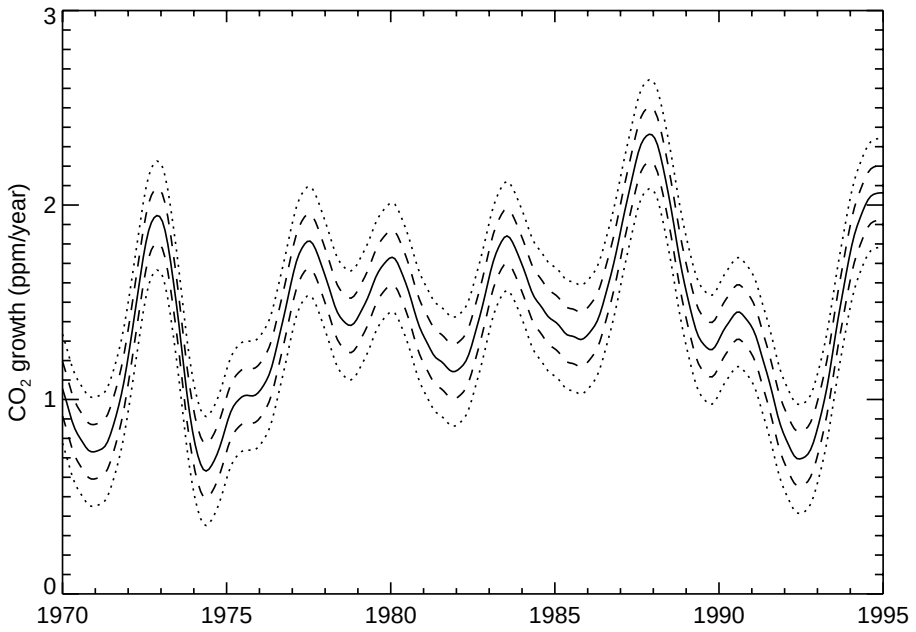
Characterising the Temporal Variability of the Global Carbon Cycle

I.G. Enting



CSIRO

Atmospheric Research

CSIRO Atmospheric Research Technical Paper no. 40

Characterising the Temporal Variability of the Global Carbon Cycle

I.G. Enting

CSIRO Atmospheric Research Technical Paper no. 40

National Library of Australia Cataloguing-in-Publication Entry

I.G. Enting
Characterising the Temporal Variability of the Global Carbon Cycle

Bibliography.
ISBN
0 643 06347 1 (print edition)
0 643 06505 9 (electronic edition)

1. Atmospheric carbon dioxide — Measurement. 2 Carbon cycle (Biogeochemistry) — Mathematical models. I. CSIRO. Division of Atmospheric Research. II. Title. (Series: CSIRO Atmospheric Research technical paper; no. 40).

577.14

Address and contact details: CSIRO Atmospheric Research
Private Bag No.1 Aspendale Victoria 3195 Australia
Ph: (+61 3) 9239 4400; fax: (+61 3) 9239 4444
e-mail: chief@dar.csiro.au

CSIRO Atmospheric Research Technical Papers may be issued out of sequence.

CSIRO Atmospheric Research Technical Paper no. 40

Characterising the Temporal Variability of the Global Carbon Cycle

I.G. Enting
CSIRO Atmospheric Research
Private Bag 1, Aspendale,
Vic 3195, Australia

Abstract

Over recent years there has been an increasing appreciation of the degree of natural temporal variability in the global carbon cycle. Analysis of this behaviour requires techniques that can estimate the variability and characterise the degree of uncertainty in such estimates. This report analyses and illustrates a number of such tools, concentrating on variations on the time-scale of seasons or longer.

1 Introduction

The natural carbon cycle involves large exchanges of carbon between atmospheric CO_2 and the major carbon reservoirs in the oceans and the terrestrial biosphere. The anthropogenic input of CO_2 to the atmosphere represents a relatively small perturbation to this natural cycle. Over recent years there has been an increasing appreciation of the extent to which there is a large natural variability in the carbon cycle and of the need for this to be taken into account when trying to interpret the anthropogenic perturbation.

The large-scale nature of variability in the carbon cycle has been apparent since the work of Bacastow (1976) who identified a correlation between the CO_2 growth rate and the Southern Oscillation index. Subsequent studies (e.g. Thompson et al., 1986) have confirmed the large-scale nature of such variability and also identified large-scale interannual variations in the amplitudes of seasonal cycles of CO_2 concentrations. More recent studies using both CO_2 and ^{13}C data have interpreted variability in composition in terms of oceanic and terrestrial fluxes, both of which were estimated to have interannual variability that was a significant fraction of the anthropogenic perturbation (Francey et al., 1995a; Keeling et al., 1995; see however Lee et al., 1998).

For a variety of reasons, the analysis of the carbon cycle requires a careful treatment of uncertainty. One requirement is the evaluation of how present uncertainties propagate into future predictions. Another problem that requires analysis of uncertainties is the assessment of the significance of apparent differences between different ways of studying the cycle. The issue is often one of estimating quantities of possible biogeochemical significance from data that are subject to a number of uncertain perturbing influences. This needs to be treated as a statistical estimation problem. Because of both the natural variability and the progressive changes in the anthropogenic perturbation, an uncertainty analysis requires the statistical characterisation of functions of time. This report investigates a range of issues involved and gives a number of il-

lustrative examples. The main aim of this report is to review a number of important tools, and so some of the examples involve methane (CH_4) which is somewhat simpler than CO_2, because of the absence of large two-way fluxes between reservoirs. Detailed applications will be presented elsewhere.

Some of the issues involved in analysing, describing and comparing quantities with such temporal variation in a function $g(t)$ are:

i. *How are (continuous) functions of time estimated from discrete data*?

ii. *How are confidence intervals assigned to statistical estimates,* $\hat{g}(t)$? What are the requirements for appropriate statistical modelling?

iii. *How can estimates of a function* $g(t)$ *(and associated uncertainty) be related to confidence intervals on growth rates*? As the mass-balance Equation (1.1) shows, fluxes are related to *rates of change of concentration*. Estimation of such growth rates involves the (ill-conditioned) problem of numerical differentiation of observational data.

iv. *How can confidence intervals be assigned to seasonal cycles*? The seasonal cycle is a major feature of the carbon cycle and estimated variations are being attributed to specific biogeochemically significant changes.

v. *What effect do correlated errors have on the estimates and their confidence intervals*? This analysis aims to cover general error distributions, with white noise (independent errors) as a particular case. However many of the standard techniques used to analyse time series have been used in forms that are applicable only for the 'white noise' case.

vi. *What are the appropriate measures of uncertainty for estimates of smoothed quantities*? There is a need to be specific about what is being estimated by smoothing observations: whether the smoothing is being treated as a way of reducing the extent to which noise affects the estimate (at the expense of losing some of the signal) or whether one is really trying to estimate a smooth function.

vii. *How can statistical studies be related to modelling*? More specifically how can statistical and deterministic modelling be interfaced? The problems involved include model calibration, stochastic forcing and assessing uncertainties in projections of the future.

viii. *What are the consistency requirements when combining disparate data*? This is important because much of our information about the carbon cycle is indirect and our understanding is based on combining a range of perspectives.

ix. *How can the global representativeness of a time series be assessed and how can the error due to non-representativeness be characterised*? The observations of the carbon cycle are primarily time series from a small number of locations. The atmospheric records are predominantly remote sites, giving good signal-to-noise characteristics but poor global coverage.

x. *More generally, how does the characterisation of temporal variability and uncertainty extend to spatial and space-time analogues*? The higher-dimensional cases can be expected to bring in new complications relative to one-dimensional time series, but many of the underlying concepts can be expected to remain applicable.

xi. *How can the uncertainties be represented*? This is a matter of finding effective ways of communicating the results of the analysis of uncertainties.

An important case, which we examine from several perspectives, is the equation for the overall carbon balance of the atmosphere:

$$\frac{d}{dt}M(t) = \Phi_{\mathrm{Fossil}}(t) + \Phi_{\mathrm{Ocean}}(t) + \Phi_{\mathrm{Biota}}(t) \tag{1.1}$$

where $M(t)$ is the atmospheric carbon content and Φ_{Fossil}, Φ_{Ocean} and Φ_{Biota} are the net carbon fluxes into the atmosphere from fossil emissions, the oceans and the terrestrial biota.

The mass-balance Equation (1.1) applies instantaneously. However, none of the quantities in the equation can be determined on arbitrarily short time-scales. Some form of time-averaging always occurs. For consistency, the same time-averaging must be applied to each of the terms in such equations if the validity of the relations is to be retained.

Some of the reasons for time-averaging are:

- to avoid trying to represent the small time-scales that are not resolved by discrete data;
- to reduce the influence of measurement error by averaging multiple measurements;
- to reduce the influence of natural time-variability by averaging over such variations;
- to reduce the influence of spatial inhomogeneities, by time-averaging over times with differing atmospheric transport;
- to allow comparisons with concentration data from ice cores where the trapping process gives concentrations that are a time-average of the atmospheric concentrations;
- to provide consistent comparisons with other data sets that are in a time-averaged form.

We use the notation $\langle . \rangle$ to denote time-averaging. In most cases we consider a linear stationary formalism that treats all times equivalently. Such averages, of a function $a(t)$, can be written as convolutions with averaging kernels, $\psi_\gamma(.)$, as:

$$\langle a(t) \rangle_\gamma = \int \psi_\gamma(t') a(t+t') dt' = \int \psi_\gamma(t-t'') a(t'') dt'' = \psi_\gamma * a \tag{1.2}$$

where $*$ denotes a convolution operation. A budget equation such as (1.1) must be consistently averaged by applying the same averaging operator to each term:

$$\langle \frac{d}{dt}M \rangle_\gamma = \langle \Phi_{\mathrm{Fossil}} \rangle_\gamma + \langle \Phi_{\mathrm{Ocean}} \rangle_\gamma + \langle \Phi_{\mathrm{Biota}} \rangle_\gamma \tag{1.3}$$

The same notation is used for the discrete case involving a sum rather than an integral.

In this report, the integral form (Equation (1.2) and its various special cases) will mainly be used as a conceptual tool. Actual time-averages (or more precisely estimates of such averages) will be obtained from sums over discrete data. We denote such digital filtering by:

$$\langle\langle z \rangle\rangle_\gamma = \sum_k \psi_{k:\gamma} z(t + k\Delta t) \tag{1.4}$$

We also define the filter response, $\Psi_\gamma(.)$, in terms of the dimensionless frequency θ as

$$\Psi_\gamma(\theta) = \sum_k \psi_{k:\gamma} e^{ik\theta} \tag{1.5}$$

In addition to time averaging, relations such as (1.1) can be transformed by other means such as Fourier transforms, Laplace transforms and spatial disaggregation. As with time-averaging, such transformations have to be applied consistently to all terms if the integrity of the relations are to be maintained.

An additional issue that applies to studies of CO_2 is the need to distinguish different related budget concepts. On a regional scale, the atmospheric carbon budget is not identical to the atmospheric CO_2 budget because of a significant role played by CO and its precursors (Enting and Mansbridge, 1991). Furthermore, the net fluxes from oceanic and biotic reservoirs are not equal to the respective rates of change of reservoirs because of the role of rivers in transporting carbon from biota to oceans without passing through the atmosphere (Sarmiento and Sundquist, 1992).

The contents of the remainder of this report are as follows: Section 2 describes the overall statistical framework used in this report and gives details of several important special cases. Section 3 demonstrates several ways of presenting analyses of uncertainty. Section 4 considers the issues of deriving indirect inferences from observational data, in terms of estimating functions derived from observations. Special cases are growth rates, seasonal cycles and fluxes determined by deconvolution. As a more complicated example, we revisit the joint CO_2-$^{13}CO_2$ budget analysis of Francey et al. (1995a). In Section 5 the use of smoothing splines is discussed. Section 6 notes analogous issues in the spatial variability of the carbon cycle. It also addresses the specific issue of representativeness of observational data: how measurements from a small number of sites can be related to global changes. Section 7 reviews the recursive estimation techniques based on the Kalman filter. Section 8 summarizes the results in terms of the requirements for undertaking consistent analyses of carbon cycle variability. An appendix lists the notation.

2 Framework

2.1 Summary

Our consideration of uncertainties in the carbon cycle is presented in statistical terms. We regard the process of analysing the cycle as being one of obtaining statistical *estimates* of quantitities

of biogeochemical significance. The underlying principle on which we base our discussion is that *any statistical analysis has to be based on a statistical model.* We consider a number of statistical models, but an underlying property is that we model observations as an additive combination of a signal and 'random' noise. Regarding the observations as having a random component means treating them as random variables, with the consequence that any estimates derived from these observations must also be regarded as random variables.

Within this general framework it is necessary to define a statistical model of the error or noise terms. Techniques such as regression analysis or other least-squares fits implicitly assume independent identically-distributed Gaussian noise. If this assumption is not true then estimation techniques based on this assumption will produce estimates that may be biased, or at least subject to larger-than-necessary uncertainty and error estimates may be quite inaccurate. This is a particularly serious problem if the error distribution is 'long-tailed' (relative to a Gaussian distribution) and/or correlated.

Enting and Pearman (1993) characterised the 'uninterpretable noise' as whatever could not be interpreted i.e. modelled deterministically. They noted that this would depend on the sophistication of the model that was used to interpret the data. This still seems to be an appropriate approach, but for illustrative purposes in this report a number of simple time series models of error statistics are used, without being related back to any particular deterministic model.

A simple class of correlated noise models can be defined by autoregressive processes. A first-order autoregressive, AR(1), noise model for the Mauna Loa CO_2 time series was used by Cleveland et al. (1983). This model, together with a number of other 'noise' models for CO_2 data, is summarized in Section 2.6. Regardless of the basis of the statistical model, an important part of the analysis should be checking the model by examining the residuals to ensure that they are consistent with the model assumptions.

Much of the analysis is based on digital filtering, either for actual calculations or as an illustrative concept. We define a number of the filters that are used in several examples:
The running mean of $2K+1$ time-steps is defined by:

$$\psi_k = \frac{1}{2K+1} \qquad \text{for } -K \le k \le K \tag{2.1.1a}$$

For even numbers, we define a $2K$-step 'running mean'

$$\psi_k = \frac{1}{2K} \qquad \text{for } -K+1 \le k \le K-1 \tag{2.1.1b}$$

$$\psi_k = \frac{1}{4K} \qquad \text{for } k = -K \text{ or } k = K \tag{2.1.1c}$$

Bloomfield (1976) defined a class of low-pass filters specified by two parameters, $\nu_c = \theta_c/2\pi\Delta t$, the 'cutoff' frequency, and K, the half-width. The half-width, K, leads to a transition band, of width $\Delta\theta = 4\pi/(2K+1)$, over which the filter response drops from near 1 to near 0. The filters are defined by:

$$\psi_0 = \frac{\alpha\theta_c}{\pi} \tag{2.1.2a}$$

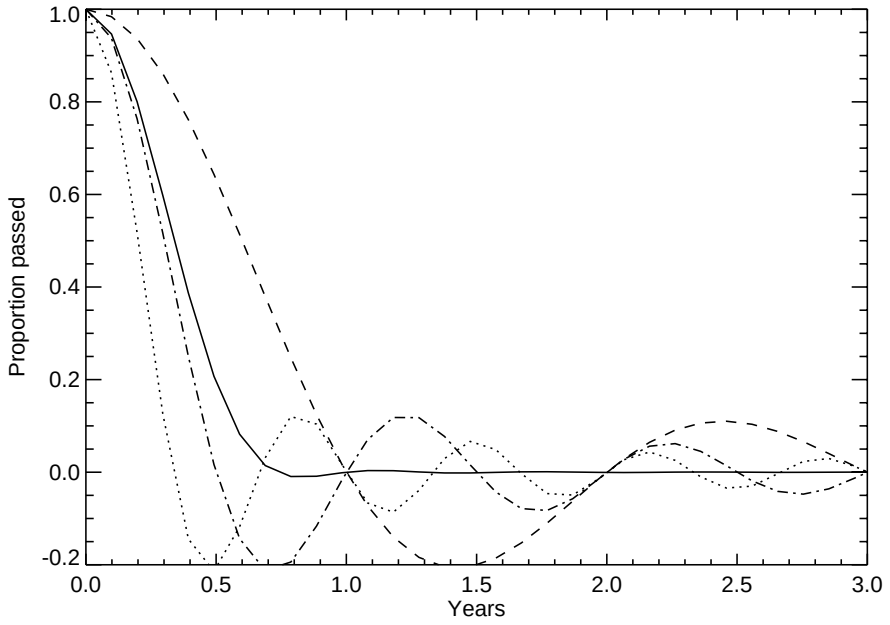


Figure 1: Frequency response of filter defined by $K = 18$, $\nu = 0.3$ case of (2.1.2a–c) (solid line) and running means of 12 months (dashed line), 24 months (chain curve) and 36 months (dotted curve).

$$\psi_k = \frac{\psi_0(2K+1)\sin(k\theta_c)\sin(2\pi k/(2K+1))}{2\pi\theta_c k^2} \qquad \text{for } 0 < |k| \leq K \qquad (2.1.2b)$$

with α determined by:

$$\sum_{k=-K}^{K} \psi_k = 1 \qquad (2.1.2c)$$

(The normalization coefficients in (2.1.2a,b) will ensure that $\alpha \approx 1$, with asymptotic equality for large K). In many of the examples in the remainder of this report, we will use the special case specified by $K = 18$ and $\nu = 0.3$ y^{-1} for $\Delta t = 1/12$ y. Figure 1 shows the frequency response for this filter along with the responses of 12-month, 24-month and 36-month running means. (Only the lower frequencies, less than 3 cycles y^{-1}, are plotted.)

In the examples below, running means will often be used when relating the discussion to the time domain and the 'Bloomfield' filters will be used for cases characterised in the frequency domain.

The use of digital filtering as a tool for estimation requires the resolution of a potential ambiguity. This is because the concept of 'averaging' arises both as a *transformation* of biogeochemical signals (e.g. the transformation from (1.1) to (1.3)) and averaging of observations as a technique for the estimation of such biogeochemical signals. If we have a time-varying biogeochemical 'signal', $g(t)$ then we may hope to estimate g by applying a 'filtering' operation "γ" to the data, i.e. $\langle\langle z\rangle\rangle_\gamma$, where (as shown in examples below) the filtering operation may be averaging, differentiation, deconvolution, or multi-component generalisations of these operations. The type of error analysis described in this report enables us to characterise how good $\langle\langle z\rangle\rangle_\gamma$ is as an

estimate of g. There is, however, another type of interpretation: $\langle\langle z\rangle\rangle_\gamma$ could be regarded as an estimate of $h(t) = \langle g\rangle_\beta$. If both the kernels γ and β involve similar time-averaging, then $\langle\langle z\rangle\rangle_\gamma$ could be much better (smaller mean square error) as an estimate of h, rather than an estimate of g. In general,

$$\mathrm{E}[(\langle\langle z\rangle\rangle_\gamma - g)^2] \neq \mathrm{E}[(\langle\langle z\rangle\rangle_\gamma - \langle g\rangle_\beta)^2] \qquad (2.1.3)$$

There is no unique 'correct' answer to this ambiguity: what is essential is to specify which quantity is being estimated.

In cases where the specific time-average of a signal, g, is defined by the context, we can use the simplified notation $\bar{g}$ to denoted such a time-average. In this notation, the distinction discussed here becomes one of distinguishing $\hat{g}$ (an estimate of g) from $\hat{\bar{g}}$ (an estimate of $\bar{g}$).

The following subsections apply the general framework described here to a range of cases. We start by calculating the propagation of known error statistics through a linear estimation procedure (Section 2.2). Section 2.3 looks at the special case where the linear estimation takes the form of a digital filter applied to a stationary time series. Direct propagation of errors may be impractical, particularly for non-linear estimation procedures. In such cases a Monte Carlo approach adding simulated noise to the observations provides an alternative (see Section 2.4). If the error distribution is unknown then the analysis will have to use error distributions derived from the data. There are some simple standard cases where the distribution is taken as known apart from a small number of parameters. The simplest is that of independent data, with identical normal distributions with unknown variance. For cases where there is even less information about the error distribution, bootstrap analyses (see Section 2.5) may provide a way of estimating uncertainties.

2.2 Error propagation

As noted above, the general form of statistical model that we use is an additive error model of the form:

$$z_k = g_k + n_k \qquad (2.2.1a)$$

where g_k is a 'signal' that is of interest and n_k is a residual 'noise' that is treated as random. We write the specific case for time series as

$$z(t) = g(t) + n(t) \qquad (2.2.1b)$$

We use the more general model (2.2.1a) to examine the simplest case which is that of linear estimation where we obtain estimates, $\hat{f}$, of some quantity, f, by taking linear combinations of the observations, z_k, with the chosen combination being the same regardless of the actual values of the z_k.

$$\hat{f}_j = \sum_k A_{jk} z_k \qquad (2.2.2)$$

This includes both the case $f = g$ where we wish to estimate g and the more general case where we wish to estimate a function f that is linearly related to g. A measure of whether $\hat{f}$ is a 'good'

estimate is given by

$$\mathrm{E}[(\hat{f}_k - f_k)(\hat{f}_j - f_j)] = [f_k - \sum_p A_{kp} g_p][f_j - \sum_q A_{jq} g_q] + \sum_{pq} A_{jp} A_{kq} \mathrm{E}[n_p n_q] \qquad (2.2.3a)$$

where $\mathrm{E}[.]$ denotes a statistical expectation. In particular, the mean square error of component k is

$$\mathrm{E}[(\hat{f}_k - f_k)^2] = [f_k - \sum_p A_{kp} g_p]^2 + \sum_{pq} A_{kp} A_{kq} \mathrm{E}[n_p n_q] \qquad (2.2.3b)$$

The second term in (2.2.3b) is the variance of the estimate and the first term is a squared bias. In special cases, it may be possible to *define* the relations so that the bias vanishes. For example if f is defined to be the 7-point running mean of g then the $K = 3$ case of the filter (2.1.1a) will give an unbiased estimate of f. However in some cases it may be desirable to accept biased estimates in order to reduce the variance (and the mean square error). The difficulty is that doing this effectively requires some knowledge of the unknown f.

