NOAA Technical Memorandum ERL ARL-173

THE STATISTICAL TREATMENT OF CO_2 DATA RECORDS

W. P. Elliott, Editor

Air Resources Laboratory
Silver Spring, Maryland
May 1989

noaa NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION / Environmental Research Laboratories

NOAA Technical Memorandum ERL ARL-173

THE STATISTICAL TREATMENT OF CO_2 DATA RECORDS

W. P. Elliott, Editor

Air Resources Laboratory
Silver Spring, Maryland
May 1989

UNITED STATES
DEPARTMENT OF COMMERCE

Robert A. Mosbacher
Secretary

NATIONAL OCEANIC AND
ATMOSPHERIC ADMINISTRATION

William E. Evans
Under Secretary for Oceans
and Atmosphere/Administrator

Environmental Research
Laboratories

Joseph O. Fletcher
Director

NOTICE

Mention of a commercial company or product does not constitute an endorsement by NOAA Environmental Research Laboratories. Use for publicity or advertising purposes of information from this publication concerning proprietary products or the tests of such products is not authorized.

CONTENTS

Preface

This collection of papers is an outgrowth of a meeting concerned with the treatment of carbon dioxide data held in Hilo, Hawaii, 24-26 March, 1988. A small group of scientists gathered to discuss various methods of extracting from raw CO_2 data information on the background concentration in the atmosphere. The meeting was sponsored by the National Oceanic and Atmospheric Administration (NOAA) and the World Meteorological Organization (WMO). Several of the papers presented at the meeting are not reproduced in this volume and some of the papers herein were not presented at the meeting. While the subject of the meeting was the treatment of CO_2 data, some of the considerations should be applicable to any trace gas measurement program. These programs will encounter many of the same problems, as well as others of their own.

Instrumentation and standard gases, topics of great importance to obtaining quality data, are not considered. Rather, the focus is on problems of selecting a set of background data from the original raw data set and of extracting from the selected data such quantities as monthly means, trends, characteristics of the annual fluctuations, and interannual variations. Problems of treating data taken intermittently, i.e., air samples collected in flasks once or twice a week are also discussed.

Background data are concentrations representative of the global atmosphere unaffected by local conditions. Since spatial gradients are expected to be quite small for a well mixed gas such as CO_2, background data should represent conditions many hundreds of kilometers upwind of the station. Ideally, two independent scientists would identify the same observations as representing background data if they were given the same records and information about the station. We cannot be sure that this would be the case now, however.

For identifying background data each observing location presents its own special problems. Observations of other quantities (gases and particles) are useful or sometimes necessary to identify those periods when background conditions prevail. Several examples of the use of other substances that have been used as guides were given at the meeting. It was also shown that calculating the trajectories of the sampled air can be helpful, as the local wind direction is not always a guide to the source of the air. In special cases the use of isotopes of carbon and other gases, such as radon, can help separate background concentrations from locally produced CO_2. The ability to detect smaller and smaller quantities of trace material raises a problem in defining background air, however. Should identifying air from North America at Mauna Loa be cause for rejecting a sample of that air from the background data set?

After a set of background data has been chosen, there is the question of forming monthly means, estimations of the annual amplitude, and other descriptors of the data. Also, gaps in the data record may exist after the selection, so interpolation schemes may be necessary. This is because CO_2

data show distinct annual cycles and simply averaging what data are available may bias a monthly average or other time period. Different interpolation methods may well produce different estimates of the relevant quantities. It is possible that the optimal interpolation algorithm is not the best treatment for determining other features of the record.

At the meeting, some differences emerged in approaches to separating seasonal variation accompanied by inter-annual variability from long term trends. A component of the time series that includes the seasonal cycle and allows for some change from year to year in this cycle can be defined in many ways. Many analyses of CO_2 time series allow quite rapid changes in the "seasonal" component so that they include, for example, the effect of Southern Oscillation anomalies. The statisticians present, however, preferred approaches in which the "seasonal" component was constrained to vary only slowly from year to year. They considered that the SO anomalies belonged in the "irregular" component, left after removal of the trend and seasonal components. Clearly the term "seasonal" has different meanings to different people. The precise definition of the seasonal component needs to be specified in all data analyses.

There are differences in data treatment among the groups measuring CO_2. At the meeting, several scientists presented methods of data treatment that are not now routinely used but which appear to be promising. Some of these are presented in this volume. It was noted that interpolation of missing data and extraction of characteristics of the records may require different techniques. Whatever technique is used, it should be fully documented so its effects can be appreciated by others.

Intermittent (flask) data present special problems. They sometimes must be fit with some sort of curve to allow estimates of monthly means, etc. to be made. Estimates of the effect of sampling frequency on the outcomes are also needed. Some locations with very small annual variations,such as the South Pole, may be characterized easily with only a few observations per month. On the other hand locations in the far north, such as Barrow and Alert, have large annual variations which are much more difficult to define.

While no definitive answers to all the problems emerged from the meeting, or will be found in this collection of papers, there are several recommendations the participants wished to make.

1. The need for careful descriptions of the data processing procedures was stressed. It is as important that other scientists understand the details of the data treatment as it is that they understand the instruments used to collect the data. Often these details are not published in Journal articles and so lost unless available in written form somewhere.

2. Background data can best be selected when consideration is given to measurements of other quantities. Anomalous concentrations of other gases or particles can be a clue that the CO_2 data should be suspect. With flask data

this is not always possible but the sampling of at least one other gas in the flask would be desirable.

3. More investigations of data processing techniques need to be carried out. There does not appear to be a single "best" method now (and perhaps never will be) but as much uniformity as possible should be a goal.

William P. Elliott
Air Resources Laboratory

1. SELECTION OF NOAA/GMCC CO_2 DATA FROM MAUNA LOA OBSERVATORY

K.W. Thoning

Cooperative Institute for Research in
Environmental Sciences
U. of Colorado
Boulder, CO

Abstract. This report describes the selection processes used by NOAA/GMCC for selecting background hourly average CO_2 data from Mauna Loa Observatory. This selection involved three steps: a preliminary selection based on within hour variability of the CO_2 concentration determined by checking the analog output of the CO_2 analyzer on a strip chart recorder; an hour-to-hour concentration difference that rejects data which change by more than 0.25 ppm from one hour to the next; and a selection based on residuals from a spline fit. Examples are shown for the 1985 data, with emphasis on January and August. Of a total of 8227 hourly averages available in 1985, 3838 values (46.7%) remain after final selection. Summer months have a larger percentage of data removed than winter months.

1.1 Introduction

The Mauna Loa Observatory (MLO) on the island of Hawaii (19.5N, 155.6W, elevation 3400 m) is one of four baseline atmospheric monitoring stations operated by the Geophysical Monitoring for Climatic Change (GMCC) program of the National Oceanic and Atmospheric Administration (NOAA). The other three observatories are located at Barrow, Alaska; American Samoa; and the South Pole. Continuous measurements of atmospheric CO_2 concentrations have been made at these sites beginning at Barrow in 1973. The purpose of these measurements is to document the long term changes in CO_2 concentrations in the atmosphere at several latitudes. This requires that the data at each site represent background conditions. However, even though each site is at a remote location, local disturbances can occur. The data must then be selected so that only those measurements which are thought to represent background conditions are used. This report will explain the selection procedures used for the CO_2 measurements from MLO.

Continuous measurements of CO_2 concentrations have been made at MLO since May of 1974 by GMCC using a non-dispersive infrared (NDIR) CO_2 analyzer. The methods used for obtaining the data are explained in Komhyr et al. (1989). Concentrations are reported in units of parts per million by volume (ppm) of dry air. The shortest time span of CO_2 data routinely used by GMCC is an hourly average. These hourly CO_2 averages are selected to identify data thought to represent background conditions. The selection procedures to be explained were used on the MLO data from 1974 to 1985 by Thoning et al. (1989). Daily, monthly, and annual averages are then computed based on these selected hourly averages.

1.2 Data Selection

The objective in selecting data is to identify those hourly CO_2 values that are representative of well-mixed air at Mauna Loa. Disturbances due to local sources or sinks of CO_2 should be removed. Although data selection can be based on independent quantities, such as wind speed and wind direction, a statistical approach was used for the Mauna Loa data. The basic assumption used was that the concentration of CO_2 in well-mixed air should change by only a small amount over several hours or more. This is based on the observation that at very remote sites, such as South Pole, Antarctica, the within-hour and hour-to-hour variability are very small (<0.1 ppm, Gillette et al., 1987). Three steps were involved in the selection procedure; examining the within-hour variability, the hour-to-hour change in concentration, and the departures from a smooth spline curve fit to a month of hourly CO averages.

Because of the large amount of hourly data for the entire record, this report will use only one year of data, 1985, as an example of the results of the selection. January and August, representing typical winter and summer months will be shown in more detail since the characteristics of the CO_2 signal are different for these times of the year. The raw data sets for these periods are shown in Figs. 1 and 2.

The CO_2 signal at Mauna Loa consists of a seasonal cycle with a peak-to-peak range of 6-7 ppm, a frequent diurnal cycle with a drawdown in CO_2 concentration of several ppm during the afternoon, and frequent, random bursts of high CO_2 concentrations usually seen during the night. The diurnal signal is caused by air being brought up the mountain that has been depleted of CO_2 by vegetation at lower elevations. The high nighttime values are due to CO_2 from volcanic vents being brought to the observatory by the downslope airflow. These signals are local phenomena, and those hourly values affected by them should be removed in the selection process.

1.2.1 Preliminary selection

The first step in the data selection scheme is a comparison of the hourly average CO_2 concentrations with an analog recording of the CO_2 analyzer output on a stripchart recorder. Hours that have significantly more variability in the air trace than in the reference gas traces are removed. Results of this selection are shown in Figs. 3 and 4 and in Table 1. Approximately 23% of the data are removed in this step. This step removes all of the hourly values with volcanic contamination, since these hours have very high variability within one hour. The diurnal drawdown in the afternoon still remains since the within-hour variability for these hours is usually very small.

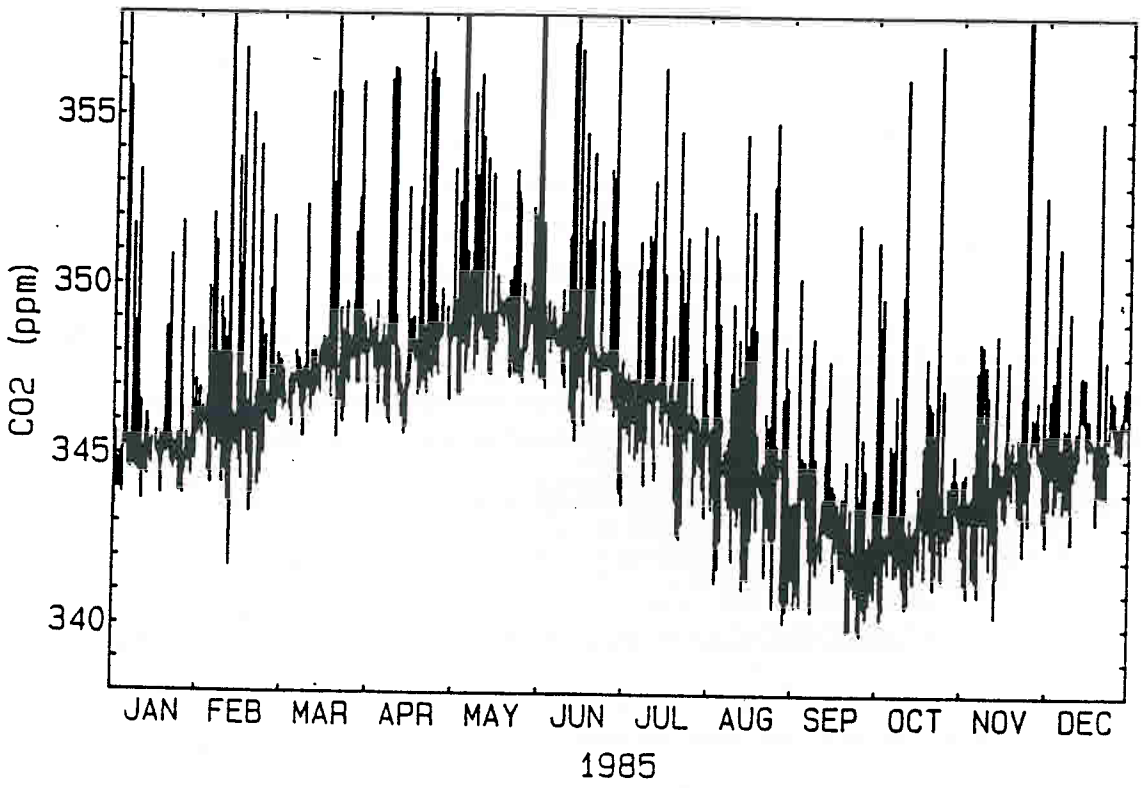

Figure 1. Hourly averaged CO_2 data from Mauna Loa Observatory for 1985. No data selection has been applied.

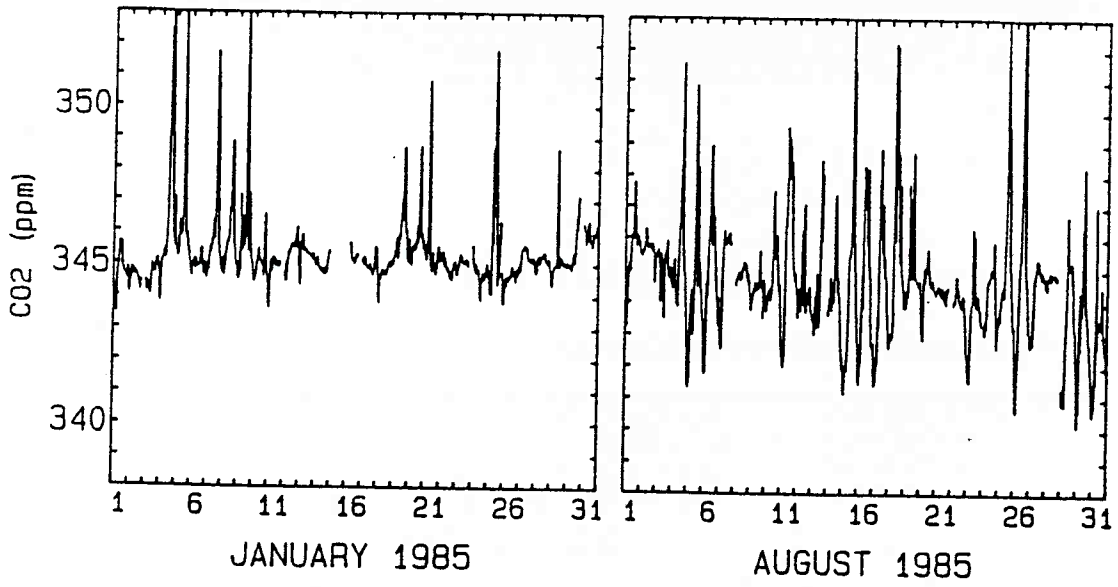

Figure 2. Hourly averaged CO_2 data for the months of January and August 1985. No data selection has been applied.

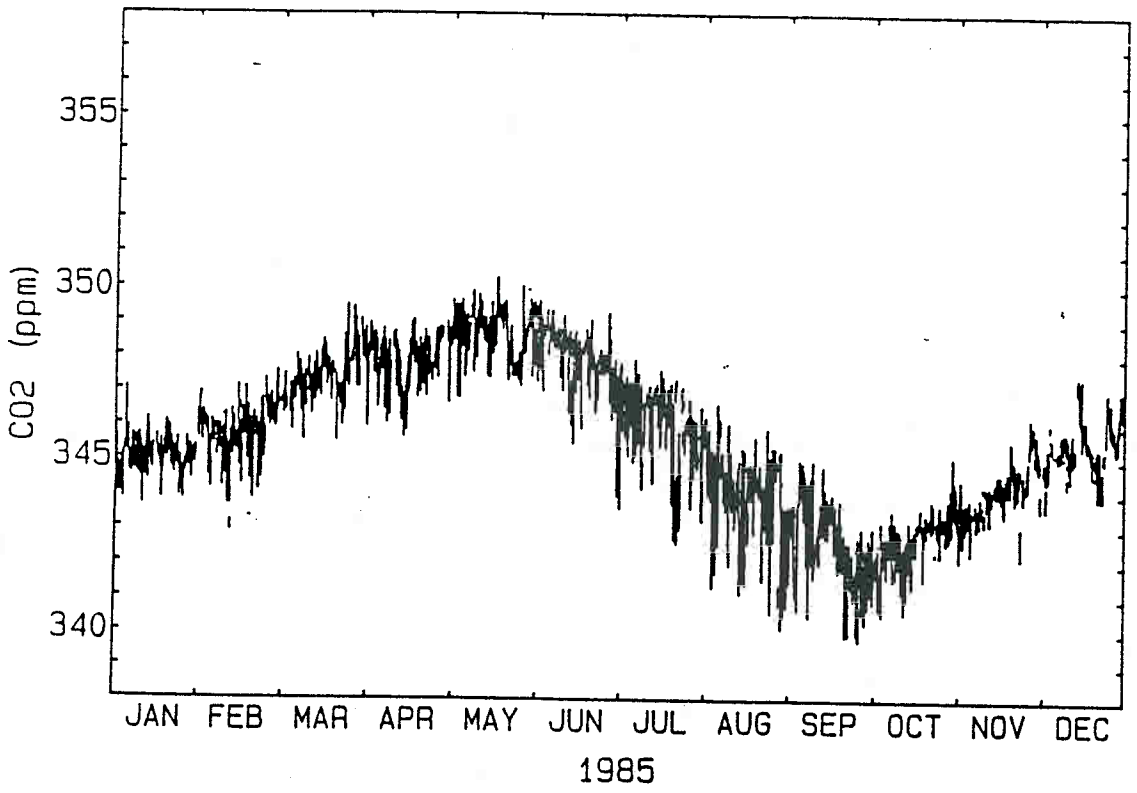

Figure 3. Hourly average CO_2 at MLO for 1985 after preliminary selection.

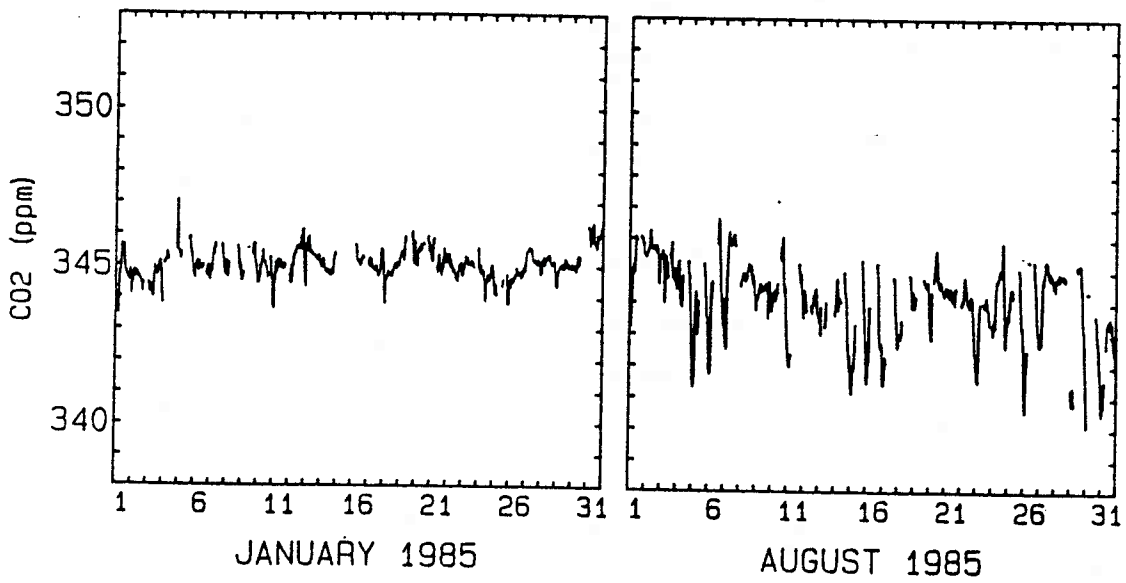

Figure 4. Hourly average CO_2 at MLO for January and August 1985 after preliminary selection.

Table 1.--Results of selection steps for 1985 MLO CO_2

	Selection step							
	RAW		Prelim.		Hr.-Hr.		Spline	
	N	Avg.	N	Avg.	N	Avg.	N	Avg.
January	682	345.39	584	345.03	543	345.05	434	345.02
August	713	344.62	558	344.05	406	344.15	223	344.69
Annual	8227	346.04	6346	345.72	5395	345.77	3838	345.92

N refers to the number of hourly averages remaining after each selection step. Avg. is the monthly or annual average based on the remaining hourly values.

Until 1986, subjective criteria were used to identify variable hours. Therefore it is likely that some inconsistency is present during the preliminary selection. However, most of the variable hours at MLO are easily determined, so that this inconsistency probably affects only a small part of the data.

Beginning with the July, 1986 data, variable hours have been determined objectively. The data acquisition unit at MLO (CAMS, Herbert et al., 1986) stores 1-minute average CO_2 analyzer voltages. The standard deviation of these 1-minute average voltages about the hourly average voltage is computed. The standard deviation is then converted to concentration (ppm). If the hourly standard deviation is above 0.3 ppm than that hour is identified as variable.

1.2.2 Hour-to-hour difference selection

The second step in the selection process is to flag hourly values whose concentrations changed by more than a specified amount from one hour to the next. Peterson et al. (1982) stipulated that if the difference between consecutive hourly average concentrations was larger than 0.25 ppm, then both of those hourly values should be rejected. A slight modification of this practice was used here. Any two consecutive hourly values where the difference was less than 0.25 ppm were retained. Thus if the CO_2 concentration was steady and then abruptly changed, no hourly values would be flagged since it is not possible to distinguish which level is background. If the concentration changed rapidly over many hours all of the hourly values would be flagged. The flags applied during the preliminary selection step were not changed.

The results of this selection step are shown in Figs. 5 and 6 and Table 1. A total of 951 hourly values were removed. This is 11.6% of the raw data, or 15.2% of the data that remained after the preliminary selection. This selection only partially removed the diurnal signal. The rate of the afternoon drawdown in CO_2 is approximately the same as the cutoff criterion of 0.25 ppm per hour, so that many hours that were affected by the drawdown

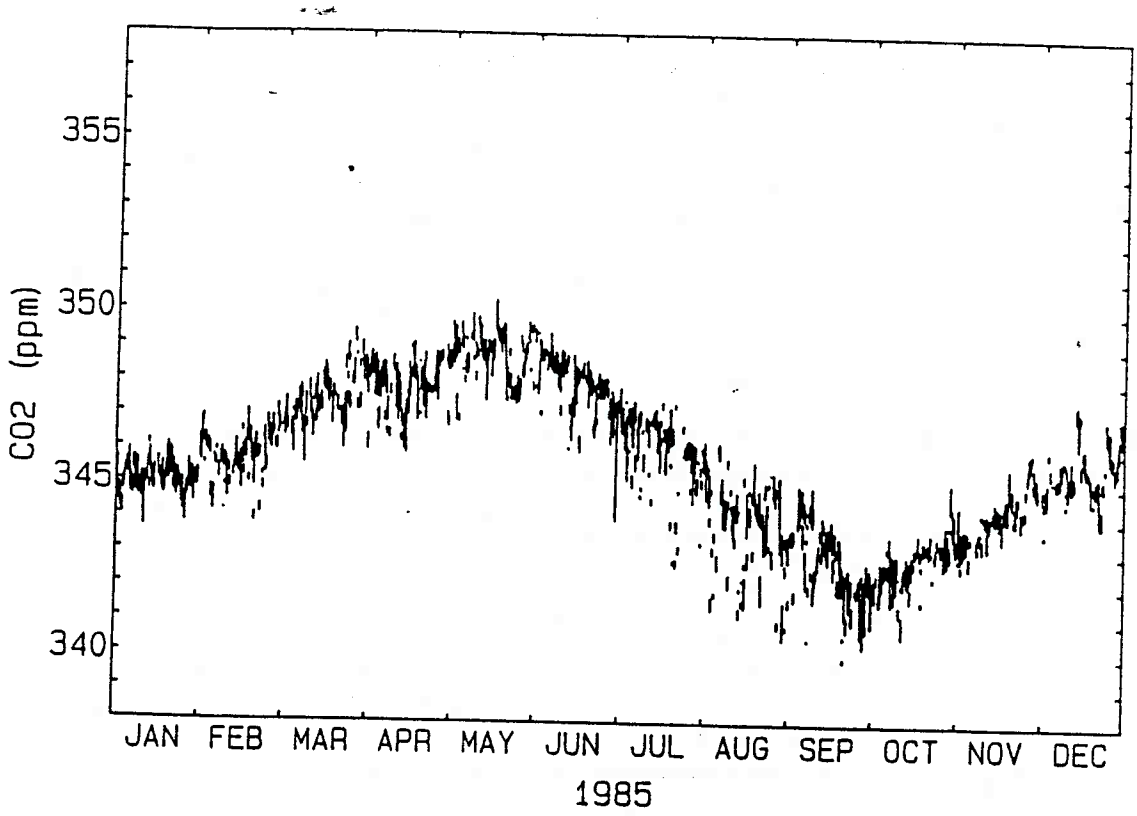

Figure 5. Hourly average CO_2 data for 1985 from MLO after preliminary selection and hour-to-hour difference selection.

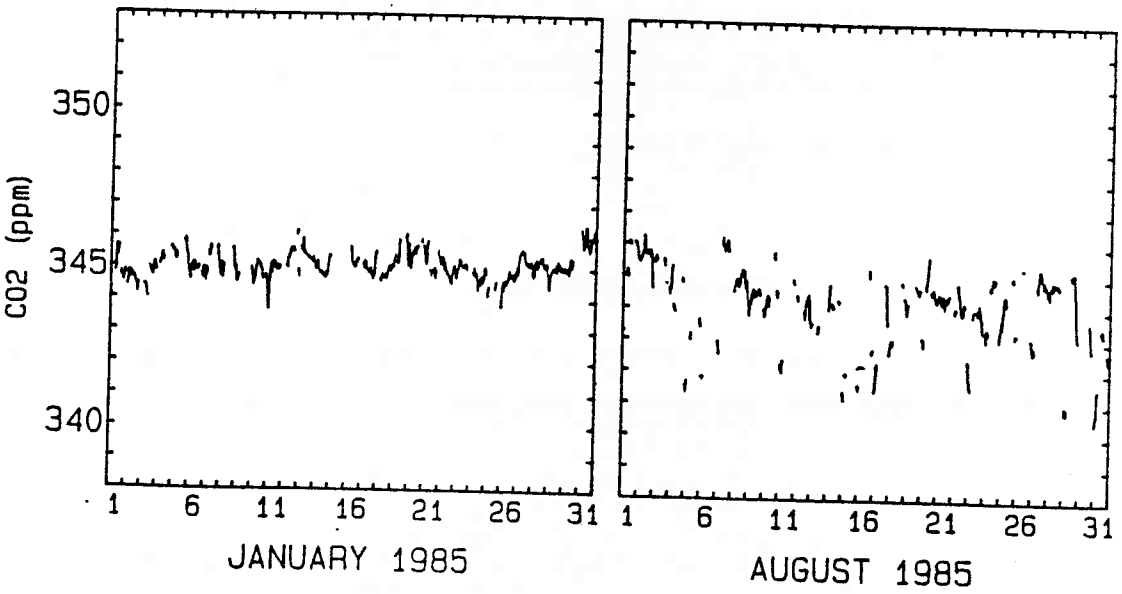

Figure 6. Hourly average CO_2 data from MLO for January and August, 1985 after preliminary and hour-to-hour difference selection.

remained. Several hours of each day when the drawdown is at its lowest point tend to have steady CO_2 concentrations.

To test the sensitivity of this selection step on the hour-to-hour difference cutoff value, the 1985 data were selected using cutoff criteria of 0.1, 0.3 and 0.5 ppm. The results from these different selections as well as the original value of 0.25 ppm are shown in Table 2. The results for the month of January changed very little, with a range from 345.02 to 345.05 ppm. The results for August changed the most, with a maximum difference in the monthly mean of 0.25 ppm between the 0.1 and 0.5 ppm cutoffs. The annual mean changed by 0.13 ppm or less. There was a large difference in the number of hourly values that were removed by the different cutoff values. The 0.1 ppm cutoff removed 45% of the available data (the data remaining after preliminary selection) for the entire year, and almost 60% of the data for August. The selection using the 0.5 ppm cutoff removed only 4% of the data for the entire year, with about 10% removed in August and less than 1% for January.

Table 2.--Results of hour-to-hour selection using different cutoff values

	Hour-to-hour difference cutoff (ppm)							
	0.1		0.25		0.3		0.5	
	N	Avg.	N	Avg.	N	Avg.	N	Avg.
January	375	345.02	543	345.05	559	345.04	579	345.03
August	226	344.30	406	344.15	433	344.13	503	344.05
Annual	3492	345.87	5395	345.77	5634	345.77	6099	345.74

N refers to the number of hourly averages remaining after the selection step. Avg. is the annual or monthly average based on the remaining hourly values.

1.2.3 Residuals from spline fit selection

Although the first two selection processes removed much of the obviously variable data, visual inspection shows that most of the afternoon drawdown in CO_2 is still present, as well as groups of hourly values that stand out from the bulk of the remaining data. Since the diurnal cycle is clearly a local phenomenon, the hourly values affected by the afternoon drawdown should not be included in the background data set. But because the afternoon drawdown does not occur 100% of the time, those values that are not affected by the diurnal signal should be retained. A final selection step was used to remove data exhibiting afternoon diurnal variations and any outliers that remain for the other times of the day.

The method used was an iterative process, working with one month of hourly data at a time. Figure 7 shows a flowchart diagram of the algorithm. First the daily averages and standard deviations were calculated using only

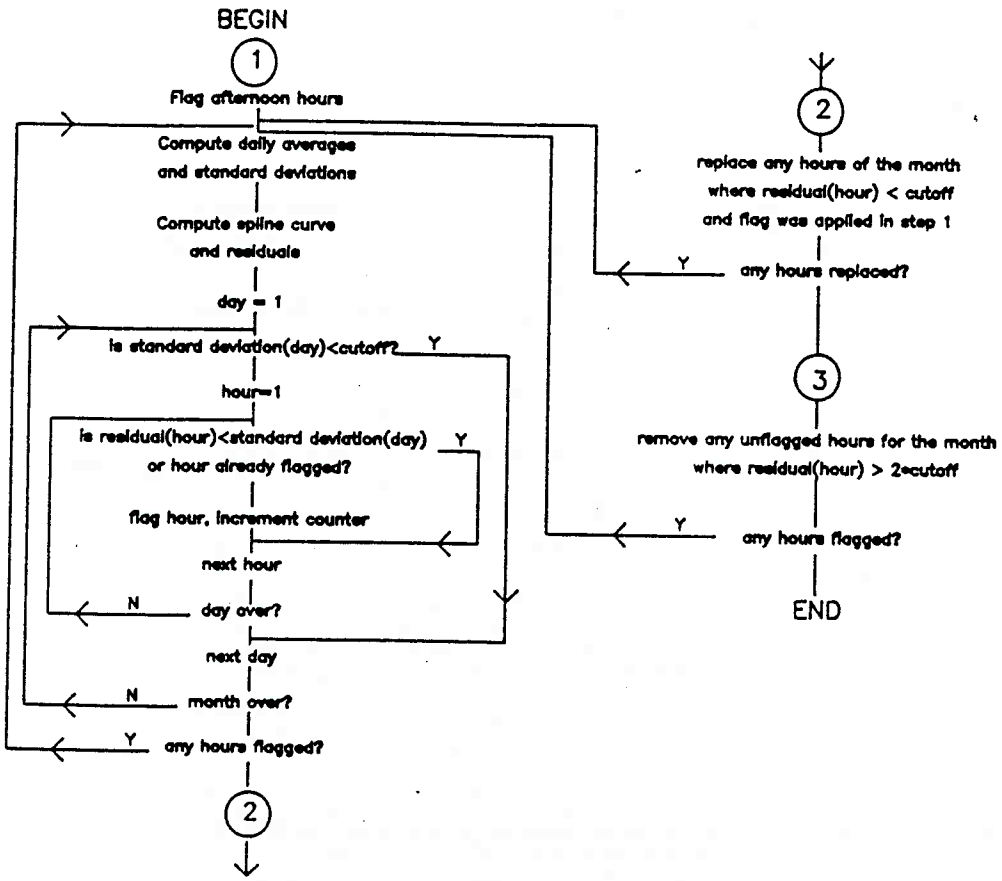

Figure 7. Flowchart diagram of algorithm used in the final selection step.

unflagged hourly values. A weighted spline curve (Reinsch, 1967) was fit to the daily averages using the corresponding standard deviations as weights. Next the residuals of the hourly data from the spline were calculated. Those hours with a residual value too large were removed, whereas those that fell within a cutoff value were retained or replaced if they had been removed during previous iterations. The selection continued until the standard deviation of the hourly averages about the daily average for each day fell below the cutoff value. Those hours when the diurnal drawdown usually occurred (11 a.m. to 7 p.m.) were temporarily removed before the first spline fit.

The cutoff value for this step was chosen to be 0.3 ppm. The results of this selection are shown in Fig. 8 and Fig. 9 and in Table 1. Superimposed on Fig. 9 are the final spline curves that were fit to the data. This step removed 1557 hourly values, or 19% of the raw data (or 29% of the data remaining after the hour-to-hour selection). Of these 1557 hours, 1061 (68%) were during the afternoon drawdown period (11 a.m. to 7 p.m.). Since 2086 hourly values were available during the afternoon period for the year, and because all of these hourly values were temporarily removed at the beginning, the final selection algorithm in effect replaced 49% of the afternoon values, while removing 6% of the original data during other times of the day.

To test the sensitivity of the selection to the cutoff value, the 1985 data were selected using cutoffs of 0.1, 0.2, and 0.5 ppm. The results of these selections are shown in Table 3. It shows that annual means and monthly means during the winter change very little with different cutoffs, but that the summer months may have significant differences. However, a cutoff value of 0.1 is very restrictive and leaves very few data for a month. If these results are ignored, then the other three data sets agree to within 0.15 ppm.

Table 3.--Results of spline fit selection using different cutoff values

	Cutoff (ppm)							
	0.1		0.2		0.3		0.5	
	N	Avg.	N	Avg.	N	Avg.	N	Avg.
January	168	345.07	345	345.03	434	345.02	515	345.06
August	103	344.80	205	344.58	223	344.69	287	344.54
Annual	1577	345.93	3184	345.97	3838	345.92	4577	345.92

N refers to the number of hourly averages remaining after the selection step. Avg. is the annual or monthly average based on the remaining hourly values.

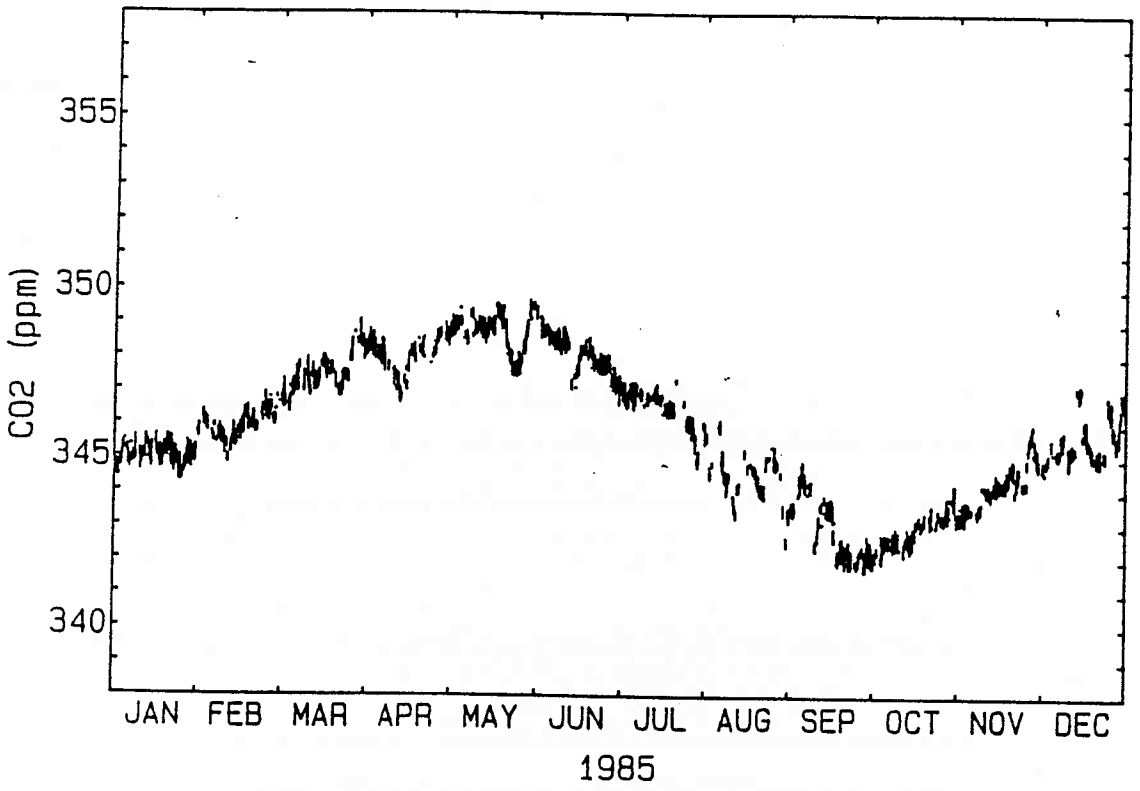

Figure 8. Hourly average CO_2 data from MLO after selection based on residuals from a spline curve fit.

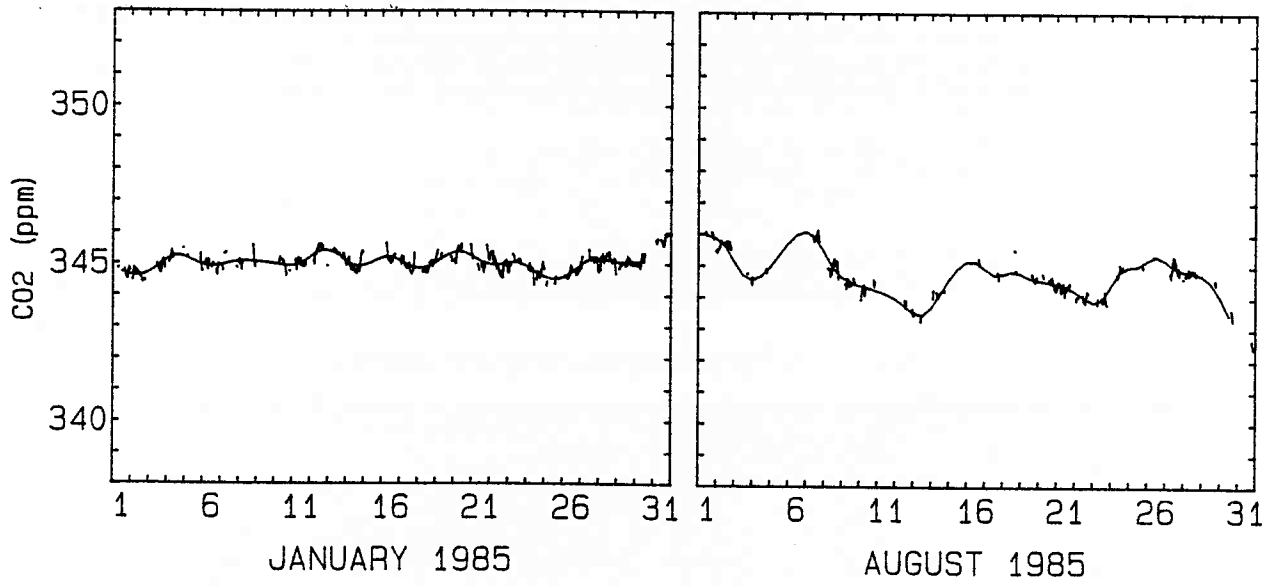

Figure 9. Hourly average CO_2 data for January and August, 1985 after selection based on residuals from a spline curve fit.

1.3 Results

The selection process described has been applied to the NOAA/GMCC Mauna Loa data from 1974 to 1986. Appendix A contains daily values computed from the selected hourly data for these years in the format adopted by the Environmental Pollution Monitoring and Research Programme of the World Meteorological Organization. Monthly means computed from daily averages and annual means calculated using monthly means are also included.

For the example year of 1985, data selection removed 53.3% of the available hourly values for the entire year, 36.4% for January, and 68.7% for August. Thus only approximately half of the data that are obtained can be considered representative of background conditions. This shows that even at remote locations it is not possible to make background measurements all of the time. The difference in the number of hourly values remaining between January and August is to be expected considering the nature of the diurnal signal at Mauna Loa. During the summer the afternoon drawdown in CO_2 is stronger and more frequent than in the winter. This requires that more values be rejected for the summer. The affect of volcanic outgassing on data selection is more random, but is more common during times of high volcanic activity, such as after the eruption of Mauna Loa in 1984.

Although this selection process could be used for data from other sites, experience shows that different sites will require different selection schemes. The preliminary selection step is used on a routine basis for the CO_2 data from each of the four GMCC observatories. The hour to hour selection was used for data from Barrow, Alaska (Peterson et al., 1982, Peterson et al., 1986) and for data from Samoa (Waterman et al., 1989). The Samoa data were also selected based on wind speed and wind direction. South Pole data from 1976 to 1982 were selected by removing values in the order of their absolute difference from a five day mean until the variance of the five day mean fell below a threshold value (Gillette et al., 1987). In each case the final selection method was tailored to the site and to the behavior of the CO_2 signal observed. For example, it is advantageous to select data by wind direction and wind speed for those sites that have a well-defined clean air sector. It is doubtful that one selection method is optimal for every site.

1.4 References

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

Herbert, G.A., E.R. Green, J.M. Harris, G.L. Koenig, S.J. Roughton, and K.W. Thaut, 1986. Control and monitoring instrumentation for the continuous measurement of atmospheric CO_2 and meteorological variables. J. Atmos. and Oceanic Tech., 3, 416-421.

Komhyr, W.D., T.B. Harris, L.S. Waterman, J.F.S. Chin, and K.W. Thoning, 1989. Atmospheric carbon dioxide at Mauna Loa Observatory: I. NOAA/GMCC measurements with a nondispersive infrared analyzer, 1974-1985. J. Geophys. Res., (in press).

Peterson, J.T., W.D. Komhyr, T.B. Harris, and L.S Waterman, 1982. Atmospheric carbon dioxide measurements at Barrow, Alaska, 1973-1979. Tellus, 34, 166-175.

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

Reinsch, C.H., 1967. Smoothing by spline functions. Numerische Mathematik, 10, 177-183.

Thoning, K.W., P.P. Tans, and W.D. Komhyr, 1989. Atmospheric carbon dioxide at Mauna Loa Observatory: II. Analysis of the NOAA/GMCC data, 1974-1985. J. Geophys. Res., (in press).

Waterman, L.S., D.N. Nelson, W.D. Komhyr, T.B. Harris, and K.W. Thoning, 1989. Atmospheric carbon dioxide measurements at Cape Matatula, American Samoa, 1976-1984. J. Geophys. Res., (in press).

APPENDIX A

Appendix A contains daily, monthly, and annual means of CO_2 concentrations from Mauna Loa Observatory for 1974 through 1986. Daily averages are computed from selected hourly averaged data. Monthly averages are computed from the daily averages, and annual means are calculated from the monthly averages.

