



Investigating bias in the application of curve fitting programs to atmospheric time series

P. A. Pickers and A. C. Manning

Centre for Ocean and Atmospheric Sciences, School of Environmental Sciences, University of East Anglia, Norwich, NR4 7TJ, UK

Correspondence to: P. A. Pickers (p.pickers@uea.ac.uk)

Received: 31 May 2014 – Published in Atmos. Meas. Tech. Discuss.: 16 July 2014

Revised: 6 November 2014 – Accepted: 6 November 2014 – Published: 23 March 2015

Abstract. The decomposition of an atmospheric time series into its constituent parts is an essential tool for identifying and isolating variations of interest from a data set, and is widely used to obtain information about sources, sinks and trends in climatically important gases. Such procedures involve fitting appropriate mathematical functions to the data. However, it has been demonstrated that the application of such curve fitting procedures can introduce bias, and thus influence the scientific interpretation of the data sets. We investigate the potential for bias associated with the application of three curve fitting programs, known as HPspline, CCGCRV and STL, using multi-year records of CO₂, CH₄ and O₃ data from three atmospheric monitoring field stations. These three curve fitting programs are widely used within the greenhouse gas measurement community to analyse atmospheric time series, but have not previously been compared extensively.

The programs were rigorously tested for their ability to accurately represent the salient features of atmospheric time series, their ability to cope with outliers and gaps in the data, and for sensitivity to the values used for the input parameters needed for each program. We find that the programs can produce significantly different curve fits, and these curve fits can be dependent on the input parameters selected. There are notable differences between the results produced by the three programs for many of the decomposed components of the time series, such as the representation of seasonal cycle characteristics and the long-term (multi-year) growth rate. The programs also vary significantly in their response to gaps and outliers in the time series. Overall, we found that none of the three programs were superior, and that each program had its strengths and weaknesses. Thus, we provide a list of recommendations on the appropriate use of these three curve fitting

programs for certain types of data sets, and for certain types of analyses and applications. In addition, we recommend that sensitivity tests are performed in any study using curve fitting programs, to ensure that results are not unduly influenced by the input smoothing parameters chosen.

Our findings also have implications for previous studies that have relied on a single curve fitting program to interpret atmospheric time series measurements. This is demonstrated by using two other curve fitting programs to replicate work in Piao et al. (2008) on zero-crossing analyses of atmospheric CO₂ seasonal cycles to investigate terrestrial biosphere changes. We highlight the importance of using more than one program, to ensure results are consistent, reproducible, and free from bias.

1 Introduction

High-precision, continuous measurements of greenhouse gases in the atmosphere were initiated over 50 years ago by Charles Keeling at Scripps Institution of Oceanography, USA (Scripps), who began measuring atmospheric carbon dioxide (CO₂) mole fraction at the Mauna Loa Observatory, Hawaii, in 1958 (Keeling, 1960). Such data sets of regular atmospheric observations made at discrete time intervals are known as atmospheric time series (Amritkar and Kumar, 1995), and typically consist of long- and short-term variations that reflect biogeochemical fluxes and atmospheric mixing processes (Thoning et al., 1989). For example, atmospheric CO₂ time series typically consist of a long-term increasing trend, which largely results from fossil fuel burning and land-use change emissions (Keeling et al., 2011),

seasonal variations that are driven mostly by terrestrial biosphere processes, synoptic-scale variations caused by changing weather systems and air masses, and irregular variations caused by volcanic eruptions, large-scale ocean–atmosphere interactions and climate fluctuations and forcings (e.g. El Niño Southern Oscillation) (Houghton, 2007).

The interpretation of atmospheric greenhouse gas mole fraction data plays a fundamental role in quantifying the sources and sinks of climatically important gas species, such as CO₂ and methane (CH₄), interpreting latitudinal concentration gradients, inferring regional fluxes, and also for assessing temporal variability, such as long-term trends and interannual growth rates (Keeling et al., 2011; Dlugokencky et al., 2011; Houghton, 2007). In order to investigate specific processes, it is often necessary to isolate and extract the variation of interest from the complete data set (Martin and Diaz, 1991). For example, an examination of trends in the amplitude of the CO₂ seasonal cycle (e.g. Chan and Wong, 1990) requires the seasonal component to be separated from any long-term trend and irregular variations, a technique known as time series decomposition (Pierce, 1979; Theodosiou, 2011).

The analysis of atmospheric time series is often a complex process because the data are usually highly autocorrelated and consist of periodic and irregular variations on both long and short timescales. Additionally, mechanical failure of the measurement instruments or down-time for other reasons can result in gaps in the time series, so that data are not always spaced at regular time intervals (Trivett et al., 1989). For these reasons, simple curve fitting procedures, such as moving averages and cubic splines, are generally inadequate for the analysis of atmospheric time series, which has led to the development of more sophisticated fitting procedures (Trivett et al., 1989).

It has been recognised previously that the application of a particular curve fitting program in the analysis or decomposition of an atmospheric time series may introduce biases that could significantly influence the results and conclusions of an investigation (Nakazawa et al., 1997; Tans et al., 1989). Thus, scientific conclusions drawn from time series analyses may depend not only on the atmospheric measurements themselves, but also on the curve fitting program used. Consequently, it has been recommended that more than one curve fitting program is employed in any given time series study, so that possible biases can be identified (Nakazawa et al., 1997).

Despite this recommendation, the vast majority of studies and published papers involving time series analysis of atmospheric greenhouse gas data appear to rely on a single curve fitting program. For example, Bacastow (1976) found a correlation between the Southern Oscillation Index (SOI) and the residual variation (i.e. any remaining variation left in a data set, once the long-term trend and seasonal variation have been removed) from both the South Pole and Mauna Loa CO₂ data sets; the correlation found in that study has formed the foundation of numerous subsequent studies (e.g.

Reichenau and Esser, 2003; Jones et al., 2001), and yet is contingent on the results from a single parametric curve fitting program. Keeling and Shertz (1992) inferred the long-term rate of decline in atmospheric O₂ mole fraction based on the application of the same curve fitting program used by Bacastow (1976). Piao et al. (2008) suggested that the Northern Hemisphere terrestrial biosphere may be sequestering less carbon than previously thought, due to an increase in carbon losses during autumn, resulting from the strong dependence of respiration to rising temperatures; this conclusion was drawn from detrended atmospheric CO₂ data derived using a single digital filtering program (Thoning et al., 1989). More recently, Minejima et al. (2012) investigated the origin of pollution events at a Japanese island site that were identified from detrended atmospheric O₂ and CO₂ data using the Thoning et al. (1989) program only.

Our intention in this paper is not to dispute the science underpinning any of the above or other studies, but rather to highlight the absence of any assessment of the suitability of the curve fitting program used in many applications. The uncertainty associated with relying on a single curve fitting program appears not to have been investigated or quantified in any of the studies cited above; hence, it is not known whether the results might have been biased by the curve fitting program employed.

The scientific import of relying on a single curve fitting program is that many studies present results showing very small trends that are barely discernible amongst the “noise” of the data. This may explain why some studies have come to contradictory conclusions; for example, Chan and Wong (1990) and Keeling et al. (1996) disagree regarding the direction of the trend in the atmospheric CO₂ seasonal cycle amplitude at Alert, Canada, and Enting (1987) and Thoning et al. (1989) reach opposite conclusions regarding correlations between the magnitude of CO₂ seasonal maxima in spring and the magnitude of seasonal minima the following autumn.

There are only a few studies that have investigated the uncertainty associated with curve fitting analyses, or compared two or more programs on the same time series. To our knowledge, the first two of such investigations (Tans et al., 1989; Trivett et al., 1989) were prompted by a meeting in March 1988 on the treatment and data processing techniques used for CO₂ time series, funded by NOAA (National Oceanic and Atmospheric Administration, USA) and the UN WMO (United Nations World Meteorological Organization) (Elliot, 1989). Both of these studies, however, provided only a preliminary assessment of the differences between some curve fitting procedures rather than an in-depth analysis. For example, although Tans et al. (1989) compared six different curve fitting procedures, they were only applied to 3-year CO₂ flask sample data sets from a single station. In Trivett et al. (1989), the discussion on the differences found between two curve fitting programs is very brief, simply stating that the seasonal cycle outputs are comparable between the

two curve fitting programs, and that the “forward step-wise multiple linear regression” curve fitting program used by the authors had limitations, such as sensitivity to outliers.

A more comprehensive analysis is presented by Nakazawa et al. (1997) who compared a digital filtering program, developed at Tohoku University, Japan, and a harmonic regression program. The authors emphasised the importance of using more than one curve fitting program in analyses of atmospheric time series, stating that an assessment of the global carbon cycle using one program could be quite different from that derived using the same data but a different program (Nakazawa et al., 1997).

Since the study by Nakazawa et al. (1997), there is no evidence in the published literature of subsequent work on curve fitting bias, either by Nakazawa and colleagues, or by other authors. Furthermore, the vast majority of studies and published papers involving time series analysis of atmospheric greenhouse gas data have continued to infer scientific conclusions from atmospheric greenhouse gas measurements based on analyses using a single curve fitting program.

Small trends in atmospheric greenhouse gas mole fraction time series can have significant consequences for the Earth system, and therefore may have substantial implications for climate change policy. Given the political and socioeconomic implications of climate change and public interest in high-profile climate science publications, it is essential to ensure that information and conclusions inferred from atmospheric time series are reproducible using a number of techniques, and are not exaggerated or attenuated by artefacts or biases of the curve fitting programs used.

The general lack of investigation into uncertainty and bias associated with using a single curve fitting program is surprising, considering their widespread use in atmospheric research over the past 30 years or so. In addition, it is unlikely that a single curve fitting program can adequately represent all atmospheric greenhouse gas mole fraction time series; in other words, a given curve fitting program may be better suited to examine data sets with particular characteristics, or for particular types of analyses. The objective of our paper, therefore, is to address, at least in part, the lack of consideration of curve fitting bias in analyses of atmospheric time series, by comparing the outputs from three widely used curve fitting programs, applied to atmospheric time series displaying diverse characteristics.

Aims and outline of paper

In this paper we investigate bias associated with the application of three curve fitting programs, known as HPspline, CCGCRV, and STL, that are widely used within the atmospheric greenhouse gas measurement community and have not been extensively compared previously. Specifically, we assess the performance of each program with respect to the complete curve fit, representation of the long-term trend and the growth rate in the long-term trend, representation of the

seasonal cycle, vulnerability to gaps and outliers in the data, and sensitivity to the input parameter settings of the programs.

We emphasise that the purpose of employing curve fitting programs to atmospheric time series is not to produce a fit that passes through the most number of data points as possible, but to extract the salient features of interest, such as seasonality, and separate these components from anomalous “noise” or other features within the time series. Thus, our objective is not to determine which of the three curve fitting programs examined is “best”, but rather to elucidate differences in the output from each program when given the same input data sets. In addition, we assess whether any of the three programs are better suited to time series exhibiting particular “characteristics” (e.g. data sets with a relatively large seasonal cycle), and to specific research applications (e.g. correlation analyses between CO₂ residual variations and large-scale climate indices).

The outline of the remainder of this paper is as follows: Sect. 2 describes each of the three curve fitting programs that are compared in this study, the contrasting data sets that the three programs have been applied to, including an explanation of why these data sets were chosen, and the experimental methods that have been employed. Section 3 presents the results and discusses our findings. Section 4 summarises the conclusions of this work, and provides specific recommendations on the appropriate use of the three curve fitting programs evaluated, as well as general recommendations for all investigations that use curve fitting programs to analyse atmospheric time series.

2 Methodology

2.1 Curve fitting programs

2.1.1 HPspline

“HPspline” is the name of a parametric curve fitting program written in Fortran, used by the Atmospheric Oxygen Research Group based at Scripps, and is an implementation of the previous “Stationfit” program developed in the 1970s by Robert Bacastow of the Carbon Dioxide Research Group, also at Scripps (Keeling et al., 1986; S. Piper, Scripps, personal communication, 2014). The current version of HPspline was developed by Martin Heimann (Max Planck Institute for Biogeochemistry, Germany) in the 1980s, and the program is now maintained by Ralph Keeling (Scripps, USA). The updated procedure includes three routines (svdcmp, svdfit and svdvar) from Numerical Recipes in Fortran (Press et al., 1996) and involves fitting data to a harmonic function, a polynomial equation, and a stiff cubic spline (Reinsch, 1967). The data are initially fitted linearly using the following expression (Keeling et al., 1989):

$$S(t) = \sum_{k=1}^h [x_k \sin(2\pi kt) + y_k \cos(2\pi kt)], \quad (1)$$

where $S(t)$ is the seasonal variation as represented by a harmonic function, h is the number of harmonics (typically four), t is time in years, $2\pi k$ is the angular frequency, and x_k and y_k are constants.

The long-term trend variation is represented by a polynomial equation, $E(t)$:

$$E(t) = a_0 + a_1 t + a_2 t^2 + \dots + a_{(n-1)} t^{(n-1)}, \quad (2)$$

where n is the number of polynomial terms (typically three), and $a_0, a_1, \dots, a_{(n-1)}$ are constants.

