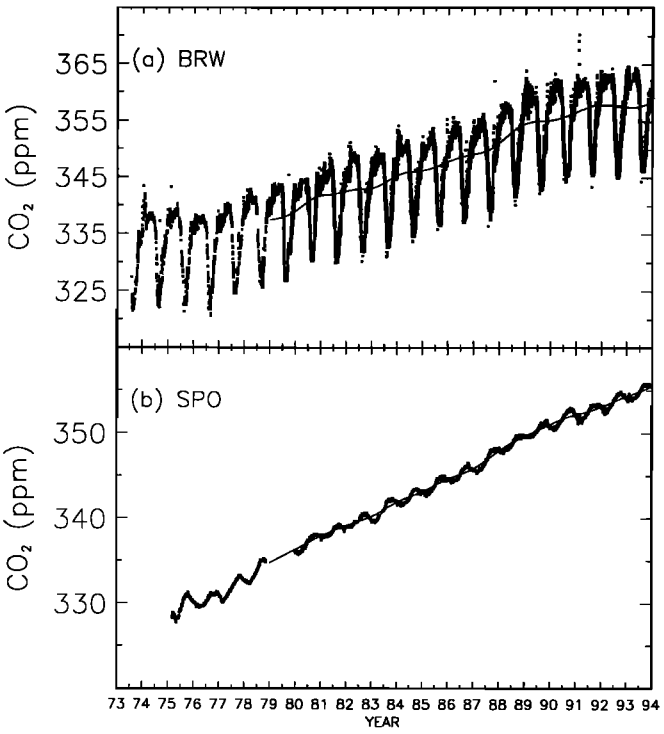

Figure 3. Daily averaged CO_2 mixing ratios determined from continuous measurements made at (a) the CMDL Barrow Observatory and (b) the South Pole Observatory. Data have been selected for background conditions. The solid curves are smooth curves and deseasonalized long-term trends fitted to the entire record (see section 3). Curves shown do not extend beyond the synchronization period, 1979-1994.

use of smoothed rather than actual data also diminishes the tendency for inverse models to produce unstable source/sink scenarios on short timescales.

The curve fitting techniques used to smooth a CO_2 measurement record, $c_{STA}(t)$, where the subscript notation "STA" (for station) indicates that the expression is specific to any one of the sampling sites listed in Table 1, have been described by *Thoning et al.* [1989]. We briefly describe the techniques here because they are used extensively in both data extension methods. To approximate the long-term trend and average seasonal cycle at a sampling site, a function of the form

$$f_{STA}(t) = a_0 + a_1 t + a_2 t^2 + \Sigma_{k=1,4}[b_{2k-1}\sin(2\pi kt) + b_{2k}\cos(2\pi kt)] \quad (1)$$

is fitted to the measurements where t is the time in years since January 1, 1979. To account for interannual variability in the seasonal cycle, the residuals, $r_{STA}(t)=c_{STA}(t)-f_{STA}(t)$, are digitally filtered through a low-pass filter with a full width at half maximum (FWHM) of approximately 40 days. The smoothed residuals from the 40-day filter, $\{r_{STA}(t)\}_{40d}$, are then combined with $f_{STA}(t)$ to produce what we call the smooth curve,

$$S_{STA}(t) = f_{STA}(t) + \{r_{STA}(t)\}_{40d}. \quad (2)$$

To account for interannual variability in the long-term trend, the residuals can be digitally filtered through a low-pass filter with FWHM of approximately 1 year. The result from the 1-year filter, $\{r_{STA}(t)\}_{1\,yr}$, is then combined with the polynomial terms in $f_{STA}(t)$ to produce the deseasonalized trend

$$T_{STA}(t) = a_0 + a_1 t + a_2 t^2 + \{r_{STA}(t)\}_{1\,yr}. \quad (3)$$

Figures 2 and 3 show the measurements $c_{STA}(t)$, the smooth curve $S_{STA}(t)$, and the long-term trend $T_{STA}(t)$ at SHM, CGO, BRW, and SPO. Table 2 summarizes the expressions used for general curve fitting.

From the smooth curve we extracted evenly spaced CO_2 values at synchronized time steps. For this work we define a synchronization period from January 1, 1979, to January 1, 1994, with a resolution of 48 equal intervals per year (referred to as weeks, where 1 week ≅7.6 days) resulting in a time period with 721 equally spaced steps. For each sampling site, values were extracted from the smooth curve at synchronized time steps following the first air sample and continuing to the last sample. We did not extract values from the curve where measurements extend beyond the synchronization period. External gaps in the synchronization period, denoted by default values (-999.999), occur when a site record begins or ends within the period. Internal gaps in the synchronization period, also denoted by default values, occur when the time between successive samples exceeds 8 weeks (confidence in the interpolated smooth curve values diminishes as gaps between successive samples increase beyond this 2-month cutoff). Internal and external gaps in the SHM record are apparent in Figure 2a. These default assignments denote weekly time steps within the synchronization period that will later be filled with either extrapolated or interpolated CO_2 mixing ratios derived from the data extension procedure.

Our confidence in values extracted from $S_{STA}(t)$ depends on the density of the data, the "scatter" in the data, and the length of the measurement period. A relative weighting scheme was designed to enable a user to give greater significance to sites with high signal to noise and/or consistent sampling. Weights derived from this scheme are assigned to extracted values from $S_{STA}(t)$ at each time step. Weights at individual sites were calculated for each year as follows:

$$w_{STA}(t) = \frac{\sqrt{N}}{RSD}$$

where RSD is the residual standard deviation of the measurements about the smoothed curve, N is the number of residuals per year used in the RSD determination, and $w_{STA}(t)$ is the calculated weight at each time step in the synchronization period. The residual standard deviations were determined with annual resolution, so that weekly weights, $w_{STA}(t_i)$, assigned within each year have the same value. We placed an upper limit on N of 96 or twice-weekly samples per year because we believed that higher sampling frequencies, such as the daily averages from the baseline observatories, often do not represent independent measurements. At Mauna Loa, Hawaii, (MLO) for instance, consecutive selected daily mean values are related to each other through the presence of large-scale weather patterns [e.g., *Heimann et al.*, 1989]. To ensure that the range of calculated weights for all sites and years was not skewed by site records with exceptionally high or low noise, we clipped all weights greater than the 70 percentile at the value of the 70

Table 2. Summary of Equations and Expressions Used in the Data Extension Procedure

Expression	Description
	General Curve Fitting
$c_{STA}(t)$	selected measurements, representative of large well mixed air masses
$f_{STA}(t) = a_0 + a_1 t + a_2 t^2 + \Sigma_{k=1,4}[b_{2k-1}\sin(2\pi kt) + b_{2k}\cos(2\pi kt)]$	(1) site climatology of $c_{STA}(t)$, approximates average seasonal cycle and trend
$S_{STA}(t) = f_{STA}(t) + \{r_{STA}(t)\}_{40d}$	(2) smooth curve fitted to $c_{STA}(t)$
$T_{STA}(t) = a_0 + a_1 t + a_2 t^2 + \{r_{STA}(t)\}_{1yr}$	(3) deseasonalized long-term trend fitted to $c_{STA}(t)$
	Benchmark Trend Method
$\Delta_{STA,T}(t) = S_{STA}(t) - T_{SPO}(t)$	(4) difference data, seasonal cycle and trend relative to reference (SPO in this case) trend
$d_{STA,T}(t) = m_0^T + \Sigma_{k=1,4}[m_{2k-1}^T \sin(2\pi kt) + m_{2k}^T \cos(2\pi kt)]$	(5) difference climatology of $\Delta_{STA,T}(t)$, approximately average seasonal cycle and offset relative to reference.
$S_{STA,T}(t) = d_{STA,T}(t) + \{r_{STA,T}(t)\}_{40d}$	(6) smooth curve fitted to $\Delta_{STA,T}(t)$
$E_{STA,BT}(t) = T_{SPO}(t) + \{\Delta_{STA,T}(t), S_{STA,T}(t)\}$	(7) extended site record, contains smoothed, extrapolated, and interpolated values
	Latitude Reference Method
$REF_{STA}(t)$	latitude reference time series, extracted from weekly latitude gradient curves at sine (latitude) of STA
$\Delta_{STA,REF}(t) = S_{STA}(t) - REF_{STA}(t)$	(8) difference data, seasonal cycle and trend anomalies relative to latitude reference
$d_{STA,REF}(t) = m_0^R + \Sigma_{k=1,2}[m_{2k-1}^R \sin(2\pi kt) + m_{2k}^R \cos(2\pi kt)]$	(9) difference climatology of $\Delta_{STA,REF}(t)$, approximatley average seasonal cycle and offset relative to reference
$S_{STA,REF}(t) = d_{STA,REF}(t) + \{r_{STA,REF}(t)\}_{40d}$	(10) smooth curve fitted to $\Delta_{STA,REF}(t)$
$E_{STA,LR}(t) = REF_{STA}(t) + \{\Delta_{STA,REF}(t), S_{STA,REF}(t)\}$	(11) extended site record, contains smoothed, extrapolated, and interpolated values

percentile and clipped all weights less than 5 percentile at the value of the 5 percentile. Lastly, we scaled the weights as follows:

$$w^*_{STA}(t) = w_{STA}(t) \bullet \frac{2.0}{w_{STA}(t)|_{5\ percentile}}$$

where $w_{STA}(t)|_{5\ percentile}$ is the weight at the 5 percentile (the minimum value of all weights) and $w^*_{STA}(t)$ is the resulting scaled weight at each time step. The scaled weights, $w^*_{STA}(t)$, assigned to smoothed data range from the fixed minimum weight of 2 to a maximum weight of approximately 12. For each site used in the data extension procedure, we have created a four-column matrix containing the synchronization years, the RSD of the measurements about the smooth curve, $S_{STA}(t)$, the number of residuals per year used in the RSD determination (N), and the scaled weights ($w^*_{STA}(t)$). Assigned default weights of value 1 indicate years where no measurements exist. Each matrix is saved to a file. Table 3 shows a portion of the weight matrix for SHM. These matrices serve two purposes. First, the scaled weights are used in the development of the latitude reference method described in section 4.2, and second, the matrices may be used as additional model input to indicate the relative confidence in each value of an extended record (see section 5). The resulting relative weighting scheme ensures that flask records with high signal to noise are weighted similarly to continuous records with high sampling frequency. As examples, both the flask and the continuous records at SPO have identical maximum weights determined for all years, while the flask and continuous records at BRW have similar weights to each other but lower than SPO.