NOAA/GMCC atmospheric carbon dioxide measurements

Program at MAUNA LOA, HAWAII
Daily mean values for the year 1974

Provisional CO2 concentration values (in ppm above 300.0)
(continuous analyzer)

Day	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Day
1							31.69	29.63	27.70	27.51	27.71	28.67	1
2							31.41	29.65	27.67	27.24	27.82	28.69	2
3							31.53	29.79	28.07	27.18	27.90	28.88	3
4						33.04	31.68	30.13	28.19	27.09	27.98	28.90	4
5						32.61	31.55	30.43	28.14	26.70	27.86	29.09	5
6						31.87	31.49	29.91	27.89	26.98	27.91	28.89	6
7						32.04	31.31	30.23	27.90	27.19	27.96	28.76	7
8						32.01	31.37	30.14	28.00	27.16	27.87	29.21	8
9						32.35	31.59	29.73	27.84	26.88	27.85	29.61	9
10						32.23	31.44	29.31	27.80	26.97	27.88	29.42	10
11						32.13	31.50	29.19	27.56	26.74	27.92	29.33	11
12						32.02	31.24	29.37	27.91	27.22		29.16	12
13						31.85	31.24	29.32	27.63	27.15	28.19	29.50	13
14						32.49	31.03	28.92	27.22	27.22	28.19	29.30	14
15						32.11	31.09	28.71	27.14		28.34	29.46	15
16						32.07	30.84	29.14		27.25	28.86	29.48	16
17					33.35	32.13	30.31	29.36	26.82	27.20	28.85	29.47	17
18					33.39	32.32	30.73	29.37	26.79	27.10	28.66	29.75	18
19					33.46	32.11	30.47	29.32	26.98	27.30	28.33	29.50	19
20					33.48	32.29	30.67	29.06	26.82	27.34	28.59	29.96	20
21					33.20	32.46	30.64	29.31	26.58		28.54	29.96	21
22					33.18	32.33	30.78	28.52	26.63		28.53	29.92	22
23					33.18	32.23	30.73	28.75	26.45	27.27	28.50	30.15	23
24					32.95	31.69	31.18	28.58	26.41	27.38	28.59	29.77	24
25					33.10	31.98	31.21	28.69	26.81	27.42		29.77	25
26					33.20	31.56	30.53	28.86	27.27	27.41	28.42	30.31	26
27					32.77	31.28	30.45	28.35	27.46	27.56	28.41	29.76	27
28					32.43	31.49	30.36	28.01	27.71	27.86	28.60	29.92	28
29					32.71	31.45	30.38	28.16	27.32	27.49	28.27	29.75	29
30					33.29	31.83	30.29	28.26	27.60	27.47	28.63	30.02	30
31					33.25		30.13	28.11		27.51		30.25	31

Monthly values

Xm	33.13	32.07	31.00	29.17	27.39	27.24	28.26	29.50	Xm
N	15	27	31	31	29	28	28	31	N
s	0.30	0.38	0.48	0.66	0.54	0.25	0.35	0.47	s

Annual mean 329.72 ppmv
(based on 8 monthly means)

Xm - Average monthly concentration
N - Number of daily values accepted
s - Standard deviation (ppmv), based on daily means

NOAA/GMCC atmospheric carbon dioxide measurements

Program at MAUNA LOA, HAWAII
Daily mean values for the year 1975

Provisional CO2 concentration values (in ppm above 300.0)
(continuous analyzer)

Day	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Day
1	29.90	30.79	32.50	33.44	33.90	34.31	33.60	30.88	28.84	27.63	29.49		1
2	30.19	30.96	32.17	33.46	34.03	34.00	33.04	30.62	29.48	27.70	29.35		2
3	30.60	30.64		33.21	34.25	34.01	32.86	30.95	29.31		28.95		3
4	30.72	31.67		33.21	33.99	33.53	32.85	30.97	29.44		28.64		4
5	30.67	31.63	31.89	32.60	33.95	33.97	32.67	30.91	29.59		29.16		5
6	30.59	31.86	31.61	32.69	33.95	34.11	32.69	30.76	29.59		28.99		6
7	30.37	31.71	31.62	32.83	33.85	33.84		30.93			29.28		7
8	30.50	31.57	31.62	32.58	33.96	33.84	32.20	30.95	29.34		28.86		8
9	30.39	31.63	31.53	32.67	33.88	33.30	32.13	31.12	29.34		29.03		9
10		31.05	31.60	33.04	33.95	33.26	32.04	30.70	29.76		29.10		10
11	30.37	31.12	31.50	33.71	34.03	33.15	31.90		29.35		28.96		11
12	30.47	30.75		33.98	33.93	32.77	32.09	30.40	29.17		29.01		12
13	30.42	31.01	31.66	34.41	33.99	32.68	31.74	30.30	28.99		28.81		13
14	30.74	31.02		34.18	34.61		31.97	30.50	28.73		29.00		14
15	30.73	30.97	31.58	33.51	34.54	32.95		30.57	27.95	28.38	29.03		15
16	30.33	31.12		33.60	34.22	33.30	31.39	30.39	27.90	28.35	29.20		16
17	30.18	31.38	31.87	32.93	33.25	33.41		30.12	27.61	28.42	29.18		17
18	30.03	31.25	32.10		33.44	33.35	30.87	30.16	27.85	28.41	29.39		18
19	30.46	31.55	31.79	33.21		33.24	31.03	29.63	27.67	28.41	29.40		19
20	30.66	31.49	32.19	33.18	33.20			29.64	28.04	28.44	29.58		20
21	30.59	31.41	32.11	33.40	33.29	32.96	30.62	29.51	27.92	28.67	29.48		21
22	30.37	31.72	32.11	33.56	33.25	33.34	30.78	29.27	27.98				22
23	30.49	32.18	32.04	33.66	33.71	33.63	31.36	28.95	28.11	28.77			23
24	30.74	32.00	31.90	33.73	33.46	33.63	31.31	28.98	28.11	28.67			24
25	31.73	31.95	31.59	33.15	33.91	33.46	31.48	28.06	27.96	28.64			25
26	31.61	31.81		32.75	34.24	33.38	31.34	28.11	27.91	28.65			26
27	31.42	31.70	31.75	33.01	34.21	32.91		28.58	27.91	28.66	30.06		27
28	31.74	31.33	31.61	32.88	34.01	32.94	30.93	28.82	28.12	28.52	30.22		28
29	31.70		31.61	32.80	34.03	32.82		29.03	28.26	28.74	30.08		29
30	31.03		32.10	33.09	34.12	32.89	31.11	28.64	27.97	28.55	29.80		30
31	30.93				34.22		31.20	28.53					31
					Monthly values								
Xm	30.69	31.40	31.84	33.26	33.91	33.39	31.81	29.90	28.56	28.45	29.28		Xm
N	30	28	24	29	30	28	25	30	29	17	25		N
s	0.50	0.41	0.27	0.47	0.36	0.45	0.80	0.97	0.72	0.32	0.41		s

Annual mean 331.13 ppmv
(based on 11 monthly means)

Xm - Average monthly concentration
N - Number of daily values accepted
s - Standard deviation (ppmv), based on daily means

NOAA/GMCC atmospheric carbon dioxide measurements

Program at MAUNA LOA, HAWAII
Daily mean values for the year 1976

Provisional CO2 concentration values (in ppm above 300.0)
(continuous analyzer)

Day	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Day
1		32.27	31.56	34.38	34.86	34.51		32.03	30.58	29.10	29.18	30.69	1
2			31.93	34.57	35.03	34.40		31.74	30.34	28.86	29.24	30.78	2
3		32.36		34.70	34.89	34.23		31.85	29.66	28.67	29.64	31.00	3
4		32.11		34.97	34.94	34.20		31.98		28.14	29.75	30.52	4
5		32.24	32.17		34.34	34.10		31.78	29.47	28.15	29.76		5
6	31.54	32.03	32.54	34.84	34.05	34.23		31.64	29.23	27.90			6
7	31.39	32.06	32.61	35.43	33.69	33.88	33.83	31.39	29.28	28.53			7
8	31.21	32.25	32.86	35.07	33.56	34.36	33.66	31.66	29.46	28.67		30.88	8
9	31.18	32.45		35.23	33.70	34.22	33.63	31.58	29.93	28.91		31.28	9
10	31.20	32.08	33.04	34.84	34.56	34.01	33.29	31.51	29.88	28.90	29.74	31.20	10
11	31.19	32.73	33.08	34.48	34.64	34.60	33.40	31.37	30.14	28.56	29.83	31.36	11
12	31.23	32.37	33.27	33.91	34.67	34.60		30.53	30.02	27.98	29.77		12
13	31.51	32.10	33.52	34.08	34.07	34.46	33.56	30.37		27.65	29.91		13
14	31.71	32.29	34.15	34.15	34.08	34.46	33.50	29.86		27.83	30.23	31.88	14
15	31.82	32.77	33.92		34.09	34.15	33.34	30.09	29.72	28.09	30.30		15
16	31.66	32.64	33.74		34.69	34.18	32.82	30.16	29.10	28.57	29.93	32.08	16
17	31.88	32.35	33.45		34.85	34.27	32.35	30.03	29.08	28.89	30.27	32.06	17
18	31.70	32.29	33.17	35.20	35.29	34.12		30.30	29.04		30.16	31.88	18
19	31.85	32.69		34.77	35.41	34.03		30.10	28.58		30.11	31.61	19
20	31.81	32.50	33.65	34.59	35.40	33.89	32.58	29.84	28.48		30.29	31.79	20
21	31.84	32.69	33.56		35.40	34.00	32.62	30.25	28.18		30.46	31.32	21
22	31.79	33.42	33.81		35.38	33.76	32.37	30.41	27.68	29.12	30.67	31.63	22
23	31.83	33.78		34.45	34.93		32.65	30.68		29.07	30.51	31.66	23
24	31.92	33.75	34.07	34.44	35.08	33.39	32.39	30.44		28.45	30.29	31.78	24
25	31.69	33.66	34.24	34.45	35.12	33.21	32.65	29.94	27.72	28.54	30.41	31.91	25
26	32.01	33.52	34.30	34.73	35.36			29.92	27.72	28.50	30.51	32.24	26
27	32.19	33.35	34.53	35.05	35.59			30.01	27.76	28.65	30.67		27
28	31.63	32.88	34.49	35.54	35.80		32.35	29.29	27.75	28.85	30.56	32.23	28
29		32.64	34.60	35.41	35.36		32.33	29.27	28.11	29.12	30.48	32.42	29
30			34.63	35.89	34.76		31.97	29.39	28.32	29.55	30.55	32.43	30
31	31.84		34.83		34.80		32.04	30.27		29.25		32.34	31

Monthly values

Xm	31.65	32.65	33.53	34.80	34.79	34.14	32.87	30.63	29.01	28.61	30.12	31.62	Xm
N	24	28	26	24	31	24	20	31	25	27	26	24	N
s	0.29	0.55	0.88	0.49	0.60	0.34	0.60	0.86	0.92	0.47	0.42	0.57	s

Annual mean 332.04 ppmv
(based on 12 monthly means)

Xm - Average monthly concentration
N - Number of daily values accepted
s - Standard deviation (ppmv), based on daily means

NOAA/GMCC atmospheric carbon dioxide measurements

Program at MAUNA LOA, HAWAII
Daily mean values for the year 1977

Provisional CO2 concentration values (in ppm above 300.0)
(continuous analyzer)

Day	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Day
1	32.38	33.05	34.32	34.92	36.82	36.58	35.64	33.61	32.06	31.08	31.85	32.49	1
2	32.28	32.93	34.10	35.43	36.85	36.62	35.37	33.60	31.77	30.92	31.69	32.66	2
3		32.82	34.61	35.91	36.69	36.55	35.53		31.74	31.06	31.76	32.63	3
4	32.35	33.12	34.78	36.14	36.84	36.21	35.50	33.48	31.68	30.93	31.72	32.82	4
5		33.26	34.70	36.07	37.17	36.14		33.28	31.72	30.76	31.95	33.34	5
6	32.09	32.87	34.62	35.80	36.78		35.44	32.79	31.42	30.82	32.04	33.02	6
7	32.20	33.02	34.10	35.97	36.66	36.28	35.32	33.02	31.47	30.86	32.25	33.11	7
8	32.19	33.06	33.83	35.90	37.01	36.15	35.20	32.78	31.36	30.76		33.09	8
9	32.16	33.83	34.19	35.84	36.64	36.16		32.29			31.97	33.45	9
10	32.16	33.83	34.26			36.04		31.97		30.85	31.84	33.49	10
11	32.46	33.66		35.65	36.66	36.27	34.69		31.21	31.31	32.07	33.29	11
12	32.36	33.32	34.66	35.37	36.24	36.13	35.04	32.58	31.11	31.32	32.25	33.14	12
13	32.78		34.79	36.01	36.54	36.06	35.11	32.98	31.76	31.61	32.59	33.38	13
14	33.00	33.29	35.26	36.14	36.87	36.37	35.01	32.92	31.24	31.14	32.47	33.09	14
15	33.42	33.06	35.33	36.31	37.06	36.04	34.96	32.73	31.67	31.12	32.46	33.37	15
16	33.39	33.33	35.74	36.29	37.04	36.08	34.93	32.57	31.56	31.43	32.54	33.35	16
17	32.93	33.61	35.50	36.24	36.84	36.13	34.99	32.59	31.71		32.39	33.26	17
18		33.28	35.56	35.91		35.80	34.85	32.55	31.06	31.28	32.40	33.53	18
19	32.84	33.11	35.49	36.07	37.12	35.53	34.77	32.08		31.31	32.50	33.63	19
20	32.80	32.75	35.59	36.20	36.85		34.44	32.26		31.57	32.59	33.98	20
21	33.11	32.94	35.50	36.21	36.80		34.37	31.86		31.24	32.48	33.92	21
22	33.21	33.25	35.24	36.40	36.99	36.50	34.46	31.98	30.70	31.14		33.79	22
23	33.24	33.06	35.06	36.23	36.96	36.55	34.30		30.53	30.97	33.03	33.79	23
24	32.92	33.18	35.26	36.17	37.13	36.32	34.23	31.90		31.34	32.70	34.10	24
25	32.91	33.25	34.85	36.20	37.03	35.73	34.27	31.63	30.46	31.35	32.84	33.71	25
26	32.89	33.26	34.86	36.44	36.84	35.64	34.20	31.61	30.79	31.29	32.69	33.77	26
27	32.87	33.38	34.64	36.38	36.67	35.80	34.08	31.77	30.74	31.30	32.77	34.02	27
28	32.73	33.51		36.07	36.81	35.90	33.96	32.48	30.75	31.39	32.83	34.08	28
29	33.22		35.11	36.06	36.74		34.25	32.28	30.87	31.79	32.70	34.35	29
30	32.75		35.21		36.61	35.74	33.91	32.58	31.03	31.76	32.43	33.87	30
31	32.85		35.11		36.74		33.99	32.26				33.97	31

Monthly values

Xm	32.73	33.22	34.91	36.01	36.83	36.13	34.74	32.52	31.27	31.20	32.35	33.47	Xm
N	28	27	29	28	29	26	28	28	24	28	28	31	N
s	0.40	0.28	0.52	0.34	0.20	0.30	0.53	0.57	0.46	0.28	0.38	0.47	s

Annual mean 333.78 ppmv
(based on 12 monthly means)

Xm - Average monthly concentration
N - Number of daily values accepted
s - Standard deviation (ppmv), based on daily means

NOAA/GMCC atmospheric carbon dioxide measurements

Program at MAUNA LOA, HAWAII
Daily mean values for the year 1978

Provisional CO2 concentration values (in ppm above 300.0)
(continuous analyzer)

Day	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Day
1	34.41		35.12	37.57	38.39	38.64	37.35	35.51	33.95	32.60	33.22	34.38	1
2		35.38	35.98	37.42		38.64	37.15	35.11		32.45	33.17	33.94	2
3		35.29	36.67	37.16	38.32	38.54	37.27	34.94		32.36	33.57	34.21	3
4	35.07	35.20	36.39	36.90	37.92	38.49	37.24	35.37		32.26	33.37	34.08	4
5	34.81	34.68	36.15	37.06	38.27	38.45	37.16	35.13		32.01		34.42	5
6	34.63	34.34	36.17	37.73	38.49	38.01	36.62	35.09	33.48	32.05	33.74	34.59	6
7		34.47	35.57	37.77	38.27	38.22	36.84	34.74	32.89	32.10	33.47	34.44	7
8		34.79	35.45	38.10	37.96	38.06	36.67	34.52		31.98	33.77	34.50	8
9	34.10	34.97	36.13		37.76	38.26	36.78	34.54		31.87	33.57	34.64	9
10	34.06	35.34	36.13	38.04	37.72	38.39	37.02	34.27	31.58	31.70	33.51	34.61	10
11	34.52	35.34	36.16	38.07	38.04	38.07	36.94	34.58	31.71	31.52	33.62	34.50	11
12	34.55	35.10	36.35	37.83	38.54	38.06	36.83	34.69	31.61	31.57	33.56	34.72	12
13	34.52	35.02	35.80	38.07	38.36	38.44	36.64	35.03	31.67	31.59	33.54	34.88	13
14	34.73	34.86	36.15	37.87	38.18	38.40	36.63	34.78		31.84	33.58	35.07	14
15	34.91	35.27	36.06	38.25	38.40	38.05	36.47	34.66	32.28	32.27	33.63	35.13	15
16	34.85	35.76	36.86	37.98	37.96	38.00	36.47	34.92	32.34	32.06	33.74	35.03	16
17	34.95	35.71	37.09	37.99	38.06			34.70		32.34	33.51	35.01	17
18	35.19	35.83	37.26	37.69	37.61		35.84	34.24		32.50	33.60	35.11	18
19	35.04	35.54	36.94		37.27		35.75	33.47	31.31	32.39	34.32	35.11	19
20	35.05	35.06	36.69	37.14	37.46	37.76	36.00	33.16	31.89	32.70	34.36	35.07	20
21	34.75	34.81	36.74	37.77	37.73	38.24	36.00	33.43		32.89	34.51	35.02	21
22		34.92	36.75	38.04	37.90	37.94	36.09	33.75		32.65	34.10		22
23	34.80	35.88	36.69	37.92	37.97	37.79	36.00	33.89	31.80	32.94	34.02	35.08	23
24	34.59	35.85	36.83	37.80	37.74	37.36	35.90	33.64	32.24	32.63	33.77	34.93	24
25	35.17	35.92	37.41	37.69	37.63	37.37	35.54	33.88	31.97	32.78	34.16	34.95	25
26	34.98	35.15	37.43	37.67	37.34	37.30	35.57	34.02	31.80	32.60	34.08	34.95	26
27	34.68		37.80	37.73	37.44	37.31	35.74	33.47	32.52	32.87	33.87	35.12	27
28		35.13	37.50	37.73	38.03	37.56	35.57	33.43	33.26	32.68	33.96	35.05	28
29	34.75		36.65	38.16	38.21	37.71		33.81	33.26	32.76		35.04	29
30	34.38		36.39	38.32	38.40	37.82	35.82	33.78	33.00	32.66	33.99	35.05	30
31	34.79				38.45		35.44	33.71				35.33	31

Monthly values

Xm	34.73	35.22	36.51	37.77	37.99	38.03	36.39	34.33	32.35	32.32	33.76	34.80	Xm
N	25	26	30	28	30	27	29	31	19	30	28	30	N
s	0.30	0.44	0.63	0.36	0.36	0.41	0.61	0.66	0.76	0.42	0.34	0.35	s

Annual mean 335.35 ppmv
(based on 12 monthly means)

Xm - Average monthly concentration
N - Number of daily values accepted
s - Standard deviation (ppmv), based on daily means

NOAA/GMCC atmospheric carbon dioxide measurements

Program at MAUNA LOA, HAWAII
Daily mean values for the year 1979

Provisional CO2 concentration values (in ppm above 300.0)
(continuous analyzer)

Day	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Day
1	35.24	36.93	36.18	38.85	39.73	39.79	38.19		34.24	34.21	34.37	36.17	1
2	35.19	36.83	36.64	38.77	39.56	40.13	38.14		34.18	33.90	34.24	36.03	2
3	35.16	36.50	36.98	38.54	39.38	40.02		36.72		33.98	34.60	36.24	3
4	35.12	36.59	37.74		39.02	39.87	38.47	36.72		33.69	34.81	36.27	4
5	35.08	36.83	37.73		38.90	39.73	38.43			33.63	34.73	36.53	5
6	35.11	36.60	37.32		38.88	39.72	38.16	36.26	33.28	33.99	34.54	36.40	6
7	35.29	36.64	37.86			39.59	38.05	36.20	33.42	33.65	34.58	36.60	7
8	35.55		37.89	38.82	38.26	39.55	37.88	36.25	33.64		34.76	36.28	8
9	35.04	36.28	38.27	39.07	37.92	39.65	37.80	36.57	34.23		35.04	36.50	9
10	35.10	36.35	38.17		38.31	39.32	38.12	36.43	35.28		35.09	36.22	10
11	35.36	36.58	38.06	38.64	38.19	39.21	38.18	36.39	35.03	33.65	34.60	36.69	11
12	35.66	36.94	38.19	38.57	38.28	39.14	38.09	36.61	35.01	33.86	34.94	36.88	12
13	36.38	36.81	37.80		38.55		38.07	36.17	34.34	34.00	34.87	36.90	13
14	36.77	36.50	37.47	38.03	38.45	38.74	37.84	35.95	33.44	33.94	35.33	36.80	14
15	36.91	36.29		38.45	39.22	38.54	37.79	36.06	33.05	34.01	35.13	36.76	15
16	36.47	36.17	38.61	38.97	39.57	38.76	38.13	35.87	33.81	34.03	35.58	36.68	16
17	36.45	36.46	37.75	38.38		38.78	37.83	36.09	33.97	34.18	35.86	36.70	17
18	36.62	36.74	36.82	38.66	39.81	39.32	38.03	35.64	33.92	34.23	35.94	36.91	18
19	36.38	37.05	36.71	39.00	39.54	39.08	37.92	35.94	33.28	34.04	35.79		19
20	36.42	36.74	37.19	38.94		39.17		35.67	34.06	33.94	35.46		20
21	36.86	36.84		39.01		39.26	37.54	35.29	33.10	33.82	35.43	37.37	21
22		36.73	37.26	38.99		39.04	37.40	34.39	32.97	33.80	35.62		22
23	36.90	36.23		39.59		38.69	37.10	34.26	33.17	33.75	35.64		23
24	36.93	36.67		39.66		38.77	36.42	33.84	33.12	34.18	35.67		24
25	36.46			39.39		38.60	36.39	33.72	33.28	34.19	35.87		25
26	36.68	36.95	38.13	38.99		38.33			33.42	34.28	35.85		26
27	37.11	36.88	38.60	39.21		38.97	35.72	33.64		34.27	35.67	37.24	27
28	37.69	36.87	38.42		39.22	38.96	36.22	34.32	33.92	34.46	35.67	37.21	28
29	37.49		38.31	39.11	39.69	39.36	36.59	34.30	34.39	34.53	35.90	37.00	29
30	37.15		39.05	40.05	40.01	39.12	37.02	34.79	34.06	34.47	36.07	37.17	30
31	37.25		39.28				36.94	34.70		34.62		37.21	31

Monthly values

Xm	36.19	36.65	37.79	38.94	39.02	39.21	37.59	35.51	33.83	34.05	35.25	36.70	Xm
N	30	26	26	23	20	29	28	27	26	28	30	24	N
s	0.85	0.25	0.75	0.45	0.64	0.47	0.74	1.01	0.64	0.27	0.54	0.38	s

Annual mean 336.73 ppmv
(based on 12 monthly means)

Xm - Average monthly concentration
N - Number of daily values accepted
s - Standard deviation (ppmv), based on daily means

NOAA/GMCC atmospheric carbon dioxide measurements

Program at MAUNA LOA, HAWAII
Daily mean values for the year 1980

Provisional CO2 concentration values (in ppm above 300.0)
(continuous analyzer)

Day	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Day
1	37.87	37.62	40.26	40.43	41.31	41.92	40.59	38.94	36.76		36.74	37.28	1
2	37.88	37.82	40.16	40.44	41.42	41.83	40.58	38.61	36.54			37.57	2
3	37.73	38.38	39.65	40.34	41.60	41.63	40.40	37.36	36.37	35.97	36.84	37.52	3
4	37.67	38.38	39.12	40.56	41.69	41.56	40.41	36.82	36.56	35.64	36.73		4
5	37.57	38.22	39.63	40.37	41.69	41.52	40.35	36.08	37.11	35.71	37.22		5
6	37.25	38.14	39.53	40.71	42.21	41.34		36.73	37.03	35.92	37.09	38.07	6
7	37.77	38.46	39.42	40.91	41.84	41.59	39.50	36.54	36.26	35.90	37.03	38.18	7
8	37.36		39.09	40.74	41.46	41.74		37.35	35.67	35.67	36.78	38.16	8
9			39.24	40.46	41.01	41.60				35.43	36.61		9
10	37.01	37.22	39.28	40.99	41.29	42.09	38.90	38.45	36.17	35.59	36.94		10
11	37.48	37.27	39.88	40.59	41.07	41.50	38.96	38.19	35.90	35.50		38.13	11
12	37.51	37.77	40.09	40.86	41.56	41.55	38.93		35.90	35.99	37.78	38.13	12
13	37.70	38.36	40.15	41.15	41.62	41.46	39.21	38.36		36.08	37.64	38.25	13
14	37.42	38.33	40.10	41.11	41.30	41.37	40.05	37.85		35.89	36.96	38.41	14
15	37.38	38.95	40.05	40.94	41.37	42.21	40.05	37.70	36.21	35.96	37.12	38.53	15
16	37.15	38.85	39.87	40.86	41.45	41.97	39.66	38.27	35.82	35.95	37.10	38.58	16
17	37.50	38.70	39.48	40.76	41.61	41.95			36.11	36.14	37.00	38.51	17
18	37.85	38.65	39.37	40.76	41.60	41.74	38.74		36.18	35.88	37.15	38.53	18
19	37.92	38.55	39.63	40.74	41.68	41.23	38.74	37.97	35.20	35.65	37.19	38.52	19
20	38.17	38.30	39.53	41.07	41.53	40.91	38.95	38.20	34.73	36.12	37.24	38.44	20
21	37.72	38.34	39.92	40.81	41.75	40.70		37.71	34.94	36.22	37.21	38.25	21
22	38.24	37.86	40.35	40.98	41.27	40.69	38.90	38.74	35.18	36.49	37.30	38.42	22
23	38.76	37.97	40.42	41.19	41.36	40.78		38.39	34.93	36.53	37.58	38.58	23
24	38.63	37.98	40.71	41.01	40.75	40.62	39.51			36.42	37.75	38.25	24
25	38.56	38.28	40.70	40.95	40.67	40.46	39.20	37.80	35.75	36.56	37.67	38.46	25
26	38.21		41.08	41.17	41.04	40.51	38.51	38.16		36.53	37.56	38.23	26
27	38.32	38.78	40.82	41.27	41.34	40.63	39.03		36.23	36.09	37.57	38.66	27
28	38.32	38.51	41.28	41.25	41.79	40.67	39.05		36.18	36.14	37.57	38.64	28
29	37.77	38.81	41.30	41.12	41.95	40.72	38.82	38.12	35.67	36.36	37.55	38.63	29
30	37.90		41.10	41.44	41.82	40.38	38.40		36.21	36.52		38.89	30
31	37.54		40.89		41.76		38.41			37.07		38.93	31

Monthly values

Xm	37.81	38.25	40.07	40.87	41.48	41.30	39.35	37.83	35.98	36.07	37.22	38.32	Xm
N	30	26	31	30	31	30	25	22	25	29	27	27	N
s	0.44	0.46	0.66	0.29	0.34	0.55	0.71	0.75	0.63	0.38	0.34	0.38	s

Annual mean 338.71 ppmv
(based on 12 monthly means)

Xm - Average monthly concentration
N - Number of daily values accepted
s - Standard deviation (ppmv), based on daily means

NOAA/GMCC atmospheric carbon dioxide measurements

Program at MAUNA LOA, HAWAII
Daily mean values for the year 1981

Provisional CO2 concentration values (in ppm above 300.0)
(continuous analyzer)

Day	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Day
1	38.46	39.55	41.50	41.54	43.17	42.46	41.22	39.68	37.93	35.61	37.70	39.68	1
2	38.36	39.51	41.35	42.01	43.18		41.11	39.55	37.84	36.01	37.88		2
3	38.32	39.45	41.29	42.53	43.21		40.89	39.67		36.62	38.33	39.68	3
4	39.51	39.67	40.85	42.39	43.21	43.00	41.30	39.03	37.32	36.42	38.33	39.80	4
5	39.81	39.74	40.63	42.35	43.10	42.83	41.29	39.23	37.17	36.52	38.33	39.72	5
6	39.72	40.33	40.60	42.72	42.93	43.10	41.51	39.32	37.36	36.48	38.25	39.55	6
7	39.46	40.75	40.86	43.06	43.14	42.99	41.47	39.02	37.74	36.77	37.99	39.31	7
8	38.76	41.39	41.09	42.86	42.98	42.90	41.48	39.42	37.12	36.60	38.26		8
9	39.07	41.69	41.19	42.88	42.93	42.77	41.12	39.31	36.87	36.65	37.81	40.03	9
10	39.04	41.32	41.56	42.75	42.95	42.82	41.20	39.16	37.09	37.10	37.75	39.84	10
11	39.01	40.28	41.28	42.53	43.01	42.77	40.94	38.92	37.34	37.12	38.21	39.60	11
12	39.24	40.39	41.59	42.59	43.14	42.78	41.32	38.89	37.39	36.78	38.45	39.40	12
13	39.61	40.09		42.29	42.91	42.79	41.23	38.83	37.35	36.90	38.57	39.49	13
14	39.46	39.88			43.15	42.70	40.87	38.88	36.87	36.99	38.30	39.60	14
15	39.17	39.79		42.48	43.08	42.62	40.59	38.69	36.76	36.96	38.07	39.41	15
16	39.02	40.01	41.42	42.73	42.97	42.75	40.61	38.61	36.38	37.25	38.09	39.60	16
17	39.05		41.55	42.54	42.96	42.51	40.26	38.49	36.44	37.34	37.89	39.36	17
18	39.60	40.24	41.42		43.03	42.18		38.40	36.98	37.49	38.30	39.73	18
19	39.54	40.17	41.98	42.54	43.04	41.99		38.15	37.54	37.15	38.76	39.70	19
20	39.34	40.19	41.57	42.60	42.93	42.27	40.99	38.23	37.56	37.55	38.61	39.86	20
21	39.72	39.71	41.66	42.62	42.89	42.30	40.82		37.21	37.21	38.62	39.88	21
22	39.62	40.21	42.40	42.94	43.02	41.98	40.60		36.93	37.79	38.53	39.91	22
23	39.53	40.70	42.56	42.50	42.77	42.22	40.54	37.40	36.83	37.55	38.82	39.97	23
24	39.38	40.87	42.71	42.21	42.80	42.19	40.74	37.09		37.66	38.95	39.82	24
25	39.33	41.61	42.74	42.37	43.01	41.79	40.42	37.39		37.51	39.48	40.17	25
26	39.49	41.63	42.46	42.18	43.03	41.66	40.54			37.72	39.37	40.44	26
27	39.39	41.71	42.22	42.12	43.15	41.69	39.98	37.56	36.65	37.75	39.31	40.62	27
28	39.99	41.80	42.41	42.33	43.08	42.04	39.75	37.92	36.71	37.85	39.26	40.74	28
29	39.99		42.44	42.30	43.02	42.45	40.26	38.06	35.97	37.92		40.47	29
30	40.09		42.58	42.37	42.85	42.33	39.91	38.00	35.71	37.84	39.67		30
31	39.59		42.05		42.90		39.68	37.99		37.72		40.41	31

Monthly values

Xm	39.34	40.47	41.71	42.48	43.02	42.46	40.78	38.60	37.04	37.12	38.48	39.85	Xm
N	31	27	28	28	31	28	29	28	26	31	29	28	N
s	0.45	0.77	0.65	0.31	0.12	0.41	0.53	0.73	0.53	0.58	0.54	0.39	s

Annual mean 340.11 ppmv
(based on 12 monthly means)

Xm - Average monthly concentration
N - Number of daily values accepted
s - Standard deviation (ppmv), based on daily means

NOAA/GMCC atmospheric carbon dioxide measurements

Program at MAUNA LOA, HAWAII
Daily mean values for the year 1982

Provisional CO2 concentration values (in ppm above 300.0)
(continuous analyzer)

Day	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Day
1	40.32	41.38	42.66	43.05	44.37	43.38	42.67	40.45		37.35		39.51	1
2	40.07	41.49	42.70	43.17	44.04	43.52	42.85	40.89	39.16	37.66	38.21	39.79	2
3	40.50	41.92	42.50	43.29	43.94	43.72	42.81	40.72	38.96	37.33	38.43	39.75	3
4	40.55	42.18	42.66	43.30	43.92	43.45	42.65	40.89	39.09	37.33	38.87	39.74	4
5	40.63	42.17	42.87	43.12	43.66	43.18	42.61	40.66	38.56	37.13	39.01	40.17	5
6		42.04	42.74	43.06	43.39	43.57	42.68	40.40	38.86	37.07	39.08	40.12	6
7	40.53	42.03	42.21	43.28	43.63	43.61	42.63	40.43			38.88	39.83	7
8	40.59	41.89	42.32	43.75	43.82	43.73	42.53	40.54	38.70	37.97	38.85	40.62	8
9	40.69	41.69	42.75	43.68	44.56	44.07	42.27	40.42	38.52	38.24	38.59	40.76	9
10	40.67	41.79	42.74	43.70	44.62	43.67	42.06		38.43	37.87	39.12	40.93	10
11	40.63	41.05	42.96	43.49	44.49	43.62	41.85	40.04	38.49	37.87	39.11	40.51	11
12	40.65	40.88	42.89	43.67	44.57	43.41	41.78			37.92	39.13		12
13	40.74	41.10	42.27	42.88	44.58	43.37	42.01	40.19		37.91	39.39	40.47	13
14	40.95	41.10	42.08	43.48	44.55	43.73	41.84	39.99	38.01	37.87	39.51	40.36	14
15	40.58	41.35	42.28	43.52	44.47	43.90	42.13	40.21		37.83		40.40	15
16	40.87	41.52	42.54	43.62	44.52	44.02	41.89	40.21	38.21	37.58		40.93	16
17	40.75	41.64	42.81	43.47	44.60	43.91	41.61	39.97	38.58	37.55		40.69	17
18	40.94		43.03	43.52	44.43	43.62	41.95	40.18	38.83	37.90	39.21	40.82	18
19	40.83	41.22	43.24	43.36	44.42	43.65	41.85	40.28	37.93	37.84	38.02	40.78	19
20	40.89	41.26	43.31	43.50	44.25	43.20	41.78	40.34	37.46	37.79	38.26	40.87	20
21	40.89	41.28	43.49	43.78	44.30	42.97	41.35	39.99	36.98	38.14	39.47	41.13	21
22	40.90	41.36	43.04	44.26	44.31	42.62	41.43	39.35	36.53	38.02	40.13	40.95	22
23	41.04		43.39	44.34	44.55		41.93	38.59	36.84	38.09	40.39	40.73	23
24	41.21	41.74	43.07	44.31	44.49	43.03	41.80	38.15	36.96	38.00	40.11	41.24	24
25	41.62	42.29	43.21	44.43	44.91	42.95	41.71	37.99	37.16	39.46	40.69	41.64	25
26	41.56	42.30		44.31	44.53	43.07	41.69	38.11	37.17	39.95	40.35	41.63	26
27	41.87	42.09	42.85	44.37	44.45	43.12	41.35	38.12	37.32	39.47	39.62	41.33	27
28	41.79	42.42	42.88	44.11	44.36	42.85	40.59		36.79	39.41	39.52	41.02	28
29	41.51		42.85	44.39	44.49	43.04		38.58	36.93	39.16	39.48	40.99	29
30	41.34		42.37	44.62	44.35	42.79	40.87		37.15	38.53	39.53	41.20	30
31	41.55		42.71		43.86		41.26	38.48		38.57		41.23	31

Monthly values

Xm	40.92	41.66	42.78	43.70	44.30	43.41	41.95	39.78	37.90	38.09	39.27	40.67	Xm
N	30	26	30	30	31	29	30	27	25	30	26	30	N
s	0.45	0.44	0.36	0.50	0.35	0.39	0.57	0.96	0.86	0.73	0.69	0.57	s

Annual mean 341.20 ppmv
(based on 12 monthly means)

Xm - Average monthly concentration
N - Number of daily values accepted
s - Standard deviation (ppmv), based on daily means

NOAA/GMCC atmospheric carbon dioxide measurements

Program at MAUNA LOA, HAWAII
Daily mean values for the year 1983

Provisional CO2 concentration values (in ppm above 300.0)
(continuous analyzer)

Day	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Day
1	41.10	42.07	41.94		45.52	45.96	44.71	42.34	40.95	39.78	40.74	42.17	1
2		42.26	42.78	44.59	45.25	45.73	45.11	41.61	41.08	39.67	40.78	42.00	2
3	40.99	42.23	42.38		45.29	45.64	44.76	41.36	40.02	39.39	41.01	42.03	3
4	41.03	42.25	42.02	44.50	44.76	45.46	45.10			39.97	41.02	42.31	4
5	41.15	42.17	42.04	44.26		45.80	45.07	42.83	37.87	39.72	40.99	42.11	5
6	41.67	42.19	42.26	44.25	46.06	45.76	45.27	43.66	39.07	39.75	41.02		6
7	41.47	42.13	42.73	44.71	46.05	45.92	45.07		39.77	39.80	41.12		7
8	41.57	42.03	42.34	44.79	46.12	46.13	45.06		40.25	39.98	41.20	42.54	8
9	41.70		42.15	44.64	45.76	45.96	44.66	43.35	40.12	39.94	41.23		9
10	41.74	42.85	42.04		45.38	45.80		42.84	40.06	40.12	41.14		10
11	41.64	42.85	41.99		45.03	45.93	44.45	42.50	39.97	39.90	41.22		11
12	41.86	42.98	41.99	44.54		45.61		41.63	39.57	40.01	41.11	42.64	12
13	41.80	42.87	42.82	44.58	45.65	45.62	44.05	41.38	39.41	40.03	41.29	42.54	13
14	41.94	42.79	43.46	44.84	45.76	45.61	44.06	41.23		39.94	41.25	42.61	14
15	42.55	43.20	43.78	44.97	46.09	45.86	43.55	41.97	39.74	40.21	41.38	43.08	15
16	41.69	43.10	43.97	45.17	46.14		43.30	41.80	40.60	40.29	41.40	43.08	16
17	41.25	42.92	44.06	45.58	46.24			42.07	39.83	40.12	41.55	42.97	17
18	40.95	42.86	43.68		46.27	45.17	43.42	42.00	39.53		41.34	42.99	18
19	40.65	42.60	43.84	45.56	46.12	44.90	43.79	42.47	40.00	40.63	41.70	43.25	19
20	40.70	42.57	44.85	45.52	46.09		43.85		40.10	40.28	41.52	43.29	20
21	41.17	42.48	45.26	45.56	46.04		44.18	42.19	40.21	40.56	41.40	43.21	21
22	41.40	42.27	45.26	45.45	45.70		43.86		39.98	40.26	41.28	43.53	22
23	41.41	42.82	45.59	45.46	45.44	44.91	43.81	41.44	40.37	40.43	41.33	43.39	23
24	41.07	43.12	45.36	45.56	45.07	45.16	43.86	41.97	40.79	40.52	42.02	43.56	24
25	41.08	43.03	45.09	45.56	46.11	44.98	43.94	42.52	39.92	40.55		43.20	25
26	41.76	43.53		45.40	46.18	44.92	43.68			40.44	41.76	43.25	26
27	41.69	43.35		45.61	46.19	44.67	43.41	42.27	40.16	40.52	41.84	43.59	27
28	41.87	42.80	42.95	45.88	45.90	44.47	42.40	42.09	40.37	40.44	42.01	43.66	28
29	41.38		43.61	46.04	45.74	44.39	42.46	42.00	40.36	40.54	42.42	43.83	29
30	41.34		44.41	46.01	45.77	44.43	42.50	40.70	40.14	40.45	42.45	43.81	30
31	41.35		44.18		45.86		41.87	39.84		40.51		43.76	31

Monthly values

Xm	41.43	42.68	43.41	45.16	45.78	45.39	43.97	42.00	40.01	40.16	41.40	43.02	Xm
N	30	27	29	25	29	25	28	25	27	30	29	26	N
s	0.41	0.42	1.22	0.55	0.41	0.54	0.91	0.80	0.62	0.33	0.43	0.58	s

Annual mean 342.87 ppmv
(based on 12 monthly means)

Xm - Average monthly concentration
N - Number of daily values accepted
s - Standard deviation (ppmv), based on daily means

NOAA/GMCC atmospheric carbon dioxide measurements

Program at MAUNA LOA, HAWAII
Daily mean values for the year 1984

Provisional CO2 concentration values (in ppm above 300.0)
(continuous analyzer)

Day	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Day
1	43.51	44.53	44.78		47.50		46.00	44.51		41.45	41.72	42.99	1
2	43.61	44.37	44.89		47.75		46.19		42.43	41.67	42.00		2
3	43.58	44.51	45.06		47.59	46.71	46.31	44.48	42.42		41.89		3
4	43.93	44.66			47.68	46.96	45.91	43.62	41.75	41.78	42.12	43.62	4
5	43.85	44.68			47.54	47.32	46.11	43.86	40.69	41.51	42.02	43.44	5
6	44.06	44.67	45.01			47.37	46.00	44.14		41.41	42.41	43.26	6
7	44.02	44.81	45.19				46.01	44.26	39.35	41.11	42.74	43.86	7
8	43.91	44.64	45.20		47.11	46.60	45.82		39.79	41.03	42.37	44.25	8
9	43.86	44.80	44.84		46.56	47.15	45.82	44.36	40.48	40.98	42.76		9
10	43.68	44.75	44.83			47.31	45.78	44.68	42.35	41.00	42.59	44.76	10
11	43.77	44.54	44.76		46.87	47.22	45.86	44.23	42.18	40.98	43.11	44.40	11
12	43.52	44.65	44.70			47.10	45.97		41.73	41.61	42.89	44.42	12
13	43.36	44.64	44.95		47.68	47.22	45.75		41.23		42.59	44.14	13
14	43.33	44.57	44.98		47.90	46.97	45.88		40.68	41.28	42.59	44.01	14
15	43.84	44.58	45.29		47.95		45.89	42.23		41.19		44.05	15
16	43.95	44.50	45.58		48.11		45.19			41.71	43.01		16
17	43.62	44.49	45.31		47.96		45.67		41.18	40.77	42.96	44.56	17
18	44.00	44.12	45.16		47.78	46.63	45.37	43.97	41.52		43.08	44.63	18
19	43.89	44.35	45.41		47.61		45.31	43.72	42.00		43.17	44.94	19
20	43.87	44.64	45.40		47.30	47.29		43.02	41.87	41.96	43.34	44.47	20
21	43.70	45.09	45.50		47.23	47.08	45.02	42.31	41.44		42.90	44.72	21
22	43.97	45.09	45.46		47.48	46.54	44.82	42.12		42.11	43.00	44.66	22
23	44.23	45.10	45.49		47.24	46.39	44.61		41.07		43.17	44.66	23
24	44.11	44.95	45.58		46.70	46.49		41.96	40.94	41.58		44.83	24
25	44.09	44.50	45.62			46.36	44.62	41.85	41.26	41.82	43.80	44.48	25
26	44.07	44.27				46.34	44.33		41.13	41.81	43.94	44.77	26
27	44.53	44.71			47.28	46.36			41.03	41.58		44.67	27
28	44.51	44.45			47.07	46.20	44.66	42.28	40.92	41.85	44.17	44.72	28
29	44.32	44.64		47.03	47.23	46.46	44.62	41.62	40.71	41.86	43.60	44.62	29
30	44.28			47.20	47.44	46.12	44.49	41.44			42.56	44.60	30
31	44.52				47.52		44.21					45.31	31

Monthly values

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	
Xm	43.92	44.63	45.17	47.12	47.44	46.79	45.44	43.23	41.26	41.48	42.83	44.36	Xm
N	31	29	23	2	25	23	28	20	24	23	27	27	N
s	0.32	0.23	0.30	0.12	0.39	0.41	0.65	1.13	0.78	0.37	0.62	0.54	s

Annual mean 344.47 ppmv
(based on 12 monthly means)

Xm - Average monthly concentration
N - Number of daily values accepted
s - Standard deviation (ppmv), based on daily means

NOAA/GMCC atmospheric carbon dioxide measurements

Program at MAUNA LOA, HAWAII
Daily mean values for the year 1985

Provisional CO2 concentration values (in ppm above 300.0)
(continuous analyzer)

Day	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Day
1	44.71		46.71	48.18	48.50	49.34	47.14	45.97	43.31	42.15	43.21		1
2	44.60	46.15	46.69	48.24	48.84		46.94	45.62	43.54	41.94	43.34	44.76	2
3	44.91	45.98	46.78	48.38	48.82	48.86	46.91	44.70	43.50	42.19	43.43	44.95	3
4	45.25			48.35	48.98	48.78	46.66	44.67			43.50	45.09	4
5	45.08	45.63	46.71	48.15	49.23	48.81	47.10		44.14	42.28	43.54	45.37	5
6	44.92	45.72	47.00	48.08	48.83	48.58	46.94		44.56	42.52	43.47	45.19	6
7	45.03	45.62	47.25	48.17	48.67	48.51	46.87	45.93	44.35	42.55		45.15	7
8	45.09	45.59	47.46	47.94	48.51	48.44		44.95	44.15	42.61	43.40	45.55	8
9		45.52			49.28	48.52	46.89	44.48	44.22	42.48	43.60	45.60	9
10	44.96	45.35	47.27	47.68	48.78	48.30	46.58	44.30		42.58	43.95		10
11	45.02	44.94	47.69		48.85	48.58	46.76		42.45	42.57	44.10	45.00	11
12	45.39	45.41	47.10	47.44	48.98	48.39	46.80	43.75	42.91		44.06	45.25	12
13	45.30	45.64	47.47	47.28	48.76	48.33	46.81	43.44	43.51	42.41	43.90	45.28	13
14	44.94	45.52	47.22	46.78	48.74	47.18	46.87	44.15	43.63	42.66	44.22	47.04	14
15		45.72	47.22	47.31	48.79	47.33	46.87		43.46	42.57	44.21	47.04	15
16	45.21	45.87	47.53	47.40	48.98	47.63	46.72	45.16	43.58	42.67	44.28	45.78	16
17	44.94	46.19	47.82	47.77	48.89	48.13	46.65	44.74	43.49	43.04	44.19	45.87	17
18	44.88	46.17	47.65	48.10	49.47	48.22	46.50	44.85	42.77	43.19	44.23	45.49	18
19	45.22	45.97		48.11	49.31	48.34	46.38	44.64	42.06	43.03	44.41	45.17	19
20	45.36	46.15	47.59	48.51	49.19	48.19	46.51	44.52	42.36	43.25	44.55	45.05	20
21	45.07	45.85	47.34	48.25	48.82	48.01	46.43	44.37	42.27		44.72	45.06	21
22	44.96			48.32	48.32	47.78	47.30	44.04	42.07	43.24	44.27	45.10	22
23	45.03	46.28	46.92		47.98	47.75		43.95	42.25	43.35	44.55		23
24	44.73	46.74	47.20		47.59	47.92		44.87	41.98			45.13	24
25	44.51	46.38	47.39	47.95	47.63	47.76	46.64	45.04		43.28	44.21	46.65	25
26	44.72	46.41	47.38	48.12	47.68	47.72	46.10	45.35	41.87	43.27	44.37	46.45	26
27	45.14	46.08	48.03	48.48	47.87	47.91	46.07	44.99	42.11	43.21	45.52	45.83	27
28	45.12	46.15	48.70	48.65	48.36	47.41	45.96	44.82	42.32	43.52	45.81	45.46	28
29	45.03		48.43		48.85	47.30	45.54	44.52	42.00	43.53	45.26	45.73	29
30	45.08		48.72	48.67	49.32	47.19	44.94	43.49	42.18	43.81	44.94	46.62	30
31	45.79		48.36		49.52			42.51		43.77		46.81	31

