A Signal Processing Approach to Analysing Background Atmospheric Constituent Data

I. G. Enting

DIVISION OF ATMOSPHERIC RESEARCH TECHNICAL PAPER No. 13

COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH ORGANIZATION, AUSTRALIA 1986

A Signal Processing Approach to Analysing Background Atmospheric Constituent Data

By I. G. Enting

Division of Atmospheric Research Technical Paper No. 13

Commonwealth Scientific and Industrial
Research Organization, Australia
1986

A signal processing approach to analysing background atmospheric constituent data

I.G. Enting

CSIRO, Division of Atmospheric Research
Private Bag 1, Mordialloc, Vic, 3195, Australia.

Abstract

A number of techniques such as spline-smoothing and complex-demodulation have previously been applied to the analysis of background atmospheric constituent data. It is pointed out that an interpretation in terms of digital filtering unifies these approaches. Earlier studies based on complex demodulation are re-examined in terms of the equivalent digital filtering operations.

1. Introduction

The variability of background atmospheric constituent data contains much information that is of potential use in understanding natural and perturbed geochemical cycles. In particular we can often seek to identify:

(i) long-term trends (possibly due to anthropogenic influences);
(ii) interannual variability (either natural or anthropogenic);
(iii) seasonal cycles (including interannual variations in cycles);
(iv) synoptic variations;
(v) variability due to the process of selecting data to obtain a 'baseline' level that is believed to be representative of large-scale air masses;
(vi) variability arising from the measurement process.

The analysis of baseline data has been undertaken using a variety of techniques such as polynomial (and other) regression, spline fitting and complex demodulation. A useful approach is to treat the analysis problem as one of signal processing with an emphasis on digital filtering. This unifies much earlier work and clarifies the determination of uncertainties. In this report we only consider the statistical analysis of a single time series and do not consider multiple-time series analysis.

2. Basic concepts

The aim of this section is to describe a number of the most important concepts for the use of signal processing techniques in the analysis of baseline data. Detailed mathematical results are given in later sections.

2a. Statistical analysis requires a statistical model

The statistical assumption that we make is to express an observed time series $z(t)$ as a sum of a signal $y(t)$ plus a noise term $\varepsilon(t)$, i.e.

$$z(t) = y(t) + \varepsilon(t) \tag{1a}$$

or for discrete records,

$$z_j = y_j + \varepsilon_j \quad . \tag{1b}$$

In the context of analysing atmospheric constituent data, the signal can be any combination of the 6 types of variations listed in Section 1, with the noise consisting of the sum of all components that are not regarded as 'signal'. It is assumed that $y(t)$ and $\varepsilon(t)$ are uncorrelated since any correlation would imply that $\varepsilon(t)$ is incorporating part of the signal $y(t)$. Much of the theory of time series analysis is confined to stationary series with zero mean but many of the filtering operations described below are of greater generality. In some aspects of the analysis it may be necessary to remove a mean and trend from the series before analysing it in order to approximate a stationary series. The additive form of equation (1) is reasonable since virtually all constituent concentrations are smooth functions of their source/sink strengths. For conserved constituents the contributions from different types of source combine additively and even for non-conserved tracers, the influences of perturbations can be linearised so that the influences of different types of perturbation will be approximately additive.

2b. The construction of estimates of the signal

The analysis procedures considered here are concerned with estimating the signal $y(t)$, given only the information in the observed record $z(t)$. The estimates of y will be denoted $\hat{y}$. If only the information from within the records is used then there is no basis for giving any special significance to particular time points. It is thus reasonable to use only stationary procedures so far as is possible. The non-stationary process of fitting monthly mean data by splines with nodes each 12 months has often been used in the analysis of CO_2 data. Enting (1986) has described some of the unsatisfactory behaviour that the non-stationarity causes in this process.

Although non-linear digital filtering of CO_2 data has been described by Cleveland et al. (1983), in this report we confine our attention to linear operations so that

$$\hat{y}_j = \sum_k G_{jk} z_k \quad . \tag{2}$$

For stationary processes, G_{kj} depends only on j-k and so we can write the estimates $\hat{y}$ in the form

$$\hat{y}_j = \sum_{k=-K}^{K} c_k z_{j-k} \quad . \tag{3}$$

The expression (3) represents the use of a digital filter to extract $\hat{y}$ which is an estimate of the signal y. For practical purposes we will only consider finite K in equation (3) and refer to K as the filter 'length' even though the number of terms is 2K+1.

Within K time units of the end of a record, equation (3) cannot be applied and if estimates are required then the more general form (2) must be used.

2c. Spectral representation

When the pure digital filtering operation is used then a spectral representation is of particular value. This is because the functions $\exp(ij\theta)$ (where $i = \sqrt{-1}$) are eigenfunctions of the filtering operation (Hamming, 1977). The corresponding eigenvalue, $H(\theta)$, is the transfer function of the filter and is defined by

$$H(\theta) = \sum_{k=-K}^{K} c_k \exp(-ik\theta) \quad . \tag{4}$$

The interpretation of θ as an angular frequency of a component of the data implies that time is being expressed in units such that the data spacing is $\Delta t=1$. This choice will be used unless otherwise indicated. Symmetric filters (i.e. $c_k=c_{-k}$) have transfer functions that are purely real so that phase shifts are 0 or π. The spectral characterisation of the time series $y(t)$, $z(t)$ and $\varepsilon(t)$ is in terms of their spectral density functions $f_{yy}(\theta)$, $f_{zz}(\theta)$, $f_{\varepsilon\varepsilon}(\theta)$ where, for a general stationary sequence, g_j, with zero mean, the spectral density can be defined (Yaglom, 1962, p55) by

$$f_{gg}(\theta) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \exp(-ik\theta) B_k \quad , \tag{5}$$

$$B_k = E[g_j g_{j-k}] \tag{6}$$

where E[•] denotes statistical expectations.

The use of the spectral density functions is subject to two problems. The first is that only $f_{zz}(\theta)$ can be estimated from the observations. Any knowledge of f_{yy} and $f_{\varepsilon\varepsilon}$ must be based on assumptions about the processes involved. The second problem is that there are definite limitations on the estimation of f_{zz} from the data. Applying a discrete Fourier transform to that data z_i and then taking the square of the amplitude leads to a very irregular sequence of estimates of $f_{zz}(\theta)$. Increasing N, the length of the record, increases the number of independent frequencies θ_j at which f_{zz} is estimated but the variance of each estimate remains of order 1 rather than decreasing as $1/\sqrt{N}$. To obtain regular estimates of the spectral density function, various smoothing functions are applied. These have the effect of reducing the variance of the estimates at the cost of a loss of ability to resolve neighbouring frequencies. The estimation of spectral density functions is considered at length by Priestly (1981).

2d. Errors

When choosing a filtering operation the obvious aim is to extract an estimate of the signal which has minimum error. A useful measure of this is the mean square error in the signal which is given by

$$E[(\hat{y}_j-y_j)^2] = \int_{-\pi}^{\pi} [|1-H(\theta)|^2 f_{yy}(\theta) + |H(\theta)|^2 f_{\varepsilon\varepsilon}(\theta)]d\theta \qquad (7)$$

In the integrand, the first term represents the error due to biasing the signal by applying a filter that fails to pass it unchanged and the second term represents the error due to the filter passing some of the noise. Since the integrand is positive everywhere, the optimal filter is obtained if the integrand is minimised for each θ. Differentiating the integrand with respect to H(θ) gives

$$-2(1-H(\theta))f_{yy}(\theta) + 2H(\theta)f_{\varepsilon\varepsilon}(\theta) = 0 \qquad (8)$$

whence

$$H(\theta)_{opt} = f_{yy}(\theta)/(f_{yy}(\theta) + f_{\varepsilon\varepsilon}(\theta)) = f_{yy}(\theta)/f_{zz}(\theta) \qquad (9)$$