In addition to the weighting matrix describe above, preparation of the measurement data creates a matrix for each of the 65 sites used in the data extension procedure. Each matrix has the same number of rows and two columns. The number of rows (721) is defined by the synchronization period and its "weekly" resolution. Matrix columns contain the synchronized time steps and the smoothed values from the curve, $S_{STA}(t)$. These matrices are used as the starting point in the development of both extension methods. Specifically, synchronized smoothed values, $S_{STA}(t)$, which represent the measurements from each site record, are used to define difference climatologies discussed in section 4.

Table 3. Portion of the Weight Matrix for SHM

Year, GMT[a]	RSD[b]	Number[c]	Weight[d]
1979.000000	-999.999	0.000	1.000
1980.000000	-999.999	0.000	1.000
1981.000000	-999.999	0.000	1.000
1982.000000	-999.999	0.000	1.000
1983.000000	-999.999	0.000	1.000
1984.000000	-999.999	0.000	1.000
1985.000000	0.940	9.000	2.449
1986.000000	0.628	21.000	5.601
1987.000000	0.889	30.000	4.729
1988.000000	0.859	33.000	5.133
1989.000000	0.709	28.000	5.728
1990.000000	0.793	24.000	4.741
1991.000000	0.885	29.000	4.670
1992.000000	0.763	37.000	6.119
1993.000000	0.836	41.000	5.879
1994.000000	0.860	20.000	3.991

The first 15 lines of informational text have been omitted. For details, refer to section 3.

[a]Synchronization year.

[b]Residual standard deviation of the measurements about the smooth curve.

[c]Number of residuals per year used in the RSD determination.

[d]Scaled weight (w^*_{SHM}).

4. Technique

The data extension methods described are not based on any atmospheric models but entirely on the data themselves. In contrast to model-based data assimilation techniques based on mass conservation, the methods described here are relatively straightforward and reproducible and result in extended records that are model independent. Data extension attempts to transfer knowledge gained during a limited period of measurements beyond the period itself. The knowledge manifests itself in what we call a trace gas site climatology that describes characteristic features in a data record such as average seasonal cycle patterns, trends, and changes in trends. From section 3 the function $f_{STA}(t)$ describes the trace gas site climatology for the measurement period at each CMDL sampling site. This function, however, may not be a reasonable description where there are no measurements. Imagine, for example, a 4-year measurement record in the high northern hemisphere above 60° latitude for the period 1987 through 1990, the site climatology derived from this 4-year record would fail to describe the decrease in the CO_2 growth rate observed globally in 1991-1992 [*Conway et al.*, 1994]. If we reference this fictitious site climatology with measurements from another site or many other sites from the CMDL network where we continue to make measurements, then perhaps we can combine knowledge of the slowdown in the CO_2 growth rate from observations at active sites with the site climatology to more accurately extend the record beyond the 4-year measurement period. The first data extension method described uses as a benchmark the trend, $T_{STA}(t)$, from a single CMDL sampling site as a reference. The second method employs a latitude reference time series derived using measurements from many sampling locations. Measurements from SHM and CGO are used to illustrate the progression of the data extension procedure for each of the two methods described.

4.1. Benchmark Trend Method

A first attempt at data extension resulted in the benchmark trend (BT) method. This method extends site records forward and backward in time by utilizing knowledge derived from individual site climatologies and information contained in a single benchmark long-term trend. Figures 4a and 5a (solid circles) show the smooth curve values defined by $S_{STA}(t)$ for SHM and CGO, respectively. To keep the method simple, we selected a network site whose long-term trend would be generally representative of a global long-term trend. The benchmark site record was required to span the synchronization period with few or no internal gaps, have high signal to noise, and be representative of very large well-mixed air masses. The SPO flask record was selected as the benchmark site.

A long-term trend, $T_{STA}(t)$, was determined for SPO (equation (3)). Trend values were then extracted at each synchronized step. Since deseasonalized long-term trend curves are minimally affected by internal gaps in site records, interpolated trend values were included resulting in a

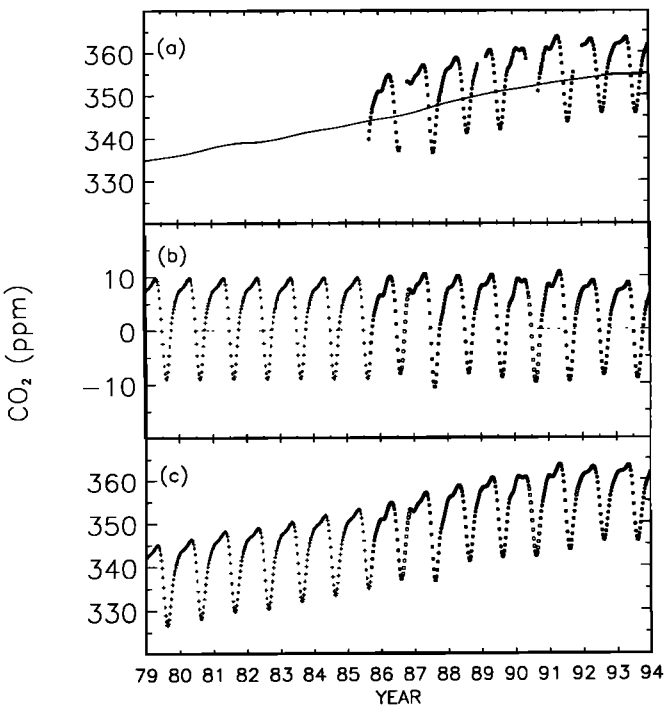

Figure 4. Progression of the data extension procedure using the benchmark trend method. (a) Weekly synchronized smooth curve values, $S_{STA}(t)$, from SHM (solid circles) and the long-term trend, $T_{STA}(t)$, derived from the benchmark site South Pole, Antarctica (SPO) (solid curve); (b) differences of smoothed values minus benchmark trend (BT) values, $\Delta_{STA,T}(t)$ (solid circles), used to define the difference climatology and interpolated differences (open circles) and extrapolated differences (pluses) defined using the difference climatology; (c) extended site record, $E_{STA,BT}(t)$, containing smoothed values (solid circles) and interpolated and extrapolated values (open circles and pluses). See section 4.1 for details.

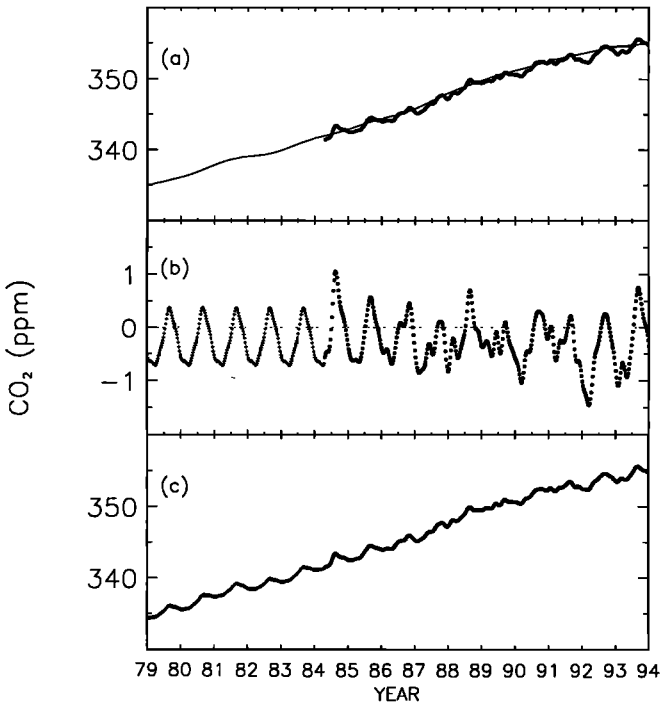

Figure 5. Same as Figure 4 but for CGO.

benchmark trend value at each step in the synchronization period. Figures 4a and 5a (solid curve) show the trend, $T_{SPO}(t)$, used in the BT method.

For each site, smoothed minus benchmark trend mixing ratios were calculated at each time step containing smoothed values,

$$\Delta_{STA,T}(t) = S_{STA}(t) - T_{SPO}(t). \quad (4)$$

The difference vector, $\Delta_{STA,T}(t)$, highlights the seasonal cycle patterns and trends at each site relative to the long-term trend at SPO. A difference climatology was then described by fitting a function to $\Delta_{STA,T}(t)$. In this curve fit, equation (1) was modified so that the average seasonal cycle and offset for the entire site record relative to the benchmark trend were approximated by the function

$$d_{STA,T}(t) = m_0^T + \Sigma_{k=1,4}\left[m_{2k-1}^T \sin(2\pi kt) + m_{2k}^T \cos(2\pi kt)\right]. \quad (5)$$

The difference climatology, $d_{STA,T}(t)$, describes the average difference between the smooth curve and the benchmark trend everywhere that actual measurements exist. As described in section 3, we can filter the residuals, $r_{STA,T}(t) = \Delta_{STA,T}(t) - d_{STA,T}(t)$, using a 40-day FWHM filter and combine the smoothed residuals with equation (5) to produce a smooth curve, $S_{STA,T}(t)$, to the difference vector,

$$S_{STA,T}(t) = d_{STA,T}(t) + \{r_{STA,T}(t)\}_{40d}. \quad (6)$$

Data extension relies on the assumption that the difference climatology described by equation (5) is valid for periods where there are no measurements (we will discuss the limitations of this assumption in section 5). On the basis of this assumption, time steps within external gaps in the synchronization period were filled using extrapolated CO_2 values determined from equation (6) where the term $\{r_{STA,T}(t)\}_{40d}$ is zero. Time steps within internal gaps in the synchronization period were filled using interpolated values from equation (6) where the contribution from the smoothed filtered residuals is nonzero. This was accomplished by first constructing residuals, $r_{STA,T}(t)$, at time steps within an internal gap by interpolating between the two residual values bounding the gap and then including these constructed residuals in the 40-day filter. The BT method results in a composite difference vector consisting of $\Delta_{STA,T}(t)$ where smooth data exist and of extrapolated and interpolated values from $S_{STA,T}(t)$ where smooth data do not exist. Figures 4b and 5b show the difference vectors for the sampling sites SHM and CGO.

Data extension using the BT method creates a matrix for each of the 65 sites used. As before, the number of rows is defined by the synchronization time period and its resolution. The first two columns are from the data preparation matrix (i.e., synchronized time steps and values from the smooth curve, $S_{STA}(t)$); two additional columns contain the $T_{SPO}(t)$ vector and the difference vector, $\{\Delta_{STA,T}(t), S_{STA,T}(t)\}$. Each matrix is saved to a file. Table 4 shows a portion of the BT method matrix for SHM.

