Modelling cycles and interdependence in irregularly sampled geophysical time series

Granville Tunnicliffe Wilson*[1], John Haywood[2], and Lynda Petherick[3]

[1]Department of Mathematics and Statistics, University of Lancaster, UK

[2]School of Mathematics and Statistics, Victoria University of Wellington, NZ

[3]School of Geography, Environment and Earth Sciences, Victoria University of Wellington, NZ

November 9, 2021

Abstract

We show how an autoregressive Gaussian process model incorporating a time scale coefficient can be used to represent irregularly sampled geophysical time series. Selection of this coefficient, together with the order of autoregression, provides flexibility of the model appropriate to the structure of the data. This leads to a valuable improvement in the identification of the periodicities within and dependence between such series, which arise frequently and are often acquired at some cost in time and effort. We carefully explain the modelling procedure and demonstrate its efficacy for identifying periodic behaviour in the context of an application to dust flux measurements from lake sediments in a region of subtropical eastern Australia. The model is further applied to the measurements of atmospheric carbon dioxide concentrations and temperature obtained from Antarctic ice cores. The model identifies periods in the glacial-interglacial cycles of these series that are associated with astronomical forcing, determines that they are causally related, and, by application to current measurements, confirms the prediction of climate warming.

*Correspondence Email:g.tunnicliffe-wilson@lancaster.ac.uk

Keywords: Irregular sampling, continuous time autoregressive model, multivariate time series, spectrum estimation, resolving cycles, causality

MSC subject classification: 60G25; 62M10; 62M20 JEL: C32; C51; C52; C53

1 Introduction

We show how a model presented in Tunnicliffe Wilson *et al.* (2015) can be applied to represent irregularly sampled time series. A common feature of the three series which we model is that they are obtained from cores which are sampled at regularly spaced depth. However, as a consequence of varying rates of deposition and compression, dating of the samples yields unequally spaced time points. We describe the measurements as irregularly sampled time series. Such geophysical series are common and the data collection generally costly in time and effort. Their analysis is much more challenging than that of regularly sampled time series for which there is long established, reliable and well understood methodology for both frequency domain analysis by spectrum estimation and time domain analysis by autoregressive models. The model we use enables the reliable estimation of similar characteristics for irregularly sampled series.

The model has the form of a continuous time linear autoregressive Gaussian process incorporating a time scale coefficient. A supplementary document to this paper supplies further background material relating to aspects of the model representation and estimation which are given in detail in Tunnicliffe Wilson *et al.* (2015). One example was presented in that work, an application to a pair of geophysical series described in Kirchner & Weil (2000), which investigates the relationship between fossil records of extinctions and originations of species and genera. That example was illustrative, with just 106 values in each series. The fitted model was of low order, 2, with just 8 coefficients and simple spectrum structure. The applications in this paper are much more challenging; there are over 600 points in the first example and several

thousand in the second. Further, the spectral structure of the series is more complex. The form of the autoregressive dependence of the model which incorporates the time scale coefficient provides the necessary flexibility for the efficient representation of this complex structure.

Turning to our data, the first example is of a record, described in Petherick *et al.* (2008), of dust flux in g $m^{-2}yr^{-1}$ (grams per square metre per year) derived from lake sediments in a region of subtropical eastern Australia, deposited over a time span from 40 cal. kyr. BP (40 thousand calendar years before the present time). Figure 1 shows a plot of this series which indicates the possibility of a periodic pattern with, what we shall call, three surges in level over the length of the record. We look at the evidence for interpreting these as a long period irregular cycle; modelling also reveals a very much higher frequency periodicity.

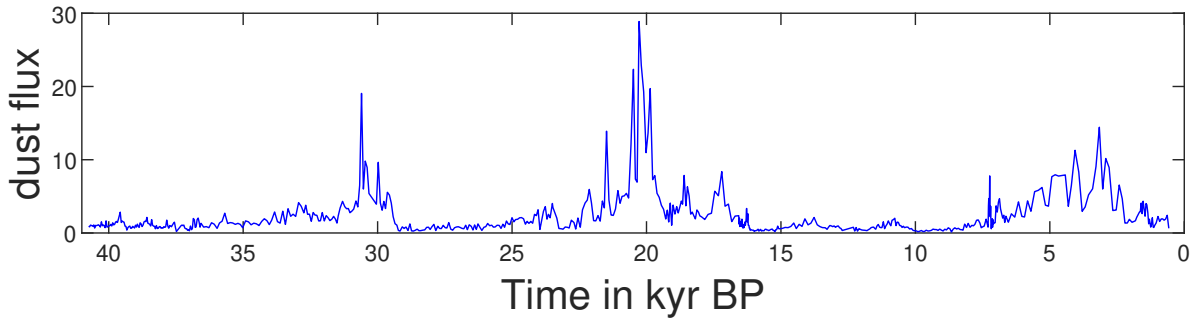

Figure 1: Dust flux g $m^{-2}yr^{-1}$ measured in lake sediments in a region of subtropical eastern Australia.