A special case of estimation that is 'biased' (i.e. biased with respect to the observations, z_k) is Bayesian estimation. This is the case when there is statistical information about f, specified by an *a priori* probability distribution $\Pr(f)$.

The general form of Bayesian estimation is based on the relation:

$$\Pr(f|z) = \frac{\Pr(z|f)\Pr(f)}{\sum_f \Pr(z|f)\Pr(f)} \qquad (2.2.4)$$

This defines what we know about f given observations z. This (posterior) knowledge of f is given in terms of its (posterior conditional) probability distribution, $\Pr(f|z)$. Often we are mainly interested in $\mathrm{E}[f]$ and its variance, i.e. the first and second moments of $\Pr(f|z)$ — of course in the simple case where $\Pr(f|z)$ is Gaussian, these moments completely characterise the distribution. In the numerator on the right, the factor $\Pr(f)$ is the *a priori* distribution for f. The factor $\Pr(z|f)$ has its mean given by the (determinstic) relation $g(f)$ and its distribution about the mean given by the 'error' distribution of e_k. For fixed z, i.e. for a given data set, the denominator is a constant normalizing factor.

2.3 Frequency domain

An important special case arises with what are known as stationary time series. In these cases, the autocovariance for two times, t_1 and t_2, depends only on the time difference, $t_1 - t_2$, and not on the actual times, i.e. they can be expressed in terms of an autocovariance function, $R_a(t_1 - t_2)$ where

$$\mathrm{E}[a(t_1)a(t_2)] = R_a(t_1 - t_2) \qquad (2.3.1)$$

Such series can be characterised by their power spectra, $f_a(\theta)$, which are the Fourier transforms of the autocovariance functions, with θ a dimensionless frequency. Specifically:

$$f_a(\theta) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \cos(k\theta) R_a(k\Delta t) \qquad (2.3.2)$$

and

$$R_a(k\Delta t) = \int_{-\pi}^{\pi} \cos(k\theta)\, f_a(\theta)\, d\theta \tag{2.3.3}$$

For the case of stationary time series, the form of the statistical model in which observations are the sum 'signal-plus-noise' is:

$$z(t) = g(t) + n(t) \tag{2.3.4}$$

where $z(t)$ are the observations, $g(t)$ is a signal of interest (with power spectrum $f_g(\theta)$) and $n(t)$ represents unwanted noise (with power spectrum $f_n(\theta)$).

The standard way to obtain an estimate, $\hat{g}(t)$, of a stationary signal, $g(t)$, is through digital filtering:

$$\hat{g}(t) = \langle z \rangle_\gamma = \sum_k \psi_{k:\gamma} z(t + k\Delta t) \tag{2.3.5}$$

The mean-square-error of such an estimate is given by

$$\mathrm{E}[(\hat{g}(t) - g(t))^2] = \int_{-\pi}^{\pi} [|1 - \Psi_\gamma(\theta)|^2 f_g(\theta) + |\Psi_\gamma(\theta)|^2 f_n(\theta)]\, d\theta \tag{2.3.6}$$

where $\Psi(\theta)$ is the frequency response of the filter. In this expression, the first term in the integral is the mean-square bias due to distortion of the signal and the second term is the variance due to failure to completely eliminate the noise.

A standard result is that the optimal filter (i.e. the one that minimises the mean-square error 2.3.6) has the response:

$$\Psi_{\mathrm{opt}}(\theta) = \frac{f_g(\theta)}{f_g(\theta) + f_n(\theta)} \tag{2.3.7}$$

Applying this formalism requires a knowledge of the error statistics, expressed as the power spectrum $f_n(\theta)$. The difficulty is that from the observations, we can only obtain *estimates* of the *combined* power spectrum $f_z(\theta) = f_g(\theta) + f_n(\theta)$. This report is intended mainly as a survey of techniques of estimation and will not specifically address the issues of modelling the signal and noise. The 'noise' or 'error' models are simply taken from previous studies, primarily Cleveland et al. (1983).

These optimal filtering calculations can be regarded as inherently Bayesian since the signal is being treated as a random process, characterised by its power spectrum $f_g(\theta)$. In the absence of an a priori model of the signals of interest, the examples in this report do not include such optimal filtering. The discussion focuses on the propagation of 'noise' through the estimation process, this being quantified (in part) by the second (variance) term in integrand of (2.3.6). The time-domain equivalant is Equation (3.3).

At this point we need to recall the ambiguity, noted above, concerning the interpretation of digital filtering. If we put $\hat{g} = \langle\langle z \rangle\rangle_\gamma$ then the mean square error in the estimate of g is given by (2.3.6). If however we define

$$h(t) = \langle\langle g \rangle\rangle_\gamma \tag{2.3.8a}$$

and put

$$\hat{h}(t) = \langle\langle z \rangle\rangle_\gamma \tag{2.3.8b}$$

then

$$\hat{h}(t) = h(t) + \langle\langle n \rangle\rangle_\gamma(t) \tag{2.3.8c}$$

This is an unbiased estimate of $h(t)$ and so the first term in the integrand of (2.3.6) vanishes.

More generally, with

$$h(t) = \langle g \rangle_\beta \tag{2.3.9a}$$

and

$$\hat{h}(t) = \langle\langle z \rangle\rangle_\gamma \tag{2.3.9b}$$

the mean square error of the estimate is

$$\mathrm{E}[(\hat{h}(t) - h(t))^2] = \int_{-\pi}^{\pi} [|\Psi_\gamma(\theta) - \Psi_\beta(\theta)|^2 f_g(\theta) + |\Psi_\beta(\theta)|^2 f_n(\theta)]\, d\theta \tag{2.3.9c}$$

which is minimised by choosing ψ_γ to have the response

$$\Psi_{\gamma:\mathrm{opt}}(\theta) = \frac{f_g(\theta)\Psi_\beta(\theta)}{f_g(\theta) + f_n(\theta)} \tag{2.3.9d}$$

In cases where the filter, ψ_γ, used in constructing the estimates is close to the transformation ψ_β that defines the function of interest, the (first) 'bias' term in the integrand (2.3.9a) will be small and the mean square error will be dominated by the (second) 'variance' term.

Given the desirability of estimating smoothed signals, $\langle g \rangle_\beta(t)$, because these can often be estimated with less bias than $g(t)$, the question still remains of what type of smoothing should be used and how it should be described. The two possibilities are to describe the filtering in the frequency domain, taking $\langle g \rangle_\beta(t)$ as a band-pass-filtered transformation of $g(t)$ or in the time-domain, taking $\langle g \rangle_\beta(t)$ as a time-average (especially a running mean) of $g(t)$. These possibilities are captured by the two types of filter used in our examples. Intermediate cases will need to be characterised in terms of the specific filter involved.

We have introduced the concept of estimating a filtered function in terms of smoothing. This is applicable when lower frequencies have the best signal-to-noise ratio. However, Equations (2.3.9a–d) are applicable for more general filters, β. In particular, where there may be long-term biases that are constant, or very slowly varying, an analysis of functions subject to band-pass filtering may be appropriate. An example where the transformation ψ_β corresponds to numerical differentiation is given in Section 4.2.

2.4 Monte Carlo

There are a number of cases in which the direct propagation of errors (as in Section 2.2) is either impossible or impractical. The most obvious cases are those involving non-linear estimation. This includes robust estimation techniques (i.e. those with reduced sensitivity to departures from the normal distribution, particularly the presence of outliers). The analysis of the seasonal cycle by Cleveland et al. (1983) is an example of the use of simulation of errors to calculate the confidence intervals for robust estimation. Cleveland et al. (1983) used $n(t)$ defined as an AR(1) process (for which realisations are readily generated) in order to determine an envelope of uncertainty for their estimated cycle amplitudes.

The effect of noise on such non-linear estimates can be assessed by simulation. If the estimate of g is given by some function (or functional) of the observations z, i.e. $\hat{g}(t) = K[z(t)]$ then one can create an ensemble of realisations of alternative estimates as:

$$\hat{g}^{(m)}(t) = K[\hat{g}(t) + n^{(m)}(t)] \tag{2.4.1}$$

where $n^{(m)}(t)$ is a synthetic time series drawn from the appropriate error distribution.

Cases where we are estimating a functional $h(.)$ of $g(t)$ need to be treated with a 'two-stage' procedure where the estimation procedure for obtaining $\hat{h}(.)$ is applied to a set of realisations, $\hat{g}(t) + n^{(m)}(t)$, based on an estimate of $g(t)$ the 'signal' component of the observations.

As noted above, Cleveland et al. (1983) applied this technique to estimate the uncertainty for their seasonal cycle cycle amplitudes. However it is suggested in Section 3 below (and illustrated in Figure 5) that displaying a set of realisations can also provide a useful way of communicating the degree of possible variation about the estimate.

A final comment is that this Monte Carlo approach can also be valuable with estimation techniques that are actually linear but which are implemented as a 'black box' that calculates $\sum_k A_{jk} z_k$ given the z_k but which cannot readily calculate $\sum_{pq} A_{jp} A_{kq} R_{pq}$ given R_{pq}. The spline fits described in Section 5 are an example of such a case.

2.5 Bootstrap

The analyses above have assumed that the statistical distribution of the errors is known. Often this will not be the case, and it will be neccessary to estimate the errors from the observations. One of the simplest 'standard' cases of such a procedure is linear regression for the case where the data errors are assumed to be independant and have identical Gaussian distributions with unknown variance σ^2. In this case, σ becomes an additional parameter to be estimated. This is a particularly simple case, because the estimates of the regression coefficients will not depend on the value of σ.

More complicated cases can be dealt with by using a variety of 're-sampling' strategies such as the 'bootstrap' approach (see Efron, 1982). The idea is that if one has a data set, $A_0 =$

$\{z_1, z_2, \ldots, z_N\}$ from which an estimate $\hat{f}(A_0)$ is obtained, then the statistical properties of this estimate can be obtained from a sequence of estimates, $\hat{f}(A_m)$ derived from a sequence of synthetic data sets, A_m. These data sets are $A_m = \{z_{1:m}, z_{2:m}, \ldots, z_{m:N}\}$ where each synthetic value $z_{j:m}$ is obtained by randomly selecting an element of A_0, *with replacement.* (For large M, the probability of any particular point being chosen at least once approaches $1 - e^{-1}$).

This 'bootstrap' approach is used in Section 5 below to estimate the degree of uncertainty in spline fits to ice-core data. 'Resampling' as a concept rather than an operational procedure, also forms the basis of the method of generalized cross-validation (GCV) of smoothing splines. In GCV, the objective is not so much the characterisation of uncertainties, but rather the determination of the (statistically) optimal degree of smoothing, which, as illustrated by the analysis in Section 2.3 will depend on the nature of both the signal and the noise.

2.6 Error models

Table 1 summarizes a number of time series models that have been proposed for the 'irregular' or 'noise' component of monthly mean CO_2 data at Mauna Loa, and in one case for the South Pole. These models are to some extent based on different assumptions about what constitutes the 'regular' component(s). Thus some of the 'noise' variability captured by some of the error models may represent variations that other studies would regard as 'signal'.

Site	Model	Var(ϵ)	Reference
Mauna Loa	$n_{j+1} = 0.63n_j + \epsilon_j$	0.066	Cleveland et al. '83
South Pole	$n_{j+1} = 0.72n_j + \epsilon_j$	0.049	Cleveland et al. '83
Mauna Loa	ϵ_j	0.23	Surendran & Mulholland '87
Mauna Loa	$n_{j+1} = 0.48n_j + 0.18n_{j-1} - 0.12n_{j-2} + \epsilon_j$	0.063	Martín & Díaz '91

Table 1: Models of irregular, or 'noise' component of monthly mean CO_2 data.

Cleveland et al. (1983) represented the irregular components of the Scripps Institution of Oceanography (SIO) Mauna Loa and South Pole records as first-order autoregressive (AR(1)) processes as indicated in Table 1. They noted that this model did not capture the ENSO-scale variations in the long-term growth.

Surendran and Mulholland (1987) analysed the SIO Mauna Loa time series by fitting an exponential and 12-month and 6-month sinusoids. They then used an AR(2) model to capture the variability associated with the ENSO phenomenon. They suggested that the residuals could be represented as white noise with a variance of 0.2303.

Martín and Díaz (1991) investigated several statistical models of the SIO Mauna Loa time series. Their first model was based solely on the CO_2 time series and took the form of an ARIMA (autoregressive integrated moving average) model:

$$[1 - \phi_1\mathbf{B} - \phi_2\mathbf{B}^2 - \phi_3\mathbf{B}^3]\nabla_{12}\nabla[c(t) - \bar{c}] = [1 - \alpha_1\mathbf{B}^{12}][1 - \alpha_2\mathbf{B}]\epsilon(t) \qquad (2.6.1)$$

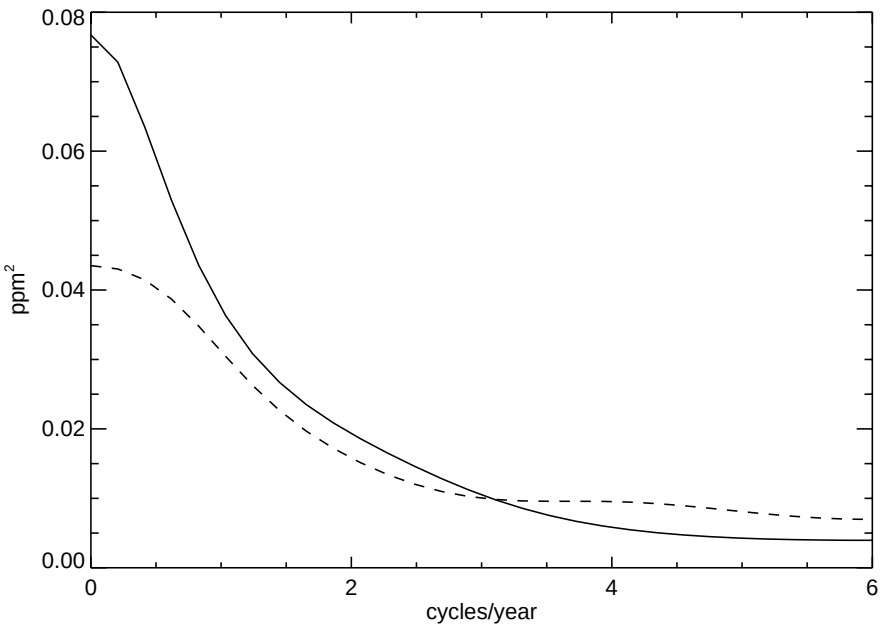


Figure 2: Power spectra for error models of monthly mean Mauna Loa time series, as described in Table 1 from Cleveland et al. (solid line) and Martín and Díaz (dashed line).

where $\mathbf{B}$ is the backward shift operator and ∇ is the difference operator. They also modelled the CO_2 data in terms of fossil fuel and sea-surface temperature and fitted the residuals as an AR(3) model:

$$[1 - 0.48\mathbf{B} - 0.18\mathbf{B}^2 + 0.12\mathbf{B}^3][n(t) - \bar{n}] = \epsilon(t) \qquad (2.6.2)$$

with $\mathrm{Var}[\epsilon] = 0.063$.

Tunnicliffe-Wilson (1989) modelled the monthly mean Mauna Loa CO_2 data series from NOAA as an ARMA (autoregressive moving average) model

$$[1 - \mathbf{B}][1 - \mathbf{B}^{12}]c(t) = [1 - 0.49\mathbf{B}][1 - 0.86\mathbf{B}]\epsilon(t) \qquad (2.6.3)$$

with $\mathrm{Var}[\epsilon] = 0.09$ but this was a model for the series as a whole and not just the irregular component.

Thoning et al. (1989) constructed an AR(1) model of the irregular component of *daily* mean data from the NOAA/GMCC Mauna Loa record. They did not quote the autocorrelation but, using the theoretical AR(1) power spectrum (Equation 2.6.4a below), it can be estimated from their Figure 6 as $a \approx 0.8$. This would imply an autocorrelation of order 0.001 for monthly data. Clearly the synoptic and subseasonal variation modelled by Thoning et al. is only a very small part of the variability identified by Cleveland et al.

Figure 2 shows some of the power spectra for these error models, plotted as a function of

frequency $\nu = \theta/2\pi\Delta t$. These use (see for example Priestley, 1981, Section 4.12.3)

$$f(\theta) = \frac{\mathrm{Var}[\epsilon]}{2\pi} \frac{1}{1 - 2a\cos(\theta) + a^2} \qquad (2.6.4a)$$

for AR(1) models, or the more general expression for AR(N):

$$f(\theta) = \frac{\mathrm{Var}[\epsilon]}{2\pi} \frac{1}{|1 - \sum_{j=1}^{N} a_j e^{ij\theta}|^2} \qquad (2.6.4b)$$

3 Representation of uncertainty

The following sections concentrate on the formal mathematical ways of characterising and calculating the variability inherent in statistical estimation of time-varying functions, using the framework developed in Section 2. The present section addresses the issues of how to communicate such results. This is done by presenting a number of examples.

EXAMPLE 1: The long-term trend at Mauna Loa

As an example we analyse the record of monthly mean CO_2 concentrations from Mauna Loa, Hawaii, as determined by the SIO measurement program (Keeling and Whorf, 1994, with additional data for later periods downloaded from the Carbon Dioxide Information and Analysis Center internet site at Oak Ridge National Laboratory). We treat the observed concentrations, $z(t)$, as consisting of a long-term trend, $g(t)$, a seasonal cycle, $s(t)$, and an irregular noise component, $n(t)$, i.e.

$$z(t) = g(t) + s(t) + n(t) \qquad (3.1)$$

The trend, which we seek to estimate, can be regarded as a global mean trend plus a contribution from medium to long-term variations in the mean spatial distribution of CO_2, primarily the latitudinal and vertical gradients. As noted in Section 2.6, Cleveland et al. (1983) modelled the SIO Mauna Loa CO_2 time series in this way with a separate signal describing the seasonal cycle and fitted the remaining irregular component to a first-order autoregressive (AR(1)) process with autocovariance $R_n(k\Delta t) = \sigma^2 a^{|k|}/(1 - a^2)$ with $\sigma = 0.257$ and $a = 0.63$.

In using digital filtering to construct an estimate, $\hat{g}(t)$, of the trend $g(t)$

$$\hat{g}(t) = \sum_k \psi_k z(t + k\Delta t) \qquad (3.2)$$

the autocovariance function for the estimate is

$$\mathrm{E}[(\hat{g}(t) - g(t))(\hat{g}(t') - g(t')] = \sum_{jk} \psi_j \psi_k R_n(t - t' + \Delta t(k - j)) \qquad (3.3)$$

(see for example Thoning et al., 1989, for examples of the use of such error estimation). The $t' = t$ case of (3.3) is the time-domain equivalent of the variance term in (2.3.6).

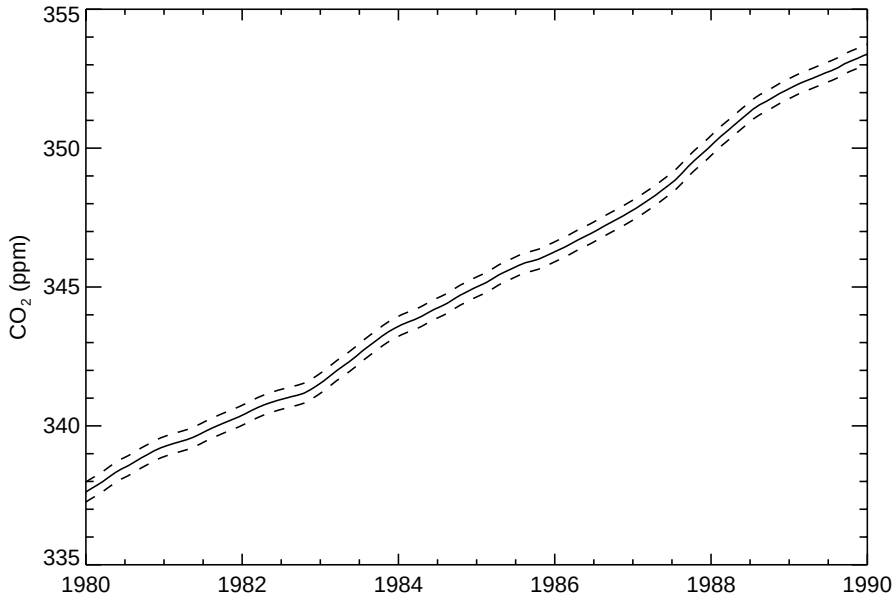


Figure 3: Long-term CO_2 trend at Mauna Loa, estimated by running mean of SIO data. The ± 2 s.d. range for the estimate is calculated using the Cleveland et al. error model. Tick marks for time on this and later plots are at the beginning of the indicated year.