The function $E(t)$ is subtracted from the data to remove the long-term trend, and the interannual variations are then fitted to a Reinsch-type cubic spline function, $R(t)$ (Reinsch, 1967), to represent any irregular variations (Bacastow et al., 1985). Simultaneously, the data are fitted to the function $(1 + \gamma t)S(t)$, where γ is a time-dependent gain factor. A non-linear fit is achieved, using the initial estimates of the harmonic coefficients (x_k and y_k) from the first fit of $S(t)$, via an iterative procedure, whereby an estimate of $S(t)$, obtained from a fit of $E(t)$, is subtracted from the data, and then $R(t)$ is fit to the residuals. Next, $R(t)$ is subtracted from the data, and the residuals are re-fit to the function $(1 + \gamma t)S(t)$. $(1 + \gamma t)S(t)$ is then subtracted from the data and $R(t)$ is re-fit to the residuals, and so the procedure continues, until convergence is obtained, usually after approximately six cycles. The overall time series can thus be represented as follows (Keeling et al., 1989):

$$P(t) = E(t) + (1 + \gamma t)S(t) + R(t), \quad (3)$$

where $P(t)$ is equivalent to the sum of the trend and the seasonal variation. Further information about the mathematical concepts underlying HPspline can be found in Bacastow et al. (1985) and Keeling et al. (1986, 1989).

2.1.2 CCGCRV

“CCGCRV” is a digital filtering curve fitting program developed by Kirk Thoning (Carbon Cycle Group, Earth System Research Laboratory (CCG/ESRL), NOAA, USA) in the late 1980s. The version of CCGCRV used in this study was written in Interactive Data Language (IDL) by Paul Krummel (Marine and Atmospheric Research Group, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia). However, there is also a C language version, the code for which is freely available at: <ftp://ftp.cmdl.noaa.gov/pub/john/ccgcrv/>.

Approximations of the seasonal cycle and the long-term trend variation are made by fitting a polynomial equation

combined with a harmonic function:

$$C(t) = a_0 + a_1 t + a_2 t^2 + \dots + a_{(n-1)} t^{(n-1)} + \sum_{k=1}^h m_k [\sin(2\pi kt + \varphi_k)], \quad (4)$$

where t is time in years, n is the number of polynomial terms (typically three), $a_0, a_1, \dots, a_{(n-1)}$ are constants, h represents the number of harmonics in the function (typically four), and m_k and φ_k define the magnitude and phase of each sinusoidal component respectively. The fit to a data set is achieved with a linear least squares regression, applying the “LFIT” routine from Numerical Recipes in Fortran (Press et al., 1996).

The next step is to calculate the residuals of the input data to $C(t)$ and filter them using a Fast Fourier Transform (FFT) algorithm, so that short-term and interannual (long-term) variations can be retained in the fitted curve (Thoning et al., 1989). This is achieved by transforming the data from the time domain into the frequency domain using the FFT, multiplying by a low-pass digital filter to remove variations of a specified frequency (see below), and then transforming the filtered data back to the time domain using an inverse FFT (Thoning et al., 1989). The low-pass filter function used is a decreasing exponential represented as follows:

$$H(f) = \exp\left[-\ln(2) \cdot \left(\frac{f}{f_c}\right)^6\right], \quad (5)$$

where f_c is the cut-off frequency of the low-pass filter, expressed in cycles yr^{-1} .

The low-pass filter is applied to the residuals twice, once with a short-term cut-off value ($f_c = f_s$) for smoothing the data, and once with a long-term cut-off ($f_c = f_l$) to remove any remaining seasonal oscillation and to represent interannual variations in the data that are not represented by the polynomial part of $C(t)$. Unless otherwise stated, we use values of 4.56 cycles yr^{-1} (i.e. a period of 80 days) and 0.55 cycles yr^{-1} (i.e. a period of 667 days) for f_s and f_l respectively, as these are the current typically used values (<http://www.esrl.noaa.gov/gmd/ccgg/mbl/crvfit/crvfit.html>). Since the FFT algorithm requires the input data to be equally spaced and without gaps, the CCGCRV program linearly interpolates between the data points at a user-specified interval (Thoning et al., 1989). Additionally, the FFT algorithm requires that the number of data points used is equal to an integral power of two; hence, the program “zero pads” the data points to obtain the necessary number by extending the data set by approximately half a year at each end (Thoning et al., 1989). The residuals are then adjusted by the program so that the end values are approximately zero. This diminishes any effect the “zero padding” may have on the ends of the filter, which especially affects the growth rate at the end points of the data set (Thoning et al., 1989).

Lastly, the features of interest (for example, seasonal cycle amplitude and long-term trend) are derived by combining the

appropriate parts of the fitting procedure: the long-term trend is obtained by combining the polynomial part only of $C(t)$ with the f_l filter (i.e. long-term trend = $C(t)_{\text{polynomial only}} + H(f_l)$), and the seasonal cycle is obtained by combining $C(t)$ with the f_s filter, and then subtracting the long-term trend (i.e. seasonal cycle = $C(t) + H(f_s) - \text{long-term trend}$). The CCGCRV fitting procedure is described in more depth in Thoning et al. (1989), and on the NOAA/ESRL website at: <http://www.esrl.noaa.gov/gmd/ccgg/mbl/crvfit/crvfit.html>.

2.1.3 STL

“STL” is an abbreviation for Seasonal Trend decomposition using LOESS (locally weighted scatterplot smoothing) and was developed by William Cleveland (AT&T Bell Laboratories, USA) in the 1980s. The version of STL used in this study was written in “R”, developed from Cleveland’s Fortran code by Brian Ripley (University of Oxford, UK) (described at <http://stat.ethz.ch/R-manual/R-devel/library/stats/html/stl.html>), and was provided to us by Sara Mikaloff-Fletcher (National Institute of Water and Atmospheric Research (NIWA), New Zealand). Unlike HP-spline and CCGCRV, STL does not employ harmonic functions, but rather is a moving average technique. A sequence of LOESS smoothers of different moving window frequencies are applied iteratively to extract the variations of interest (Carslaw, 2005). The implementation consists of two recursive loops: the inner loop applies a seasonal smoother to the annual cycle-subseries (defined as series containing values for each month, e.g. the first subseries contains only January values, the second subseries contains only February values, and so on), followed by a trend smoother, while the outer loop computes the fitted values, which are weighted according to a “nearest neighbour routine”, with extreme values down-weighted during the next iteration of the inner loop (Cleveland et al., 1990). In this manner, the procedure progressively refines approximations of the trend and seasonal components until convergence is achieved, which typically occurs after less than 10 iterations of the outer loop (Carslaw, 2005).

LOESS assigns a neighbourhood weight, $v_i(x)$, to each data point using the tricube weight function, W (Cleveland et al., 1990), according to the following:

$$v_i(x) = W\left(\frac{|x_i - x|}{\lambda_q(x)}\right), \quad (6)$$

where x_i is the measurement of the independent variable, x is the computed value of the fit and $\lambda_q(x)$ is the distance of the q th farthest x_i from x , where q is a positive integer that controls the smoothness of the LOESS regression curve. Next, a polynomial of degree d is fit to the data with weight $v_i(x)$ at (x_i, y_i) , where y_i is the measurement of the dependent variable. The value of the locally fitted polynomial at x is $\hat{g}(x)$.

The inner loop of the STL procedure consists of six steps as follows (Cleveland et al., 1990):

- Step 1: a detrended series is computed by subtracting the long-term trend variation from the entire time series. For the initial pass through the inner loop, a value of zero is used for the trend. This step incorporates the trend into the cycle-subseries component until step 4, where it is removed in the detrending process.
- Step 2: the annual cycle-subseries are then smoothed by LOESS, using a first degree polynomial and q equal to the value of the seasonal smoothing window ($swin$ in years; set by the user). Smoothed values are computed for the range of values $v = -n_{(p)} + 1$ to $N + n_{(p)}$, where $n_{(p)}$ is the frequency of the input data series (i.e. 12, for monthly time series) and N is the total number of data points in the time series. Thus, the number of smoothed values is $2n_{(p)}$ greater than the annual cycle-subseries prior to the LOESS smoothing ($v = 1$ to N). This extension of the smoothed values by $n_{(p)}$ data at each end is to account for the subsequent loss of $n_{(p)}$ data in Step 3.
- Step 3: a low-pass moving average filter is applied twice to the smoothed cycle-subseries, where the length of the moving average is $n_{(p)}$. This is followed by the application of a third low-pass moving average filter with length = 3 and then LOESS smoothing. These three moving averages result in the loss of $n_{(p)}$ data at each end of the time series, which is accounted for by the extension of the seasonal smoothing in step 2 by $2n_{(p)}$ data points.
- Step 4: the smoothed cycle-subseries is detrended to prevent low-frequency variation from being included in the seasonal component of the decomposition.
- Step 5: a deseasonalised series is computed by subtracting the computed seasonal component from the entire time series.
- Step 6: this deseasonalised series is smoothed using LOESS with q equal to the value of the trend smoothing window ($twin$ in months; set by the user). This produces a trend component which is used in step 1 of subsequent iterations of the inner loop.

The outer loop of the STL procedure down-weights any outliers in the data by assigning robustness weights, ρ_v , to the series using a bisquare weight function, B (Cleveland et al., 1990):

$$\rho_v = B(|R_v|/h), \quad (7)$$

where R_v is the residual component (i.e. the data with the trend and seasonal components removed) and $h = 6 \times \text{median}(\|R_v\|)$. In subsequent iterations of the inner loop, the neighbourhood weight, $v_i(x)$, used in the LOESS smoothing of steps 2 and 6, is multiplied by the robustness weight, ρ_v , of the preceding pass of the outer loop. For more

Table 1. Atmospheric time series used in this study, including site location and altitude, gas species, time period and data source. Note that all data are monthly mean baseline-only data, i.e. representative of “clean” background air.

Station name	Location and altitude	Gas species	Percent of time series interpolated	Period	Data source
Alert Station (ALT) Canada	82.45° N, 62.52° W 210 m a.s.l.	CO ₂	< 2	1988–2009	b
		CH ₄	< 2	1988–2009	b
Baring Head (BHD) New Zealand	41.41° S, 174.87° E 85 m a.s.l.	CO ₂	10 ^a	1970–2011	c
		O ₃	26 ^a	1991–2010	d
Cape Grim Observatory (CGO) Australia	40.68° S, 144.68° E 94 m a.s.l.	CH ₄	< 2	1984–2011	e

^a Percentage interpolation required for BHD data is relatively high, not because the data do not exist, but because only baseline data (i.e. data that are consistent with the concept of a well-mixed atmosphere) are reported to the WDCGG database. ^b WDCGG (Doug Worthy, Environment Canada, Canada). ^c Britton Stephens, National Center for Atmospheric Research, USA; Gordon Brailsford and Antony Gomez, National Institute of Water and Atmospheric Research Ltd., New Zealand).

^d WDCGG (Sylvia Nichol, National Institute of Water and Atmospheric Research Ltd., New Zealand). ^e Paul Krummel, CSIRO Marine and Atmospheric Research, Australia, and the Australian Bureau of Meteorology (Cape Grim Baseline Air Pollution Station), website: <http://www.csiro.au/greenhouse-gases/>.

details regarding the STL program we refer the reader to Cleveland et al. (1990).

When choosing values to use for the *swin* and *twin* parameters, Carslaw (2005) points out that the seasonal and trend components should not compete for the same variation in the data, and low-frequency fluctuations should be retained in the trend component and not in the remainder. In this study, unless otherwise stated, we used a *swin* value of 5 years and a *twin* value of 25 months, which are the values typically used by our international colleagues (S. Mikaloff-Fletcher, personal communication, NIWA, 2011).

A major limitation of the currently available STL R programming code is that it can only be applied to equally spaced data with no gaps (Manning et al., 1990). Note that this is not a limitation of the STL procedure itself, but rather that missing values are not allowed in the current version of the R code to limit computational speed requirements. Such data sets can be derived by pre-treating the original data set using an interpolation or smoothing technique. However, such interpolation techniques may create biases or artefacts in the time series, particularly if there are large gaps. In order to mitigate this limitation, all comparison tests were carried out on time series consisting of monthly mean data that were already equally spaced in time, and had few or no missing values.

2.2 Time series

The three curve fitting programs described above were applied to semi-continuous atmospheric measurement data, provided by international colleagues and available to download from the World Data Centre for Greenhouse Gases (WDCGG) website at: <http://ds.data.jma.go.jp/gmd/wdcgg/> (see Table 1). In order to investigate the ability of each curve fitting program to represent particular features of at-

mospheric time series, we selected a suite of input data sets that provide a diverse variety of seasonal and trend characteristics, and that are notable for their long duration (several decades). For example, we chose the atmospheric CO₂ data set from Alert Station, Canada (ALT), because it has a relatively large seasonal cycle amplitude and a very regular but asymmetric seasonal pattern with prolonged and rounded maxima and contrasting sharp minima, which are characteristic of high northern latitude CO₂ data sets. Previous studies (e.g. Nakazawa et al., 1997; Tans et al., 1989; Trivett et al., 1989) have found that curve fitting programs often struggle to represent the deep troughs of the seasonal minima of such time series.