Our second example is to model a pair of series: the atmospheric carbon dioxide (CO_2) concentration in parts per million by volume (ppmv) and temperature anomaly in degrees Celsius (°C), derived from Antarctica ice core measurements as compiled in Lüthi *et al.* (2008) and associated with the time span from 800 kyr BP. Figure 2 shows plots of these series. They are quite evidently similar in their appearance, which follows the pattern of glacial-interglacial cycles. We demonstrate the application of our model to represent both the obvious long period cyclical patterns and the comparatively short term causal interdependence. By application to current measurements, the model confirms the prediction of climate warming associated with

increases in atmospheric CO_2.

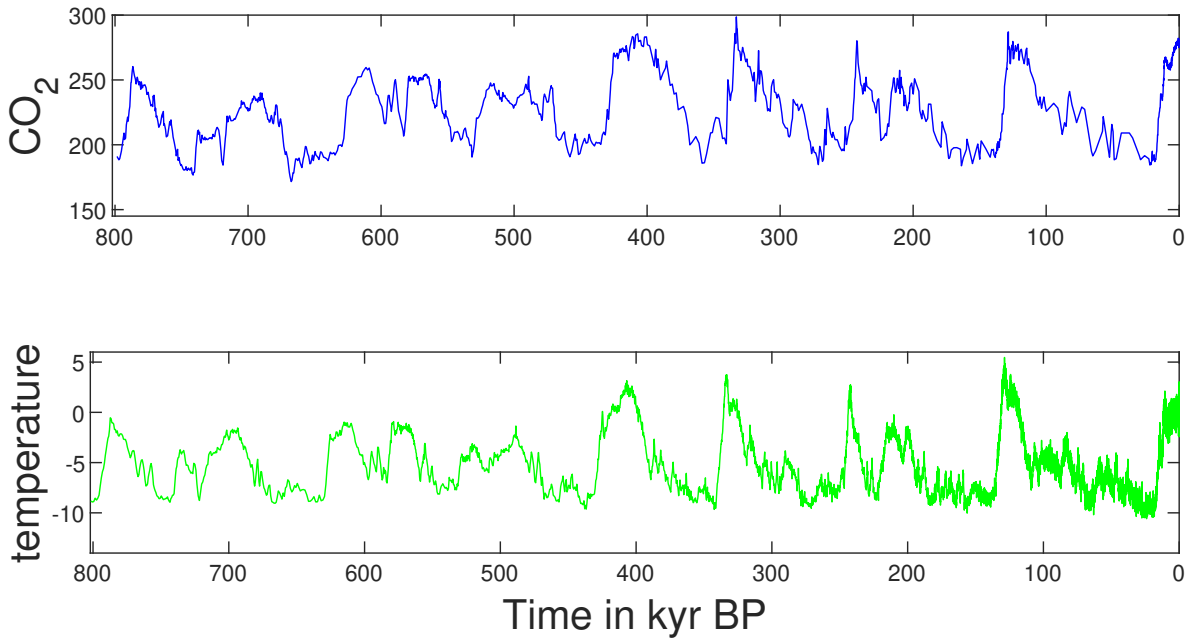

Figure 2: Antarctic ice core measurements of atmospheric carbon dioxide concentration ppmv and temperature °C.

The main contribution of this paper is methodological; the detailed account of the model building steps, the treatment of the novel modelling issues which arise with such challenging data and the statistical considerations which are important if a faithful representation and valid interpretation is to be made of the time series data. We also demonstrate the value of the model in providing answers to a range of questions about the series structure and properties. We believe this will be of value to readers and encourage them to apply the methodology in similar contexts.

In the next section of this paper we set out the model, then in the following two sections demonstrate modelling of the two sets of data. We use the first of these, the dust flux example, to explain at length our modelling procedure, including the likelihood criteria and plots that we use to select and specify our model, and to demonstrate the efficacy of our procedure by further analyses. In the second example we devote a shorter section to building a bivariate model and

displaying its properties. This is followed by a section which focuses on evidence for the causal dependence of each series upon the other. In connection with this we examine the predictive properties of the model on the geophysical time scale. We also illustrate its application on the shorter term, current, time scale. This shows that the model is consistent with presently accepted measures of the dependence of global temperatures on atmospheric carbon dioxide concentrations, and their extrapolation over coming decades.

A further concern regarding our model is independent validation of its ability to provide a faithful representation of the series structure. The methodology used in Kirchner & Weil (2000) was direct spectrum estimation of irregularly sampled time series by the Lomb-Scargle method, Lomb (1976), Scargle (1982, 1989), from which lagged cross-correlations between the series were derived. The model developed in Tunnicliffe Wilson *et al.* (2015) confirms the pattern of this cross-correlation. We have therefore considered the use of this direct spectrum estimation for more general validation of our model. We have, however, determined that for our first example the Lomb-Scargle method does not provide a good estimate of the series spectrum. We provide evidence for this in the supplementary document, along with other analyses to demonstrate that the model does indeed provide a faithful representation of the series structure.