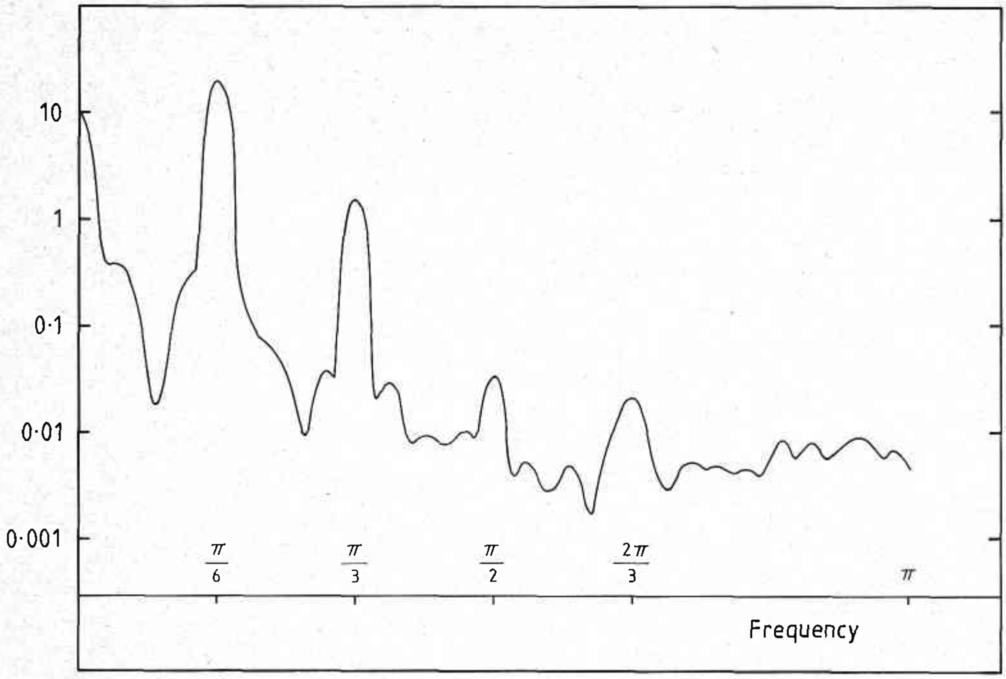
(see Yaglom, 1962, p132).

Figure 1a shows the estimated spectrum obtained from the monthly mean CO_2 concentrations at Mauna Loa by Thompson et al. (1986) by constructing a smoothed periodogram. The various features that can be seen in the spectrum are

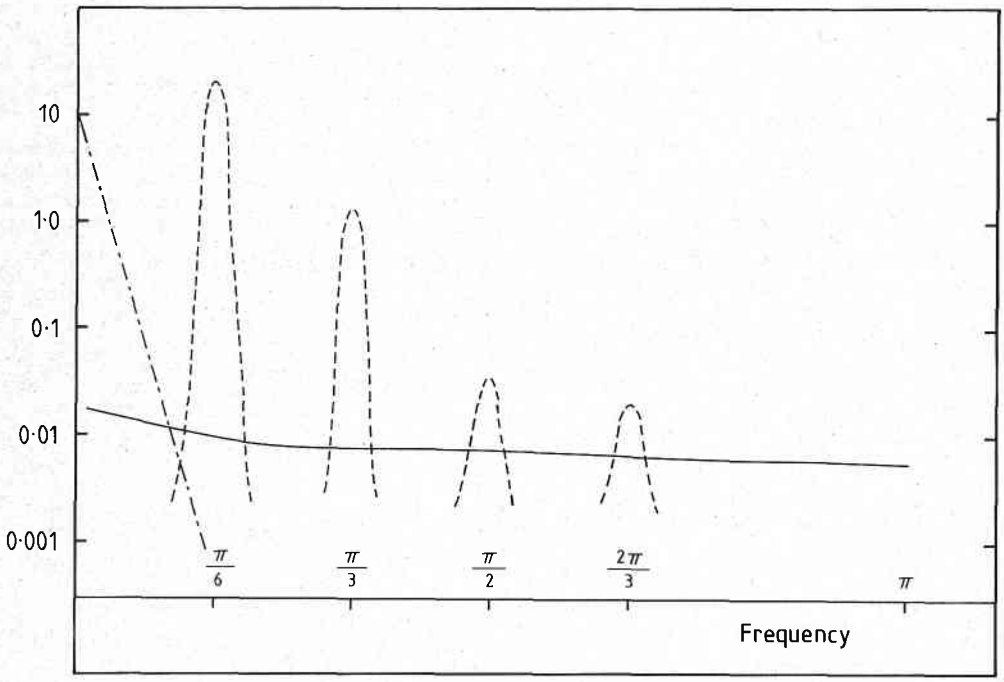
(i) peaks corresponding to a 1 year cycle and its harmonics
(ii) a low frequency component giving the interannual variability of the increase
(iii) a background noise component.

Because this is a set of smoothed values, no significance can be assigned to the widths of the seasonal peaks as a characterisation of the interannual variability of the seasonal cycle.

Figure 1b shows one possible division of the spectrum into different contributions. The chain curve on the left could be regarded as representing the long-term increase of CO_2, including the interannual variations in this increase. The set of dashed peaks represent the seasonal cycle, including its interannual variations. The solid curve represents a 'background' variation whose spectrum is close to a flat white noise spectrum. Hamming (1977) has



(a)



(b)

Figure 1: (a) Spectrum of the monthly mean CO_2 concentration at Mauna Loa, Hawaii, as estimated by Thompson et al. (1986).
(b) Schematic decomposition of the Mauna Loa spectrum into long-term (chain curve), seasonal (dashed) and approximately white noise (solid) components.

noted that such a nearly white-noise spectrum can often be expected to arise from the effects of 'aliasing' when sampling from a process with a decaying spectrum.

The type of division shown in Figure 1b forms the basis of the statistical model that underlies the analysis i.e. the model assumes that the time series consists of signal y and noise ε with spectral densities f_{yy} and $f_{\varepsilon\varepsilon}$ respectively. Any statistical analysis must, explicitly or implicitly, be based on a model of this type. Furthermore the statistical analysis will not, in itself, give any indication of whether the model is scientifically reasonable or useful.

2e. Limitations of filtering

The ability to separate a signal from noise is limited by two factors. The first is the theoretical difficulty that even the optimal filter defined by (9) has a mean square error

$$E((\hat{y}_{opt}-y)^2) = \int_{-\pi}^{\pi} (f_{yy}(\theta)f_{\varepsilon\varepsilon}(\theta)/f_{zz}(\theta))d\theta \qquad (10)$$

(Yaglom, 1962, equation 5.31).

This will be zero only if $y(t)$ and $\varepsilon(t)$ have non-overlapping spectra i.e. at least one of $f_{yy}(\theta)$ and $f_{\varepsilon\varepsilon}(\theta)$ is zero at each frequency. If this condition does not hold then it is impossible to separate the signal from the noise completely and unambiguously. An obvious extension of this result is that good estimates of the signal can be obtained if the spectra of the signal and the noise are only weakly overlapping while only poor estimates of the signal will be obtainable if the signal and noise have common peaks.

The second limitation on filtering is that we usually want to work with finite length filters i.e. finite K in equation (3), while the optimal filter is usually of infinite length. The design of a practical filter will thus involve various compromises. There is extensive literature on the design of digital filters (see for example Hamming, 1977).

2f. Smoothing splines as filters

The (cubic) smoothing spline is a function $\hat{y}(t)$ chosen so as to minimise the quantity

$$\phi = \sum_{j=1}^{N} (z_j - \hat{y}(t_j))^2 + \lambda \int_{t_1}^{t_N} \left(\frac{\partial^2 \hat{y}}{\partial t^2}\right)^2 dt \quad . \qquad (11)$$

The use of smoothing splines in the analysis of CO_2 data was introduced by Bacastow (1976). Since (11) is a least squares expression, the best fit $\hat{y}$ is given by a linear combination of the z_k as expressed by the general equation (2). Recent analysis (Cox, 1983; Silverman, 1984) has shown that, except near the ends of the records, the linear relation is accurately approximated by the filter relation (3). Cox (1983, equation 4.4) has shown that for equally spaced data, the limiting form of the transfer function is

$$H(\theta)_{spline} = (1 + \lambda\Delta t\theta^4)^{-1} \qquad (12)$$

making the spline process a low-pass filter. In this definition, both the angular frequency, θ, and the time step, Δt, are defined in terms of the same time units that are used in expression (11). Thus θ is defined on the range $[-\pi/\Delta t, \pi/\Delta t]$.