Monthly values

Xm	45.03	45.88	47.47	48.01	48.72	48.11	46.59	44.59	43.00	42.88	44.19	45.62	Xm
N	29	25	27	25	31	29	27	27	27	27	28	28	N
s	0.26	0.40	0.57	0.47	0.52	0.55	0.51	0.76	0.86	0.51	0.65	0.68	s

Annual mean 345.84 ppmv
(based on 12 monthly means)

Xm - Average monthly concentration
N - Number of daily values accepted
s - Standard deviation (ppmv), based on daily means

NOAA/GMCC atmospheric carbon dioxide measurements

Program at MAUNA LOA, HAWAII
Daily mean values for the year 1986

Provisional CO2 concentration values (in ppm above 300.0)
(continuous analyzer)

Day	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Day
1	46.86	46.32	47.17	48.76	50.31	50.44	48.99	45.85	44.55	44.92	45.38	46.50	1
2	46.27	46.34	46.98	48.76	50.87	50.31	49.28	46.38		44.18	45.09	46.43	2
3	46.53	46.44	46.93	48.69	51.01	50.34	48.83	46.37	43.14	44.00	45.30	46.46	3
4		46.48	46.60	48.97	50.70	50.42	48.81	46.69	44.32	44.03	45.34	46.68	4
5		46.69	46.56	48.92	50.28	50.25	48.58	46.44	44.30	43.93	45.14	46.74	5
6	46.52	46.44	46.70	49.49	50.25	49.76	48.59	46.40	44.13	43.95	45.38	46.92	6
7	46.96	46.80	46.66	49.47	50.05	49.78	48.62	46.45	44.36	43.88	45.30	47.02	7
8	46.39	46.85	46.74	49.71	50.52	50.11	48.59	46.43	45.11	43.67	45.31	46.72	8
9	46.32	46.94	47.01	49.79	50.44	49.93	48.44	46.28	45.38	43.96	45.50	46.57	9
10	46.25	47.38	46.71	49.92			47.56	46.06	45.11	44.42	45.61	46.74	10
11	46.26	47.03	47.55	49.48	50.42	49.89	46.54	45.73	45.17	44.26	45.66	46.83	11
12	46.13	46.90	47.88	49.62	50.55	49.70	46.45	46.23	44.94	44.22	45.88	46.88	12
13	46.17		47.37	49.23	50.08	49.79	46.77	46.32	45.35	44.40	45.82	46.76	13
14	46.46	46.98	47.39	49.28			47.24	45.92	45.56	44.38	45.77	46.62	14
15	46.43	46.62	47.02	49.28	50.40		47.62	45.91	45.51	44.12	45.64	46.69	15
16		47.43	47.19	49.55	50.21		47.67	45.66	45.65	43.98	45.47	46.61	16
17		47.72	47.66	49.44	50.42		47.62	45.28	44.41	44.16	45.51	46.75	17
18	46.28	47.72	47.50	49.37	50.45	49.63	47.34	44.77	45.71	44.21	45.32	46.80	18
19	46.34	47.26	48.69	49.38	50.49	49.68	47.68			44.08	45.66	47.01	19
20	46.46	47.12	49.08	49.23	50.51	49.76	47.83	45.18	45.43	44.46	46.08	46.99	20
21		46.77	49.49	49.34	50.45	49.77	47.58	46.32	44.94	44.09	45.63	47.04	21
22	46.99	46.70	49.54	49.75	50.40	49.45	47.48	45.92	43.97	44.51	45.55	46.94	22
23	46.97	46.48	49.56	49.94	50.53	49.25	47.56	45.59	44.15	44.40	45.57		23
24	46.99		49.19	50.06	50.17	49.49	47.57	45.92	44.27	43.84	45.62	47.56	24
25	46.42	47.65	49.27	50.51	49.98	49.40	47.73	46.06	44.35	44.08	45.67	47.70	25
26	46.43	47.57	49.06	50.47	49.99	49.29	47.82	45.86	44.41	44.84	46.32	47.10	26
27	46.39	47.51	48.98	50.32	49.79	49.12	47.90	46.04	44.79	44.69	46.31	47.22	27
28	46.41	47.28	48.77	50.46	50.05	49.07	47.65	45.25	45.06	45.14	46.27	47.02	28
29	46.32		48.85	50.34	50.15		47.09	45.12	45.11		46.34	47.03	29
30	46.37		48.75	50.39	50.71	48.86	47.05	44.93	44.88	45.16	46.56	46.92	30
31	46.41		48.77		50.98		47.09	45.25		45.23		46.93	31

Monthly values

Xm	46.47	46.98	47.92	49.60	50.38	49.73	47.79	45.89	44.79	44.31	45.67	46.87	Xm
N	26	26	31	30	29	24	31	30	28	30	30	30	N
s	0.26	0.45	1.06	0.54	0.30	0.44	0.72	0.51	0.61	0.41	0.38	0.29	s

Annual mean 347.20 ppmv
(based on 12 monthly means)

Xm - Average monthly concentration
N - Number of daily values accepted
s - Standard deviation (ppmv), based on daily means

2. TRENDS AND SEASONAL CYCLES OF ATMOSPHERIC CO_2 OVER ALERT, SABLE ISLAND, AND CAPE ST. JAMES, AS ANALYZED BY FORWARD STEPWISE MULTIPLE REGRESSION TECHNIQUE

N.B.A. Trivett, K. Higuchi,
and S. Symington

Atmospheric Environment Service
4905 Dufferin Street
Downsview, Ontario M3H 5T4
Canada

2.1 Introduction

The Canadian CO_2 program was initiated almost 20 years ago in 1969 at Ocean Station P as a joint effort between Scripps Institute of Oceanography (SIO) at La Jolla, California and the Institute for Ocean Sciences (IOS) at Sidney, British Columbia (see Wong et al., 1984). As a result of the Stockholm Conference in 1974, the Atmospheric Environment Service (AES) established two additional stations in 1975: one in the north at Alert in the Canadian Arctic, and one in the east on Sable Island off the coast of Nova Scotia. In 1982 the weather ship at Ocean Station P was taken out of service, with subsequent termination of atmospheric CO_2 measurements there. However, Cape St. James on Queen Charlotte Islands off Canada's west coast was established as a replacement for Ocean Station P in the CO_2 program. In 1984 responsibility for the Canadian CO_2 program was divided between AES (land based measurements) and IOS (ocean based measurements). IOS continued to analyze the flasks collected by AES and was responsible for maintaining the calibration scale used in the Canadian program. These duties were taken over by AES in 1988.

Although there has been an increase in the number of stations with continuous monitoring programs for atmospheric concentration of CO_2, grab flask sampling networks are still the most convenient and economical method for obtaining information about the chemical composition of the atmosphere on a global scale. While the type of flask and sampling methodology employed vary somewhat from program to program, each grab sample is essentially a "snapshot" of the chemical state of the atmosphere at a point in time and space. As a result, time series of atmospheric CO_2 measurements obtained from flask networks are very variable, and a few missing samples can leave large time gaps in the data set. This makes quality control of the flask data very difficult.

There are several problems associated with quality controlling grab flask sample data. In addition to the normal data loss due to broken flasks or obviously contaminated samples, data may also be rejected as outliers because they might have resulted from field sampling problems and/or errors in laboratory analysis procedures. Furthermore, data may be rejected as unrepresentative of the "baseline" or background conditions (or some other criteria appropriate to the specific network). Application of these quality control criteria usually results in a data set which is relatively "clean", in

a sense that it represents fairly accurately the background condition at the observational site, but significantly reduced in size. When an average sampling frequency is about one per week (as is the case with the Canadian CO_2 flask sampling program), any reduction which creates temporal gaps larger than those already existing in the original data set poses a serious problem in obtaining a reasonable and unbiased facsimile of the "real" seasonal cycle and trend in the atmospheric CO_2. A solution to this problem is to employ some sort of interpolation scheme or curve fitting procedure to filter out the high frequency sampling noise and still follow the annual trends and strong seasonal variation in individual years. Monthly and annual CO_2 values are then usually derived from the interpolated data.

Many analysis schemes that are applied to CO_2 data from the "continuous" measurement system in baseline programs are not easily applied to discrete flask data values. Time series analysis techniques such as harmonic analysis or digital filtering usually require regularly spaced data in the time domain. Even with continuous data, some interpolation (usually linear) is required to fill in the inevitable missing data gaps. This is not a serious problem if the time gap is less than the periodicity of meteorological processes or other factors that influence CO_2 values at a sampling site. Now, it has been shown that the variability of atmospheric CO_2 concentration values on a 4 to 7 day period is highly correlated with atmospheric synoptic events (Halter and Peterson, 1981; Halter and Harris, 1983; Higuchi and Daggupaty, 1985). For a weekly flask program one missing sample means interpolating over a two week period. A simple linear interpolation in this case might not be adequate or appropriate. Cubic splines have been used with some success, but a cubic spline still has problems coping with large gaps in the data set (see Wong et al. (1984) and references contained therein).

The interpolation or curve fitting scheme used in each network could have a significant effect on the global interpretation of atmospheric CO_2 trends and seasonal cycles, and their interannual variability. In this paper, we evaluate the long term trend and seasonal cycle of the atmospheric CO_2 concentration measurements obtained at three Canadian stations (Alert, Sable Island, and Cape St. James) using a forward selection stepwise multiple linear regression, and compare some of the results with those obtained by Wong et al. (1984) who used a cubic spline technique.

2.2 Data Selection

Until the middle of 1986, the CO_2 standards maintained at IOS were CO_2-in-N_2, which were calibrated against standards maintained by SIO. The basic scale used by IOS was the I index of Keeling et al. (1976), and adjustments to CO_2-in-air (1985 scale) were performed using procedures outlined by Wong et al. (1984) and SIO. In August 1986, IOS switched to CO_2-in-air on the 1985 WMO scale for its standards. All the data reported in this paper have been converted to the 1985 scale.

Both 5 litre and 2 litre flasks are used in the Canadian CO_2 monitoring network. The 5 litre flasks, which were used at Ocean Station P, are now being used at Cape St. James, while the 2 litre flasks are used at Alert and

Sable Island. At each station, two flask samples are taken on a quasi-periodic basis (on average, one per week). During the chemical analysis procedure, two aliquots are extracted from each of the two flasks. Concentration values obtained from these two aliquots are then averaged to give a mean concentration for each flask. The difference between these two aliquots is termed the "within flask" difference, while the difference between the means of the flask sample pair is denoted the "between flask" difference. The exact criteria used for data classification are given in Table 1. Flask pairs that are classified as A or B are retained for time series analysis to determine trends and seasonal cycles.

2.3 Data Analysis

A time series of atmospheric CO_2 can be thought of as a synthesis of four basic components: (1) the long term trend, (2) the annual or seasonal cycle, (3) the interannual variation, and (4) "noise", or intrannual high frequency fluctuations correlated with weather events. Within the context of our objective given in Introduction, the most appropriate interpolation scheme would simulate the first 3 components and filter out the 4th component. The long term trend may be fitted by a polynomial equation of suitable degree, while the seasonal cycle may be fitted to a series of harmonics. In this paper we present results obtained from an application of a least squares stepwise multiple regression technique to atmospheric CO_2 time series obtained from the three Canadian monitoring stations.

A series of trignometric sine and cosine terms with periods ranging from 1 month to 12 years were fitted to the CO_2 data from each station using a multiple regression with important parameters selected in a forward step-wise manner, as described in Draper and Smith (1981). The independent variables used were time, t, t^2, t^3, and a series of harmonic terms $\sin(\Omega t)$ and $\cos(\Omega t)$, where $\Omega = \frac{2\pi}{p}$. The values of p took on integer values from 1 to 12, representing cycles with periods of 1 year to 12 years; p also ranged from 1/12 to 1 in increments of 1/12, representing cycles of 1 month to 12 months. While polynomial terms and cyclic terms of periods ranging from 1 month to 12 years are available for selection, step-wise multiple regression technique retains only those terms in the equation which contribute significantly to reducing the unexplained variance. This approach was chosen so as not to prejudice the form of the equation by preselecting which polynomial and harmonic terms were to be fitted to the time series.

Data values that were more than 3 standard errors of fit away from the fitted curve were flagged and identified as outliers. These rejected data were not used for any further statistical analysis. This curve fitting procedure was continued until no further outliers were identified. It took 3 iterations to identify all the outliers in the Alert data, while it took 2 and 4 iterations for Cape St. James and Sable Island, respectively. Data points lying outside the 2 standard errors of fit away from the last curve to be fitted were flagged as non-baseline and requiring further study. As with any curve fitting procedure, care must be taken to avoid overfitting with too many

Table 1.--Data classification procedure

The sampling procedure is to obtain two flasks per daily sample. From each flask, two aliquots are analyzed.

Primary classification segregates each flask into one of 5 categories:

ALIQUOT PER FLASK	DESCRIPTION	CATEGORY
0	missing/missing	Missing
1	missing/other > CO2 limit*	Lab analysis error
1	missing/other < CO2 limit	Botched one analysis
2	aliquot diff > limit **	Noisy instrument
2	aliquot diff < limit	Acceptable

Secondary Classification uses the daily flask pair categories and further classifies the daily flask average.

A: -Best Data
-2 flasks, 2 aliquots each
-both flasks have A classification
-between flask averages is less than the limit set***

B: -2 flasks
-one aliquot missing but other is less than CO2 limit
-between flask averages is less than the limit set

S: -Probable sampling error
-one flask may be broken (missing)
-within flask diff is less than 0.3 ppm
-between flask diff is greater than 0.6 ppm

L: -Probable laboratory analysis error
-one aliquot missing and other is analyzed greater than CO2 limit or average of flask is greater than CO2 limit

* CO2 limit was an arbitrary value (380 ppm) set to determine contamination
** Within flask difference limit was set to 0.3 ppm
*** Between flask difference limit was set to 0.6 ppm

iterations. Table 2 gives the final regression equations generated by this technique.

It should be pointed out that the technique of forward stepwise multiple linear regression used here to analyze flask data is not without problems. As with any least squares technique, it is sensitive to outliers. It is a common procedure to use the standard error of fit to identify and eliminate outliers in an iterative way, resulting in a tendency to make the data fit the curve, not the curve to the data. Furthermore, by initially including so many harmonic terms in the regression equation as dependent variables available for selection, some data sets can produce quite irregular curves; this was the case with data from Cape St. James after the first iteration. One principle advantage with the multiple regression approach, however, is that it gives trend, seasonal cycle, and interannual variation in one single "operation"; and outliers and non-baseline data, as defined, are easily identified using the standard error of fitted curve.

2.4 Discussion

2.4.1 Seasonal cycle and amplitude change

Seasonal cycles for the stations are given in Figs. 1 to 3, along with data as classified by the multiple regression technique. As might be expected, the curve generated for Alert (with its isolated location and minimal local sources) exhibits the best behaviour and has the best statistics (Table 3). With a few exceptions, the fitted curve catches the peaks and troughs in the seasonal cycle. The deep troughs of 1983 and 1987 are missed by about 1 ppm. The high CO_2 concentration values observed during the winter of 1982/83 are identified by the analysis technique as non-baseline; these are being investigated to see if they are the result of unusual episodes of long range transport from Europe and the Soviet Union.

The CO_2 data from Sable Island, which is affected by air masses from the heavily populated and industrialized eastern United States, and from the relatively clean Atlantic Ocean, are much more variable than the data from Alert. Fifteen data points were rejected as bad data, and 32 classified as non-baseline (Table 3). The data from Sable Island are being stratified by wind directions, and classified by potential source regions using 5-day back trajectories; the results of this analysis are not yet available.

Cape St. James is the "acid test" of the multiple regression technique. However, with such large gaps in the data set, it is rather difficult to verify the equation. In general, the fit looks good, except for the peaks in 1982 and 1985; the results are comparable to the seasonal cycle obtained for Ocean Station P by Wong et al. (1984), who used a cubic spline technique.

Changes in the amplitude of the seasonal cycle as determined from the regression equations in Table 2 are given in Table 4 for Alert, Sable Island, and Cape St. James. Because of the nature of the flask data (with relatively large temporal gaps between samplings) and the limitations of the curve

Table 2.--Multiple regression equations, describing seasonal cycles and secular trends for Alert, Sable Island, and Cape St. James. Variable t is given as the time of year expressed as a fraction of the year minus 1900 (e.g., 84.5 for the middle of 1984), and $T = 2\pi t$.

---

(1) Alert: $CO2(ppm) = 219.83 + 1.4911*t + 6.1800*\sin(T) - 2.2282*\sin(2T) + 1.6659*\cos(T) + 1.2985*\cos(2T) - 0.8536*\cos(3T) + 0.6362*\sin(T/8) + 0.4777*\sin(3T)$

(2) Sable: $CO2(ppm) = 219.35 + 1.4921*t + 5.2468*\sin(T) - 3.3389*\cos(T) - 2.2375*\sin(2T) - 0.8400*\cos(3T) + 0.6964*\cos(2T) + 0.6140*\sin(4T) + 0.5985*\sin(T/8)$

(3) Cape: $CO2(ppm) = 222.95 + 1.4443*t + 6.1283*\sin(T) - 2.5371*\sin(2T) + 1.7978*\cos(T) - 0.6295*\cos(3T) + 0.4741*\cos(2T) + 0.3868*\sin(12T/5) + 0.3748*\sin(T/3) + 0.2753*\sin(4T)$

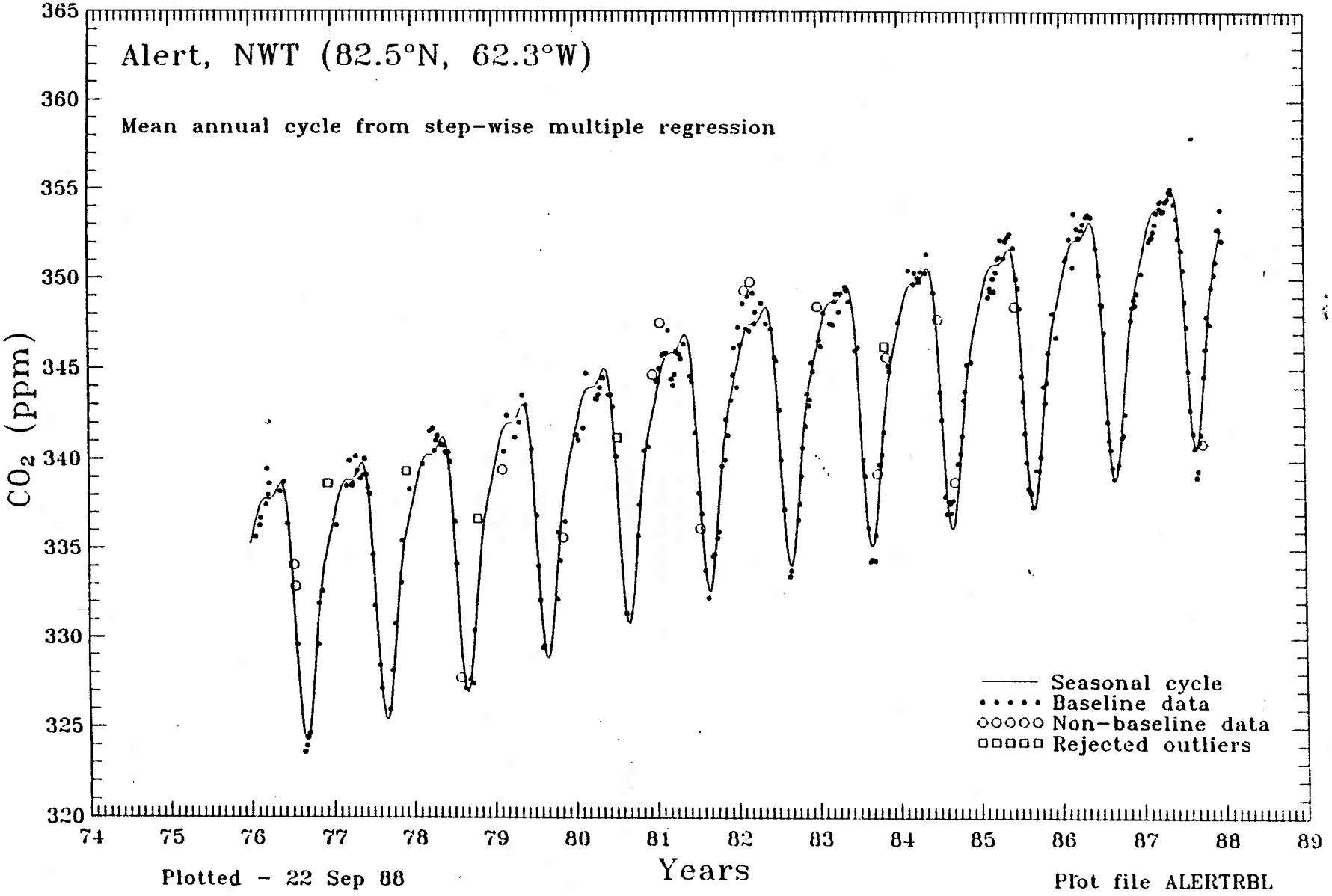

Figure 1. The seasonal cycle of atmospheric CO_2 concentration (WMO 1985 scale) at Alert, as determined by a forward stepwise multiple regression (FSMR) equation (see Table 2) derived by a fit to results obtained from analysis of 2 litre flask samples.

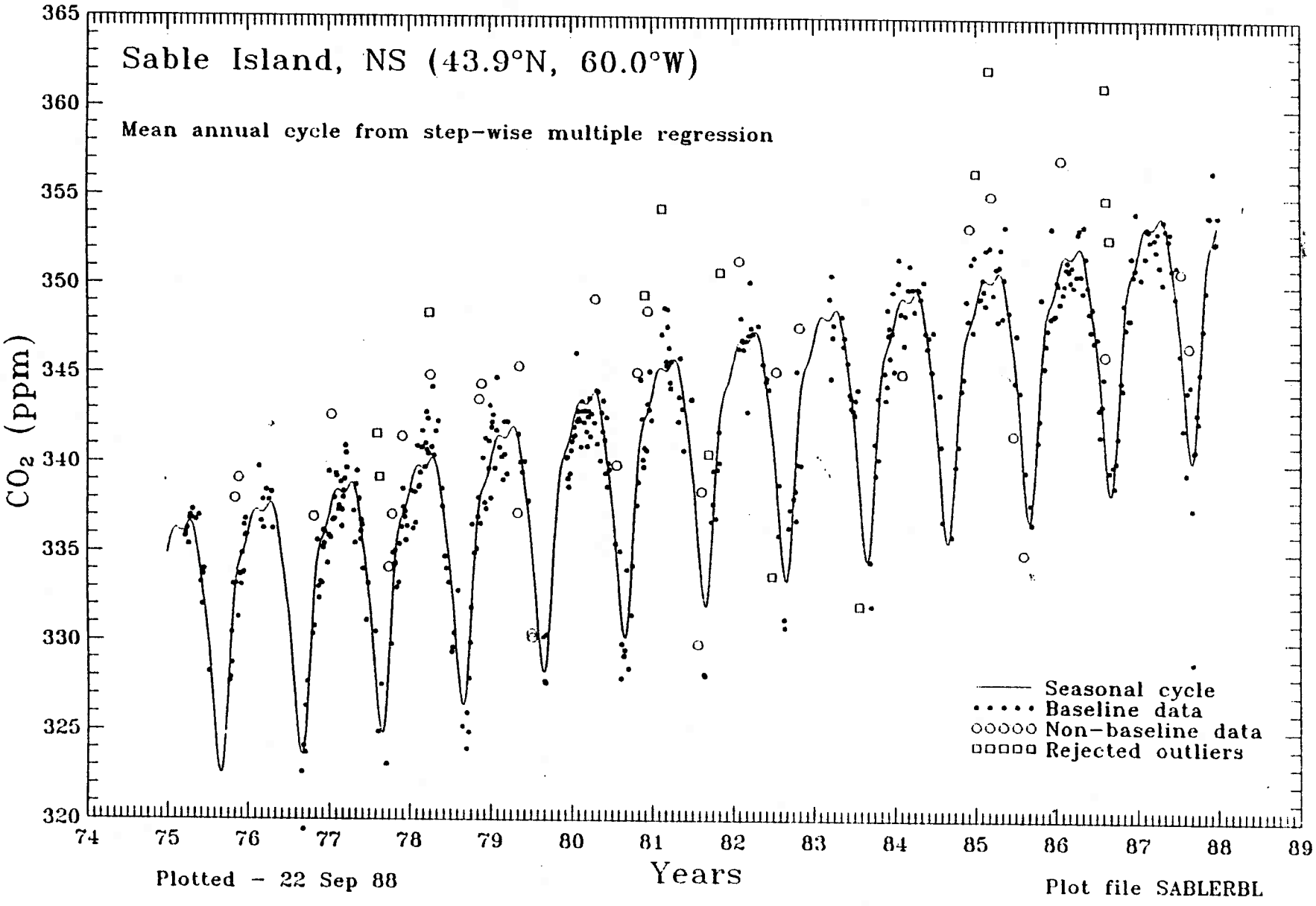

Figure 2. Same as Figure 1, except for Sable Island.

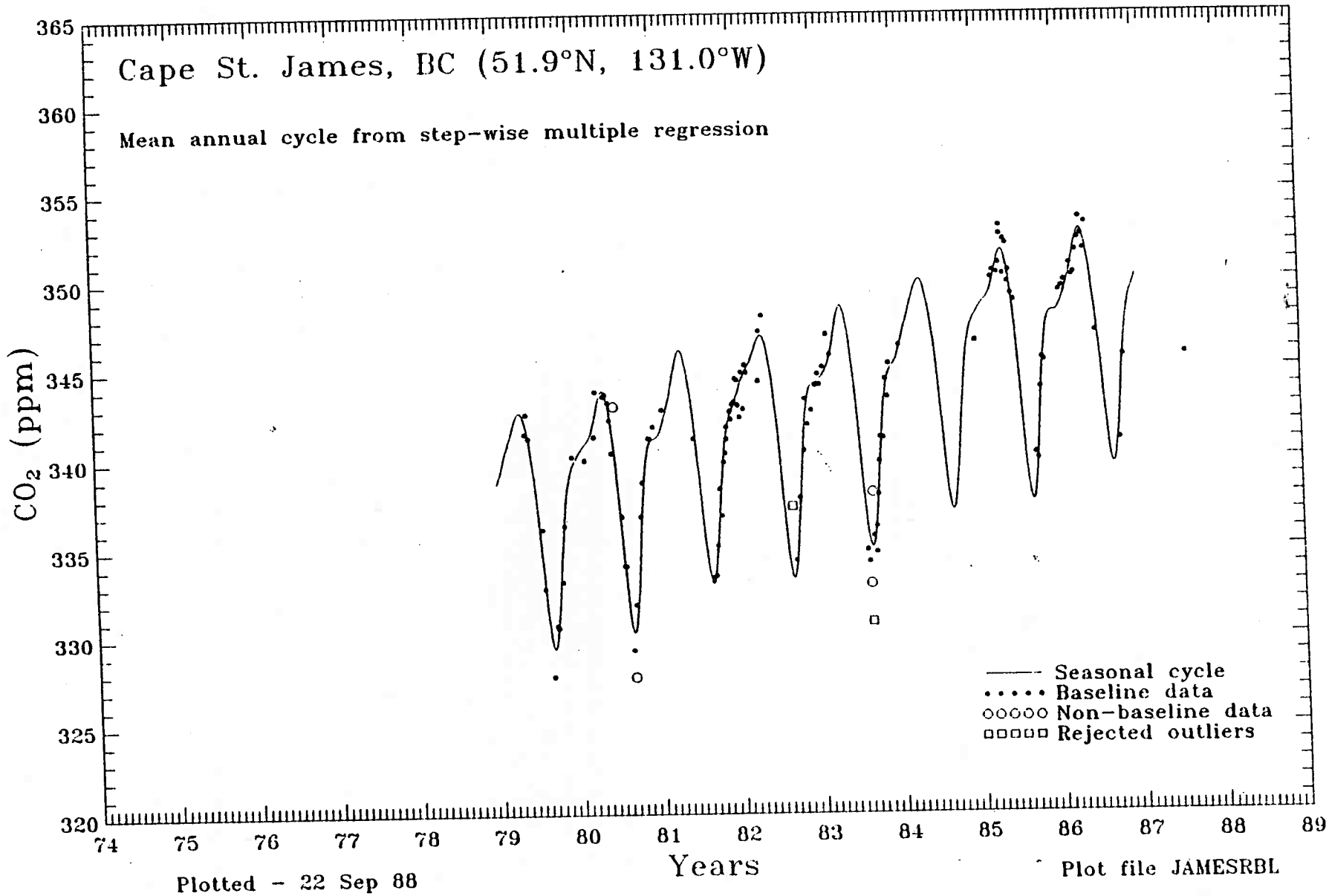

Figure 3. Same as Figure 1, except for Cape St. James and FSMR fitted to results obtained from analysis of 5 litre flask samples.

Table 3.--Correlation coefficients and standard deviations for the fitted curves are listed below, along with the number of samples classified in each group.

Station	Correlation Coefficient	Standard Deviation	Number of Samples Classified as Baseline	Nonbaseline	Rejected
Alert	0.990	0.979	321	17	5
Sable	0.952	2.136	472	32	15
Cape	0.984	1.118	115	4	6

Table 4.--Changes in the amplitude of the seasonal cycle as determined from the equations listed in Table 2.

Year	Alert	Sable Island	Cape St. James
1975		14.14	
1976	14.48	14.14	
1977	14.39	14.05	
1978	14.28	13.91	
1979	14.21	13.81	13.30
1980	14.21	13.81	13.60
1981	14.29	13.90	13.05
1982	14.40	14.04	13.58
1983	14.47	14.14	13.50
1984	14.47	14.14	12.87
1985	14.40	14.05	13.98
1986	14.28	13.91	13.07
1987	14.21	13.81	

fitting procedure used, values of the seasonal amplitude given in Table 4 should be viewed with a great deal of caution in terms of their reflection of reality. Consequently, quantitative changes in the amplitude from year to year should be viewed as very tentative; no statistical significance test has been performed. However, it is interesting to note that all three stations experienced a large rise in their seasonal amplitudes in 1982/83, and a general decrease thereafter. Given the fact that these three stations are widely separated geographically, spatial homogeneity in the behaviour of this property does give a sense of qualitative reality to the general trend in the change in seasonal amplitude.

2.4.2 Long-term increases in atmospheric CO_2

Table 5 shows the mean annual concentrations for each of the stations. The values listed under spline technique were taken from Wong et al. (1984). The annual averages under forward stepwise multiple regression (FSMR) technique were calculated from the monthly averages obtained linearly from daily estimates using the equations listed in Table 2. Secular trends obtained by polynomial curve fitting in a least squares sense to the seasonally adjusted concentration values are shown in Figs. 4 to 6 for Alert, Sable Island, and Cape St. James, respectively.

The mean annual rate of increase obtained from the slope of a least squares regression line through the annual averages is 1.49 ppm/year for Alert, 1.48 ppm/year for Sable Island, and 1.43 ppm/year for Cape St. James. These agree well with the average growth rate reported by Gammon (1985) from the continuous infrared analyzer measurements at Point Barrow, Mauna Loa, Samoa, and South Pole stations for the period 1976-1982. The mean annual growth rate for Mould Bay (the nearest NOAA/GMCC to Alert) reported by Komhyr et al. (1985) is 0.97 ppm/year. However, this is based on only 3 years of record. In the same paper, the reported growth rate for Point Barrow, Alaska (averaged over 12 years) is 1.38 ppm/year. This is close to the values we obtained for Alert. (Given the shortness of the record and large gaps in it, the rate of increase for Cape St. James must be viewed as tentative.)

The mean annual increases obtained by fitting a linear regression to the mean annual data obtained by cubic spline interpolation (Wong et al., 1984) are comparable but higher at 1.56 ppm/year for Alert, 1.57 ppm/year for Sable Island, and 1.44 ppm/year for Ocean Ship P. The data for Ocean Station P seem to show more rapid increase in atmospheric CO_2 between 1974 to 1976 than later in the time series. The mean annual rate of increase is 1.54 ppm/year for 1976 to 1980.

Table 5.--Mean annual concentrations of CO_2 at the Canadian BAPMoN baseline stations. The averages listed under Spline are from Wong et al. (1984), and are in the WMO 1983 scale; Ocean Station P has been substituted for Cape St. James*. The mean annual increase is given as the slope of a least squares regression line through the annual averages.

	Year	Spline	FSMR
(1) Alert:	1976	334.10	333.68
	1977	335.19	334.81
	1978	336.92	336.30
	1979	337.99	338.13
	1980	340.30	340.10
	1981	341.71	341.92
	1982		343.41
	1983		344.57
	1984		345.59
	1985		346.74
	1986		348.23
	1987		350.06
Slope:		1.56	1.49
(2) Sable:	1975	332.19	332.23
	1976	331.17	333.28
	1977	333.87	334.45
	1978	335.87	335.94
	1979	336.52	337.75
	1980	338.73	339.69
	1981	340.90	341.50
	1982		342.99
	1983		344.17
	1984		345.21
	1985		346.39
	1986		347.88
	1987		349.69
Slope:		1.57	1.48
(3) Cape:	1975	330.10*	
	1976	330.87*	
	1977	332.55*	
	1978	334.17*	
	1979	335.14*	337.79
	1980	337.26*	338.96
	1981		340.89
	1982		342.15
	1983		343.24
	1984		345.27
	1985		346.45
	1986		347.57
Slope:		1.44	1.43

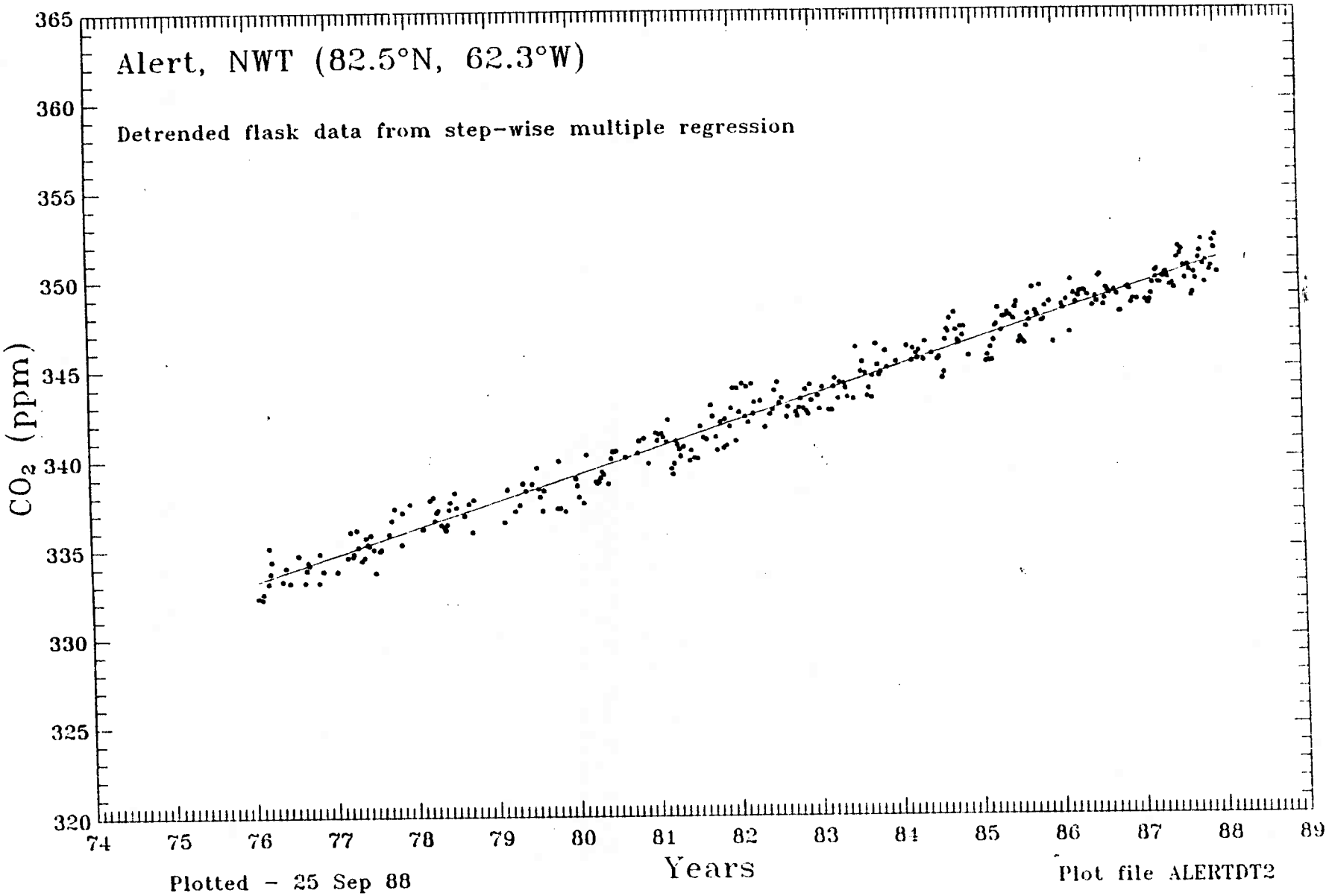

Figure 4. Secular trend of atmospheric CO_2 at Alert obtained by fitting a polynomial curve to seasonally adjusted CO_2 concentration values.

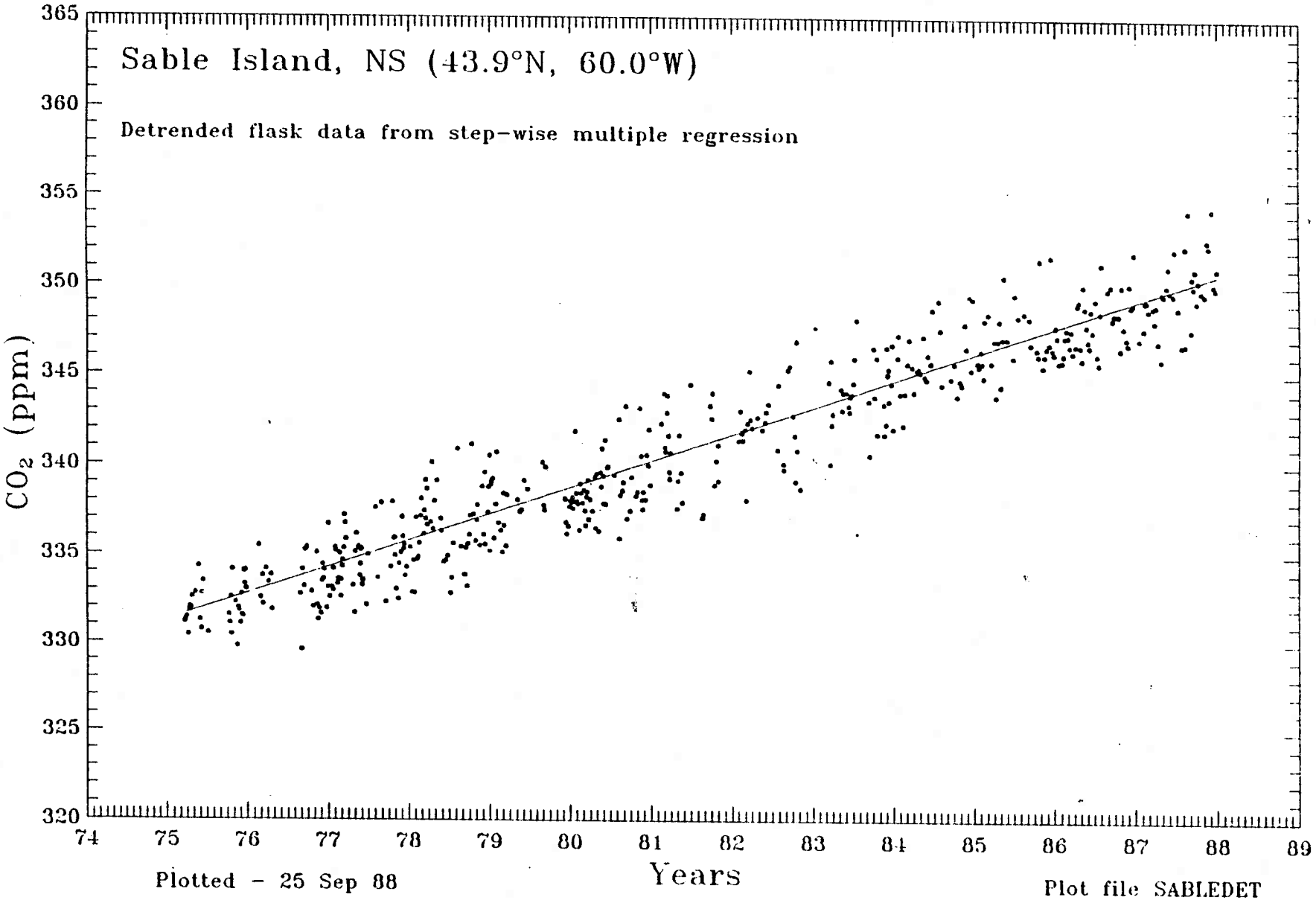

Figure 5. Same as Figure 4, except for Sable Island.

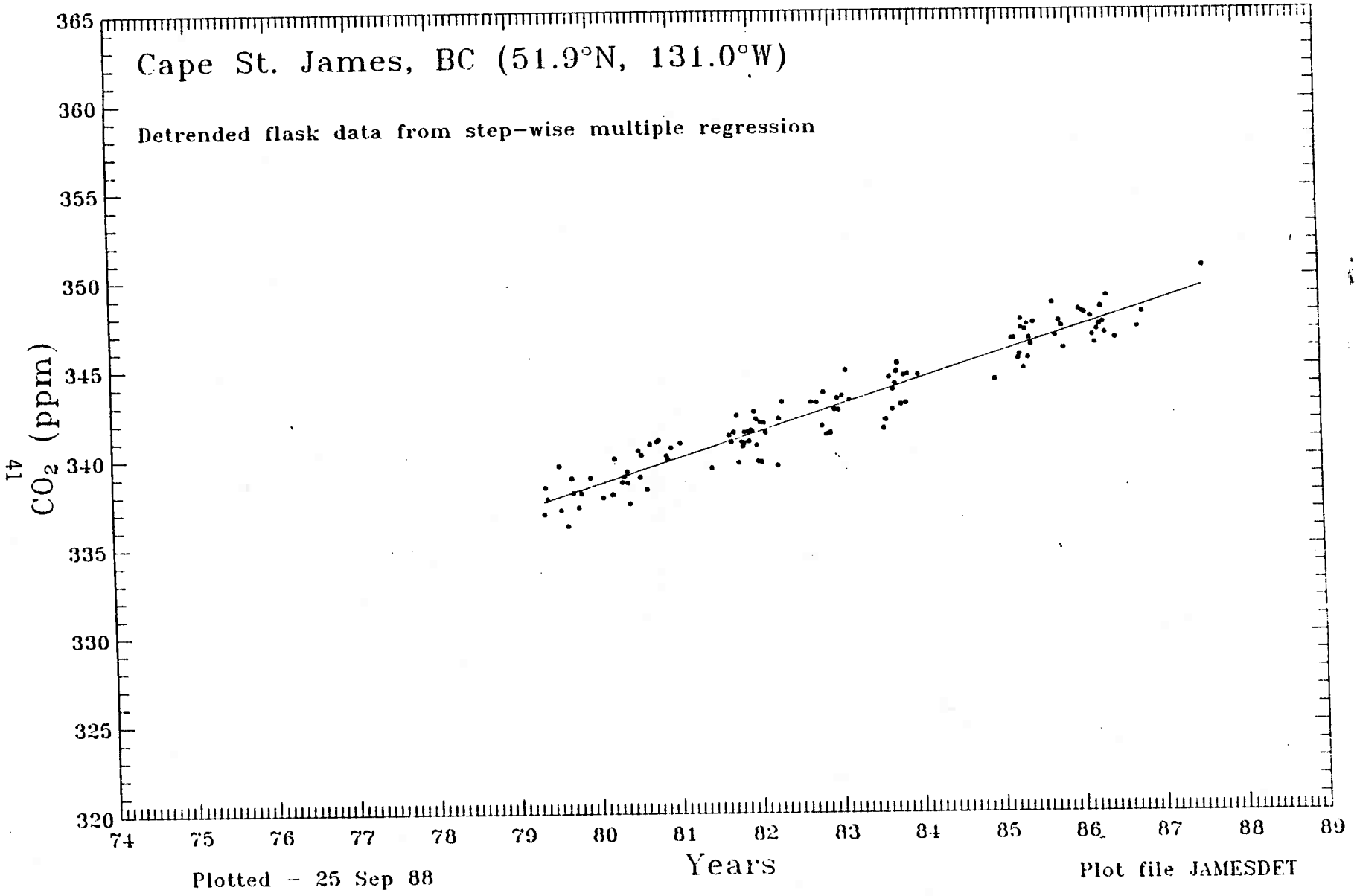

Figure 6. Same as Figure 4, except for Cape St. James.