Finally, an extended CO_2 site record, $E_{STA,BT}(t)$, containing smoothed, extrapolated, and interpolated mixing ratios can be generated by combining the two column vectors containing the benchmark trend and differences as follows:

$$E_{STA,BT}(t) = T_{SPO}(t) + \{\Delta_{STA,T}(t), S_{STA,T}(t)\} \quad (7)$$

When data are present, the above expression returns the original smoothed values, i.e., $E_{STA,BT}(t) = S_{STA}(t)$. When data are not present, the extended CO_2 mixing ratio is a combination of the globally representative benchmark trend, an average offset over the entire site record relative to the benchmark trend and an average seasonal cycle. Figures 4c and 5c show the extended site records resulting from the BT data extension method for SHM and CGO. At both sites, extrapolated CO_2 values filling external gaps clearly show the combination of the benchmark trend and the average seasonal cycle. Table 2 summarizes the expressions used in the BT method.

Table 4. Portion of the SHM Result Matrix Derived From the Benchmark Trend Method

Date, GMT[a]	$S_{STA}(t)$[b]	$T_{SPO}(t)$[c]	$\{\Delta_{STA,T}(t), S_{STA,T}(t)\}$[d]
1979.000000	-999.999	334.885	7.226
1979.020874	-999.999	334.913	7.076
.	.	.	.
.	.	.	.
.	.	.	.
1985.666626	-999.999	343.961	-7.125
1985.687500	-999.999	343.991	-6.400
1985.708374	339.792	344.020	-4.228
1985.729126	341.663	344.049	-2.386
.	.	.	.
.	.	.	.
.	.	.	.
1993.979126	361.324	355.127	6.197
1994.000000	361.766	355.150	6.616

The first 15 lines of informational text have been omitted. For details, refer to section 4.1.

[a]Synchronized time steps.

[b]Smooth values from curve fit.

[c]SPO benchmark trend.

[d]Difference vector.

On the basis of our benchmark criteria the MLO flask record could have also been used as a benchmark site. As an exercise, the BT method was implemented once using $T_{SPO}(t)$ and once using $T_{MLO}(t)$. Comparison of the extended records using the two benchmark sites showed that either would be acceptable as the benchmark site. Small differences in the results using the BT method at SHM and CGO that were observed using either SPO or MLO were due to differences in the long-term trend at each benchmark site. Should one benchmark site be preferred? In Figure 6 we compare the trends derived from extrapolated values at SHM (which started during 1985) using SPO as a benchmark site (dashed curve) and using MLO as a benchmark site (dashed-dotted curve) with the long-term trend from smoothed values at Cold Bay, Alaska (CBA) which is nearby in latitude (solid curve) from January 1979 to September 1985. As expected, trends derived from extrapolated values agree better with trends observed at nearby sites when the benchmark site is closer in latitude. This alone cannot account for the differences shown in Figure 6. From 1985 to 1993 the average difference of the annual means at MLO and SPO was greater than the average difference prior to 1985. As a result, the difference climatology derived from the period 1985 to 1993 does not adequately describe the climatology prior to 1985. In this instance the SHM trend derived using SPO as the benchmark site is offset toward higher CO_2 values from the SHM trend derived using MLO as the benchmark site.

4.2. Latitude Reference Method

The benchmark trend approach to data extension does not capture interannual variability in the seasonal cycle that is apparent from the smoothed values in Figures 4a and 5a. We compared the SHM extended record, $E_{SHM,BT}(t)$, from 1979 to 1986 with the smoothed record from CBA, $S_{CBA}(t)$, at a similar latitude and found that the CBA record shows significant interannual variability in both the shape and the amplitude of the seasonal cycle (Figures 7a and 7b). Variability of this nature is observed at all sampling locations and is representative of the year-to-year differences in regional source/sink strengths and atmospheric transport. Knowledge of atmospheric CO_2 variability recorded at sites nearby in latitude was essentially ignored in the BT method that used the long-term trend at a single benchmark site as a reference to each site climatology. Clearly, this method underutilized valuable information presented by the many CO_2 measurements available from the CMDL network. We know, for instance, that the annual mean CO_2 gradient from Antarctica to the Arctic was ≈25% greater during 1988-1991 than during the preceding seven years. We would like to incorporate such knowledge into the extended records but cannot do so by using the trend from a single benchmark site. In the latitude reference method we have refined the data extension procedure so that extrapolated and interpolated values at individual sites incorporate knowledge of CO_2 variability at sites nearby in latitude.

Making use of the extensive latitudinal distribution of the CMDL network has allowed us to narrow the spatial extent of the reference used with each site climatology. The latitude reference (LR) method considers measurements from many network sampling records when constructing a difference climatology. Specifically, difference climatologies are defined in terms of smoothed values from individual site records, $S_{STA}(t)$, and both the trend and the seasonal cycle patterns from a reference time series defined at the latitude of STA. The

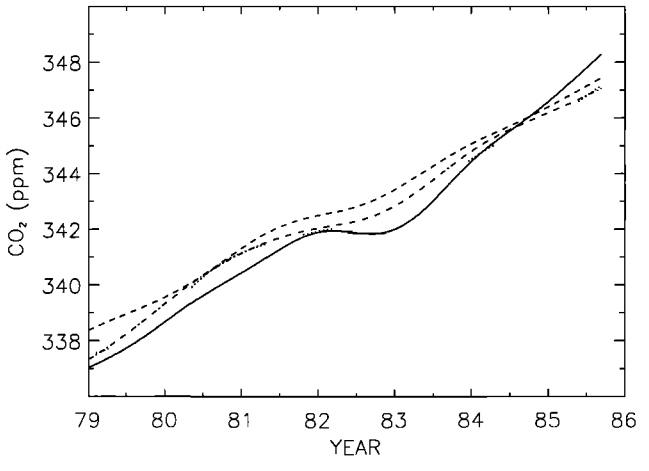

Figure 6. Comparison of deseasonalized long-term trends, $T_{STA}(t)$, derived from extrapolated values at SHM with the trend derived from smooth curve values at Cold Bay, Alaska (CBA) which is nearby in latitude. The solid curve is the long-term trend derived from smooth curve values at the comparison site CBA. The dashed curve is the trend derived from extrapolated values using SPO as the benchmark site. The dashed-dotted curve is the trend derived from extrapolated values using Mauna Loa, Hawaii (MLO) as the benchmark site (see section 4.1). The dotted curve is the trend derived from extrapolated values derived using the latitude reference method (see section 4.2).

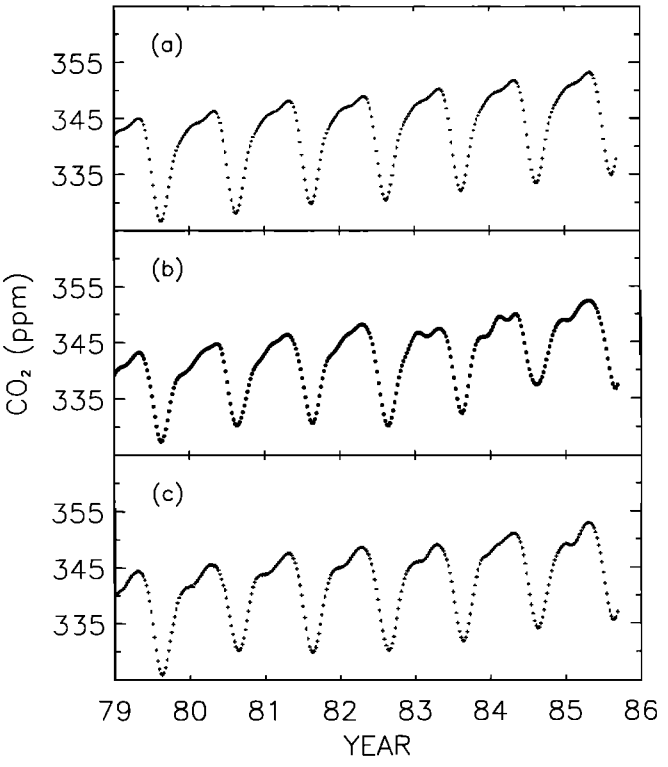

Figure 7. (a) Extrapolated CO_2 values derived using the benchmark trend method at SHM for January 1979 to September 1985; (b) weekly synchronized smoothed values at CBA over the same period; (c) extrapolated CO_2 values derived using the latitude reference method at SHM over the same time period.

latitude reference time series, discussed shortly, is constructed from measurements made from active CMDL sites during the synchronization period. The LR method is an attempt to extrapolate and interpolate CO_2 mixing ratios that are more representative of a site's latitude by incorporating knowledge obtained from sampling records nearby in latitude. Extended carbon dioxide trend and seasonal cycle patterns at SHM, for example, should be influenced by patterns recorded at nearby sites such as CBA.

We again use SHM and CGO to illustrate the progression of the data extension procedure. Figures 8a and 9a show smoothed values from the flask sampling sites SHM and CGO. First, we defined the large scale north-south gradient for each week of the synchronization period. Weekly latitudinal distributions (CO_2 (ppm) versus sine (latitude)) comprised of smooth values from marine boundary layer sites were compiled. Since the gradients were to represent very large scale zonal characteristics, only marine boundary layer (MBL) sites with records obtained from

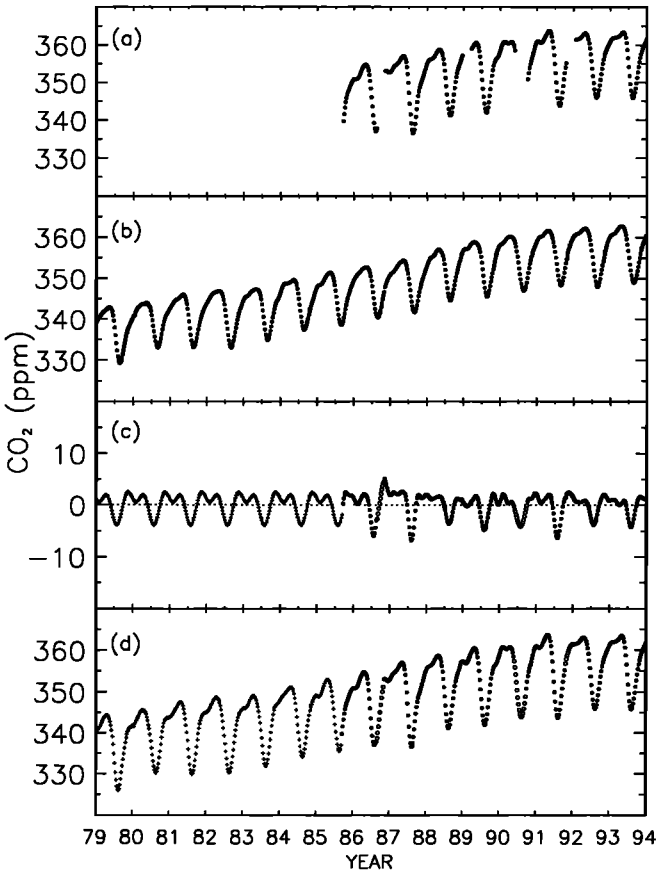

Figure 8. Progression of the data extension procedure using the latitude reference method. (a) Weekly synchronized smooth curve values, $S_{STA}(t)$, from SHM; (b) latitude reference time series values described by $REF_{STA}(t)$ at the latitude of SHM; (c) differences of smoothed values minus reference values, $\Delta_{STA,REF}(t)$ (solid circles), used to define the difference climatology and interpolated differences (open circles) and extrapolated differences (pluses) defined using the difference climatology; (d) extended site record, $E_{STA,LR}(t)$, containing smoothed values (solid circles) and interpolated and extrapolated values (open circles and pluses). See section 4.2 for details.