It must be noted that near the ends of the records the filtering expression (12) breaks down. The equivalent filter becomes asymmetric in time so that $H(\theta)$ is no longer purely real. Thus frequencies that are passed are subject to a phase shift of $\mathrm{Arg}[H(\theta)]$ in this region. Silverman (1984) has extended Cox's analysis to the case of unequally spaced data and has shown that the effective filter is only weakly dependent on the data density. This property makes smoothing splines suitable for interpolating the non-uniform records that arise from performing some form of baseline data selection. This filtering interpretation of smoothing splines suggests that λ should be chosen on the basis of the filtering properties expressed by equation (12) and its generalisations. Generally in CO_2 studies λ has effectively been estimated by eye, a procedure which is satisfactory since the filtering properties depend only weakly on λ over much of the frequency range. There are many statistical procedures for estimating λ from the data (Golub et al., 1979 and references therein) but these assume independent errors ε_i and behave very poorly if the ε_i are autocorrelated (Diggle and Hutchinson, 1986; Diggle in discussion section of Silverman, 1985). For studies of relatively long-term variations it is usually possible to obtain equally spaced data, possibly by using spline interpolation. Once such data are obtained, it is probable that smoothing splines will not be the optimal filters for further analysis and that specifically designed digital filters will give a superior separation of signal and noise. One limitation on the use of splines as low-pass filters is the relatively broad transition band. This could be reduced either by using specifically designed filters or, by using higher-order splines. Cox (1983) has noted if mth derivatives are used in the constraint term of (11) then the transfer function is of the form

$$H(\theta) = (1 + a\lambda\theta^{2m})^{-1} \quad . \qquad (13)$$

The sharpness of the transition from pass-band to stop-band increases with m.

2g. Complex demodulation

Complex demodulation (Bloomfield, 1976) is a technique for analysing a signal that is assumed to include a component of the form

$$y(t) \approx A(t)\cos(\omega t + \phi(t)) \qquad (14)$$

where $A(t)$ and $\phi(t)$ are slowly varying functions. The procedure gives estimates $\hat{A}(t)$ and $\hat{\phi}(t)$ which can, if desired, be combined to give an estimated signal

$$\hat{y}(t) = \hat{A}(t)\cos(\omega t + \hat{\phi}(t)) \quad . \qquad (15)$$

Thompson et al. (1986) used complex demodulation of CO_2 data to extract estimates $\hat{A}(t)$ and $\hat{\phi}(t)$ for both the annual cycle and its first harmonic. These quantities were constructed both as possible sources of information concerning the carbon cycle and as part of the process of examining long term trends. When examining long-term trends, the estimated signals $\hat{y}_{12}(t)$ and $\hat{y}_6(t)$ with periods 12 and 6 months respectively were constructed according to equation (15). These signals were then subtracted from the original data $z(t)$ to give a decycled data set

$$z_D(t) = z(t) - \hat{y}_{12}(t) - \hat{y}_6(t) \quad . \qquad (16)$$

An estimate, $\hat{y}_*(t)$, of the combination of the long-term trend plus interannual variability was obtained by applying an appropriate low-pass filter to $z_D(t)$. Much of the complexity of this procedure used by Thompson et al. (1986) can be avoided by noting that the signals $\hat{y}_{12}(t)$ and $\hat{y}_6(t)$ could be obtained from the original data by band-pass filtering. The mathematical description of complex-demodulation as a filtering operation is given in Section 4 below. If the transfer functions of the band-pass filters are denoted $H_{12}(\theta)$ and $H_6(\theta)$ then z_D is constructed by operating on z(t) with a band-reject filter whose transfer function is

$$H_D(\theta) = 1 - H_{12}(\theta) - H_6(\theta) \quad . \qquad (17)$$

Applying a low-pass filter with transfer function $H_L(\theta)$ means that $\hat{y}_*(t)$ can be obtained from z(t) by applying a filter whose transfer function is

$$H_*(\theta) = H_L(\theta)(1 - H_{12}(\theta) - H_6(\theta)) \quad . \qquad (18)$$

The suitability of this combination can be seen from the arguments of topic (e) above. A basic low-pass filter, H_L, is modified to ensure that the transfer function is very close to zero in bands where the original signal has peaks (i.e. 12 month and 6 month periods) that are regarded as noise to be excluded from the final signal. However, for the purposes of extracting long-term variations, it seems that direct application of appropriate low-pass filters may be more suitable (see Section 4 below).

2h. Filters for linear operations

A slight generalisation of the discussion above is needed if we require some operation to be performed on the signal. In particular, we may have a set of observations

$$z(t) = y(t) + \varepsilon(t) \qquad (19)$$

and wish to obtain an estimate of w(t) which is related to y(t) by some process described by a stationary linear operator

$$\text{i.e} \quad w(t) = \quad y(t) \quad . \qquad (20)$$

If we construct

$$\hat{w} = \quad z$$

for some stationary linear operator then the error expression (7) generalises to

$$E[(\hat{w}-w)^2] = \int_{-\pi}^{\pi} [|L(\theta) - F(\theta)|^2 f_{zz}(\theta) + |F(\theta)|^2 f_{\varepsilon\varepsilon}(\theta)]\,d\theta \qquad (21)$$

where $L(\theta)$ and $F(\theta)$ are the transfer functions for and respectively. The case in which this is of most interest is when is the differentiation operator as in the work of Bacastow (1976) and Thompson et al. (1986).

3. Low-pass filters

3a. General characteristics

One of the most basic types of filter is the low-pass filter. Ideally a low-pass filter has a transfer function

$$H(\theta) = 1 \quad , \ \theta < \omega_c \tag{22a}$$
$$\quad\;\; = 0 \quad , \ \theta > \omega_c \ . \tag{22b}$$

Such a filter would be used to extract the low-frequency part of a time series. It can also be used as a basis for constructing a high-pass filter with transfer function $1-H(\theta)$, and the difference of two low-pass filters (with different cutoff frequencies) will be a band-pass filter. Low-pass filters also play an important role in the process of complex demodulation which is described in Section 4. The sharp cutoff at ω_c as specified by equation (22) represents an ideal situation that would require an infinitely long filter. Any finite length filter will depart from this ideal in that there will be few, if any, points at which the transfer function will have exactly the values 0 or 1 and there will be a finite 'transition band' in which the transfer function drops from values near 1 to nearly 0. An additional problem that may occur is the appearance of oscillations in the transfer function, an effect known as 'Gibbs phenomenon.'

A low-pass filter will have the value 1 at only a finite number of frequencies due to the effects noted above and so, even in the pass band, the signal will be subject to some distortion. One measure of the quality of a low-pass filter is the number of derivatives of $H(\theta)$ that are zero at $\theta=0$.

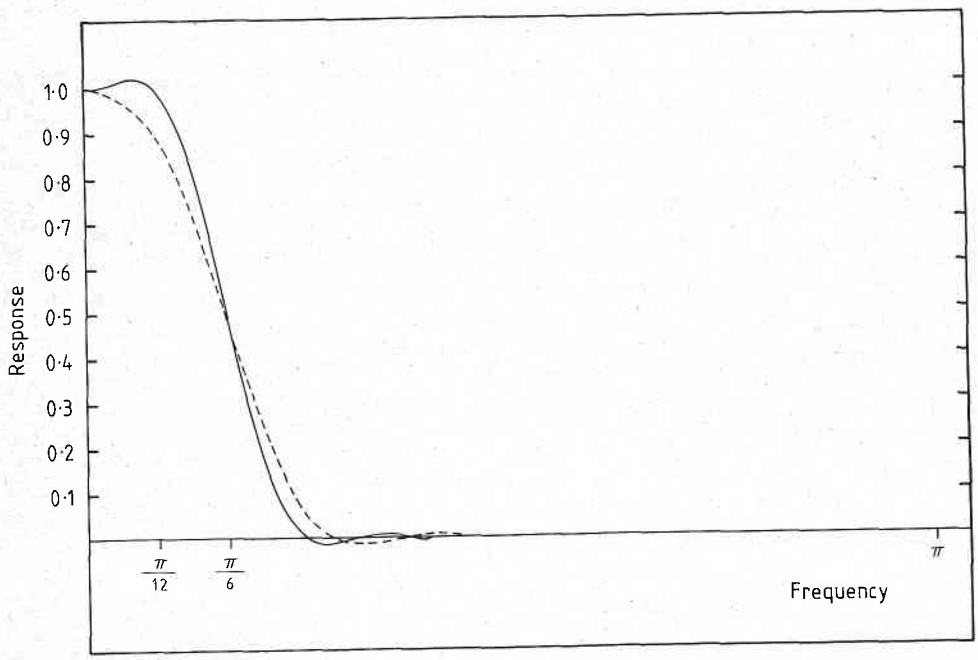


Figure 2: Transfer functions of the low-pass filters used by Thompson et al. (1986) when analysing decycled data. Solid curve, $\omega_c=\pi/7$, K=14 used for Mauna Loa data, dashed curve $\omega_c=\pi/6$, K=10 used for all other sites.