2.5 References

Draper, N.R., and H. Smith, 1981. Applied Regression Analysis. 2nd ed., John Wiley & Sons, New York.

Gammon, R.H., E.T. Sundquist, and P.J. Fraser, 1985. History of carbon dioxide in the atmosphere. In: Atmospheric Carbon Dioxide and the Global Carbon Cycle, United States Department of Energy, Office of Energy, Research, DOE/ER-0239.

Halter, B.C., and J.T. Peterson, 1981. On the variability of atmospheric carbon dioxide concentration at Barrow, Alaska during summer. Atmos. Environ., 15, 1391-1399.

Halter, B.C., and J.M. Harris, 1983. On the variability of atmospheric carbon dioxide concentration at Barrow, Alaska during winter. J. Geophys. Res., 88, 6858-6864.

Higuchi, K., and S.M. Daggupaty, 1985. On variability of atmospheric CO_2 at station Alert. Atmos. Environ., 19, 2039-2044.

Keeling, C.D., R.B. Bacastow, A.E. Bainbridge, C.A. Ekdahl, Jr., P.R. Guenther, L.S. Waterman, and J.F.S. Chin,. 1976. Atmospheric carbon dioxide variations at Mauna Loa Observatory, Hawaii. Tellus, 28, 538-551.

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

Wong, C.S., Y.H. Chan, J.S. Page, R.D. Bellegay, and K.G. Pettit, 1984. Trends of atmospheric CO_2 over Canadian WMO background stations at Ocean Weather Station P, Sable Island, and Alert. J. Geophys. Res., 89, 9527-9539.

3. CO_2 MEASUREMENTS BY CSIRO (AUSTRALIA)

I.G. Enting

CSIRO, Division of Atmospheric Research
Private Bag No.1, Mordialloc
Victoria 3195, Australia

3.1 Introduction: Programs

A number of programs of measuring atmospheric CO_2 are operated by CSIRO (Australia) in the Division of Atmospheric Research (formerly Division of Atmospheric Physics).

These are:

i. An aircraft program involving the collection of flask samples which are returned to Aspendale for analysis. Regular samples have been obtained over southeastern Australia on a continuing basis since 1972 and over the Great Australian Bight since 1974. Samples have been collected over other regions for shorter periods.

ii. In situ CO_2 analysis has been undertaken at Cape Grim, Tasmania (40°41′S, 144°41′E) on a continuous basis since 1976.

iii. An in situ analyzer has operated on an intermittent basis on Macquarie Island (54°29′S, 158°58′E) for a few days each month since 1979.

iv. Glass flasks have been collected from Mawson, Antarctica (67°37′S, 62°52′E) since 1978, and from Cape Grim on a quasi-regular basis since the in situ program began. Flask samples have also been obtained from ships traversing the Southern Ocean to supply Australian Antarctic bases and for a short period flask samples were obtained from Wilbinga, Western Australia (31°25′S, 115°33′E). In each case the flasks were returned to the Division of Atmospheric Research in Aspendale for analysis by NDIR techniques.

v. The Division of Atmospheric Research obtains flask samples for the analysis of other atmospheric constituents by gas chromatography. The CO_2 concentration in such samples is regularly measured but gas chromatography yields lower precision than the NDIR measurements. The CO_2 concentration obtained in this way is primarily used as a quality check on the sampling procedure. Similarly, CO_2 is measured in samples obtained for mass-spectrometric measurement of stable isotopes of CO_2. These samples are obtained in 5L glass flasks from a global network of 6 to 8 stations.

3.2 Documentation

There are four main classes of documentation of the procedures and results of the Division's CO_2 programs.

i. Technical reports of the Divisions of Atmospheric Physics and Atmospheric Research.

ii. Scientific publications in international journals.

iii. Program reports and data summaries in the regular report of the Cape Grim baseline monitoring station.

iv. Regular submission of results to agencies for distribution and archiving of data.

The three technical reports relevant to NDIR measurements are Beardsmore et al. (1978), Pearman et al. (1983), and Beardsmore et al. (1984). The first two cover aircraft data after 6 and 10 years of operation. The third technical report covers the surface observational programs. A fourth technical report (Pearman et al., 1988) is currently in preparation covering some of the studies reported elsewhere in this volume on refining the baseline selection criteria, as well as describing changes in sampling and analysis procedures.

The reports of the Cape Grim baseline monitoring station have appeared at intervals of 1 or 2 years under the titles Baseline 1976, Baseline 1977, Baseline 1978, Baseline 1979-1980, Baseline 1981-1982, Baseline 83-84, Baseline 85, and Baseline 86. In each case the year(s) in the title reflect the time period of the observations reported. Baseline 87 is currently in preparation. The main scientific papers reporting the procedures and results of the CO_2 observational programs are Pearman and Beardsmore (1984) describing aircraft observations and Beardsmore and Pearman (1987) describing surface observations.

The results of our observational programs are reported on a regular basis to WMO, Data and Interpretation Section, Geneva, Carbon Dioxide Information and Analysis Center (CDIAC), Oak Ridge, TN, and the National Climatic Data Center, Asheville, NC. Supplementary information supplied with these data is generally taken from the reports listed above. Further details of CSIRO CO_2 data held by these agencies is given in Section 3.6 below.

3.3 Procedures: Aircraft

3.3.1 Automatic sampling

Automatic sampling over the Great Australian Bight commenced in December 1973 from Boeing 727 aircraft (various series) operated by Trans Australia Airlines (subsequently Australian Airlines). The samples are collected on normal commercial services using automatic equipment carried aboard specially modified aircraft. Sampling has generally been at the rate of 3 to 5 samples per month - usually there have been 3 modified aircraft in service at any one time. There has been periodic replacement of the aircraft over the period of the program. The sampling system has been described by Beardsmore et al. (1978), and has remained essentially unchanged. A single 3.8L sample is obtained in a steel flask at a

unchanged. A single 3.8L sample is obtained in a steel flask at a specified time (usually 1 hour) after passing through a specified pressure level (usually at 8.5 km). The actual sampling location is determined later from records of aircraft movements. The sample is identified as definitely/possibly stratospheric or tropospheric.

3.3.2 Manual sampling

Much of the DAR air sampling program has made use of flights by Australia's air regulatory authority (Dept. of Civil Aviation, Air Transport Group of Dept. of Transport, Dept. of Aviation). Fokker F-27 aircraft were used from 1972 to 1978 and Fokker F-28 aircraft from 1979 onwards with a greatly reduced sampling rate after 1986. These flights were supplemented by specially chartered light aircraft for the period (1978-9) of the change-over from F-27 to F-28 aircraft and for the period since 1986. The F-28 flights provided data for both the mid-troposphere and upper troposphere. Other flights provided mid-troposphere (3.5 - 5.5 km) data and occasional boundary layer data. On each flight from 6 to 10 0.5L flask samples are taken for NDIR analysis for CO_2. In many cases a small amount of the air is also analyzed by gas chromatography for CO, CO_2 and CH_4. In addition, on the more recent flights, 5L flasks samples have been obtained for stable isotope analysis of the CO_2. On each of the aircraft, the flasks are filled to approximately twice atmospheric pressure using a pump connected to one of the passenger ventilation outlets.

3.3.3 Analysis of aircraft samples

The flask samples obtained from our aircraft sampling program are analyzed in the DAR laboratory. Pearman et al. (1983) list the various NDIR analyzers used in the DAR CO_2 program at various periods and describe the analyzer characteristics including carrier-gas errors. All of our NDIR measurements use locally-produced tertiary calibration gas mixtures. The concentration of these gases are regularly checked by comparison with secondary standard gas mixtures supplied by the WMO Central CO_2 Laboratory (Scripps Institution of Oceanography). Initially CO_2-in-nitrogen standards were used and measured concentrations were corrected for carrier gas effects. Since about November 1982 these standards have been replaced by CO_2-in-air standards. These standards and our working gases have used real air rather than synthetic air.

3.4 Cape Grim in situ Procedures

In situ monitoring of atmospheric CO_2 has been conducted at Cape Grim since April 1976. Prior to September 1977 the CO_2 was only monitored for alternate 6-hour periods. Subsequently full-time monitoring could be selected when baseline conditions were expected and after September 1978 monitoring has been full-time except for operational changes or problems. Operations commenced in a large trailer van in 1981 and were transferred to the permanent laboratory of the Cape Grim Baseline Air Pollution Station (operated by the Bureau of Meteorology). The temporary CO_2 system is referred to as Mark I and the permanent laboratory system as Mark II.

Beardsmore et al. (1984) describe the monitoring equipment used in each system. Four different NDIR analyzers have been used - the periods of operation are given by Beardsmore and Pearman (1987). Since that account was published, the routine monitoring was transferred to URAS 2T Serial No. 30827270 from May 17, 1987. On this date, sampling commenced from an intake 70 m above the highest point of the cape, in contrast to the 10 m intake previously used in the Mark II system and the 3 m intake used in the temporary Mark I system. Beardsmore et al. (1984) give details of the local tertiary standard gases used at Cape Grim. Calibration of the tertiary working standards relative to the secondary standards from Scripps Institution of Oceanography is performed at the DAR laboratory at Aspendale. In addition to the tertiary standards used as reference gases once per hour, additional tertiary standard gases are used, less frequently, as surveillance gases to check on the longer-term stability of the reference gases. Routine operation uses two reference gases whose concentrations span the baseline level. In each hour there is a 10 minute calibration period - 5 minutes for each reference gas. Because of the long equilibration time of the system, the first one to three minutes of each of the five minute periods is ignored.

3.5 Data Processing

3.5.1 Flask data

The procedures for analyzing flask data are basically as described by Beardsmore et al., 1978 and Pearman et al., 1983. The raw data are in the form of analogue chart recordings for the flask sample and a pair of reference gases. These are converted to concentrations by linear interpolation with a carrier gas correction (when CO_2/N_2 reference gases were used prior to November 1982). Corrections reflecting recalibrations of the secondary standards and refinements to the system of primary standards have been applied as required. The individual flask concentrations calculated in this way are reported in the various technical papers, along with statistical summaries as appropriate.

3.5.2 Cape Grim

In the initial stages of operation the data were recorded as analogue charts. Hourly CO_2 data were extracted manually and recorded with hourly mean wind-speed and direction.

This manual procedure was superseded in February 1978 by automatic logging of data including the NDIR signal and wind-speed and direction (see Baseline 1978). The CO_2 data were recorded as 2 x 2 minute integrations for the last 4 minutes of each 5 minute calibration period and 4 x 12 minute integrations for the last 48 minutes of each 50 minute air measurement period. The data were recorded on audio cassettes and returned to Aspendale for further processing.

The automatic data logging system was upgraded in July 1981 with the move to the permanent laboratory building (the Mark II system). The main

change was that minute-long integrations were recorded. For both the Mark I and Mark II systems the within-hour values are combined to give hourly means to be used for the production of baseline selected CO_2 data sets.

The data acquisition process was further upgraded with the installation of 2 HP series 1000 mini-computers (model A700) at Cape Grim. Analogue data outputs are scanned every 6 seconds and at the end of each minute, one minute averages are calculated and archived. Walford (1987) reported further details of the GRIMCO data acquisition system. The GRIMCO system took over CO_2 data acquisition when the Mark II data logger failed due to a lightning strike on November 14, 1985.

The data processing procedures for the Mark I and Mark II systems are described by Beardsmore et al. (1984). Flowcharts of the processes are given in the appendix of that report. An account of the Mark I data processing procedure is also given in Baseline 1979-1980. For the Mark I system the sequence of operation was

i. read cassettes and transfer files

ii. insert line marking into file

iii. use working cylinder calibrations to produce concentration values

iv. apply carrier gas correction

v. compute hourly means

In the Mark II system these operations were combined into fewer steps; the final output was a file of unlaundered hourly means compatible with the Mark I output.

In each case, the initial processing was followed by editing which consisted of

i. remove hours for which instrumental problems (e.g. power failure) were noted in operation log book

ii. insert backup data for CO_2, windspeed, etc., if available.

Subsequent data processing of the files of "laundered" hourly means applied the various selection criteria discussed above.

3.6 Data Reporting

3.6.1 Cape Grim in situ

Summary CO_2 data from the in situ measurements of Cape Grim have been reported in successive issues of "Baseline". This began with Baseline 1976 presenting a graph of monthly mean CO_2 data selected according to wind

speed and direction from April 1976 to March 1977. In Baseline 1977, the record was extended to December 1977 with the data given both as a graph and as a table showing mean concentration to a precision of 0.1 ppmv and number of baseline hours in each month. In each case a fixed multiplicative carrier gas correction had been applied and the data were reported in the WMO 1974 CO_2 calibration scale. Baseline 1978 and Baseline 1979-80 extended these tables to include 1978 and 1979 respectively and reported concentrations to a precision of 0.01 ppmv.

No CO_2 data were reported in Baseline 1981-1982 because of the change from the 1974 to 1981 calibration scale and the change from the Mark I to Mark II system at Cape Grim.

Baseline 83-84 reported on studies of selection criteria and noted the adoption of the new selection criteria based on wind-speed and direction. These criteria were applied retrospectively to the whole data set and the data were presented in the WMO 1981 scale. The Mark I data included a correction for an unexplained step change. The data presented were monthly means of selected CO_2 concentrations (to a precision of 0.1 ppmv), number of days with baseline data and standard deviations of hourly mean values contributing to the monthly mean. Baseline 85 tabulated the same quantities for 1985.

More detailed presentation of the Cape Grim in situ data are given by Beardsmore et al., (1984). They tabulate daily means, number of hours data in day and standard deviations of hourly values, as well as monthly means (i.e., averages of daily means), number of days per month and standard deviations of daily values within month. These quantities are given for selected data and for all available data, for both Mark I and Mark II (through to 1982) systems. In the case of the Mark I system, data are presented both with and without the correction that was applied to take into account step change (relative to both surface and aircraft flask data) in the system's behaviour during 1978. The in situ data from Cape Grim is available from CDIAC as numeric data package NDP-010. Monthly means of selected CO_2 data through 1986 have been submitted to WMO.

3.6.2 Flask data

The primary documentation of the flask programs is the set of technical reports. For the aircraft data, Beardsmore et al. (1978) and Pearman et al. (1983) present data for each individual aircraft flask for the first 6 and 10 years respectively, using the WMO 1974 and 1981 CO_2 scales respectively. The 10-year data set is available from CDIAC as numeric data package NDP-007. Beardsmore et al. (1984) report data for individual flasks from Cape Grim (to 1981), Mawson (to 1982), Wilbinga (till end of sampling in 1981) and Southern Ocean ships (summers only, up to Feb. 1982). Monthly means and standard deviations of Cape Grim flask data are given in Baseline 1979-80 (till 1979; in 1974 scale), Baseline 83-84 (till 1984, in 1981 scale) and Baseline 85 (1985 data with note on estimated correction of 0.03 ppmv to convert to WMO 1985 scale). Beardsmore and Pearman (1987) report further southern ocean flask data.

3.7 References

Baseline 1976, Baseline Air Monitoring Report 1976. Dept. of Science Canberra, 1978.

Baseline 1977, Baseline Air Monitoring Report 1977. Dept. of Science Canberra, 1979.

Baseline 1978, Baseline Air Monitoring Report 1978. Dept. of Science and Technology, Canberra, 1981.

Baseline 1979-80, Baseline Air Monitoring Report 1979-80. Dept. of Science and Technology, Canberra, 1983.

Baseline 1981-1982, Baseline Atmospheric Program (Australia) 1981-92. R.J. Francey (Ed.), Dept. of Science and Technology and CSIRO, 1984.

Baseline 83-84, Baseline Atmospheric Program (Australia) 1983-1984. R.J. Francey and B.W. Forgan (Eds.), Dept. of Science/Bureau of Meteorology and CSIRO/Division of Atmospheric Research, 1985.

Baseline 85, Baseline Atmospheric Program (Australia) 1985. R.J. Francey and B.W. Forgan (Eds.), Dept. of Science/Bureau of Meteorology and CSIRO/Division of Atmospheric Research, 1987.

Baseline 86, Baseline Atmospheric Program (Australia) 1986. B.W. Forgan and R.J. Fraser (Eds.), Dept. of Science/Bureau of Meteorology and CSIRO/ Division of Atmospheric Research, 1988.

Beardsmore, D.J., G.I. Pearman, P.J.B. Fraser, and J.G. O'Toole, 1978. The CSIRO (Australia) Atmospheric Carbon Dioxide Monitoring Program: The first six years of data. Technical Paper No. 35, Division of Atmospheric Physics, CSIRO.

Beardsmore, D.J., G.I. Pearman, and R.C. O'Brien, 1984. The CSIRO (Australia) Atmospheric Carbon Dioxide Monitoring Program: Surface data. Technical Paper No. 6, Division of Atmospheric Research, CSIRO.

Beardsmore, D.J., and G.I. Pearman, 1987. Atmospheric carbon dioxide measurements in the Australian region: data from surface observations. Tellus, 39B, 42-66.

Pearman, G.I., D.J. Beardsmore, and R.C. O'Brien, 1983. The CSIRO (Australia) Atmospheric Carbon Dioxide Monitoring Program: Ten years of aircraft data. Technical Paper No. 45, Division of Atmospheric Physics, CSIRO.

Pearman, G.I., and D.J. Beardsmore, 1984. Atmospheric carbon dioxide measurements in the Australian region: Ten years of aircraft data. Tellus, 36B, 1-24.

Pearman, G.I., D.J. Beardsmore, and F.R. de Silva, 1988. The effects of changes in instrumentation and monitoring techniques on the carbon dioxide concentrations measured at Cape Grim. Technical Paper (in preparation), Division of Atmospheric Research, CSIRO.

Walford, P., 1987. Introduction to the Cape Grim computer system (GRIMCO), pp. 34-36 of Baseline 85.

4. STUDIES OF BASELINE SELECTION CRITERIA FOR CAPE GRIM, TASMANIA

I.G. Enting

CSIRO Division of Atmospheric Research
Private Bag No.1, Mordialloc
Victoria 3195, Australia

4.1 Introduction

The aim of baseline monitoring of CO_2 (or any other constituent) in the atmosphere is to obtain concentration values that are representative of large-scale air masses. The extent to which this aim can be achieved will depend on the monitoring site. Generally it will be easier to determine representative concentrations for air masses originating over oceans rather than land (see for example the discussion by Fraser et al., 1983). The selection criteria for obtaining representative data will vary from station to station and at some sites it may be necessary to let the selection criteria vary seasonally. For the CSIRO CO_2 measurement program, consideration of the appropriate baseline selection criteria forms part of the ongoing research activity of the program. As measurement accuracy and precision are refined by improvements in techniques, instrumentation and standards it becomes appropriate to refine the baseline selection criteria so that reductions in operational variability are matched by reductions in variability due to differences between air masses. This account summarizes three accounts of distinct studies of baseline selection at Cape Grim, Tasmania (40°41′S, 144°41′E) where CO_2 has been measured on a continuous basis since 1976.

The following section (based on Beardsmore and Pearman, 1987) describes the current operational criteria for baseline selection and includes some of the analyses that led to the adoption of these criteria. Section 3 describes a number of sets of parallel CO_2 measurements made at Cape Grim under various operational conditions. The account is based on the report by Pearman and Beardsmore (1988). The subsequent section describes a statistical analysis relating variability in CO_2 data to calculated back-trajectories. The last section summarizes the research report of Enting and Robbins (1988).

4.2 Current Selection Criteria

For most of the monitoring programs conducted at Cape Grim, "baseline" conditions are defined as being

i. winds from the sector 190° to 280° and

ii. (for some observational programs) condensation nucleii count < 600 particles cm^{-3}.

Since CO_2 is monitored on a continuous basis, however, the baseline selection criteria can make use of the statistical character of the CO_2 record itself. The current selection criteria use the wind direction (190° to 280°)

and consistency of the CO_2 record. The consistency criterion is applied to the hourly mean values and requires that an hour, m, that is accepted as baseline must be part of a sequence of N consecutive hours such that the hourly mean for each of the other N-1 hours differs from the mean of hour m by less than a cutoff d. The current selection criteria use N = 5 hours and d = 0.3 ppmv. Prior to the paper of Beardsmore and Pearman (1987), all earlier tentative data that had been published had been selected using the wind direction criterion as well as that of wind-speed > 5 m s^{-1}. The use of the wind speed criterion was dropped following the use of concentration consistency as suggested by the study by Beardsmore and Pearman (1987). They showed that there was little change in either the monthly mean concentration or the standard deviation of the set of hourly means when the wind-speed criterion was used in conjunction with the other two criteria. The main reason for this was that few additional data points were excluded on the basis of wind-speed once the wind-direction and consistency criteria had been applied.

The study by Beardsmore and Pearman reports the effects on the monthly mean and standard deviations of hourly values of various combinations of selection criteria based on wind sector (190° to 280°), wind-speed (> 5m s^{-1} if this criterion is used) and N-hour consistency to $\pm$ d (N = 5 or 10, d = 0.3, 0.5 or 1.0 ppmv).

4.3 Inter-Instrument Comparisons

During 1985 and 1986 a number of studies were conducted at Cape Grim to compare concentrations derived from 2 separate analysis systems operating in parallel. These studies are reported by Pearman and Beardsmore (1988). The first analyzer "ALPHA" was URAS 2T Serial No. 30107275 which has been the primary CO_2 monitoring instrument for CO_2 at Cape Grim from September 1984 to May 1987 as well as for several earlier periods. The second analyzer "BETA" was URAS 2T Serial No. 30827270 which had been the primary CO_2 monitoring instrument from June 1981 to July 1983 and from November 1983 to September 1984. Prior to the experiments, analyzer BETA was modified to improve its electrical, optical and temperature stability. None of these changes seemed to make any significant improvement to the noise levels in analyzer BETA. What was observed, however, was a mean difference of 0.06 ppmv between ambient CO_2 concentrations as measured by the two instruments, even though they were both using the same pair of calibration gases. This difference remains unexplained. As part of the sequence of comparisons, changes were made to the operation of BETA, comprising:

i. changes of techniques of drying sample and calibration air streams;

ii. more complete temperature equivalence between sample and calibration gas streams;

iii. additional particle filtering (down to 1 micron) to reduce possible scattering in the measurement cell;

iv. the use of no-flow conditions for measurements.

None of these changes was found to make any detectable change in the difference between the two analyzers. The most significant comparison was between samples taken from the 10 m and 70 m intakes. (Heights are relative to the observation deck of the station which is level with the highest point of the cliff-top, 92 m above sea level, see Fig. 1). Figure 2 shows a record of hourly mean values for 6 to 12 November, 1986. The record shows a mean difference between the analyzers, somewhat less than half of which can be attributed to the 0.06 ppmv difference between measurements obtained when the two analyzers are measuring identical airstreams. The remaining difference presumably represents an actual vertical gradient. It will also be seen that the record from the 10 m intake shows greater short-period variability (only partly accounted for by the fact that ALPHA is a less stable instrument than BETA) as well as having a pronounced diurnal cycle. In contrast the record from the 70 m intake is dominated by the longer term synoptic changes. Figures 3a and 3b show the 1500 h 1000 hPa trajectories for the period covered by Fig. 2, indicating the passage of a cold-front on 11 November.

As a result of this study, the routine in situ CO_2 measurements at Cape Grim have used the 70 m intake and analyzer BETA since 27 May 1987.

4.4 Trajectory Studies

As reported by Forgan (1988) 4-day back-trajectory information for Cape Grim (1500 h local time, 1000 hPa) is now available on a routine basis both as hourly co-ordinates and summary information. Enting and Robbins (1988) reported the results of a pilot study designed to test the utility of trajectory information in refining baseline selection criteria.

The study covered the years 1984 and 1985, and looked at daily mean data constructed from hours satisfying the CO_2 baseline selection criteria wind from the 190° to 280° sector and 5-hour consistency to within 0.3 ppmv). Baseline CO_2 data were recorded on 360 of the 731 days. For 26 of the days, the mean is based on only 1 or 2 hours of data. (The 5-hour consistency criterion can produce isolated baseline-selected hours if these hours correspond to extrema or to approximate points of inflexion. Isolated hours can also be produced by the operation of the wind-direction criterion). There were 4 occasions with gaps of 10 days or more in the record. Further details of the data coverage for 1984-5 and the longer period 24 July 1981 to 31 December 1985 are given by Enting and Robbins. The variability of the daily mean CO_2 concentrations was expressed in terms of deviations about spline fits to the data. In order to test for associations between trajectories and variability in the CO_2 concentrations, the 1500 hr trajectories for the 360 days were classified as:

i. rejected: (82 days) trajectories that passed over mainland Australia, Tasmania or New Zealand;

ii. doubtful: (86 days) trajectories that were near land at the 4-day point, came close to land or were parallel to coastlines for a considerable

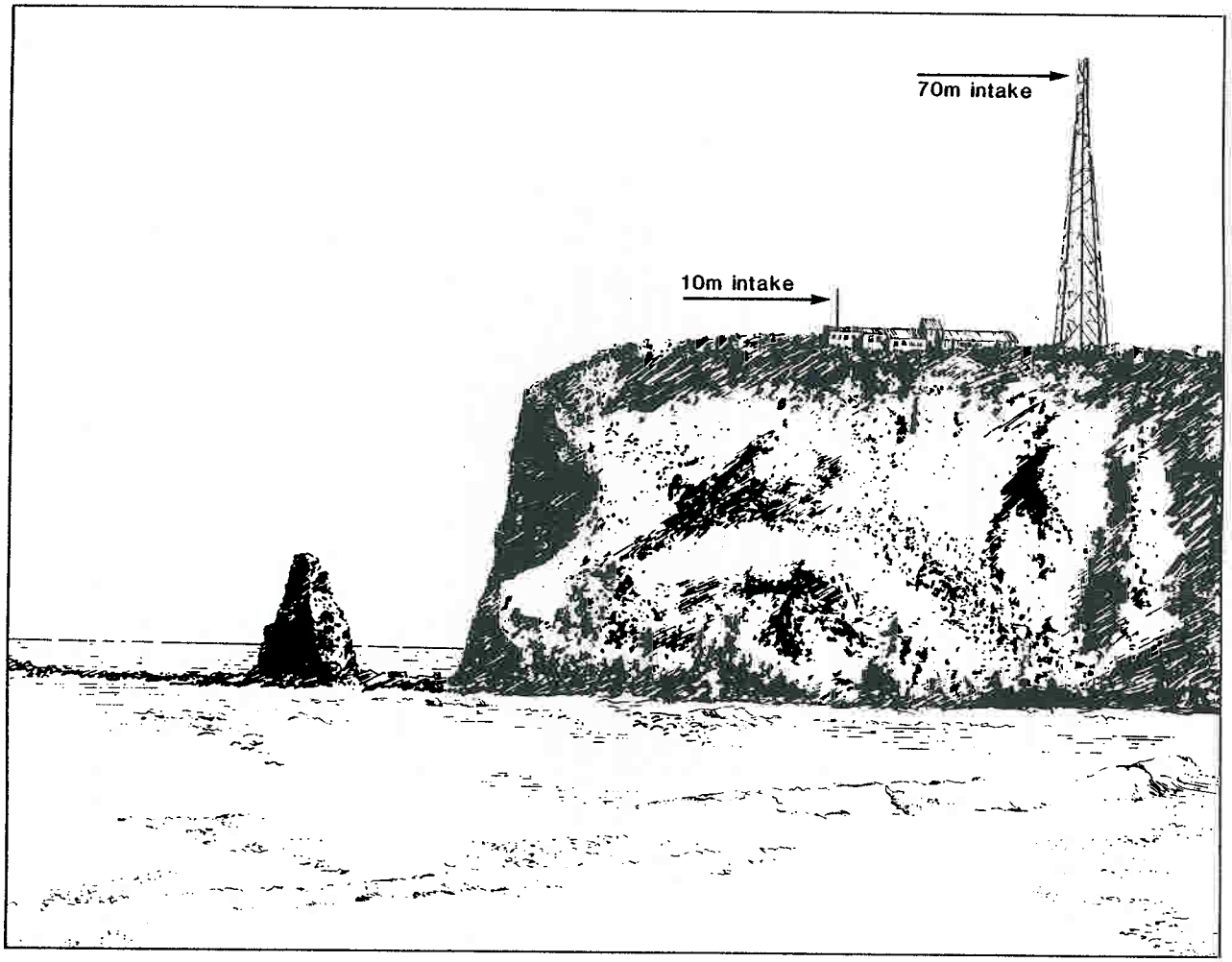

Figure 1. Cape Grim baseline station as seen from the south, showing 10 m intake and telecommunications tower with 70 m intake.

distance. Ocean trajectories reaching Cape Grim at small angles to the west coast of Tasmania were also classified as doubtful;

iii. accepted: (192 days) all others.

Figure 4 gives examples of the classification.

The CO_2 data for days with "accepted" trajectories were characterized by having proportionally fewer daily means constructed from only 1 or 2 hours of data when compared to the 1984-5 selected data set as a whole (7 out of 192 compared to 26 out of 360 - further details are given by Enting and Robbins, 1988). When compared to the set of all selected data, the data corresponding to "accepted" trajectories had more long gaps (both in absolute number and in proportion). There were 11 gaps of 10 days or more, the longest being 41 days with all other gaps less than 23 days.

The main results of the study were the residuals of the various subsets of the data relative to a spline fit to the full set of selected data. The spline used was chosen to give 50% attenuation of sinusoidal variations with a period of 30 days. Thus variations of period T days will be attenuated by an amount $1/(1 + (30/T)^4)$ according to the asymptotic formula describing the behavior of the smoothing splines.

The RMS residuals from this spline fit were 0.223 ppmv for all data, 0.128 ppmv for data corresponding to accepted trajectories, 0.170 ppmv for data corresponding to doubtful trajectories and 0.380 ppmv for data corresponding to rejected trajectories. Residuals relative to smoother splines are given by Enting and Robbins. They also give an RMS residual value of 0.090 ppmv for residuals of the "accepted" data relative to a spline fit (with 50% attenuation at 30 days) to the "accepted" data itself. The reduction in the residuals (relative to the corresponding fits to all data) would seem to indicate that fitting the whole data set gives a small bias and that the 0.090 ppmv is the appropriate measure of the variability (on periods less than 30 days) in the "accepted" data.

The differences in the RMS residuals for the various classes of trajectory indicate that there is scope for refining the baseline selection criteria. The trajectories can not give a complete solution to this problem because they are only available for 1500 h on each day. Data rejected on the basis of the calculated trajectory may in fact come solely from the other times of the day when the actual trajectory may have been acceptable. There is some indication of this type of effect in the distribution of residuals of the rejected data relative to the spline fit. This shown in Fig. 5, along with the normal distribution having the same standard deviation. Clearly the distribution of residuals is non-normal. It could however be regarded as a broad "outlier" distribution superimposed on a core of normal data whose distribution differs little from that corresponding to accepted trajectories. An important aim of future studies is to define selection criteria that enable us to retain these "core" data while excluding the "outliers". Such studies are currently in progress.

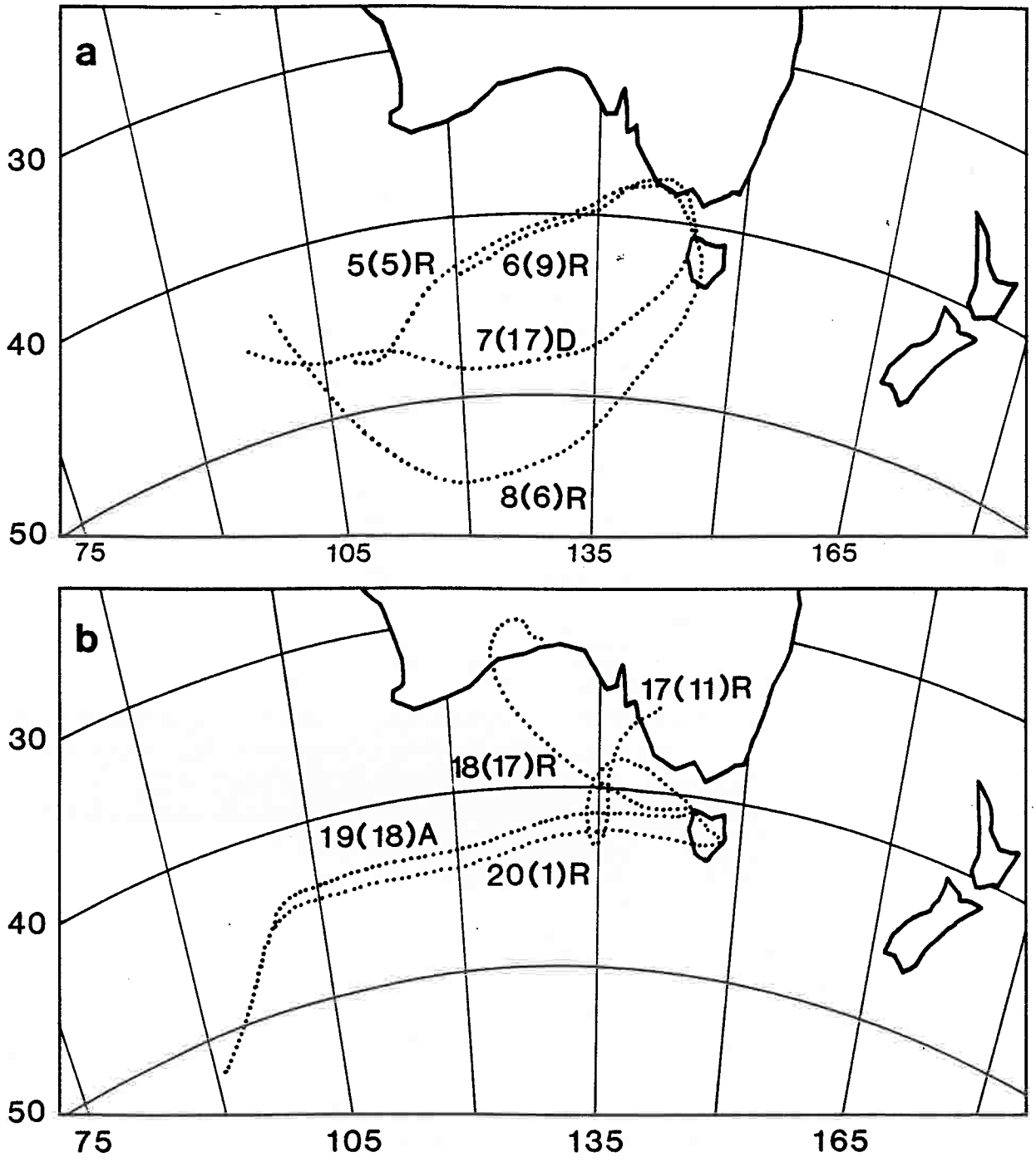

Figure 4. (a) and (b). Trajectories for April 1984, showing classification into Accepted (A), doubtful (D) and rejected (R). First number is April 1984 data, number in brackets is number of hours in day that satisfied baseline CO_2 selection criteria.

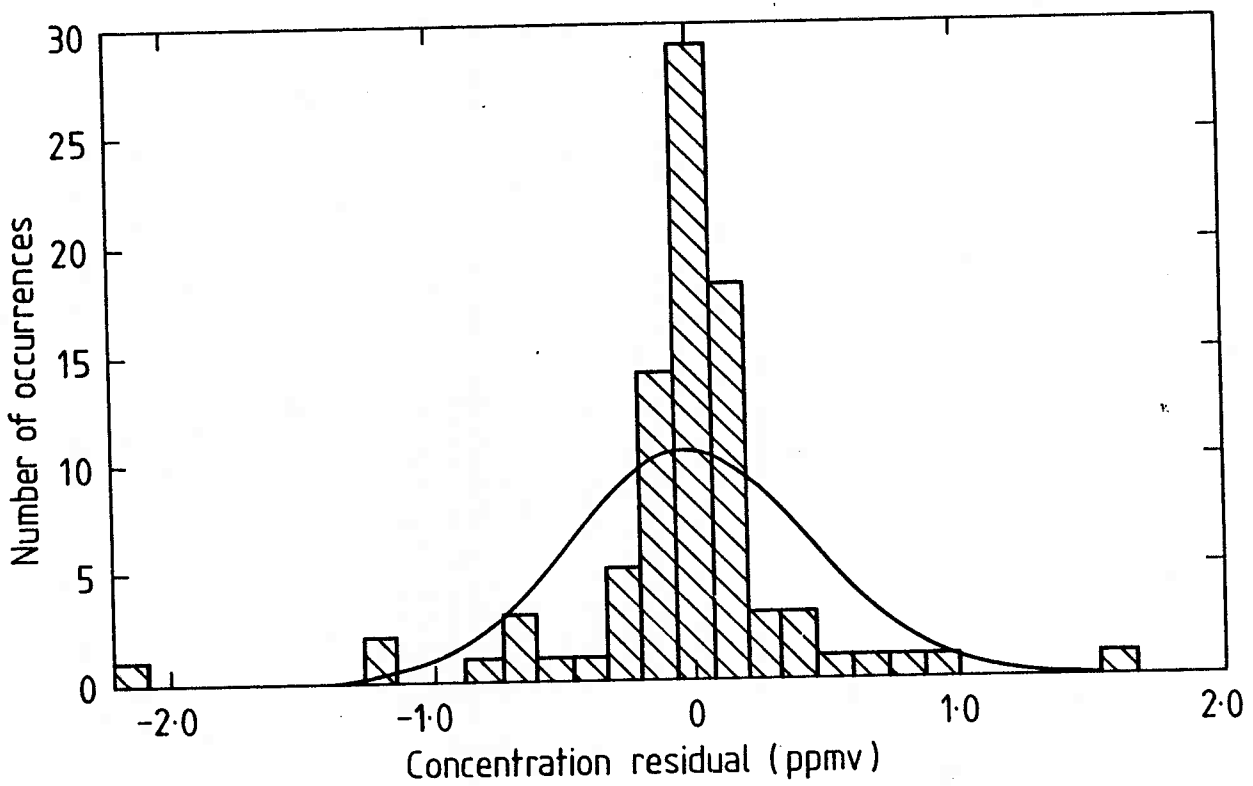

Figure 5. Distribution of residuals from spline (fitted to all selected data and using 50% cutoff at 120 days) for selected CO_2 data corresponding to rejected trajectories.

4.5 References

Beardsmore, D.J., and G.I. Pearman, 1987. Atmospheric carbon dioxide measurements in the Australian region: Data from surface observatories. Tellus, 39B, 42-66.

Enting, I.G., and F.J. Robbins, 1988. Trajectory analysis of selected CO_2 data. Section 3.6 of Baseline 86. P. Fraser and B. Forgan (Eds.), CSIRO and Dept. of Administrative Services.

Forgan, B., 1988. Section 6.16 of Baseline 86, P. Fraser and B. Forgan (Eds.), CSIRO and Dept. of Administrative Services.

Fraser, P.J., G.I. Pearman, and P. Hyson, 1983. The global distribution of atmospheric carbon dioxide II. A review of provisional background observations 1978-1980. J. Geophys. Res., 88, 3591-3598.

Pearman, G.I., D.J. Beardsmore, and F.R. de Silva, 1988. The effects of changes in instrumentation and monitoring techniques on the carbon dioxide concentrations measured at Cape Grim. Division of Atmospheric Research, Technical Paper (in preparation), CSIRO.

5. THE USE OF LOESS AND STL IN THE ANALYSIS OF ATMOSPHERIC CO_2 AND RELATED DATA

William S. Cleveland
Jean E. McRae

AT&T Bell Laboratories
Murray Hill, NJ 07974

Abstract. Loess is a procedure for putting smooth curves and surfaces through noisy data. STL is based on loess; it decomposes a second time series into trend, seasonal, and remainder components. This paper shows how loess and STL can be used in the analysis of atmospheric CO_2 and related data, and describes computer routines that implement the procedures.

5.1 Introduction

5.1.1 What loess and STL do

Loess is a procedure for putting smooth curves or surfaces through noisy data (Cleveland et al., 1988; Cleveland and Devlin, 1988). Figure 1 shows an example. The circles are daily averages of measurements of CO_2 made at Mauna Loa by NOAA (Komhyr and Harris, 1977; and Thoning et al., 1986). The time period covered is 4612 days but since there are missing values, there are only 4193 observations. The curve is a loess smoothing of the data. Figure 2 shows a different loess smoothing that captures just the seasonal variation and the long-term trend in the data.

In Figs. 1 and 2, a dependent variable, CO_2 concentration, has been smoothed as a function of one independent variable, time. But loess can also be used to smooth data as a function of two or more independent variables. Figures 3 and 4 show an example. The data are temperature measurements at 348 locations in an eastern portion of the U.S. The measurement sites are shown in Fig. 3; the coordinates are those of an Alber projection. Contours of a fitted loess surface are shown in Fig. 4.

STL (Cleveland et al., 1988) is a procedure for decomposing a seasonal time series into trend, seasonal, and remainder components. Figure 5 shows an example. The data, monthly averages of CO_2 measurements made at Mauna Loa by Scripps (Keeling et al., 1982), are graphed in the top panel. The next three panels show trend, seasonal, and remainder components from STL. The trend component describes the long-term behavior in the series. The seasonal component describes the yearly periodic behavior. This component is shown again in Fig. 6 where the values for each month are graphed together against the years; the horizontal line is the average of all values for the month, and the ends of the vertical lines that emanate from the horizontal line indicate the monthly values. The remainder component is the remaining variation in the time series. The three components are a decomposition of the series in that the series is equal to the sum of the three components.

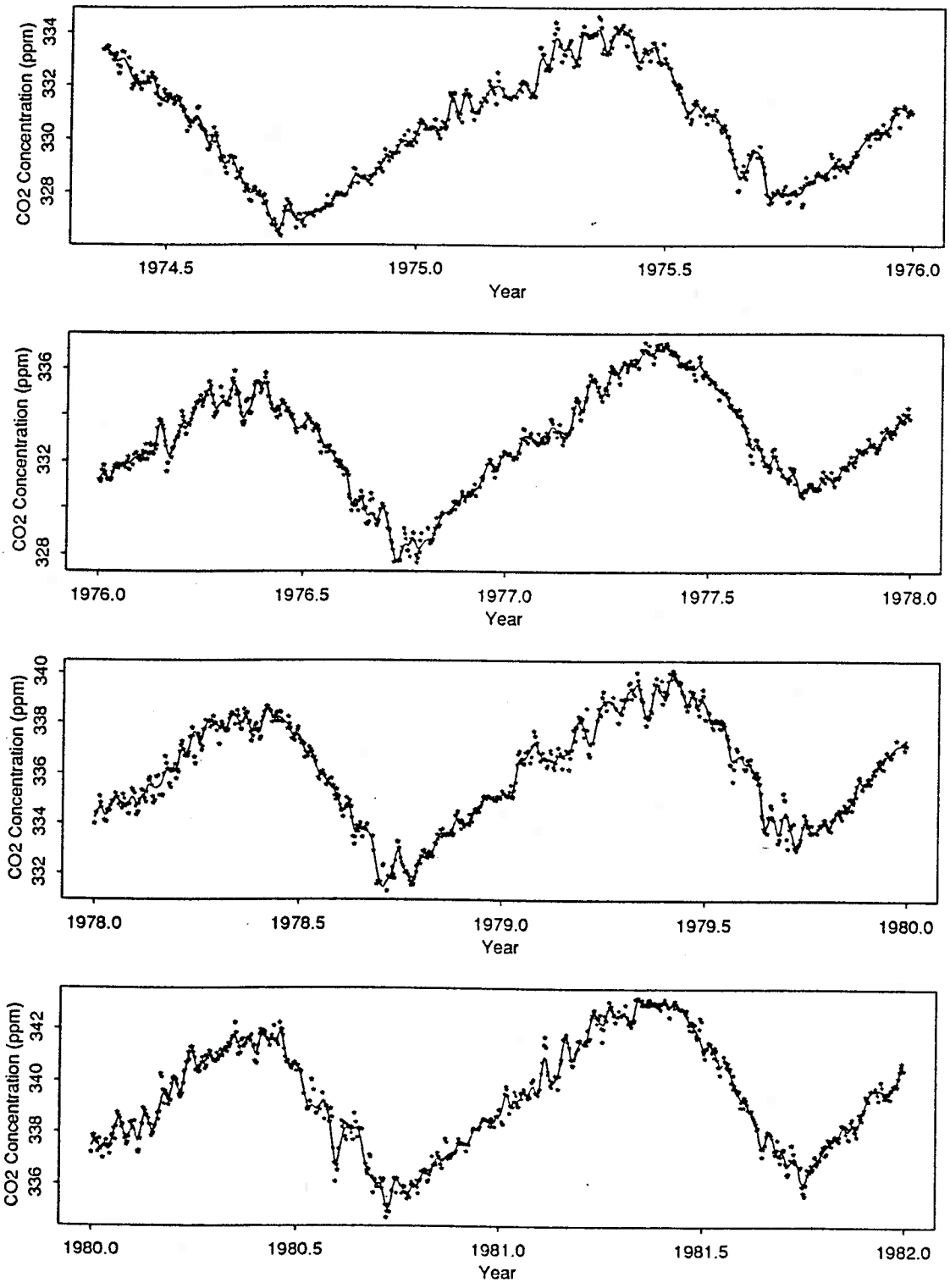

Figure 1. Loess smoothing of daily CO_2 data.