Other atmospheric time series we have chosen to examine exhibit more complex seasonal patterns. For example, the ALT CH₄ data set, shown in Fig. 1a, is characterised by a double seasonal maximum during winter. In contrast, the Baring Head, New Zealand (BHD) CO₂ seasonal cycle exhibits a variable pattern, such that it can be difficult to determine seasonal cycle characteristics. This is partly because, in contrast to the ALT CO₂ data set, the BHD CO₂ seasonal cycle has a much smaller amplitude, owing to a considerably smaller terrestrial biosphere in the Southern Hemisphere. We also examine the BHD ozone (O₃) data set, which exhibits a seasonal cycle with relatively high interannual variability, in that the magnitude of the maxima and minima fluctuate significantly from year to year, although the shape of the seasonal cycle is relatively consistent.

In addition to complex seasonal patterns, some of the time series were chosen because they exhibit different trend characteristics. For example, there is little variability in the growth rate of the ALT and BHD CO₂ long-term trends, whereas the ALT and Cape Grim Observatory, Australia (CGO) CH₄ long-term trend growth rates vary considerably.

Table 2. Range of input parameter settings that were used to test program sensitivity. Values in bold text indicate typical settings used by colleagues and also throughout this study, unless specifically stated otherwise. The right-hand column shows the input smoothing parameter values that were found to produce the “best fit” to the decomposed components of an artificial data set (discussed in Sect. 3.5).

Program name	Parameter	Range of values tested	Artificial data “best fit” values
HPspline	Spline stiffness parameter, $SD2$, in ppm yr^{-2}	1, 30 , 99999,	500
	Number of harmonic terms	2, 4 ,	4
	Degree of polynomial function	1, 3 , 5	3
CCGCRV	Short-term cut-off period (f_s), in days	5, 80 , 200	250
	Long-term cut-off period (f_l), in days	100, 667 , 1200	1500
	Number of harmonic terms	2, 4 , 6	4
	Degree of polynomial function	1, 3 , 5	3
STL	Seasonal smoothing window ($swin$), in years	1, 5 , 50	8
	Trend smoothing window ($twin$), in months	5, 25 , 500	45

As mentioned above, the currently available version of STL requires equally spaced data, so we only used monthly mean time series and we interpolated the data to fill any gaps (see Table 1). Interpolation was carried out by applying HPspline to the original time series, and using values of the curve fit as surrogate data where there were missing values. Since our objective is to investigate bias associated with curve fitting programs, and not to infer scientific conclusions, interpolating missing data has not had any influence on our results and conclusions. In using monthly mean data, however, we were not able to assess the programs with respect to fitting higher-frequency variations such as diurnal cycling or synoptic-scale variability.

2.3 Experimental methods

The three curve fitting programs were tested for their ability to represent each time series as a whole, and for under-estimation and over-estimation of the curve fits at the seasonal maxima and minima. We also assessed the proportion of data “captured” by the curve fitting programs, defined as when the curve fit passes within the $\pm 1\sigma$ standard deviation uncertainties of the data (SD). The decomposed components of the time series (i.e. seasonal, long-term trend, growth rate of the long-term trend and residual components) were also compared, and the programs were assessed for their ability to cope with outliers and gaps in the data, which were introduced artificially. Each of the programs was also tested for sensitivity to the input parameters of the fitting programs, such as the number of harmonics used in the fit. Table 2 lists the range of input parameter setting values that we have used, but to ensure the robustness of our conclusions on the behaviour of the three curve fit programs, we also tested many intermediate values of these input parameter settings. Analyses of the seasonal cycle amplitude of the detrended time series were performed using the ALT CO₂ and BHD CO₂ time series only, allowing for both Northern and Southern Hemisphere representation of the terrestrial biosphere sea-

sonal cycle, whereas all the other curve fitting program tests were performed on all five time series.

The use of statistical analyses in the few existing atmospheric curve fitting comparison studies has been limited, with previous authors relying heavily on visual interpretation of graphical representations of the curve fit outputs in order to describe the differences between programs (e.g. Nakazawa et al., 1997; Tans et al., 1989; Trivett et al., 1989). We have attempted to employ some statistical techniques in our analysis. However, time series are complex and highly autocorrelated, which makes the use of many statistical tests redundant or inappropriate. For example, t -tests can be used to determine whether the means of two populations are significantly different, but this is somewhat meaningless when applied to an atmospheric time series, which consists of three components: a long-term trend, seasonal cycle, and residual variations. Therefore, we have cautiously employed t -tests in comparisons of the individual decomposed components only, such as the mean long-term trend, and also for the analysis of quantifiable seasonal characteristics, such as the mean seasonal cycle amplitude.

Carslaw (2005) states that it is difficult to assess the relative performance of different curve fitting programs, as there is no obvious point of reference against which different programs can be compared. Thus, in order to provide a robust, quantitative framework for comparing outputs from the three curve fitting programs, we have used $\pm 1\sigma$ standard deviations of the monthly mean data as an uncertainty estimate of the data, and to provide a quantitative point of reference, to which we have compared differences in the curve fitting program outputs. Differences between curve fitting program outputs that were larger than the uncertainty of the data were deemed to be significant within the context of the data set.

3 Results and discussion

3.1 Curve fits

We first ran all three curve fitting programs on all five data sets listed in Table 1, using the input parameters shown in bold in Table 2. The differences between the three program curve fits are smallest for the BHD CO₂ and CGO CH₄ time series. These two time series are the least challenging for the programs, owing to the relatively small seasonal cycle amplitudes of the data. The largest differences between the program curve fits are for the ALT CO₂ and ALT CH₄ time series (with differences of up to 2 ppm and 15 ppb respectively), which both have relatively large seasonal cycle amplitudes (ALT CH₄ shown in Fig. 1a). The programs produce much more similar curve fits to each other for the ALT CO₂ data than for the ALT CH₄ data, and this is likely because of the more complex seasonal pattern exhibited by CH₄ as mentioned in Sect. 2.3 above. Figure 1b shows the residuals of the ALT CH₄ curve fits, which are distributed around zero for all three programs, and are almost always largest for HPspline and smallest for CCGCRV.

Across all five time series, the differences between the curve fits are notably largest between HPspline and CCGCRV, and the smallest differences are either between HPspline and STL (for ALT CO₂ and BHD CO₂), or CCGCRV and STL (for ALT CH₄, BHD O₃, and CGO CH₄). The largest curve fit differences between the three programs most often coincide with the timing of the seasonal maxima and minima (as shown in Fig. 1a), where the curve fitting programs have a tendency to either under- or over-estimate the seasonal inflexion points. As shown in Fig. 2, the differences between the curve fits generated by the three programs are often significant because they exceed the uncertainty of the monthly mean measurements (as represented by the 1 σ standard deviations of the monthly means).

Comparing how closely the programs fit to the data points can provide useful insight into the appropriate use of a curve fitting program, even though the purpose of fitting curves to atmospheric time series is not to fit as closely to the data points as possible, as mentioned previously. The curve fits from CCGCRV are consistently closest to the data points for all five time series (see Table 3). In the case of the BHD CO₂ time series, the CCGCRV curve fit is within $\pm 1\sigma$ standard deviation for 99.4 % of the data. The HPspline curve fits are the most distant from the data for all five time series; the closest agreement between the HPspline curve fit and the data is for the BHD CO₂ time series, where the curve fit captures 90.0 % of the data. This pattern is congruent with the residuals of the curve fits, which are smallest for CCGCRV, largest for HPspline, and intermediate for STL across all five time series (shown in Fig. 1b for ALT CH₄).

It is important not to arrive at the erroneous conclusion that CCGCRV performs “better” than STL or HPspline because it produces a curve fit that is closest to the input data points. As

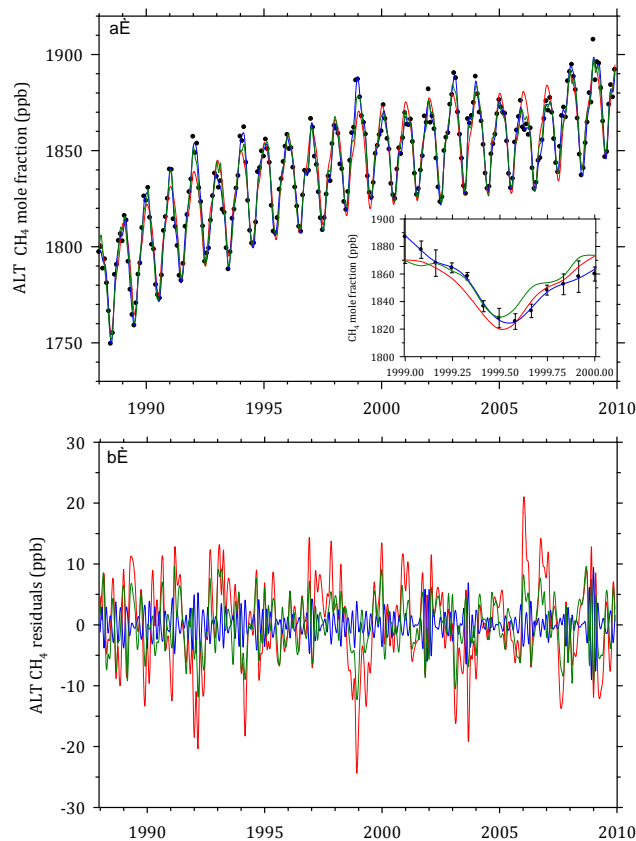


Figure 1. (a) HPspline (red line), CCGCRV (blue line) and STL (green line) curve fits to monthly means of CH₄ mole fraction measured at Alert, Canada (ALT). The inset shows a single year of data, with the SD of the observed monthly means represented by the black error bars. Error bars are not shown on the full size plot to retain clarity. In general, all three curve fits lie within the SD limits. HPspline exceeds the SD limits for 34.1 % of the data, STL exceeds the SD limits for 16.3 % of the data, and CCGCRV exceeds the SD limits for 2.3 % of the data. As with CO₂, the curve fitting programs tend to exceed the SD limits at the seasonal maxima and minima, where the programs over- or under-fit the seasonal inflexion points. (b) Residuals of the HPspline (red line), CCGCRV (blue line) and STL (green line) curve fits to monthly means of CH₄ mole fraction measured at ALT. HPspline produces the largest residuals, CCGCRV produces the smallest residuals, and STL produces intermediate residuals.

stated previously, the purpose of applying curve fitting programs to atmospheric time series is to separate the time series into trend, seasonal and residual components. By producing a curve fit that is closer to the data, CCGCRV retains more short-term variation in the seasonal and trend components of the fit, thus resulting in smaller residuals. In contrast, HPspline is much less “flexible”, meaning that the curve fits do not follow closely to the original data points as often, particularly for time series with large interannual variations in the seasonal cycle, such as the ALT CH₄ time series. Hence, for HPspline, a larger proportion of the variation in the data set

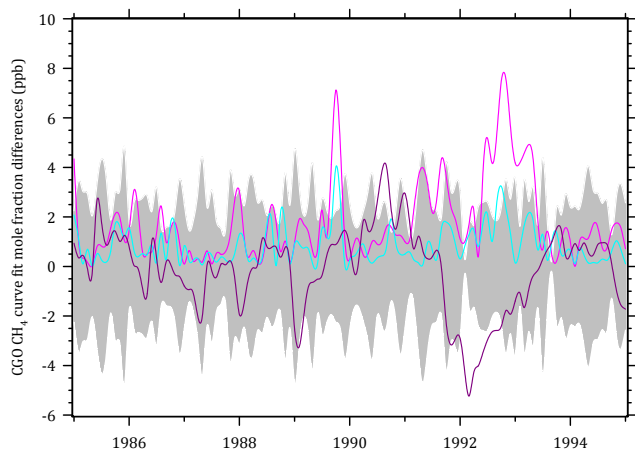


Figure 2. Comparison of the curve fit mole fraction differences between HPspline and CCGCRV (pink), CCGCRV and STL (cyan), and HPspline and STL (purple) for a subset of the CH₄ time series at Cape Grim, Australia (CGO) (between 1985 and 1995). The SD of the monthly mean measurements (grey shading) are also shown for comparison. The SD consist of discrete data points at monthly intervals, however, we have chosen to represent them as a continuous band to aid visual comparison to the curve fit differences. The largest differences are between HPspline and CCGCRV (pink), although differences between all three programs sometimes exceed the SD of the observations.

is assigned to the residual component of the decomposition compared to the CCGCRV and STL decompositions. This lack of flexibility in the HPspline curve fits is attributed to the spline stiffness component of the program.

Carslaw (2005) states that seasonality is a concept that cannot be explicitly defined, and there is no definitive boundary between what constitutes seasonal and residual variation, hence it is vulnerable to subjective interpretations. A curve fitting program that incorporates some interannual variability within the seasonal component of the fit cannot be said to perform either better or worse than a program that assigns the same interannual variation to the residual or long-term components of the fit. Even so, some curve fitting programs may be better or worse suited to certain types of analyses, or for decomposing certain types of data, based on how the trend, seasonal and residual variations in a time series are extracted. What is apparent from our analyses is that CCGCRV and STL attribute more interannual and short-term variability in the data to the seasonal component of the time series, whereas HPspline attributes more of such variability to the residual component.