We apply (3.2) with a 12-month running mean (actually $\psi_k = 1/12$ for $k = -5$ to 5 and $\psi_k = 1/24$ for $k = \pm 6$). Applying (3.3) for $t = t'$ gives $\mathrm{Var}[\hat{g}(t)] \approx 0.18^2$ ppm^2. Figure 3 shows the estimated trend with the 2 s.d. range of ± 0.36 ppm. However, while this range gives an indication of how much the 'true' trend might be expected to differ from the estimate, it does not show the possible patterns of departure. The covariance function for the estimates (Equation 3.3) contains this information. Figure 4 plots this autocovariance function for the trend estimated by the running mean, assuming the Cleveland et al. model of $n(t)$. The error estimates are highly correlated for periods of up to 6 months. This correlation time reflects the combined effects of the spread of the filter and the correlations in the 'noise' component. The information in Figure 4 could also be presented in the frequency domain as a power spectrum for the uncertainty in the estimate.

To convey more directly the information about the degree of possible variability due to uncertainties in the estimate we use the Monte Carlo procedure from Section 2.4. Figure 5 shows a set of curves produced by the Monte Carlo procedure (with each successive sample displaced by 0.5 ppm to enable the cases to be distinguished).

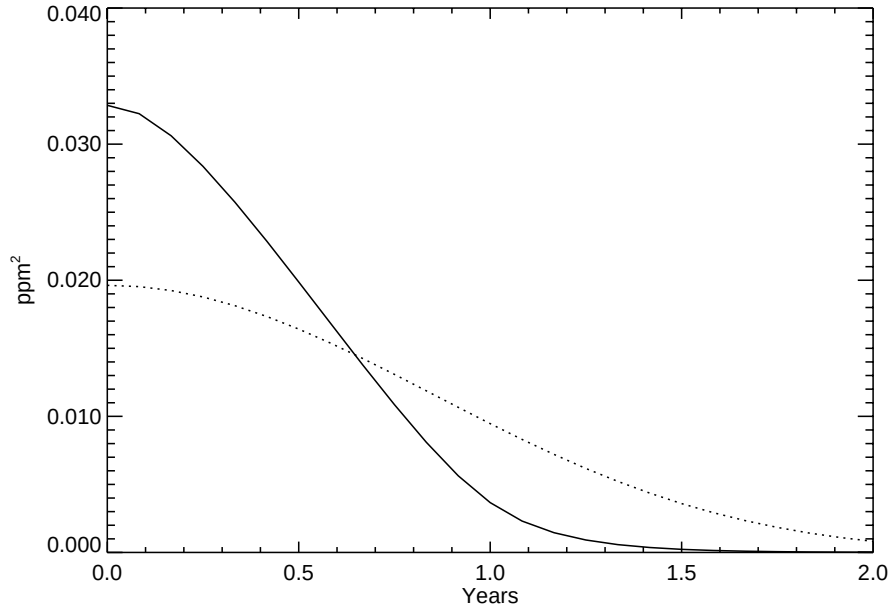


Figure 4: Autocovariance for estimates of the Mauna Loa CO_2 trend, calculated using Equation (3.3) with error model taken from Cleveland et al. (1983). The solid curve is for estimates calculation by the 12-month running mean (i.e. for the estimates shown in Figure 3. The dotted curve is for estimates obtained from the $K = 18$, $\nu = 0.3$ case of (2.1.2a–c).

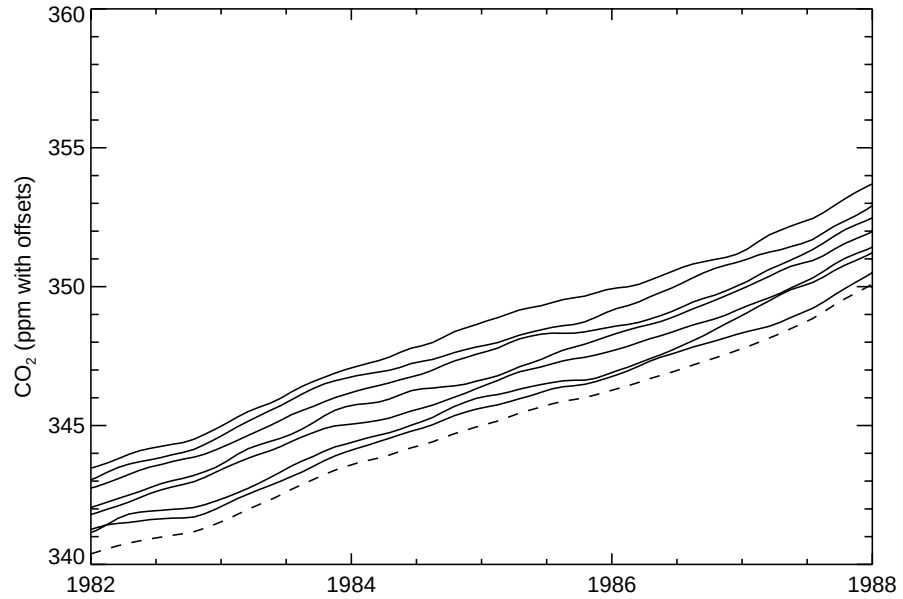


Figure 5: Estimate of Mauna Loa CO_2 trend (dashed curve) with sample of possible alternative functions. These are calculated by adding a synthetic 'noise' term, defined by the Cleveland et al. (1983) model and then applying the running mean filter. Each successive case is displaced by an additional 0.5 ppm, to separate the plots.

4 Functions and functionals

4.1 Issues

For the most part, the discussion in Section 2 was quite general and did not depend on whether the quantity that was being estimated was a particular component of the observed data or whether it was a more general function of some signal in the data. Nevertheless it seems worthwhile to address the issue of estimating functionals of signals in the observations, both as a general issue and in terms of some important specific cases.

Among the common principles involved in estimating functionals, we note that:

- explicit linear estimation allows explicit propagation of uncertainty. The discussion in Section 2.2 specifically included this degree of generality;
- for non-linear estimation (or linear problems that can not be readily manipulated) the Monte Carlo approach remains applicable.

Particular cases of estimating functionals are illustrated in the following subsections:

i. estimation of a time-averaged (i.e. smoothed) version of a signal (see Sections 2.1, 4.2 and 4.3);

ii. estimation of the growth rate (see Section 4.3);

iii. deconvolution of an observed signal, where the signal interpreted as the convolution of a forcing with a response function (see Section 4.4);

iv. estimation of seasonal cycles (see Section 4.5 and Enting et al., 1999).

v. extraction of a vector signal from multi-component data (see Section 4.6 for an example).

4.2 Running means

One particular transformation that is often of interest is the mean signal over some specified period. A prominent example is the IPCC discussion of the atmospheric carbon budget for the 1980s (Schimel et al., 1995). The time-average for a specific N-month period is, of course, just a single point from the N-month running mean and so the discussion of specific time-averages is just a special case of our general approach.

For the moment, we regard the averaging period (N months) as fixed and use ψ_β to denote this running mean of the signal. Thus

$$\bar{g} = \psi_\beta * g \qquad (4.2.1)$$

The actual estimates of $\bar{g}$ will be obtained from the data using a filter ψ_γ. The properties of a 'good' estimate are low variance (as calculated using Equation 3.3) and low bias, i.e. $\psi_\gamma \approx \psi_\beta$. However when ψ_γ is a running mean, the filter passes a signicant amount of high frequency power (see the examples in Figure 1) and thus may pass a significant amount of noise. However reducing this contribution to the variance of the estimates of $\bar{g}$ introduces a 'bias' as ψ_γ departs from ψ_β. The best trade-off is determined by the expressions (2.3.6) and (2.3.7) defining the optimal filter, but this requires an *a priori* estimate of the power spectrum of the signal.

To analyse the importance of reducing the variance in estimates of running means, we consider a specific class of estimates which we define as

$$\hat{\bar{g}} = \bar{\hat{g}} \qquad (4.2.2a)$$

where

$$\hat{g} = \psi_\alpha * z \qquad (4.2.2b)$$

This means that

$$\hat{\bar{g}} = \psi_\gamma * z \qquad (4.2.2c)$$

with

$$\psi_\gamma = \psi_\beta * \psi_\alpha \qquad (4.2.2d)$$

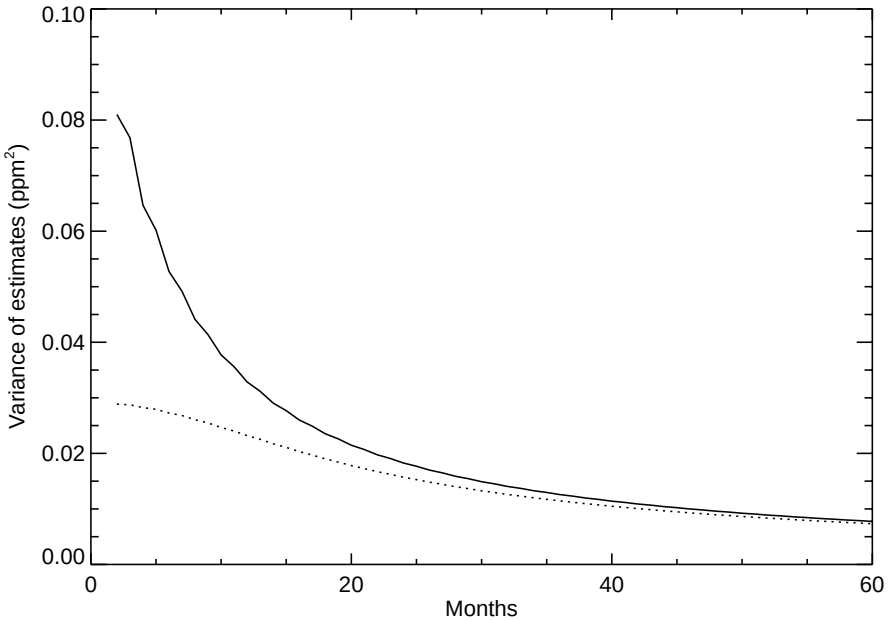


Figure 6: Variance when estimates of N-month running mean signal are derived derived by filtering observations which include noise characterised by Cleveland et al. AR(1) model. Solid curve shows variance (as a function of N) when N-month running mean signal is estimated as N-month running mean of observations. Dotted curve shows variance when N-month running mean signal is estimated as running mean of observations filtered with $K = 12$, $\nu = 0.5$ cycle y^{-1} 'Bloomfield' filter.

Figure 6 shows (as a plot against N) the variance of estimates obtained using an N-month running mean (solid curve) and the variance using the convolution of the running mean and a low-pass 'Bloomfield' filter (Equations 2.1.2a–c, with $K = 12$ and $\nu = 0.5$ cycle y^{-1}). For a fixed amount of additional smoothing (provided by ψ_α) this smoothing has progressively less effect on the variance of the estimate as the averaging period increases. (In practice, the running means of greatest interest are for whole numbers of years, so that the running mean acts to filter out the seasonal cycle). Recall that for even N we define the 'running mean' as a filter with $N + 1$ points, the two end-points having coefficients $1/2N$ (Equations 2.1.1b,c). This gives an extra smoothing for even N which leads to the odd-even oscillation apparent in the solid curve in Figure 6.

4.3 Rates of change

As noted in the introduction, a very important case where we need to consider transformations of observational data is when we require rates of change. The realisations of estimates of CO_2 concentrations shown in Figure 5 show that great variability and uncertainty in the growth rate can be accomodated within the confidence intervals shown in Figure 3.

A conceptual starting point for discussing rates of change is to note that, for smoothing defined by a convolution, the derivative of a smoothed function is equivalent to having smoothed (with the same kernel) the derivative of the original function. This can be seen by differentiating (1.2) as

$$\frac{\partial}{\partial t}\langle f(t)\rangle_\gamma = \int \psi_\gamma(t')\frac{\partial}{\partial t} f(t+t')\,dt' = \langle \frac{\partial}{\partial t} f(t)\rangle_\gamma \qquad (4.3.1)$$

As a specific example, the derivative of the 2-year mean is equivalent to the 2-year mean of the derivative, or in the abbreviated notation, $\bar{\dot{g}} = \dot{\bar{g}}$.

For discussing estimation from discrete data, it is useful to transform (4.3.1) as

$$\frac{\partial}{\partial t}\langle f(t)\rangle_\gamma = \int \psi_\gamma(t')\frac{\partial}{\partial t} f(t+t')\,dt' = \int \psi_\gamma(t')\frac{\partial}{\partial t'} f(t+t')\,dt' = -\int f(t+t')\frac{\partial}{\partial t'}\psi_\gamma(t')\,dt' \qquad (4.3.2)$$

assuming ψ_γ is differentiable on $[-T, T]$ and goes to zero at $\pm T$. Equation (4.3.2) expresses the smoothed derivative in terms of a kernel that differentiates and smooths the original function. This kernel is given by the negative of the derivative of the smoothing kernel. Discretised forms of such kernels can be used with Equation (3.3) to calculate the uncertainty of derivatives estimated from discrete data.

One important reason for explicitly considering estimation of derivatives is that the problem of numerical differentiation is well-known as an ill-conditioned inverse problem, i.e. one which is subject to large amplification of any errors in the input. While this might suggest that the derivative needs to be smoothed more than the concentration curve in order to obtain meaningful estimates, this argument needs to be treated with caution. If one takes the time derivative of the relation $z(t) = g(t) + n(t)$ then the derivatives of g and n would have power spectra $\theta^2 f_g(\theta)$ and

$\theta^2 f_n(\theta)$ respectively, so that from (2.3.7) the optimal estimate of $\dot{g}$ will be obtained by filtering $\dot{z}$ with the same filter that extracts the optimal estimate of g from z.

A special case that often occurs is that of estimating the time-average growth rate over some interval. Without loss of generality in the case of stationary times series, we take the intervals as $[-T, T] = [-M\Delta t, M\Delta t]$. Since we have:

$$2T\bar{\dot{g}} = \int_{-T}^{T} \dot{g}\, dt = g(T) - g(-T) \tag{4.3.3}$$

an obvious class of estimator of the mean derivative is:

$$\hat{\bar{\dot{g}}} = [\hat{g}(T) - \hat{g}(-T)]/2T \tag{4.3.4}$$

Using the analysis from Equations (2.3.8–9) with

$$\psi_{k:\beta} = [\delta_{k,M} - \delta_{-k,M}]/2M\Delta t \tag{4.3.5}$$

shows that the optimal estimate of the time-averaged growth rate of $g(t)$ is the difference of the optimal estimates of g, divided by the time-interval. The uncertainty in any estimate of the mean growth rate derived from differencing estimates of the function will be

$$\mathrm{Var}[\hat{\bar{\dot{g}}}] = [R_{\hat{g}}(0) - R_{\hat{g}}(2T)]/T \tag{4.3.6}$$

where $R_{\hat{g}}$ is the autocovariance function of $\hat{g}$, the estimate of g.

The analysis of growth rates using smoothing splines is discussed as a special case in Section 5.2 below.

EXAMPLE 2: The long-term CO_2 growth rate at Mauna Loa

In order to illustrate some of the above issues involved in estimating growth rates, we consider the rate of increase of CO_2 at Mauna Loa. We analyse the SIO data set (Keeling and Whorf, 1994, plus electronic update as noted above) and use the Cleveland et al. error model as in Example 1. We consider an initial smoothing of the record defined by the Bloomfield low-pass filters with coefficents ψ_k specified by Equations (2.1.2a–c) with $K = 18$ and $\nu_f = 0.3$ cycles y^{-1}, i.e. $\omega_f = 0.6\pi$ y^{-1}. The filter that generates the estimates of the (smoothed) derivative is given by

$$\psi'_k = (\psi_{k-1} - \psi_{k+1})/2\Delta t \tag{4.3.7}$$

Figure 7 shows the estimated growth rate, together with the ± 1 s.d. and ± 2 s.d. uncertainty ranges (i.e. point-by-point uncertainties). Figure 8 shows the autocovariance of these estimates (so that the range in the previous figure are $\pm\sqrt{R(0)}$ and $\pm 2\sqrt{R(0)}$) . The curve is obtained from (3.3) using (4.3.7) for the filter and the Cleveland AR(1) model for $R(.)$. It will be noticed that there is a prominant negative autocorrelation beyond 1 year. This is to be expected in estimates of derivatives. Positive errors in the estimated derivative must be followed by compensating negative errors (and vice versa) to keep the integral of the estimated derivative consistent with the confidence intervals on the original function.

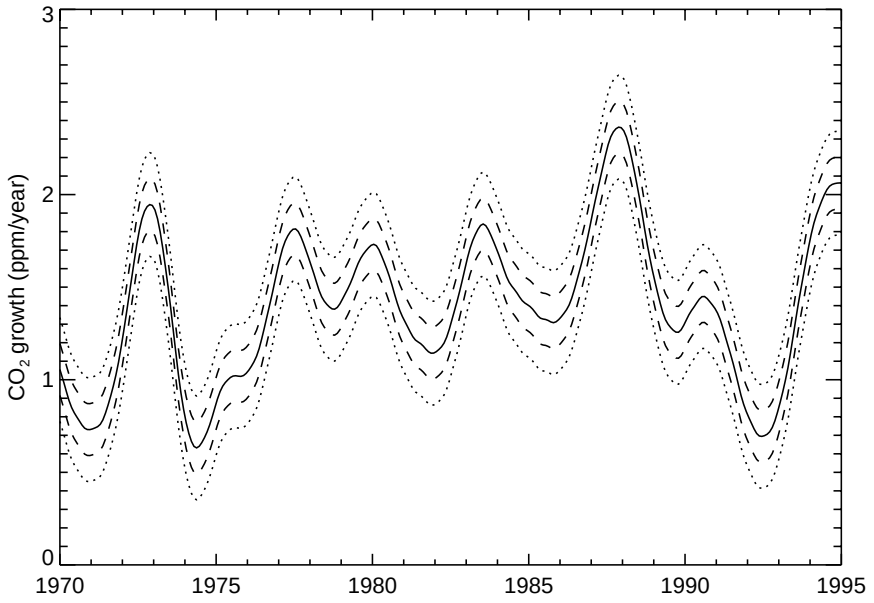


Figure 7: Estimated CO_2 growth rate at Mauna Loa based on 2-month differencing of concentration estimates derived from $K = 18$, $\nu = 0.3$ case of (2.1.2a–c) (solid curve), with point-wise uncertainties at ± 1 s.d. (dashed curves), and ± 2 s.d. (dotted curves).

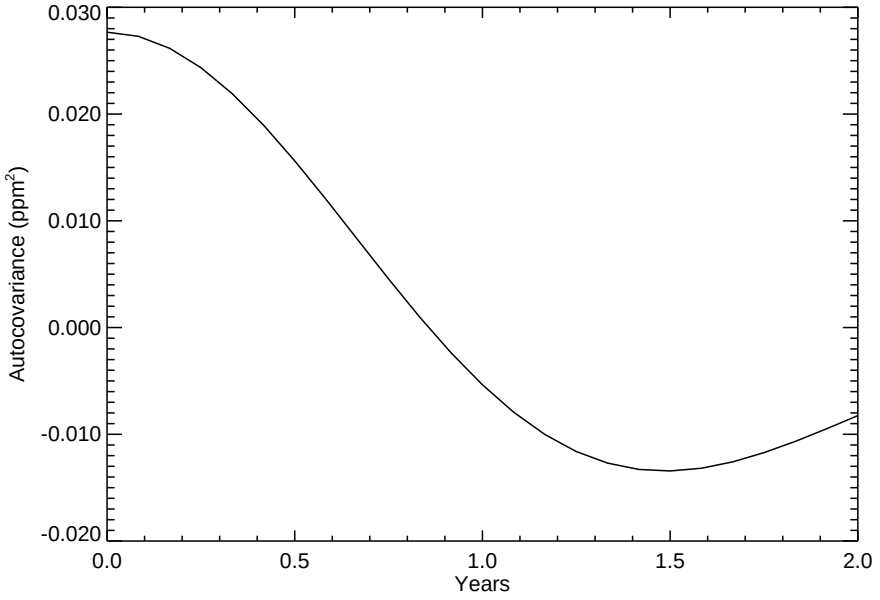


Figure 8: Autocovariance for estimates of CO_2 growth rate at Mauna Loa (as in Figure 7). Calculated using (3.3) with Cleveland et al. error model and 2-month difference (Equation 4.3.7) of $K = 18$, $\nu = 0.3$ case of (2.1.2a–c).