For our second, bivariate, example we examine the dependence between the series. For regularly sampled multivariate time series this may be characterized using what is known as Granger causality, Granger (1980). This is generally determined by fitting vector autoregressive models to the series. In this paper we extend the methodology to our bivariate series example and also outline some of the issues relating to this characterization of causality in our supplementary document.

2 The time series model

2.1 The first order continuous time autoregressive model

As stated earlier, we use a linear stationary continuous time Gaussian process model to represent the series $x(t)$ of which we have observation $x_i = x(t_i)$ at observation time points t_i, for $i = 1,\ldots,n$. Such a model determines the covariances $\gamma_{j,k} = \mathrm{cov}(x_j, x_k)$ and therefore the well determined distribution of the observations. Models for spatial processes in two or more dimensions can be similarly used, by specifying a suitable covariance function, to determine the joint distribution of measurements at a discrete set of spatial points. However, for processes in the single dimension of time it is natural and of computational as well as analytic advantage, to specify the process in terms of its evolution from the past.

The simplest of such processes is the first order continuous time autoregressive, or CAR(1) process $x(t)$ with variance σ_x^2 and autocorrelation function $\mathrm{corr}[x(s),x(t)] = \exp(-\alpha|s-t|)$ for some $\alpha > 0$. The prediction of a future value conditional upon the present is given by

$$x(t+h) = \exp(-\alpha h)x(t) + e_h(t) \tag{1}$$

where the prediction error $e_h(t)$ has variance $[1-\exp(-2\alpha h)]\sigma_x^2$ and is independent of $x(t)$. This enables the joint distribution of the sample to be expressed in terms of the independent quantities $x(t_1)$ and $e_{h_i}(t_i) = x(t_{i+1}) - \exp(-\alpha h_i)x(t_i)$ for $i = 1,\ldots,n-1$, where $h_i = t_{i+1} - t_i$. This transformation, from $x(t_i)$ to $e_{h_i}(t_i)$, is known as the prediction error decomposition, and can be extended to all our models.

The CAR(1) process model is the differential limit of the prediction equation (1) as $h \to 0$:

$$dx(t) = -\alpha\, x(t)\, dt + \sigma_e\, dW(t), \tag{2}$$

where $\sigma_e^2 = 2\alpha\sigma_x^2$ and $W(t)$ is a Wiener process with incremental variance $\mathrm{Var}[W(t+dt) -$

$W(t)] = dt$. Our model is a direct generalization of this process.

2.2 A discrete basis for a continuous time process

The CAR(1) process defined by (2) is Markovian, the prediction of a future value depending only on the current value and not on any values previous to that. We now extend this process model to include dependence on these past values. We do this with terms taken from a discrete basis of the continuous range of past values; any linear function of the past can be expressed as a combination of these basis values. In the supplementary document we give reference to the derivation of this property, with illustrations, in the context of the prediction theory for continuous time processes. The basis is determined by a linear operator Z, which when applied to a stationary continuous time process $x(t)$ gives the process $y(t)$ which may be defined by either an integral or differential expression. The integral form is:

$$y(t) = Zx(t) = -x(t) + 2\kappa \int_{h=0}^{\infty} \exp(-\kappa h) x(t-h) dh. \tag{3}$$

The coefficient κ is an important (inverse) time scale parameter. Its value is initially chosen in any modelling application so that the exponential decay in the integral conforms with the appropriate time scale of past dependence in the series. It may then be adjusted to improve the predictive performance of the model. The operator Z is known as a generalized shift because when it is applied to a series which is a pure cycle of period $T \geq 1/\kappa$, it generates a cycle with the same amplitude and period with a time delay close to $2/\kappa$. Furthermore the process $y(t)$ has the same autocovariance function as $x(t)$ and generally has a similar appearance to a version of $x(t)$ with a time lag of $2/\kappa$. The differential form of the operator is

$$dy(t) = \kappa[x(t) - y(t)]dt - dx(t), \tag{4}$$

which is useful for defining a continuous time state space formulation of our model presented in

Tunnicliffe Wilson *et al.* (2015). This is reproduced in the supplementary document, together with an explanation of how it is used to fit the model to an observed time series.

The operator Z may be applied repeatedly to form $Z^k x(t)$ and for $k > 0$ these, together with $x(t) = Z^0 x(t)$, form a basis for the present and past of the process, $x(t-h)$ for $h \geq 0$. $Z^k x(t)$ may similarly be defined for $k < 0$, to extend the basis to future values of the process. This is a complete basis whatever the value of $\kappa > 0$. There is also an orthogonality property of the basis, that if $x(t)$ follows the CAR(1) process (2) with $\alpha = \kappa$ the variables $Z^k x(t)$ are uncorrelated at any time t. The value of κ actually used in an application is chosen for an efficient representation of the predictor over the span of relevant past values.