3.2 Long-term trends

Overall, the long-term trend curves produced by the three programs agree well when the mean slopes of the trends for the entire time series are compared (e.g. for ALT CH₄, see Fig. 3b). This is not surprising for CCGCRV and HP-

Table 3. Percentage of data points “captured” by HPspline, CCGCRV and STL for the five atmospheric time series. “Captured” points are those for which the fitted curve passes within $\pm 1\sigma$ standard deviations of the monthly mean for each data point.

Station	Gas species	CCGCRV (%)	STL (%)	HPspline (%)
ALT	CO ₂	95.8	88.6	76.9
	CH ₄	97.7	83.7	65.9
BHD	CO ₂	99.4	95.9	90.0
	O ₃	83.5	73.4	67.5
CGO	CH ₄	99.1	94.5	73.6

spline, which both use third-degree polynomial functions as part of the trend calculation. It is reassuring that the STL trend curves also agree well, since STL decomposes the trend variation using a very different process. There are, however, some large differences in the trends apparent on short-term timescales, particularly for the more variable ALT CH₄ (see Fig. 3a) and BHD O₃ data sets (differences of up to 10 and 1.5 ppb respectively). For these time series, the HPspline trends are smoother than those produced by CCGCRV and STL, which incorporate more high-frequency variation into the trend component of the fit. Many of these short-term differences between the HPspline trends and those of CCGCRV and STL are statistically significant. Additionally, these short-term differences may bias estimates of the mean long-term trend of a time series when they occur at the ends of the time series, although this is not the case for the five time series we have examined here. Our results indicate that the stiffness of the spline component of HPspline causes the program to produce smoother trends than those produced by CCGCRV and STL.

Figure 4 shows the long-term trend growth rate results for CGO CH₄ mole fraction, which are analogous to the long-term trend results. Again, the HPspline growth rates are much smoother than those produced using CCGCRV and STL; this result also applies to the ALT CH₄ and BHD O₃ time series (with differences between programs of up to 19.5 and 3 ppb yr⁻¹ respectively), owing to the exclusion of high-frequency variations from the trend component by the stiff cubic spline. Figure 4 also shows a “ringing effect” superimposed on the HPspline growth rate curve that increases in amplitude towards the ends of the time series, and is an artefact of the stiff spline. The largest differences in growth rate are between the HPspline and CCGCRV curves, many of which are statistically significant on short-term timescales. There are, however, no significant differences between the mean growth rates of the entire data set for any of the five time series. We also find that STL sometimes produces growth rate curves that have relatively very large spurious variations at the ends of the data sets, which likely arise due to the loss

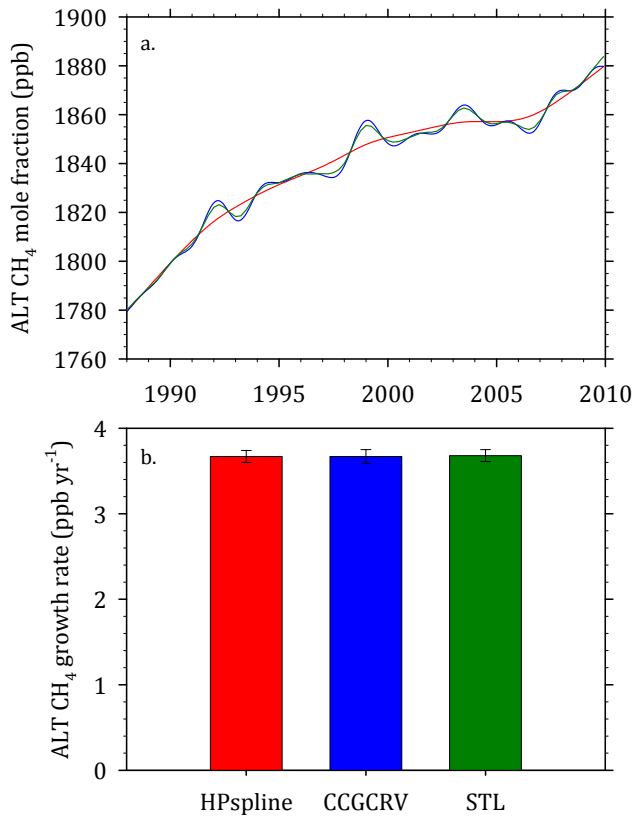


Figure 3. (a) Long-term trends of monthly mean CH₄ observations at ALT produced by HPspline (red line), CCGCRV (blue line) and STL (green line). (b) Although there are large differences between the long-term trends of the three curve fitting programs over short timescales, the mean slope of the long-term trends (long-term growth rate) for the entire time series are very similar for all three programs. Error bars shown indicate the standard error of the regressions of the mean long-term trends.

and subsequent extension of the data during Step 3 of the STL fitting procedure (see Sect. 2.1.3 above).

For both the trends and growth rates, the CCGCRV curves consistently display more interannual variation than the STL curves, which is likely due to the shorter typical CCGCRV trend smoothing period (f_1) of 667 days, compared to a typical STL trend smoothing value (*twin*) of 25 months (~ 760 days). Although there are large differences between the trend and growth rate curves on short-term timescales for some of the time series, it is reassuring that there are no significant differences between the trend and growth rate means for the whole time series, since for some atmospheric greenhouse gases, such as CO₂, even very small differences in the long-term atmospheric accumulation over the past decades could propagate into very large differences in future projections of atmospheric CO₂ mole fraction. Our findings do indicate, however, that care must be taken in studies that examine and report the most recent behaviour in the accumulation of greenhouse gases in the atmosphere, for example by the

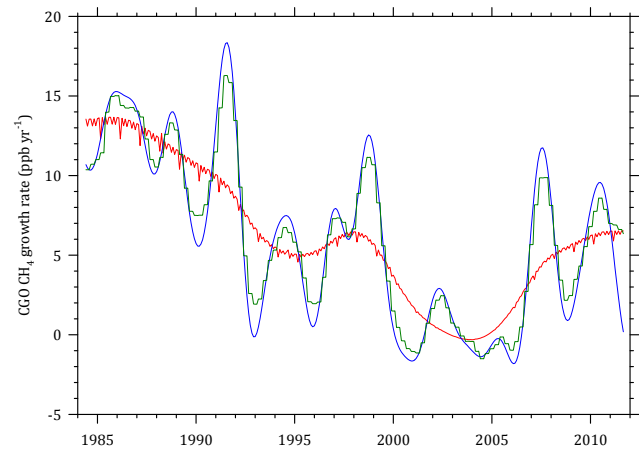


Figure 4. Growth rates of the long-term trend in monthly mean CH₄ observations at CGO produced by HPspline (red), CCGCRV (blue), and STL (green). As with the long-term trends, the mean growth rates calculated for the whole time series are similar for all three curve fitting programs, although there are large differences on short timescales. HPspline calculates a growth rate that is much smoother than those calculated by CCGCRV and STL, due to the stiffness of the spline component of the HPspline fitting procedure. The “ringing” effect superimposed on the HPspline growth rate curve is caused by the stiff spline and increases in magnitude towards the ends of the time series.

WMO (2014) and the Global Carbon Project (Le Quéré, et al., 2014), since such results are sensitive to the curve fit program used, both because of the short timescales involved (a few years or less) and because of possible end effects.

3.3 Seasonal cycles

Comparing how effectively the curve fitting programs represent interannual variations in the seasonal cycle demonstrates that HPspline is least able to follow interannual variability in the magnitude of the seasonal minima and maxima for all five time series; e.g. the HPspline curve captures less than 50 % of the ALT CO₂ maxima and CGO CH₄ minima data points (see Fig. 5). CCGCRV is able to capture interannual seasonal variability the most effectively, and STL has intermediate effectiveness, capturing at least 70 % of the maxima and minima for all the time series. Unlike Nakazawa et al. (1997), Trivett et al. (1989) and Tans et al. (1989), we find that CCGCRV and STL are able to adequately represent the deep summer CO₂ minima at ALT (the programs fitted 91 % and 95 % of these minima respectively), a feature that is characteristic of high-latitude Northern Hemisphere stations; HPspline was also able to represent 78 % of the deep summer CO₂ minima at ALT, but under-estimated some of the deepest CO₂ minima significantly.

Comparison of the magnitudes of the seasonal cycle maxima and minima show that HPspline produces higher seasonal maxima values and lower minima values, whilst con-

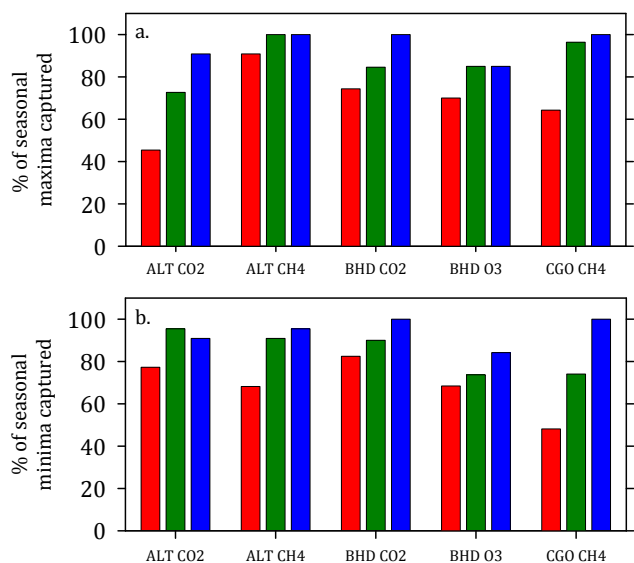


Figure 5. Percentage of seasonal maxima (a) and seasonal minima (b) “captured” by HPspline (red bars), STL (green bars) and CCGCRV (blue bars) for all five time series. The term “captured” refers to the curve fit passing within the $\pm 1\sigma$ standard deviation limits of the monthly mean data at the seasonal inflexion points. With one exception (ALT CO₂ minima), CCGCRV always captures the greatest number of seasonal maxima and minima (approximately 94% across all five time series), STL captures approximately 86% across all five time series, and HPspline captures the least: approximately 68% across all five time series. This difference between the three programs reflects their comparative “flexibility”, which is partially determined by the input smoothing parameters of the program settings.

versely, STL produces lower seasonal maxima values and higher seasonal minima values; CCGCRV produces intermediate values of both the seasonal maxima and minima. The differences between the mean magnitudes of the seasonal inflexion points are statistically significant to the 95% confidence level in some cases. For example, the differences in the mean seasonal maxima calculated by HPspline and STL for the ALT CO₂ and BHD CO₂ time series are significant at 0.23 and 0.10 ppm respectively. In addition, the STL seasonal minima occur on average 6 days earlier than those of HPspline, which is also statistically significant.

Figure 6a shows the differences in the ALT CO₂ seasonal cycle amplitude produced by the three curve fitting programs. The mean seasonal cycle amplitudes produced by HPspline, CCGCRV and STL for the ALT CO₂ data set are 15.3, 15.2 and 15.1 ppm, respectively. As with the seasonal maxima and minima, overall, HPspline produces the largest seasonal cycle amplitudes, STL produces the smallest, and CCGCRV produces intermediate values, for both the ALT and BHD CO₂ time series. The mean seasonal cycle amplitudes produced by HPspline, CCGCRV and STL for the BHD CO₂ data set are 1.24, 1.18 and 1.06 ppm, respectively; the dif-

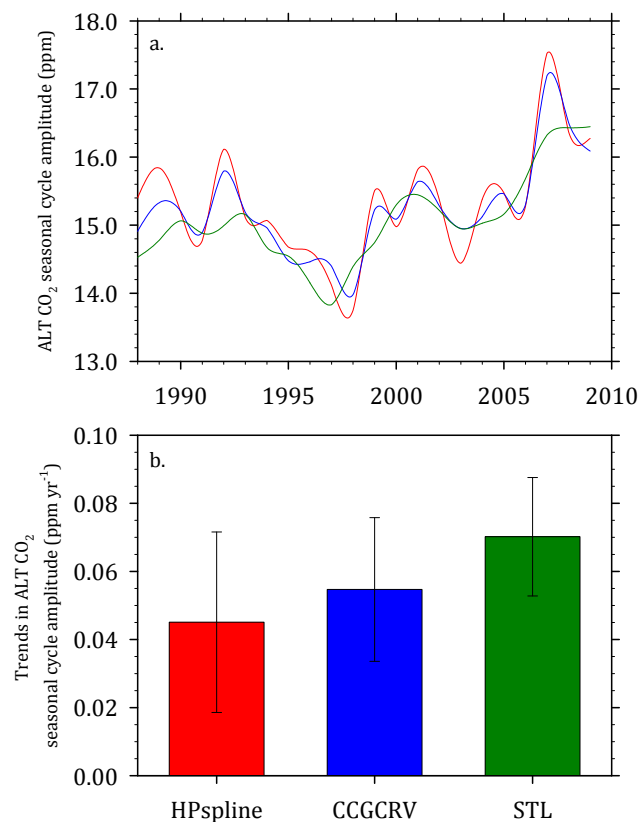


Figure 6. (a) Seasonal cycle amplitude of monthly mean CO₂ observations at ALT produced by HPspline (red), CCGCRV (blue) and STL (green). (b) Mean linear trends in the seasonal cycle amplitude of monthly mean CO₂ at ALT, calculated using the detrended output of the three curve fitting programs. Even though there are relatively large differences between the amplitudes of individual years, the mean trends are not significantly different from one another, as shown by the black error bars, which denote the standard error of the mean amplitude linear regressions.

ference in the HPspline and STL amplitudes of 0.18 ppm is statistically significant. For both time series, all three programs indicate that the value of the seasonal maxima is increasing and the value of the seasonal minima is decreasing; hence, the three programs also show that the seasonal cycle amplitude of CO₂ is increasing in magnitude over time at both ALT (see Fig. 6b) and BHD. These positive trends in the seasonal cycle amplitude are statistically significant for all three programs at both sites, except for the HPspline ALT CO₂ amplitude trend, which is not significant. There are no significant differences between the CO₂ seasonal cycle amplitude trends produced by the three programs, either at ALT (indicated by the error bars in Fig. 6b) or at BHD, which in part is owing to the relatively large interannual variability in the seasonal cycle amplitude.