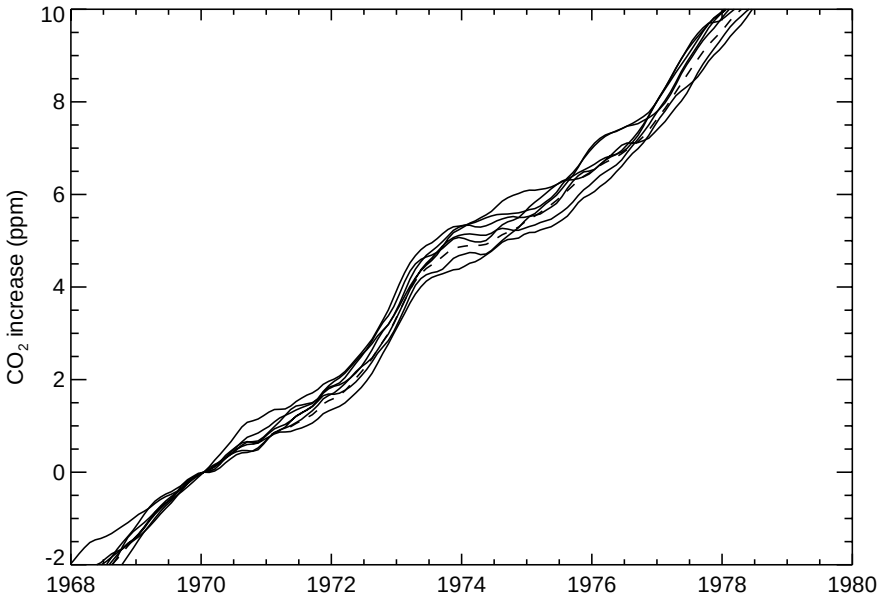


Figure 9: Multiple realisations of CO_2 concentration estimates (using 12-month running mean), normalised to the (mid-January) 1970 value. The changes for N-years from 1970 correspond to N times the N-year mean growth rate for the period 1970–1970+N

Figure 9 illustrates the issues described by Equation (4.3.6) involved in estimating time-averages of rates of change. The curves are Monte Carlo realisations of 12-month running means of the Mauna Loa CO_2 record, normalized to January 1970. The time-average for the period 1970 to 1970+τ is given by τ^{-1} times the increase from 1970. As described by (4.3.6) there is an overall tendency for the variance to decrease as τ increases, due to the τ^{-1} averaging factor. For small τ, the $-R_{\hat{g}}(\tau)$ term partly offsets this tendency since $R_{\hat{g}}(\tau)$ is largest at small τ — the autocorrelation in the estimates keeps the curves in Figure 9 together for the initial period after 1970.

4.4 Deconvolutions

An early application of response functions to the study of the carbon budget was by Oeschger and Heimann (1983) who expressed the atmospheric carbon content, $M(t)$, in terms of the sources, $S(t)$, as

$$M(t) = M_0 + \int^t G(t - t')\, S(t')\, dt' \qquad (4.4.1)$$

where $G(t)$ represents the response of the natural systems and is the amount of a unit input to the atmosphere that remains in the atmosphere after a time t.

Enting and Mansbridge (1987) noted that (4.4.1) has a formal inverse of the form:

$$S(t) = \frac{d}{dt}M(t)/G(0) - [M(t) - M_0]\,\dot{G}(0)/G(0)^2 - \int K(t-t')\,[M(t') - M_0]\,dt' \quad (4.4.2)$$

where the kernel $K(.)$ can be derived from $G(.)$. If $G(.)$ is a sum of n exponentials, then $K(.)$ is a sum of $n-1$ exponentials.

Relation (4.4.2) corresponds to

$$S(t) = \frac{d}{dt}M(t)/G(0) - \Phi_{\rm response}(t) \quad (4.4.3)$$

where $\Phi_{\rm response}$ is the flux associated with the response of the natural system. This flux is a functional of $M(t)$.

A key result is that if the time-averaging takes the form of a convolution, then (4.4.1), which is conceptually an instantaneous relation, can be consistently averaged as

$$\langle M(t)\rangle_\gamma = M_0 + \int_{-\infty}^{t} G(t-t')\,\langle S(t')\rangle_\gamma\,dt' \quad (4.4.4)$$

The result (4.4.4) is a particular case of the general result that convolution operations commute. The specific case is derived by noting that, in our notation, taking a time average of such a convolution gives

$$\begin{aligned}\langle \int_{-\infty}^{t} G(t-t'')\,S(t'')dt''\rangle_\gamma &= \int_{-T}^{T}[\psi_\gamma(t')\int_{-\infty}^{t+t'} G(t+t'-t'')\,S(t'')dt'']dt' \\ &= \int_{-T}^{T}[\psi_\gamma(t')\int_{-\infty}^{t} G(t-\tau)\,S(\tau+t')d\tau]dt' \quad (4.4.5)\end{aligned}$$

with $\tau = t'' - t'$. Thus

$$\langle \int_{-\infty}^{t} G(t-t'')\,S(t'')dt''\rangle_\gamma = \int_{-\infty}^{t} G(t-t'')\,\langle S(t'')\rangle_\gamma dt'' \quad (4.4.6)$$

Most deconvolutions of carbon cycle data are performed using (4.4.3) with a numerical model to calculate $\Phi_{\rm response}$. This is commonly done by running a model in 'inverse mode' with the atmospheric concentration forced to track a specified concentration record $C(t)$ (see for example Siegenthaler and Oeschger, 1987; see also discussion in Enting et al., 1994). The significance of the formal solution (4.4.2) is that it emphasises the fact that the function $M(t)$ occurs in more than one term in the expression for $S(t)$ and so the errors in these terms cannot be regarded as independent. Cases where response functions have been used more directly in carbon cycle studies are in the model developed by Wigley (1991), IPCC modelling by the group at the Max Planck Institut für Meteorologie (see Enting et al., 1994), and the Bern group (Joos et al., 1996; Bruno and Joos, 1997; Joos and Bruno, 1998) and in the Kalman filtering analysis by Trudinger (1999).

Many deconvolution studies of CO_2 have been based on smoothing spline fits to observational data, particularly from ice cores. However the analysis in this section suggests that the main

uncertainties arise from the need for differentiation. Therefore, when discussing smoothing splines in Section 5 below, we concentrate on the determination of growth rates (Section 5.2) and do not explicitly address deconvolution studies.

EXAMPLE 3: Deconvolution of methane data

In order to illustrate the issues discussed above, we consider the problem of deducing the total source of atmospheric methane, this being somewhat simpler than the corresponding calculations for CO_2. The budget equation is written as

$$S(t) = \frac{d}{dt}M(t) + M(t)/\tau \tag{4.4.7}$$

where τ is an atmospheric (chemical) lifetime, which we set to 10 years as used in the IPCC Radiative Forcing Report by Prather et al. (1995). Note that this is a 'whole atmosphere lifetime' and not an 'adjustment time', because the budget refers to the whole atmosphere and not just a perturbation. Equation (4.4.7) can be seen to be a special case of (4.4.2) with $G(t) \propto \exp(-t/\tau)$, since $K(.)$ vanishes if $G(.)$ is a single exponential.

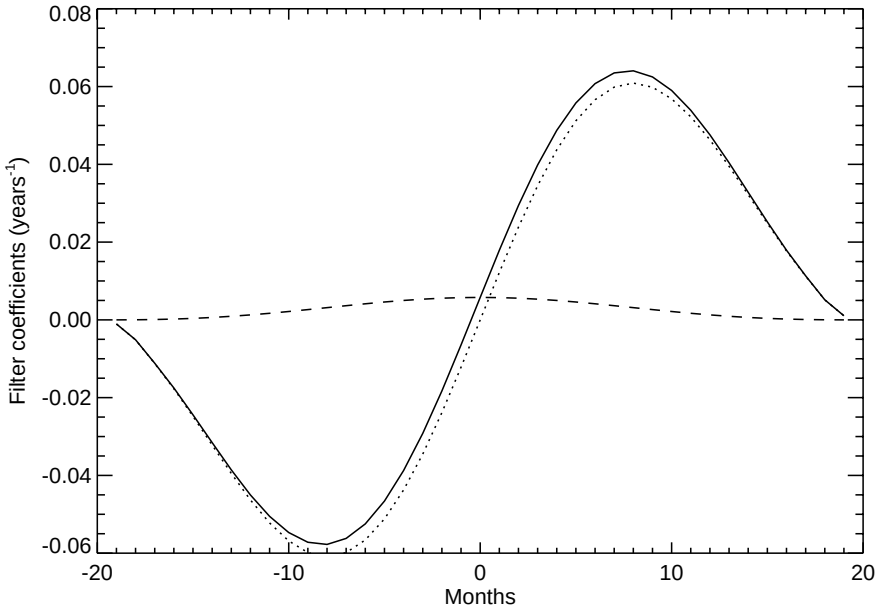


Figure 10: Deconvolution kernel for methane source estimates (solid curve) and the two terms in the (smoothed) budget (4.4.8): the derivative (dotted curve) and the M/τ (dashed curve). The curves are shown as continuous for clarity, but the filter coefficients are defined only for integer values.

We consider the time average

$$\langle S(t)\rangle_\gamma = \langle \frac{d}{dt}M(t)\rangle_\gamma + \langle M(t)\rangle_\gamma/\tau \tag{4.4.8}$$

with the averaging using the same filter as in Example 2.

This implies that we estimate $\langle S(t)\rangle_\gamma$ with a deconvolution kernel defined by

$$\psi_k^* = (\psi_{k-1} - \psi_{k+1})/2\Delta t + \psi_k/\tau \tag{4.4.9}$$

Figure 10 shows the form of this filter, based on a 'Bloomfield' filter (Equations 2.1.2a–c above) with $K = 18$ and $\nu_c = 0.3$ cycles per year. The solid curve plots the filter coefficients, the dotted curve is the filter defining the derivative and the dashed curve is the filter defining the (smoothed) M/τ term. It is clear that in this case, the uncertainties in the estimates obtained by deconvolution will be dominated by the uncertainties in estimating the derivative. For CO_2 where the time constants are longer than the atmospheric lifetime of CH_4, the uncertainties in deconvolutions will be dominated to an even greater extent by the uncertainties inherent in numerical differentiation.

4.5 Seasonal cycles

One of the most prominant signals in atmospheric CO_2 concentrations is the seasonal cycle. This is the largest signal after the long-term trend, especially in the northern hemisphere. The cycle and its variability is analysed in a forthcoming report by Enting et al. (1999) and so only a summary is presented here. The technique of complex demodulation (Bloomfield, 1976; Priestley, 1981, Section 11.2.2) is used to obtain an estimate of the principle component, $s(t)$, of the cycle (frequency $\omega_A = 2\pi$ y^{-1}) as $s(t) = A(t)\cos(\omega_A t - \phi(t))$ with $A(t)$ and $\phi(t)$ slowly varying.

The technique of estimating amplitude and phase of the cycle by complex demodulation is non-linear. Therefore, the Monte Carlo technique outlined in Section 2.4 is a convenient way of determining the range of uncertainty in the estimates. However, in spite of the non-linearity, Enting et al. (1999) were able to obtain analytic expressions that gave an approximate error analysis. The approach was to follow Enting (1987b) and note that the technique of complex demodulation can be expressed as a digital filtering problem. Consequently, the formalism in Section 2.3 above can be applied to the analysis of the cycle. The basis of the equivalence is that if

$$\hat{s}(t) = \hat{A}(t)\cos(\omega_A t - \hat{\phi}(t)) \tag{4.5.1}$$

then the complex demodulation estimates for $\hat{A}$ and $\hat{\phi}$ were equivalent to defining the estimate $\hat{s}$ as

$$\hat{s}(t) = \sum_k \psi_{k:c} z(t + k\Delta t) \tag{4.5.2}$$

The error analysis is based on combining this with the estimate, $\hat{r}$, of the phase-shifted cycle

$$\hat{r}(t) = \hat{A}(t)\sin(\omega_A t - \hat{\phi}(t)) \tag{4.5.3a}$$

which is given by

$$\hat{r}(t) = \sum_k \psi_{k:s} z(t + k\Delta t) \tag{4.5.3b}$$

where the ψ_k are the coefficients of the low-pass filter ψ used in complex demodulation and $\psi_{k:s} = \sin(k\theta_A)\psi_k$ and $\psi_{k:c} = \cos(k\theta_A)\psi_k$ are transformations of this low-pass filter. Given a stationary error model, we can apply (3.3) to determine the (constant) variances of the estimates $\hat{r}$ and $\hat{s}$. From the definitions of ψ_c and ψ_s we have $\Psi_s(\theta_A) = \sqrt{-1}\,\Psi_c(\theta_A)$. If ψ is highly smoothing (i.e. $\Psi(\theta)$ is non-zero for only a narrow band of low frequencies) then $\Psi_c(\theta)$ and $\Psi_s(\theta)$ are non-zero only in a small range of frequencies around θ_A and so $|\Psi_c(\theta)| \approx |\Psi_s(\theta)|$ for all frequencies, and so we have (approximately) $\mathrm{Var}[\hat{s}] = \mathrm{Var}[\hat{r}] = \sigma^2$. For ψ_k symmetric, $\mathrm{Cov}[\hat{s}, \hat{r}]$ is zero by symmetry.

Therefore we can write

$$\hat{A}^2 = \hat{s}^2 + \hat{r}^2 = (A\cos(\omega_a t) + x)^2 + (A\sin(\omega_A t) + y)^2 \qquad (4.5.4a)$$

where x and y are normally distributed with variance σ^2 whence

$$\mathrm{E}[\hat{A}^2] = A^2 + 2\sigma^2 \qquad (4.5.4b)$$

Enting et al. (1999) use a binomial expansion to show that

$$\mathrm{E}[\hat{A}] = A + \text{ a bias of order } \sigma^2/2A \qquad (4.5.4c)$$

whence

$$\mathrm{Var}[\hat{A}] \approx \sigma^2 \qquad (4.5.4d)$$

EXAMPLE 4: The seasonal cycle at Mauna Loa

The seasonal cycle of CO_2 is a prominent feature of northern hemisphere records. Over the 4 decades of the Mauna Loa record, it has been apparent that the amplitude of the cycle is changing. A number of studies have expressed this change as a long-term trend, but a more detailed analysis suggests that most of the longer-term changes come from periods of relatively rapid change in the late 1970s and late 1980s (Thompson et al., 1986). Enting et al. (1999) have extended the complex demodulation analysis to a large number of CO_2 records. In this section we present the Mauna Loa case as an illustrative example.

Again, the data set and error model are those used in Example 1. The filter ψ is specified by (2.1.2a–c) with $K = 18$ and $\nu_c = 0.3$ cycles per year. Applying (3.3) gives $\mathrm{Var}[\hat{s}] \approx \mathrm{Var}[\hat{r}] \approx 0.01884$ so that the standard deviation of the amplitude estimates is 0.14 ppm. Figure 11 shows the estimated amplitude for Mauna Loa. Studies by Enting et al. (1999) indicate that the range of estimates obtained by different observational programs at Mauna Loa is consistent with this uncertainty range. This suggests that a significant fraction of the noise is due to measurement and sampling procedures rather than coherent atmospheric variability.

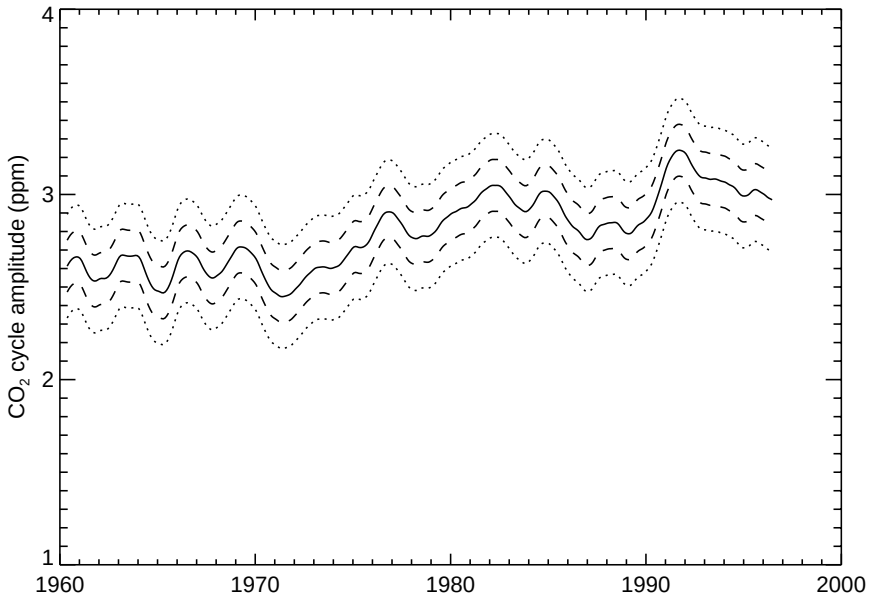


Figure 11: Estimate of the amplitude of the seasonal cycle of CO_2 at Mauna Loa, with ± 1 s.d. and ± 2 s.d. uncertainty ranges.

4.6 Isotopes

The basis of the use of isotopes in biogeochemical studies is the relation between elemental fluxes and isotopic fluxes. We represent the carbon balance of the atmosphere as

$$\frac{d}{dt}M = \sum_Y \Phi_Y \qquad (4.6.1a)$$

where the sum is over all processes, Y. The corresponding ^{13}C balance is

$$\frac{d}{dt}(MR_A) = \sum_Y R_Y \Phi_Y \qquad (4.6.1b)$$

where R_A is the isotopic ratio of the atmosphere and R_Y is the (average) isotopic ratio of flux Φ_Y. (The subdivision into processes, Y, has to be sufficiently detailed to allow meaningful averaging of isotopic ratios to obtain R_Y.) Often it is possible to obtain good estimates of the isotopic ratios R_Y (or equivalently the δ_Y defined below) and, assuming that the R_Y are known and are not all equal, Equation (4.6.1b) gives additional constraints on the fluxes Φ_Y.

The ^{13}C budget can be re-expressed as a linear combination of Equations (4.6.1a,b) in terms of a ^{13}C anomaly M_x^* defined by

$$M_x^* = \alpha(M_x^{13}/R_r - M_x) = \delta_x M_x \qquad (4.6.2)$$

where R_r is a reference ratio and α is a scale factor that defines the units of δ_x. The anomaly budget is given by

$$\frac{d}{dt}M_x^* = \frac{d}{dt}(M\delta_A) = \sum_Y \delta_Y \Phi_Y \qquad (4.6.3)$$

in terms of anomaly fluxes

$$\Phi_Y^* = \alpha(\Phi_x^{13}/R_r - \Phi_Y) = \delta_Y \Phi_Y \tag{4.6.2}$$

$$\frac{d}{dt}(M\delta_A) = \sum_Y \delta_Y \Phi_Y \tag{4.6.3}$$

using

$$\delta_Y = \alpha(R_Y - R_r)/R_r \tag{4.6.4}$$

The notation δ_Y is meant to be suggestive of the δ^{13}C used to communicate measurements of isotopic ratios. (The principle is essentially that of Tans (1980) treating an anomaly, $\delta_X M_X$, as a conserved tracer.) The correspondence between δ from (4.6.4) and the conventional δ^{13}C will not be exact and will depend on the choice of the constants R_r and α used in the definitions. If we denote the 'standard' definition of δ^{13}C (in units of ‰) by δ' and ^{13}C:^{12}C ratios by r, we have

$$r = (1 + \delta'/1000)r_s \tag{4.6.5a}$$

where r_s is the standard ^{13}C:^{12}C ratio of a material with δ^{13}C = 0. The anomaly δ values defined here are

$$\delta = \alpha \left[\frac{r}{1+r}/R_r - 1 \right]$$

Putting $R_r = r_s/(1 + r_s)$ gives

$$\delta = \alpha \frac{r - r_s}{(1+r)r_s} \tag{4.6.5b}$$

Putting $\alpha = 1000(1+r)$ would make δ exactly equivalant to the standard definition of δ' (in ‰). However, we need to use the same α value for all terms in Equation (4.6.3) and so the factor $1+r$ needs to be replaced by a fixed value. If all the δ values lie between -10 and -30 then using the central value and replacing $1 + r$ by $1 + 0.98r_s$ will lead to errors of only 1%. Therefore, in all the calculations listed below, we substitute δ' values in place of the corresponding δ values.

When considering carbon isotopes, we need to recall that the anthropogenic perturbation is a small change in a large natural cycle with gross fluxes, $\Phi^-_{\rm Ocean} \approx \Phi^+_{\rm Ocean}$ and $\Phi^-_{\rm Biota} \approx \Phi^+_{\rm Biota}$, to and from oceans and biota. The net fluxes $\Phi_{\rm Ocean} = \Phi^+_{\rm Ocean} - \Phi^-_{\rm Ocean}$ and $\Phi_{\rm Biota} = \Phi^+_{\rm Biota} - \Phi^-_{\rm Biota}$ are small residuals of the two opposing gross fluxes.

The net isotopic change due to two opposing gross fluxes can be re-expressed as:

$$\Phi_X^+\delta_X^+ - \Phi_X^-\delta_X^- = (\Phi_X^+ - \Phi_X^-)\delta_X^+ + \Phi_X^-(\delta_X^+ - \delta_X^-) \tag{4.6.6}$$

In (4.6.6) the first term on the right is the isotopic signal carried by the net flux and the second term is an isotopic signal (an 'isoflux') due to isotopic disequilibrium. Note that the δ_X^+, δ_X^- refer to the isotopic compostion of the fluxes and that if these are composite fluxes the δ_X^+, δ_X^- need

to be defined as flux-weighted averages. Often the isotopic composition of exchanges between reservoirs will be offset from the isotopic composition of the source reservoir, so a transfer $\Phi_{P\to Q}$ between reservoirs P and Q will have an isotopic composition given by

$$\delta_{P\to Q} = \delta_Q + \eta_{P\to Q}$$

where $\eta_{P\to Q}$ specifies the isotopic fractionation in the $P \to Q$ transfer.