The importance of this is that if $H(0)=1$ and all derivatives up to order k (inclusive) are zero at the origin, then the filter passes a polynomial of degree k exactly. In particular, the transfer function of a symmetric filter will have a gradient of zero at the origin. If it also has a value of 1, it will pass a linear term exactly and so it will not be necessary to extract such terms prior to filtering.

3b. Low-pass filters using sigma factors

Bloomfield (1976) has defined a two-parameter set of low-pass filters based on the use of a smoothing technique to reduce the Gibbs oscillations. The ideal low-pass filter of equation (22) has filter coefficients

$$c_o = \omega_c/\pi \tag{23a}$$

$$c_k = c_{-k} = \sin(k\omega_c)/\pi k \quad . \tag{23b}$$

The strong ripples in the transfer function of this filter can be reduced by convolving it with an appropriate 'window'. This corresponds to multiplying each filter coefficient by a smoothing factor known as a sigma factor. Bloomfield uses

$$c_o = \alpha\omega_c/\pi \tag{24a}$$

$$c_k = c_{-k} = c_o \sin(k\omega_c) \sin(2\pi k/(2K+1))/(2\pi\omega_c k^2/(2K+1))$$

$$\text{for } k=1 \text{ to } K \quad , \tag{24b}$$

with α chosen so that $H(0)=1$. The two filter parameters are thus the cutoff frequency ω_c and the length K. The subroutine LOPASS (Bloomfield, 1976, p149) implements this filter. Bloomfield quotes the effective width of the transition band as

$$\delta = 4\pi/(2K + 1) \quad . \tag{25}$$

The two low-pass filters used directly by Thompson et al. (1986) were of this type and had (ω_c, K) given by $(\pi/7, 14)$ for analysing the Mauna Loa data and $(\pi/6, 10)$ for all other sites considered. The transfer functions for these filters are shown in Figure 2. It will be seen that the length 10 filter has noticeably poorer behaviour in the pass band. However both filters are symmetric and so have zero gradient at the origin and consequently pass linear trend terms exactly. This property was of potential importance in the study by Thompson et al. (1986) since their series were differentiated after low-pass filtering but in practice they removed the linear terms separately prior to filtering. The filters shown in Figure 2 have $H(\theta) \approx \frac{1}{2}$ at $\pi/6$ which corresponds to the annual cycle in monthly data. Thus they cannot be used directly to extract a long-term trend from a series containing an annual cycle. Thompson et al. (1986) used these filters in conjunction with band-reject filters that removed signals with frequencies around $\pi/6$ and $\pi/3$ (i.e. 12 month and 6 month periods). The details are given in Section 4 below where it is noted that the effective filters had lengths ranging from 28 to 41 months.

It is of interest to see whether the process of extracting a long-term trend might be better accomplished by direct use of appropriate low-pass filters. Figures 3 and 4 show the transfer functions of two sets of

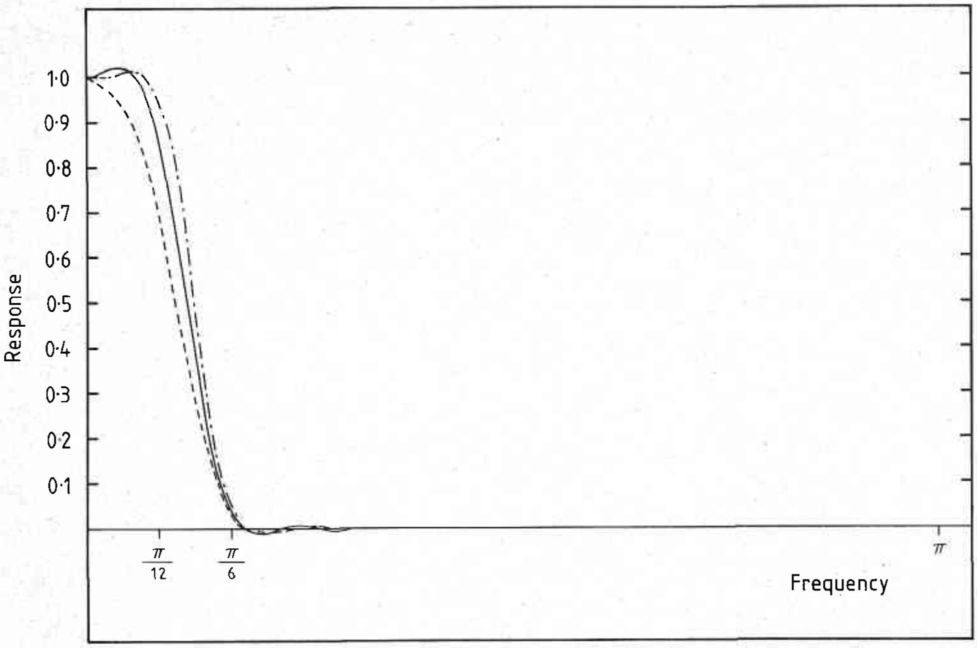


Figure 3: Transfer functions of low-pass filters defined by equation 24, using $\omega_c=\pi/6-2\pi/(2K+1)$ for K=16 (dashed), K=20 (solid) and K=24 (chain).

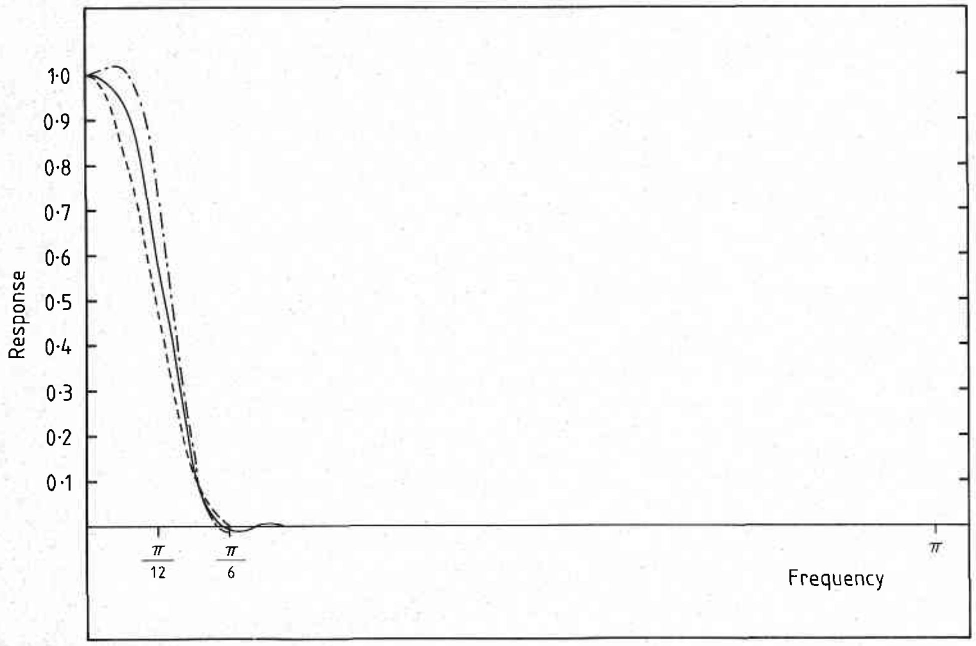


Figure 4: Transfer functions of low-pass filters defined by equation 24 using $\omega_c=\pi/7-2\pi/(2K+1)$ for K=16 (dashed), K=20 (solid) and K=24 (chain).

filters defined by equations (24a,b). The expression (25) was used to define the transition band width as a function of K denoted $\delta(K)$. Figure 3 corresponds to filters with cutoff

$$\omega_c = \frac{\pi}{6} - \delta(K)/2 \qquad (26)$$

for the various values of K. Although this has $\pi/6$ at the end of the nominal transition band, it will be seen that $H(\pi/6)$ is of order 0.05 which may be undesirably large when analysing signals with a strong annual cycle. In order to ensure that the annual cycle was more definitely in the stop band, the set of filters shown in Figure 4 used

$$\omega_c = \frac{\pi}{7} - \delta(K)/2 \qquad (27)$$

for various K values. It will be seen that $H(\pi/6)$ is significantly reduced.