Many of the differences in the seasonal output from the three programs are also scientifically significant in addition to being significant based on statistics alone. Previous stud-

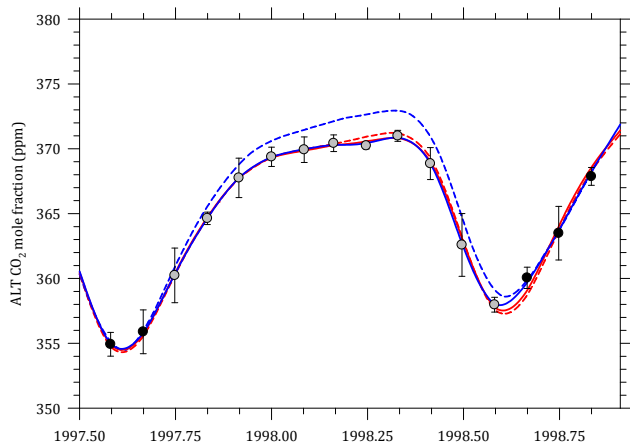


Figure 7. A subset of HPspline (red dashed line) and CCGCRV (blue dashed line) curve fits to monthly mean CO₂ observations at ALT (black dots), where data from October 1997 to August 1998 have been removed in order to create an artificial 11-month gap in the time series (solid grey dots). The curve fits for the two programs with no artificial gap are also shown by the two solid lines for comparison. Error bars indicate the SD limits of the monthly mean observations. Since STL is unable to recognise gaps in the time series, the program processes the data as if no gap exists and so it cannot be plotted against the other two program outputs. CCGCRV is severely affected by the gap as indicated by the significant increase in the seasonal maximum. HPspline is relatively robust to the gap in the time series, which only causes a small change in the curve fit.

ies that have examined changes in the seasonal cycle of greenhouse gases over long time periods, such as Piao et al. (2008), typically find small trends, for example of less than a day per year in the shift in phasing of the seasonal cycle. Therefore, differences in the phasing of the program outputs on the order of several days (e.g. as found for CO₂ seasonal minima) indicate that results from such studies may be significantly biased by the choice of curve fitting program. In addition, the current WMO/GAW (Global Atmosphere Watch) compatibility goal for measurement stations is ± 0.1 ppm (± 0.05 ppm in the Southern Hemisphere) for CO₂ and ± 2 ppb for CH₄ (Brailsford, 2012); thus, introducing uncertainties greater than these values during the data analysis process simply because of the choice of curve fitting program is scientifically significant and should be avoided if possible.

3.4 Gaps and outliers

The programs were assessed for their ability to cope with gaps in the time series by introducing artificial gaps with durations from 3 to 11 months into the time series. Figure 7 shows the effect of an 11-month gap in the ALT CO₂ time series on the HPspline and CCGCRV curve fits. STL is not shown in Fig. 7, since the programming code requires that the data are regularly spaced and is unable to recognise the artificial gap. This is effectively because STL does not take

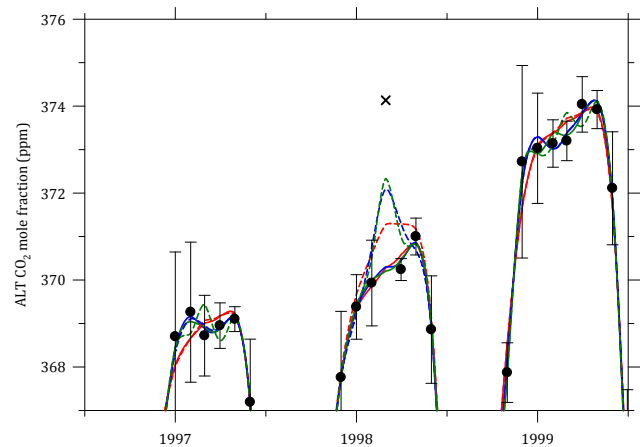


Figure 8. A subset of HPspline (red dashed line), CCGCRV (blue dashed line) and STL (green dashed line) curve fits to monthly mean CO₂ observations at ALT (black dots) from 1996 to 2000, where the monthly mean for March 1998 has been replaced with an artificial outlier (black cross) that has a value 1% greater (~ 4 ppm) than the original measured value. The curve fits for the three programs with no artificial outlier are also shown by the three solid lines for comparison. Error bars indicate the SD limits of the monthly mean observations, and the y axis has been scaled to aid visual comparison of the curve fit differences. As shown, both CCGCRV and STL are significantly affected by the outlier at the point of occurrence. Additionally, the timing of the seasonal maximum has shifted two months earlier for the STL and CCGCRV curves, to coincide with the occurrence of the outlier in March 1998. STL is also affected by spurious variation in the adjacent years, which shifts the timing of the seasonal maximum earlier by two months in 1997. HPspline is relatively robust to the influence of the outlier, however, the timing of the seasonal maximum is also shifted earlier in the year by two months during 1998.

into consideration the time stamps of the data, only the frequency of the input data, $n_{(p)}$, which is defined by the user (see Sect. 2.1.3) and is assumed to be constant throughout the time series. When run with exactly a year of data missing, STL processes the data as if there were no gap at all, whereas gaps that are shorter or longer than 12 months cause very large fitting anomalies in all output after the gap.

CCGCRV curve fits are significantly affected by gaps in the time series, as are the long-term trend and growth rate components of the CCGCRV decomposition. In contrast, HPspline is relatively unaffected by gaps in the time series for all components of the decomposition. This indicates that the Reinsch spline part of HPspline is more robust to gaps than the CCGCRV filtering, since if either the polynomial or harmonic functions were the vulnerable component, which are common to both programs, one would expect the HP-spline trend to be affected similarly to CCGCRV. Varying the time of year of the gaps has no effect on the response of the curve fitting programs, with the exception of the ALT CO₂ time series, for which the anomalies caused by the gaps are

larger when the gaps incorporate the seasonal maxima. This is most likely owing to the asymmetric shape of the ALT CO₂ seasonal cycle in which the seasonal maximum constitutes a large proportion of the year. We also note that our tests were conducted on relatively long time series of 20 or more years; for shorter time series, we would expect both HPspline and CCGCRV to be less robust to gaps.

The three programs were also tested for vulnerability to outliers by replacing one of the time series data points with a data point that was either 1 % greater or less than the original value. These tests reveal that all three programs are affected to some degree by even such relatively small outliers, and by only a single outlier in time series of 20 or more years' duration. Across all five time series, CCGCRV is generally the most sensitive of the three programs to outliers, although STL is also significantly affected, and sometimes more so than CCGCRV (see Fig. 8). In addition, anomalies in the CCGCRV curve fits only occur at the point in the time series where the outlier also occurs, whereas the STL curve fits are characterised by anomalies in the preceding and subsequent years also. The detrended output of the decomposition is more severely affected by the outliers than the trend and growth rate outputs for both CCGCRV and STL, which suggests that the f_s and $swin$ smoothers are more susceptible to outliers than the f_1 and $twin$ smoothers.

As with the artificial gaps, HPspline is the most robust program to outliers in the time series, although some of the outlier anomalies in the HPspline curve fit are significant (see Fig. 8). Importantly, an outlier that occurs near to the inflexion points of the seasonal cycle can influence the magnitude and/or the timing of the seasonal maximum or minimum for any of the programs (see Fig. 8), and can cause significant biases in the seasonal analysis of a time series. Although HPspline is relatively robust to outliers, the curve fit reveals small anomalies that “echo” throughout the entire data set, coincide with the timing of the outlier within the seasonal cycle, and diminish in magnitude with increasing time (both forwards and backwards in time from the occurrence of the outlier). We believe that this “echo” is an artefact of the spline, since it is not present in the trend (which is independent of the spline), and disappears from the curve fit and detrended outputs when the flexibility of the spline is increased (that is, by increasing the spline stiffness parameter, $SD2$). Therefore, while less severely affected, a larger proportion of the HPspline curve fit could be biased by a single outlier in the time series which may not be obvious on first inspection, whereas the effect of an outlier will be more easily recognisable in the CCGCRV and STL outputs (and thus, it is likely to be easier to filter the outlier as a spurious point and re-compute the curve fits).

The “echo” effect resulting from outliers in the HPspline curve fits becomes more apparent when an outlier is placed at the beginning of the time series (i.e. within the first seasonal cycle). Both CCGCRV and STL are affected by such an outlier at the time that it occurs, but the curve fits for

the rest of the time series remain unaffected. HPspline, however, is more severely affected by this outlier, both at the time that it occurs, and throughout the rest of the time series than when it occurs in the middle of the time series. This effect, where a curve fitting program is more susceptible to time series anomalies when they occur at the ends of the data set, is known as an “end effect”. HPspline is not the only curve fitting program found to be susceptible to end effects, as STL also occasionally exhibited significant end effects in the long-term trend growth rate (Sect. 3.2 above).

3.5 Program input parameters

The ranges of input parameters tested are shown in Table 2. These ranges were chosen to test the limits of the three curve fitting programs, and are therefore not necessarily appropriate for all analyses of atmospheric time series. The input parameters tested include the “stiffness” of the spline component of the HPspline fitting program (i.e. the $SD2$ setting), the f_s and f_1 smoothing parameters for CCGCRV, the $swin$ and $twin$ smoothing parameters for STL, and the number of harmonic and polynomial terms included in the HPspline and CCGCRV programs. For the five time series that we used in these tests, our results show that changing the number of polynomial terms in the CCGCRV and HPspline fitting procedures has no significant effect, while changing the number of harmonics only has a small effect on the HPspline curve fits and residuals but no effect on the CCGCRV output. Hence, only the spline stiffness (HPspline) and smoothing parameters (CCGCRV and STL) have any significant influence on the curve fits and decomposed outputs from the programs.

Reducing the spline stiffness of HPspline (increasing $SD2$) significantly increases the flexibility of the program, allowing a much greater amount of interannual variability to be incorporated into the curve fits (see Fig. 9a, red lines), although still less than the interannual variability that is incorporated into the CCGCRV and STL curve fits (with typical smoothing values for the latter two programs – see bold values in Table 2). This results in significant differences in other components of the decomposition, such as the seasonal cycle amplitude (see Fig. 9b) and the long-term trend (see Fig. 10).

For CCGCRV, using a smaller period for f_s only affects the curve fit and detrended output significantly when a very small value is used (see Fig. 9a and b, blue lines), since CCGCRV is already able to track much of the variability in the data sets. Increasing f_s , however, has a greater effect (also shown in Fig. 9a and b, blue lines). Varying f_1 has a significant effect on the long-term trend (see Fig. 10, blue lines) and growth rate components of the fit, particularly when relatively small values (e.g. $f_1 = 200$) are used, and some higher-frequency variations are included in the trend.