EXAMPLE 5: Joint CO_2-$^{13}CO_2$ budget

This example reviews the uncertainty analysis in the calculations of Francey et al. (1995a). This study is important is its own right as an indication of the variability of the carbon cycle. The calculation is also important because it is essentially equivalent to the double deconvolution method that is applied to the analysis of CO_2 and ^{13}C records from ice cores. The global constraint from ^{13}C in synthesis inversions, both steady-state (Enting et al., 1995) and time-dependent (Rayner et al., 1999), comes from essentially the same set of equations. For CO_2, the specific cases of (4.6.1a and 4.6.3) that are of most interest take the form:

$$\frac{d}{dt}M = \Phi_{\rm Fossil} + \Phi_{\rm Ocean} + \Phi_{\rm Biota} \qquad (4.6.7a)$$

$$\frac{d}{dt}(M\delta_A) = \delta_F\Phi_{\rm Fossil} + \delta_O^+\Phi_{\rm Ocean} + \delta_B^+\Phi_{\rm Biota} + \Phi_{\rm Ocean}^-(\delta_O^+ - \delta_O^-) + \Phi_{\rm Biota}^-(\delta_B^+ - \delta_B^-) \quad (4.6.7b)$$

Enting et al. (1993) represented these equations in the form of a sum of two-component vectors $[a, b]$ with a in ‰GtCy^{-1} and b in GtCy^{-1}. The vector sum equivalent to (4.6.7a,b) is

$$\begin{aligned} -\dot{M}[1,\delta_A] - [0, M\,\dot{\delta}_A] + \Phi_{\rm Fossil}[1,\delta_F] + \Phi_{\rm Biota}[1,\delta_B^+] + \Phi_{\rm Ocean}[1,\delta_O^+] \\ +[0,\Phi_{\rm Ocean}^-(\delta_O^+ - \delta_O^-)] + [0,\Phi_{\rm Biota}^-(\delta_B^+ - \delta_B^-)] = [0,0] \end{aligned} \qquad (4.6.7c)$$

A graphical representation of this relation is shown in Figure 12.

The simultaneous Equations (4.6.7a,b) can be solved explicitly so that (eliminating $\Phi_{\rm Biota}$):

$$\Phi_{\rm Ocean}(\delta_O^+ - \delta_B^+) = \dot{M}(\delta_A - \delta_B^+) + M\dot{\delta}_A - \Phi_{\rm Fossil}(\delta_F - \delta_B^+) - \Phi_{\rm Ocean}^-(\delta_O^+ - \delta_O^-) - \Phi_{\rm Biota}^-(\delta_B^+ - \delta_B^-) \qquad (4.6.7d)$$

We now apply the concepts developed earlier in this report to re-examine how well we can estimate $\Phi_{\rm Ocean}$ on various time-scales, working through (4.6.7d) term by term. Francey et al. (1995a) noted a distinction between two types of error. Firstly there was a short-term 'noise' that they aimed to remove by time-averaging. Time-averaging was also proposed as a solution to the issue of spatial representativeness (see Section 6.2 below). Secondly, there was the possibility of unknown long-term biases (particularly in the isoflux terms) that, it was argued, were sufficiently slowly-varying not to affect the signal on sub-decadal time-scales, but which might give a mean offset to the estimates.

The error analysis of (4.6.7d) is complicated by the appearence of δ_B^+ in multiple terms and the apparent non-linearity as the δ factors multiply mass terms. Such non-linearity can, in principle,

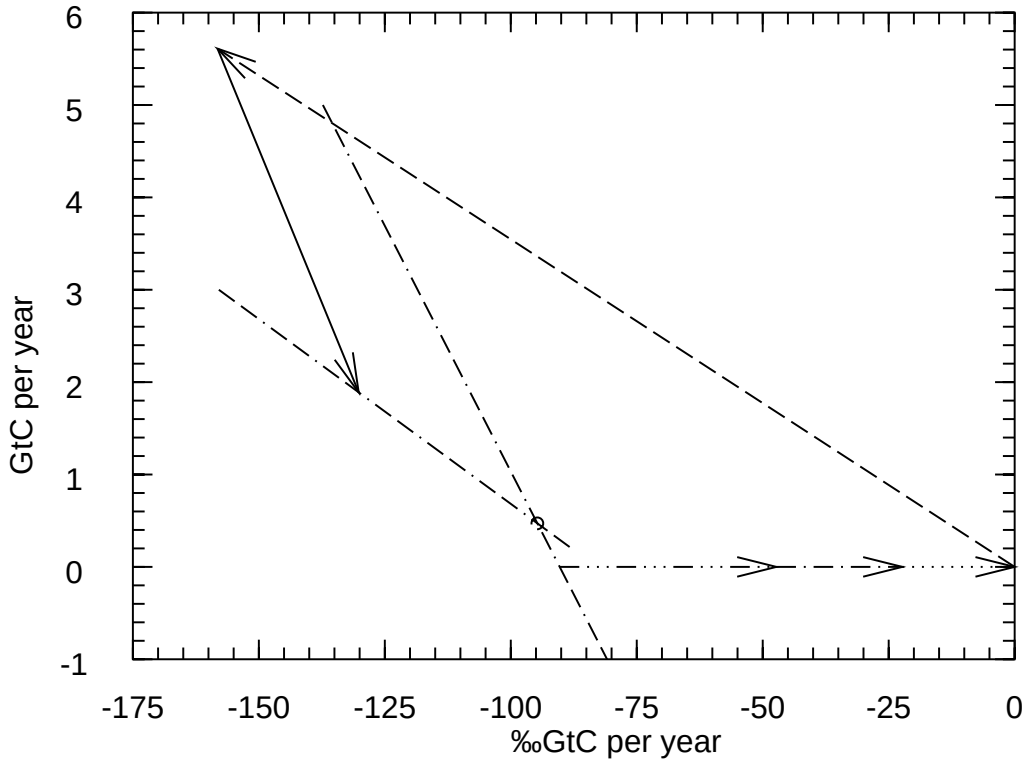


Figure 12: Vector representation of joint CO_2-$^{13}CO_2$ budget. Arrows are: — $\dot{M}$; ..... $\dot{\delta}_2$ – – – $\Phi_{\rm Fossil}$ – · – . biotic disequilibrium – · · · – · · · oceanic disequilibrium. The solution for $\Phi_{\rm Biota}$ and $\Phi_{\rm Ocean}$ is determined by the intersection of the two lines – – · of known slope (i.e. known δ_X) and unknown length, representing net ocean fluxes and net biotic fluxes.

lead to covariance effects when time-averaged. To resolve the role of the non-linearity it is easiest to go back to the original (linear) equations of the budgets of carbon and ^{13}C anomaly and to time-average *before* transforming the equations.

In these terms we can *define* $\bar{\delta}_Y$ as

$$\bar{\delta}_Y = \bar{\Phi}^*_Y / \bar{\phi}_Y \tag{4.6.8}$$

In other words, as with averages over component fluxes, time-averaged δ values need to be defined as flux-weighted time-averages.

In order to address the issue of non-linearity we note that δ^+_B arises when choosing a multiple of Equation (4.6.7a) that cancels $\Phi_{\rm Biota}$ in (4.6.7b). For the purposes of this analysis we modify this and subtract a multiple of (4.6.7a) specified by $\delta_{\rm ref} \approx \delta^+_B$ without requiring exact equality. The results of uncertainty in δ^+_B then appear as a multiple of $\Phi_{\rm Biota}$. This would vanish if we had exact equality, but that given $\delta_{\rm ref}$ is only an approximation to δ^+_B there is an error contribution $\propto \Phi_{\rm Biota}$. Combining the use of $\delta_{\rm ref}$ with the use of the flux-weighted time-average δ values

defined by (4.6.8) re-casts (4.6.7d) as:

$$\Phi_{\rm Ocean}(\bar{\delta}_O^+ - \delta_{\rm ref})$$

$$= \dot{M}(\bar{\delta}_A - \delta_{\rm ref}) + M\dot{\bar{\delta}}_A - \Phi_{\rm Fossil}(\bar{\delta}_F - \delta_{\rm ref}) - \Phi^-_{\rm Ocean}(\bar{\delta}_O^+ - \bar{\delta}_O^-) - \Phi^-_{\rm Biota}(\bar{\delta}_B^+ - \bar{\delta}_B^-) - \Phi_{\rm Biota}(\bar{\delta}_B^+ - \delta_{\rm ref}) \quad (4.6.9)$$

As working approximations, we use

$$\delta_B^- = \delta_A - 18 \qquad \text{in ‰} \quad (4.6.10a)$$

and

$$\delta_{\rm ref} = \hat{\delta}_A - 17.5 \quad (4.6.10b)$$

Note that we are treating $\delta_{\rm ref}$ as an 'error-free' prescribed function that 'just happens' to have a fixed difference from an estimate of δ_A. Therefore we must, in principle, consider the degree of uncertainty in the difference between δ_A and $\delta_{\rm ref}$.

With these preliminaries we can work through (4.6.9) term-by-term. There are several different types of quantity involved. Some of the quantities are from direct observational data and the uncertainties can be treated as described in previous sections. The fossil data are inventory-based estimates derived from annual data. The isofluxes (disequilibrium contributions) involve more complicated analyses. In each case there is a long-term variation in isotopic disequilibrium that can be described by the convolution of a model response with the history of change in atmospheric δ^{13}C over the industrial period. However in each case there is also a short-term variability. For the biotic fluxes, much of this comes from interannual variation in the gross-flux, $\Phi^-_{\rm Biota}$. For the oceans there will be short-term interannual variability driven by changes in disequilibrium due to temperature dependence of isotopic fractionation. The error statistics of these isofluxes are modelled as the sum of two independent AR(1) processes: one with a very large autocorrelation to represent the uncertainty in the long-term model-derived estimate and one with a small autocorrelation to characterise the unknown interannual variability.

A number of these quantities are only available in annual mean form. The inventory data for $\Phi_{\rm Fossil}$ are given as annual totals, and the biotic gross fluxes have such large seasonal cycles that the balance described in Figure 12 is only meaningful for de-seasonalised data. Because of this restriction, the calculations are performed with $\Delta t = 3$ months. This recognises that a time-step of 1 month is carrying more information than is meaningful for the analysis. The use of a time-step of 1-year is inadequate given that the analysis above indicates that to estimate a mean flux for a single year requires additional smoothing. The calculations are performed for the years 1983 to 1994 inclusive, except when additional smoothing leads to data loss from the ends of the record. For each term on the right of (4.6.9) we require a best estimate (see Figure 13) (with some consistent degree of time-averaging) and an autocovariance function for the estimate (see Figure 14).

$\frac{d}{dt}M(\delta_A - \delta_{\rm ref})$ This term is derived from observations. For the present analysis we use CO_2 from Mauna Loa (Keeling and Whorf, 1994, plus electronic update) and δ^{13}C from Cape Grim (Francey et al., 1995b and personal communication). In principle this raises issues

of consistency in the departures from global representativeness. An outline discussion is given in Section 6 below. However, the comprehensive approach to having the correct spatial representation of the data is to perform the analysis within the framework of a synthesis inversion based on calculated responses from an atmospheric transport model. It will be recalled that one of the objectives of the present report is to identify the issues associated with the time series statistics of inputs to such calculations.

The δ^{13}C data are taken as monthly means with ± 1 s.d. of 0.02‰ independent errors. A small number of gaps in the record are filled by linear interpolation. When there are overlapping data for the two different mass spectrometers, the average is used. Full consideration of the change of instruments would take us beyond the stationary analysis used here. In the discussion above we noted that in principle we need to consider errors from the $\delta_A - \delta_{\rm ref}$ term. However, these errors will be of order 0.1% of the product, while the analysis in Section 4.3 indicates that errors in estimates of $\dot{M}$ will be of order 10–20%. Thus for the error autocovariance, we propagate the Cleveland et al. error model through the filtering process described below. For concentrations and their autocovariances, the conversion factor 0.471 ppm/GtC is used. The autocovariance of the term is shown as the solid curve in Figure 14.

$M\frac{d}{dt}\delta_A$ This term is based on the same set of observations as the previous term. Again, the error is dominated by the contribution from the derivative term and is calculated as described in Section 4.3, starting from a 0.01‰ s.d. white noise on the monthly mean δ^{13}C data.

$\Phi^-_{\rm Biota}(\delta^+_B - \delta^-_B)$ This is usually derived from modelling with the terrestrial biota responding to the atmospheric δ^{13}C record derived from ice cores. (Quay et al. (1992) estimated 12‰GtCy^{-1}; Enting et al. (1993) used a 5-box biosphere model to estimate 26.5‰GtCy^{-1}; Fung et al. (1997) calculate an isotopic disequilibrium of 0.33‰ in 1998, corresponding to an isoflux of $\approx$ 18‰GtCy^{-1}). Heimann and Maier-Reimer (1996) quote a 'best-estimate' of 23.4‰GtCy^{-1}, but this is a composite that incorporates the equivalent of the present budgeting calculation and so is not an independent estimate. The function used here (25.6‰GtCy^{-1} in 1990 and increasing at 0.2‰GtCy^{-1}y^{-1} over the 1980s) is taken from BDM (box-diffusion model) calculations by Trudinger (personal communication). The curve is shown in Figure 13 as the – · – line. For the long-term uncertainty, the intermodel differences suggest a standard deviation of around 3‰GtCy^{-1}. We use 9.0(‰GtCy^{-1})^2 as the $R(0)$ value of an AR(1) autocovariance with an inter-season correlation $a = 0.99$. The long-turnover times for the biota will buffer the isotopic composition against significant interannual variability. However, there is scope for significant interannual variability in the isoflux due to interannual variability in $\Phi_{\rm Biota}$. One measure of how large this variability might be is obtained from the study by Dai and Fung (1993) modelling ecosystem responses to variations in temperature and precipitation. This suggests short-term net flux variations of order 0.25 GtCy^{-1} with NPP variations of about twice this amount, or 1% of the total. We represent this as an AR(1) process with $R(0) = 0.2$, $a = 0.7$.

$\Phi^-_{\rm Ocean}(\delta^+_O - \delta^-_O)$ The oceanic isoflux can be obtained from observations (as in Tans et al., 1993) or from modelling as in double deconvolution. As with the biotic disequilibrium, we regard the uncertainty as having a long-term 'model-error' component and a short-term 'natural variability' component. The model error component is assigned a 10%

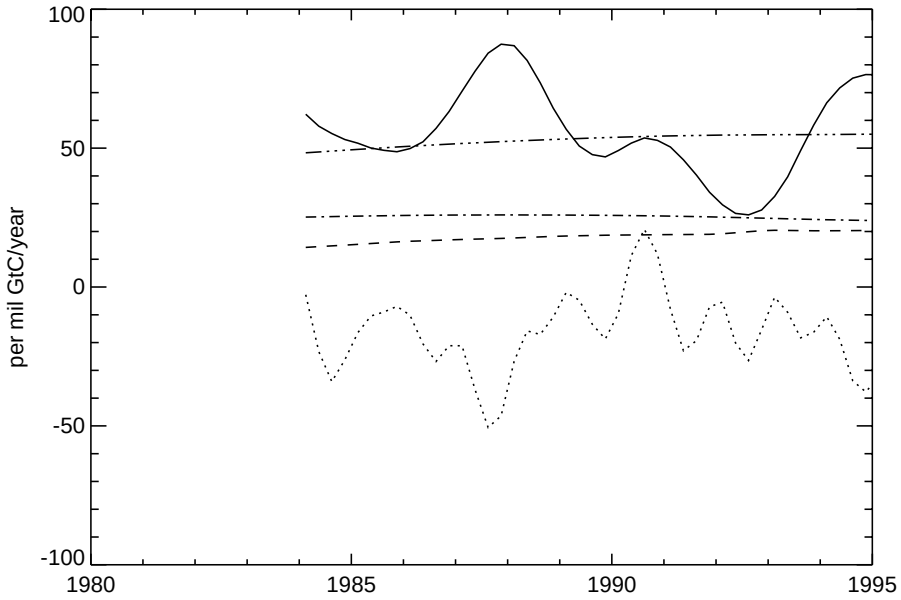


Figure 13: Terms in Equation (4.6.9) used to obtain ocean flux estimates. Lines as for Figure 12.

uncertainty. The short-term variability is related to the temperature dependence of the difference between the air-sea and sea-air fractionation factors. This is of order 0.1‰ $^{\circ}\mathrm{C}^{-1}$. With a gross flux of order 80 GtCy^{-1}, we represent this as an AR(1) process with $R(0) = 64$ and $a = 0.7$.

$\Phi_{\mathrm{Fossil}}(\delta_F - \delta_{\mathrm{ref}})$ Both the fossil flux Φ_{Fossil} and its isotopic composition δ_F are derived from inventory studies, with the main isotopic differences being associated with the type of fuel. We attribute most of the uncertainty to the Φ_{Fossil} factor and split it into a possible long-term methodological bias (most plausibly due to under-reporting) and a year-to-year variability due to effects such as mis-match between reporting times and emission times. Each component is assigned a 0.5 GtCy^{-1} uncertainty. The effect of this term is small because $\delta_F \approx \delta_{\mathrm{ref}}$. Uncertainty in the fossil emissions will translate directly into uncertainties in estimates of Φ_{Biota}.

$\Phi_{\mathrm{Biota}}(\delta_B^+ - \delta_{ref}^B)$ This term is negligible compared to most of the other terms in (4.6.9) because we expect Φ_{Biota} to be at most a few GtCy^{-1} and δ_{ref} to differ from the 'ideal' value of δ_B^+ by less than 1‰. This term is omitted from the subsequent uncertainty analysis.

Figure 13 shows the initial time series for each term in (4.6.9). Figure 14 shows the autocovariances of each of these time series inputs.

Figure 15 shows the estimated ocean fluxes. These are obtained by combining the contributions shown in Figure 13 (with the signs as shown in (4.6.9)) and then dividing by $\delta_O^+ - \delta_{\mathrm{ref}} \approx 15‰$. Figure 16 shows the autocovariance for the ocean flux estimate. These are obtained by summing the autocovariances shown in Figure 14, and then dividing by (15‰)^2.

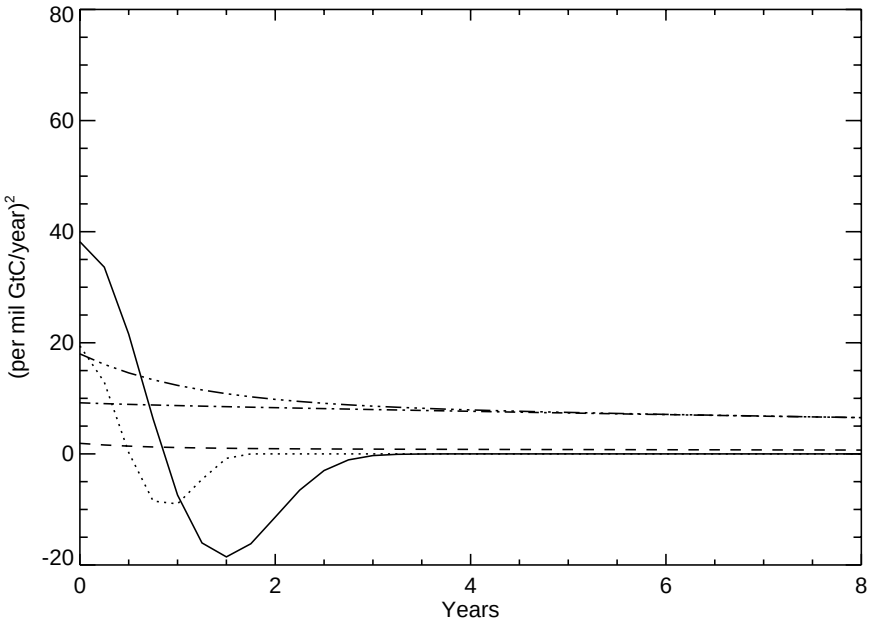


Figure 14: Autocovariance of terms in (4.6.9). Lines as for Figure 12.