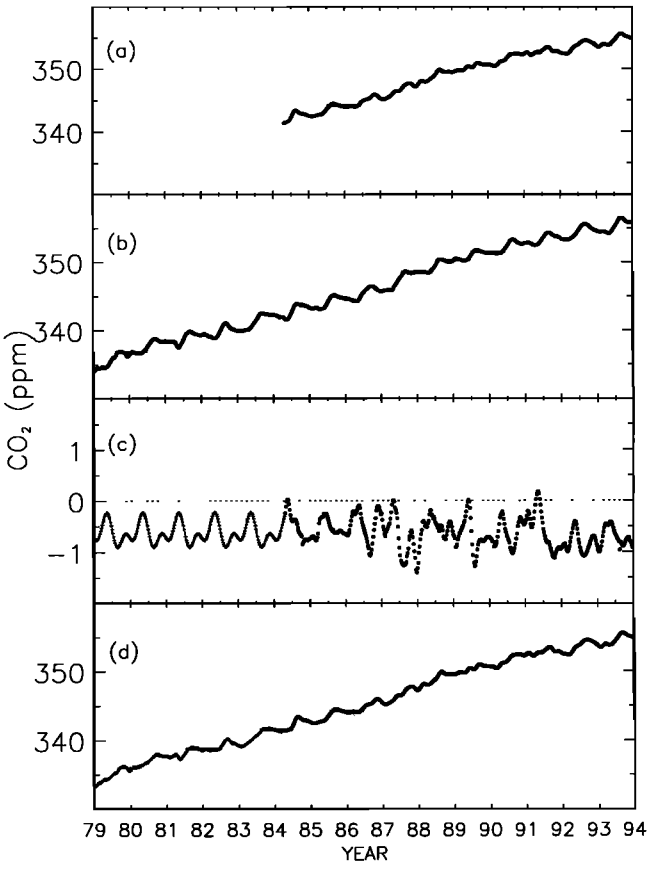

Figure 9. Same as Figure 8 but for CGO.

generally well-mixed air masses (marked "y" in the last column of Table 1) were used. The number of sites contributing to each weekly distribution varied with network expansion and attrition. A locally weighted regression model or "loess" model [*Cleveland and Devlin*, 1988; *Cleveland*, 1979] was then applied to each distribution to approximate the weekly north-south latitude gradient. The loess smoothing procedure replaces each point in the weekly latitudinal distribution with a value determined from a locally weighted least squares quadratic fit to smooth values contained in a window about the point. For this application the window is set to include all values in the distribution. Assigned weights, as defined in section 3, were passed to the loess model with each value in the distribution to ensure that CO_2 values from sites with high signal to noise and/or consistent sampling had greater influence on the resulting loess curve fit. At each time step, values were extracted from the loess curve at intervals of 0.05 sine of latitude from 90°S to 90°N. This was accomplished at the poles by copying the southern and northernmost CO_2 values from the distribution to fictitious sites at 90°S and 90°N, respectively and assigning these values minimum weight. When values from SPO were present, this "trick" was not required for 90°S, but always in the north a value from either BRW, Mould Bay, Northwest Territories (MBC), or Alert, Northwest Territories (ALT) was copied to the "site" at 90°N. Figure 10 shows some of these latitudinal curves for 1986. The area of the symbol is proportional to the relative weight of each value. We have selected the loess model because it is robust, well documented, easily available, and provides a suitable local fit about each point in the distribution. We have also tried a weighted cubic least squares polynomial fitted to the weekly distribution. A

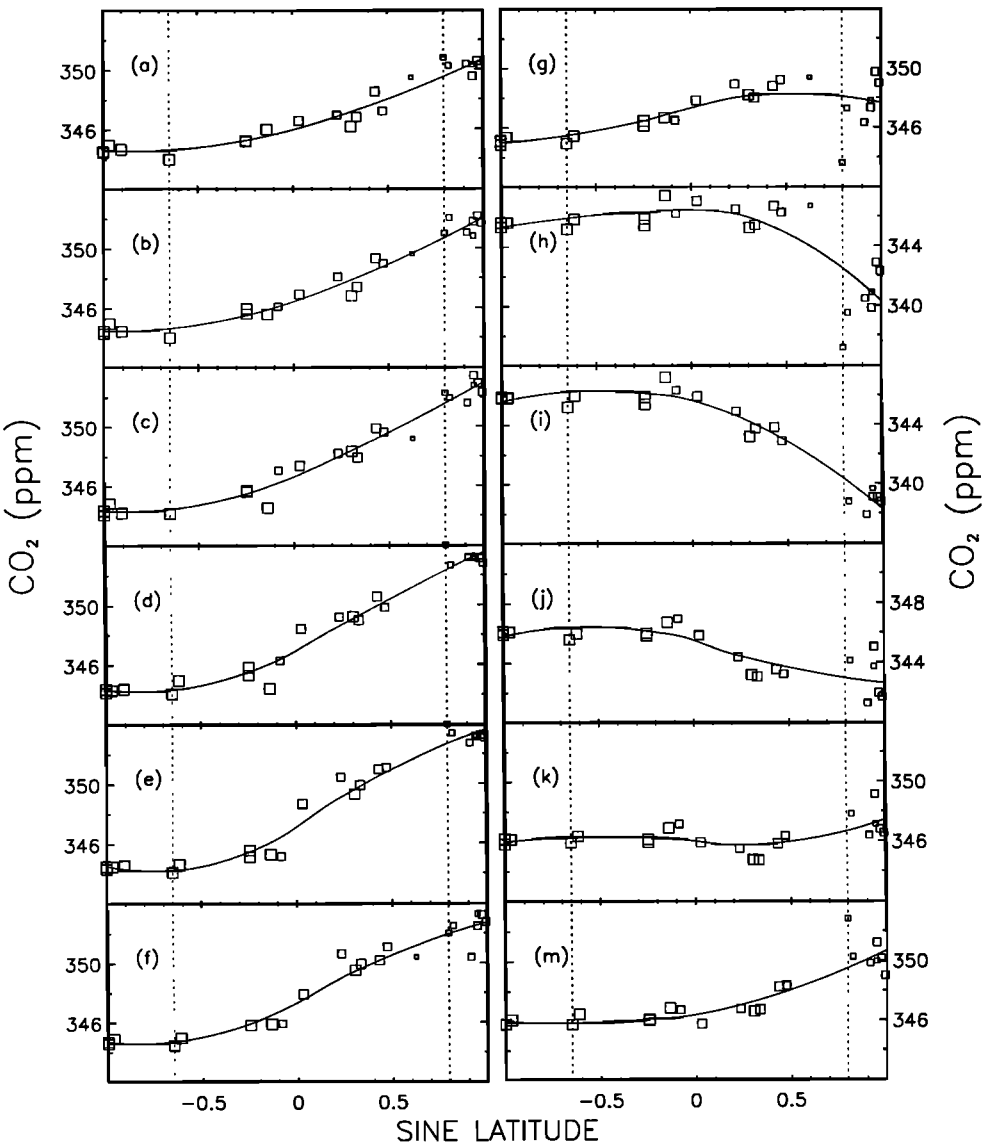

Figure 10. Weighted loess curves fitted to weekly latitudinal distributions of marine boundary layer (MBL) smoothed values. (a-m) Weekly distribution and curve for every fourth week (approximately monthly) beginning January 1, 1986. The area of the symbol is proportional to the relative weight of each value (see section 3). The vertical dotted lines indicate the sine (latitude) positions of CGO (41°S) and SHM (53°N). Weekly curves describing the north-south gradients are used in constructing latitude reference time series (see section 4.2).

polynomial is attractive because of its simplicity and robustness, but it produces a fit that is more global in nature. This is a concern when the number of sites defining the north-south gradient is small. In 1979 and 1980, for example, when the number of sites in a weekly distribution was small, the polynomial tended to overshoot or undershoot inflection points resulting in curves that probably did not adequately represent the actual latitudinal gradient. The loess curve also exhibited this behavior when the number of sites in the distribution was small but to a much lesser extent. The integration of data records from other laboratories will improve the density of measurements during these early years and better constrain these weekly north-south gradients.

The resulting set of loess curves allows us to construct reference MBL CO_2 time series at each time step in the synchronization period at any latitude. For example, at CGO we construct a reference time series, $REF_{CGO}(t)$, by extracting a CO_2 mixing ratio from the loess curves fitted to each weekly north-south gradient at the sine (latitude) of CGO (see Figure 10). The latitude reference, $REF_{CGO}(t)$, is thus influenced mostly by MBL sites nearby in latitude to CGO, by CGO itself during its measurement period, and to a lesser extent by all other valid MBL mixing ratios used in the loess curve fits. Figures 8b and 9b show this reference time series, $REF_{STA}(t)$, at the latitudes of SHM and CGO, respectively.