2.3 The continuous time Z operator autoregressive model of order p

Our model is to extend (2) by a finite span of elements of the basis, defined above, for past values of the process. The predictor of the increment is then:

$$dx(t) = \{\xi_1 x(t) + \xi_2 Z x(t) + \cdots + \xi_p Z^{p-1} x(t)\}\, dt + dW(t), \tag{5}$$

where the innovation term $dW(t)$ in (5) is the increment of a Wiener process $W(t)$ with $\mathrm{Var}\{dW(t)\} = V_e\, dt$. The notation we use here is that in Tunnicliffe Wilson *et al.* (2015).

This model may only be an approximation to the true structure of a stationary process, which would in theory require an infinite order p. In practice, how well, and with what order, an approximation may be achieved, will depend also upon κ. This determines the rate of decay of the weight applied to past values of $x(t)$. For lower values of κ the weight extends further into the past, so can pick up slower variations in the series; for higher values it picks up faster variations. We call the model the continuous time Z operator autoregressive (CZAR) model of order p. We shall sometimes write CZAR(p,κ) to emphasize the dependence of Z on κ.

Fitting this model to observed time series requires the determination or specification of both

p and κ, and estimation of the model coefficients. The continuous time model is fitted to the discrete, but irregularly sampled, data by integrating the state space form of the model over the variable length intervals between successive observations to enable prediction of each observation from past values. The likelihood of the model can then be formed from the prediction error decomposition which generates the orthogonal prediction errors and their standard errors. This, together with the model integration, is computed using state space methodology; details are given in the supplementary document. The coefficients are found which maximize the likelihood of the predictions on the Gaussian assumption, which is implicit in the Wiener process increments of the model. This assumption could be relaxed to allow $W(t)$ to have independent or orthogonal increments, resulting in corresponding properties of the prediction errors, whilst still using the Gaussian form of likelihood. Further investigation into the consistency and distribution of parameter estimates would then be required.

2.4 The multivariate model with observation noise

We extend the model (5) to multivariate time series of dimension m by taking $x(t)$ to be the vector of series elements at time t, applying the operator Z element-wise to $x(t)$, and the coefficients ξ_k to be $m \times m$ matrices. The disturbance term $dW(t)$ is the increment of an m dimensional vector Wiener process in which the term V_e of its incremental variance is a general $m \times m$ positive matrix. For simplicity we do not change the notation of (5) for the vector or matrix forms. However, for the multivariate form of the model we do also allow a structural parameterization of the dependence between the series by introducing a matrix coefficient on the left of (5) as $\xi_0\, dx(t)$ with unit diagonal elements of ξ_0. The incremental variance matrix V_e is then constrained to be a diagonal set of positive variances. For our bivariate example ξ_0 may be of upper or lower triangular form without loss of generality, with just one off-diagonal free coefficient. This is exactly equivalent to the non-structural form with unconstrained V_e.

Finally, we extend our model with what we will call observation noise terms. These arise

from a range of possible sources, such as measurement error or sample size variability. They are represented, when necessary, by an additive term, independent for each observation, resulting in an actual measurement of the form

$$y_i = x(t_i) + \varepsilon_i, \tag{6}$$

where $x(t)$ is the underlying, hidden, or latent process which is not directly observed. We assume that the ε_i have constant variance σ_ε^2, though changes through time could be modelled.

A completely uncorrelated, or white noise, continuous time process does not exist, though it may be approximated by a CAR(1) with a very high value of α, so that the autocorrelations decay very rapidly. If an irregular series is modelled without an observation noise term when one is in fact present, it may therefore still be well represented if the model terms allow a component of the autocorrelation decaying within the span of the intervals between observations. Such a component, if seen in the low lag model autocorrelations, can indicate that an observation noise term is required. This will be seen in our Antarctic ice core series example.

3 Modelling the dust flux series

In this section we first explain and illustrate how we investigate and select appropriate values of p and κ for fitting the CZAR model to the dust flux series. Our explanations carefully cover the likelihood principles and criteria that are applicable for this purpose. The selection of the coefficient κ adds a new dimension to the more usual situation for autoregressive models where the order p alone is to be selected. This leads us to select and compare two possible models with different values of both κ and p.

The comparison is largely based upon the displayed spectra of the fitted models. It would be useful to be able to validate these spectra by other, direct, methods, such as the widely used Lomb-Scargle spectrum estimate. However, as we show in the supplementary document, this

alternative produces dissimilar results to our modelling approach; we investigate why, and conclude that it is not a reliable alternative for the dust flux series. We do, however, present there a limited approach based on data binning, and a simulation exercise, which provide evidence that our models do faithfully represent the series structure.