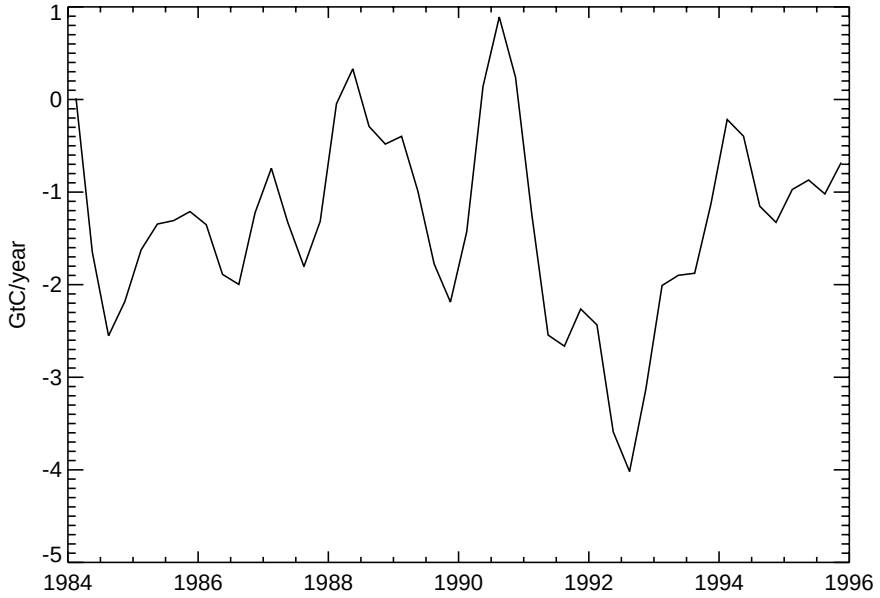


Figure 15: Ocean flux estimate(s)

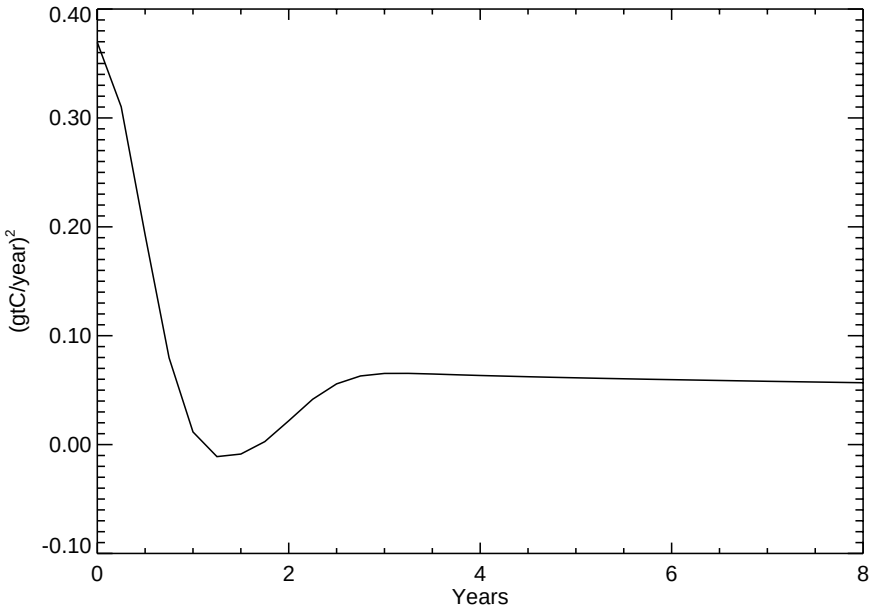


Figure 16: Autocovariance of ocean flux estimate(s)

The isotopic disequilibrium terms are input as smooth curves. In the absence of observations, all short-term variability is treated as noise. Similarly, the fossil data are smoothly interpolated from annual values. The CO_2 and δ^{13}C observations were each input as monthly mean values and smoothed with a 'Bloomfield filter' and (when required) differenced to obtain growth rates. The autocovariances for the estimated growth rates were calculated as described in Section 4.3, using the Cleveland et al. (1983) error model for the CO_2 data and a 0.1‰ standard deviation white noise error for the δ^{13}C data.

These results need to be regarded as indicative of exploratory calculations rather than as definitive statements about the carbon cycle. The point-wise uncertainties in Φ_{Ocean} (of order ± 0.6 GtCy^{-1} are well below those estimated for longer-term averages in the synthesis inversion calculations by Enting et al. (1995). This difference will reflect a combination of 'representativeness' error, and the assumed *a priori* precision of the isoflux estimates. Figure 14 is of particular interest because it shows the relative importance of the different error contributions on a range of time-scales. In particular, within the present assumptions, the uncertainties in multi-year averages will be dominated by the uncertainties in the long-term dynamics of isotopic disequilibrium.

The interannual variability of the oceanic CO_2 uptake has recently been discussed by Lee et al. (1998). They noted significant differences between the degree of variability estimated by Keeling et al. (1995), Francey et al. (1995a) (similar analyses with different data sets), Rayner et al. (1999) (essentially a similar analysis using an atmospheric transport model to deal with the 'representativeness' problem) and the Lee et al. analysis (based on interannual variations in wind-speed and p_{CO2}). An assessment of the significance of these discrepancies requires combining the present approach (detailed statisitical analysis) with the spatial inversions of Rayner et al. (1999).

5 Smoothing splines

5.1 Principles

Spline curves are defined as piecewise polynomial functions: i.e. as polynomials between a set of points known as the nodes. A spline of order n is a continuous function with continuous derivatives up to order $n-1$ and the nth derivatives constant over the intervals between the nodes.

A number of early analyses of CO_2 time series made use of fixed-node splines. These had fewer nodes than data points and used the node spacing to define a low-pass filtering effect: variations on scales less than about the node spacing were suppressed because they could not be represented by such a spline. With the node spacing defined to give a desired frequency cutoff, the fit was defined as the spline that gave the least-squares fit to the data. An analysis by Enting (1986) showed that this type of fixed-node spline fit had a very great phase-dependence for frequencies in the cutoff band: the results depended strongly on the position of the nodes, and not just on their spacing.

Smoothing splines have a node at each data point. The filtering is achieved by an additional smoothness constraint. They are formally defined as the curves, $g(t)$, that minimize the functional:

$$J = \sum_{j=1}^{N} [z(t_j) - g(t_j)]^2 + \lambda \int_{t_1}^{t_N} \left[\frac{d^2}{dt^2} g(t) \right]^2 dt \qquad (5.1.1)$$

Formally the minimum in (5.1.1) is over all functions with bounded second derivatives. Actual calculations use the result that the curve that minimises (5.1.1) is a cubic spline with nodes at the data points, t_j.

Enting (1987a) has reviewed some of the properties of smoothing splines, using normalisations consistent with those used in the computer routines given by de Boor (1978). The key result is that smoothing splines act approximately like low-pass filters with response functions:

$$\Psi_{\text{spline}}(\omega) = 1/[1 + (\omega/\omega_c)^4] \qquad (5.1.2)$$

where the 50% cutoff frequency is given by

$$\omega_c = (\lambda \Delta t)^{-1/4} \qquad (5.1.3a)$$

corresponding to a period

$$T_c = 2\pi/\omega_c \qquad (5.1.3b)$$

In the time domain, the filtering has been described by Silverman (1984, 1985). (Note that the notation of these two papers differs from that used here — and from each other! Note in particular that the density of points — Silverman's f — is $1/N\Delta t$ rather than $1/\Delta t$.) The spline curve is shown to be approximately equivalent to convolving the data with a kernel given by $h^{-1}\kappa((t-t')/h)$ with

$$\kappa(x) = \exp(-|x|\sqrt{2}) \sin(|x|/\sqrt{2} + \pi/4)/2 \qquad (5.1.4a)$$

and

$$h = (\lambda \Delta t)^{1/4} = \omega_c^{-1} \tag{5.1.4b}$$

The equivalent filter has coefficients:

$$c_k = \omega_c \Delta t \times \kappa(k \omega_c \Delta t) \tag{5.1.4c}$$

Note that (5.1.2) to (5.1.4) are asymptotic results. One point at which they will break down is near the ends of records. Silverman (1984) show that this can be approximated by using a slightly modified kernel reflected about the end-point. Enting (1989) gives examples that show the limitations of this approximation. These asymptotic results will break down if λ is so large that all points are affected by the ends of the records or if λ is so small that T_c becomes comparable to Δt.

A common modification of smoothing splines is to use a weighted sum of residuals in the objective function, defining the spline as the function that minimises

$$J = \sum_{j=1}^{N} [z(t_j) - g(t_j)]^2 / u_i^2 + \lambda \int_{t_1}^{t_N} \left[\frac{d^2}{dt^2} g(t) \right]^2 \, dt \tag{5.1.5}$$

If all the weights are equal to u then the cutoff frequency becomes

$$\omega_c = (u^2 \lambda \Delta t)^{-1/4} \tag{5.1.6}$$

This equivalance can be used as a way of applying different filtering to different parts of a single record. Dividing the data set into two or more time intervals and using different choices of constant weight within each interval is (approximately) equivalent to using different time-averaging within each interval. This can be convenient, but may lead to difficulties in defining consistent time-averaging for subsequent analysis.

Smoothing splines have a number of advantages:

- as with all splines, they interpolate between the data points, i.e. the fit $g(t)$ can be evaluated at all times t (in the interval $[t_1, t_N]$), and not just at the data points;
- they can handle unequally-spaced data and the filtering properties are relatively insensitive to the data spacing; and
- the spline curve is defined as a differentiable function.

One disadvantage is that the de Boor (1978) smoothing spline algorithm breaks down at about $N \approx 1000$ data points (when using 32-bit arithmetic). This problem is noted by de Boor who suggests that the fixed-node splines might be more appropriate when such a large number of data points are involved. As noted above, Enting (1986) has pointed out problems with this approach. Numerical experiments (H. Granek and C. Trudinger, unpublished) show that the

problems with smoothing splines can be postponed to larger N by changing the calculation from a 32-bit to a 64-bit floating point representation, or by replacing the matrix inversion procedure used by de Boor (1978) with a procedure that is less sensitive to rounding error.

Granek (1995) has discussed a hybrid form that uses the same smoothness constraint as for smoothing splines but works with a reduced number of nodes. The fit is still determined by minimizing (5.1.1) but over a 'smaller' class of functions defined by splines with the specified nodes. A heuristic argument by Granek illustrates why such a spline should be very close to the smoothing spline: Minimizing over a restricted class of functions will not change the solution if the restricted class includes the minimum for the wider class. Since minimising (5.1.1) over all functions produces a smooth function, the same result can be produced by minimizing over an appropriate set of smooth functions. The set of reduced-node splines will not in general contain the smoothing spline, but it should contain splines that are very similar. Thus minimising over the restricted set should give a curve similar to the smoothing spline. Numerical and analytic studies by Granek (1995) have confirmed and quantified this argument.

There are a number of ways in which particular smoothing spline (i.e. the particular value of λ) can be chosen:

1. As discussed above, the value of λ can be chosen on the basis of the requisite frequency response by using (5.1.2). This is the approach that we have mainly used in our laboratory. We also make use of the ability to use the data weights to modify the frequency response in different parts of the record on the basis of Equation (5.1.6).
2. Alternatively, the spline can be chosen to be the smoothest (i.e. largest λ) subject to the sum of squares of residuals not exceeding some pre-specified value.
3. Less commonly, a requisite smoothness can be pre-specified and λ chosen to give the closest fit consistent with this smoothness.
4. A more sophisticated way of choosing λ is the technique of generalised cross-validation (GCV) described by Craven and Wahba (1979). This approach aims to find the spline with the minimum mean-square error. As will be seen from the discussion on digital filtering in Section 2.3, this requires a knowledge of the error distribution. The GCV technique uses estimates derived (conceptually) from the set of residuals for each point $k = 1, N$ calculated for a spline using all points but k. The expressions for the optimal λ are asymptotically correct for large N (subject to constraints on the distribution of nodes). Theoretical and numerical examples show that the GCV estimates are near-optimal for N of order 50 or more.

Approaches (2) and (3) involve iterative solution of problem (1), i.e. finding the smoothing spline for specified λ and repeatedly adjusting λ and then recalculating the spline until the solution with the required properties is found. Similarly, GCV involves iterative adjustment of the paramater λ.

It is straightforward to apply the various techniques described in Section 2 for quantifying the uncertainties in splines, regarded as estimates of unknown functions. In particular, the explicit

expression for the (approximate) equivalent filter allows the use of (3.3). In addition, Monte Carlo simulations can be used to simulate the error distribution in spline fits and, for unknown error statistics, bootstrap analyses can be undertaken.

The literature on splines also addresses the issue of estimating uncertainty. Silverman (1985) analyses splines in a Bayesian framework, where λ characterises the *a priori* distribution of 'smoothness'. Within this framework, he expresses the variance of the spline estimates as $2^{-3/2}\sigma^2\lambda^{-1/4}(\Delta t)^{3/4}$ where σ^2 is the variance of the data. Silverman notes the general form of confidence intervals for linear functionals such as the derivative. However he doesn't give explicit expressions and for practical calculations, the easiest approach is to use the results of (4.2.2) with the approximation (5.1.4c). Silverman also notes the possibility of determining confidence intervals for non-linear functionals (e.g. locations and/or values of maxima) by Monte Carlo simulation of samples from the posterior probability distribution. The main limitation of Silverman's analysis is that it assumes independent errors on the data.

The use of smoothing splines to estimate functions raises the question of what is actually being estimated. In terms of the discussion in Section 2.3, the spline fit is not well descibed as a band-pass filter because the transition band around ω_c is very broad. Similarly, it is not simply an average over a period T_c because the equivalent filter (Equations 5.1.4a–c) differs greatly from a running mean.

EXAMPLE 6: Fitting ice core and firn data

Measurements of trace gas concentrations in bubbles trapped in polar ice provide an invaluable record of past atmospheric changes. However this is a smoothed record of actual atmospheric changes. Although the majority of the trapping happens in a short time before the final close-off (Enting, 1985), the age distribution of the gas is broadened by the processes of diffusion through the firn (Trudinger et al., 1997).

The data set analysed here is 69 data points over the period 1006–1978 as measured by Etheridge et al. (1996). Figure 17 shows smoothing spline fits to the data for various degrees of smoothing with detail for the recent period shown in Figure 18. In the recent period, Δt of order 2 years leads to cutoff periods of T_c = 16, 42, and 112 years for $\lambda = 20$, 1000 and 50000. In the pre-industrial period, these same values of λ give $T_c = 30$, 79 and 210 years.

Since there is little information available about the 'noise' structure of these ice-core records, we have used the bootstrap method to explore the issue. Figure 19 shows the spline fit to the original data set and then 25 realisations of spline fits to re-sampled subsets, all calculated with $\lambda = 1000$. One feature of the realisations is that the distribution has a strong clustering about the original fit and then a long-tailed component, generated by cases when the resampling omits a sequence of consecutive data points. This behaviour is even more pronounced for fits using smaller λ.

There is a minor technical point that arises in connection with re-sampling calculations (and occasionally with 'normal' spline fitting). The definitions of smoothing splines in terms of the function that minimises (5.1.1) remains valid even when the t_j are not all distinct. However

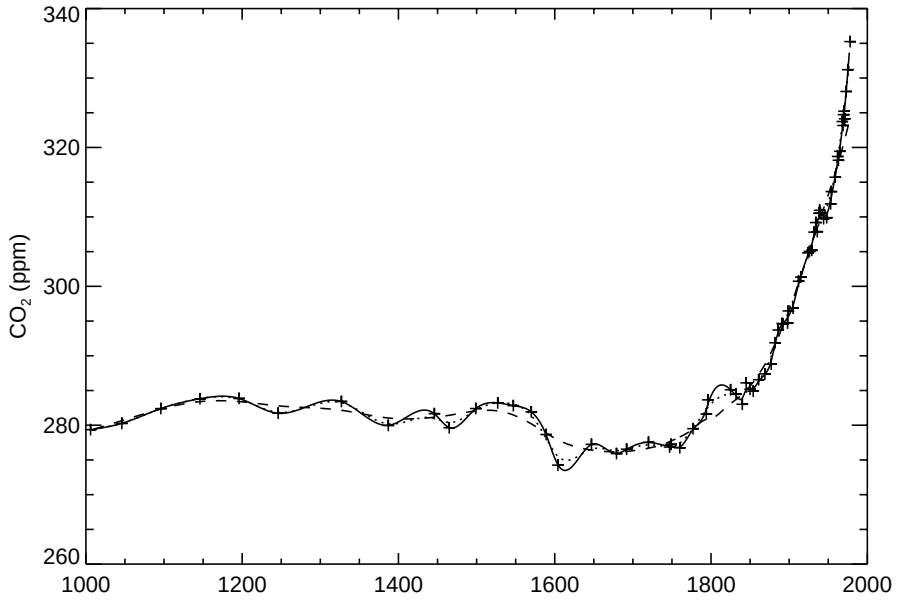


Figure 17: Smoothing spline fits to ice-core data for various cutoff frequencies defined by $\lambda = 20$ (solid curve), 1000 (dotted curve), 50000 (dashed curve).

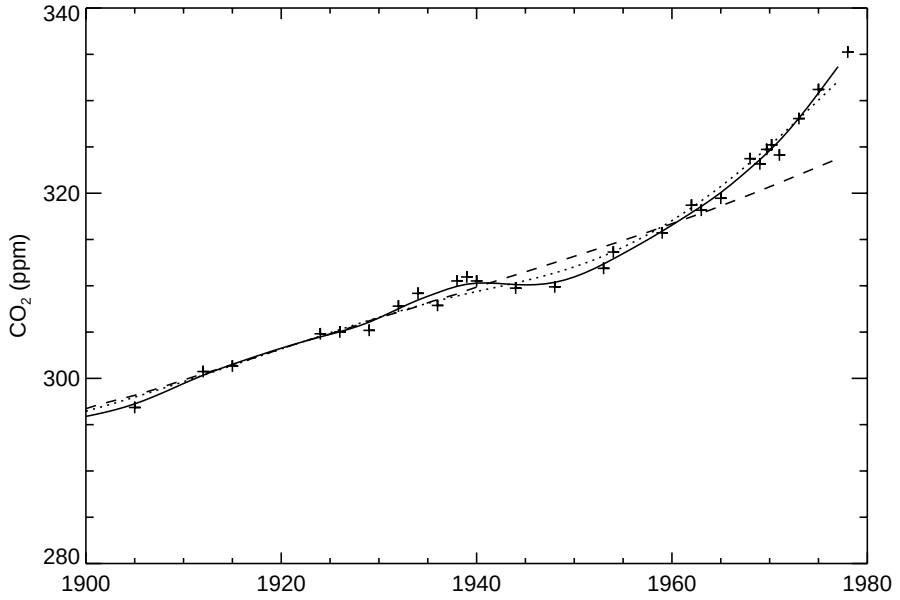


Figure 18: Smoothing spline fits to ice-core data (detail 1900-1980) for various cutoff frequencies defined by $\lambda = 20$ (solid curve), 1000 (dotted curve), 50000 (dashed curve).

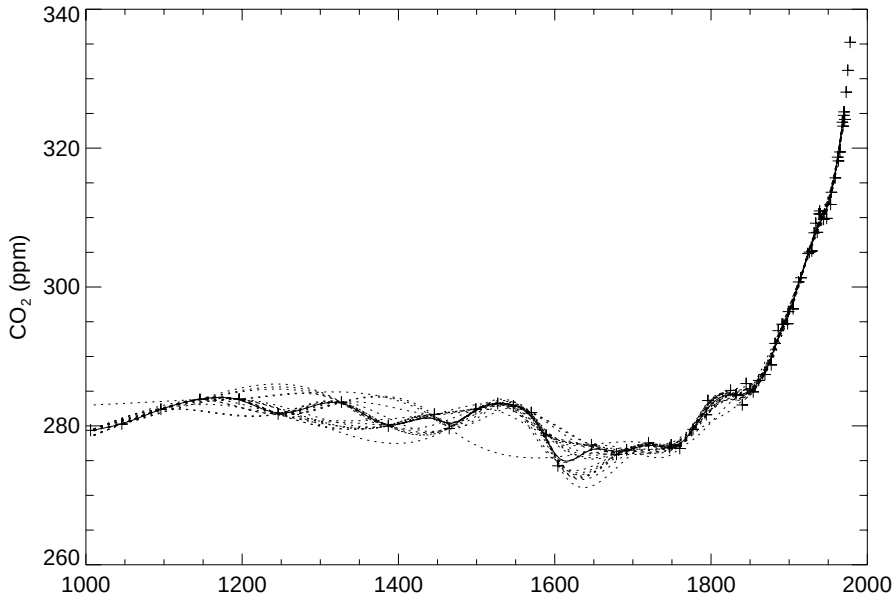


Figure 19: Bootstrap analysis of smoothing spline fits to ice-core data for various cutoff frequencies.

the de Boor (1978) algorithm fails in this case (and will encounter numerical instability if the internode spacing is too small.) The way to use the de Boor code to produce the correct spline is to replace y_j values whose t_j coincide by their average y value and apply a weight $1/\sqrt{n}$ for each point with n coincident values in the original data set. Alternatively, for splines whose set of nodes is specified independently of the set of t_j (i.e. in the algorithms for the approach described by Granek, 1995) the problem does not arise.

5.2 Growth rates and deconvolutions

As noted in Section 5.1, an important advantage of smoothing splines is their ability to define values continuously through the interval fitted, to the extent that the spline function is differentiable (with discontinuous third derivatives at the nodes). This makes smoothing splines particularly useful for problems involving rates of change. As shown by Equation (4.4.2) source deconvolution problems effectively involve differentiation. Furthermore, the methane deconvolution example discussed in Section 4.4 suggests that the uncertainties in deconvolution calculations will often be dominated by the uncertainties in estimating the derivatives. Therefore in this section we do not consider deconvolutions explicitly, but rather concentrate on the use of splines to estimate growth rates.