3c. More general low-pass filters

The low-pass filters in the previous section have particularly good cut-off properties in the stop-band but this is achieved at the expense of loss of control of the transition band width for any given filter length. In some cases this may be undesirable and it may be appropriate to sacrifice some of the relative smoothness of the filters defined by (24a,b) in order to obtain a shorter filter for a given transition band width. The class of filters that we consider here are approximations to the ideal filter with

$$\begin{aligned} H(\theta) &= 1, & \theta &\leq \omega_p & (28a)\\ &= 0, & \theta &\geq \omega_s & (28b)\\ &= (\omega_s-\theta)/(\omega_s-\omega_p), & \omega_p &\leq \theta \leq \omega_s . & (28c) \end{aligned}$$

Fourier analysis shows that the corresponding filter coefficients are

$$c_o = (\omega_p + \omega_s)/2\pi \qquad (29a)$$

$$c_k = c_{-k} = [\cos(k\omega_p) - \cos(k\omega_s)]/(\pi k^2(\omega_s-\omega_p)) \;, \; k \neq 0 \quad . \qquad (29b)$$

Figure 5 shows the transfer functions for filters obtained from various truncations of (29b) using $\omega_p=\pi/12$ and $\omega_s=\pi/6$ and without normalising to give $H(0)=1$. Again it will be seen that even at the intended beginning of the stop band, $H(\pi/6)$ has a value that may be significant when analysing data containing a strong annual cycle. Figure 6 shows the results of studies analogous to those of Figure 4 in which the transition band is manipulated so that $H(\pi/6)$ is very small. For the present class of filters this is a rather more ad hoc procedure than that described above. Because of the Gibbs oscillations in the transfer function, much of the 'stop band' has significant departures from zero. What is required for analysing data with a strong annual cycle is to have a transfer function with a zero near $\pi/6$. The example shown in Figure 6 corresponds to $\omega_p = \pi/12$, $\omega_s = \pi/7$, with K=13.

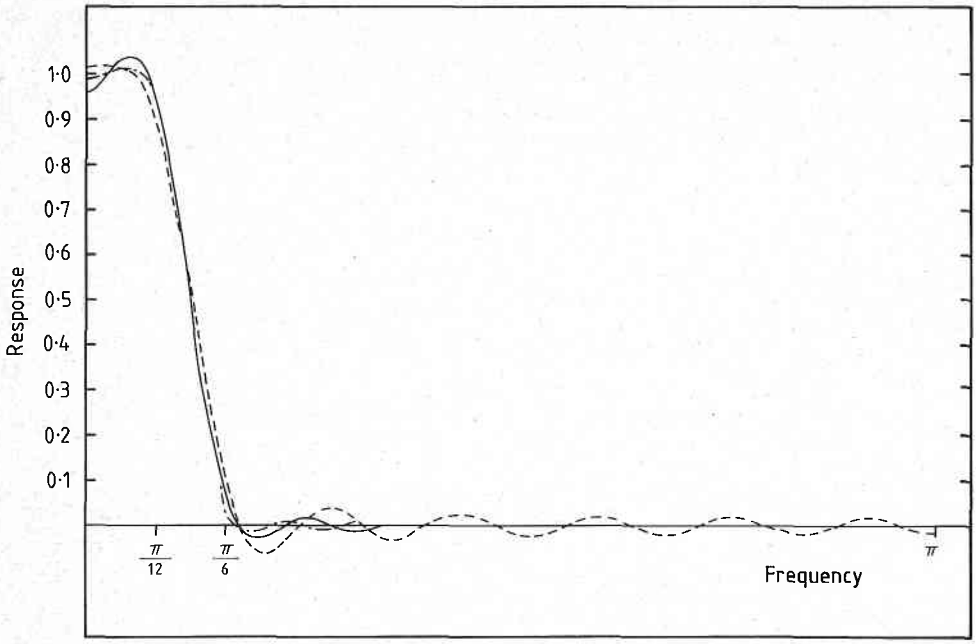


Figure 5: Low-pass filters fitted to transition band from $\pi/12$ to $\pi/6$ as per equation (29) using K=12 (dashed), K=16 (solid) and K=20 (chain).

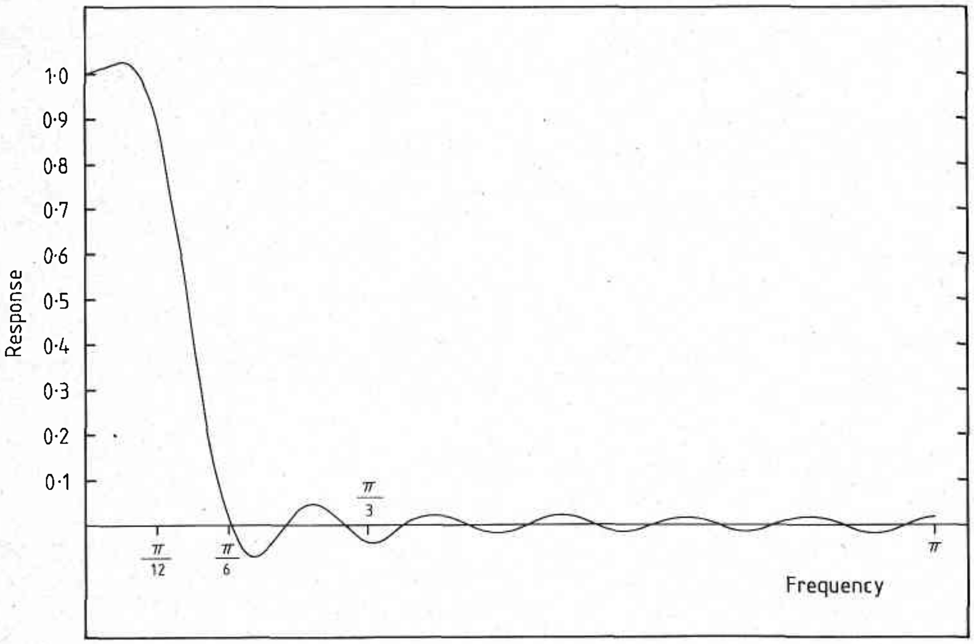


Figure 6: Low-pass filter designed to have a zero near $\pi/6$.