From this point, the framework of the LR method parallels the BT method described in section 4.1. For each site, smoothed mixing ratios minus values from $REF_{STA}(t)$, at the sine (latitude) of the sampling location, were calculated for each time step containing smooth values,

$$\Delta_{STA,REF}(t)=S_{STA}(t) - REF_{STA}(t). \quad (8)$$

These differences, $\Delta_{STA,REF}(t)$, highlight features that distinguish the individual site from the reference time series derived from active network sites. The real information content of a particular station is embodied in these differences: they portray what cannot be predicted for that site based on measurements elsewhere. A difference climatology was then described by fitting a function to $\Delta_{STA,REF}(t)$. Equation (1) was again modified so that the average offset for the entire site record relative to the latitude reference values and the average seasonal cycle (also relative to the latitude reference values) were approximated by the function

$$d_{STA,REF}(t) = m_0^R + \Sigma_{k=1,2}\left[m_{2k-1}^R \sin(2\pi kt) + m_{2k}^R \cos(2\pi kt)\right]. \quad (9)$$

The difference climatology, $d_{STA,REF}(t)$, describes the average difference between the smooth curve and the latitude reference everywhere that actual measurements exist. Again, we can filter the residuals, $r_{STA,REF}(t)=\Delta_{STA,REF}(t)-d_{STA,REF}(t)$, using a 40-day FWHM filter and combine the smoothed residuals with equation (9) to produce a smooth curve,

$$S_{STA,REF}(t) = d_{STA,REF}(t) + \{r_{STA,REF}(t)\}_{40d}. \quad (10)$$

The data extension procedure again relies on the assumption that the difference climatology described by equation (9) is valid for periods when there are no measurements. Time steps within external gaps in the synchronization period were filled using extrapolated CO_2 values determined from equation (10) where the term $\{r_{STA,REF}(t)\}_{40d}$ is zero. Time steps within internal gaps in the synchronization period were filled using interpolated values from equation (10) where the contribution from the smoothed filtered residuals is nonzero exactly as in the benchmark trend method. Figures 8c and 9c show the difference vectors for the sampling sites SHM and CGO. Comparing the SHM difference vector (Figure 8c, solid circles) with the difference vector derived from the BT method (Figure 4b, solid circles), we find that differences derived from the LR method are more tightly distributed about a zero offset. This is to be expected since the latitude reference time series exhibits both seasonal cycle patterns and a long-term trend, while the BT reference is only a benchmark long-term trend without a seasonal cycle. This contrast is greatest at sites farthest from the benchmark site used in the BT data extension method. An interesting detail visible in Figure 8c is that the seasonal cycle at SHM is both larger and a few weeks ahead in phase compared to the latitude reference values. This can be explained qualitatively because the seasonal cycle is mostly caused by the terrestrial biosphere and SHM is close to the northern Asian mainland. This feature will thus provide a quantitative constraint on photosynthesis/respiration in northern Asia. Another interesting feature in Figure 8c is the double maximum during winter that provides an additional constraint on respiration.

The latitude reference method creates a matrix for each of the 65 sites used. Each matrix has the same number of rows and columns as before. The columns contain the first two columns of the data preparation matrix, the latitude reference time series vector at the site's latitude, $REF_{STA}(t)$, and the difference vector, $\{\Delta_{STA,REF}(t), S_{STA,REF}(t)\}$. Again, each matrix is saved to a file. Table 5 shows a portion of the LR method matrix for SHM.

Table 5. Portion of the SHM Result Matrix Derived From the Latitude Reference Method

Date, GMT[a]	$S_{STA}(t)$[b]	$REF_{STA}(t)$[c]	$\{\Delta_{STA,REF}(t), S_{STA,REF}(t)\}$[d]
1979.000000	-999.999	339.193	1.206
1979.020874	-999.999	339.613	0.929
.	.	.	.
.	.	.	.
.	.	.	.
1985.666626	-999.999	338.578	-2.521
1985.687500	-999.999	338.846	-1.878
1985.708374	339.792	339.484	0.308
1985.729126	341.663	340.390	1.273
.	.	.	.
.	.	.	.
.	.	.	.
1993.979126	361.324	360.064	1.260
1994.000000	361.766	360.645	1.121

The first 15 lines of informational text have been omited. For details, refer to section 4.2.

[a]Synchronized time steps.

[b]Smooth values from curve fit.

[c]Latitude reference at the sine (latitude) of SHM.

[d]Difference vector.

Extended CO_2 site records, $E_{STA,LR}(t)$, derived from the LR method can be generated by combining the two column vectors containing latitude reference values and differences as follows:

$$E_{STA,LR}(t) = REF_{STA}(t) + \{\Delta_{STA,REF}(t), S_{STA,REF}(t)\} \quad (11)$$

Figures 8d and 9d show the extended site records resulting from the latitude reference method for SHM and CGO. Table 2 summarizes the expressions used.

The primary feature that distinguishes the LR extended records (Figures 8d and 9d) from the BT extended records (Figures 4c and 5c) is the interannual variability in the seasonal cycle of extrapolated CO_2 mixing ratios. This variability is introduced into the data extension procedure by considering year-to-year seasonal cycle differences observed at nearby sites. If we compare the CBA smoothed record with the SHM extended record from 1979 to 1986 (Figures 7b and 7c), we note that features in the nearby CBA record display similarities to the extrapolated values at SHM. The pattern in the CO_2 summer minimum at CBA for the period 1979 to 1986, for example, is well reproduced in the SHM extended values. Also, variability in the shoulder during the winter CO_2 buildup observed at CBA is evident in the SHM extended values. These features are not identically reproduced at SHM because of influences from sites other than CBA nearby in latitude.

Long-term trends derived using extrapolated values from the LR method are more similar to trends from active sites nearby in latitude than to a globally representative long-term trend. This is apparent in Figure 6 where we compare the trend derived using extrapolated values from the LR method at SHM (dotted curve) with the long-term trend from smoothed values at nearby CBA (solid curve) for the period January 1979 to September 1985. The trend at SHM determined using the LR method more closely resembles the smooth trend at CBA than does the trend determined using the BT method (dashed and dashed-dotted curves). The deviation in 1979, where data are sparse, is partially due to the artifacts of the latitudinal curve fitting described earlier.

A small bias in the constructed time series may be introduced when measurements at a particular site differ significantly from values at sites nearby in latitude. Such is the case at CGO where CO_2 mixing ratios are consistently lower than measurements from Antarctic sites to the south and AMS to the north [*Conway et al.*, 1994] providing evidence for the existence of a southern ocean sink [*Tans et al.*, 1989]. When CGO measurements are included in the weekly north-south latitude distribution, the loess curves at the latitude of CGO are appropriately pulled toward lower CO_2 values. The function $d_{CGO,REF}(t)$ is based on times when $REF_{CGO}(t)$ is influenced by CGO itself, but at all time steps, $S_{CGO,REF}(t)$ is added to $REF_{CGO}(t)$ even when CGO data are absent. The integration and extension of measurement records from other laboratories should reduce this condition by improving spatial resolution and temporal continuity.

Both the benchmark trend and the latitude reference methods result in extended records that capture large spatially averaged CO_2 characteristics. Neither method can construct anomalous CO_2 values that may have occurred at individual sampling sites because of localized transport or temporary sources or sinks. Both methods assume that the difference climatology is invariant in time. For example, the latitude reference method assumes that on average the difference climatology that describes the relationship between measurements from an individual site and measurements from MBL sites nearby in latitude will not change as the site record is extended beyond the actual measurement period. We know from direct observations that this is not the case. As an extreme example, at the sampling location Tae-ahn Peninsula, Korea (TAP), which is relatively close to heavily populated regions, the annual mean and the seasonal cycle of the difference from the latitude reference values are quite different from year to year. We expect the assumption of invariance of the differences to be "conservative" in terms of sources and sinks that might be derived with models. Imagine that CO_2 mixing ratios at MBL sites at temperate northern latitudes during a certain year are 1 ppm higher than in other years, but we have no data from the continents that year. Our assumption of invariance leaves differences between the continental and the marine boundary layers the same as has been observed in other years. We expect models will be forced to infer that the cause of the higher annual mean concentrations is spread over all longitudes more or less equally. In other words, lack of data "forces" the result that the cause of the 1 ppm anomaly is in this case more or less equally of marine and terrestrial origins. This strategy may have to be revised in light of the increasing availability of $\delta^{13}C$ data.

5. Discussion

The two data extension methods described above construct atmospheric CO_2 values by combining knowledge from a long-term reference with information from individual sites resulting in extended records containing smoothed, extrapolated, and interpolated values. While the benchmark trend method constructs a reference long-term trend from a single benchmark site, the latitude reference method constructs a reference time series utilizing MBL sites selected from the CMDL sampling network and continuous programs. Extrapolated CO_2 values derived from the LR method account for variability in the interhemispheric gradient as well as seasonal cycle patterns and long-term trends derived from measurements at sites nearby in latitude. Extended values derived from the LR method should better represent temporal variations in CO_2 than values derived from the BT method. To test this assumption, we reconstructed several site records using both methods and compared the results with observations.

Several sampling sites were active during most of the synchronization period resulting in extended records containing few or no extrapolated and interpolated values. The flask data set at CBA for example, has no external or internal gaps and thus has no extrapolated or interpolated values in the extended record (Figure 11a). As an exercise, we excluded portions of the CBA record prior to the data preparation (see section 3) and reconstructed the record once using the benchmark trend method and once using the latitude reference method. Both extended records were then compared with the actual record. First, we excluded all data prior to January 1987. The resulting extended data sets from both data extension methods produced extrapolated mixing ratios from 1979 through 1986 (Figures 11b and 12b). The mean differences of extrapolated values minus smoothed values for the 8-year period were (0.56 ± 1.14) ppm for the BT method and (0.01 ± 0.93) ppm for the LR method