The STL curve fits and decomposed outputs can also be significantly influenced by varying the program input smoothing parameters. Using larger values for both $swin$ and

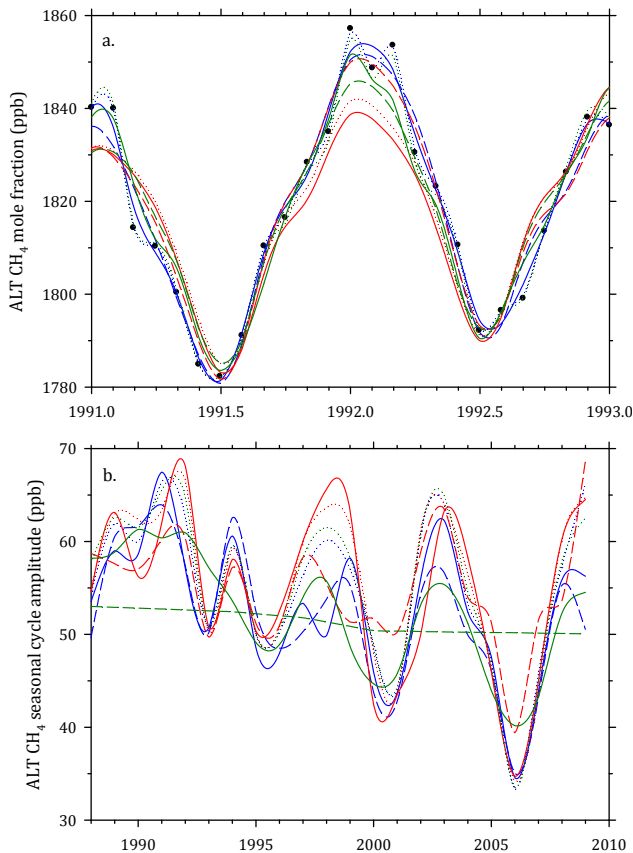


Figure 9. (a) HPspline (red), CCGCRV (blue) and STL (green) curve fits to monthly means of CH_4 mole fraction (black dots) measured at ALT. For clarity, only the subset 1991–1993 is shown. Solid lines denote curve fits produced with the typical input smoothing parameter settings (see bold values in Table 2). Dotted lines denote curve fits produced with the following input parameter settings: $SD2 = 750$ for HPspline (red), $f_s : f_1 = 5 : 667$ for CCGCRV (blue), and $swin : twin = 1 : 25$ for STL (green). Dashed lines denote curve fits produced with the following input parameter settings: $SD2 = 99\,000$ for HPspline (red), $f_s : f_1 = 200 : 667$ for CCGCRV (blue), and $swin : twin = 25 : 25$ for STL (green). (b) Seasonal cycle amplitude of monthly mean CH_4 observations at ALT produced by HPspline (red), CCGCRV (blue) and STL (green), using the typical input smoothing parameter values (solid lines). Dotted and dashed lines for HPspline (red), CCGCRV (blue) and STL (green) are produced using the same input parameter settings as in (a).

$twin$ shifts the curve fit upwards, resulting in higher values for the annual mean mole fractions, as well as higher seasonal maximum and minimum values; correspondingly, using smaller values for the $swin$ and $twin$ parameters results in lower values of annual means and seasonal maxima and minima. Decreasing only the $swin$ parameter produces a more flexible curve fit, while increasing the $swin$ parameter produces a less flexible curve fit (see Fig. 9a, green lines). Changes in the $swin$ parameter can cause large differences in the seasonal cycle amplitude, as shown in Fig. 9b (green

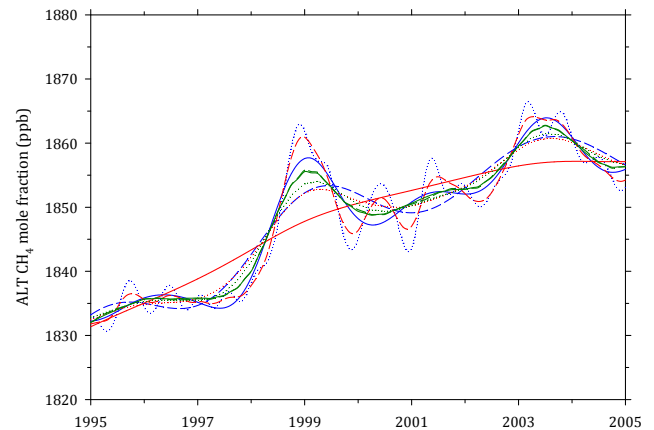


Figure 10. A subset of the long-term trends of monthly mean CH_4 observations at ALT produced by HPspline (red), CCGCRV (blue), and STL (green), using the typical input smoothing parameter values (solid lines; see bold values in Table 2). Dotted lines denote long-term trends produced with the following input parameter settings: $SD2 = 750$ for HPspline (red), $f_s : f_1 = 60 : 200$ for CCGCRV (blue), and $swin : twin = 5 : 5$ for STL (green). Dashed lines denote long-term trends produced with the following input parameter settings: $SD2 = 99\,000$ for HPspline (red), $f_s : f_1 = 80 : 1200$ for CCGCRV (blue), and $swin : twin = 5 : 500$ for STL (green).

lines). In contrast, decreasing the $twin$ parameter has only a moderate effect on the long-term trend output, and increasing the $twin$ parameter has almost no effect (see Fig. 10, green lines).

In order to more directly compare the three curve fitting programs, we manipulated the input smoothing parameters in an attempt to make the programs produce curve fits and decomposed outputs that were as similar to each other as possible. One combination of input smoothing parameter settings that resulted in very similar results across all three programs, whilst still maintaining a relatively high level of flexibility in the curve fits, was: $SD2 = 99\,999$ for HPspline; $f_s = 91$ days and $f_1 = 667$ days for CCGCRV; and $swin = 3$ years and $twin = 22$ months for STL. It should be noted that there are many other combinations of input smoothing parameter settings that can also generate very similar results from the three curve fitting programs, and that we have presented just one combination in this work. In general, the three programs can be forced to produce relatively similar curve fits, although the level of agreement depends on the interannual variability of the input time series. Notably, we find no combination of smoothing parameter values that produces similar curve fit results as well as similar decomposed components of the fitting procedure, since one combination of smoothing parameter values may produce similar curve fits but different trend and detrended outputs, and vice versa.

Altering the input smoothing parameters of the three programs causes many of the outputs from some of the previous tests in Sects. 3.1 to 3.4 to change. For example, we find

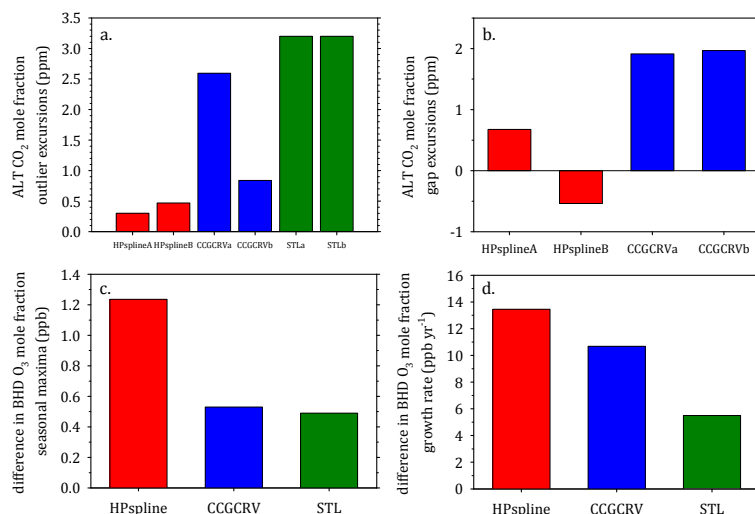


Figure 11. Four plots demonstrating the dependency of the curve fits on the program input smoothing parameters. **(a)** Excursions in the ALT CO₂ curve fits caused by an outlier (1 % greater than the original data point) in March 1998, where for HPsplineA, $SD2 = 30$; for HPsplineB, $SD2 = 9999$; for CCGCRVa, $f_s : f_1 = 80 : 667$; for CCGCRVb, $f_s : f_1 = 200 : 667$; for STLa, $swin : twin = 5 : 25$; and for STLb, $swin : twin = 25 : 25$. **(b)** Excursion in the ALT CO₂ curve fits caused by an 11 month gap during 1998, where the program input smoothing parameters are varied as for **(a)**. Note that the STL R program cannot fit across gaps and for this reason was not included in this plot. **(c)** Differences in the BHD O₃ 1998 seasonal maximum generated by varying the input smoothing parameters with the following values: HPspline: $SD2$ values of 30 and 1000; CCGCRV: $f_s : f_1$ periods of 80 : 667 and 200 : 667; STL: $swin : twin$ values of 5 : 25 and 25 : 25. **(d)** Differences in the BHD O₃ 2001 long-term trend growth rate generated by varying the input smoothing parameters with the following values: HPspline: $SD2$ values of 30 and 9999; CCGCRV: $f_s : f_1$ periods of 80 : 667 and 80 : 200; STL: $swin : twin$ values of 5 : 25 and 5 : 5.

that using the input smoothing parameter values that result in the most similar curve fits, STL is the most flexible program and CCGCRV is only slightly less flexible, which is converse to the previous result, when the typical smoothing parameter settings were used. With smoothing parameters producing the most similar curve fits, STL is also more severely affected by outliers than CCGCRV, which again is in contrast to the previous outlier test results using typical values. Figure 11 summarises four examples of how the curve fitting outputs can be substantially influenced by changing the input smoothing parameters of each program, and demonstrates how it is possible to obtain entirely different results from a time series using the same curve fitting program with different input smoothing parameter values. Furthermore, as demonstrated in Fig. 9a and b, the differences in curve fitting output caused by changing the input smoothing parameter values are often greater than those that result from using different programs. Thus, not only might a time series analysis be biased by the choice of curve fitting program, but also by the choice of input smoothing parameters.

Figure 11 also indicates that in a few circumstances, the curve fitting programs produce the same outputs, despite using very different values of the input parameter smoothing values. For example, it appears that STL has an almost identical response to a 1 % outlier when the $swin$ value is varied from 5 to 25 years (Fig. 11a), and CCGCRV has a very sim-

ilar response to an 11-month gap in the time series when the f_s period is varied from 80 to 200 days (Fig. 11b).

To test the difference in fitting behaviour between the three programs further, we applied each program to an artificially created data set, composed of a relatively large seasonal cycle amplitude that varies interannually, a slowly varying long-term trend, and random noise (with a normal distribution). The black dots of Fig. 12a and b respectively show the long-term trend and seasonal cycle components of this artificial data set. The purpose of this test was first to determine how accurately each of the programs could decompose an artificial time series into its component parts, which are known for an artificial data set (unlike the decomposed components of a real time series, for which the “true” decomposed components are unknown), and second to assess which input smoothing parameter settings produce the most accurate decompositions.

We find that using the typical input smoothing parameter values, CCGCRV and STL assign too much short-term variation in the long-term trend and seasonal cycle components. The following input smoothing parameter values are found to produce the most accurate decomposition for the trend, seasonal and residual components of the time series composition: $SD2 = 500$ for HPspline; $f_s = 250$ days and $f_1 = 1500$ days for CCGCRV; and $swin = 8$ years and $twin = 45$ months for STL (see Table 2 for how these values compare to the typically used values). For CCGCRV and

STL, both the seasonal and trend smoothing parameters are increased compared to the typical values (making the programs less flexible), in order to produce the most accurate decomposition of the data. For HPspline, changing the $SD2$ parameter has little effect on the decomposition, although a higher value (more flexible) produces a slightly more accurate decomposition.

Using these non-typical input smoothing parameter settings, all three programs are able to successfully represent the long-term trend, with the largest differences occurring at the ends of the time series (see Fig. 12a). For the seasonal variation, CCGCRV and STL are able to represent the seasonal cycle slightly more successfully than HPspline, which includes too much short-term variation in the seasonal component (see Fig. 12b). The differences between the three programs for the seasonal component are much larger than those for the trend component.

The means of the residual component are similar across the three programs (not shown), although there are large differences on short timescales, and we expect that any correlation of the residual component with climate indices (e.g. the El Niño Southern Oscillation index) would produce very different results for each of the three programs. The curve fits are also relatively similar to each other, but do not produce a close fit to the artificial time series data points. These results emphasise that the purpose of time series decomposition is not necessarily to fit the data as closely as possible, since this may introduce more short-term variation into the trend and seasonal components than is actually present in the data, and may also underestimate the magnitude of the residual component. Thus, although we have found HPspline to be the least flexible of the three curve fitting programs examined, the input smoothing parameter values for both CCGCRV and STL had to be substantially altered from the typical values used in the community in order to produce an accurate decomposition. In contrast, the typical smoothing value for HPspline was only changed by a relatively small amount in order to produce the best fit to the decomposed data, and this change had little effect on the results compared to the typical value.