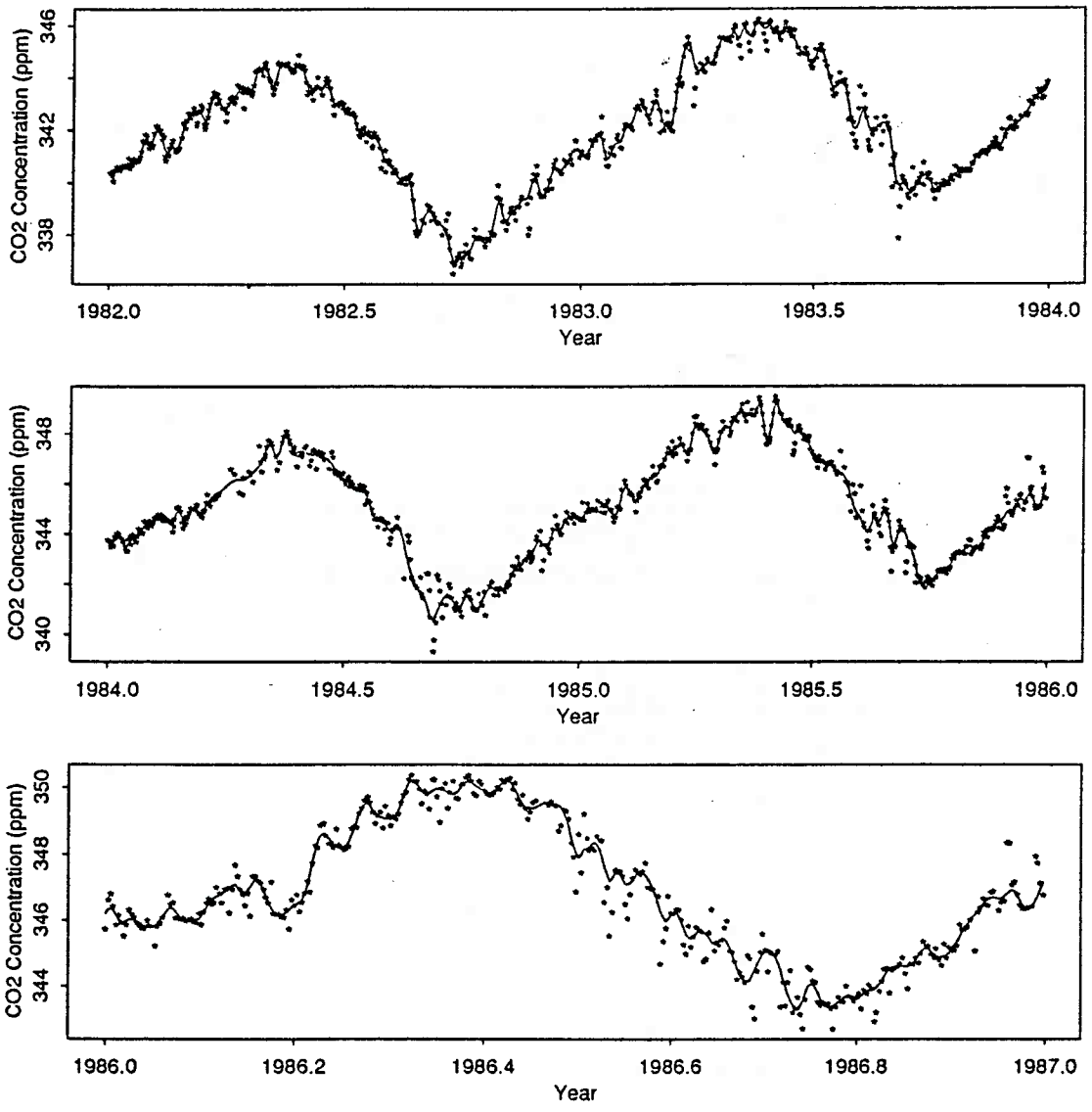

Figure 1. Con't.

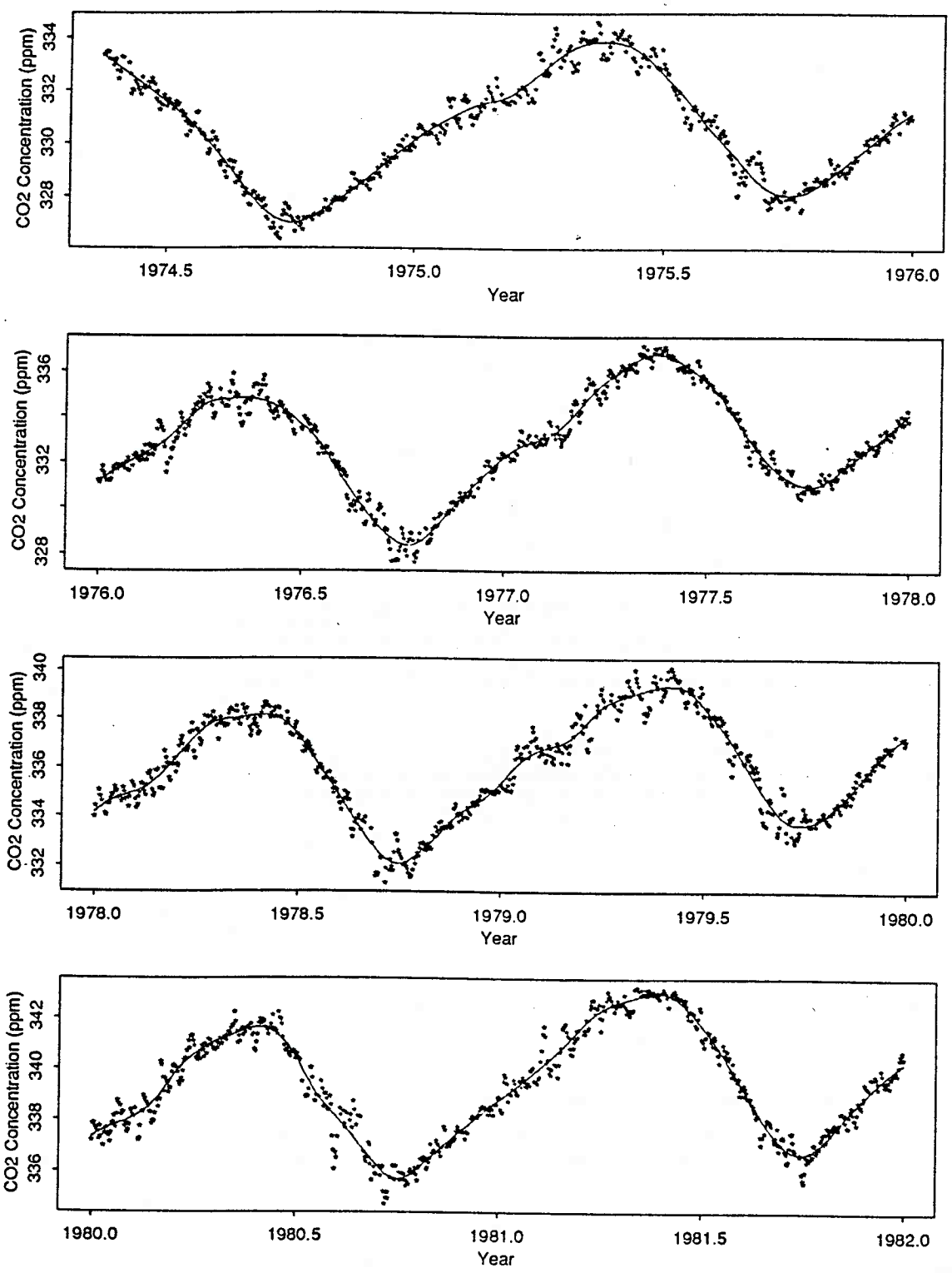

Figure 2. Loess smoothing of daily CO_2 data.

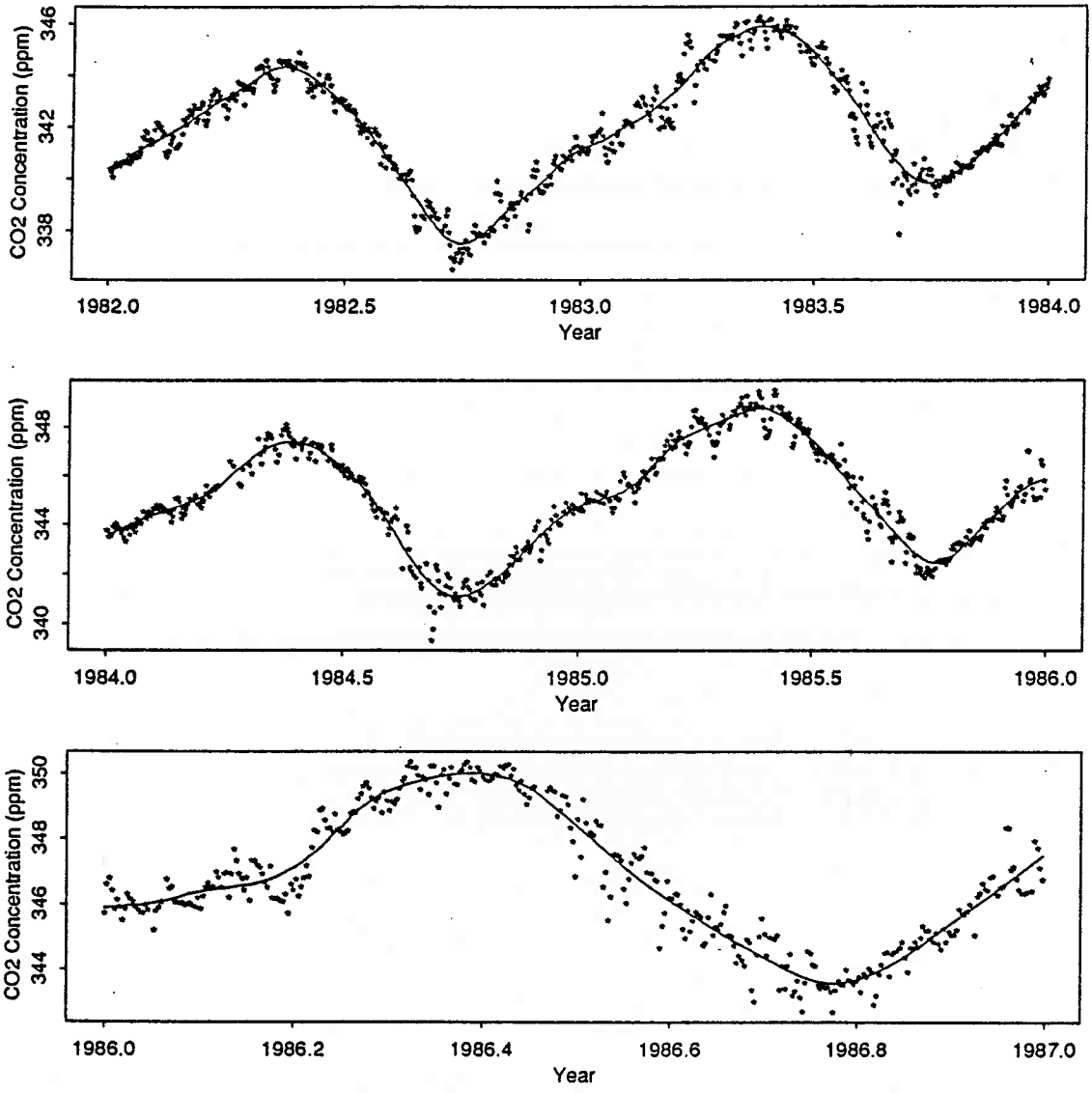

Figure 2. Con't.

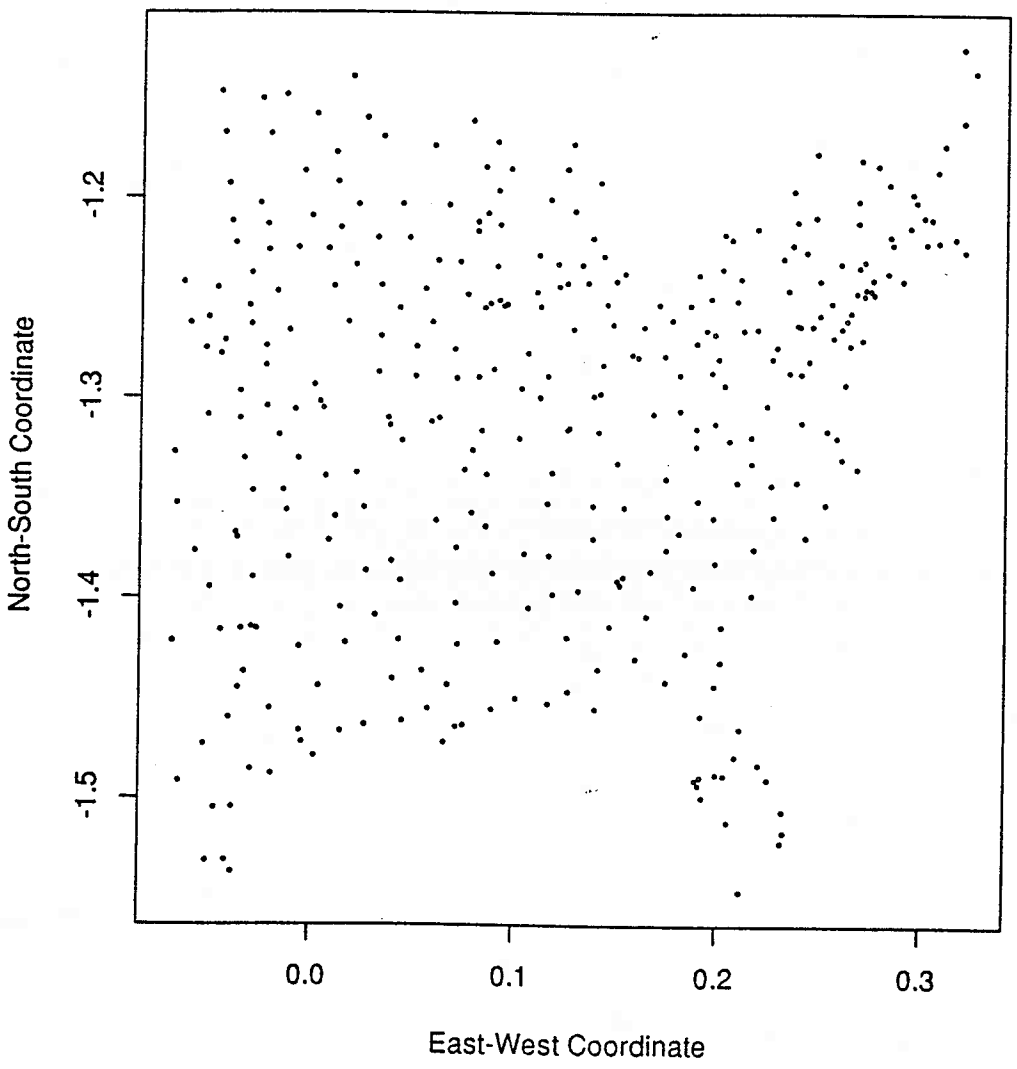

Figure 3. Locations of temperature measurements in eastern portion of U.S.

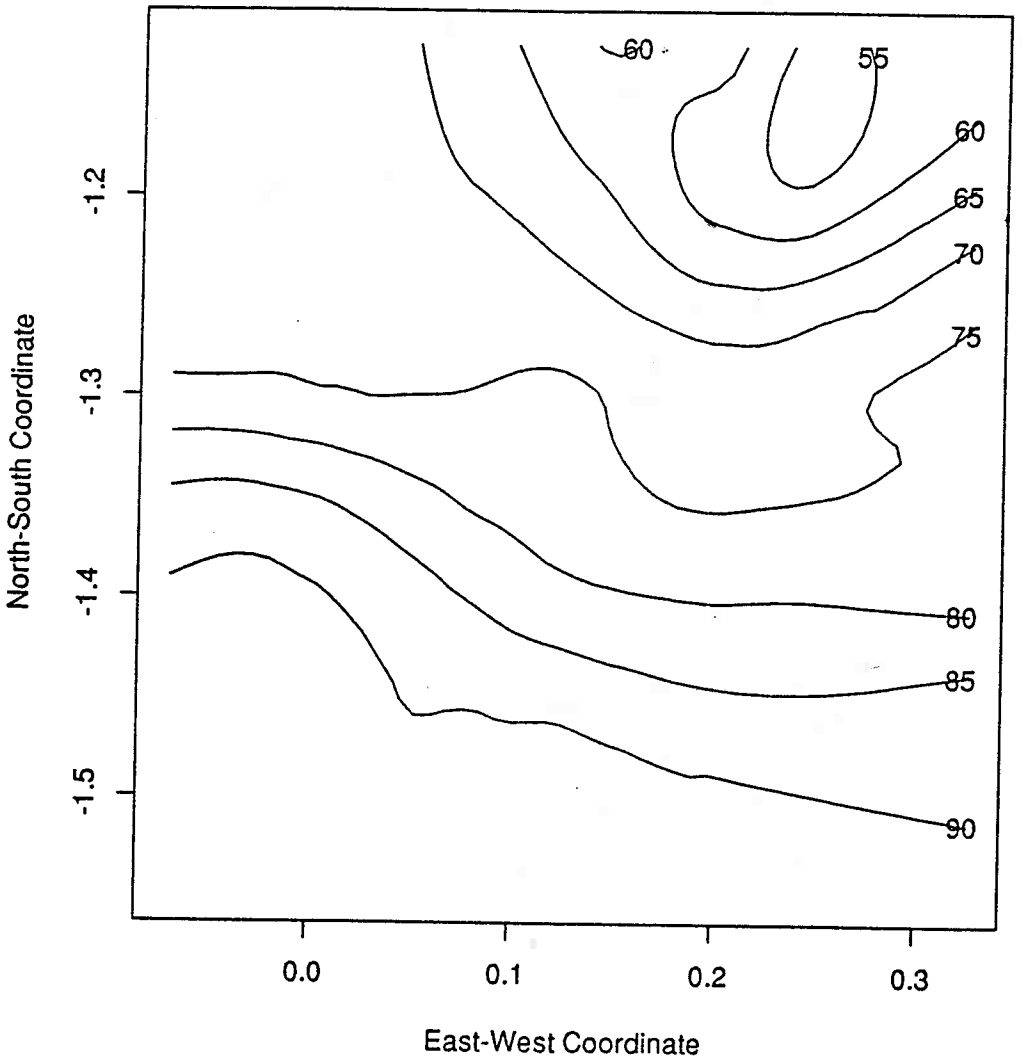

Figure 4. Contours of loess surface fitted to temperature data.

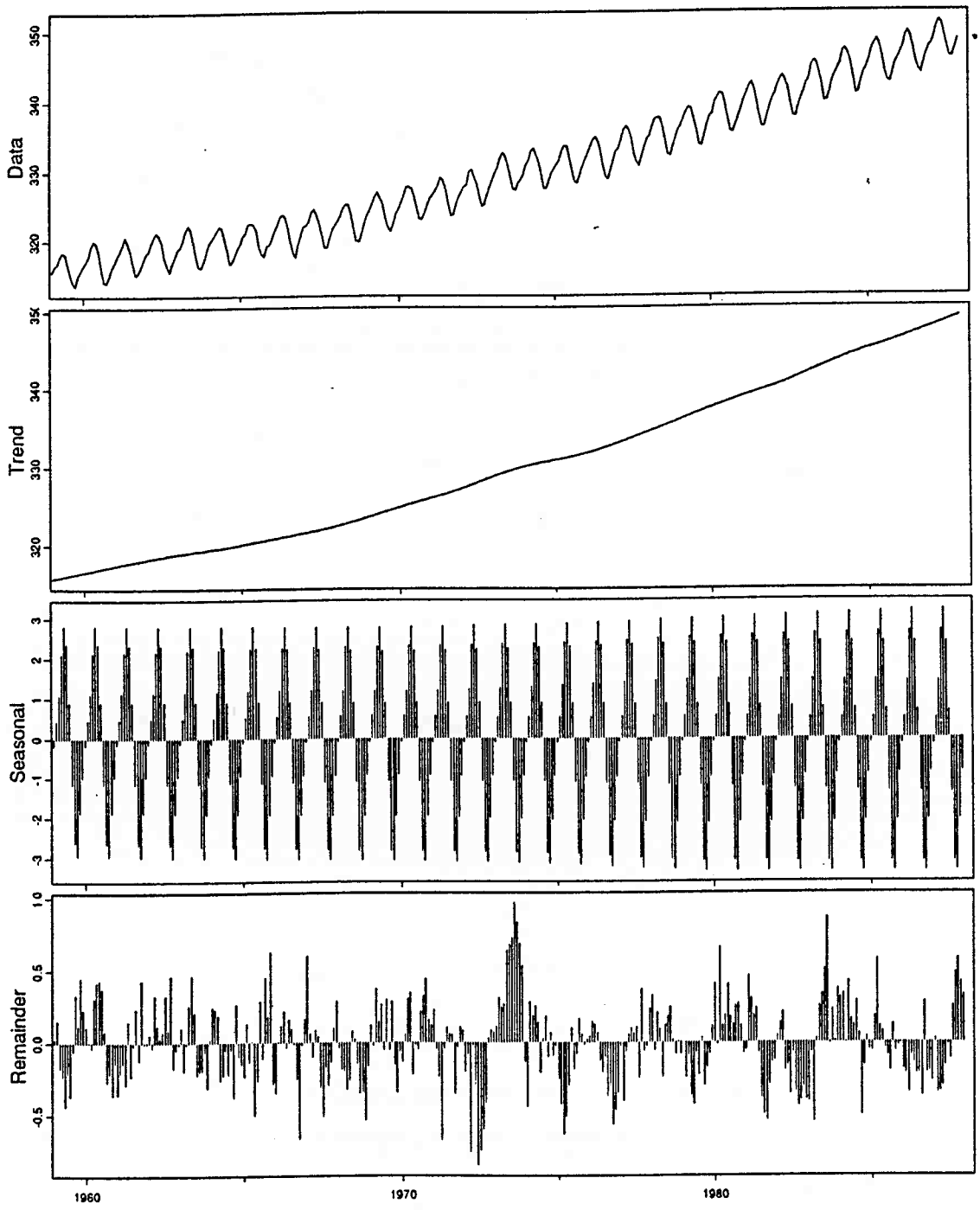

Figure 5. STL decomposition of monthly CO_2 data.

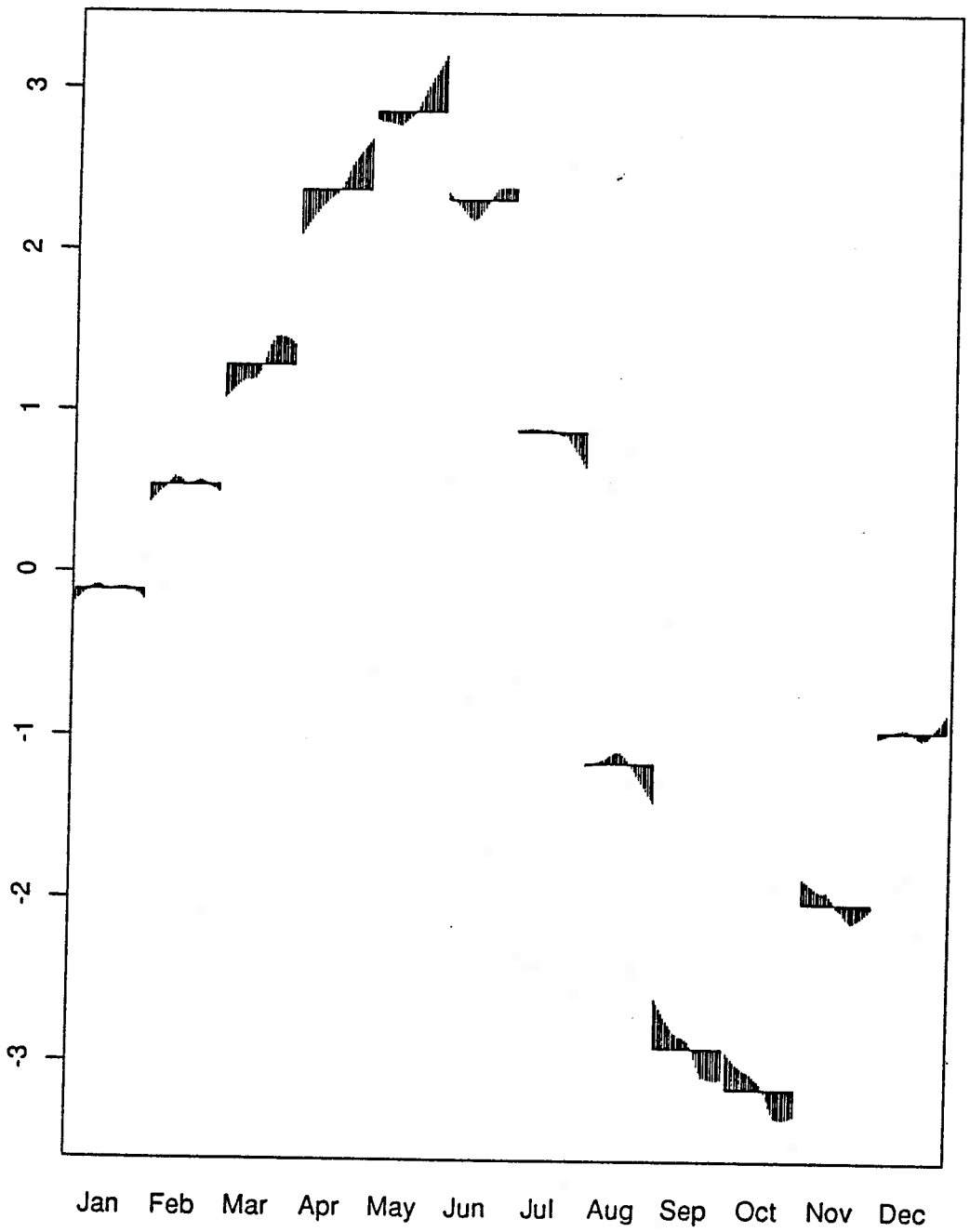

Figure 6. Seasonal component from STL decomposition.

5.1.2 Salient features

Loess has a robustness feature that prevents outliers from distorting the smoothed values. Also, loess can track peaks and valleys that are locally-quadratic in behavior. These features are important in applications to geophysical data and are not present in most other implementations of smoothing procedures such as cubic smoothing splines (Reinsch, 1967). There is a smoothing parameter that controls the smoothness of the loess curve or surface. This parameter is a positive number and as it increases, the smoothed values become smoother. In Fig. 1, f = 15/4193 and loess tracks much of the synoptic variation in the data. (For reasons that will be clear later, we have expressed f as a fraction whose denominator is the number of daily CO_2 measurements.) In Fig. 2, f has been increased to 101/4193 and now the loess curve is smoother and tracks the seasonal and long-term trend but not the synoptic variation.

STL has two smoothing parameters, n_t and n_s. The first controls the smoothness of the trend component. The second controls the smoothness of the seasonal values across time at each position of the cycle; for example, for the monthly Scripps data, this means the smoothness of the values of each monthly series through the years. An important property of STL is that compared to other decomposition procedures (Zellner, 1978), it is extremely simple. Consequently, programs that implement it are simple in structure and run fast even for time series with thousands of observations. Finally STL has a robustness feature that prevents outliers from distorting the trend and seasonal components.

5.2 Using Loess

Detailed accounts of the definition of the loess procedure can be found in other references (Cleveland et al., 1988; Cleveland and Devlin, 1988). Here, we will focus on the use of the procedure in practice.

5.2.1 One independent variable

Suppose the measurements of the dependent variable are y_i for i = 1 to n. Suppose, for the moment, that there is just one independent variable and that the measurements are x_i for i = 1 to n. The loess curve, $\hat{y}(x)$, is defined at any value, x, along the scale of the independent variable. The basic idea of the procedure is this. Let f be the value of the smoothing parameter, a positive number. Suppose for the moment that $f \leq 1$. Let q be fn truncated to an integer. The q values among the x_i that are closest to x are found, and (x_i,y_i) for each x_i in this collection is given a neighborhood weight based on how close x_i is to x; points whose x_i are closer to x are given larger weight. Then a linear or quadratic polynomial is fit to the (x_i,y_i) using weighted least squares, and the value of the polynomial at x is $\hat{y}^{(0)}(x)$, the initial

estimate. This initial estimate is computed at $x = x_i$ for $i = 1$ to n. The residuals for this initial smoothing are

$$\hat{\varepsilon}_i^{(0)} = y_i - \hat{y}^{(0)}(x_i).$$

Robustness weights are computed that decrease as $|\hat{\varepsilon}_i^{(0)}|$ increases; thus points yielding residuals with large magnitudes receive less weight. The updated loess estimate, $\hat{y}^{(1)}(x)$, is similar to $\hat{y}^{(0)}(x)$ except that (x_i,y_i) receives a weight equal to the product of the neighborhood and robustness weights. $y^{(1)}(x)$ is computed at the x_i, and new residuals and robustness weights are computed; the second update, $\hat{y}^{(2)}(x)$ is then computed using the new weights. These robustness iterations can be carried out any number of times, but two are usually sufficient; that is, the final estimate is $\hat{y}(x) = \hat{y}^{(2)}(x)$.

The value of f can also be greater than 1; in this case, all of the points get positive weights that all increase to 1 as f increases. So for very large values of f, the loess smoothing is just a globally linear or quadratic surface fitted to the data.

5.2.2 Two or more independent variables

The loess procedure for two or more independent variables is a straightforward extension of the procedure for one independent variable. We need only one additional item: a notion of distance in the space of the independent variables.

Suppose there are two independent variables, w and x, whose measurements are w_i and x_i for $i = 1$ to n. The definition for more than two will be obvious from the definition for exactly two. To compute the loess surface at (w,x), a notion of distance from (x,w) to (x_i,w_i) is needed. We will use one of two definitions of distance. One is just Euclidean distance

$$\sqrt{(w - w_i)^2 + (x - x_i)^2}\ .$$

The other is to standardize the variation in the measurements and then use Euclidean distance for the standardized variables. We will standardize by the 10% trimmed standard deviation, which is defined by the following: Sort the data from smallest to largest, delete the smallest 10% of the data and the largest 10% of the data; and then compute the sample standard deviation of the remaining observations. Let s_x be the 10% trimmed sample standard deviation of x_i and define s_w similarly. Then the standardized measurements are x_i/s_x and w_i/s_w, and the distance is

$$\sqrt{\left(\frac{w - w_i}{s_w}\right)^2 + \left(\frac{x - x_i}{s_x}\right)^2}$$

Now we proceed exactly as in the one-dimensional case - find the q closest (w_i,x_i) to (w,x), compute weights, fit a linear or quadratic surface, and update based on the magnitudes of the residuals.

5.2.3 Choosing parameters in the use of loess

If there is one independent variable, the data analyst must make four choices:

(1) The value of f.

(2) Locally-linear or locally-quadratic fitting.

(3) The number of robustness iterations.

(4) The values of x at which to compute $\hat{y}(x)$.

If there are two or more independent variables, there is one more choice:

(5) Euclidean distance or standardized Euclidean distance.

5.2.3.1 The value of f

As we stated earlier, $\hat{y}(x)$, gets smoother as f increases. The choice of f depends on the pattern of the data and the goal of the analysis. We suggest trial and error. The objective is to choose f as large as possible, which makes $\hat{y}(x)$ as smooth as possible, without distorting the pattern in the data. For example, for the loess smoothing in Fig. 1, we wanted to track the synoptic variation, so f was taken to be quite small, f = 15/4193. (4193 is the total number of observations, n, so f = 15/4193 means the span of the smoother, q, is 15 days.) In Fig. 2, the goal was to reproduce the long-term trend and seasonal oscillations, and ignore the higher-frequency variation, so f was taken to be much larger, f = 101/4193.

5.2.3.2 Locally-quadratic vs. locally-linear fitting

Locally-quadratic fitting is sensible to use when the pattern in the data being smoothed has peaks and valleys. Locally-linear fitting tends to chop off peaks and valleys unless f is quite small; in other words, one has to put up with a curve or surface that is locally noisy in order to track the pattern in the data. (Cubic smoothing splines (Reinsch, 1967) have the same problem.) Locally-quadratic fitting can yield a curve or surface that is both smooth and tracks peaks and valleys. For the loess smoothings in Figs. 1 and 2, locally-quadratic fitting was used because of the many peaks and valleys. For the loess smoothing in Fig. 4, locally-linear fitting was used since the curvature is moderate.

5.2.3.3 The values of x

For one independent variable, it makes sense to compute loess, $\hat{y}(x)$, at a collection of equally-spaced values from the minimum x_i to the maximum. This was done in Figs. 1 and 2; in both cases the spacing is one day.

For two independent variables, it makes sense to compute loess on a rectangular grid and then graph the surface in some way such as a contour

plot. For the temperature data, loess was computed at a 50x50 rectangular grid. Let h_i for $i = 1$ to 348 be the horizontal coordinates of the measurement locations in Fig. 5, and let v_i be the vertical coordinates. The horizontal coordinates of the grid are 50 equally spaced values from the minimum h_i to the maximum, and the vertical coordinates are 50 equally spaced values from the minimum v_i to the maximum.

5.2.3.4 The number of robustness iterations

Choosing the number of robustness iterations requires careful judgment of the data. Let y_i for $i = 1$ to n be fitted values, that is, a loess curve or surface evaluated at the values of the independent variables. Let

$$\hat{\varepsilon}_i = y_i - \hat{y}_i$$

be the residuals. If the smoothing is likely to result in outliers, that is, residuals whose magnitudes are large compared with the others, then it is sensible to use two robust iterations. If there are no outliers, then the number can be zero; however, in this case, it does no harm to use two. For Fig. 1, the number was taken to be 2 since the residuals can contain errant observations because the rejection procedures used to define the daily data are not infallible (Thoning et al., 1988). For the temperature data, the number was also 2 since a few observations deviate radically from the pattern of the majority of the data.

There is, however, a situation where it is best to set the number of robustness iterations to zero. Suppose that the residuals contain systematic variation. This is the case in Fig. 2 because we chose to eliminate the synoptic variation in the smoothing which means this variation winds up in the residuals. We do not want to reduce the weights for observations that happen to occur at the peak or trough of a synoptic cycle; so in this case it is sensible to use no robustness iterations. We are not operating in an inconsistent manner by choosing two iterations in Fig. 1 and none in Fig. 2 because the residual variation is quite different in the two cases. The errant observations can distort $\hat{y}(x)$ locally when the high-frequency synoptic variation is included in the smoothing, but do not have an appreciable effect when the smoothing eliminates the synoptic variation.

5.2.3.5 Euclidean distance or standardized Euclidean distance

If the values of the independent variables are the physical locations of measurements of the dependent variable, then it makes sense to preserve the metric of the measurement space and use Euclidean distance. If the independent variables are not positions, in particular if they have different units, it makes sense to use the standardized Euclidean distance. For the smoothing used in Fig. 4, Euclidean distance was used.

5.3 The Use of STL

A detailed account of the definition of STL can be found elsewhere (Cleveland et al., 1988). Here, we will focus on the use of the procedure in practice. Suppose the time series is X_v for $v = 1$ to N. The program

to be described in Section 5.4 that carries out STL requires the series to be measured at equally-spaced points in time, which means missing values are not allowed. As we will see later, missing values can be filled in with loess and then the STL program can be applied to the expanded data. However, while the program requires no data be missing, the basic idea of STL does not, and a program could be written that allows missing values, although speed would be sacrificed somewhat.

Suppose n_p is the number of values, or period, in each cycle of the seasonal component. For the Scripps CO_2 concentrations in Fig. 1, $n_p = 12$; that is, there is a yearly periodicity. The result of STL is a decomposition,

$$X_v = T_v + S_v + R_v,$$

where T_v is the trend component, S_v is the seasonal component, and R_v is the remainder.

5.3.1 The basic idea

STL is based on the loess smoother. There is a succession of applications of loess that loops back and forth between trend and seasonal smoothings; this trend-seasonal looping is nested inside iterations in which robustness weights are defined for the trend and seasonal smoothings.

For simplicity of description let us suppose that the time series X_v is measured monthly and $n_p = 12$. Recall that in the definition of loess, q is fn truncated to an integer. Here we work directly with q rather than using f.

Suppose T_v, S_v, and ρ_v are the current values of the trend component, the seasonal component, an the robustness weights, respectively. Suppose that the ρ_v have just been computed. Then T_v and S_v are updated in the following way. First, T_v is subtracted from X_v to form a detrended series. Each monthly subseries of the detrended series is smoothed separately by a loess robustness smoothing with locally-linear fitting, with q set equal to n_s, an odd integer, and with robustness weights gotten from the corresponding monthly subseries of ρ_v. That is, the January values are smoothed, the February values are smoothed next, and so forth. These smoothed values are put together to form a temporary seasonal smoothing that is adjusted to remove any trend that might have gotten into the component. This temporary component is subtracted from X_v to form a deseasonalized series. This deseasonalized series is smoothed directly by a loess robustness smoothing with $q = n_t$, an odd integer, with locally-linear fitting, and with robustness weights, ρ_v to form a temporary trend component. These seasonal and trend smoothings loop back and forth a total of n_i times. The resulting trend and seasonal components are the updated values T_v and S_v. Now the remainder

$$R_v = X_v - T_v - S_v$$

is computed and the updated robustness weights are computed on the basis of the R_v. The robustness iterations are carried out a total of n_o times.

5.3.2 Choosing parameters in the use of STL

To use STL, the data analyst must make four choices:

(1) n_t

(2) n_s

(3) n_i

(4) n_o

5.3.2.1 n_t

As n_t increases, the smoothness of the trend component increases. As with the value of f in loess, n_t must be chosen on the basis of the data being analyzed and on the basis of the goal of the analysis. For the decomposition of the Scripps CO_2 data in Fig. 1, n_t= 51.

5.3.2.2 n_s

As n_s increases, the smoothness of each subseries of the seasonal component increases. For example, for the seasonal component of the Scripps data, each monthly subseries of the seasonal component get smoother as n_s increases. For the decomposition shown in Figs. 5 and 6, n_s = 19.

5.3.2.3 n_i

Convergence of the temporary components in the trend-seasonal smoothings is usually quite fast, so n_i = 2 is usually sufficient and n_i = 4 affords much protection. The smaller value is sensible if there are two robustness iterations and the large if there are none.

5.3.2.4 n_o

Much the same thinking applies to n_o as to the number of robustness iterations for loess. If there is likely to be outliers that are large compared with the variation in the remainder, then n_o = 2 is sensible. If there is likely to be no outliers then n_o = 0 is sensible if there is substantial systematic variation in the remainder. For the Scripps data of Figs. 5 and 6, n_o = 0 since the data are particularly well-behaved and since there is systematic variation in the remainder.

5.3.3 Filling in missing values

The daily NOAA data in Figs. 1 and 2 have missing values. To decompose these data, the loess curve of Fig. 1 was evaluated for each day and the loess smoothed values were used in the decomposition. That is, the loess smoothed values were the input of STL. Before doing this, however, February 29 was deleted so that n_p = 365 consistently. The resulting decomposition is shown in Fig. 7. The seasonal smoothing length is n_s = 19, the same value used for the decomposition of the Scripps data. The trend smoothing length is

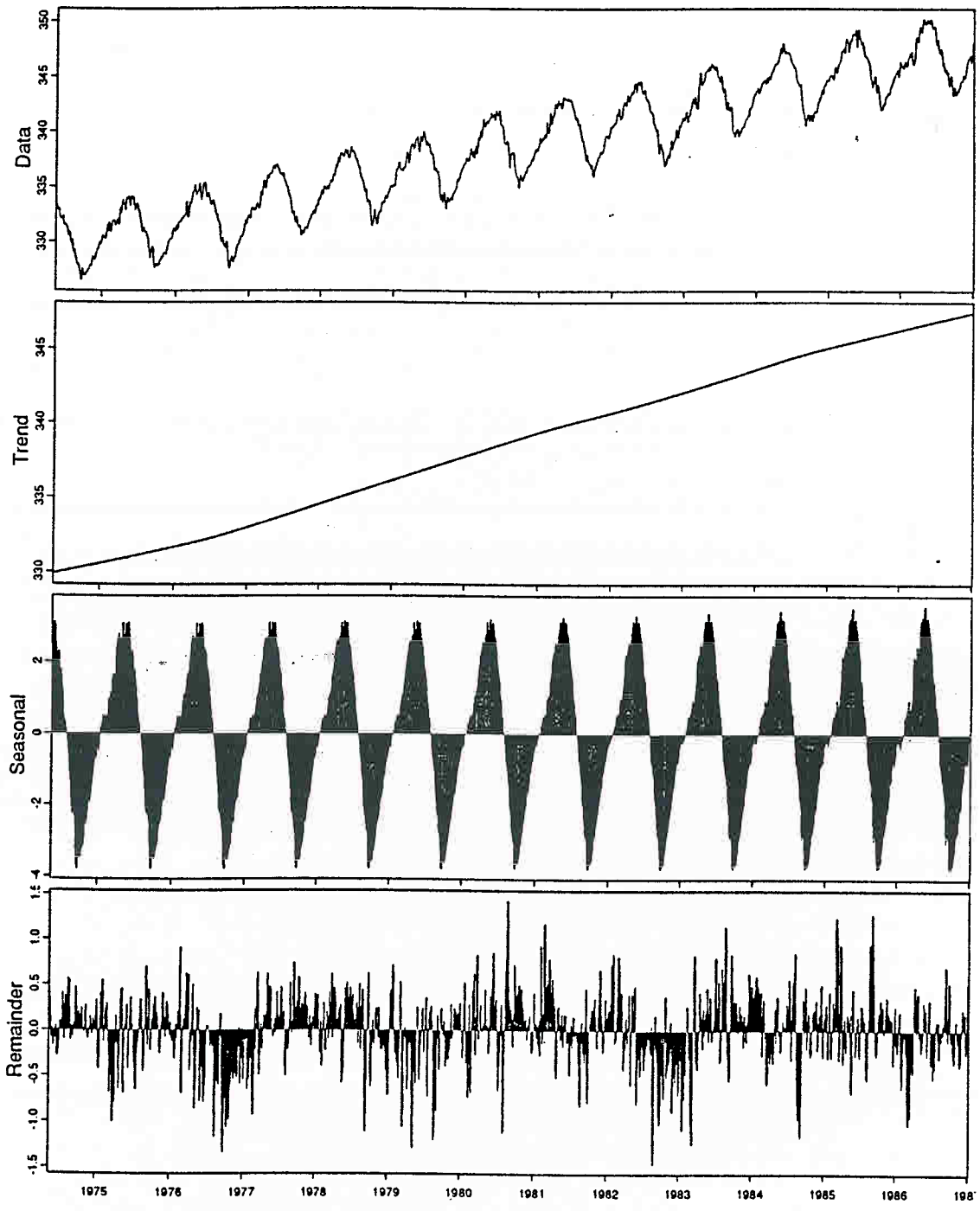

Figure 7. STL decomposition of loess smoothing of daily CO_2 data.

n_t = 1501 to capture just the very long-term trend. The number of robustness iterations is n_o = 0 because the data, the result of a robust loess smoothing, do not contain outliers; outliers in the raw data have been dealt with by the loess smoothing and go into the residuals. Finally, the number of trend-seasonal loops is n_i = 4.

5.3.4 A post-smoothing for the seasonal component

Consider the smoothing for the seasonal component of the NOAA daily series. Each daily subseries of the seasonal component is smoothed separately. This means that each subseries will be smooth across years. For example, the values of the seasonal component for July 4 will be a smoothly changing series from one year to the next. But the smoothing does not guarantee that the seasonal component from one day to the next will be smooth. The reason why STL does not impose such smoothness is that there are many time series for which the seasonal component is not smooth from one time point to the next. But for the NOAA daily data we quite clearly want to have a component that is smooth from one day to the next. However, as Fig. 7 shows, there is local roughness in the seasonal component.

There is a simple solution to a locally-rough seasonal component--a post-smoothing in which the seasonal component from STL is smoothed by loess. The smoothed values are the final seasonal components. In carrying out this smoothing we want to be sure to use locally-quadratic fitting because there are many peaks and valleys in the seasonal component. Also, there is no need for robustness iterations since STL produces a well-behaved component apart from the local roughness. Finally, we are unlikely to want to smooth too much since the roughness will typically have a small variance. In fact, for the Scripps CO_2 monthly data, there is no discernible local roughness in the seasonal component from STL, as Fig. 5 shows, so no post-smoothing was used.