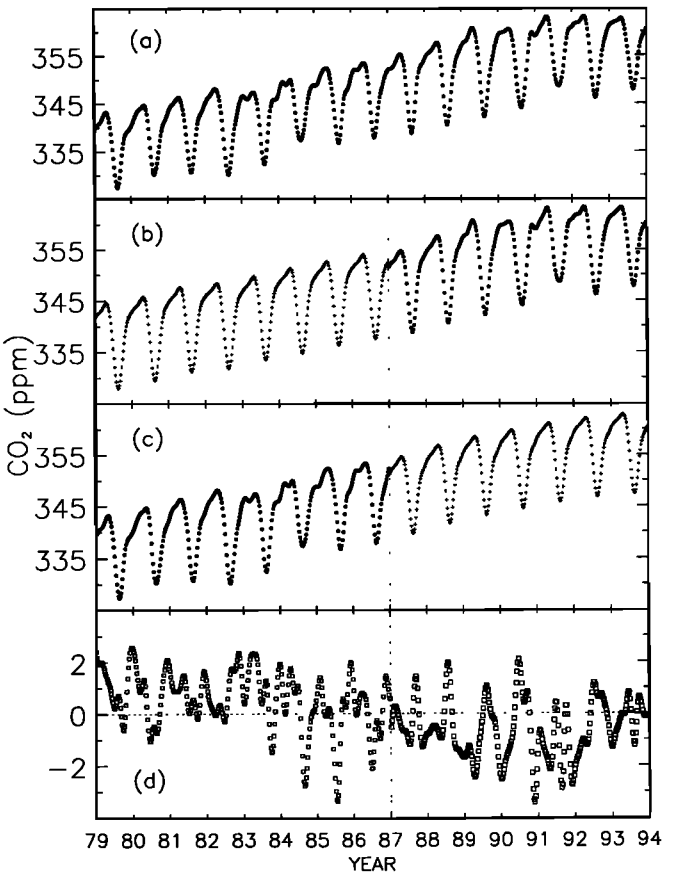

Figure 11. Reconstruction of a portion of the flask sampling record at CBA using the benchmark trend method. (a) CBA record; (b) reconstruction of the CBA record assuming the measurements begin in 1987. Plotted are smoothed values (solid circles) and extrapolated values (pluses). (c) Reconstruction of CBA assuming record ends in 1987. Plotted are smoothed values (solid circles) and extrapolated values (pluses). (d) Differences of extrapolated values (Figures 11b and 11c) minus smoothed values (Figure 11a). The vertical dotted line distinguishes the two scenarios.

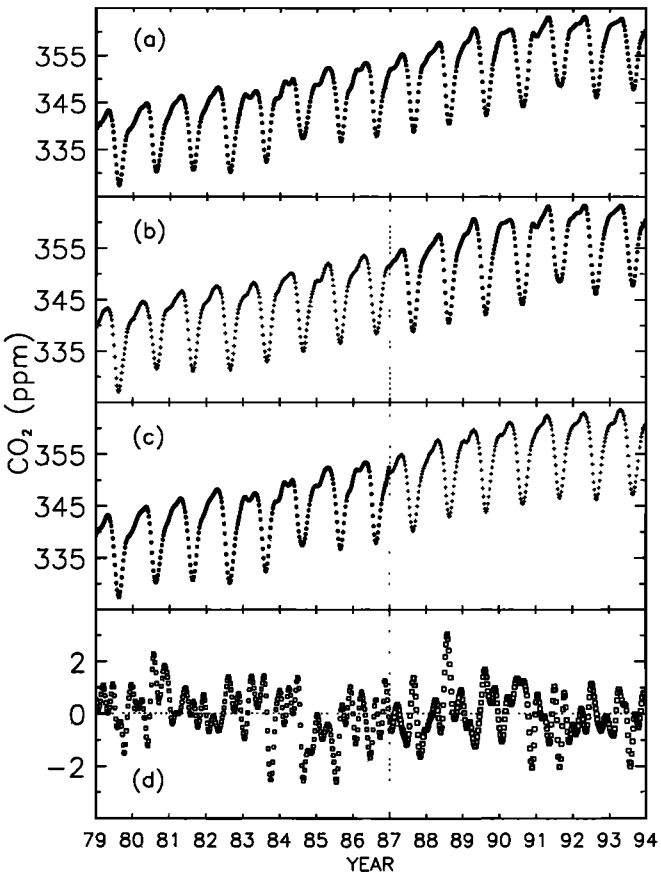

Figure 12. Reconstruction of a portion of the flask sampling record at CBA using the latitude reference method. (a) CBA record; (b) reconstruction of the CBA record assuming the measurements begin in 1987. Plotted are smoothed values (closed circles) and extrapolated values (pluses). (c) Reconstruction of the CBA record assuming the measurements end in 1987. Plotted are smoothed values (closed circles) and extrapolated values (pluses). (d) Differences of extrapolated values (Figures 12b and Figure 12c) minus smoothed values (Figure 12a). The vertical dotted line distinguishes the two scenarios.

(Figures 11d and 12d). Second, we assumed that sampling at CBA was discontinued in January 1987. The resulting extended data sets from both extension methods contained extrapolated values from 1987 through 1993 (Figure 11c and 12c, respectively). The mean differences of extrapolated values minus smoothed values were -(0.59 ± 1.09) ppm for the BT method and -(0.02 ± 0.87) ppm for the LR method (Figures 11d and 12d). Several other sites with sufficiently long records showed similar results. Figure 13 shows the distribution of mean differences for 14 sites with long records. Site reconstruction using the latitude reference method agrees better with the actual site record than does reconstruction using the benchmark trend method. Extrapolated values derived from the LR method and smoothed values differed on average by ~0.2 ppm, while the BT difference was ~0.5 ppm. The strong tendency for BT-derived extrapolated values to be greater than the smoothed values for the period 1979 through 1986 and less than the smoothed values for the period 1987 through 1993 reflects the inability of the BT method to detect the changes in the interhemispheric gradient mentioned earlier. Similar results were produced when the exercise was repeated using different portions of the records. These results suggest that extrapolated values obtained using the latitude reference method better represent the temporal variations in CO_2 observed at the sampling locations than extrapolated values derived from the benchmark trend method.

The remaining discussion examines several aspects of the latitude reference method to better understand the limitations of such an approach. In developing the LR method, we have made assumptions that require comment. Model calculations suggest and measurements verify that atmospheric CO_2 does vary longitudinally [e.g., *Fung et al.*, 1983; *Conway et al.*, 1988; *Nakazawa et al.*, 1992]. We assume that the relationship of an individual site to the latitude reference is invariant in time. Observations show that this assumption is often not supported. We have already discussed TAP as an obvious case, but most other sites suggest that their relationship to nearby sites is continually changing. We make these assumptions for two reasons. First, although the CMDL cooperative air sampling network is the most extensive in the world, it still does not provide the sampling coverage required to adequately characterize the variability of atmospheric CO_2 with longitude; and second, forward and backward extrapolation necessarily requires assumptions about CO_2 trends, changes in trends, and seasonal cycle patterns at locations where we have no data. What impact might these assumptions have on the results?

The function $d_{STA,REF}(t)$, equation (9), adequately represents the relationship between an individual site and a reference trend and seasonal cycle pattern as long as the relationship remains nearly the same over time. For example, $d_{STA,REF}(t)$ derived from measurements at SCS made during the early 1990s, a period of rapid increase in fossil fuel consumption in China and Southeast Asia, will perhaps accurately describe the present climatology at SCS but may misrepresent the climatology during

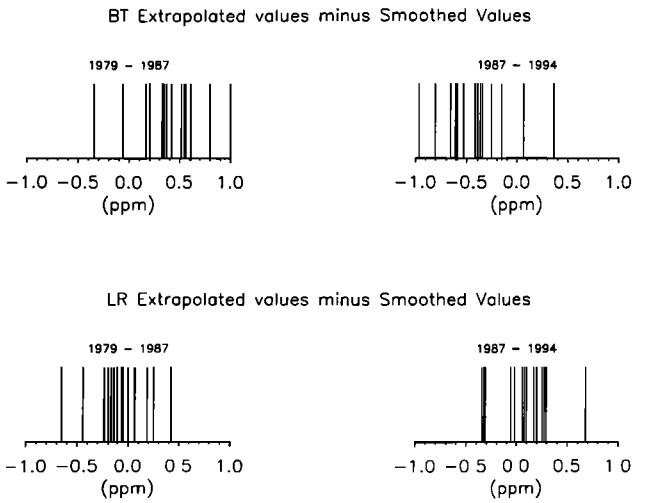

Figure 13. The distribution of mean differences of extrapolated minus smoothed CO_2 values from the reconstruction of 14 CMDL sampling sites with long records using (top) the benchmark trend method and (bottom) the latitude reference method. Each vertical line intersects the horizontal axis at the mean difference at one of 14 sites. (left top and bottom) Results when assuming the records begin in 1987; (right top and bottom) results assuming the records end in 1987.

the 1980s when rapid economic development was just getting started. On a shorter timescale, it is reasonable to expect time variations of the coefficients of $d_{STA,REF}(t)$ related to El Niño events at many network sites. It is possible to allow long-term variations in the curve fit parameters of $d_{STA,REF}(t)$ if enough data are available. We have experimented with the inclusion of time-dependent coefficients in $d_{STA,REF}(t)$ while being aware of the dangers of extrapolating polynomial fits beyond the range of data. A conservative approach would be to add a linear time-dependent coefficient to $d_{STA,REF}(t)$. We have chosen not to include such variation by assuming that $d_{STA,REF}(t)$ is invariant over the synchronization period.

The CMDL cooperative air sampling network has been continually changing since its inception in 1968. Figure 1 shows the latitudinal distribution of sampling density where network expansion, attrition, and gaps within individual sampling site records are evident. What impact, if any, do changes in network sampling density have on extended CO_2 values at any given site? As alluded to earlier, the weighted loess curve fit used in determining the latitude reference time series, $REF_{STA}(t)$, is sensitive to the addition, deletion, or absence of a MBL site when the number of sites is small. This is particularly evident in the 1979-1981 period. Curve fits are also sensitive to changes in network density in the low latitudes where values are heavily weighted by area. To illustrate this sensitivity, we plot the weekly latitude reference mixing ratios extracted from each latitudinal curve fit at the latitude of CHR as a function of time (Figure 14). We connect weekly values with a line to draw attention to perturbations in the time series. In 1980 (Figure 14, inset a), sampling problems at Guam, Mariana Islands (GMI), Key Biscayne, Florida (KEY), and Seychelles, Mahe Island (SEY) resulted in occasions when data from one or more of these sites were missing from the weekly latitudinal distribution. Because the sampling density was sparse during this period, gaps in records strongly impacted the curve fitted to the latitudinal gradient. In March and April 1984, sampling at Christmas Island (CHR) and CGO began providing additional constraints on the gradient. Measurements from CHR and CGO influence the fits as can be seen from the impact on the latitude reference time series when sampling at the two sites began (Figure 14, inset b). Other peculiarities in the data construct are evident in more recent years. In 1986, 1987, and 1990, internal gaps at SEY were responsible for the apparent perturbations. Although by 1986 the sampling network was well established, sampling changes in the low latitudes continued to strongly influence the latitudinal gradient because of the region's large surface area.