3.1 The influence of the sampling times distribution on the choice of κ

Our first step in modelling this series is to examine the variability of the sampling times of the data. This is summarized in Figure 3 by two histograms. There are 663 measurements in this series and the first histogram shows how the sampling points are distributed over time, the second how the sampling intervals vary. The mean interval length is 60.7 years, though there is

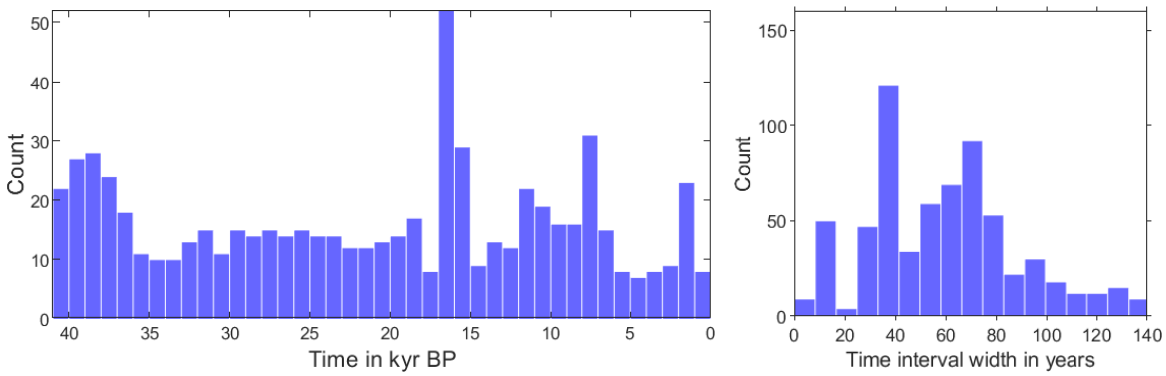

Figure 3: The distribution of sampling points over time, and of the width of the sampling intervals.

considerable variability around this, which emphasizes the need for a model which can address the irregular sampling. Beyond the upper limit on the histogram of intervals there are 6 further values, the largest being 414 years. The median interval length of 62 years is close to the mean.

The sampling interval distribution is important because it describes the information available for determining the range of frequencies which may be estimated by any procedure or model. For regularly sampled series the (fixed) sampling interval constrains the range of frequencies which can be uniquely identified to those less than one half the sampling rate r, the

reciprocal of the interval length. This upper limit is known as the Nyquist frequency. Any cycle in the series which is of higher frequency cannot be distinguished from, i.e. is aliased with, one which is of a lower frequency. For irregularly sampled time series there is no obvious such limit and Shapiro & Silverman (1960) show that sampling times realized from a Poisson process are alias-free. Publicly available software for Lomb-Scargle spectral estimation requires the user to specify an upper limit for the estimated frequency, set as a multiple of the Nyquist frequency derived as if the observations were equally spaced. For Poisson sampling times such a limit will exclude some higher frequencies. The flexible multi-parametric CZAR model (5) admits a spectrum which is not limited in that way; it extends without limit, decaying asymptotically as f^{-2} with increasing frequency f. Moreover, see Tunnicliffe Wilson *et al.* (2015), the coefficient κ provides a means of limiting the effective frequency range of the model spectrum. For a regularly sampled series this may be constrained to lie approximately within the range of the Nyquist frequency by requiring that $\kappa \leq r\pi/2$, where r is the sampling rate of the series.

If the series is irregularly sampled, we suggest using the mean sampling rate for r in this upper limit for κ, but if the sampling interval varies substantially it may be possible to reliably identify some cyclical components of higher frequency by choosing κ somewhat greater than $r\pi/2$. This argument applies to the dust flux series, and suggests using a value higher than that of $\kappa = 0.026$, which is derived from the mean sampling rate of $r = 0.0165$.

3.2 The selection of κ and p using likelihood criteria

The only visually evident possibility of a cycle in the logarithm of the series in Figure 4 is one of low frequency arising from just three unevenly spaced surges in level within the entire series range. A property of the model (5) is that it can accurately approximate any form of spectrum given any specified value of κ using a sufficiently high model order p. But allowing for a wide range of spectrum frequencies by specifying a high value of κ, would limit the ability of the model to resolve a low frequency cycle except by using a higher value of the model order. We

will approach this problem by fitting models using a range of values of κ.

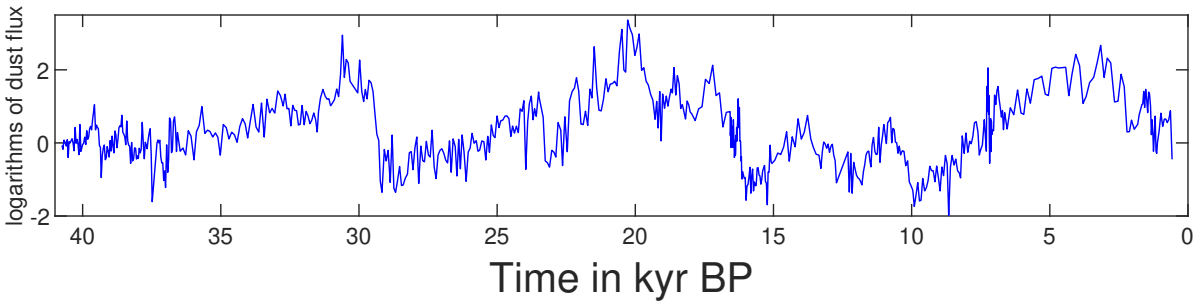

Figure 4: Logarithms of dust flux measurements derived from lake sediments in a region of subtropical eastern Australia.