The issue raised in Section 5.1 as to 'what is being estimated' remains a problem when considering derivatives. An additional complication is that relations such as $\bar{\dot{g}} = \dot{\bar{g}}$ are only approximately true because the spline is only approximately a stationary filter. Of course the greatest

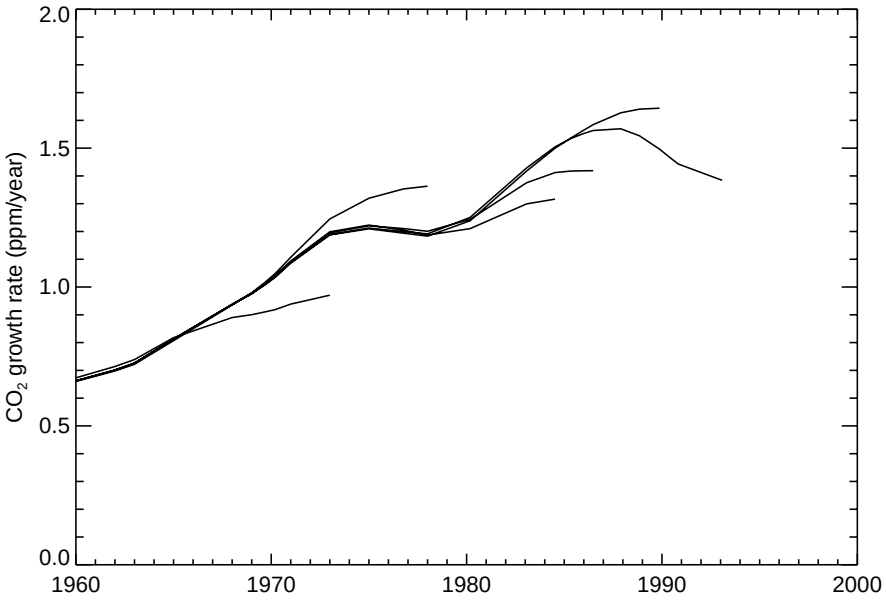


Figure 20: Rates of CO_2 increase estimated from derivatives of a sequence of spline fits to a composite data set from ice-cores and firn air. Successive curves (going left to right) differ by fitting 3 more data points. The smoothing corresponds to $T_c \approx 19$ years.

departure from stationarity occurs at the end of the record. The following example explores this aspect of estimating growth rates using splines.

EXAMPLE 7: Growth rates and end effects in ice core data.

The data set is from Etheridge et al. (1996), combining both the data from bubbles in ice-cores (as used above) and data from air extracted from the firn and additional 13 points spanning effective dates of 1976–1993. Figure 20 shows a sequence of derivatives of spline fits to these data. All the splines used $\lambda = 40$ whence $T_c \approx 19$ years. The initial spline fit was based on all the data. A sequence of additional curves was calculated by successively truncations of the data set, each case omitting the last three points of the previous case. The common feature is that each truncated case shows a flattening relative to the extended cases. This is of course to be expected since the definition of the spline is in terms of (constrained) minimisation of the absolute value of the second derivative (i.e. flatter first derivative), a criterion that becomes relatively more important in regions (ends of gaps) where there are fewer data to impose any curvature.

6 Spatial analogues

6.1 The atmospheric transport inversion problem

As well as varying in time, atmospheric CO_2 has spatial variations that can be interpreted to provide information about the carbon cycle. The principle is that the spatial distribution of concentrations can provide information about the spatial distribution of surface fluxes and that this provides information about the processes involved. However, since the distribution of concentrations reflects the effects of both surface fluxes and atmospheric transport, a model of atmospheric transport is needed for any such analysis.

The analysis of spatial data raises many of the same problems as the analysis of temporal data, with the extra complication of additional dimensions.

In Section 2.3 it was noted that spectral analysis of concentration records gives only estimates of the combined spectra of signal and noise. Additional information or assumptions are required in order to decide which parts of the observed variations represent meaningful signals. When considering spatial variations, the situation is far worse — there are not sufficient data available from which to estimate multi-dimensional power spectra to characterise spatial (or space-time) variability. Indeed, for spatial variations it seems likely that the concept of statistically stationary distributions is likely to be of limited use, and that more complicated (and less attainable) characterisations of the variability will be required for a comprehensive analysis of the uncertainties in interpretation of spatial variations. Some initial studies of the relation between spatial and temporal variations in CO_2 has been undertaken by Dargaville (1999) using a singular vector decomposition.

The relation between concentrations and sources can be expressed as

$$c(r,t) = c_{-\infty} + \int G(r,t,r',t')S(r',t')dr'dt' \qquad (6.1.1)$$

where the Green's function $G(.)$ embodies the transport characteristics of the atmosphere. Since (6.1.1) is not a convolution integral, even if $S(r,t)$ could be regarded as stationary in space and/or time, this would not imply that $c(r,t)$ had such stationarity. Inversions of CO_2 data use $G(.)$ based on numerical models. There is a distinction between those models where the transport comes from analyses of observed winds, and those where the transport is based on GCM calculations. The latter will not reproduce specific synoptic events and so on synoptic time scales (6.1.1) will at best only be correct in an average sense. Many studies have implicitly assumed that this is not a serious problem if monthly-mean data are fitted. It would seem that there has been little, if any, systematic testing of this assumption.

There are two main types of inversion calculation, the *synthesis* techniques based on discretisation of (6.1.1) and *mass balance* techniques based on the differential form:

$$S(r_x,t) = \frac{d}{dt}c(r_x,t) + T_{r_x,t}[c(r',t')] \qquad (6.1.2)$$

Equations (6.1.2) applies at a set of locations, r_x, generally the entire surface of the earth, at which $c(r_x, t)$ is known and $S(r_x, t)$ is unknown.

Synthesis inversions are based on a discretisation of (6.1.1) in space and time and the solution of the resulting linear equations permits direct propagation of error statistics in the manner described in Section 2.2. This was first applied to a cyclo-stationary case by Enting et al. (1993). In the cyclo-stationary case, the data were expressed as Fourier components of the seasonal cycle. For a stationary noise distribution, the Fourier component will have independent errors. This provides a useful simplification in the computer codes and simplifies the error modelling. Enting et al. (1993, 1995) based their data uncertainties for the seasonal components on the residuals of regression analyses and their uncertainty estimates for annual means on sampling experiments reported by Tans et al. (1990).

Mass balance inversions where $c(r_{\rm surface})$ is needed, firstly to define the rate of change and secondly to provide the boundary condition for the model that is used to calculate the transport term. It is not computationally practical to propagate error statistics through the transport operation $T[.]$. Uncertainty analyses for mass balance inversion have used the bootstrap approach (Conway et al., 1994).

The analogies between spatial and temporal varibility can also be used to help clarify some of the issues involved in discussing the resolution of inversion calculations. We do this in terms of inverse length scales, κ, and distinguish 4 different scales that need to be considered:
$\kappa_{\rm grid}$: This is the cutoff defined by the discretisation of the transport model;
$\kappa_{\rm S:N}$: This is the wavenumber above which $f_n(\kappa) > f_s(\kappa)$, i.e. the estimation of the signal is degraded because the noise is of greater amplitude than the signal (c.f. Equation 2.3.1);
$\kappa_{\rm smooth}$: the resolution at which we wish to know the answers (c.f. Equation 2.3.9); and
$\kappa_{\rm basis}$: (for synthesis inversions) the resolution imposed by the basis used to describe the source distribution.

Clearly we need:

$$\kappa_{\rm smooth} \leq \kappa_{\rm basis} \leq \kappa_{\rm grid}$$

and

$$\kappa_{\rm smooth} \leq \kappa_{\rm S:N}$$

6.2 Representativeness

The budget Equation (1.1) applies to the whole atmosphere, but in practice the concentration data are only available at a relatively small number of locations, predominantly at the earth's surface. On sufficiently long time-scales the variations will be characteristic of the atmosphere as a whole. The following argument addresses the issue of what might be meant by 'sufficiently long'. We consider the simplest problem of how well a surface average will represent an average over the whole atmosphere and use a simple 'toy model' approximation of purely diffusive

transport, following Enting and Newsam (1990). In this model, with different horizontal and vertical diffusion and a specific form of enhancement of east-west diffusion the coordinates in the differential equation separate. This means that the present problem can be characterised as a one-dimensional diffusion model, corresponding to the $m = n = 0$ case of the solution described by Enting and Newsam.

The equation for a mixing ratio c in terms of pressure coordinates in the range $[0, 1]$ is

$$\frac{\partial}{\partial t}c = K_p \frac{\partial^2 c}{\partial p^2} \tag{6.2.1}$$

where K_p is an effective diffusion coefficient and subject to the boundary conditions:

$$K_p \frac{\partial c}{\partial p} = 0 \qquad \text{at } p = 0 \tag{6.2.2a}$$

and

$$K_p \frac{\partial c}{\partial p} = s(t) \qquad \text{at } p = 1 \tag{6.2.2b}$$

The solution for a frequency ω has the form

$$c(p, t) = A_\omega e^{i\omega t} \cosh(\lambda_\omega p) \tag{6.2.3a}$$

with the constraint (following from 6.2.1)

$$i\omega = K_p \lambda_\omega^2 \tag{6.2.3b}$$

The change in the atmospheric content of the constituent described by c is given by the (pressure weighted) integral of the change in mixing ratio:

$$\frac{d}{dt}\bar{c} = \frac{d}{dt}\int_0^1 c(p, t)\, dp = K_p \frac{\partial c}{\partial p}|_{p=1} = s(t) \tag{6.2.4}$$

In these units, $s(t)$ is the source that determines the global budget. We can use the solution to assess the scale of the error involved in using the rate of change of surface concentration $\frac{d}{dt}c(p = 1, t)$ to represent the changes in the atmospheric mean $\frac{d}{dt}\bar{c}$.

At frequency ω we have

$$\frac{d}{dt}c(p = 1, t)/s(t) = \frac{i\omega \cosh(\lambda_\omega)}{\lambda_\omega \sinh(\lambda_\omega)} = \lambda_\omega \coth(\lambda_\omega) = 1 + (\lambda_\omega)^2/3 - (\lambda_\omega)^4/45 + \ldots \tag{6.2.5}$$

To avoid a major bias due to lack of vertical representativeness, we require

$$|(\lambda_\omega)^2| = |i\omega/K_p| \leq 1$$

Using the value $K_p = \frac{4}{9} \times 10^3$ Pa^{2}s^{-1} from Enting and Newsam (1990) gives

$$\omega = 2\pi\nu \leq 1.4$$

or ν^{-1} of order 4 years or longer.

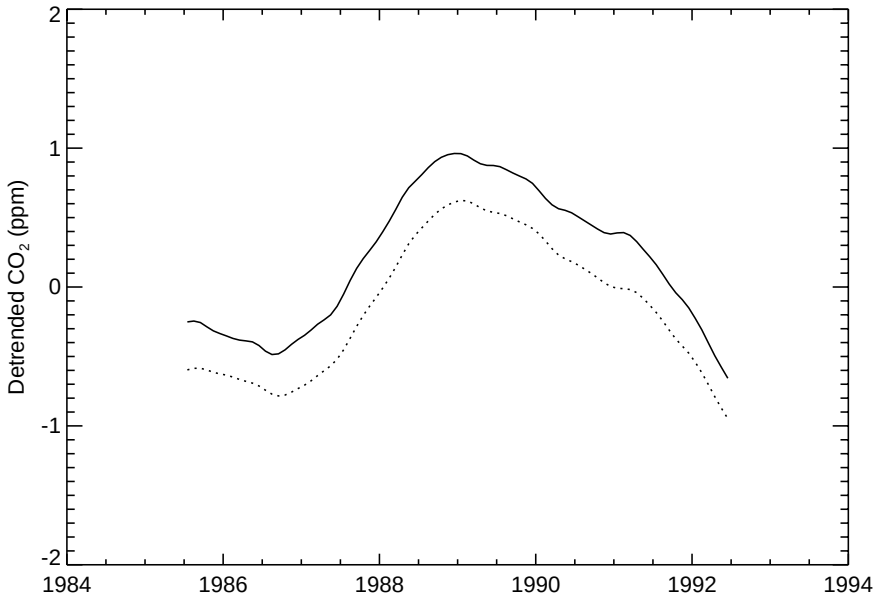


Figure 21: Comparison of the interannual variability in the surface mean (solid line) and the whole-atmosphere mean (dotted line) of detrended and decycled CO_2 concentrations for 1985–93.

In view of the approximate nature of the calculation, the 4-year time-scale cannot be taken too literally. A more precise analysis needs to be undertaken, using model response functions such as those used in time-dependent inversions (e.g. Rayner et al., 1999). In such analyses it may be convenient to use multi-exponential parameterisations of the responses in the manner described by Mulquiney and Norton (1998).

While there has not been such an analysis of the frequency-dependence of the relation between surface and global means, there have been calculations of the actual relation for recent times. Figure 21 shows concentrations (detrended by subtracting $345 + 1.5(t - 1985)$ and decycled using a '12-month' running mean) from a mass-balance inversion calculation by Dargaville et al. (1999). The solid line shows the global mean surface concentration. This is based on NOAA/CMDL data. The dotted line is the whole-atmosphere mean calculated by the model when run with the surface concentration as the specified boundary condition. The model is that described by Law et al. (1992), using wind fields from ECMWF analyses (Trenberth, 1992). Although the calculation is based on winds that differ from year to year, the main result obtained by Dargaville et al. (1999) is that these variations in transport have little effect on CO_2 distribution. Figure 21 shows that the main difference is that the global means have a small phase lag relative to the surface mean, as well as being slightly smoother.

7 Recursive estimation

One widely-used class of estimation techniques for time series is the recursive estimation techniques based on the Kalman filter (see for example: Gelb, 1974; Young 1984). The approach is based on the definition of a state-space model in which the system is represented by a time-evolving state that is to be estimated from indirect observations.

Some of the advantages of the state-space-modelling/recursive-estimation approach are:

- The recursive estimation procedure is computationally efficient;
- The approach integrates the deterministic modelling with the statistical analysis;
- The technique applies to non-stationary time series without any change. In particular, missing data can be accomodated;
- The technique is inherently Bayesian, so that it can incorporate *prior* information; and
- The technique can be applied to multiple data sets.

Some of the important disadvantages of Kalman filtering are:

- The Kalman filter gives a one-sided estimate and so (except in special cases such as estimating fixed parameters) the Kalman filter estimate has a higher variance than a two-sided estimate;
- While the estimates also have their uncertainties calculated, this is restricted to $\hat{x}(t)$ and $\mathrm{Var}[\hat{x}(t)]$. Correlations between estimates at different times are not calculated unless the state-space is enlarged to include lags.

The state-space model is defined by the time evolution equation for the state $\mathbf{x}(t)$:

$$\mathbf{x}(t + \Delta t) = \mathbf{F}(t) + \mathbf{u}(t) + \mathbf{w}(t) \tag{7.1}$$

and by the relation between the 'state' and the observations $\mathbf{z}(t)$

$$\mathbf{z}(t) = \mathbf{H}(t)\mathbf{x}(t) + \mathbf{n}(t) \tag{7.2}$$

where $\mathbf{F}$ is the evolution matrix and $\mathbf{u}$ is a specified deterministic forcing. $\mathbf{w}$ is a stochastic forcing with a specified covariance matrix $\mathbf{Q}$. The matrix $\mathbf{H}$ projects the state onto the set of observations and $\mathbf{n}$ is a noise term with a specified covariance matrix, $\mathbf{R}$. Therefore the Kalman filtering formalism requires a specification of $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{F}$, $\mathbf{H}$ and $\mathbf{u}$, each of which may vary in time.

The formalism produces estimates $\hat{\mathbf{x}}(t)|\mathbf{z}(t' \leq t)$, i.e. estimates of the state vector at time t based on the observations at time t and all previous observations. However the validity of the estimates, and their associated covariance matrices, depends on the validity of the statistical model

defined by (7.1–2). In view of the potential difficulty in obtaining the requisite *a priori* information, it may be appropriate to adopt a 'conservative' approach and construct the model with a larger-than-expected *a priori* uncertainty to avoid the risk of biasing the estimation by forcing it to fit unrealistically precise prior values. The main mechanism for doing this is through the matrix $\mathbf{Q}$ which defines the stochastic forcing.

Among the applications of Kalman filtering to biogeochemical problems have been:

- A time series analysis of the Mauna Loa record by Surendran and Mulholland (1986). They estimated growth rates, in order to determine an airborne fraction;
- Deconvolution analyses by Trudinger (1999 and personal communication), making use of the technique's ability to handle multi-component non-stationary data, in order to intrepret CO_2 and ^{13}C records from ice and firn;
- Global inversions to estimate CFC lifetimes (Hartley and Prinn, 1993); and
- Estimation of methane emissions from north-west Europe using a regional transport model (Stijnen et al., 1997).

8 Concluding remarks

Some of the key requirements for making consistent comparisons between different views of the global carbon cycle (or other biogeochemical system) are:

- *The comparisons must refer to the same quantities.* For CO_2 for example, there needs to be consistent usage of storage or flux budgets;
- *The comparisons must refer to the same time period.* In addition to the strong natural variability, the anthropogenic perturbations to the carbon cycle and other biogeochemical cycles are changing over time; and
- *When comparing averaged quantities (i.e. almost always) the form of averaging needs to be equivalent.* This applies to spatial averaging as well as temporal averaging.

The examples given in the preceeding sections of this report illustrate the implications for time series analysis that these consistency requirements impose.

Among the important issues in time series analysis that are identified in this report are:

- an analysis of the autocovariance structure of time series estimates is an important part of describing the significance of any time series estimation procedure;

- the error amplification from numerical differentiation is a major source of error on short time-scales and dominates the uncertainty in many deconvolution calculations; and
- there is a need for specific definitions of what is being estimated.

It is expected that many of the principles illustrated in this report will have wider utility in biogeochemical studies, particularly when generalised to non-stationary processes. Two likely examples are the analysis of spatial and spatio-temporal variability and the analysis of synoptic and sub-synoptic time variability. For the present, the principles outlined here are being incorporated into re-analyses of our synthesis inversion calculations.

Acknowledgements

The author wishes to thank Rachel Law for providing the data for Figure 21 prior to publication. Various sections of this report are based on joint work and/or discussions with colleagues: Aihong Zhong for analyses of the seasonal cycles; Cathy Trudinger for splines and state-space modelling; Roger Francey for ^{13}C and carbon budgeting and Fortunat Joos for discussions on deconvolution. David Karoly, Peter Rayner, Ian Galbally, Graeme Pearman and Cathy Trudinger have made valuable comments on the manuscript.

References

Bacastow, R.B. (1976) Modulation of atmospheric carbon dioxide by the Southern Oscillation. *Nature*, **261**, 116–118.

Bloomfield, P. (1976) *Fourier Analysis of Time Series: An Introduction*. (Wiley: New York).

de Boor, C. (1978) *A Practical Guide to Splines*. (Springer-Verlag: New York).

Bruno, M. and Joos, F. (1997) Terrestrial carbon storage during the past 200 years: A Monte Carlo analysis of CO_2 data from ice core and atmospheric measurements. *Glob. Biogeochem. Cycles*, **11**, 111–124.

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

Conway, T.J., Tans, P.P., Waterman, L.S., Thoning, K.W., Kitzis, D.R., Masarie, K.A. and Zhang, N. (1994) Evidence for interannaul variability of the carbon cycle from the National Oceanic and Atmospheric Administration/Climate Monitoring and Diagnostics Laboratory global air sampling network. *J. Geophys. Res.*, **99D**, 22831–22855.

Craven, P. and Wahba, G. (1979) Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.* **31**, 377–403.

Dai, A. and Fung, I.Y. (1993) Can climate variability contribute to the "missing" CO_2 sink. *Glob. Biogeochem. Cycles*, **7**, 599–609.

Dargaville, R.J. (1999) Ph.D. Thesis (University of Melbourne), (in preparation).

Dargaville, R.J., Law, R.M. and Pribac, F. (1999) Implications of interannual variability in atmospheric transport on modelled CO_2 concentrations and source estimates. (in preparation).

Efron, B. (1982) *The jackknife, the bootstrap and other resampling plans.* (SIAM: Philadelphia).

Enting, I.G. (1985) A lattice statistics model for the age distribution of air bubbles in polar ice. *Nature*, **315**, 654–655.

Enting, I.G. (1986) Potential problems with the use of least squares spline fits to filter CO_2 data. *J. Geophys. Res.*, **91D**, 6666–6670.

Enting, I.G. (1987a) On the use of smoothing splines to filter CO_2 data. *J. Geophys. Res.*, **92D**, 10977–10984.

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

Enting, I.G. (1989) Time series studies of CO_2 data in the Division of Atmospheric Research, CSIRO, Australia. pp91–100 of *The Statistical Treatment of CO_2 Data records*. Ed. W.P. Elliott. NOAA Technical Memorandum ERL ARL-173. (U.S. Dept. Commerce: Washington).

Enting, I.G. and Mansbridge, J.V. (1987) Inversion relations for the deconvolution of CO_2 data from ice cores. *Inverse Problems*, **3** L63–L69.

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

Enting, I.G. and Newsam, G.N. (1990) Inverse problems in atmospheric constituent studies: II. Sources in the free atmosphere. *Inverse Problems*, **6**, 349–362.

Enting, I.G. and Pearman, G.I. (1993) Average global distributions of CO_2. pp31–64 of *The Global Carbon Cycle*. Ed. M. Heimann. NATO ASI Series. Vol. I 15. (Springer-Verlag:

Berlin).

Enting, I.G., Trudinger, C.M., Francey, R.J. and Granek, H. (1993) *Synthesis inversion of atmospheric CO_2 data using the GISS tracer transport model.* Division of Atmospheric Research, Technical Paper No. 29. (CSIRO: Australia).

Enting, I.G., Wigley, T.M.L, and Heimann, M. (1994) *Future Emissions and Concentrations of Carbon Dioxide: Key Ocean/Atmosphere/Land Analyses*. Division of Atmospheric Research, Technical Paper No. 31. (CSIRO: Australia).

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

Enting, I.G., Zhong, A. and Trudinger, C.M. (1999) *The Seasonal Cycle of Atmospheric Carbon Dioxide: Analytic and Synthetic Studies.* (CSIRO Atmospheric Research Technical Paper, in preparation)

Etheridge, D.M., Steele, L.P., Langenfelds, R.L., Francey, R.J., Barnola, J.-M. and Morgan, V.I. (1996) Natural and anthropogenic changes in atmospheric CO_2 over the last 1000 years from air in Antarctic ice and firn. *J. Geophys. Res.*, **101D**, 4115-4128.