4. Complex Demodulation

4a. Overview

The procedure of complex demodulation has been described by Bloomfield (1976). The aim is to characterise a quasi-sinusoidal oscillation whose amplitude and phase are slowly varying in time. The various steps involved in analysing a series z_k which is believed to have a component with frequency near ω are:

(i) Multiply by $\exp(-i\omega t)$:

i.e. $$x_k = z_k \exp(-i\omega k) \quad . \qquad (30)$$

(ii) Low-pass filter :

$$u_k = \sum_{j=-K}^{K} c_j x_{k-j} = \sum_j c_j z_{k-j} \exp(i\omega(j-k)) \quad . \qquad (31)$$

Complex demodulation can be used in constructing an estimated signal y(t) but the conventional approach is to take the complex series u(t) and interpret this as an amplitude A(t) and phase $\phi(t)$ via

$$A(t) = 2|u(t)| \qquad (32)$$

$$\phi(t) = \text{Arg}\ (u(t)) \qquad (33)$$

i.e. $$u(t) = \tfrac{1}{2} A(t) \exp(i\phi(t)) \quad . \qquad (34)$$

The series A(t) and $\phi(t)$ are regarded as the slowly varying amplitude and phase of a function

$$A(t)\cos(\omega t + \phi(t))$$

$$= \tfrac{1}{2} A(t)[\exp(i\omega t + i\phi(t)) + \exp(-i\omega t - i\phi(t))]$$

$$= u(t) \exp(i\omega t) + u(t)^* \exp(-i\omega t)$$

$$= y(t) \quad . \qquad (35)$$

If an estimate of the actual signal y(t) is required, this is most directly obtained by the following two steps:

(iii) Multiply by $\exp(i\omega t)$:

$$v_k = \exp(i\omega k) u_k = \sum_j c_j z_{k-j} \exp(i\omega j) \quad . \qquad (36)$$

(iv) Multiply real part by 2 :

$$y_k = v_k + v_k^* = \sum_j 2c_j \cos(\omega j)\, z_{k-j} \quad . \qquad (37)$$

Thus in filtering terms, the procedure is equivalent to applying a filter whose coefficients are derived from those of the original low-pass filter by multiplying the jth coefficient by $2\cos(\omega j)$.

The effect of this procedure on a single frequency, λ, is given by

$$z(t) = A\exp(i\lambda t) + A^* \exp(-i\lambda t) \tag{38}$$

$$x(t) = A\exp(i(\lambda-\omega)t) + A^* \exp(i(-\omega-\lambda)t) \tag{39}$$

$$u(t) = H(\lambda-\omega)A\exp(i(\lambda-\omega)t) + H(-\omega-\lambda)\ A^* \exp(i(-\omega-\lambda)t) \tag{40}$$

where ω is the demodulation frequency and $H(\cdot)$ is the transfer function of the low-pass filter with coefficients c_j.
Therefore

$$v(t) = AH(\lambda-\omega)\exp(i\lambda t) + A^* H(-\omega-\lambda)\exp(-i\lambda t) \tag{41}$$

$$y(t) = [H(\lambda-\omega) + H(-\omega-\lambda)][A\exp(i\lambda t) + A^* \exp(-i\lambda t)]$$

$$= z(t)[H(\lambda-\omega) + H(-\omega-\lambda)] \quad . \tag{42}$$

Thus if the filter described by H(.) is a symmetric low-pass filter with cutoff ω_c, then for $\omega_d > \omega_c$, complex demodulation at frequency ω_d corresponds to a band pass filter with pass bands

i.e. $$-\omega_d - \omega_c \leq \omega \leq \omega_c - \omega_d \tag{43}$$

or $$\omega_d - \omega_c \leq \omega \leq \omega_c + \omega_d \quad . \tag{44}$$

When analysing the seasonal cycle of CO_2, Thompson et al. (1986) used both the A(t), ϕ(t) representation and the reconstructed y(t) form in different parts of their analysis.

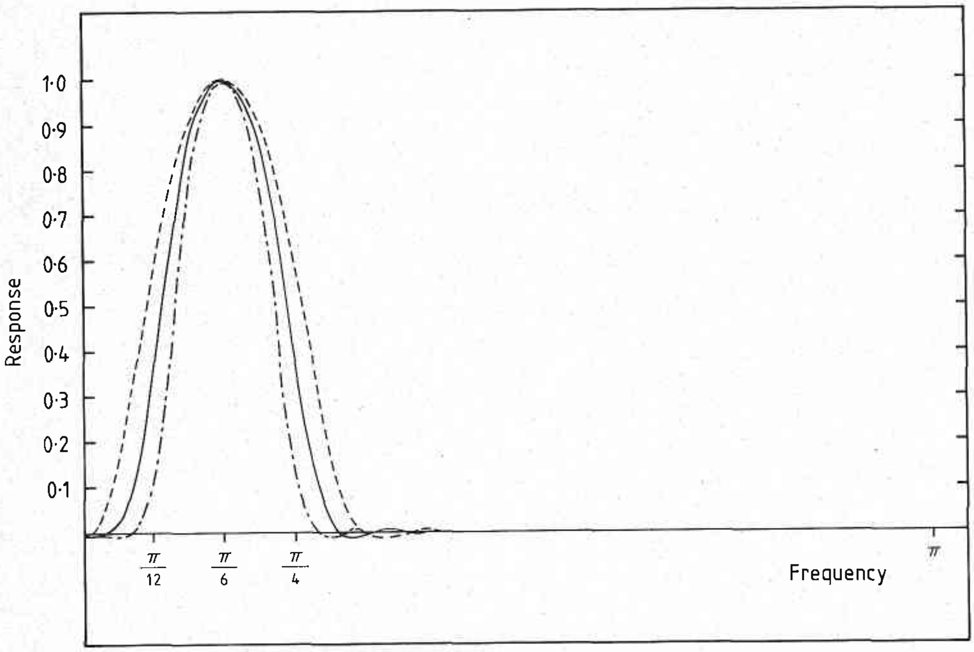


Figure 7: Some of the band-pass filters effectively used by Thompson et al. (1986) in extracting the annual cycle for CO_2 data. The nominal pass band is $\pi/6-\omega_c \leq \omega \leq \pi/6 + \omega_c$ with $\omega_c = 10\pi/N$. The filter lengths are K=18 (dashed), K=22 (solid) and K=28 (chain).

4b. Details of past analysis

The complex demodulation analyses undertaken by Thompson et al. (1986) were performed using computer routines based on those given by Bloomfield (1976). These routines involve 3 parameters: ω_d, the demodulation frequency, ω_c, the cutoff frequency of the low-pass filter and K, the length of the low-pass filter. This is an appropriate degree of generality. However the routines given by Bloomfield restrict K to be a factor of the length of the series, N, i.e.

$$K = N/m \qquad \text{for some integer } m \qquad (45)$$

and also require

$$\omega_c = (m + 2n)\pi/N \qquad \text{for some integer } n \quad . \qquad (46)$$

Table 1 lists the various sites from which CO_2 data were analysed by Thompson et al. (1986) and gives the characteristics of the filters involved in the complex demodulation. It also includes the characteristics of the low-pass filters used in the extraction of interannual variations from the decycled series. The series were demodulated at $\omega_d = \pi/6$ and $\omega_d = \pi/3$. Figure 7 shows some of the band pass filters that were equivalent to the use of these demodulations at $\pi/6$. As noted in Section 2g above, the signals obtained from demodulation at $\pi/6$ and $\pi/3$ were subtracted from the original data. This is equivalent to constructing a band-reject filter with transfer function $1-H_{12}(\theta)-H_6(\theta)$. Some of the transfer functions of these band-reject filters are shown in Figure 8.