For the NOAA daily data, f was taken to be 51/4609 for the post-smoothing. (The total number of values of the seasonal component is 4609, so that span, q, of the smoothing is 25.) The final decomposition is shown in Fig. 8.

5.4 Computer Routines

Fortran subroutines are available that carry out loess and STL. In this section, a few comments are made about the routines, and ways of obtaining the code are described.

All of the routines have parameters that can be set to certain values to speed up computations. While the details vary from one routine to another, the basic idea is the same; exact loess computations are carried out at a set of points smaller in number than the full set where the loess surface is evaluated, and then the evaluation is carried out for the full set using interpolation.

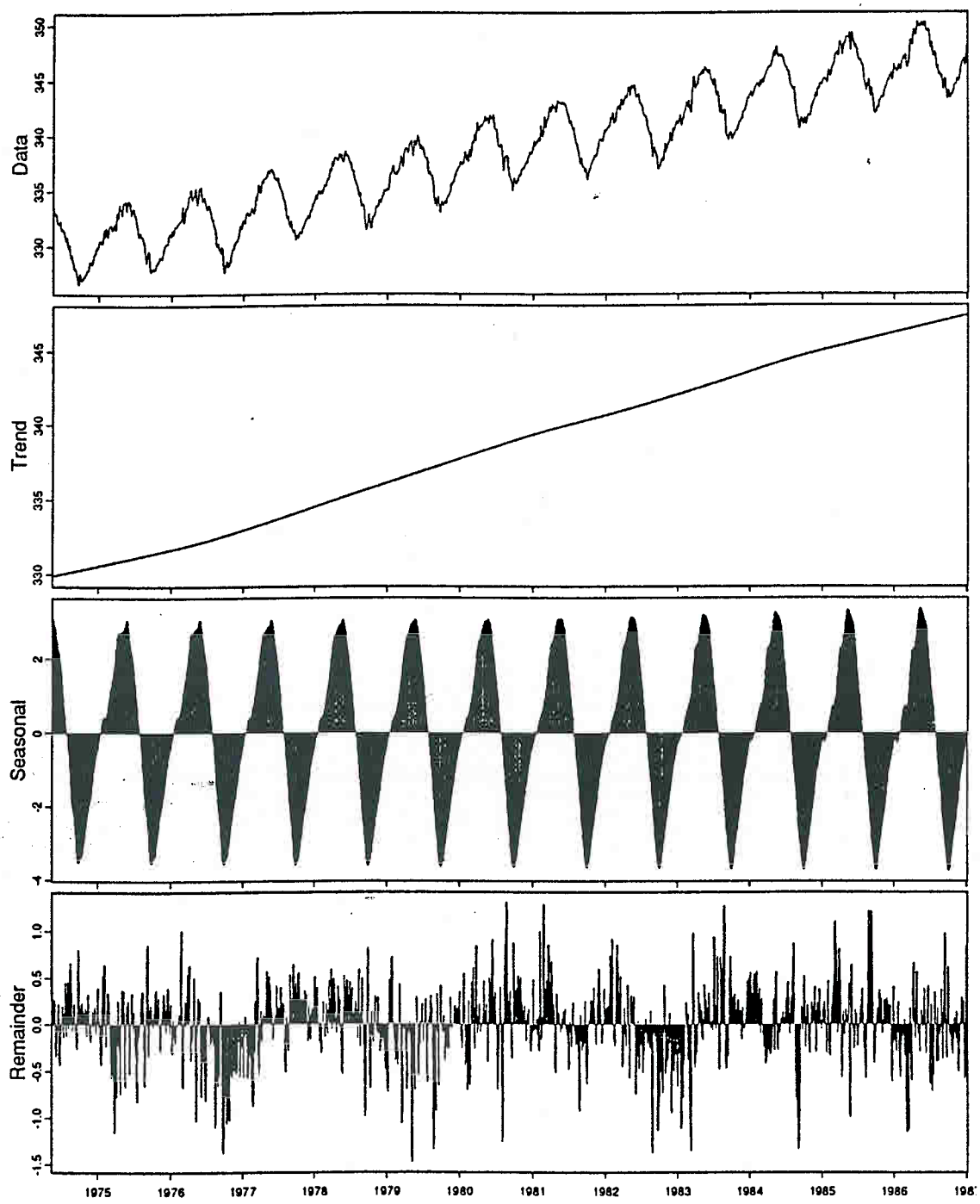

Figure 8. STL decomposition with post-smoothing of seasonal component.

5.4.1 Lowess

The classical program implementing the first version of the loess procedure (Cleveland, 1979) is a routine called lowess. It has two limitations: it accommodates only one independent variable, and implements only locally-linear fitting and not locally-quadratic fitting. The speed parameter is delta. A reasonable default value for delta is the range of the x_i divided by 100.

5.4.2 Loess, lostat, and loster

The routine loess carries out the full set of procedures described here. The speed parameter is fcell. A reasonable value for fcell is f/10 unless f is quite small in which case f/5 might be reasonable.

The routines lostat and loster provide information for making statistical inferences. It is important to note that these statistical procedures are relevant only if the deviations from the surface can be assumed to be independent, identically-distributed random variables with a normal, or more generally, a symmetric distribution. This would be a reasonable assumption for the temperature data in this paper but not for the CO_2 data.

5.4.3 Stl

The routine stl carries out the STL procedures as described here. The speed parameters are nsjump and ntjump. A reasonable default value for the first is the integer closest to ns/10 and a reasonable value for the second is the integer closest to nt/10.

5.4.4 Obtaining the routines

The Fortran subroutines may be obtained in one of two ways. The simplest is to send electronic mail to a login at Bell Laboratories:

UUCP: research!netlib

INTERNET: netlib@research.att.com

with the message

send lowess from go

to get lowess, with the message

send loess from a

to get loess, or the message

send stl from a

to get stl. A program reads the message and automatically distributes the code to your return address. Also, programs may be obtained on a floppy disk by writing to

John Kimmel
Wadsworth Publishing Co.
25742 Woodbrook Rd.
Laguna Hills, CA 92653

5.5 Acknowledgments

We are grateful to Pieter Tans of the University of Colorado for providing us with the NOAA CO_2 data and to Thomas Boden of the Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory for providing us with the Scripps CO_2 data.

5.6 References

Cleveland, R.B., W.S. Cleveland, J.E. McKae, and I.J. Terpenning, 1988. STL: A seasonal-trend decomposition procedure based on loess. Journal of Official Statistics, (to appear).

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

Cleveland, W.S., S.J. Devlin, and E. Grosse, 1988. Regression by local fitting: Methods, properties, and computational algorithms. J. of Economics, 37, 87-114.

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

Keeling, C.D., R.B. Bacastow, and T.P. Whorf, 1982. Measurements of the concentration of carbon dioxide at Mauna Loa Observatory, Hawaii. In Carbon Dioxide Review: 1982, W.C. Clark (Ed.), 317-385, Oxford University Press, New York.

Komhyr, W.D., and T.B. Harris, 1977. Measurements of atmospheric CO_2 at the U.S. GMCC baseline stations. Air Pollution Measurement Techniques: Special Environmental Report 10, WMO 460, 9-19. World Meteorological Organization, Geneva.

Reinsch, C., 1967. Smoothing by spline functions. Numerische Mathematik, 10, 177-183.

Thoning, K.W., P.P. Tans, and W.D. Komhyr, 1988. Atmospheric carbon dioxide at Mauna Loa Observatory: II. Analysis of the NOAA/GMCC data. 1974-1985. J. Geophys. Res. (submitted).

Zellner, A. (Ed.), 1978. Seasonal Analysis of Economic Time Series. U.S. Bureau of the Census, Washington, DC.

6. ANALYSIS OF SELECTED TIME-SERIES OF ATMOSPHERIC CARBON DIOXIDE CONCENTRATIONS

G. Tunnicliffe Wilson
Lancaster University, U.K.

Abstract. The pattern of variation in atmospheric carbon dioxide concentrations is modelled on a range of time-scales. The time series analysis is based on fitting autoregressive-moving average models. Filters defined by these models are then used to extract the trend and seasonal variations in the data. A robust filter is used for level estimation and interpolation of hourly data.

6.1 Introduction

The data supplied were derived mainly from measurements of atmospheric carbon dioxide concentrations made at the Mauna Loa Observatory by the Geophysical Monitoring for Climatic Change division of NOAA's Air Resources Laboratory. They consisted of hourly, daily and monthly mean values, with some missing values in the hourly and daily means. Subsets of the data were selected and subjected to time series modelling and analysis, as part of an exercise to illustrate the properties of different data processing procedures.

6.2 The Monthly Data

The supplied time series corresponds to the period May 1974 to September 1987 inclusive. There were no missing values.

The method of analysis used was that advocated by Box and Jenkins (1976) for seasonal time series. The autocorrelations of the data, and of various differences of the data, were examined. This led to the selection of a standard model, expressed in the operator notation of Box and Jenkins as

$$(1-B)(1-B^{12})\ Z_t = (1-qB)(1-rB^{12})\ a_t$$

where B is the back-shift operator, Z_t is the given series, and a_t is a white noise series. The parameters q and r are estimated to have values 0.49 and 0.86 respectively, and the residual standard deviation (of the series a_t) was 0.30. This measures the standard deviation of one-step ahead forecasts.

The model was subject to checking by examination of the residual autocorrelations. These should show no statistical evidence of being different from zero, but in fact some slight residual autocorrelation was found at lag 8. This will be discussed later.

The interpretation of the coefficients is in terms of forecast adaptation. The model is developed to capture the statistical structure for the purpose of forecasting, using a minimum of free parameters. For this

model the parameter q measures the speed of adaptation of the model to level changes, and q = 0.49 implies rapid adaptation with a time constant of 1/(1-q), or about 2 months. This reflects the fact that much of what appears to be irregular fluctuation in the series from month to month, is in fact a series of cumulative independent increments with the nature of a random walk. This is to be expected from physical considerations, because CO_2, once released into the atmosphere will have a fairly long residence time on the time scale considered here.

The coefficient r=0.86 similarly measures the speed of adaption of the model to changes in the seasonal pattern, the time constant in this case being 1/(1-0.86) or about 7 years. The fact that r is estimated to be different from 1.0 is statistical evidence that the seasonal pattern is changing over the length of these data. Compare this with the model fitted to data from 1964 to 1978, reported in the paper by Tunnicliffe Wilson and Belcher (1986), for which the estimate of r was virtually 1.0, implying constant seasonality.

However, the model was refitted to the logarithms of the present series, and the estimate of r then moved to 0.998. The implication is that in the original series, the small changes evidenced in the additive seasonality may be ascribed to a constant but multiplicative seasonal pattern. The best power transformation was estimated, and was not statistically different from the logarithmic.

The models were fitted to less than the full span of the series, the last two years being retained for checking purposes. The values for these final 24 points were very well forecasted by the model, confirming its correctness.

One possible criticism of the model is that it provides no separate parameter which may be tuned to adapt to changes in trend. The time constant for this is effectively the sum of the two already mentioned. Slight underestimation of this constant, particularly when it is fairly large, does not seriously affect the forecasting ability of the model, and so is not, statistically, of great importance. In this case all the evidence points to a constant trend throughout the data. The evidence for a logarithmic transformation does suggest that on the original scale the trend is exponential, but with variations over the fitted span of only about 5%, this would appear almost linear even without the random walk fluctuations.

A more general model was also fitted, of the form:

$$(1-B)1-B^{12})Z_t = (1-q_1B - \ldots - q_{13}B^{13})\ a_t$$

which permits separate adaptation to level, trend and seasonality and incidentally encompasses a wide variety of structural models. This gave the same conclusion, that the trend is constant. This model also "mops up" residual autocorrelation previously observed at around lag 8, though provides no reason for its presence. The forecasts were slightly different, and a little improved, and a possible candidate for the inadequacy of the first model was thought to be calendar effects on both the emission of CO_2 and the compilation of the monthly mean levels. Some attempt was made to represent these by introducing explanatory dummy variables for the numbers of weekends

in the month, etc. This was after noticing the weekly cycle evident in the daily mean levels. The effects were just significant, but failed to remove the residual autocorrelations which caused the concern. The irregularity of these effects throughout the year would explain why fixed explanatory variables were inadequate--see the account of the daily data.

Finally consider the decomposition of the series to separate the trend and seasonality. There is no unique structural model, i.e., a model containing independent stochastic features for the trend, seasonal and error components of the model, which is compatible with the fitted Box-Jenkins model. Burman (1980) discusses and exploits the possibilities. In fact, taking r=1.0 implies that we can decompose the series as a fixed trend, fixed seasonal pattern, and an independent random walk and random component. These components may optimally be estimated by two sided filters. What is presented in Fig. 1, is the result of one sided filters applied to present and past series values. They are uniquely defined, following the model forecast properties. The model produces from any time point, a forecast pattern which is the sum of a linear trend and a seasonal pattern. It is possible therefore to decompose the series into components which, separately forecasted, would yield the linear trend and the seasonal part. The components which achieve this and yet follow the lowest order models, define the decomposition. The remaining component in this case is a random series. The filters are defined in Tunnicliffe Wilson and Belcher (1986).

This decomposition was therefore carried out based on the first model fitted. Figure 1 shows the trend component, with the constant value 325 subtracted, the seasonal component, and the noise or irregular component with the value 5 subtracted, otherwise the vertical scale is the same for all three components.

It would be possible to subtract a fixed linear trend from the trend component, in order better to reveal the fluctuations, and similarly for the seasonal component. However, the components as presented provide a fair picture from which some understanding might be gained of the sources of variation in the monthly mean levels. The first difference of the level component may be useful when discussing variation in CO_2 emissions which might be evidenced in the data.

6.3 The Daily Data

Similar techniques were applied to a subset of the daily data. The fifth set of 300 points was selected, the first day being Nov. 28, 1977. There were 28 values missing, so that a period of 328 days was covered, the last day being Oct. 21, 1978.

No seasonal correlation was expected within this period, but sample autocorrelations of the series and its first difference (allowing for missing values) revealed a seasonal period of 7 days, confirmed by periodogram analysis of the first differences.

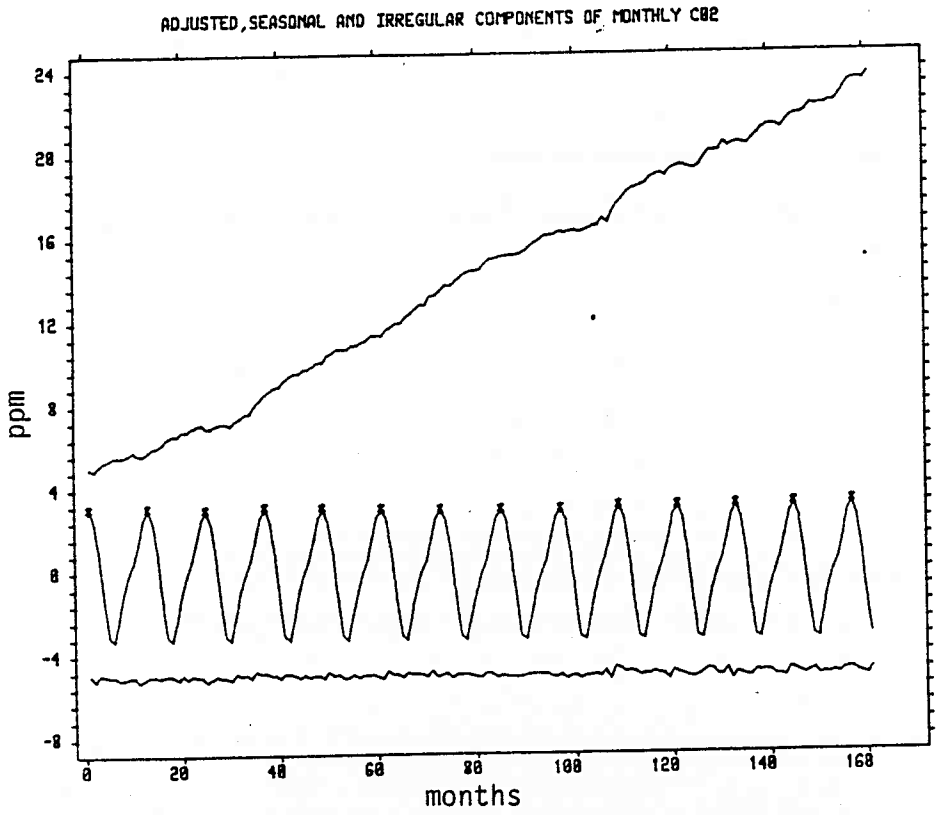

Figure 1. Components of monthly CO_2 data. Top curve is adjusted trend, middle curve is the seasonal component, bottom curve is irregular component. See text for explanation of vertical scale.

A Box-Jenkins model was fitted, and extended to allow for some low lag residual autocorrelation, to a model of the form:

$$(1-B)(1-B^7)\ Z_t = (1-q_1B - \ldots - q_4B^4)(1-rB^7)\ a_t$$

with estimated coefficients q_1=0.05, q_2=0.24, q_3=0.17, q_4=0.25, r=0.79.

It was clear that analysis of data for a period of under one year, could not define an annual pattern, which appeared only as a levelling off then a sharp downward trend change. This limitation on the analysis could be overcome by progressing to an analysis of a much longer period.

The main comment on the coefficients is that r is not close to one, and indicates a changing weekly pattern.

Thirty-five values at the end of the selected 328 were held back from the estimation, to be used for checking. The forecasts of these values were not good for lead times more than a few days, mainly, of course, because of the leveling out of the downward trend which could not be predicted without seasonal information.

A decomposition of the data was carried out on the same basis as that for the monthly data, and is shown in Fig. 2. The most obvious point is that the estimated weekly pattern is quite variable, growing through the winter months, and declining in the summer months. Moreover the pattern changes through the winter, with the peak moving. Asterisks have been used to mark every seventh day on this graph. Confidence intervals could be calculated to indicate how well the weekly pattern is estimated, since there is quite a lot of variability in the data on short time scales. On this graph the trend component has had 330 subtracted and the residual component has had 2 subtracted.

It may be that the sharp trend change has thrown the model and the output from the filters to some extent--the series could be split into two parts to check this. The weekly pattern is however real enough, as the periodogram indicates. The question is just how stable it is.

The noise or irregular component is correlated up to three days lag, and presuming that this is not the result of data smoothing, it would indicate some persistence in the concentration of CO_2 in the air being measured. Again the decomposition gives scope for considering various causes of the variations observed in the series.

The missing values were interpolated at the estimation stage of the Box-Jenkins model, as the optimal values consistent with this model.

6.4 The Hourly Data

These data provided a quite different challenge. The values selected were for the ten days numbered 201 to 210 of 1985, i.e., 240 values in all. The missing values and the inconsistent day-to-day variability of the data

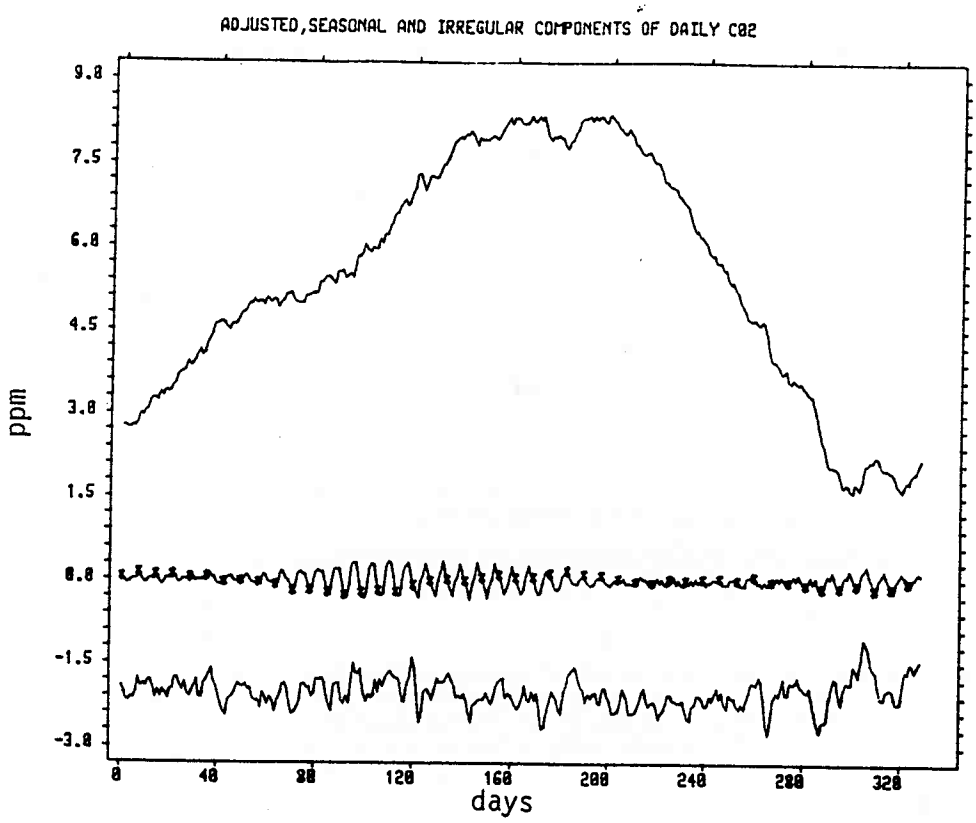

Figure 2. Components of daily CO_2 data. Top curve is adjusted trend, middle curve is the oscillatory component, bottom curve is the irregular component. See text for explanation of vertical scale.

make the kind of analysis previously applied to the monthly and daily data somewhat less appropriate. In any case this example provided an opportunity to try out some robust smoothing procedures which have been developed.

These procedures are based on a structural model, that the data consist of the sum of a signal and a noise component. The signal is defined to be a process of (small) independent increments, the noise to be independent errors. When the increment and error distributions are both Normal, the optimal two-sided filter for the signal is the two-sided exponentially weighted moving average, which can be implemented by very simple recursive formulae. What we have attempted is to use a numerical procedure to update the conditional distribution of the signal on a forward pass, followed by a reverse pass to produce what is termed the smoothed estimate, conditional on all the data. This requires convolution formulae to be evaluated at each time point, and by doing this numerically rather than analytically, distributions other than the Normal may be specified. Commonly we simply extend the Normal distribution to have long uniform tails. In the case of the error distribution, this means that the result is insensitive to outliers or extreme values. We also extend the increment distribution to have uniform tails, but of very low density. To some extent this is to avoid a numerical problem of underflow causing a joint distribution of zero density--rather difficult to normalize as a conditional distribution! A recent presentation of this kind of approach is by Kitagawa (1987). He uses very powerful computing facilities, but our results were obtained using a desk-top machine.

The property of the resulting filter is that provided the data are well behaved, i.e., look as if the errors and increments are normal, then effectively a two-sided exponential smoother results. However, extreme values are ignored, and there is also a small probability of a sudden jump in level of the smoothed signal--quite impossible with a standard filter. Clearly some tuning is necessary to obtain acceptable results, and we are learning how to do this. The smoothed signal is also the expected value computed from the conditional distribution, and is not always an adequate representation, since the distribution may be bimodal over a stretch of time.

In Fig. 3 the upper frame shows the data with smoothed line superimposed, and the lower frame shows the smoothed signal in greater detail.

Note first that the missing values were replaced by scattered values at 370, 380, and 390. They could be easily dealt with in a more logical manner--this device just avoids a sudden jump in the signal to 999.99 (the missing value indicator), and they are treated as outliers. Note also that at the left of the graph, the smoothed signal moves erratically, due to the uncertainty at this point--there is no consistent clear level. At similar points later in the series the smoothed signal is well tied down on both sides, so stretches across gaps without difficulty.

The smoothed signal shows a slow decline in the level. The span of the filter over "normal" sections is about 4 days. It may be that standard two-sided filters with this span would produce similar results. The kind of extension to be considered for the procedure presented here, is to use an

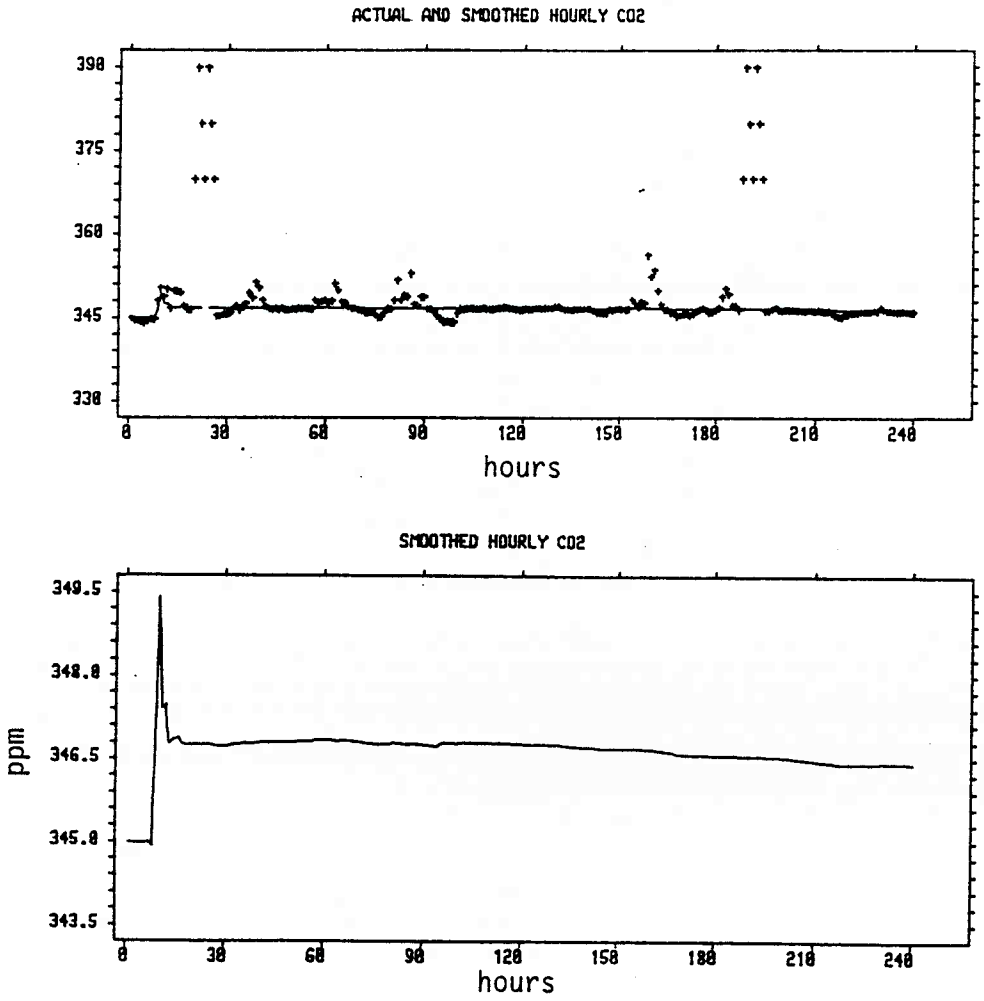

Figure 3. Hourly CO_2 data. Top frame -- actual data; bottom frame -- smoothed signal.

asymmetric error distribution to allow for the fact that large positive excursions are more common than negative ones.

Clearly, in this context much more thought could be given to specifying the kind of statistical structure that is present in the data. However, the numerical filter is a relatively simple and automatic consequence of this specification, although the computational demands are much heavier than the two-term exponential smoother.

6.5 Conclusions

The philosophy of the approach to this analysis is to ensure that an adequate statistical model has been constructed for the data, and the data processing procedures are a consequence of this modelling exercise. The objectives of the data analysis will evidently affect the procedures used. Some data decompositions have been carried out which might be expected both to produce a useful processed signal, and an understanding of the sources of variation in the data.

6.6 References

Box, G.E.P., and G.M. Jenkins, 1976. Time Series Analysis: Forecasting and Control, 2nd Edition, Holden-Day, San Francisco.

Burman, J.P., 1980. Seasonal adjustment by signal extraction. J. Royal Statist. Soc. A, 143, 321-337.

Kitagawa, G., 1987. Non-gaussian state space modelling of non-stationary time series. J. Amer. Stat. Assoc., 82, 1032-1041.

Tunnicliffe Wilson, G., and J. Belcher, 198 . Modeling, estimation and testing of time-dependent climatological parameters. K. Cehak (Ed), Third International Conference on Statistical Climatology, 112-118. Vienna, Osterreichische Gesellchaft fur Meteorologie.

7. TIME SERIES STUDIES OF CO_2 DATA IN THE DIVISION OF ATMOSPHERIC RESEARCH, CSIRO, AUSTRALIA

I.G. Enting

CSIRO, Division of Atmospheric Research
Private Bag No. 1, Mordialloc
Victoria 3195, Australia

Time series studies in the DAR CO_2 program have concentrated on the longer-term variability in the CO_2 record comprising periods of one month or more. The primary basis of our analyses has been the decomposition of the data into long-term changes, seasonal cycles plus other, less structured, residual variation. This decomposition has been performed in a variety of ways. The motivation for such analyses is the attempt to understand the atmospheric CO_2 budget by assuming that different types of processes influence the atmospheric CO_2 concentrations on characteristically different time-scales. In some studies, we have carried this approach further by decomposing the longer-term changes into an overall growth with interannual variations superimposed.

Our studies in DAR have involved the application of a range of techniques to accomplish these aims. In parallel with these applications we have conducted a number of theoretical investigations, by mathematical analysis, numerical simulation or simply by reviewing the statistical literature in order to assess the applicability of the techniques that we have been using.

Along with many other groups, some of our early studies represented the longer-term changes using cubic splines with nodes fixed at 12-month intervals. A numerical study by Enting (1986) revealed the problems that can arise when using this technique. The results of such a fixed-node spline fit to a sinusoid with a period of twice the node spacing will depend dramatically on the phase. If the nodes of the spline coincide with the zeroes of the sinusoid then the least-squares spline fit is zero except for end effects. If the nodes of the spline coincide with the peaks and troughs of the sinusoid then the least-squares spline fit reproduces the sinusoid almost exactly. Spline fits to sinusoids whose periods are approximately twice the node spacing exhibit "beating" in time as the extrema and zeroes successively tend to coincide with the nodes of the splines.

The alternative form of spline fitting that has often been used to extract long-term variations from CO_2 records is the Reinsch spline or smoothing spline. Enting (1987) reviewed a number of recent results from the statistical literature (see in particular Silverman, 1984) and made some suggestions for the appropriate use of these splines in analysis of CO_2 data and similar records. The relevant points are

i. The notation used in the analysis of these splines differs from author to author. Enting (1987a) used the notation of de Boor (1978) so that

the notation used in the analysis is consistent with the computer programs given by de Boor.

ii. Smoothing splines can be used both to smooth a time series and to interpolate a time series with non-uniform data spacing.

iii. Except near the ends of the records, the smoothing spline is equivalent to a linear digital filter.

iv. There are asymptotic expressions for the effective filtering characteristics. In particular, the response function is

$$H(\theta) = 1/(1 + (\theta/\theta_{0.5})^4) \tag{1}$$

where the 50% cutoff frequency $\theta_{0.5}$ is related to the mean data spacing Δt and the stiffness parameter λ appearing in the spline formalism by

$$\theta_{0.5} = (\lambda\, \Delta t)^{-1/4} \tag{2}$$

In the time domain, spline smoothing is equivalent (asymptotically) to smoothing with the filter coefficients:

$$c_k = \theta_{0.5}\, \Delta t\, K(\theta_{0.5}\, \Delta t) \tag{3}$$

where

$$K(x) = \exp(-|x|/\sqrt{2})\, \sin(|x|/\sqrt{2} + \pi/4) / 2 \tag{4}$$

v. End effects become important at a distance of order $1/\theta_{0.5}$ from the ends of the records. The asymptotic effective smoothing kernel (4) is modified by adding its reflection about the end-point.

vi. The filtering characteristics are relatively insensitive to local variation in the data spacing.

It was suggested by Enting (1987a) that the main virtue of smoothing splines was their ability to interpolate unequally spaced data sets. If a low-pass filter is required for equally-spaced data then a sharper transition band can be obtained using a specifically designed digital filter. To illustrate this point, Fig. 1 compares the response functions of (a) a smoothing spline (b) a low-pass filter constructed as specified by Bloomfield (1976) (c) the response function, $\exp(-\ln 2(\theta/\theta_{0.5})^4)$, used for filtering in some NOAA/GMCC studies.

Enting (1987a) compared the actual effective smoothing kernel to expression (3) for a case corresponding to removing variations with periods below about 2 years from monthly mean data. The agreement was excellent. However in the discussion at the Hilo time-series meeting it was noted that the expressions (1) and (3) above are only asymptotically valid. In practice, departures from the formulae (1) and (3) can be expected if:

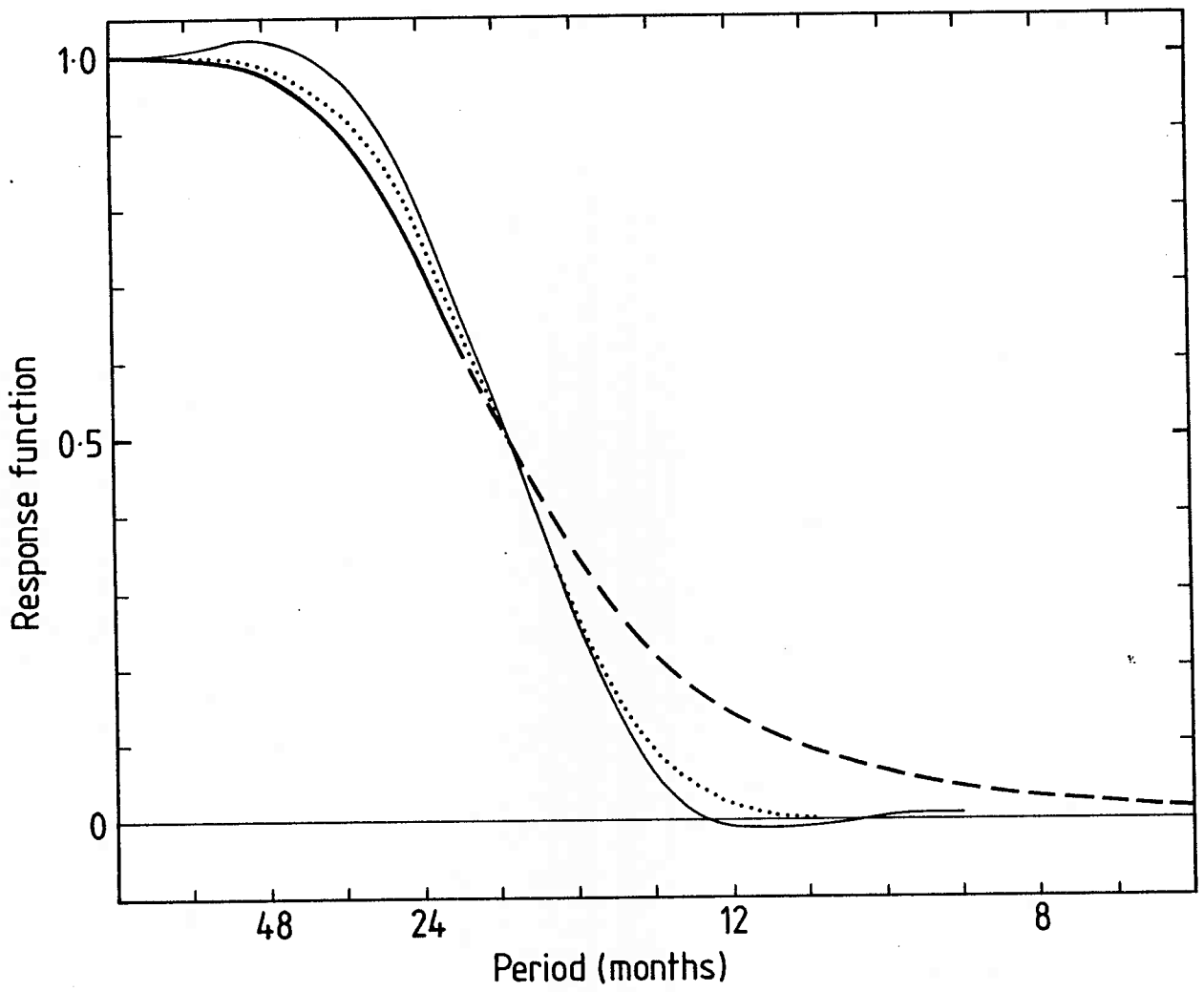

Figure 1. Comparison of filter transfer function $H(\theta)$ for filters with a common 50% cutoff point. (a) dashed curve, smoothing spline ($H(\theta) = 1/(1+(\theta/\theta_{0.5})^4$); (b) solid curve, low-pass filter of type described by Bloomfield (1976); (c) dotted curve, response function exp ($-\ln 2(\theta/\theta_{0.5})^4$) used in some NOAA/GMCC studies, applied in the frequency domain.

i. the cutoff period $T_{0.5} = 2\pi / \theta_{0.5} = 2\pi (\lambda \Delta t)^{1/4}$ becomes small enough to be comparable to Δt i.e., $\lambda \leq (\Delta t)^3/16\pi^4$; or

ii. the cutoff period becomes so large that all of the record is subject to end effects from both ends simultaneously i.e., $\lambda \leq (t_N - t_1)^4/\Delta t$.

A number of further cases were calculated to explore the breakdown of the asymptotic expressions. The calculations used 100 data points at unit spacing and calculated the effective kernels by fitting spline curves to data consisting of a single "1" embedded in a set of "0" values. Figure 2 shows the effective smoothing kernel (solid curve) for λ values 10^6, 10^4, and 10^2 compared to the asymptotic expression (3) (dashed curve, not shown in cases where it is indistinguishable from the solid curve). It will be seen that the asymptotic expression breaks down for the case of very smooth splines (i.e., large λ values). For the smaller λ values shown, the agreement is excellent. For $\lambda = 1$ in which case only 3 points contribute 85% of the weight of the kernel, the actual kernel agreed with expression (3) to within 0.001. For $\lambda = 10^{-2}$ the agreement broke down with the central weight being 0.891 compared to the theoretical prediction of 1.118. This case corresponds to $T_{0.5} = 1.986\,\Delta t$.

The accuracy of the correction for end effects was also tested. Figure 3 shows the effective smoothing kernel (solid curve) for constructing estimates at various points (shown by the arrows) near the end of the record, using the case of intermediate smoothing ($\lambda = 10^4$). The theoretical curves are shown as dashed; it is clear that in this case the agreement is not particularly good.

As noted above, we have preferred to use specifically designed digital filters for extracting trends and cycles from CO_2 records involving equally-spaced data. Furthermore we have preferred to carry out the filtering operations in the time domain rather than in the frequency domain. Working in the time domain removes any questions concerning the way in which the ends of the record are being treated. If our estimated signal y is obtained from the time series z_k by the filtering operation:

$$y_k = \sum_{j=-M}^{N} c_j z_{k-j} \qquad (5)$$

then clearly we cannot obtain estimates until N steps after the beginning of the record nor can we obtain estimates within M steps of the end of the record. In practice similar restrictions will apply to estimates obtained by using Fourier transform techniques for filtering but a precise characterization of the end effects seems more difficult to obtain.

In most of our studies we have used symmetric filter (M=N in expression 5 above). The low-pass filters have been of the form given by Bloomfield (1976). Band-pass filters have been constructed as the difference of two low-pass filters.

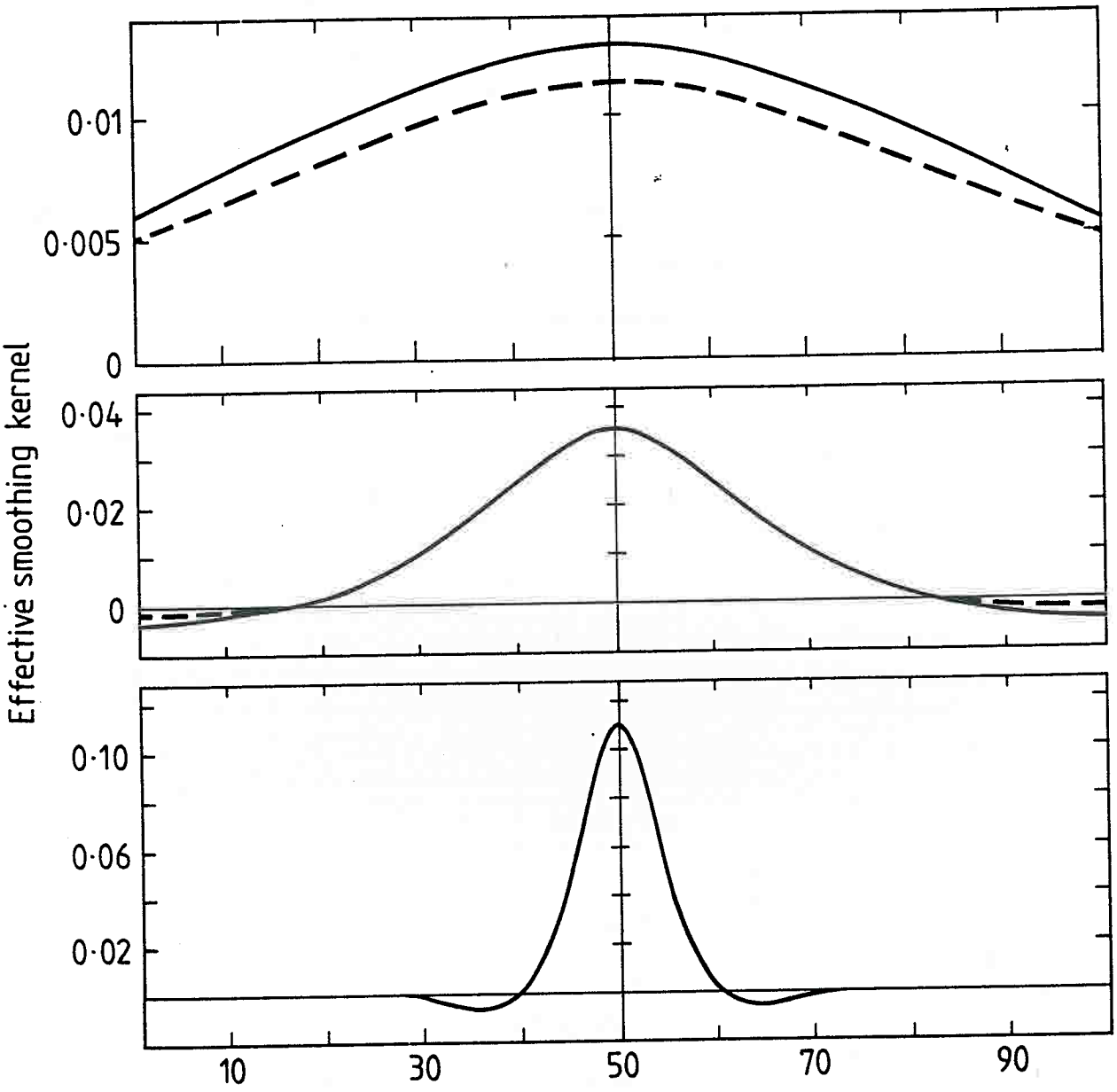

Figure 2. Comparison of actual effective smoothing kernels for spline fits (solid curves) with asymptotic expressions (dashed curves, not shown if indistinguishable from solid curves) for $\lambda = 10^6, 10^4, 10^2$ from top to bottom, chosen as illustrations of the limits of validity of the asymptotic expressions.