Both the coefficient κ and the order p affect the fit of the estimated model (5) to the observed series, but in very different ways. As is generally the case in statistical modelling, the number of free parameters must be limited to avoid over-fitting. For (5) this entails limiting p. To do this we use maximum likelihood model estimation and its associated standard procedures of model testing, particularly likelihood ratio tests. We express these using the deviance, defined for a model of order k in terms of the likelihood L_k as $D_k = -2\log L_k$. In the supplementary document we explain how the deviance is related to the prediction error sum of squares of the fitted model. On the assumption that the true model of the series is CZAR with known κ and p, the reduction in deviance $D_p - D_q$ has a chi-square distribution on $q-p$ degrees of freedom for $q > p$. In practice we are aware that there may be no true finite order model and that p must be selected by some strategy applied to the deviance sequence D_k to produce a good approximating model. Different strategies have been proposed, the choice of which might appear to be subjective. The important point is that in large samples they select models which truly represent the structure of the series; in finite samples the effects of differently selected orders is a compromise between the lower resolution of the structure using a lower order, and the increased uncertainty of its estimated properties using a higher order.

Akaike (1973) introduced an information criterion (AIC) $D_k + 2k$, the minimum of which

determined p. Moreover, Shibata (1980, 1981) showed that for a process which was not a finite order autoregression, the AIC was in fact optimal for the estimation of its spectrum, although previously, Shibata (1976) had shown that it did not consistently estimate the true order p of a finite order model. Hannan & Quinn (1979) proposed the modified criterion $D_k + 2\log\log(n)\,k$, where n is the observed data set size, which does both consistently estimate a finite order model and the spectrum of an infinite order model. We will call this the HQIC.

There are, however, patterns in D_k that can lead to consideration of a higher order model than that determined by order selection criteria. To explain this we first present some background to order selection methods. Successively increasing the model order and testing each new coefficient for significance corresponds to the strategy of Box & Jenkins (1970) of selecting a cut-off point for the partial autocorrelation sequence in the context of discrete time autoregressive modelling. For our model it corresponds to successively testing the reduction $D_{k-1} - D_k$ for significance. Against this is the fact that, as k increases, eventually a false positive test will result if the significance level is held fixed. Again in the context of discrete time autoregressive modelling a plot for increasing order k of the mean residual sum of squares (MRSS), or estimated prediction error variance, was a valuable visual tool to indicate a good approximating model, by the point at which it leveled out. This is equivalent to looking for the leveling point of the bias corrected (residual) deviance or MCD defined as $D_k + k$. For $k > p$ and a model of finite order p, this is an unbiased estimate of the expected value of D_p. It is also useful for comparing models with different values of κ as well as the order p.

However, the subjectivity of the MRSS led to its replacement by the final prediction error, FPE, criterion, Akaike (1973), which estimates the error variance when the fitted model is used to predict an independent realization of a series of the same structure. There is then a further cost due to statistical uncertainties in the additional estimated model parameters as the order increases. The AIC is an information theoretic generalization of the FPE, but for Gaussian models it is equivalent to the FPE. It increases if the deviance reduction $D_{k-1} - D_k < 2$ at any

lag k. For a model of finite order p there is an expected increase of 1.0 in the AIC for successive increases in k beyond p.

However, in finite samples, associated with some seasonal or other periodic structure in the series, there may be an isolated and meaningful substantial reduction $D_{k-1} - D_k$ at some lag k, preceded by a run of values less than 2 over which the AIC steadily increases. If the reduction in AIC at lag k is not to the level of the previous minimum, the AIC will prefer a lower order.

We shall therefore use the AIC, $D+2p$, in our model building to indicate an approximating model order, together with the MCD, $D+p$ to indicate the comparative fit for different values of both κ and p. Also we shall be aware that a substantial reduction in either of these at a lag higher than that selected by the AIC, indicates an alternative approximating model order. Over-fitting by the selection of too high an order will tend to dilute the evidence for model features, for example by inflating standard errors of model spectra. But such features may not be evident at all if the model is under-fitted.