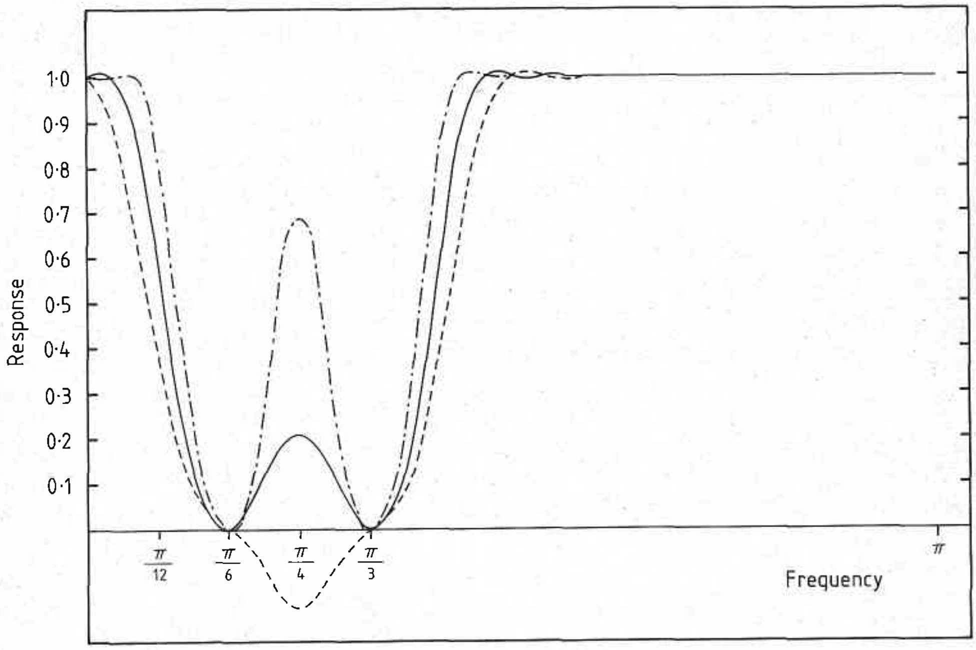
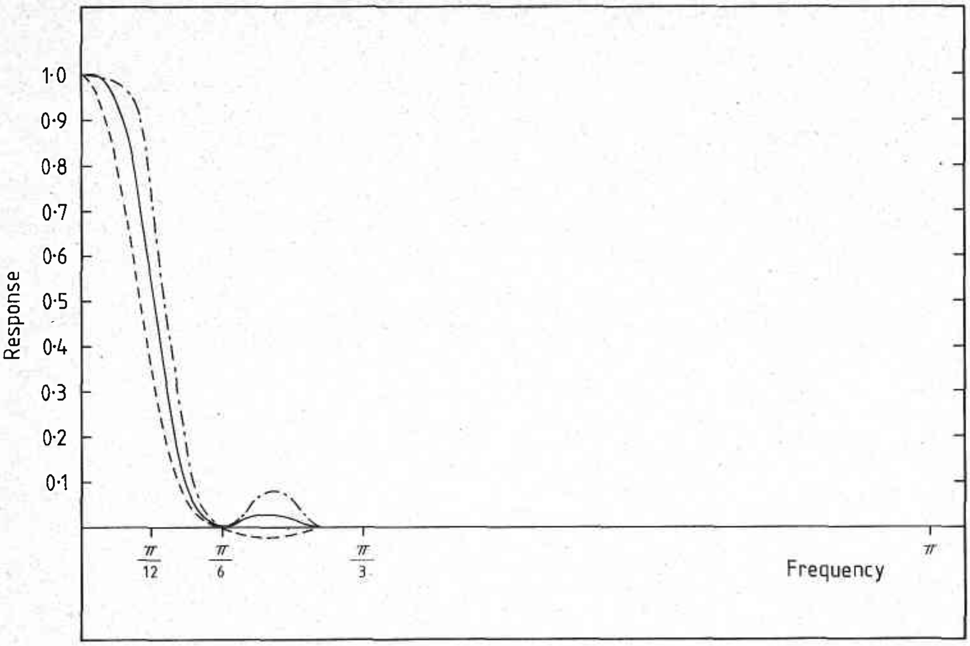
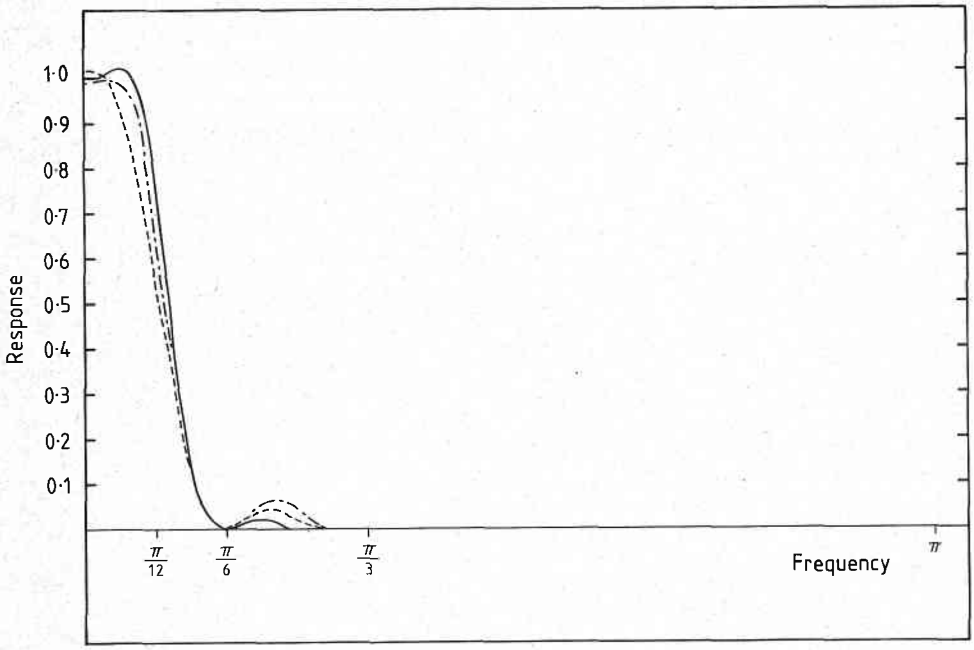


Figure 8: Some of the band-reject filters effectively used by Thompson et al. (1986) in decycling CO_2 data. The reject bands are $\pi/6-\omega_c \leq \omega \leq \pi/6 + \omega_c$ and $\pi/3-\omega_c \leq \omega \leq \pi/3+\omega_c$ with $\omega_c = 10\pi/N$. The filter lengths are K=18 (dashed), K=22 (solid) and K=28 (chain).



(a)



(b)

Figure 9: The effective filters used by Thompson et al. (1986) in extracting long-term trends from the CO_2 data. These filters represent successive application of band-reject filters of the type shown in Figure 8 and low-pass filters as shown in Figure 2. The filter lengths are shown in the form (K_{demod} + $K_{low-pass}$). (a) Effective filters used for Barrow (28+10) chain, Samoa (22+10) solid, and Cape Kumakahi (18+10), dashed. (b) Effective filters used for Mauna Loa (27+14) solid, South Pole (19+10) dashed, and Bass Strait (26+10) chain.

The interannual variations were obtained by applying one of the low-pass filters whose transfer functions $H_L(\theta)$ are shown in Figure 2. The resulting transfer function is $H_L(\theta)[1-H_{12}(\theta)-H_6(\theta)]$. The effective filter is given by the convolution of the low-pass filter with the band reject filter and so the length of the effective filter is the sum of the length of the final low-pass filter (i.e. 10 or 14) and the length of the demodulating filter (from 18 to 27).

The final transfer functions of the filters equivalent to the trend extraction process of Thompson et al. (1986) are shown in Figures 9a,b. Comparison with Figure 4 suggests that direct application of an appropriate low-pass filter will give as good, if not better, separation of the seasonal cycle and the long-term trend.

Site	Series length, N	Demodulating filters		low-pass filters		Effective lengths
		ω_c	K_{demod}	ω_c	K_L	
Kumakahi	108	$10\pi/108$	18	$\pi/6$	10	28
Samoa	132	$10\pi/132$	22	$\pi/6$	10	32
Bass Strait	156	$10\pi/156$	26	$\pi/6$	10	36
Barrow	168	$10\pi/168$	28	$\pi/6$	10	38
South Pole	228	$16\pi/228$	19	$\pi/6$	10	29
Mauna Loa	324	$20\pi/324$	27	$\pi/4$	14	41

Table 1: Details of filters used by Thompson et al. (1976) in decycling series and extracting long-term variations. The series lengths are the number of months. The demodulating filters are the low-pass filters used for demodulation at both $\pi/6$ and $\pi/3$ (i.e. 12 month and 6 month cycles). The low-pass filters were applied to the decycled series. The effective filter length is $K_{demod} + K_L$.

5. Data reduction

5a. Monthly means

In most studies of atmospheric constituents, it is necessary to obtain summaries of the data at regular intervals for the purposes of comparison with other records and to exhibit the major trends in the data. The general requirement is to reduce the size of the raw data set and possibly to produce a uniformly spaced data set from non-uniform raw data. One very common way in which this is done is simply to take monthly means of the data.