3.6 Re-analysis of zero-crossing trends of Piao et al. (2008) using HPspline and STL

In order to demonstrate the importance of whether scientific conclusions are unduly influenced by the choice of curve fitting program, we have used HPspline and STL to replicate the zero-crossing analysis in Piao et al. (2008), who used CCGCRV. In brief, Piao et al. (2008) used CCGCRV to detrend CO_2 time series from 10 Northern Hemisphere field stations from the Globalview- CO_2 database (GLOBALVIEW- CO_2 , 2004), linearly interpolated the detrended data to obtain values of the spring downwards zero-crossing dates (SDZ) and the autumn upwards zero-crossing dates (AUZ), and then for each year, subtracted the SDZ from the AUZ to obtain the carbon uptake period (CUP)

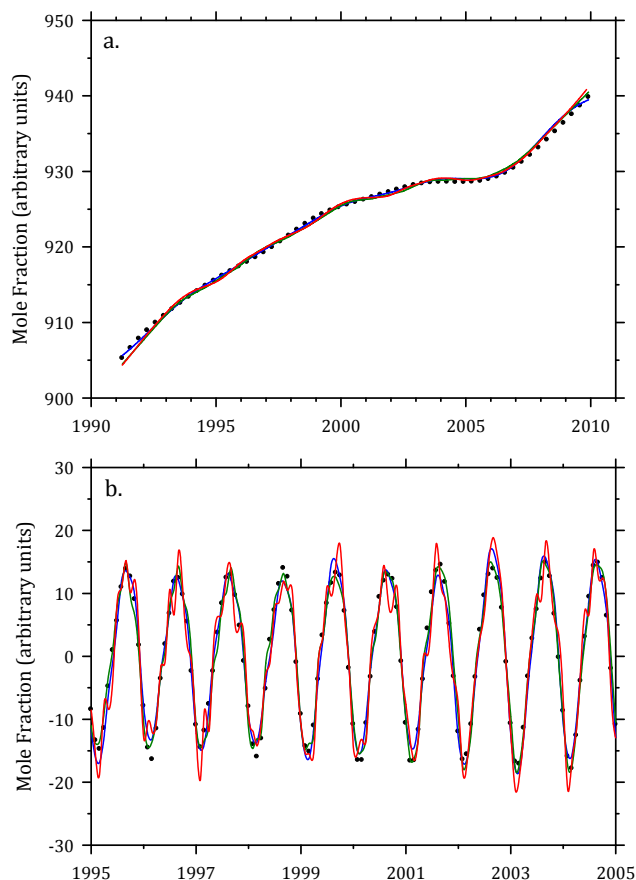


Figure 12. (a) Long-term trends produced by HPspline (red), CCGCRV (blue) and STL (green) for an artificial data set. The actual values of the long-term trend component used to generate the artificial data set are shown by the black dots. The input smoothing parameter values used to achieve these long-term trends are: $SD2 = 500$ for HPspline, $f_s : f_l = 250 : 1500$ for CCGCRV and $swin : twin = 8 : 45$ for STL. (b) A subset of the seasonal cycles produced by HPspline (red), CCGCRV (blue) and STL (green) for the artificial data set. The actual values of the seasonal component used to generate the artificial time series are shown by the black dots. The input smoothing parameter values used to achieve these seasonal cycles are the same as in (a).

of the terrestrial biosphere. Trends in the CUP were determined using linear regressions, and the probability that these trends were statistically significant was calculated.

Figure 13 shows the results of Piao et al. (2008) alongside our re-analysis of the CUP trends with exactly the same input time series, but using STL and HPspline output instead of CCGCRV output. Trends that are positive indicate that the CUP is getting longer, and suggest that the net terrestrial biosphere carbon sink is getting larger, while negative trends indicate the opposite. There are small differences between the CUP trends calculated using the three different curve fitting programs at all of the stations, and a statistically significant difference between the STL and HPspline trends

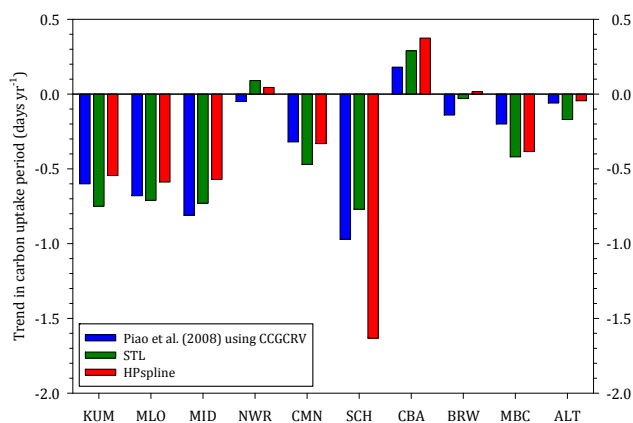


Figure 13. Carbon update period (CUP) trends calculated from 10 atmospheric CO₂ time series in the Globalview-CO₂ (GLOBALVIEW-CO₂, 2004) database. Blue bars indicate the original trends taken from Piao et al. (2008), who used CCGCRV, with green and red bars indicating the re-analysed CUP trends using STL and HPspline respectively. The atmospheric measurement stations are: Cape Kumukahi, USA (KUM), Mauna Loa, USA (MLO), Sand Island, USA (MID), Niwot Ridge, USA (NWR), Mt. Cimone, Italy (CMN), Schauinsland, Germany (SCH), Cold Bay, USA (CBA), Barrow, USA (BRW), Mould Bay, Canada (MBC) and Alert Station, Canada (ALT). Stations are all in the mid-to-high latitudes of the Northern Hemisphere, presented in order of increasing latitude from left to right. The mean standard error of the HPspline linear regressions is ± 0.3 days yr⁻¹, and the mean standard error of the STL linear regressions is ± 0.1 days yr⁻¹.

at SCH (determined based on the standard error of the linear regressions used to calculate the CUP trends). No other station showed statistically significant differences in the output from the three different curve fitting programs, and at all 10 stations, the mean CUP trends agree well for all three curve fitting programs.

Our re-analysis does, however, reveal that the number of stations exhibiting negative and statistically significant CUP trends is dependent on the curve fitting program used. The analysis of Piao et al. (2008) found that nine out of the ten stations have negative CUP trends (see Fig. 13), and that three of these nine negative CUP trends are statistically significant. In contrast, only eight of the STL trends are negative, although five of these are statistically significant (as is one of the positive trends), and for HPspline, 7 of the trends are negative, only one of which is statistically significant.

Although we have found the conclusions from the curve fitting analysis in Piao et al. (2008) to be robust when using any of the three curve fitting programs presented here, the analyses of individual stations are dependent on the curve fitting program used. This one example highlights the importance of investigating the influence of curve fitting bias on the scientific conclusions of an analysis by employing more than one curve fitting program wherever possible.

4 Conclusions and recommendations

We have investigated bias in the application of three commonly used curve fitting programs to monthly mean atmospheric time series from three stations: Alert Station, Canada, Cape Grim Observatory, Australia, and Baring Head, New Zealand. Our comparisons show that there are often significant differences between the outputs of these three programs, and that the outputs are also very sensitive to the choice of program input smoothing parameters. We have also found that the differences between the program outputs depend on the amount of interannual variability in the time series and the seasonal cycle amplitude. For time series with gradual year-on-year changes and/or a relatively small seasonal cycle amplitude, the programs produce much more similar outputs to each other than for time series that are characterised by relatively high interannual variability and/or a relatively large seasonal cycle amplitude. More specifically, we draw the following conclusions from our study:

1. CCGCRV was found to be the most flexible program, HPspline was the least flexible, and STL demonstrated intermediate flexibility, where flexibility describes the amount of short- and long-term variability in the time series that the three programs are able to represent. Hence, the HPspline and CCGCRV curve fits were found to be consistently the least similar across all five time series. The difference in flexibility is also reflected in the residual components of the decomposition, which were consistently largest for HPspline, smallest for CCGCRV, and intermediate for STL. Even when the *SD2* spline stiffness setting of HPspline was increased to its maximum value (minimum stiffness), it was not possible to make the HPspline curve fits as flexible as the CCGCRV and STL curve fits. The fact that HPspline was the least flexible program does not necessarily mean that it is not the most appropriate program to use, as we have demonstrated by applying the three programs to an artificial data set; both CCGCRV and STL produced decompositions of the artificial data set with too much short-term variation in the long-term trend and seasonal components of the decomposition.
2. The mean slopes of the long-term trend and associated growth rate curves agreed well between all three programs over long time periods, although there were some significant differences between the curves for individual years. HPspline tends to generate smoother trend and growth rate curves than the other two programs, as it assigns a greater proportion of the long-term variability to the residual component of the decomposition.
3. For some time series, statistically significant differences were found between the HPspline and STL magnitudes of the seasonal maxima, minima and amplitudes, and the timing of the seasonal inflexion points. This is in

part attributable to STL assigning a smaller proportion of the variation in the time series to the seasonal component of the decomposition. STL and CCGCRV over- and under-estimated the seasonal maxima and minima less frequently than HPspline, which for some time series captured less than 50 % of the seasonal inflexion points. All three programs, however, found that the magnitude of the CO₂ seasonal cycle amplitude at both ALT and BHD has increased over time.

4. The current version of STL cannot be used when gaps exist in the time series, and requires the data to be at evenly spaced intervals. CCGCRV was significantly affected by gaps of 11 months, whereas HPspline was relatively robust to gaps in the time series.
5. All three programs were found to be vulnerable to relatively small outliers (1 % larger or smaller than the original data point) in the time series, although, in general, CCGCRV was found to be the most vulnerable, and STL was more severely affected than HPspline. All three programs were more vulnerable to outliers that occurred near the seasonal inflexion points, and HPspline was the most vulnerable program to outliers that occurred at the ends of the time series.
6. Changing the number of harmonic and polynomial terms in the fitting procedures had no significant effect on the program outputs; however, changing the values of the input smoothing parameters did significantly affect the outputs from all three curve fitting programs. It was not possible to force the three curve fitting programs to produce the same decomposed components of the time series (i.e. trend, seasonal, and residual components) as well as the same curve fits by manipulating the input smoothing parameters.
7. Changing the input smoothing parameters of the programs caused changes in some of the outputs of the previous curve fitting tests, including which program was the most flexible and which was the most susceptible to outliers. Importantly, differences between the results produced by the same curve fitting program run with different input smoothing parameter values were often greater than differences between the results produced by different programs using the typical input smoothing parameter settings.
8. Using STL and HPspline with the typical input parameter settings, we carried out a re-analysis of trends in the terrestrial biosphere carbon uptake period (CUP), determined with an atmospheric CO₂ seasonal cycle zero-crossing analysis as given in Piao et al. (2008), who used CCGCRV. We found that the overall scientific results from Piao et al. (2008) were robust for all three curve fitting programs, but the difference in the HPspline and STL CUP trends at one measurement station

was statistically significant, and the number of negative and statistically significant CUP trends across the suite of 10 stations was dependent on the choice of curve fitting program.

Based on the results of our study, we provide the following list of general recommendations on the use of HPspline, CCGCRV and STL with atmospheric time series of particular characteristics, and for certain types of analyses. Although these recommendations are based on a reasonably comprehensive suite of analyses and comparisons that have been summarised in this paper, they are, however, based on results from examining only five atmospheric time series of monthly averaged data, and therefore may not hold true for all time series. The five time series we used, however, were specifically selected to represent a broad range of characteristics, such as relative magnitude of the seasonal cycle amplitude, and interannual variability in both the seasonal and trend variation.

1. We strongly recommend that users choose appropriate values of the input smoothing parameters based on the characteristics of the time series, and not based on the values that are typically used by colleagues, or that have been used historically. We also highly recommend that users conduct sensitivity tests to ensure that the scientific results of an analysis are not unduly biased by the choice of input smoothing parameter settings.
2. Advice on how to choose appropriate input smoothing parameter settings for CCGCRV and STL is provided in Thoning et al. (1989) and Cleveland et al. (1990) respectively. We recommend that the f_s versus f_1 and $swin$ versus $twin$ values are sufficiently different from each other so that the CCGCRV and STL short- and long-term smoothing procedures do not compete for the same variation in the data.
3. We strongly recommend that users employ more than one curve fitting program, to ensure that the results of an analysis will not be unduly biased by the choice of curve fitting program. This is particularly important for analyses that are more vulnerable to curve fitting bias, such as those that examine relatively small trends or changes in time series that are very variable, and for analyses where a high degree of accuracy is required, such as a comparison of time series from two locations, for example.
4. For analyses where it is appropriate or useful to assign more variation to the residual component of a time series decomposition, and less variation to the trend and seasonal components, we recommend the use of HPspline. A key example is investigating correlations between atmospheric time series and large scale climate phenomena or climate indices such as the El Niño/Southern Oscillation.

5. For analyses where it is important that year-to-year variations in the seasonal and trend components are retained, we recommend CCGCRV, for example in studies such as Piao et al. (2008) examining changes in features of the seasonal cycle. STL is also appropriate for such analyses, although the user should be aware that STL assigns less variation to the seasonal component, and more variation to the trend component compared to CCGCRV, depending on the input smoothing parameter settings used.
6. For analyses where it is important to accurately represent the magnitude or timing of the seasonal inflexion points, and/or the seasonal cycle amplitude, we recommend CCGCRV and STL for all time series data, including analyses where the time series is characterised by deep, short minima or maxima, such as high-latitude Northern Hemisphere atmospheric CO₂ time series. We only recommend the use of HPspline in the above-mentioned seasonal analyses for time series with a relatively constant seasonal cycle shape and phasing throughout.
7. For analyses where it is important to fit the data as closely as possible, we recommend the use of CCGCRV, and discourage the use of HPspline.
8. For studies reporting the most recent growth rate in the accumulation or decline of gases in the atmosphere, we strongly recommend the use of more than one curve fitting program, since growth rate calculations are particularly sensitive to possible end effects.
9. For calculations of mean long-term trends, results can be sensitive to trend values at the ends of the time series, which in turn can sometimes be quite different across the three curve fitting programs. Therefore, for such calculations, we recommend the use of more than one curve fitting program.
10. For analyses of interannual variations of the long-term trend, particularly for time series with variable long-term trend growth rates, such as atmospheric CH₄ time series, we recommend CCGCRV and STL, but not HPspline. We also warn users that STL occasionally generates spurious variations at the ends of the long-term trend growth rate curves.
11. For analyses where the time series contains gaps, or the data are not evenly spaced through time, we recommend HPspline, but not CCGCRV or STL. In particular, users should be aware that the currently available R version of STL is unable to fit across gaps in time series. For analyses including time series with gaps, and where the focus of the investigation is such that another recommendation in this list advises use of CCGCRV or STL, it may be appropriate to first use an interpolation technique to populate the gaps in the time series, and then carry out the analyses with CCGCRV or STL.
12. For analyses where the time series contains outliers, if the purpose is to identify the outliers and remove them from the time series, we recommend CCGCRV, because the program is sensitive to outliers, but only at the time they occur, meaning that outliers are easily recognisable and can subsequently be removed.
13. If the purpose of an analysis of a time series containing outliers is to produce curve fits, or to decompose the data without removing the outliers, we recommend STL or CCGCRV. This is because although HPspline tends to be affected to a lesser degree than STL at the time of the outlier occurrence, the program output is affected throughout a much larger proportion of the time series than the STL and CCGCRV outputs. Additionally, for time series that contain outliers near the ends, we discourage the use of HPspline because it is sensitive to end effects caused by outliers.