An additional comparison of measurements with extrapolated values derived from the LR method was possible due to overlapping independent CO_2 data records from the Alert sampling site in Canada. The CMDL flask sampling effort at Alert began in 1985 as a cooperative effort with the Atmospheric Environment Service (AES) CO_2 program of Canada. The AES CO_2 program has maintained a flask sampling effort at Alert since 1975. In Figure 15a, smooth curve values, $S_{STA}(t)$, from the AES CO_2 flask sampling program at ALT are plotted as a function of time. In Figure 15b the extended CMDL ALT record is shown. Sampling began in June 1985 as indicated by the vertical dotted line. Figure 15c shows the calculated differences

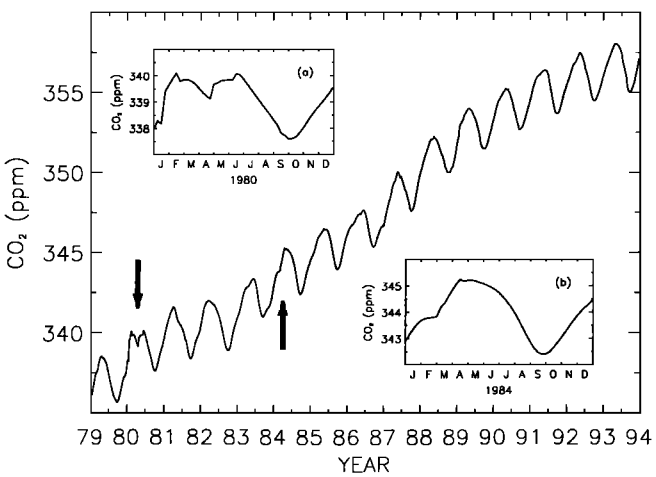

Figure 14. Latitude reference time series, $REF_{STA}(t)$, at the latitude of Christmas Island (CHR). The weekly values are connected with a line to draw attention to perturbations in the time series. Inset a illustrates how the changes in the weekly latitudinal distribution in 1980 because of sampling problems at Guam, Mariana Islands (GMI), Key Biscayne, Florida (KEY), and Seychelles, Mahe Island (SEY), affect reference values at the latitude of CHR. Inset b illustrates how the commencement of sampling at CHR in March and CGO in April 1984 affects the reference values at the latitude of CHR.

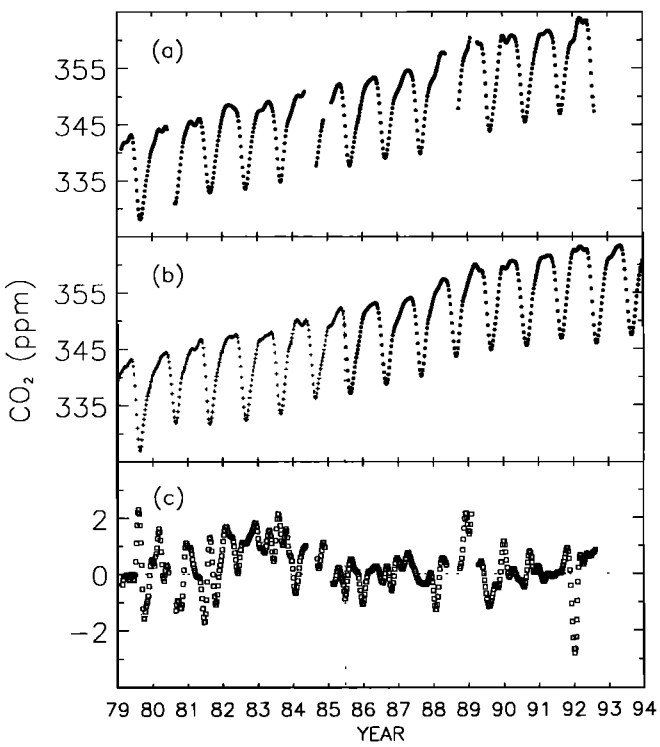

Figure 15. Comparison of Atmospheric Environment Service (AES) smoothed data with CMDL smoothed and extrapolated values from independent sampling programs at Alert, Northwest Territories (ALT). (a) Smoothed values from the AES flask sampling program at ALT for 1979 to late 1992; (b) extended CMDL record, $E_{STA,LR}(t)$, at ALT containing smoothed (solid circles) and extrapolated values (pluses); (c) differences of AES smoothed values (Figure 15a) from CMDL extended record (Figure 15b). NOAA/CMDL sampling at ALT began in June 1985 as indicated by the vertical dotted line.

of the AES smoothed record minus the CMDL extended record for the period 1979-1992. As a reference, the mean difference of AES smoothed values minus CMDL smoothed values from mid-1985 to late-1992 was (0.1 ± 0.7) ppm. The mean difference of AES smoothed values minus CMDL extrapolated values for the period 1979 to mid-1985 was (0.5 ± 0.9) ppm. Although not rigorous, the direct comparison of the two smoothed ALT records suggests that calibration differences between laboratory reference standards are small. As a result, the comparison of CMDL extrapolated ALT values with AES smoothed ALT data indicates that the backward extrapolation of the CMDL ALT record using the latitude reference method produces an extended record that reasonably reproduces direct observations.

As a test of how data extension might effect bias because of a continually changing network composition, we have used the extended records derived from the latitude reference method in a limited model application and have compared the outcome to earlier results. The two-dimensional (latitude, height) transport model described by *Tans et al.* [1989] has been used elsewhere [*Conway et al.*, 1994] to deduce source and sink patterns of CO_2 as a function of time and latitude using the CMDL cooperative air sampling network. For this exercise, the two-dimensional model was run in two operating modes. In the first mode, the model used only the smooth CO_2 values $S_{STA}(t)$ from the extended records where measurements exist; results from this mode of operation served as the basis of the comparison. In the second mode, the model used smoothed, extrapolated, and interpolated values from the extended records, $E_{STA,LR}(t)$. The impact of using the extended records in this application was then assessed by direct comparison of the model results from the two operating modes.

The two-dimensional model requires that mixing ratios at the surface be defined at all times and latitudes. In mode 1, curves are fitted to biweekly latitudinal distributions comprised of smoothed values only [*Tans et al.*, 1989]. Weights were assigned to each smoothed value according to the weighting scheme described in section 3. The weight matrices (Table 3) also contain the residual standard deviation (RSD) of the measurements about the smooth curve and the number of residuals per year used in the RSD determination. These additional variables should allow the choice of alternative weighting schemes. The resulting biweekly meridional curve fits were then used as inputs to the model. The curve fits in this application are, in general, different from the latitudinal loess fits because the latter are limited to using only MBL sites whereas the former are not.

Figure 16a shows zonally averaged, global, and tropical (≈17°S to 17°N) CO_2 source strengths as a function of time derived from the two-dimensional model using smooth values only (mode 1). The fossil fuel CO_2 source has been removed. Since different sampling network distributions would likely give rise to different derived CO_2 source strength scenarios, we have used a bootstrap analysis to estimate the uncertainty in these derived sources due to the CMDL network distribution itself. For this analysis we constructed 100 alternative "networks" by selecting sites at random, with restitution, from the 65 actual network sites listed in Table 1. Each of the 100 bootstrap networks contained a total of 65 sites, but sites may have been present more than once or not at all. To ensure the stability of the meridional curve fits described above, each "network" was required to include either the BRW flask or continuous record,

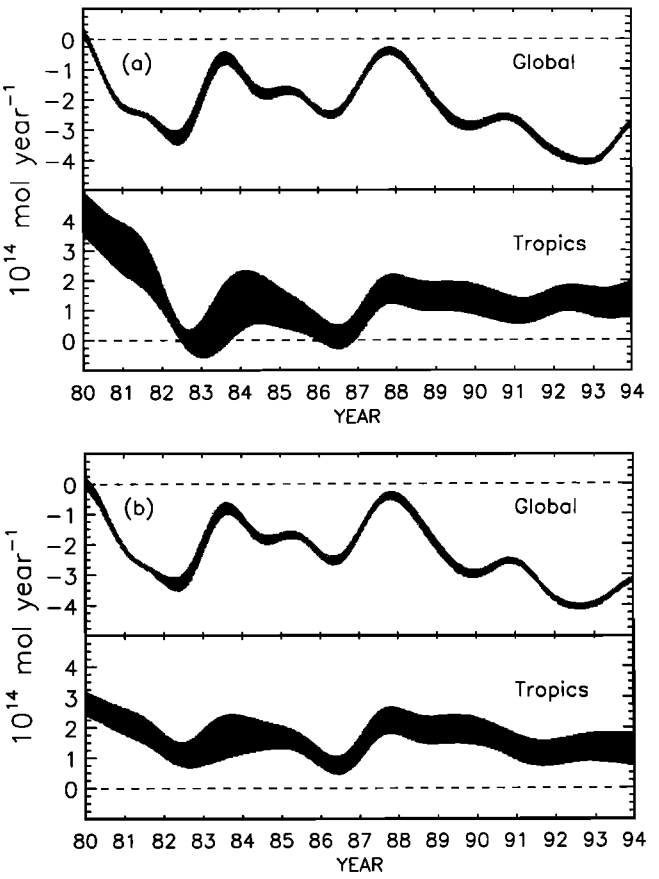

Figure 16. Global and tropical (≈17°S to 17°N) smoothed biweekly natural CO_2 source strength scenarios (i.e., fossil fuel source removed) derived from the two-dimensional model. (a) Results from the model using smoothed CO_2 values, $S_{STA}(t)$, only. (b) Results from the model using the extended records, $E_{STA,LR}(t)$. Estimated uncertainties determined using bootstrap analysis (see section 5). 1×10^{14} moles of CO_2 is equivalent to 1.2 Gt carbon.

the CBA and Cape Kumukahi, Hawaii (KUM) records, either the Tutuila, American Samoa (SMO) flask or continuous record, and both the SPO flask and the continuous records. The two-dimensional model was then run 100 additional times in mode 1 using each of the constructed networks. Results from each model run were accumulated allowing statistics to be compiled on several derived parameters. In Figure 16 the uncertainty (shaded band) about a source strength curve (solid curve) is one standard deviation (±1σ) from the mean of the source strength determined from the 100 model runs. The uncertainty estimates shown in Figure 16a suggest that while the CO_2 source strength in the tropics is known on average to within about 25%, the global source strength is known to within ~5%. The large relative uncertainty in the derived tropical source is due primarily to the large potential variability in the number of sites within the 17°S to 17°N latitude zone from one "network" to the next. This can lead to substantial differences in the local curvature as a function of latitude. The derived global source strength on the other hand, depends on the time derivative of the global integral of the CO_2 mixing ratio and is thus less sensitive to "unusual" local distributions that may arise from the bootstrap analysis.