In order to characterise the effects of taking monthly means, the case of uniformly spaced data with n points per month is considered. The process of taking monthly means of such data is equivalent to taking n-point running means to smooth the data and then sampling this smoothed data at intervals of one month. The transfer functions for various n-point running means have been given by Hamming (1977, p30, Fig.3.2-2). However Hamming's discussion is in terms of units such that $\Delta t=1$ in each case. For comparison of running means each averaging over 1 month, the appropriate Δt is 1/n months when using n-point means. Hamming's expressions for the transfer functions for n points (n odd) convert to

$$H_n(\theta) = \frac{\sin(\theta/2)}{n \sin(\theta/2n)} \quad . \tag{47}$$

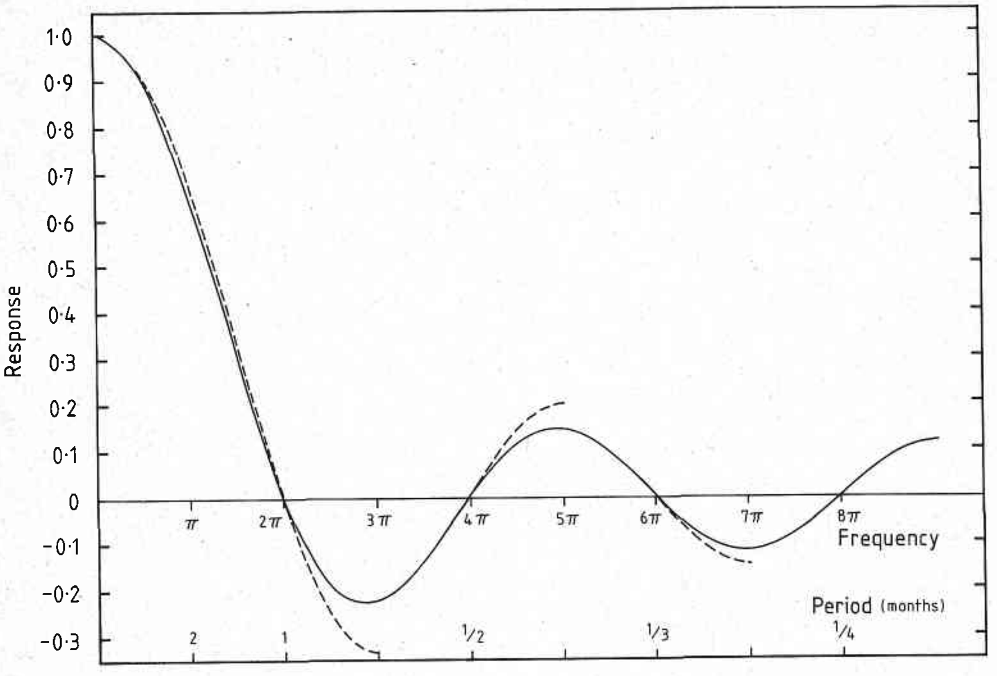


Figure 10: Transfer functions for n-point running means applied to data with n points per month, with angular frequencies in units of months^{-1}. The solid curve is for n=9. The curves for n=3, 5 and 7 follow this closely except near their respective cutoff points of 3π, 5π and 7π. The departures from the n=9 curve are shown as dashed.

Each transfer function applies over the interval [-nπ, nπ]. The n=9 case is shown as the solid curve in Figure 10. Near the origin all the curves behave similarly and so Figure 10 only shows (as dashed curves) the n=3, 5 and 7 cases near their cutoff points. Sampling any record at 1 month intervals introduces an aliasing effect in that any components remaining in the record with angular frequencies greater than π (ie periods less than 2 months) are aliased onto lower frequencies). A noticeable characteristic of the transfer functions shown in Figure 10 is their relatively rapid drop-off near the origin so that, for example, cycles with period 4 months are reduced by 10% by the averaging process.

5b. Smoothing splines

Enting (1987) describes how the work of Cox (1983) and Silverman (1984, 1985) can be used to determine the appropriate ways of using smoothing splines in the analysis of baseline data. Spline functions can be used in data reduction by the same process of smoothing followed by sampling that is described in Section 5a above. Silverman has shown that Cox's asymptotic results for the transfer function still hold (asymptotically) when the data are no longer uniformly spaced but are described by a mean density of data points. Since the spline fit to discrete data defines a continuous function $\hat{y}(t)$, spline fitting provides a way of producing a uniformly spaced record from non-uniform raw data. The smooth decay of the splines transfer function (equation 12) shows that by incorporating smoothing, spline fitting can avoid problems of local polynomial interpolation. Hamming (1977, p48 Fig 3.7-1) shows how such interpolations can introduce severe distortions of high frequency noise terms. As well as giving an interpolating function that tends to suppress high frequency noise, smoothing splines also have desirable properties at low frequencies. The transfer function given by equation (12) has its first three derivatives at the origin equal to zero, indicating that its initial decay is much slower than the transfer functions for the running means described above.

As noted by Enting (1987) (see also Section 2f above) the signal processing approach suggests that spline fitting should be used by choosing λ on the basis of the spectral properties of the data, interpreted as signal plus noise. Enting (1987) notes that several of the most readily available computer routines construct smoothing splines by minimising the integrated squared second derivative in equation (11) subject to a fixed sum of squares. Thus λ is not used explicitly. However routine *SMOOTH* of de Boor (1978) can be modified to solve the minimisation of expression (11). Enting also notes that the normalisations used in representing expression (11) differ between authors - indeed different normalisations are used in the 1984 and 1985 papers by Silverman. As in Enting (1987), expression (11) above uses the normalisation chosen by de Boor (1978) and so relates directly to his computer routines.

6. Summary

The discussion in Section 2 describes a 'signal processing' approach to the analysis of baseline atmospheric constituent data that provides a unified framework withih which a number of earlier analysis techniques can be compared. It is suggested that in many cases a direct use of digital filtering is the most desirable approach. Obviously, in any data analysis, both the techniques used and the underlying model assumptions should be fully specified. Digital filters should be described either by specific sets of coefficients or by their transfer functions, or both. Smoothing splines can be very conveniently characterised in terms of the factor λ in equation (11). This quantity determines the asymptotic filter coefficients equivalent to the spline fit and therefore also gives the asymptotic transfer function. It is, however, important to specify the normalisation used in defining λ (Enting, 1987). The signal processing approach outlined in Section 2 indicates that any estimates of confidence levels for the signals that are extracted must be based on assumptions about the error spectrum.

The complex demodulation procedures used by Thompson et al. (1986) have been reviewed in some detail, partly because the technique does not seem to have been previously described in terms of band-pass filtering and partly because the work of Thompson et al. omitted details of the filters that were used. The discussion in Section 4 above suggests that the routines from Bloomfield (1976) as used by Thompson et al. (1986) are unduly restrictive. In particular there is no inherent reason why all of the time series could not have been analysed using the same filter rather than using a different effective filter for each different length of series.

Acknowledgements

The details of the previous complex demodulation analyses were provided by Dr. M.L. Thompson.

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. John Wiley and Sons (New York), 258pp.

Cleveland, W.S., A.E. Freeny and E. Graedel (1983) The seasonal component of atmospheric CO_2: Information from new approaches to the decomposition of seasonal time series. J.Geophys.Res., 88C, 10934-11046.

Cox, D.D. (1983) Asymptotics for M-Type smoothing splines. Ann. Statist. 11, 530-551.

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

Diggle, P. and M.F. Hutchinson. (1986) Spline smoothing with autocorrelated errors. Presented at 26th Summer Research Institute of Australian Mathematical Society. Canberra January 1986.

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

Enting, I.G. (1987) On the use of smoothing splines to filter CO_2 data. (in preparation).

Golub, G.H., M. Heath and G. Wahba. (1979) Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics, 21, 215-223.

Hamming, R.W. (1977) Digital Filters Prentice Hall (Englewood Cliffs, N.J), 226pp.

Priestly, M.B. (1981) Spectral Analysis and Time Series. Academic Press (London), 890pp.

Silverman, B.W. (1984) Spline smoothing: The equivalent variable kernel method. Ann.Statist. 12, 898-916.

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

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

Yaglom, A.M. (1962) An Introduction to the Theory of Stationary Random Functions (Trans. R.A. Silverman). Prentice Hall (Englewood Cliffs, N.J.), 235pp.