3.3 Likelihood function plots and model selection for the dust flux series

Before presenting our models for the dust flux series, we note once more that κ does not have the same role as the other model coefficients, but assists in this aim by allowing a more parsimonious model approximation with a lower order. Ideally we would generate our order selection plots for each value of κ over a fine grid covering the plausible range. However, generating each plot for a series of thousands of values with parameter numbers up to, in the case of a bivariate series model, well over 100, is currently not practical in its time requirement, though low level language re-coding of programs could reduce this substantially. We will therefore alternate between (i) plots over a finer range of κ values for a small number of values of the order, sufficiently high to plausibly provide a good approximation to the structure, and (ii) plots over increasing order for a small number of fixed values of κ indicated by the former plots. For our examples, for economy of space, we will show only a selection of such plots. Only the rel-

ative likelihood is important when comparing different statistical models, and correspondingly only differences in deviance values, so in the various plots we correct all deviance values by a fixed quantity chosen to simplify the y–axis scale.

For the dust flux series, the graph on the left of Figure 5 shows two plots of the MCD over the range $0.005 \leq \kappa \leq 0.050$, the upper limit being approximately twice that derived above from the mean sampling rate. The lower limit is simply chosen to be much less than this, by a factor of 10. The two plots are for the orders $p = 10$ and $p = 20$. The higher of these is specified, hopefully, to allow the model sufficient flexibility to fit any structure in the series, the lower is less flexible but is more parsimonious, and may be more efficient in its parametrization.

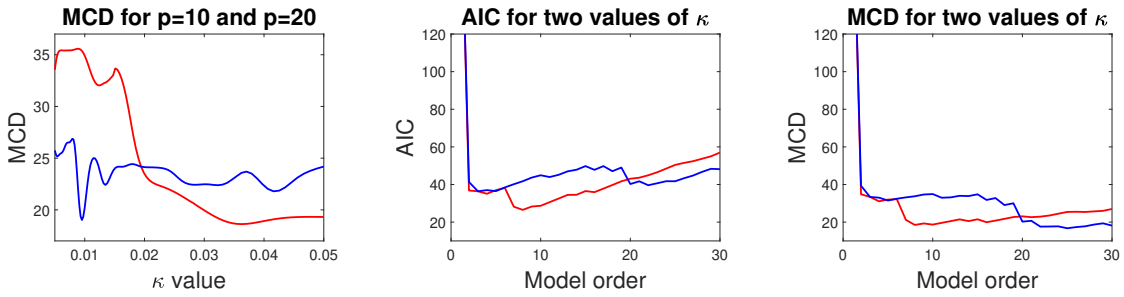

Figure 5: On the left, plots of the MCD of models for the dust flux series with the rate parameter over the range $0.005 \leq \kappa \leq 0.050$ and for model orders $p = 20$, in blue, and $p = 10$, in red. In the center, plots of the AIC for models of order up to $p = 30$, for $\kappa = 0.035$, in red and $\kappa = 0.010$ in blue. On the right, the same as the center, but with the AIC replaced by the MCD.

We see that for the model order $p = 10$ a high value of $\kappa \approx 0.035$ is preferred, whereas for $p = 20$ a narrow range around the lower value of $\kappa = 0.01$ is indicated, with minimum MCD very close in value to that for $p = 10$. Following these considerations, we move on to selecting a model order, other than the two used for this graph, by plotting the dependence of the AIC and MCD upon the model order for two preferred values of $\kappa = 0.035, 0.01$. From the graph in the center of Figure 5 the overall minimum of the AIC appears for $\kappa = 0.035$ at $p = 8$. For $\kappa = 0.01$ the minimum is at $p = 3$, very slightly lower than at $p = 5$. We note, however, a

substantial fall in the AIC for $\kappa = 0.01$ at $p = 20$, due to a reduction in deviance of 10.8 at that order, significant at the level 0.001. The MCD plots for the two values of κ in the graph on the right of the figure, confirms the evidence for considering the model of order $p = 20$ for $\kappa = 0.01$ as an alternative to that selected by the AIC. For both models we list the estimated parameters and their t values in the supplementary document. More illuminating for revealing the structure are the model spectra. However, we note that for the model with $\kappa = 0.035$ and $p = 8$ the t value of the penultimate coefficient ξ_7 is 4.28 and that of the last coefficient ξ_8 is -2.03, reflecting the prominent fall in the AIC at lag 7 and its minimum at lag 8. For the model with $\kappa = 0.01$ and $p = 20$ the t value of the final coefficient ξ_{20} is -3.47, supporting our argument above for the selection of this model.

3.4 Spectra of the fitted models

We show the model spectra in Figure 6. These are displayed on a log–log scale to permit the wide range of peaks to be better visualized, though the frequency axis is annotated with the corresponding periods. Throughout our graphical presentation of model properties we will show error limits to indicate uncertainty in the estimated quantities. Except where otherwise stated, these will be two standard error limits derived from the large sample variance matrix of the model parameter estimates.