We next ran the model using smooth, extrapolated, and interpolated CO_2 values from the extended records (mode 2). Although the extended records have uniform distribution in time, weighted latitudinal curve fits were still needed to define surface mixing ratios at every latitude. The weights assigned to the smoothed values were identical to those used in mode 1 and described in section 3. In the model, extrapolated and interpolated values were assigned unity weight, lower than the fixed minimum weight of value 2 for smoothed values. The resulting biweekly meridional curve fits were then used as inputs to the model.

Figure 16b shows the zonally averaged, global, and tropical natural source strength of CO_2 derived from the mode 2 model run. Uncertainties were estimated using the bootstrap analysis technique, but because extrapolated values derived from the latitude reference method depend on the composition of the sampling network itself, the technique was modified. As before, we constructed the 100 alternative "networks" by selecting sites at random, with restitution, from the 65 actual network sites. The same minimum constraint on the network composition used in mode 1 was imposed. We next constructed entirely new sets of extended records by running the date extension procedure 100 times, once using each unique network distribution. Finally, the model was run 100 additional times in mode 2 using the extended records constructed from each of the alternative networks. Results from the model runs were again accumulated allowing statistics to be compiled on several derived parameters.

Direct comparison of the two-dimensional model results suggests that there is bias introduced by changes in the CMDL network distribution. Carbon dioxide source strengths derived from the mode 1 model run tend to suggest more extreme local minima and maxima than comparable sources derived from the mode 2 model run. This is particularly evident from the zonal averages such as the tropical source shown in Figure 16a. Source strengths derived from the mode 2 model run tend to attenuate the extremes suggesting less source strength variability. To illustrate this effect, we contrasted the mode 1 derived tropical source (Figure 16a) with the mode 2 tropical source (Figure 16b). This comparison also shows that mode 1 bootstrap uncertainty estimates that are sensitive to changes in the network distribution exhibit greater variability than the mode 2 uncertainty estimates, particularly from 1980 to 1987 when the network was expanding rapidly. When using the extended records, the extrapolated values derived from the extension procedure provide more stability to the source determination despite the fact that extrapolated values were given the lowest weight in determining the surface distribution as input to the two-dimensional model. As expected, hemispheric and globally averaged source strengths also show this effect but to a lesser degree because averaging is over larger spatial scales (contrast Figures 16a and 16b, global source).

We believe that sources derived from the two-dimensional model run that excluded extrapolated and interpolated values were indeed biased by additions, deletions, and discontinuities in the CMDL network (Figure 1). We discussed how a model may be forced to infer sudden changes in sources or sinks simply because a measurement record such as the Baltic Sea (BAL) suddenly begins. This problem is exacerbated by the fitting of smooth curves as a function of latitude. Roughly speaking, a source will correspond to a bulge (negative second derivative) in the latitudinal curve fit and a sink to a positive second derivative. The addition of a new site may cause the appearance of a slight bulge in the curve fit accompanied by areas of a slightly more positive second derivative on either side or vice versa. For example, if, as in the case of BAL, the model infers that there was suddenly an increase in sources at latitudes of northern Europe, then the model would likely infer a partially compensating source decrease elsewhere. The use of extended records can partially correct for bias that would result from the sudden addition or omission of sites. The model "sees" a sampling distribution that although not perfect has complete temporal and spatial coverage of the CMDL network. We anticipate that as data from other measurement programs are added and records become longer, the source/sink scenario produced by the two-dimensional model will improve further.

6. Conclusion

We have proposed a data integration and extension procedure that will result in a global atmospheric CO_2 database that should aid modelers in the pursuit of a better understanding of the global carbon budget. This global database will provide unprecedented spatial resolution and temporal continuity of atmospheric CO_2. In this paper we have developed the data extension procedure using the CMDL cooperative air sampling network. The extension procedure attempts to apply knowledge embodied in a CO_2 site climatology obtained from a site record to time periods beyond the record itself. This is accomplished by combining site climatologies determined from individual site data with long-term references derived from data from CMDL sampling locations active before, during, and after individual measurement records. Trace gas site climatologies describe characteristic features in a site record such as average seasonal cycle patterns, trends, and changes in trends. Two extension methods have been presented. The benchmark trend method extends CO_2 site records using information from site climatologies and a benchmark trend derived from a single network site that is representative of the global trend in atmospheric CO_2. The latitude reference method constructs reference time series from measurements of samples collected at active sites in the marine boundary layer. These references are then used with site climatologies to produce extended records containing more realistic seasonal cycles and trend patterns. The data extension procedure is designed to provide modelers with extended atmospheric CO_2 records consisting of high-precision measurements from an extensive sampling network and extrapolated and interpolated CO_2 values constructed using knowledge gained from the many years of network sampling measurements.

An essential feature of the data extension procedure proposed here is that extended records will continue to change with time as new data are added. Data extension will be as much an ongoing activity as continued monitoring. It is our expectation that the extended records will improve somewhat with time as the latitude reference time series become defined better with time.

Acknowledgments. We thank Neil Trivett of the Atmospheric Environment Service of Canada for providing carbon dioxide measurements from the sampling site at Alert, N.W.T. Canada. We also thank I. Fung, I. Enting, E. J. Dlugokencky, and P. S. Bakwin for their helpful comments on the text. This work is supported in part by the Atmospheric Chemistry Project and the Ocean-Atmosphere Carbon Exchange Study of NOAA's Climate and Global Change Program and by the Environmental Protection Agency.

References

Bolin, B., and C.D. Keeling, Large-scale atmospheric mixing as deduced from the seasonal and meridional variations of carbon dioxide, *J. Geophys. Res.*, *68*, 3899-3920, 1963.

Ciais, P., P.P. Tans, J.W.C. White, M. Trolier, R.J. Francey, J.A. Berry, D.R. Randall, P.J. Sellers, J.G. Collatz, and D.S. Schimel, Partitioning of ocean and land uptake of CO_2 as inferred by $\delta^{13}C$ measurements from the NOAA/CMDL global air sampling network, *J. Geophys. Res.*, 100, 5051-5070, 1995.

Cleveland, W.S., Robust locally-weighted regression and smoothing scatterplots, *J. Am. Stat. Assoc.*, *74*, 829-836, 1979.

Cleveland, W.S., and S.J. Devlin, Locally-weighted regression: An approach to regression analysis by local fitting, *J. Am. Stat. Assoc.*, *83*, 596-610, 1988.

Conway, T.J., P. Tans, L.S. Waterman, K.W. Thoning, K.A. Masarie, and R.H. Gammon, Atmospheric carbon dioxide measurements in the remote global troposphere, 1981-1984, *Tellus*, *40*(B), 81-115, 1988.

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 NOAA/CMDL global air sampling network, *J. Geophys. Res.*, *99*, 22,831-22,855, 1994.

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

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

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

Gillette, D.A., W.D. Komhyr, L.S. Waterman, L.P. Steele, and R.H. Gammon, The NOAA/GMCC continuous CO_2 record at the south pole, 1975—1982, *J. Geophys. Res.*, *92*, 4231-4240, 1987.

Heimann, M., C.D. Keeling, and C.J. Tucker, A three dimensional model of atmospheric CO_2 transport based on observed winds, 3, Seasonal cycle and synoptic time scale variations, in *Aspects of Climate Variability in the Pacific and Western Americas*, *Geophys. Monogr.* vol. 55, edited by D.H. Peterson, pp. 277-303, AGU, Washington D. C., 1989.

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 Western Americas*, *Geophys. Monog.* vol. 55, edited by D.H. Peterson, pp. 305-363, AGU, Washington D. C., 1989.

Komhyr, W.D., R.H. Gammon, T.B. Harris, L.S. Waterman, T.J. Conway, W.R. Taylor, and K.W. Thoning, Global atmospheric CO_2 distribution and variations from 1968-1982 NOAA/GMCC CO_2 flask sample data, *J. Geophys. Res.*, *90*, 5567-5596, 1985.

Komhyr, W.D., T.B. Harris, L.S. Waterman, J.F.S. Chin, and K.W. Thoning, Atmospheric carbon dioxide at Mauna Loa Observatory, 1, NOAA Global Monitoring for Climatic Change measurements with a nondispersive infrared analyzer, 1974-1985, *J. Geophys. Res.*, *94*, 8533-8547, 1989.

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

Nakazawa, T., S. Murayama, K. Miyashita, S. Aoki, and M. Tanaka, Longitudinally different variations of lower tropospheric carbon dioxide concentrations over the North Pacific Ocean, *Tellus*, *44*(B), 161-172, 1992.

Newsam, G.N., and I.G. Enting, Inverse problems in atmospheric constituent studies, I, Determination of surface sources under a diffusive transport approximation, *Inverse Prob.*, *4*, 1037-1054, 1988.

Pearman, G.I., and P. Hyson, Activities of the global biosphere as reflected in atmospheric CO_2 records, *J. Geophys. Res.*, *85*, 4468-4474, 1980.

Peterson, J.T., W.D. Komhyr, L.S. Waterman, R.H. Gammon, K.W. Thoning, and T.J. Conway, Atmospheric CO_2 variations at Barrow, Alaska, 1973-1982, *J. Atmos. Chem.*, *4*, 491-510, 1986.

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

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

Thoning, K.W., P.P. Tans, and W.D. Komhyr, Atmospheric carbon dioxide at Mauna Loa Observatory, 2, Analysis of the NOAA/GMCC data, 1974-1985, *J. Geophys. Res.*, *94*, 8549-8565, 1989.

Waterman, L.S., D.W. Nelson, W.D. Komhyr, T.B. Harris, K.W. Thoning, and P.P. Tans, Atmospheric carbon dioxide measurements at Cape Matatula, American Samoa, 1976-1987, *J. Geophys. Res.*, *94*, 14,817-14,829, 1989.

Yamazaki, K., and M. Chiba, A 3-D global simulation of the advective transport of passive tracers from various northern hemisphere sources, *Tellus*, *45*(B), 160-178, 1993.

K. A. Masarie (corresponding author), Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309. (e-mail: kmasarie@cmdl.noaa.gov.)

P. P. Tans, NOAA, Climate Monitoring and Diagnostics Laboratory, 325 Broadway, Boulder, CO 80303.

(Received October 13, 1994; revised February 27, 1993; accepted February 27, 1995.)