Francey, R.J., Tans, P.P., Allison, C.E., Enting, I.G., White, J.W.C. and Trolier, M. (1995a) Changes in oceanic and terrestrial carbon uptake since 1982. *Nature*, **373**, 326–330.

Francey, R.J., Allison, C.E. and Welch, E.D. (1995b) The 11-year high precision *in situ* stable isotope record from Cape Grim 1982–1992. pp16–25 of *Baseline Atmospheric Program Australia 1992.* Ed. A.L. Dick and P.J. Fraser. (Dept. of Environment, Sport and Territories (Bureau of Meteorology) and CSIRO Division of Atmospheric Research).

Fung, I., Field, C.B., Berry, J.A., Thompson, M.V., Randerson, J.T., Malmström, C.M., Vitousek, P.M., Collattz, G.J., Sellers, P.J., Randall, D.A., Denning, A.S., Badeck, F. and John, J. (1997) Carbon 13 exchanges between the atmosphere and the biosphere. *Glob. Biogeochem. Cycles*, **11**, 507–533.

Gelb, A. (Ed) (1974) *Applied Optimal Estimation*. (MIT Press: Cambridge Mass.).

Granek, H. (1995) Generalized smoothing splines in CO_2 analysis. *J. Geophys. Res.*, **100D**, 16857–16865.

Hartley, D. and Prinn, R. (1993) Feasibility of determining surface emissions for trace gases using an inverse method in a three-dimensional chemical transport model. *J. Geophys. Res.*, **98D**, 5183–5197.

Heimann, M. and Maier-Reimer, E. (1996) On the relations between the oceanic uptake of CO_2 and its carbon isotopes. *Glob. Biogeochem. Cycles*, **10**, 89-110.

Joos, F. and Bruno, M. (1998) Long-term variability of the terrestrial and oceanic carbon sinks and the budgets of the carbon isotopes ^{13}C and ^{14}C. *Glob. Biogeochem. Cycles*, **12**, 277–295.

Joos, F. Bruno, M., Fink, R, Siegenthaler, U., Stocker, T.F., Le Quéré, C. and Sarmiento, J.L. (1996) An efficient and accurate representation of complex oceanic and biospheric models of anthropogenic carbon uptake. *Tellus*, **48B**, 397–417.

Keeling, C.D. and Whorf, T.P. (1994) Atmospheric CO_2 records from sites in the SIO sampling network. pp 12–26 of *Trends '93: A Compendium of Data on Global Change.* Ed. T.A. Boden, D.P. Kaiser, R.J. Sepanski and F.W. Stodd. (ORNL/CDIAC, Oak Ridge). Data set updated from CDIAC internet site.

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

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

Lee, K., Wanninkhof, R., Takahashi, T., Doney, S.C. and Feely, R.A. (1998) Low interannual variability in recent oceanic uptake of atmospheric carbon dioxide. *Nature*, **396**, 155–159.

Martín, F. and Díaz, A. (1991) Different methods of modeling the variability in the monthly mean concentrations of atmospheric CO_2 at Mauna Loa. *J. Geophys. Res.*, **96D**, 18689–18704.

Mulquiney, J.E. and Norton, J.P. (1998) A new inverse method for trace gas flux estimation 1. State-space model identification and constraints. *J. Geophys. Res.*, **103D**, 1417–1427.

Oeschger, H. and Heimann, M. (1983) Uncertainties of predictions of future CO_2 concentrations. *J. Geophys. Res.*, **88C**, 1258–1262.

Prather, M., Derwent, R., Ehhalt, D., Fraser, P., Sanhueza, E. and Zhou, X. (1995) Other trace gases and atmospheric chemistry. pp73–126 of *Climate Change 1994. Radiative Forcing of Climate Change and An Evaluation of the IPCC IS92 Emission Scenarios.* Ed. J.T Houghton, L.G. Meira Filho, J. Bruce, Hoesung Lee, B.A. Callander, E. Haites, N. Harris and K. Maskell. (Published for the IPCC by CUP.: Cambridge UK).

Priestley, M.B. (1981) *Spectral Analysis and Time Series.* (Academic Press: London).

Quay, P.D., Tilbrook, B. and Wong, C.S. (1992) Oceanic uptake of fossil fuel CO_2: Carbon-13 evidence. *Science*, **256**, 74–79.

Rayner, P.J., Enting, I.G., Francey, R.J. and Langenfelds, R. (1999) Reconstructing the recent carbon cycle from atmospheric CO_2, $\delta^{13}C$ and O_2/N_2 observations. *Tellus* (in press)

Sarmiento, J.L. and Sundquist. E.T. (1992) Revised budget for the oceanic uptake of anthropogenic carbon dioxide. *Science*, **356**, 589–593.

Schimel, D., Enting, I.G., Heimann, M., Wigley, T.M.L., Raynaud, D., Alves, D. and Siegenthaler, U. (1995) CO_2 and the Carbon Cycle. in *Climate Change 1994; Radiative Forcing of Climate Change*. Ed. J.T. Houghton, L.G. Meira Filho, J. Bruce, Hoesung Lee, B.A. Callander, E. Haites, N. Harris and K. Maskell. (Published for the IPCC by CUP: Cambridge, UK).

Siegenthaler, U. and Oeschger, H. (1987) Biospheric CO_2 emissions during the past 200 years reconstructed by deconvolution of ice core data. *Tellus*, **39B**, 140–154

Silverman, B.W. (1984) Spline smoothing: the equivalent variable kernel method. *Annals of Statistics*, **12**, 896–916.

Silverman, B.W. (1985) Some aspects of the spline smoothing approach to non-parameteric regression curve fitting. (with discussion). *J. Roy. Statist. Soc.*, **B47**, 1–52.

Stijnen, J., Heemink, A.W., Janssen, L.H.J.M. and van der Wal, J.T. (1997) Estimation of methane emissions in Europe using Kalman smoothing. pp133-142 of *Proceedings of the CKO/CCB Workshop on Bottom-up and Top-down Estimates of Greenhouse Gases in Bilthoven on 27-6-1997.* Report no. 728001 006. (Rijksinstituut voor Volkgezendheid en Milieu: Netherlands).

Surendran, S. and Mulholland, R.J. (1986) Estimation of atmospheric CO_2 concentration using Kalman filtering. *Int. J. System. Sci.* **17**, 897–909.

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

Tans, P.P. (1980) On calculating the transfer of carbon-13 in reservoir models of the carbon cycle. *Tellus*, **32**, 464–469.

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

Tans, P.P., Berry, J.A., and Keeling, R.F. (1993) Oceanic $^{13}C/^{12}C$ observations: a new window

on oceanic CO_2 uptake. *Glob. Biogeochem. Cycles*, **7**, 353–368.

Thompson, M.L., Enting, I.G., Pearman, G.I. and Hyson, P. (1986) Interannual variation in atmospheric CO_2 concentration. *J. Atmos. Chem.* **4**, 125–155.

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

Trenberth K.E. (1992) *Global Analyses from ECMWF (and an Atlas of 1000 to 10 mb Circulation Statistics)*. NCAR Technical Note NCAR/TN-373+STR. (National Center for Atmospheric Research: Boulder, Colorado).

Trudinger, C.M., Enting, I.G., Etheridge, D.M., Francey, R.J., Levchenko, V.A. and Steele, L.P. (1997) Modeling air movement and bubble trapping in firn *J. Geophys. Res.*, **102D**, 6747–6763.

Trudinger, C.M. (1999) Ph.D. Thesis, Monash University. (in preparation).

Tunnicliffe-Wilson, G. (1989) Analysis of selected time-series of atmospheric carbon dioxide concentrations. pp82–90 of *The Statistical Treatment of CO_2 Data records*. Ed. W.P. Elliott. NOAA Technical Memorandum ERL ARL-173. (U.S. Dept. Commerce: Washington).

Wigley, T.M.L. (1991) A simple inverse carbon cycle model. *Glob. Biogeochem. Cycles*, **5**, 373–382.

Young, P. (1984) *Recursive Estimation and Time-Series Analysis*. (Springer-Verlag: Berlin).

Notation

$\langle g(t)\rangle_\gamma$ Time average of $g(t)$, formed by convolution with kernel ψ_γ.

$\bar{g}(t)$ Time average of function $g(t)$ when averaging kernel is defined by context.

$\dot{g}(t)$ Time derivative of function $g(t)$.

$\hat{g}(t)$ Statistical estimate of function $g(t)$.

$\langle\langle z\rangle\rangle_\gamma$ Result of applying digital filter with coefficients $\psi_{k;\gamma}$ to time series $z(t)$.

$a(t)$ General time series.

$A(t)$ Amplitude of seasonal cycle.

$\mathbf{B}$ Backwards shift operator: $\mathbf{B}z_t = z_{t-1}$.

$f_a(\theta)$ Power spectrum of time series $a(t)$.

$\mathbf{F}(t)$ Time evolution matrix in state-space model.

$g(t)$ General time series representing a 'signal' of interest.

$G(t)$ Green's function.

$\mathbf{H}(t)$ Matrix defining relation between observations and state vector in state space model.

K_p Vertical diffusion constant (for toy model in Section 6.2).

M Atmospheric carbon content.

$n(t)$ Error (noise) time series. ($\mathbf{n}(t)$ in state-space models with multi-component observations).

N Number of data points in time series.

p Atmospheric pressure.

r_s Standard ^{13}C:^{12}C ratio for defining δ^{13}C for measurements isotopic differences.

$R(t)$ Autocovariance of time series.

$R_g(t)$ Autocovariance of function $g(t)$.

R_r Reference ^{13}C:C ratio for defining isotopic anomalies.

R_X Isotopic ratio of flux Φ_X.

$s(t)$ Seasonal component of CO_2 concentration.

$\mathbf{u}(t)$ Deterministic forcing in state-space model.

$\mathbf{w}(t)$ Stochastic forcing in state-space model.

$\mathbf{x}(t)$ State vector in state-space model.

$z(t)$ Observational data. ($\mathbf{z}(t)$ in state-space models with multi-component observations).

δ_A Carbon isotope ratio anomaly (relative to R_r) of the atmosphere.

δ_X Isotopic enhancement (or depletion if negative) of flux Φ_X, relative to reference with ratio R_r.

δ' Standard δ^{13}C measure of isotopic anomaly, expressed in terms of ^{13}C:^{12}C ratios.

Δt Time spacing in time series.

β Generic label for specifying kernels that are used to produce functionals by the process of convolution.

γ Generic label for averaging kernels and digital filters.

ϵ White-noise forcing in ARIMA models (and special cases).

λ_ω Vertical scale (in pressure coordinates) of concentration variations at frequency ω, from toy model described in Section 6.2.

$\phi(t)$ Phase of seasonal cycle.

Φ_X Net carbon flux due to process X.

ψ_k kth coefficient of digital filter.

$\psi_{k:\gamma}$ kth coefficient of digital filter γ.

$\Psi_\gamma(\theta)$ Frequency response of filter with coefficients $\psi_{k:\gamma}$.

θ Dimensionless frequency, range $[-\pi, \pi]$.

λ Weighting factor used in defining smoothing splines.

ω Angular frequency, range $[-\pi/2\Delta t, \pi/2\Delta t]$.

ω_A Angular frequency of annual cycle, 2π y^{-1}.

ω_c Angular frequency for 50% attenuation by smoothing spline.

ν Frequency, range $[-1/2\Delta t, 1/2\Delta t]$.

CSIRO Atmospheric Research Technical Papers

This series has been issued as *Division of Atmospheric Research Technical Paper* (nos. 1–19); *CSIRO Division of Atmospheric Research Technical Paper* (nos. 20–37) and *CSIRO Atmospheric Research Technical Paper* from no. 38.

Regular electronic publication commenced with no. 45. Earlier technical papers are progressively being made available in electronic form. A current list of technical papers is maintained at *http://www.dar.csiro.au/info/TP.htm.* Papers may be issued out of sequence.

No. 1 Galbally, I.E.; Roy, C.R.; O'Brien, R.S.; Ridley, B.A.; Hastie, D.R.; Evans, W.J.F.; McElroy, C.T.; Kerr, J.B.; Hyson, P.; Knight, W.; Laby, J.E. *Measurements of trace composition of the Austral stratosphere: chemical and meteorological data.* 1983. 31 p.

No. 2 Enting, I.G. *Error analysis for parameter estimates from constrained inversion.* 1983. 18 p.

No. 3 Enting, I.G.; Pearman, G.I. *Refinements to a one-dimensional carbon cycle model.* 1983. 35 p.

No. 4 Francey, R.J.; Barbetti, M.; Bird, T.; Beardsmore, D.; Coupland, W.; Dolezal, J.E.; Farquhar, G.D.; Flynn, R.G.; Fraser, P.J.; Gifford, R.M.; Goodman, H.S.; Kunda, B.; McPhail, S.; Nanson, G.; Pearman, G.I.; Richards, N.G.; Sharkey, T.D.; Temple, R.B.; Weir, B. *Isotopes in tree rings.* 1984. 86 p.

No. 5 Enting, I.G. *Techniques for determining surface sources from surface observations of atmospheric constituents.* 1984. 30 p.

No. 6 Beardsmore, D.J.; Pearman, G.I.; O'Brien, R.C. *The CSIRO (Australia) Atmospheric Carbon Dioxide Monitoring Program: surface data.* 1984. 115 p.

No. 7 Scott, John C. *High speed magnetic tape interface for a microcomputer.* 1984. 17 p.

No. 8 Galbally, I.E.; Roy, C.R.; Elsworth, C.M.; Rabich, H.A.H. *The measurement of nitrogen oxide (NO, NO_2) exchange over plant/soil surfaces.* 1985. 23 p.

No. 9 Enting, I.G. *A strategy for calibrating atmospheric transport models.* 1985. 25 p.

No. 10 O'Brien, D.M. *TOVPIX: software for extraction and calibration of TOVS data from the high resolution picture transmission from TIROS-N satellites.* 1985. 41 p.

No. 11 Enting, I.G.; Mansbridge, J.V. *Description of a two-dimensional atmospheric transport model.* 1986. 22 p.

No. 12 Everett, J.R.; O'Brien, D.M.; Davis, T.J. *A report on experiments to measure average fibre diameters by optical fourier analysis.* 1986. 22 p.

No. 13 Enting, I.G. *A signal processing approach to analysing background atmospheric constituent data.* 1986. 21 p.

No. 14 Enting, I.G.; Mansbridge, J.V. *Preliminary studies with a two-dimensional model using transport fields derived from a GCM.* 1987. 47 p.

No. 15 O'Brien, D.M.; Mitchell, R.M. *Technical assessment of the joint CSIRO/Bureau of Meteorology proposal for a geostationary imager/sounder over the Australian region.* 1987. 53 p.

No. 16 Galbally, I.E.; Manins, P.C.; Ripari, L.; Bateup, R. *A numerical model of the late (ascending) stage of a nuclear fireball.* 1987. 89 p.

No. 17 Durre, A.M.; Beer, T. *Wind information prediction study: Annaburroo meteorological data analysis.* 1989. 30 p. + diskette.

No. 18 Mansbridge, J.V.; Enting, I.G. *Sensitivity studies in a two-dimensional atmospheric transport model.* 1989. 33 p.

No. 19 O'Brien, D.M.; Mitchell, R.M. *Zones of feasibility for retrieval of surface pressure from observations of absorption in the A band of oxygen.* 1989. 12 p.

No. 20 Evans, J.L. *Envisaged impacts of enhanced greenhouse warming on tropical cyclones in the Australian region.* 1990. 31 p. [Out of print]

No. 21 Whetton, P.H.; Pittock, A.B. *Australian region intercomparison of the results of some general circulation models used in enhanced greenhouse experiments.* 1991. 73 p. [Out of print]

No. 22 Enting, I.G. *Calculating future atmospheric CO_2 concentrations.* 1991. 32 p. Also electronic edition (718 kB pdf), at *http://www.dar.csiro.au/publications/Enting_2000d.pdf*

No. 23 Kowalczyk, E.A.; Garratt, J.R.; Krummel, P.B. *A soil-canopy scheme for use in a numerical model of the atmosphere — 1D stand-alone model.* 1992. 56 p.

No. 24 Physick, W.L.; Noonan, J.A.; McGregor, J.L.; Hurley, P.J.; Abbs, D.J.; Manins, P.C. *LADM: A Lagrangian Atmospheric Dispersion Model.* 1994. 137 p.

No. 25 Enting, I.G. *Constraining the atmospheric carbon budget: a preliminary assessment.* 1992. 28 p. Also electronic edition (571 kB), at *http://www.dar.csiro.au/publications/Enting_2000b.pdf*

No. 26 McGregor, J.L.; Gordon, H.B.; Watterson, I.G.; Dix, M.R.; Rotstayn, L.D. *The CSIRO 9-level atmospheric general circulation model.* 1993. 89 p.

No. 27 Enting, I.G.; Lassey, K.R. *Projections of future CO_2.* with appendix by R.A. Houghton. 1993. 42 p. Also electronic edition (860 kB pdf), at *http://www.dar.csiro.au/publications/Enting_2000e.pdf*

No. 28 [Not published]

No. 29 Enting, I.G.; Trudinger, C.M.; Francey, R.J.; Granek, H. *Synthesis inversion of atmospheric CO_2 using the GISS tracer transport model.* 1993. 44 p.

No. 30 O'Brien, D.M. *Radiation fluxes and cloud amounts predicted by the CSIRO nine level GCM and observed by ERBE and ISCCP.* 1993. 37 p.

No. 31 Enting, I.G.; Wigley, T.M.L.; Heimann, M. *Future emissions and concentrations of carbon dioxide: key ocean/atmosphere/land analyses.* 1993. 120 p.

No. 32 Kowalczyk, E.A.; Garratt, J.R.; Krummel, P.B. *Implementation of a soil-canopy scheme into the CSIRO GCM – regional aspects of the model response.* 1994. 59 p.

No. 33 Prata, A.J. *Validation data for land surface temperature determination from satellites.* 1994. 36 p.

No. 34 Dilley, A.C.; Elsum, C.C. *Improved AVHRR data navigation using automated land feature recognition to correct a satellite orbital model.* 1994. 22 p.

No. 35 Hill, R.H.; Long, A.B. *The CSIRO dual-frequency microwave radiometer.* 1995. 16 p.

No. 36 Rayner, P.J.; Law, R.M. *A comparison of modelled responses to prescribed CO_2 sources.* 1995. 84 p.

No. 37 Hennessy, K.J. *CSIRO Climate change output.* 1998. 23 p.

No. 38 Enting, I.G. *Attribution of greenhouse gas emissions, concentrations and radiative forcing.* 1998. 27 p. Also electronic edition (557 kB pdf), at *http://www.dar.csiro.au/publications/Enting_2000c.pdf*

No. 39 O'Brien, D.M.; Tregoning, P. *Geographical distributions of occultations of GPS satellites viewed from a low earth orbiting satellite.* (1998) 28p.

No. 40 Enting, I.G. *Characterising the temporal variability of the global carbon cycle.* 1999. 53 p. Also electronic edition (636 kB pdf),
at *http://www.dar.csiro.au/publications/Enting_2000a.pdf*

No. 41 (in preparation).

No. 42 Mitchell, R.M. *Calibration status of the NOAA AVHRR solar reflectance channels: CalWatch revision 1.* 1999. 20 p.

No. 43 Hurley, P.J. *The Air Pollution Model (TAPM) Version 1: technical description and examples.* 1999. 41 p. Also electronic edition (276 kB pdf),
at *http://www.dar.csiro.au/publications/Hurley_1999a.pdf*

No. 44 Frederiksen, J.S.; Dix, M.R.; Davies, A.G. *A new eddy diffusion parameterisation for the CSIRO GCM.* 2000. 31 p.

No. 45 Young, S.A. *Vegetation Lidar Studies.* Electronic edition in preparation.

No. 46 Prata, A. J. *Global Distribution of Maximum Land Surface Temperature Inferred from Satellites: Implications for the Operation of the Advanced Along Track Scanning Radiometer.* 2000. 30 p. Electronic edition only (8307 kB pdf),
at *http://www.dar.csiro.au/publications/Prata_2000a.pdf*

No. 47 Prata, A. J. *Precipitable water retrieval from multi-filter rotating shadowband radiometer measurements.* 2000. 14 p. Electronic edition only (1554 kB pdf),
at *http://www.dar.csiro.au/publications/Prata_2000b.pdf*

No. 48 Prata, A.J. and Grant, I.F. *Determination of mass loadings and plume heights of volcanic ash clouds from satellite data.* (in preparation)

No. 49 (in preparation)

No. 50 (in preparation)

No. 51 Meyer, C.P., Galbally, I.E., Wang, Y.-P., Weeks, I.A., Jamie, I. and Griffith, D.W.T. *Two automatic chamber techniques for measuring soil-atmosphere exchanges of trace gases and results of their use in the OASIS field experiment.* Electronic edition only (pdf),
at *http://www.dar.csiro.au/publications/Meyer_2001a.pdf*

Address and contact details: CSIRO Atmospheric Research
Private Bag No.1 Aspendale Victoria 3195 Australia
Ph: (+61 3) 9239 4400; fax: (+61 3) 9239 4444
e-mail: chief@dar.csiro.au

Last updated 27 March 2001