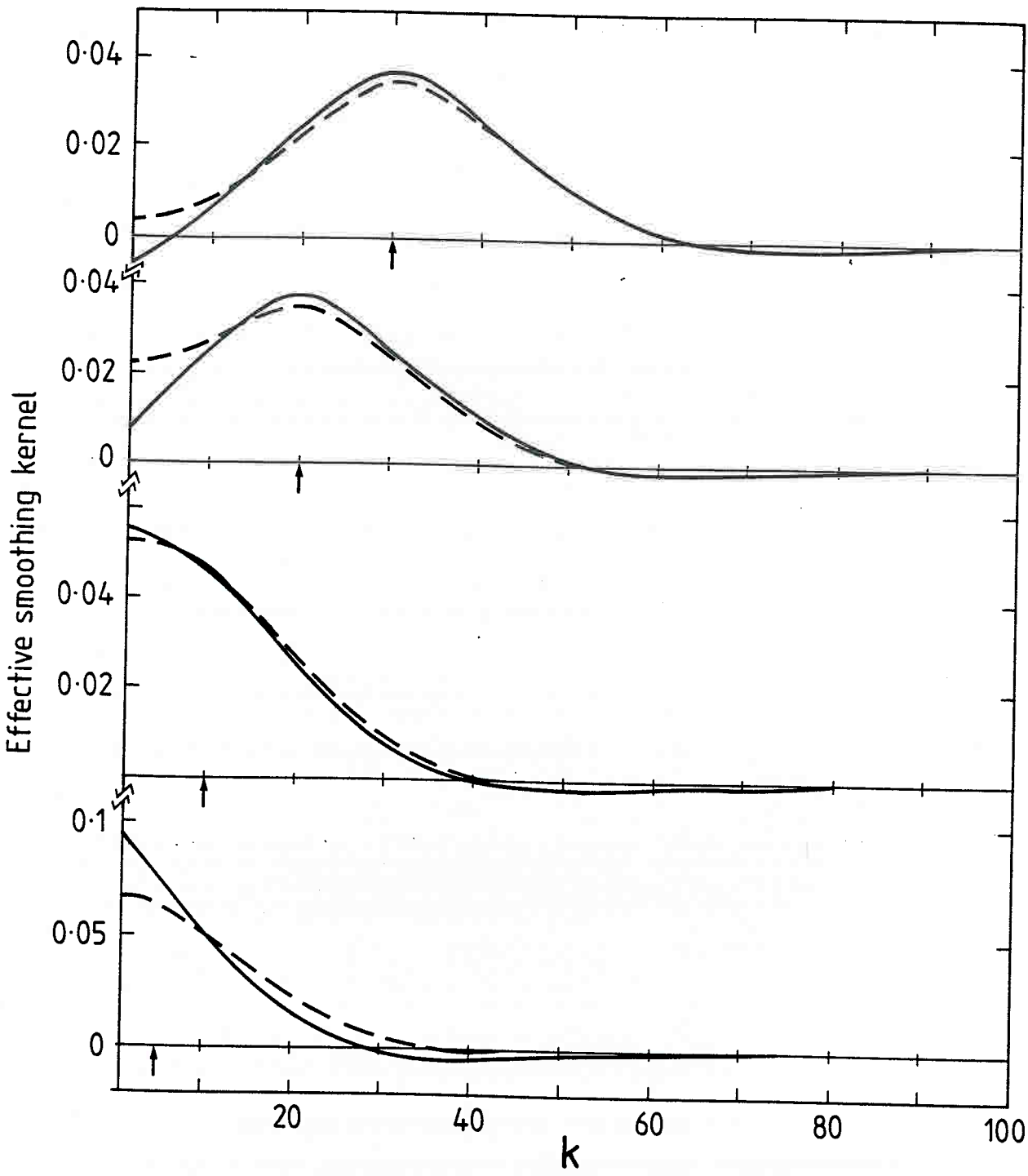

Figure 3. End effects in spline smoothing. Comparison of actual effective smoothing kernels (solid curves) with theoretical values (dashed curves) obtained by reflecting expression (3) about the end point. In each case the curves give the kernel for constructing the smoothed value at the point shown by the arrow.

One aspect of CO_2 data that has been the subject of a number of our studies is the amplitude of the seasonal cycle. In this regard it should be noted that there are differences in definition concerning what aspects of the CO_2 variation should be regarded as seasonal. For example the decomposition of CO_2 records by Cleveland et al. (1983) into "trend", "seasonal" and "irregular" allows only very smooth changes in the amplitude of the cycle from one year to the next; a very high degree of repeatability is required for the "seasonal" component. In contrast, many of the other studies described below allow much more variation in the "seasonal" component from one year to the next. It seems that in these cases "seasonal" has been defined implicitly as "apparently driven by seasonal variations in the environment". Thus it is accepted that the "seasonal" component is allowed to show an interannual variability comparable to that exhibited by climatic variations.

An important early DAR study of variations in seasonal cycles was that of Pearman and Hyson (1981). They used a regression analysis to search for evidence of an increase in the amplitude of the seasonal cycle of CO_2 at Mauna Loa. In this case only smooth (i.e. linear) variations in the amplitude were taken into consideration. In a subsequent study (Thompson et al. 1986) we used the technique of complex demodulation (Bloomfield, 1976) to obtain estimates of the amplitude and phase of the "seasonal cycle" of CO_2 at Mauna Loa, the South Pole, Barrow, Cape Kumukahi, Samoa and Bass Strait. No obvious coherent patterns of variation were observed. The method of complex demodulation regards the "seasonal" component as being of the form $y_s(t) = A(t)\cos(\theta t + \phi(t))$ where the amplitude, $A(t)$ and the phase $\phi(t)$ are regarded as being slowly varying functions of time. Enting (1987b) noted that the estimate $y_s(t)$ obtained by using the estimates $A(t)$, $\phi(t)$ is equivalent to the result of passing the original time series through a band pass filter. The band-width of the filter (which will have an inverse relation to the length of the filter) determines the degree of smoothness that is imposed on the interannual changes in the amplitude and phase of the "seasonal" component.

This digital filtering approach was used by Enting (1987c) to study seasonal variations at the South Pole, Cape Grim and Bass Strait. For such southern hemisphere sites, a considerable degree of interannual variation is expected because the seasonal cycle results from the algebraic sum of a number of partly competing contributions. Thus small changes in a single contribution could give proportionally larger changes in the resultant. The study failed to find any coherent pattern in the interannual variation of the seasonal cycles of these three southern hemisphere sites. The digital filtering analysis of the Mauna Loa seasonal cycle (Enting, 1987b) found a pattern of amplitude changes similar to that found by the complex demodulation study of Thompson et al. (1986) and also by the detrending approach of Bacastow et al. (1981, 1985). The claim by Enting (1987b) to have detected a characteristic biotic signature in the pattern of amplitude changes requires qualification. The scatter plot comparison presented by Enting (1987b) gives a biased comparison between summers and winters by virtue of the filter characteristics. An unbiased comparison still reveals the same characteristic signature for the first half of the Mauna Loa record (see Enting and Manning, this volume).

As noted above when we have used digital filters we have generally restricted ourselves to symmetric filters (i.e., $c_k = c_{-k}$). In some cases it is desirable to use asymmetric filters so as to minimize the loss of data from one or other end of the record. Enting and Robbins (1988) have considered the design of such filters according to the Wiener minimum mean-square-error criterion. As an example of the results of our study, Fig. 4 shows the mean-square error for a band-pass filter, designed to extract the "seasonal" component from a time series of monthly values in which there was a "seasonal" component, a white noise contribution, a band of low-frequencies and a quadratic polynomial trend that was required to be removed exactly. For a filter with coefficients c_j for $-M \leq j \leq 24$, Fig. 4 shows the mean square error as a function of M. It will be seen that quite skewed filters e.g., M = 6 give little degradation in the quality of the estimates but that smaller values of M (i.e., estimates required in closer to real time) involve significantly larger errors. These errors are even more serious for the case of Kalman filters which generally involve the M = 0 case which is not shown on Fig. 3.

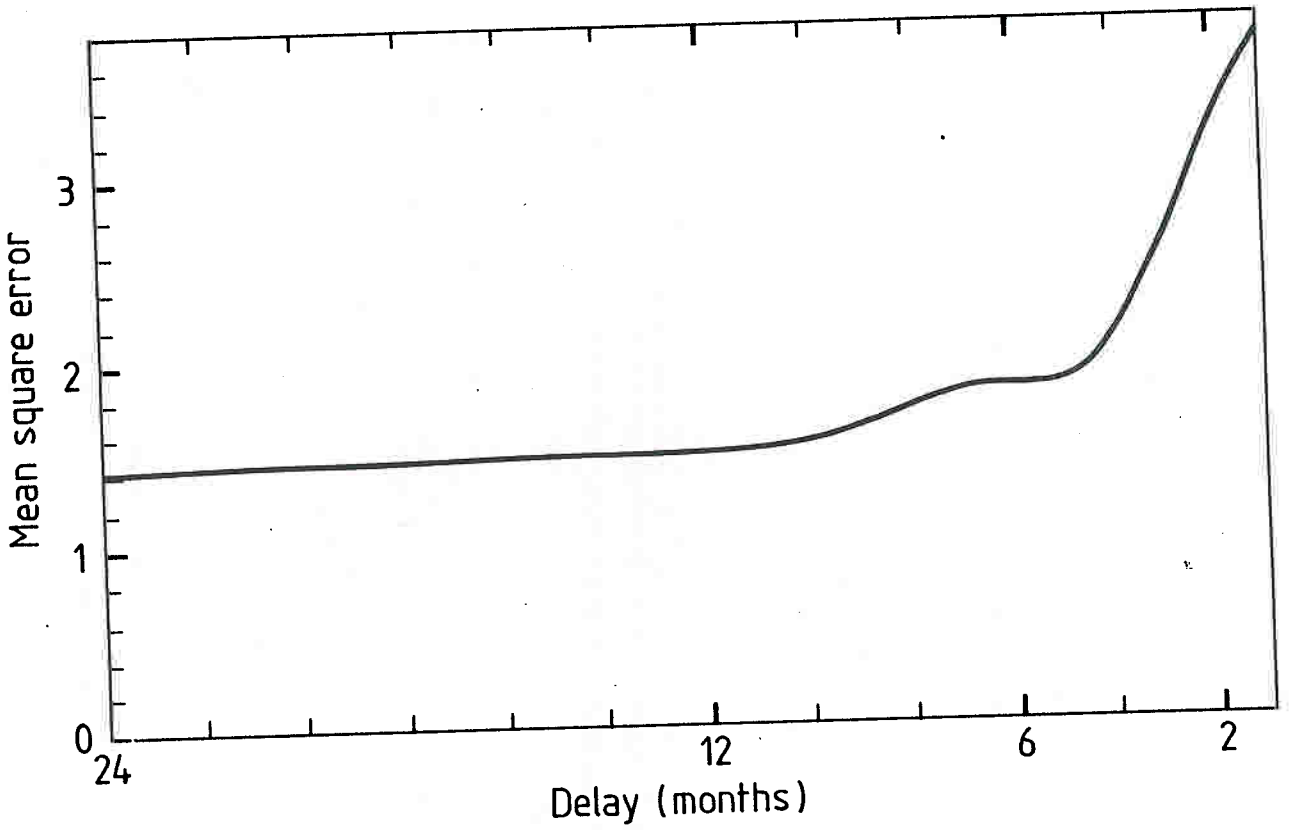

Figure 4. Mean square error (arbitrary units) for asymmetric filters with coefficients c_j for $-M \leq j \leq 24$ as function of M. Estimates are for band-pass filter designed to remove quadratic polynomial exactly while performing optical extraction of band of frequencies around 1 cycle/year in presence of low frequencies of similar power and white noise of 10% of that power.

References

Bacastow, R.B., C.D. Keeling, and T.P. Whorf, 1981. Seasonal amplitude increase in atmospheric CO_2 concentration at Mauna Loa, Hawaii. Papers presented at WMO/ICSU/UNEP Conference on Analysis and Interpretation of Atmospheric CO_2 data, WMO Publication WCP-14, pp 169-176. World Meteorological Organization, Geneva.

Bacastow, R.B., C.D. Keeling, and T.P. Whorf, 1985. Seasonal amplitude increase in atmospheric CO_2 concentration at Mauna Loa, Hawaii, 1959-1982. J. Geophys. Res., 90, 10529-10540.

Bloomfield, P., 1976. Fourier Analysis of Time Series: An introduction. John Wiley, New York, 258 pp.

Cleveland, W.S., A.E. Freeny, and T.E. Graedel, 1983. The seasonal component of atmospheric CO_2: Information from new approaches to the decomposition of seasonal time series. J. Geophys. Res., 88, 10934-10946.

de Boor, C., 1978. A Practical Guide to Splines. Springer-Verlag, New York, 392 pp.

Enting, I.G., 1986. Potential problems with the use of least squares spline fits to filter CO_2 data. J. Geophys. Res., 91, 6668-6670.

Enting, I.G., 1987a. On the use of smoothing splines to filter CO_2 data. J. Geophys. Res., 91, 10977-10984.

Enting, I.G., 1987b. The interannual variation in the seasonal cycle of carbon dioxide concentration at Mauna Loa. J. Geophys. Res., 92, 5497-5504.

Enting, I.G., 1987c. The seasonal cycle of atmospheric CO_2 in the Australian region: A signal processing analysis. Baseline 85, B.W. Forgan and P.J. Fraser (Eds.), Dept. of Science and CSIRO, pp. 19-25.

Enting, I.G., and F.J. Robbins, 1989. Asymmetric filters for analyzing time series of atmospheric constituent data. Tellus, 41A, 109-114.

Enting, I.G., and M.R. Manning, 1989. The seasonal cycle of CO_2 at Mauna Loa: A re-analysis of a digital filtering study (this volume).

Pearman, G.I., and P. Hyson, 1981. The annual variation of atmospheric CO_2 concentration observed in the northern hemisphere. J. Geophys. Res., 86, 9839-9843.

Silverman, B.W., 1984: Spline smoothing: The equivalent variable kernel method. Annals Statistician, 12, 898-916.

Thompson, M.L., I.G. Enting, G.I. Pearman, and P. Hyson, 1986: Interannual variation of atmospheric CO_2 concentration. J. Atmos. Chem., 4, 125-155.

8. SOME COMMENTS ON KALMAN FILTERING OF CO_2 DATA

I.G. Enting

CSIRO, Division of Atmospheric Research
Private Bag No. 1, Mordialloc
Victoria 3195, Australia

8.1 Introduction

Kalman filtering has recently been applied to the analysis of CO_2 data in a series of papers by Mulholland and co-workers (Surendran and Mulholland, 1986, 1987; Mulholland et al., 1987). Because of the recent introduction of this technique into CO_2 studies it seems worthwhile listing some of the advantages and disadvantages of the approach.

The principle advantages of the technique are

i. it can handle non-stationary processes. This is particularly useful in observational programs where ongoing refinements lead to a reduction in "noise" levels;

ii. the recursive form of the estimation is computationally compact;

iii. estimation starts from the beginning of the record with the variance of the estimates decreasing as more data become available;

iv. estimates are available for times right through to the end of the record - indeed the formalism also produces single-step predictions as part of the estimation procedure.

The disadvantages of the Kalman filtering approach are:

i. the production of immediate estimates as soon as each data point becomes available leads to a higher mean-square error in the estimates that would be the case if lagged estimates were used. (For further details see the account of CSIRO time studies, Enting, this volume, and the study by Enting and Robbins, 1988, on the error vs lag relation for asymmetric filters.) The modification of the Kalman filter algorithm to produce lagged estimates is relatively complex.

ii. the formalism requires a state-space model of the process and such a model may be difficult to construct.

iii. the state-space representation may have many more components than seem meaningful because of the requirement for a first-order autoregressive model.

8.2 Modelling

The state-space model required for Kalman filtering is best understood within the context of the modelling spectrum defined by Karplus (1977). This spectrum ran from black-box models of gross responses through to fully-detailed, generally deterministic, white-box models. The position on the spectrum was determined by the amount of deductive (as opposed to inductive) information embodied in the models. Enting (1987) gave examples of models of atmospheric CO_2 concentration from various points along the modelling spectrum. The examples were (from darkest to lightest):

i. statistical curve fitting, e.g. fitting an exponential curve through the Mauna Loa CO_2 data.

ii. assuming a constant airborne fraction i.e. a fixed proportion of fossil carbon releases remain in the atmosphere.

iii. treating the atmospheric response to CO_2 releases in terms of a response function (e.g. Oeschger and Heimann 1983; Enting and Mansbridge 1987a,b).

iv. box models of the global carbon cycle.

v. two-dimensional parameterized models of atmospheric transport.

vi. three-dimensional atmospheric transport models based on general circulation models.

Within this spectrum the state-space models used in Kalman filtering will generally lie somewhere in between simple statistical curve fitting and the response function formalisms. An example of how autoregressive models can be developed to provide increasing explanatory power (i.e. move towards the lighter end of the spectrum) is given by the work of Tiao et al. (1976). They sought to explain O_3 levels (daily maxima of hourly averages) in Los Angeles. Their initial model had a strong autoregressive component with each day's O_3 level depending significantly on the previous day's level, even though the O_3 peaks largely dissipated overnight. The autoregressive model described the statistical characteristics of the O_3 record even though there was no mechanism by which one day's O_3 level could influence levels for the next day. In actual fact the autoregressive component was acting as a proxy variable for meteorological persistence. Tiao et al. produced two additional regression models involving increasing numbers of relevant environmental variables. As the amount of additional information was increased, the estimated O_3 autoregressive contribution decreased and in addition the models explained more of the variability in the O_3 records.

8.3 Some Models of Atmospheric CO_2 Increase

The most prominent feature of the records of atmospheric CO_2 concentration over the last three decades is the overall increase. This

increase can be analyzed using Kalman filtering and we present some examples to illustrate this. Surendran and Mulholland (1986, 1987) model an exponential growth in CO_2 concentrations by regarding this exponential behavior as an intrinsic characteristic of the carbon cycle. The models presented here treat the CO_2 increase as being due to inputs of fossil carbon. In terms of the model spectrum described above, the present models represent a slightly lighter shade of grey than the models used by Surendran and Mulholland.

The general form of the Kalman filtering formalism allows the state (and observations) to be vector quantities with the observations being noise (the vector $\underline{v}(k)$) plus some projection (specified by a matrix $\underline{H}$) of the state, i.e.

$$\underline{z}(k) = \underline{H}(k)\ \underline{x}(k) + \underline{v}(k) \qquad (1)$$

The formalism requires the model describing the time-evolution of the state to be of the form:

$$\underline{x}(k+1) = \underline{F}(k)\ \underline{x}(k) + \underline{G}(k)\ \underline{w}(k) + Q(k)\ \underline{u}(k) \qquad (2)$$

where $\underline{F}$ is the autoregressive evolution matrix, $\underline{w}$ is a white-noise contribution, $\underline{G}$ describes the influence of $\underline{w}$ on the state, $\underline{u}$ is a forcing term and $\underline{Q}$ defines the influence of $\underline{u}$ on the state $\underline{x}$.

In constructing a model of the CO_2 increase, we concern ourselves with the excess (above pre-industrial) CO_2 level in the atmosphere due to releases of CO_2 from fossil carbon. For year k, this quantity is denoted x(k) and it is our (single-component) state vector for the models. The observed data z(k) are the annual mean CO_2 concentrations from Mauna Loa (Keeling, 1986) with the pre-industrial value subtracted.

We assume the observations represent the excess CO_2 plus a single-component noise term v(k), i.e.

$$z(k) = x(k) + v(k) \qquad (3)$$

R, the variance of the noise term v(k) is taken as 0.25 $(ppmv)^2$ for all times k. The models presented here will not consider any random forcing, $\underline{w}$, and so $\underline{G}$ and $\underline{w}$ are not required. Such forcing would probably play an important part in any refinements to these models to represent phenomena such as ENSO-induced variations. The deterministic forcing is given by $\underline{u}(k)$ which in these models is a single component vector giving the fossil carbon release for year k+1. This means that the 1970 fossil carbon release, (i.e. the integrated release from January to December) contributes to the 1970 annual mean which would be more closely tied to July to June integration. We justify this on the grounds that Mauna Loa is a northern hemisphere site and so its CO_2 variations should lead global averages by some months. The preliminary and illustrative nature of the present studies does not justify any attempt to refine this aspect of the models.

The first model that we consider is based on the assumption that a constant proportion of fossil carbon remains in the atmosphere for all time once it is injected. We take this proportion as being 0.55. The evolution model thus has

$$F = 1$$
$$Q = 0.471 \times 0.55 = 0.259$$

where the factor 0.471 converts Gt of carbon (the units of u(k)) to ppmv (the units of x and z). Using F = 1 reflects the assumption that once the fixed proportion of CO_2 is assumed to have gone into the oceans, the residue is assumed to remain in the atmosphere.

The other numerical values required are an initial value for x (which is taken as 30 ppmv) and a variance for this initial value (which we take as 100 $(ppmv)^2$).

The Kalman filtering formalism is then used to estimate the state vector $\underline{x}(k)$ for each k (with k = 1 corresponding to 1959) by using the $\underline{z}(j)$ for $j \leq k$. The estimation equations are given by Anderson and Moore (1979). Figure 1 shows the residuals i.e. z(k) - x(k).

The second model that we consider assigns a mean lifetime, τ, to the excess CO_2. The model for the evolution is

$$F = \exp(-1.0/\tau)$$
$$Q = 0.471$$

This expression for F assumes that an initial input of CO_2 into the atmosphere is taken up by the oceans with a characteristic time τ. Figure 2 shows the residuals z(k) - x(k) for models using τ values of 50, 55 and 60 years.

A comparison of Figs. 1 and 2 suggests that the second class of model gives a similar statistical fit to the data to that obtained using the assumption of a constant airborne fraction. This can be explored in more detail by adjusting the model parameters to minimize the summed squares of residuals. The minima are 3.35 for an airborne fraction of 0.563 and 3.84 for a lifetime τ = 52.3 years.

These comparisons suggest that, contrary to expectation, if one wants to describe the behavior of atmospheric CO_2 using a single number then it would be slightly preferable to use an airborne fraction rather than an atmospheric lifetime. However the plots of the residuals make it clear that both models are inadequate and that there are systematic long-term variations in the CO_2 record on decadal time-scales. Surendran and Mulholland (1987) model these variations using a sinusoid with a period of 137.5 months but note that the record is too short to support an earlier claim by Rust et al. (1979) that this periodicity reflects the sunspot cycle.

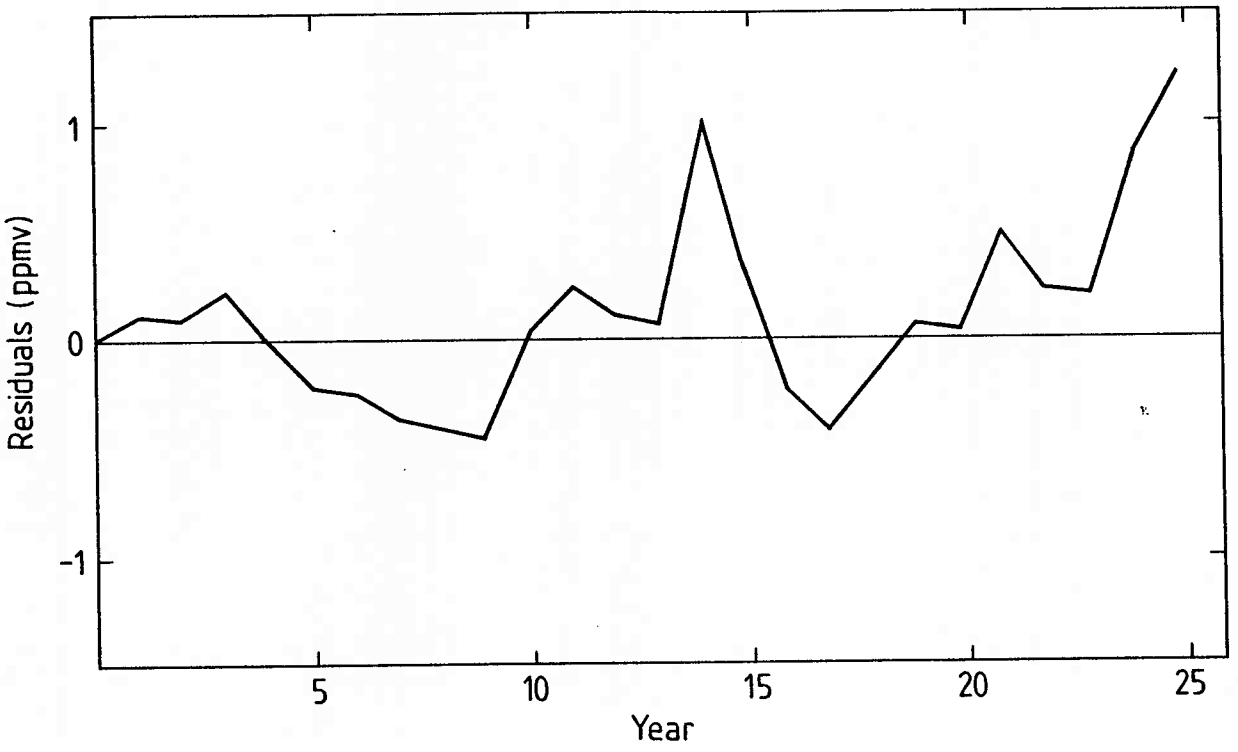

Figure 1. Residuals from state-space model based on assumption of a constant airborne CO_2 fraction of 55%.

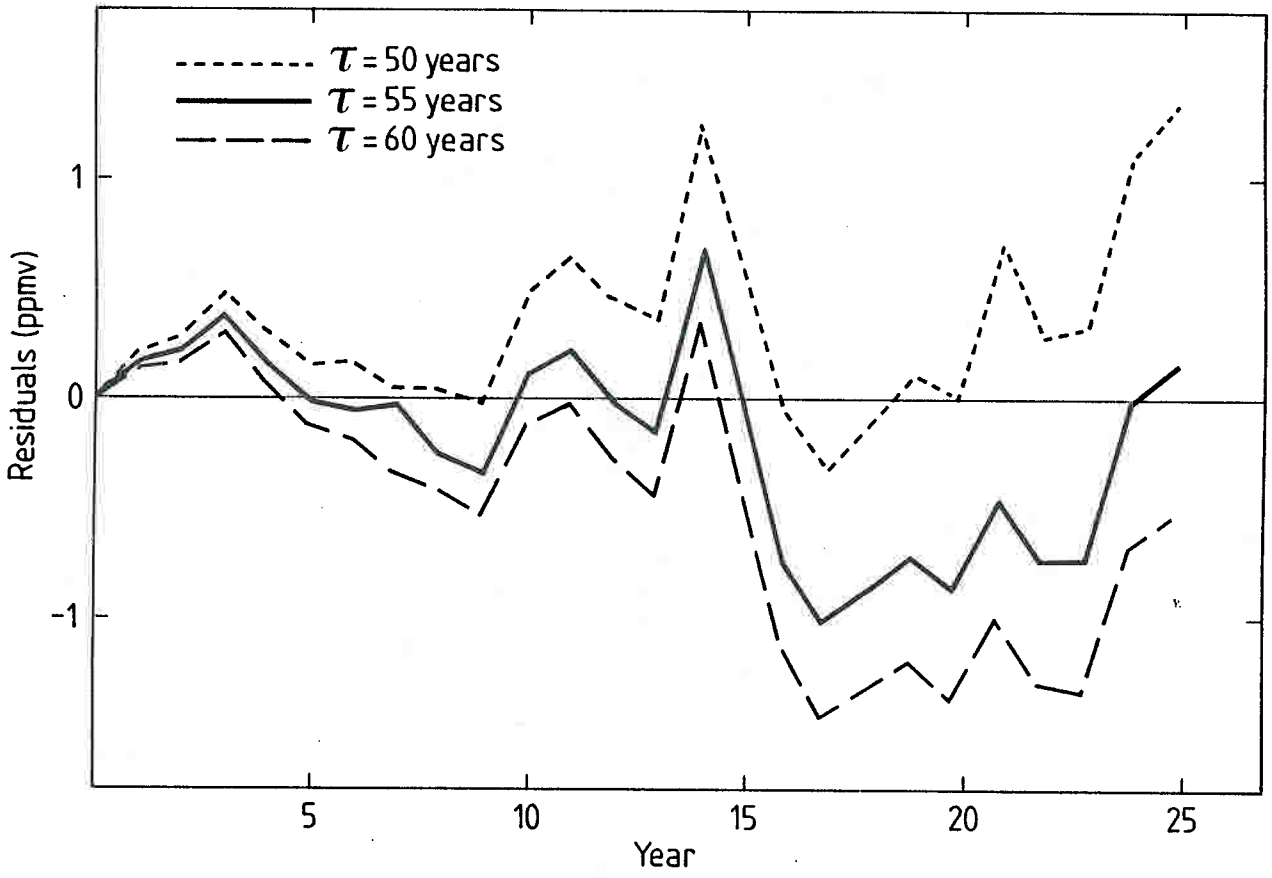

Figure 2. Residuals from state-space model assuming oceanic uptake of CO_2 is characterized by an atmospheric lifetime, τ, for τ = 50 years (dotted curve), τ = 55 years (solid curve) and τ = 60 years (dashed curve).

8.4 Separating Trends from Seasonal Cycles

As noted in the discussion of CSIRO time series studies (Enting, this volume), time series of atmospheric constituent data are often decomposed into long-term trends plus seasonal cycles (plus residuals). It is of some interest to see how this task would be accomplished by the Kalman filter described by Surendran and Mulholland (1987). The state space representation that they used involved a 4-component state vector:

$x_1(k)$ = long-term growth component at time k
$x_2(k)$ = Rate of change of x_1
$x_3(k)$ = Seasonal component at time k
$x_4(k)$ = Rate of change of x_3

with the unit time step being one month. Their original model evolution matrix was

$$F = \begin{bmatrix} 1 & [\exp(ft)-1]/f & 0 & 0 \\ 0 & \exp(ft) & 0 & 0 \\ 0 & 0 & \cos(\theta t) & -\theta \sin(\theta t) \\ 0 & 0 & -\theta \sin(\theta t) & \cos(\theta t) \end{bmatrix}$$

with $f = 0.00329$ month^{-1} (giving a 4% per annum increase in the long-term component x_1) and $\theta = \pi/6$ radians month^{-1} (so that the x_3 component has a period of 1 year), with F(t) evaluated at t = 1 month. The observed time series z(k) representing monthly mean CO_2 concentrations with a pre-industrial value of 298 ppmv subtracted was regarded as the sum of the long-term component and the seasonal cycle, i.e. $x_1 + x_3$ (plus white noise). The vector components x_2 and x_4 are auxiliary variables that define the model but which are not seen directly in the observed time series z(k). Thus in terms of equation (1),

$$\underline{H} = [1, 0, 1, 0]$$

In the absence of forcing terms u or w the Kalman filtering estimation formula reduces to

$$\underline{x}(k+1) = \underline{F}(k)\,\underline{x}(k) + \underline{L}(k+1)\,[\underline{z}(k+1) - \underline{H}(k+1)\underline{F}(k)\underline{x}(k)] \qquad (4)$$

Even if $\underline{H}$ and $\underline{F}$ (and $\underline{R}$, the variance of the noise term $\underline{v}$) are constant in time, the gain matrix, $\underline{L}$, is time dependent, reflecting the increasing amount of information that becomes available as the length of the record increases. (The expression for $\underline{L}$ is given by Anderson and Moore, 1979). However the gain matrix approaches a limiting steady state. Surendran and Mulholland (1987)

further adjusted the estimation adaptively on the basis of the residuals and produced a final estimated gain matrix of:

$$\underline{L}(\infty) = [0.586 , 0.125 , 0.0649, 0.135]'.$$

The behavior of this steady state limit can be analyzed by calculating the impulse response function i.e. the sequence $\underline{x}(k)$ when $z(k)$ consists of a single 1 at time N, surrounded by zero values at all other times. The effective response function for the Kalman filter, regarded as a filter for extracting long-term trends, is obtained by taking the discrete Fourier transform (DFT) of the estimates $x_1(N+m)$ arising from the impulse input. The effective response function describing the estimation of the seasonal components is obtained from the DFT of $x_3(N+m)$. The response functions describing the estimation of x_1 and x_3 are shown in Figs. 3 and 4 respectively. The solid curve is the real part and the dashed curve is the imaginary point in each case.

The power of the compact recursive formalism embodied in equation (4) is shown by the very sharp frequency separation that is achieved. The limitations are shown by the large imaginary components in each filter for periods near to one year. These will cause the filter to perform badly if used on data with significant interannual variability in the seasonal cycle. The response function for the extraction of long-term trends also extracts a large proportion of any variability corresponding to periods of less than one year.

8.5 Conclusions

Obviously Kalman filtering and the associated state-space modelling process are additional tools that could be useful in CO_2 studies. Given the small number of studies that have been conducted, it is too early to guess at their ultimate utility in this area. However there are two particularly encouraging features of Kalman filtering that suggest that it may be a very valuable approach. The first is the way in which deterministic information can be combined with purely statistical description of the behavior of atmospheric CO_2. The formalism makes it possible, in principle, to bridge the gap between purely statistical models based on the statistical structure of the data and process-oriented models. The second important feature of Kalman filtering is the ability to incorporate non-stationary processes. Many records of atmospheric constituent data are obviously non-stationary. These two advantages seem to be sufficiently important to outweigh the disadvantages listed in the introduction and to justify the extra effort involved in obtaining lagged estimates to reduce the variance in the filter output.

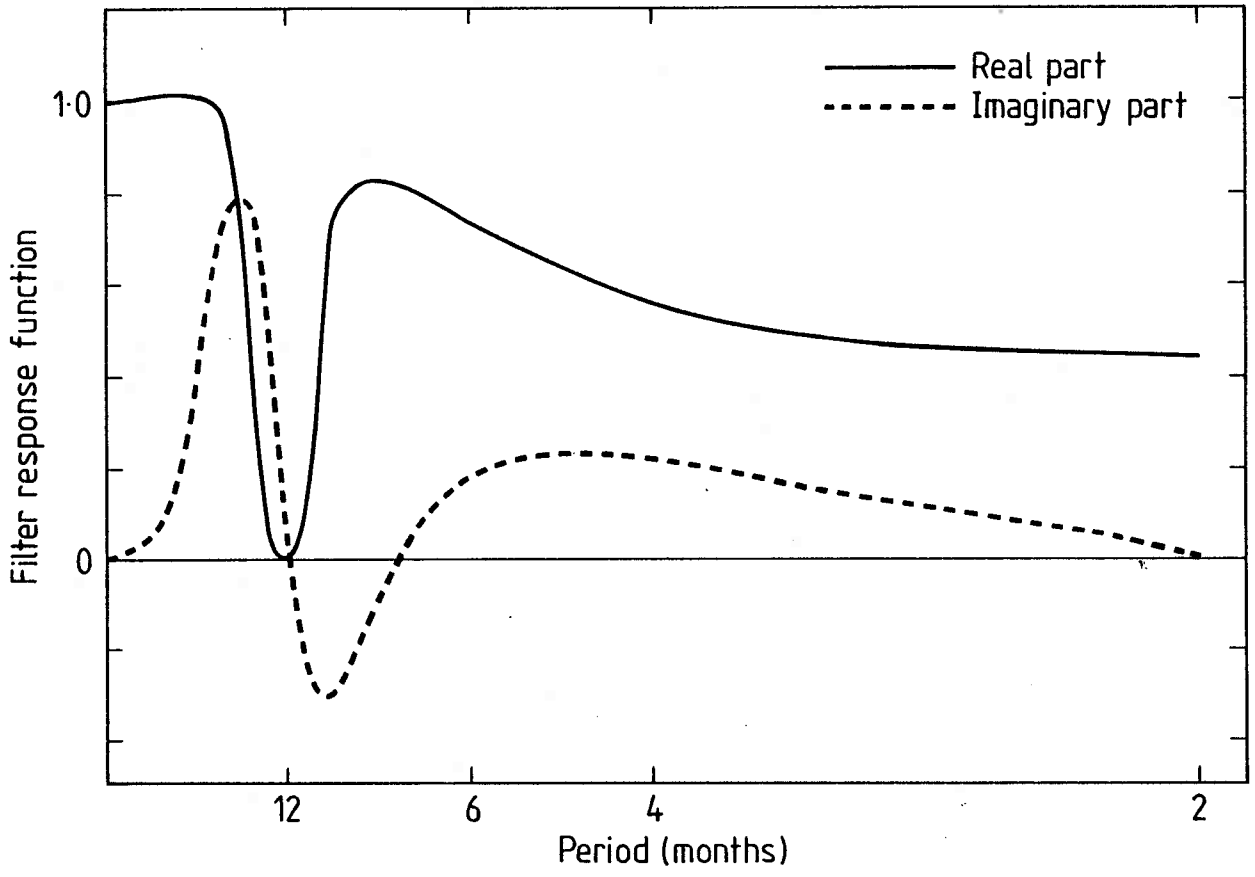

Figure 3. Filter response functions for asymptotic limit of the Kalman filter of Surendran and Mulholland (1987) used to extract long-term trends. Solid curve is real part, dashed curve is imaginary part.

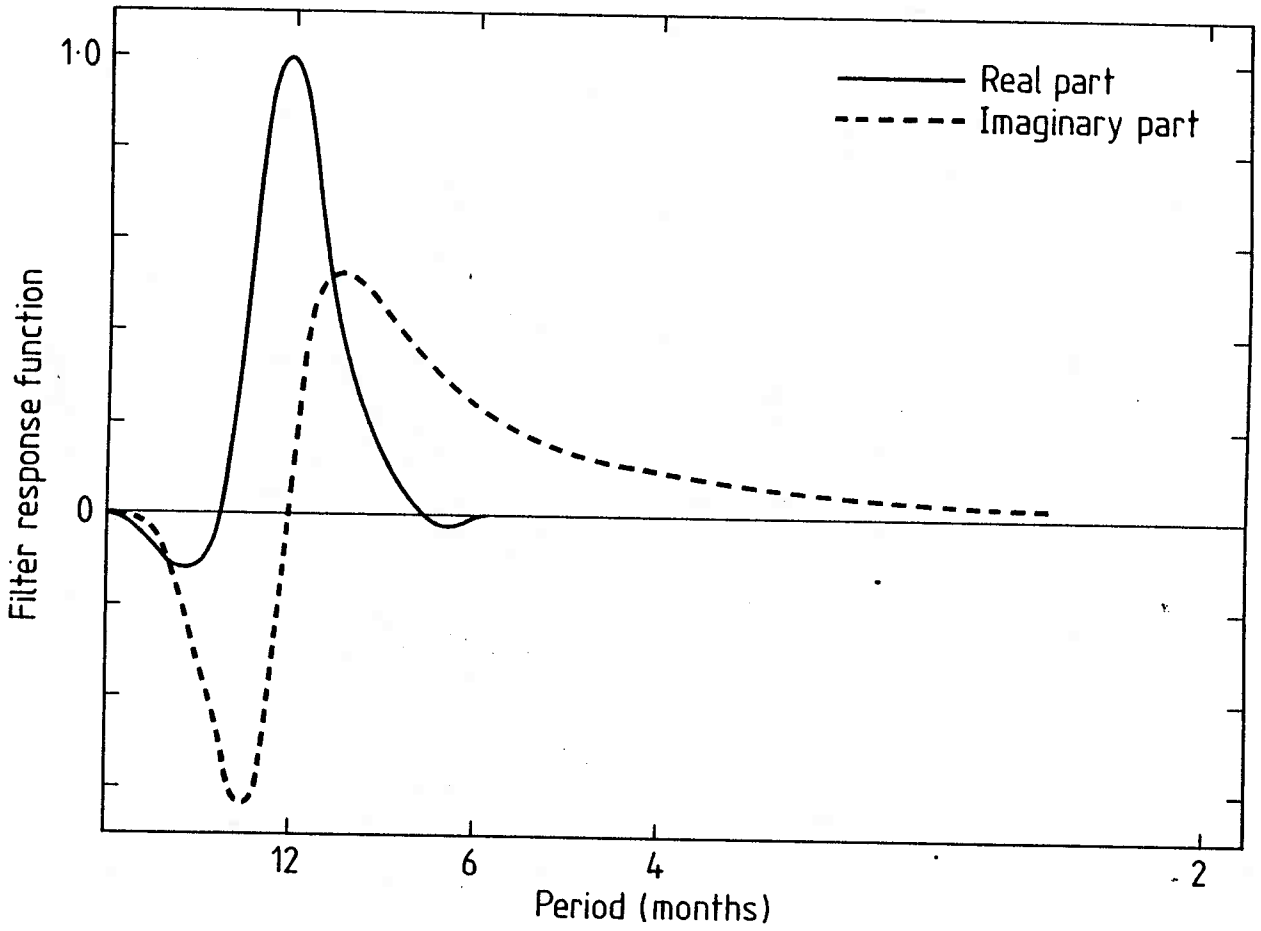

Figure 4. Filter response functions for asymptotic limit of the Kalman filter of Surendran and Mulholland (1987) used to extract estimate of seasonal component. Solid curve is real part, dashed curve is imaginary part.

8.6 References

Anderson, B.D.O., and J.B. Moore, 1979. Optimal filtering. Prentice-Hall, Englewood Cliffs, New Jersey.

Enting, I.G., 1987. A modelling spectrum for carbon cycle studies. Math. Comput. Simulation, 29, 75-85.

Enting, I.G., and J.V. Mansbridge, 1987a. The incompatibility of ice-core CO_2 data with reconstructions of biotic CO_2 sources. Tellus, 39B, 318-325.

Enting, I.G., and J.V. Mansbridge, 1987b. Inversion relations for the deconvolution of CO_2 data from ice cores. Inverse Problems, 3, L63-69.

Enting, I.G., and F.J. Robbins, 1989. Asymmetric filters for analysing time series of atmospheric constituent data. Tellus, 41A, 109-114.

Karplus, W.J., 1977. The spectrum of mathematical modelling and systems simulation. Math. Comput. Simulation, 19, 3-10.

Keeling, C.D., 1986. Atmospheric CO_2 concentrations - Mauna Loa Observatory 1958-1986. Report NDP-001/R1. Carbon Dioxide Information Center, Oak Ridge.

Mulholland, R.J., J.S. Read, and W.R. Emanuel, 1987. Asymptotic analysis of airborne fraction used to validate global carbon models. Ecological Modelling, 36, 139-152.

Oeschger, H., and M. Heimann, 1983. Uncertainties of predictions of future atmospheric CO_2 concentrations. J. Geophys. Res., 88, 1258-1262.

Rust, B.W., R.M. Rotty, and G. Marland, 1979. Inferences drawn from atmospheric CO_2 data. J. Geophys. Res., 84C, 3115-3122.

Surendran, S., and R.J. Mulholland, 1986. Estimation of atmospheric CO_2 concentration using Kalman filtering. Int. J. Systems Sci., 17, 897-909.

Surendran, S., and R.J. Mulholland, 1987. Modeling the variability in measured atmospheric CO_2 data. J. Geophys. Res., 92D, 9733-9739.

Tiao, G., M.S. Phadke, and G.E.P. Box, 1976. Some empirical models for the Los Angeles photochemical smog data. J. Air Pollut. Control Assn., 26, 485-490.

9. BACKGROUND ATMOSPHERIC CO_2 PATTERNS FROM WEEKLY FLASK SAMPLES AT BARROW, ALASKA: OPTIMAL SIGNAL RECOVERY AND ERROR ESTIMATES

Pieter P. Tans[1], Kirk W. Thoning[1],
William P. Elliott[2], and Thomas J. Conway[3]

Abstract. Atmospheric CO_2 flask sampling time series were simulated by choosing one or more hourly CO_2 concentration values per week from the continuous analyzer record of the GMCC observatory at Point Barrow, Alaska using a partially random procedure. These simulated time series were analyzed using various curve-fitting techniques. After data selection, monthly averages were calculated for each simulated flask data set. These averages were compared with the parent data set from which the simulated flasks were derived. The statistical uncertainty of a monthly flask average was estimated to be between 0.4 and 0.6 ppm with the optimal curve-fitting technique. Systematic biases of up to 1.3 ppm were present depending on the month and on the curve-fitting technique that was used. Increasing the sampling from once per week to twice per week did not significantly reduce the biases. The seasonal cycle was characterized by a fit with four harmonics. The harmonic coefficients were determined very tightly by only three years of flask data.

9.1 Introduction

The Geophysical Monitoring for Climatic Change (GMCC) program of the National Oceanic and Atmospheric Administration (NOAA) measures the concentration of atmospheric carbon dioxide (CO_2) at four observatories located at Point Barrow, Alaska (BRW); Mauna Loa, Hawaii (MLO); American Samoa (SMO); and the South Pole (SPO). At these sites nondispersive infrared (NDIR) analyzers are employed for continuous in-situ measurements of CO_2. In addition, whole-air samples are taken in glass flasks from a worldwide network of sampling sites. These samples are then shipped to the GMCC laboratory in Boulder, Colorado, where they are analyzed by an NDIR CO_2 analyzer (Komhyr et al., 1983, Komhyr et al., 1985, Conway et al., 1988).

The purpose of these measurements is not only to document the global increase of CO_2, but to obtain quantitative estimates of the globally important sources and sinks of this gas. A careful study of the spatial and temporal details of the CO_2 data can reveal the outlines of the global biogeochemical cycle of carbon. In addition, we might begin to see feedback effects on the carbon cycle due to the changing global climate. For example, the observed spatial gradients in CO_2 concentration suggest the presence of a seasonal sink of carbon in the southern oceans (Tans et al., 1989).

---

[1]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder
[2]NOAA, Air Resources Laboratory, Silver Spring, Maryland
[3]NOAA, Geophysical Monitoring for Climatic Change, Air Resources Laboratory, Boulder, Colorado

For this "global" interpretation of the measurements to be valid it is necessary that they characterize the CO_2 concentrations over large areas. The specific measurement sites have been chosen because they allow us to sample "background" air for at least a substantial fraction of the time. Our definition of background air is that it be representative of a large area, with horizontal dimensions of the order of 1000 km. In operational terms it is best to have sampling sites with steady concentrations, because sources and sinks in the vicinity of the site will tend to superimpose high frequency variability on the concentration record.

In addition to the natural variation in CO_2 concentration, statistical and systematic errors and uncertainties inherent in the observations will also lead to errors in our estimates of the global sources and sinks of CO_2. There are many possible sources of systematic errors, such as a "local" contribution to the signal, the accuracy and precision of the gas calibration standards, a possible interference (e.g. water vapor) in the analyzer, changes in the CO_2 concentration in flasks during storage (e.g. due to the oxidation of stopcock grease), analytical errors, or incorrect procedures by the sample taker. A local contribution to the signal can in most cases be removed from continuous analyzer records by selecting periods of steady CO_2 concentrations. With typically only one sample pair per week the careful selection for "background" conditions that is done with the continuous analyzers is basically impossible in the case of flask sampling. For instance, a possible presence of a diurnal cycle at the site could be partially reflected in the flask record, depending on the time of day when the samples are obtained.