Key examples of further work that would improve our understanding of possible curve fitting bias include comparing curve fitting program outputs using higher-frequency time series, such as weekly, daily or hourly averages, conducting curve fitting comparisons using additional curve fitting programs to the three tested here, and conducting comparisons on shorter time series.

Our results clearly show that significant bias and uncertainty can be introduced in the application of curve fitting programs to atmospheric time series. It is thus important that investigators ensure that curve fitting programs are appropriate for the application for which they are used, use more than one program to analyse the same data so that any biases can be identified, and test the sensitivity of the results to the input smoothing parameters chosen.

Great care is taken by experimental scientists to ensure that atmospheric greenhouse gas measurements are very precise, reproducible, and compatible with other measurement sites. The same care and attention is essential in the analysis of these data, and in the application of curve fitting programs, to ensure the robustness and reproducibility of scientific interpretation and conclusions drawn.

Acknowledgements. We thank Paul Krummel (CMAR, CSIRO, Australia) for providing the CCGCRV IDL code and for making the Cape Grim Observatory CH₄ data available on the CSIRO website, and Sara Mikaloff-Fletcher (NIWA, New Zealand) for providing the STL R script used in this work. We also thank Steve Piper (Scripps Institution of Oceanography, USA) and Mike Salmon (UEA, UK) for assistance with HPspline. The Baring Head CO₂ data were kindly provided by Gordon Brailsford (NIWA, New Zealand) and Britton Stephens (NIWA, New Zealand and NCAR, USA). The Alert Station CO₂ and CH₄ data were made

available via WDCGG by Doug Worthy (Environment Canada, Canada), and the Baring Head O₃ data were made available via WDCGG by Sylvia Nichol (NIWA, New Zealand). We also thank Britton Stephens, Gordon Brailsford, Martin Manning (Victoria University of Wellington, New Zealand), Paul Krummel and Sara Mikaloff-Fletcher for their helpful input during the initial phases of this project. Finally, we thank Shilong Piao (Peking University, China), Jonathan Barichivich (UEA, UK), and Ángel Gómez-Peláez (Izaña Atmospheric Research Center, Meteorological State Agency of Spain) for helpful discussions on the use of curve fitting programs, and Robert Jenkins (UEA, UK) for advice regarding the statistical analyses. P. A. Pickers is supported by a UK Natural Environment Research Council Ph.D. studentship (NE/K500896/1).

Edited by: N. Zeng

References

- Amritkar, R. E. and Kumar, P.: Interpolation of missing data using nonlinear and chaotic system analysis, *J. Geophys. Res.*, 100, 3149–3154, doi:10.1029/94JD01531, 1995.
- Bacastow, R. B.: Modulation of atmospheric carbon dioxide by the Southern Oscillation, *Nature*, 261, 116–118, doi:10.1038/261116a0, 1976.
- Bacastow, R. B., Keeling, C. D., and Whorf, T. P.: Seasonal amplitude increase in atmospheric CO₂ concentration at Mauna Loa, Hawaii, 1959–1982, *J. Geophys. Res.-Atmos.*, 90, 10529–10540, doi:10.1029/JD090iD06p10529, 1985.
- Brailsford, G. W. (Ed.): Report of the Sixteenth WMO/IAEA Meeting on Carbon Dioxide, Other Greenhouse Gases, and Related Measurement Techniques (GGMT-2011), Wellington, New Zealand, World Meteorological Organization Global Atmosphere Watch Report Series No. 206, available at: <http://www.wmo.int/pages/prog/arep/gaw/gaw-reports.html>, 2012.
- Carlsaw, D. C.: On the changing seasonal cycles and trends of ozone at Mace Head, Ireland, *Atmos. Chem. Phys.*, 5, 3441–3450, doi:10.5194/acp-5-3441-2005, 2005.
- Chan, Y. H. and Wong, C. S.: Long-term changes in amplitudes of atmospheric CO₂ concentrations at Ocean Station P and Alert, Canada, *Tellus B*, 42, 330–341, 1990.
- Cleveland, R. B., Cleveland, W. S., McRae, J. E., and Terpenning, I.: STL: A seasonal-trend decomposition procedure based on Loess, *Journal of Official Statistics*, 6, 3–33, 1990.
- Dlugokencky, E. J., Nisbet, E. G., Fisher, R., and Lowry, D.: Global atmospheric methane: budget, changes and dangers, *Philos. T. Roy. Soc. A*, 369, 2058–2072, doi:10.1098/rsta.2010.0341, 2011.
- Elliott, W. P.: The statistical treatment of CO₂ records, NOAA Technical Memorandum ERL ARL-173, Maryland USA, 1989.
- Enting, I. G.: The Interannual Variation in the Seasonal Cycle of Carbon Dioxide Concentration at Mauna Loa, *J. Geophys. Res.*, 92, 5497–5504, doi:10.1029/JD092iD05p05497, 1987.
- GLOBALVIEW-CO2: GLOBALVIEW-CO2: Cooperative Atmospheric Data Integration Project – Carbon Dioxide, CD-ROM, NOAA GMD, Boulder, Colorado, 2004.
- Houghton, R. A.: Balancing the global carbon budget, *Annu. Rev. Earth Planet. Sc.*, 35, 313–347, doi:10.1146/annurev.earth.35.031306.140057, 2007.
- Jones, C. D., Collins, M., Cox, P. M., and Spall, S. A.: The carbon cycle response to ENSO: A coupled climate-carbon cycle model study, *J. Climate*, 14, 4113–4129, doi:10.1175/1520-0442(2001)014<4113:TCCRTE>2.0.CO;2, 2001.
- Keeling, C. D.: The concentration and isotopic abundances of carbon dioxide in the atmosphere, *Tellus*, 12, 200–203, 1960.
- Keeling, C. D., Guenther, P. R., and Whorf, T. P.: An Analysis of the Concentration of Atmospheric Carbon Dioxide at Fixed Land Stations and over the Oceans based on Discrete samples and Daily Averaged Continuous Measurements, Scripps Institution of Oceanography, 1986.
- Keeling, C. D., Bacastow, R. B., Carter, A. F., Piper, S. C., Whorf, T. P., Heimann, M., Mook, W. G., and Roeloffzen, H.: A three-dimensional model of atmospheric CO₂ transport based on observed winds: 1. analysis of observational data, *American Geophysical Union, Geophys. Monogr.*, 55, 165–236, 1989.
- Keeling, C. D., Chin, J. F. S., and Whorf, T. P.: Increased activity of northern vegetation inferred from atmospheric CO₂ measurements, *Nature*, 382, 146–148, doi:10.1038/382146a0, 1996.
- Keeling, C. D., Piper, S. C., Whorf, T. P., and Keeling, R. F.: Evolution of natural and anthropogenic fluxes of atmospheric CO₂ from 1957 to 2003, *Tellus B*, 63, 1–22, doi:10.1111/j.1600-0889.2010.00507.x, 2011.
- Keeling, R. F. and Shertz, S. R.: Seasonal and interannual variations in atmospheric oxygen and implications for the global carbon cycle, *Nature*, 358, 723–727, doi:10.1038/358723, 1992.
- Le Quéré, C., Moriarty, R., Andrew, R. M., Peters, G. P., Ciais, P., Friedlingstein, P., Jones, S. D., Sitch, S., Tans, P., Armeth, A., Boden, T. A., Bopp, L., Bozec, Y., Canadell, J. G., Chevallier, F., Cosca, C. E., Harris, I., Hoppema, M., Houghton, R. A., House, J. I., Jain, A., Johannessen, T., Kato, E., Keeling, R. F., Kitidis, V., Klein Goldewijk, K., Koven, C., Landa, C. S., Landschützer, P., Lenton, A., Lima, I. D., Marland, G., Mathis, J. T., Metzl, N., Nojiri, Y., Olsen, A., Ono, T., Peters, W., Pfeil, B., Poulter, B., Raupach, M. R., Regnier, P., Rödenbeck, C., Saito, S., Salisbury, J. E., Schuster, U., Schwinger, J., Séférian, R., Segsneider, J., Steinhoff, T., Stocker, B. D., Sutton, A. J., Takahashi, T., Tilbrook, B., van der Werf, G. R., Viovy, N., Wang, Y.-P., Wanninkhof, R., Wiltshire, A., and Zeng, N.: Global carbon budget 2014, *Earth Syst. Sci. Data Discuss.*, 7, 521–610, doi:10.5194/essdd-7-521-2014, 2014.
- Manning, M. R., Lowe, D. C., Melhuish, W. H., Sparks, R. J., Wallace, G., Brenninkmeijer, C. A. M., and McGill, R. C.: The use of radiocarbon measurements in atmospheric studies, 13th Radiocarbon International Conference, 32, 37–58, 1990.
- Martin, F. and Diaz, A.: Different methods of modeling the variability in the monthly mean concentrations of atmospheric CO₂ at Mauna Loa, *J. Geophys. Res.*, 96, 18689–18704, doi:10.1029/91JD01710, 1991.
- Minejima, C., Kubo, M., Tohjima, Y., Yamagishi, H., Koyama, Y., Maksyutov, S., Kita, K., and Mukai, H.: Analysis of ΔO₂/ΔCO₂ ratios for the pollution events observed at Hateruma Island, Japan, *Atmos. Chem. Phys.*, 12, 2713–2723, doi:10.5194/acp-12-2713-2012, 2012.
- Nakazawa, T., Ishizawa, M., Higuchi, K., and Trivett, N. B. A.: Two curve fitting methods applied to CO₂ flask data, *Environmetrics*, 8, 197–218, doi:10.1002/(SICI)1099-095X(199705)8:3<197::AID-ENV248>3.0.CO;2-C, 1997.

- Piao, S., Ciais, P., Friedlingstein, P., Peylin, P., Reichstein, M., Luysaert, S., Margolis, H., Fang, J., Barr, A., Chen, A., Grelle, A., Hollinger, D. Y., Laurila, T., Lindroth, A., Richardson, A. D., and Vesala, T.: Net carbon dioxide losses of northern ecosystems in response to autumn warming, *Nature*, 451, 49–52, doi:10.1038/nature06444, 2008.
- Pierce, D. A.: Signal extraction error in nonstationary time series, *Ann. Stat.*, 7, 1303–1320, doi:10.1214/aos/1176344848, 1979.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P.: *Numerical Recipes in Fortran 90. The Art of Parallel Scientific Computing*, 2nd Edition, Cambridge University Press, 1996.
- Reichenau, T. G. and Esser, G.: Is interannual fluctuation of atmospheric CO₂ dominated by combined effects of ENSO and volcanic aerosols?, *Global Biogeochem. Cy.*, 17, 5–1, 2003.
- Reinsch, C. H.: Smoothing by spline functions, *Numer. Math.*, 10, 177–183, doi:10.1007/bf02162161, 1967.
- Tans, P., Thoning, K. W., Elliot, W. P., and Conway, T. J.: Background atmospheric CO₂ from weekly flask samples at Barrow, Alaska: Optimal signal recovery and error estimates, NOAA Technical Memorandum, ERL ARL-173, Air Resources Laboratory, Silver Spring, Maryland, USA, 112–123, 1989.
- Theodosiou, M.: Forecasting monthly and quarterly time series using STL decomposition, *Int. J. Forecast.*, 27, 1178–1195, doi:10.1016/j.ijforecast.2010.11.002, 2011.
- Thoning, K. W., Tans, P. P., and Komhyr, W. D.: Atmospheric carbon dioxide at Mauna Loa Observatory 2. Analysis of the NOAA GMCC data, 1974–1985, *J. Geophys. Res.-Atmos.*, 94, 8549–8565, doi:10.1029/JD09iD06p08549, 1989.
- Trivett, N. B. A., Higuchi, K., and Symington, S.: Trends and seasonal cycles of atmospheric CO₂ over Alert, Sable Island, and Cape St. James, as analyzed by forward stepwise multiple regression technique, NOAA Technical Memorandum ERL ARL-173, Air Resources Laboratory, Silver Spring, Maryland, USA, 27–42, 1989.
- WMO Greenhouse Gas Bulletin, no 10, *The State of Greenhouse Gases in the Atmosphere Based on Global Observations through 2013*. Geneva: World Meteorological Organization, 2014.