Because of either errors in technique or real short-term atmospheric variability, there are always "outliers" in the flask records, data not considered representative of background conditions. Lacking sufficient supporting data besides CO_2 concentration as a function of time, we feel it is best to exclude outliers on statistical grounds (Gillette and Steele, 1983). To that end we fit a preliminary smoothed curve to the data, calculate the standard deviation of the residuals with respect to the curve and eliminate those data points that are further than three standard deviations from the curve. The residuals of the remaining points have been found to be normally distributed (Conway et al., 1988). There is a pronounced annual cycle in the data from most sampling sites, and the interpolation between data points that results from curve fitting gives a better estimate of the monthly mean than that obtained by simply averaging the flasks during a particular month. This is especially true when data are missing, resulting in uneven weighting of different periods. Curve fitting is therefore an integral part of our data selection and analysis. Estimates of the monthly, seasonal, and annual mean CO_2 concentrations are determined from the fitted curve. However, the scheme used to fit a curve to the flask data and to reject outliers may introduce a bias into the resulting averages.

In this paper we will discuss the statistical uncertainties of the curve fits, and the systematic biases that may result from outlier rejection, and the influence of the time of day of sampling. Specifically, we want to determine the precision with which monthly, seasonal and annual values can be estimated by the curve fits to the flask records, as well as the amplitude and phase of the annual cycle, the interannual differences and the long term

trends. Although our analysis is confined to CO_2 data, the results can serve as a guide for background sampling of other well-mixed atmospheric gases such as methane and nitrous oxide, each of which, however, will have its own unique features.

The method we used to determine the statistical uncertainties in flask data was to construct a number of surrogate flask records, all obtained from the same parent data set of continuous analyzer (hereafter called CA) observations. After these surrogate records have been treated with various curve fitting and outlier rejection schemes they are compared with the full original CA record. We also considered the effect of an increased sampling frequency for the flasks. These series of tests were performed for data from Point Barrow, Alaska (Peterson et al., 1982, Peterson et al., 1986), because it is a relatively noisy site and the sharp drawdown of the CO_2 concentration at the end of the summer can be hard to reproduce in a curve fit through relatively sparse data.

9.2 Construction of Artificial Flask Records

The CA records are obtained from an hourly measurement cycle consisting of 5 minutes each of two reference gases of known concentration and 50 minutes of ambient air measurement. The voltage output of the CO_2 analyzer is averaged to form one-minute means. From these one-minute means the hourly average voltage and the standard deviations of the one-minute values about the hourly average for ambient air are calculated, as well as average voltages for the two reference gases (due to the response time of the analyzer, the first 3 minutes after each gas change are ignored). The sensitivity of the analyzer, in units of ppm per volt, is calculated from the known reference gas concentrations and their corresponding measured voltages. Hourly average concentration for the ambient air is obtained by multiplying the hourly average voltage by the analyzer sensitivity. Hourly variability is obtained by multiplying the standard deviation of the minute averages by the analyzer sensitivity.

Because our interest in the observations lies in the background concentrations, certain hours must be rejected even after obvious instrument malfunctions have been eliminated. Each station has its own set of background criteria, but they all include eliminating hours during which the standard deviation of the one-minute values is higher than a set value, and eliminating hours when the wind conditions are likely to bring air that is affected by local CO_2 sources or sinks. The first criterion, low within-hour variability, removes most of the hours that are not considered to represent background conditions. We will compare the surrogate flask data, whose construction will be described in detail below, to the preliminary selected background (to be called PSCA) data, or a subset thereof. The PSCA data are obtained from the CA record by removing all hours with high within-hour variability. The PSCA data approximate fairly closely the atmospheric background signal at the site.

For the comparison of curve fitting and outlier rejection methods we constructed two sets of six surrogate flask records from three years (1983-1985) of the CA record at Barrow. We picked one hourly average value

per week, where the first series is picked from Mondays only, the second from Tuesdays, etc. At this point no attention is paid to whether there is large within-hour variability. The actual hour from which the value is taken is determined by a random number generator, but we restricted the range to the hours of the day when flasks at Barrow are usually taken (between 11 a.m. and 2 p.m. local time). If there is an instrument malfunction, that hour is treated as a missing flask value. This led to six parallel flask records because calibrations of the NDIR CO_2 analyzer interfered one day per week. This first set we called the "A" series. The second set of six, the "B" series, is similar to the first but extra noise was added to the hourly average values in the following way. We have distributions of the within-hour standard deviations for each month, since the distribution of the variability depends on the season. A value for the within-hour standard deviation is randomly picked from the distribution for the appropriate month. Then the noise contribution for the simulated flask is determined by picking randomly from a normal distribution characterized by the standard deviation found above with zero mean. Series B is not just series A with noise; the hours themselves were picked at random again.

There are two advantages in comparing parallel series of surrogate flask data with the parent data set from which they have been derived. First, the CA record constitutes a well defined basis for comparison of the flask records and second, the large number of parallel series that can be created allows us to characterize some statistical properties of the flask records and to extract some systematic effects from the noise. The full, unselected CA record can be thought of as representing the actual atmosphere at Barrow, the "background" values of which we try to recover from the flask records.

9.3 Treatment of the Flask Records

In order to extract the information of interest, i.e. monthly and annual means, seasonal cycle characteristics, and long term trend, it is necessary to identify data that are probably erroneous and/or not representative of background conditions. The flask program has been described in detail by Komhyr et al. (1985), and by Conway et al. (1988). Typically, a pair of flasks is filled once (or twice) per week at each site using a portable battery operated pump. Each flask sample represents about a 30 second average of the atmospheric CO_2 concentration. The concentration values of the pairs are averaged if their difference does not exceed 0.5 ppm. The averaged data are then examined for values considered unrepresentative of background conditions. Our simulated flasks have only one value so we treat them as the mean values of satisfactory pairs. However, roughly six percent of the real flask data are identified as outliers and outliers in the simulated flasks data sets will also be rejected. The artificial flask records were subjected to the following smoothing and data selection procedures:

1. Straightforward monthly averages, with subjective outlier rejection by eye.

2. Cubic spline fit, where the "stiffness" of the fit is determined by a statistical test ("runs test") of the residuals, with outliers outside of three standard deviations rejected. This method has recently been used for the GMCC flask samples and is described by Conway et al. (1988).

3. Low-pass digital filtering in the frequency domain. The transfer function is given by

$$H(f) = \exp\left[-\ln 2 (f/fc)^4\right]$$

where f is the frequency, and fc is the cutoff frequency, defined as the frequency at which $H(f)=0.5$. The cutoff is chosen as 0.22 yr^{-1}, corresponding to a period of 80 days. This method has been described by Thoning et al. (1989). Outliers with a residual greater than three standard deviations are rejected.

4. A fit with four harmonics plus a linear trend, and a cubic spline to smooth the residuals of the harmonic fit. The stiffness of the spline is again determined via the runs test. A similar combination of harmonics and a spline has been described by Bacastow et al. (1985) and Keeling et al. (1985). In our case we reject outliers with residuals greater than three standard deviations.

5. A fit with four harmonics plus a linear trend, and digital filtering (same cutoff frequency as in procedure 3) of the residuals of the harmonic fit. Rejection of outliers with residuals greater than three standard deviations.

6. Use LOWESS, a subroutine for robust locally weighted regression (Cleveland, 1985; see also Cleveland and McCrae, this volume), to produce a smooth curve through the data, with automatic rejection of data with residuals greater than six times the median residual.

The curve fitting and data rejection for procedures 2,3,4 and 5 were done iteratively, until no samples remained that were greater than 3 residual standard deviations from the curve. These six different methods have been applied to both the "A" and the "B" series. The results are nearly identical for both sets, so we will report on the "A" series only. Figure 1 shows the daily averages of the PSCA record, together with one of the artificial flask records and a smooth curve drawn through the flasks via method 2. The concentrations are given as part per million by volume in dry air.

Monthly average CO_2 concentrations were calculated by directly averaging the flasks (method 1) or by averaging weekly values of the smooth curves in methods 2-6. These averages were compared with the monthly averages obtained from the preliminary selected CO_2 concentrations between the hours of 11:00 a.m. and 2:00 p.m. (a subset of the PSCA data set). The means and standard deviations of the differences are tabulated in Table 1a and 1b. There are three years of data for six parallel records, so there are 18 numbers for each month from which we calculate the mean and standard deviation. The results of this comparison are plotted in Fig. 2 for methods 1, 2, and 5.

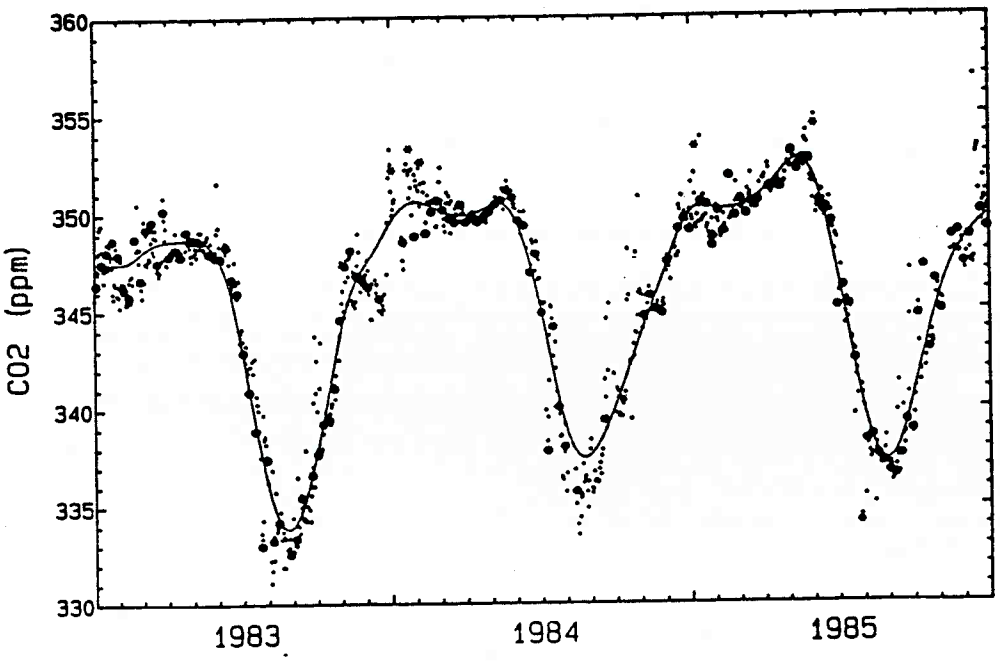

Figure 1. Preliminary selected daily average CO_2 concentrations (·) at Point Barrow, Alaska, derived from hourly concentration data between 11:00 a.m. and 2:00 p.m., when flasks are usually taken. One of the simulated flask data sets is also shown (●), that has been derived from the same set of hourly CO_2 data. The curve fit is a cubic spline (method 2, see text) through the flask data. Flask values rejected as outliers are indicated by an asterisk (*).

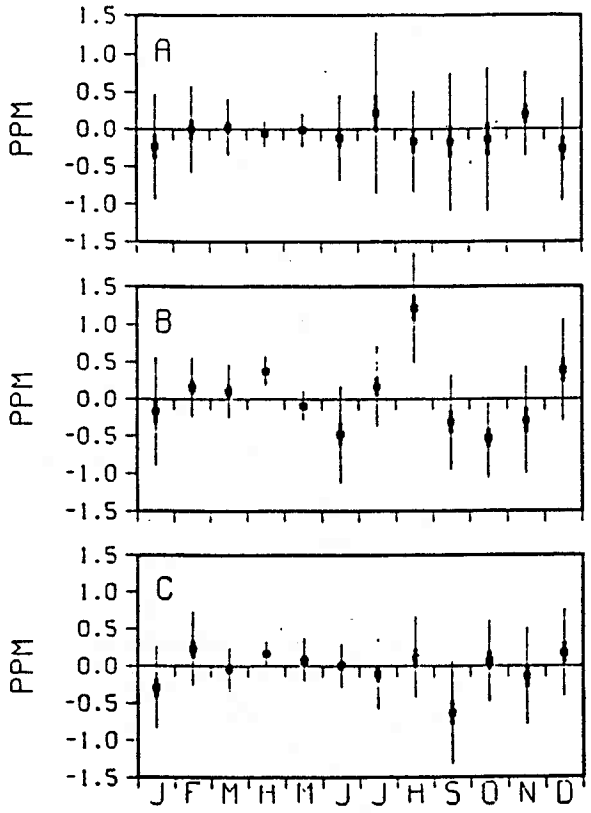

Figure 2. Differences between monthly average CO_2 concentrations determined from simulated flask records and the preliminary selected monthly averaged calculated for the continuous data (between 11:00 a.m. and 2:00 p.m.) from which the flasks have been derived. The standard deviation (thin error bar) of the difference and the mean is plotted, as well as the standard error of the mean (thick error bar). A. Straight averages, method 1; B. spline fit, method 2; C. harmonics and linear trend plus digital filter, method 5.

Table 1a.--Average difference between monthly mean CO_2 concentration determined by the "A" series flasks and the mid-day PSCA record (flask minus PSCA).

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
ave	-0.24	-0.01	0.02	-0.07	-0.02	-0.12	0.21	-0.17	-0.18	-0.15	0.19	-0.28
spl	-0.18	0.15	0.10	0.37	-0.10	-0.48	0.16	1.21	-0.32	-0.54	-0.30	0.38
fft	-0.19	0.32	-0.07	0.21	0.02	-0.11	-0.05	0.32	-0.56	-0.14	-0.18	0.19
hspl	-0.35	0.40	-0.19	0.10	0.23	-0.11	0.01	0.41	-0.76	-0.19	0.03	0.21
hfft	-0.30	0.22	-0.05	0.16	0.07	0.00	-0.11	0.10	-0.64	0.06	-0.15	0.17
low	-0.37	-0.06	-0.02	0.08	-0.22	-0.47	0.63	1.28	-0.34	-0.64	-0.38	-0.09

Table 1b.--Standard deviations of monthly means as determined by the "A" series flasks.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
ave	0.71	0.59	0.39	0.18	0.23	0.58	1.08	0.68	0.93	0.96	0.57	0.69
spl	0.74	0.41	0.37	0.21	0.21	0.66	0.55	0.74	0.65	0.54	0.73	0.69
fft	0.73	0.39	0.30	0.17	0.25	0.31	0.50	0.60	0.72	0.55	0.65	0.61
hspl	0.52	0.70	0.39	0.29	0.42	0.26	0.52	0.64	0.61	0.53	0.91	0.74
hfft	0.56	0.51	0.30	0.17	0.31	0.30	0.49	0.56	0.70	0.56	0.66	0.59
low	1.09	0.50	0.37	0.18	0.23	0.33	0.59	0.80	0.66	0.44	0.70	0.50

Clearly, the method of straight averaging produces no significant biases, but the standard deviations are rather large (Fig. 2A). From July through October the uncertainty (one sigma) of the flask monthly averages is close to 1 ppm. Smoothing reduces the variance of the estimates, but at the cost of introducing biases. The spline fit and LOWESS have a tendency to underestimate the summer drawdown of CO_2 (Figs. 1 and 2B). The spline could be made more flexible by changing the stiffness criterion, but that introduces structure to the curve during winter and spring that seems unwarranted by the data. LOWESS suffers from the same problem. It cannot fit the summer drawdown unless the (adjustable) local region over which LOWESS calculates an average is made so small that hardly any smoothing results.

The combination of harmonic fits with some smoothing of the residuals of the harmonic fits gives better results. The harmonics capture most of the drawdown, after which the demands on the smoothers are less severe. An additional advantage is that we obtain at each site the harmonic coefficients that describe most of the average behavior of the seasonal cycle. The phase and amplitude of each harmonic can be compared with the predictions of

atmospheric transport models that incorporate seasonal sources and sinks of CO_2. Our preferred method is the combination of harmonics and digital filtering (method 5) because it has no significant bias in August (Fig. 2C) when the seasonal CO_2 concentration is at its lowest. We will use this method in the remainder of this paper. The statistical uncertainty of the monthly averages as determined from the flask data is 0.4 ppm (one sigma) during the first half of the year and 0.6 ppm during the second half.

9.4 Influence of Data Selection

An intriguing feature of fitted curves is that the monthly average concentrations for September are underestimated in most of the flask records. Closer inspection reveals that this difference is related to data selection. Data which differ from the curve by more than three standard deviations of the residuals have been ignored in the flask records. When we fit a smooth curve (digital filtering, cutoff frequency 0.22 yr-1) to the PSCA data using only hourly data between 11:00 a.m. and 2:00 p.m. and discard values with residuals greater than three sigma, there is no significant difference between this version of the continuous data and the mean flask results obtained by method 5.

It can be argued that we should not ignore data with large residuals in order to avoid systematic biases such as we found above. We would agree with that in the case of these simulated flasks, but not in the case of the real flasks. The simulated flasks do not mimic some additional errors that can be present in the real flasks, such as operator error. We feel that even though we cannot easily quantify operator error, the chance is too great that real flasks containing additional errors could be retained when outliers are not rejected, thereby introducing even greater biases.

The time of the day when flask samples are taken can also produce a bias. The flask samples are usually obtained around noon, which can result in a bias if there is a diurnal variation at the site. We have compared the PSCA data comprising the full 24 hours with the mid-day hours only. During July and August the mid-day (between 11:00 a.m. and 2:00 p.m.) average is 0.3 ppm lower than the 24-hour average. This is probably due to photosynthetic activity in the vicinity of the station that is not recognized by the preliminary selection criterion of high within-hour variability.

9.5 Determination of Recurring Seasonal Characteristics

We characterized the seasonal variation of the CO_2 concentration by fitting a linear combination of four harmonics and a linear trend to the three years of data in each of the artificial flask series:

$$C(t) = a + b\,t + \sum_{i=1}^{4} c_i \sin(2\pi i t + \phi_i)$$

In Table 2 we compare the values that we find for the harmonic amplitudes and phases of the "A" series with those of the mid-day hours of the PSCA data. The fit yields significant coefficients for all four harmonics with only three years of data. The coefficients are close to those determined from the continuous data. Data selection introduces differences in the coefficients for the continuous data. In the second column marked "PSCA", data with residuals greater than 3 ppm have been ignored which especially seems to affect the phase of the harmonics.

Table 2.--Seasonal characteristics found from PSCA data and "A" series of surrogate flask data

	PSCA mid-day	PSCA mid-day no outliers	Flasks mean	Flasks st.dev.
trend (ppm/year)	1.44	1.44	1.36	0.13
amplitude 1 (ppm)	6.55	6.60	6.64	0.09
phase 1 (degrees)	22.2	19.8	22.2	0.8
amplitude 2	3.04	2.98	3.07	0.06
phase 2	-23.5	-26.8	-22.4	2.9
amplitude 3	1.29	1.15	1.11	0.16
phase 3	-42.3	-50.8	-47.	7.
amplitude 4	0.43	0.35	0.38	0.17
phase 4	-85.	88.	-85.	16.

9.6 Some Additional Artificial Flask Records

We tested how some alternative sampling strategies might affect the results. We constructed six artificial flask series in which the flasks were chosen at random from the entire 24 hours of the day. This set of flask series has the same systematic biases with respect to the full PSCA data that the "A" flask series has with respect to the mid-day PSCA data. We also increased the sampling frequency of the artificial flasks to twice a week for both the "A" series and the series derived from the full 24 hours of the day. The statistical uncertainty of the monthly averages decreases by $\sim\sqrt{2}$, as expected, but the systematic differences with the PSCA data remain. We conclude that there is not much gain in sampling twice a week, provided there are no frequent measurement problems or sampling errors. In the latter case more frequent sampling would increase the experimental control of the results, although for that purpose it would probably be better to collect two pairs of flasks in rapid succession once a week. Gillette and Hanson (1983) concluded from statistical considerations that the optimum sampling frequency would be twice a week or even more frequently. However, they did not consider biases that result from the rejection of outliers.

9.7 Summary and Conclusions

Simulated flask data time series have been constructed for the continuous analyzer record at Barrow, Alaska. From these flask records we have tried to reconstruct background estimates of the CO_2 concentration records, where background is defined by the preliminary selected continuous data. After removal of outliers and the application of various curve-fitting techniques the monthly means derived from flasks exhibit a standard deviation varying between 0.2 and 1.1 ppm. The flask derived monthly means also exhibit offsets varying from -0.8 to 1.2 ppm. The latter feature arises because of shifts in the curve that occur after the rejection of data outliers.

The optimal curve-fitting method found used four yearly harmonics plus a linear trend, with digital filtering of the residuals. This method had generally low noise and it captured the point of deepest CO_2 drawdown in August without significant bias and without introducing spurious variability in other parts of the record. The harmonic coefficients were determined very tightly by only three years of flask data. In addition, the harmonic coefficients themselves are valuable as a definition of the annual cycle, since they can be compared to the harmonics of annual cycles generated by three-dimensional carbon cycle models.

9.8 References

Bacastow, R.B., C.D. Keeling, and T.P. Whorf, 1985. Seasonal amplitude increase in atmospheric CO_2 concentration at Mauna Loa, Hawaii, 1959-1982. J. Geophys. Res., 90, 10529-10540.

Cleveland, W.S., 1985. The Elements of Graphing Data. Wadsworth, Monterey.

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

Gillette, D.A., and K.J. Hanson, 1983. Sampling strategy to obtain data used in models of global annual CO_2 increase and global carbon cycle. J. Geophys. Res., 88, 1345-1348.

Gillette, D.A., and A.T. Steele, 1983. Selection of CO_2 concentration data from whole-air sampling at three locations between 1968 and 1974. J. Geophys. Res., 88, 1349-1359.

Keeling, C.D., T.P. Whorf, C.S. Wong, and R.D. Bellagay, 1985. The concentration atmospheric carbon dioxide at Ocean Weather Station P from 1969 to 1981. J. Geophys. Res., 90, 10511-10528.

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

Peterson, J.T., W.D. Komhyr, T.B. Harris, and L.S. Waterman, 1982: Atmospheric carbon dioxide measurements at Barrow, Alaska, 1973-1979. Tellus, 34, 166-175.

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

Tans, P.P., T.J. Conway, and T. Nakasawa, 1989. Latitudinal distribution of the sources and sinks of atmospheric carbon dioxide derived from surface observations and an atmospheric transport model. J. Geophys. Res., in press.

Thoning, K.W., P.P. Tans, and W.D. Komhyr, 1989. Atmospheric carbon dioxide at Mauna Loa Observatory: II Analysis of the NOAA GMCC data, 1974-1985. J. Geophys. Res., in press.

10. THE SEASONAL CYCLE OF CO_2 AT MAUNA LOA: A RE-ANALYSIS OF A DIGITAL FILTERING STUDY

I.G. Enting

CSIRO, Division of Atmospheric Research
Private Bag No. 1, Mordialloc
Victoria 3195, Australia

and

M.R. Manning

DSIR Institute of Nuclear Sciences
Private Bag, Lower Hutt, New Zealand

In a recent study, one of us (Enting, 1987) examined the interannual variation in the amplitude of the seasonal cycle of CO_2 at Mauna Loa. The procedure used was to apply a band-pass filter to the time series of monthly means so as to extract an estimate of the seasonal cycle on a year-to-year basis. Visual inspection of the resulting estimate (Fig. 3 of Enting, 1987) suggested that abnormally large spring peaks were usually followed by abnormally low fall troughs, but that there was much less correlation between fall troughs and the peaks in the following spring. Enting (1987, Fig. 4) presented a pair of scatter plots, showing the CO_2 cycle in May vs the following October and October vs the following May, as further evidence of the apparent asymmetry. It was further suggested that the differences represented a characteristic biotic signal.

We now report a re-analysis of the data which we have undertaken for several reasons:

i. Further data are now available from CDIAC in the form of revision 1 to numeric data package NDP-001. The new NDP-001/R1 (Keeling, 1986) extends the data by about two years and makes minor changes (typically 0.01 ppmv) to the monthly values, presenting the data as monthly means corrected to the 15th of the month.

ii. The data as analyzed by Enting included a data entry error of 1 ppmv in one of the interpolated values.

iii. There was concern (by the present authors and other attendees at the Hilo meeting) that the scatter-plot analysis would be questionable because of the longer-term changes in the amplitude of the seasonal cycle at Mauna Loa.

iv. Most seriously, one of us (Manning, personal communications) noted that the scatter-plot analysis presented by Enting (1987) represented a biased comparison between summers and winters.

This last point requires further explanation and the appropriate starting point is to consider how the filter treats white noise.

The estimates, $\hat{y}_j$ used by Enting (1987) were of the form

$$\hat{y}_j = \Sigma_k c_k z_{j-k} \tag{1}$$

where the z_m are the input time series data and c_k are filter coefficients.

If the input is white noise e_j with unit variance i.e.

$$E(e_i e_j) = \delta_{ij} \tag{2}$$

then the autocovariance of the estimate y_j is

$$E(\hat{y}_j \hat{y}_{j+m}) = \Sigma_k c_k \Sigma_{k'} c_{k'} E(e_{k-j} e_{k'-m-j}) \tag{3}$$

$$= \Sigma_k c_k c_{k+m}$$

where E(.) denotes statistical expectations. For symmetric filters this autocovariance represents the convolution of the filter with itself. For the filter used by Enting (1987) the autocorrelation from white noise is shown in Fig. 1. It will be seen that there is a negative correlation for lags of five months but very little correlation at lags of seven months. Because of the asymmetry in the cycle, the peak to trough comparisons in Fig. 4 of Enting (1987) involve lags of five months across summers and seven months across winters. Thus the comparison was biased and the qualitative results correspond to those expected from applying the filter to a fixed cycle plus white noise.

In our re-analysis of the data we present what we regard as unbiased comparisons by using six-month lags in each case. We also address point (iii) above by taking into account the longer-term changes in the amplitude of the seasonal cycle of CO_2 at Mauna Loa. Figure 2 shows an estimate of the amplitude of the 12-month component of the Mauna Loa record obtained using the technique of complex demodulation (Bloomfield, 1976). Figure 2 extends the plot given by Thompson et al. (1986) by including more recent data and also by showing estimates for the period before 1966 which were not presented by Thompson et al. because of their interest in inter-station comparisons. The main feature of this record is the change around 1976 from variations about an approximately fixed mean amplitude to a period of larger (and possibly increasing) amplitudes after 1980. The fact that most of the longer-term increase in amplitude occurred towards the end of the 1970's, also appears in the analysis of the cycle by Cleveland et al. (1983).

On the basis of these longer-term changes in amplitude, we present (in Fig. 3) revised versions of the scatter plots for signal estimates using the same filter as in Enting (1987). In each case lags of six months are used with Fig. 3a showing May vs the following November while Fig. 3b shows November vs the following May. In each case the first 15 data points are shown as solid circles, corresponding to the apparent steady-state prior to

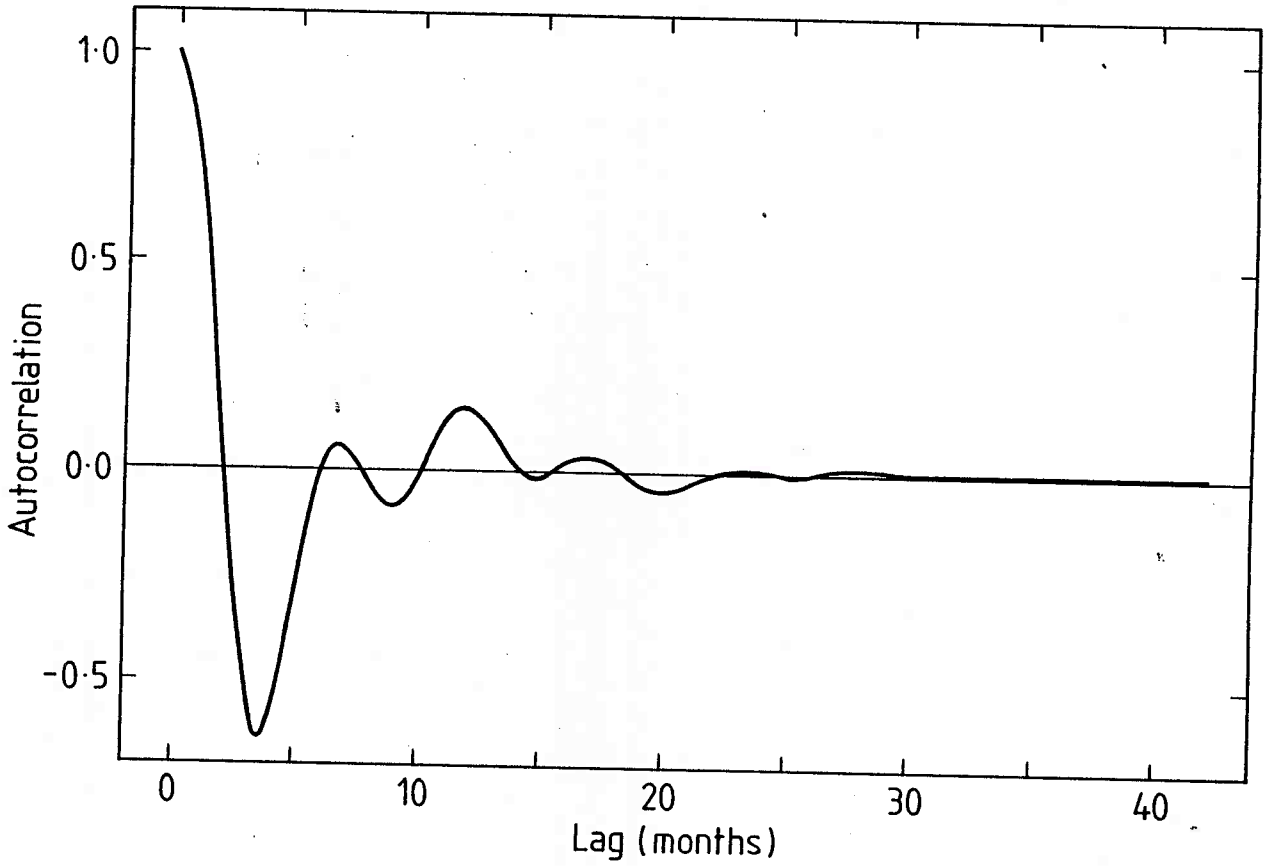

Figure 1. Auto correlation obtained by applying the filter used by Enting (1987) to white noise. The differences between the correlations for lags of five months as opposed to seven months indicates that the scatter-plots of Enting (1987) comparing cycle peaks represented a biased comparison.

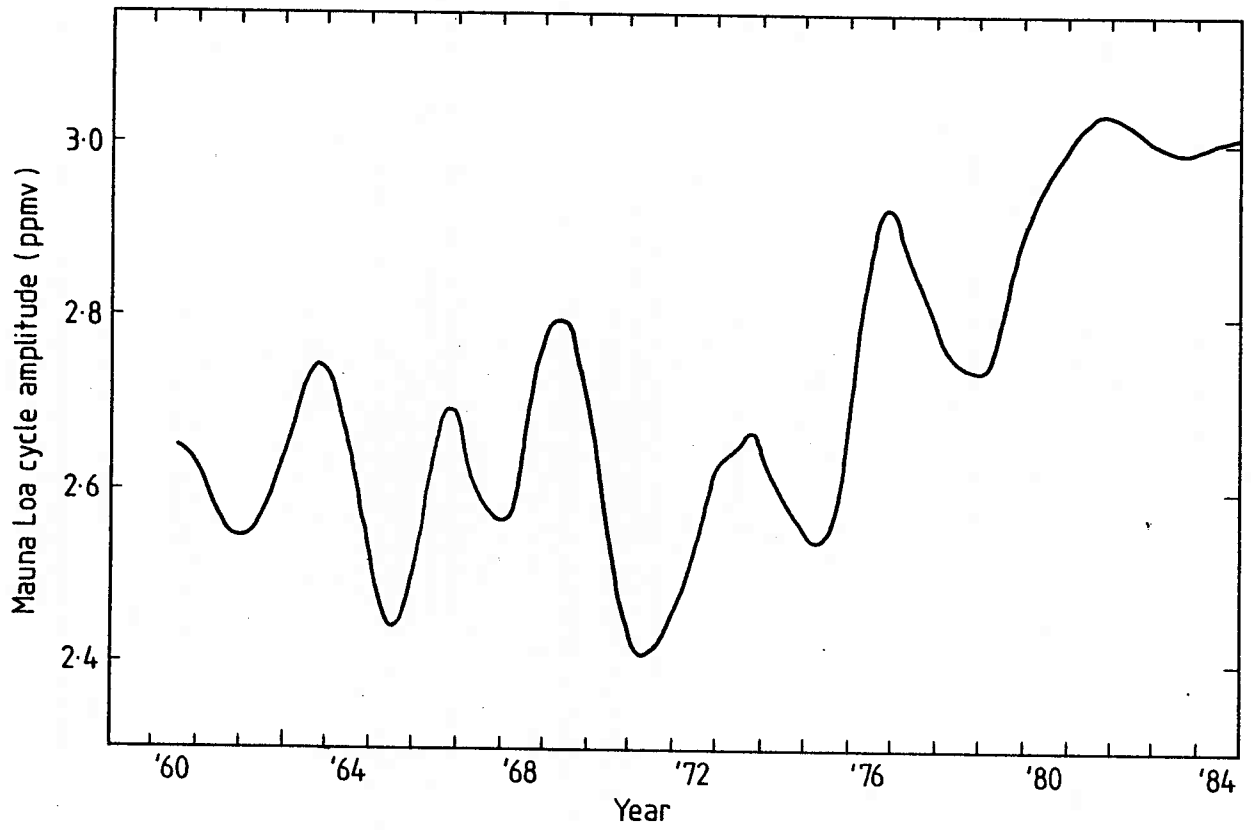

Figure 2. Amplitude of the twelve-month component of the CO_2 cycle at Mauna Loa as estimated using complex demodulation.

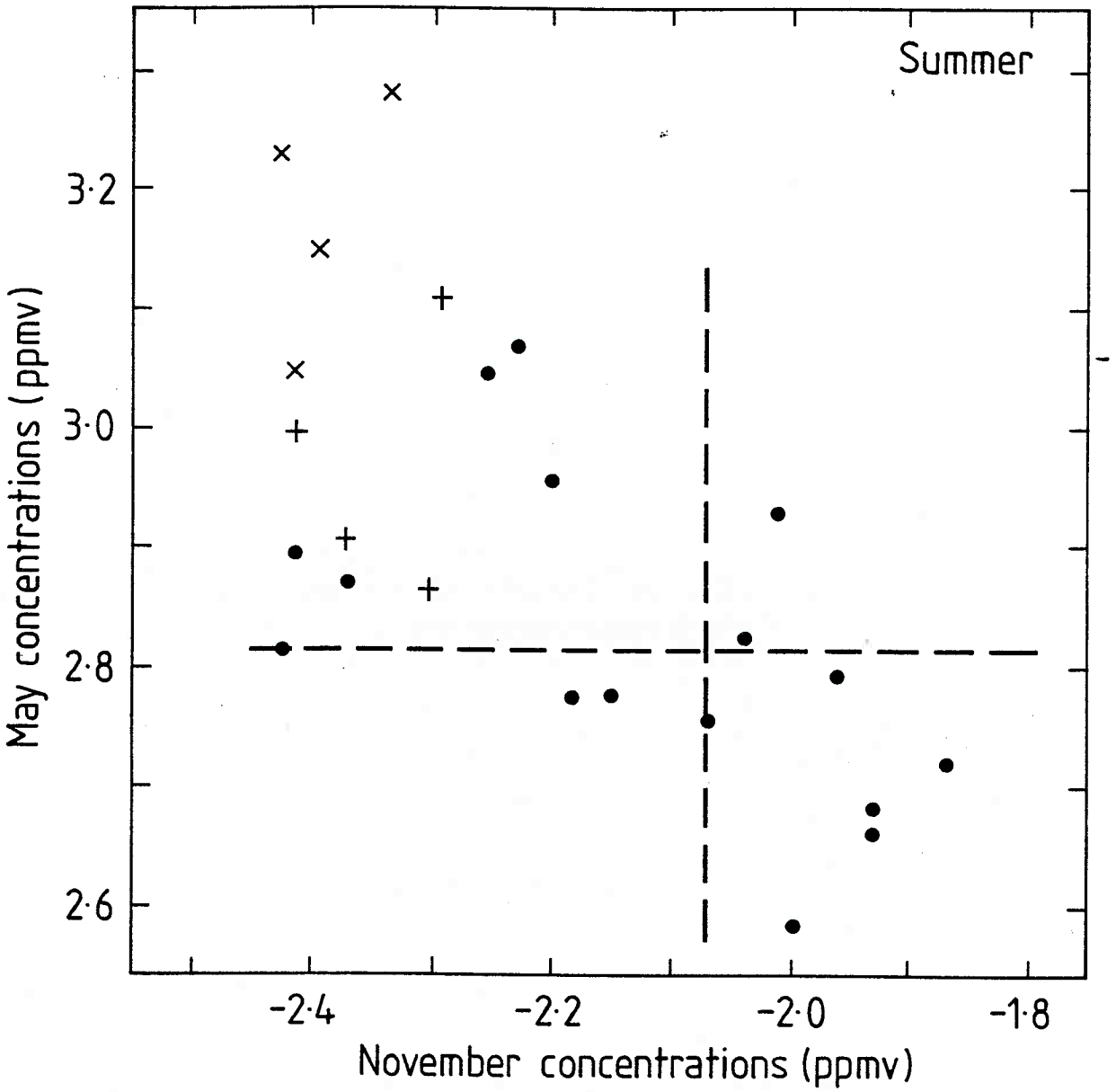

Figure 3a. Scatter plots for lags of six months, May to following November. The data for the first fifteen years (1960-1974) is shown as dots (·); the next five years (1975-1979) as pluses (+); and the last four years as crosses (x).

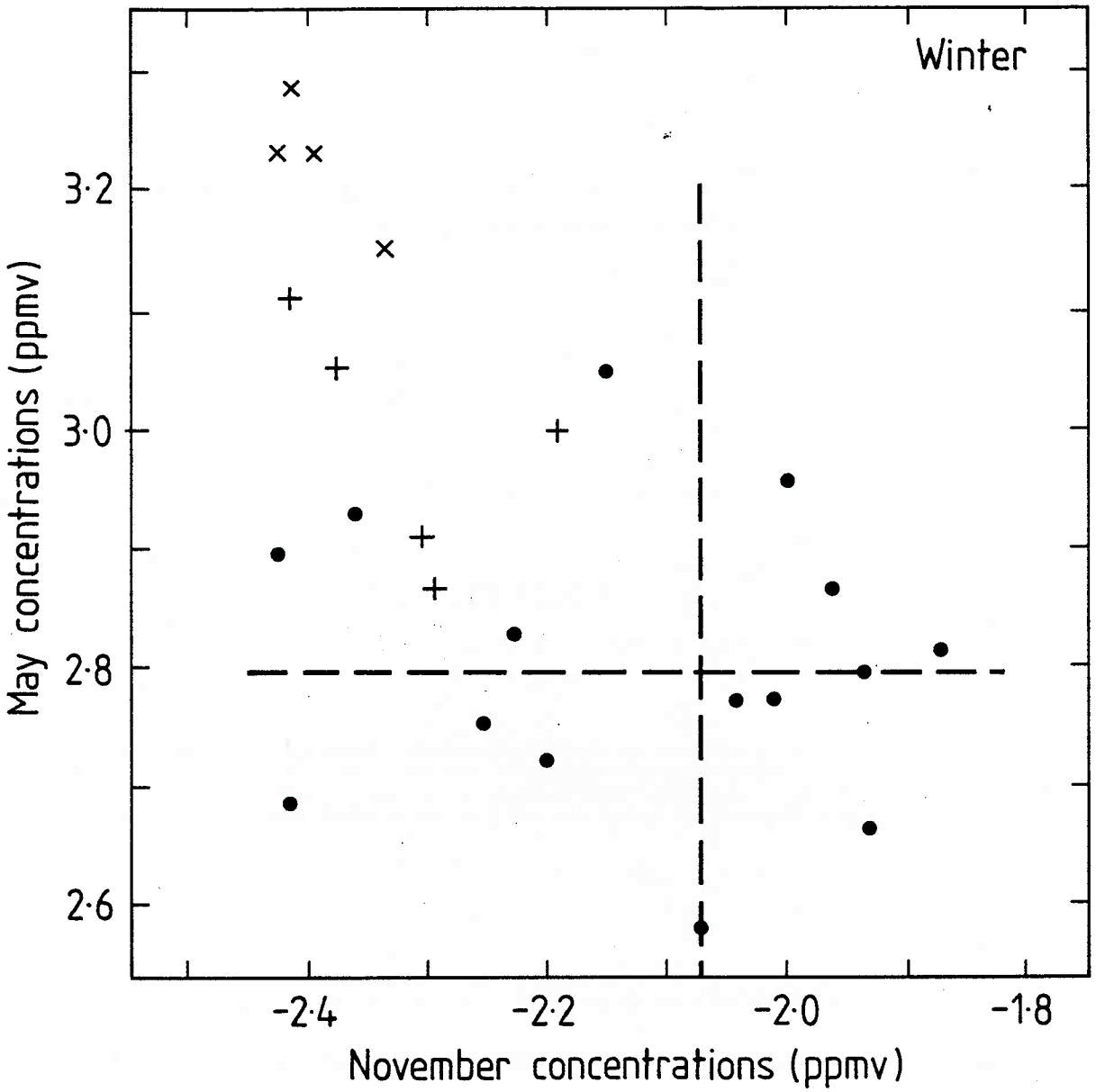

Figure 3b. Scatter plots for lags of six months, November to following May. The data for the first fifteen years (1960-1974) is shown as dots (·); the next five years (1975-1979) as pluses (+); and the last four years as crosses (x).

1976. The next five points are shown with + signs, corresponding to a possible transition period from 1976 to 1980 and the remaining points are shown with x signs.

If the data sets are taken as a whole in each case then there is rather little qualitative difference between the two scatter plots. A negative correlation appears in each case although it seems a little stronger across the summers. However, when the data are distinguished according to the three time periods described above it is immediately clear that the record is not statistically homogeneous and that the first 15 data points represent smaller amplitudes than the later data. It would also seem (albeit on the basis of a small number of points) that there are definite differences between the "transitional" and the "post-1980" data. This difference could imply that the late 1970"s represented a transition to a new quasi-steady state with generally larger amplitudes. Alternatively the two post-1976 segments of the data could differ simply because they represent different segments of a period in which amplitudes continued to increase.

Returning to the analysis of the shorter-term amplitude variations, the circles in Fig. 3 tend to confirm the correlation structure that was detected in the visual inspection of the cycle estimate. The dashed lines in Fig. 3 represent the median values for May and November data sets for the first 15 years. (The May medians differ because the two data sets differ by one point). The "winter" points are distributed between the four quadrants in numbers 3,3,3 and 4 (i.e. as uniformly as possible with 13 points) while the "summer" points are distributed between the quadrants as 5,5,2 and 1 points. Since only 15 comparisons are considered, the differences between the distribution of pre-1976 data for summers and winters do not have a high degree of statistical significance. The use of medians in defining the quadrants of the scatter plots ensures that any asymmetry will be such that diagonally opposite quadrant have approximately equal numbers of points. Monte Carlo simulation using 10000 sets of 15 points suggests that the probability of having a skewness of 10:3 or more is about 0.08 for those cases in which the medians are different points if the data are independent. Since the data points are not independent, the 92% estimated significance must be reduced. An analysis (J.V. Mansbridge, personal communication) of the complete set of relative rankings of the May and November coordinates using Spearman's ranked correlation test (Conover, 1971) suggests the correlations across summers were significantly different from zero at the 95% confidence level, again assuming independent data. Correlations across winters (as measured by Spearman's rho) were only 10% as strong as the correlations across summers. Spearman's test indicates an 80% probability of such a small correlation having arisen by chance, from random data.

The overall conclusion from our re-analysis of the Mauna Loa seasonal cycle is that the earlier part of the record shows a seasonal asymmetry in the interannual variation of the seasonal cycle. There is a greater correlation across summers than across winters in the manner which Enting (1987) suggested would be characteristic of interannual variations in biotic productivity. It would seem that for further investigation of the statistical characteristics of these short-term interannual variations in the seasonal cycle, we require

an extended period of observations during which no longer-term changes in amplitude occur.

References

Bloomfield, P., 1976. Fourier Analysis and Time Series: An introduction. John Wiley, NY.

Cleveland, W.S., A.E. Freeny, and T.E. Graedel, 1983. The seasonal component of atmospheric CO_2: Information from new approaches to the decomposition of seasonal time series. J. Geophys. Res., 88, 10934-10946.

Conover, W.J., 1971. Practical Non-parametric Statistics. John Wiley, NY.

Enting, I.G., 1987. The interannual variation in the seasonal cycle of carbon dioxide concentration at Mauna Loa. J. Geophys. Res., 92, 5497-5504.

Keeling, C.D., 1986. Atmospheric CO_2 concentrations - Mauna Loa Observatory Hawaii 1958-1986. Report NDP-001/R1. Carbon Dioxide Information and Analysis Center, Oak Ridge.

Thompson, M.L., I.G. Enting, G.I. Pearman, and P. Hyson, 1986. Interannual variation of atmospheric CO_2 concentration. J. Atmos. Chem., 4, 125-155.