The first of these estimated spectra has peaks only at high frequency, that at the higher frequency corresponding to an irregular cycle with a period of approximately 42 years. The error limits indicate that this is a true peak. Though of small amplitude compared with the range of variation of the series, this may be a signal reflecting known drivers of climate variability such as the Pacific Decadal Oscillation, Mantua & Hare (2002). At low frequencies, but very much higher amplitudes, no peak is seen, implying that there is no regularity to the three surges seen in the series. The second spectrum seen in Figure 6 shows peaks corresponding to a wide range of periods. The same high frequency peak is seen. But there is now a low frequency

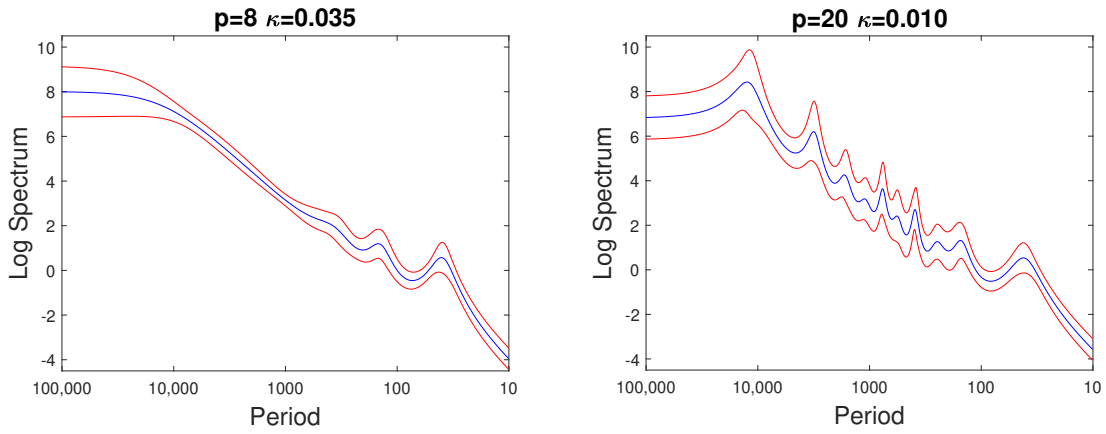

Figure 6: Plots of spectra for two models fitted to the flux series for different values of the rate parameter κ and model order p.

peak corresponding to a period of approximately 12500 years and reflecting the mean period between the surges seen in Figure 1. The spectrum peak is broad, reflecting the uneven spacing of the surges and consequent uncertainty in the period. If the surges were sporadic rather periodic, completely randomly in time, as points in a homogeneous Poisson process, we would not expect to see any evidence of a spectral peak, as in the spectrum on the left. A possible geophysical explanation is that the observed cyclicity is driven by the interactions between different orbital forcing cycles, Short *et al.* (1991), one of the shorter periods of these being close to 14000 years. The next peak in the spectrum corresponds to a period of approximately 3000 years, close to the third harmonic of the main peak. The peaks with periods between 12000 and 100 years are mostly of low amplitude compared with the error limits.

In summary, both models confirm a high frequency cycle and the second model provides evidence in favour of a low frequency periodicity. The evidence for this in the first model is limited by its use of the higher value of κ. The lower value of κ allows both peaks to be revealed but a consequence of the higher order is the introduction of intermediate smaller fluctuations in the spectrum. Overall, the evidence from these models for the high and low frequency spectral peaks is of value in the wider context of modelling similar data.

4 Modelling the CO_2 and temperature series relationship

Having described our procedure for order selection in the previous section, we limit ourselves here to presenting the plots on which we base our model selection of $\kappa = 0.0008$ and $p = 14$ for what will be an initial model of these series. From this model, however, we find strong evidence in the autocovariance functions of the need for observation noise terms. Inclusion of these terms in a further and final model is important to avoid mistaken inference about the causal relationships between the series, a point which we explain in the supplementary document. However, we postpone consideration of this causal relationship to a final section, and complete this section with graphs of the spectral properties of the model.

4.1 The sampling time distributions of the series

Each series consists of measurements of the variable of interest, together with the age of the core sample from which that measurement is derived. They are shown in Figure 2. Details of the compilation of these records are found in Lüthi *et al.* (2008) and references contained therein. The ages associated with the measurements of the two series are different apart from one coincidental value. The CO_2 series is compiled from four segments of measured values as explained in Lüthi *et al.* (2008) and consists of 1096 values. The temperature series consists of 5788 values. The spread of the measurements of the age and the distribution of the intervals between measurements are shown in Figures 7 and 8. Again, the histograms of intervals have some values beyond the limits shown; 6 values for the CO_2 series with the largest being 6029, and 94 values for the temperature series, with the largest being 1364, indicating some long gaps in the records. Thus the mean interval for the CO_2 series is 729 whilst the median is 581; the mean interval for the temperature series is 138.5 with the median being much lower at 58.1. The sampling of the series is clearly very irregular in time.

The interest in modelling these series is largely to discover the nature of the dependence

between them, which is expected to be strong, given the remarkable similarity of the movements in the two series. There is also clearly strong dependence within each series, in the form of what are recognized to be glacial-interglacial cycles. We therefore aim for a model which will represent well both the longer term pattern of these cycles, each spanning some 100 kyr, and short term behaviour on the scale of decades.

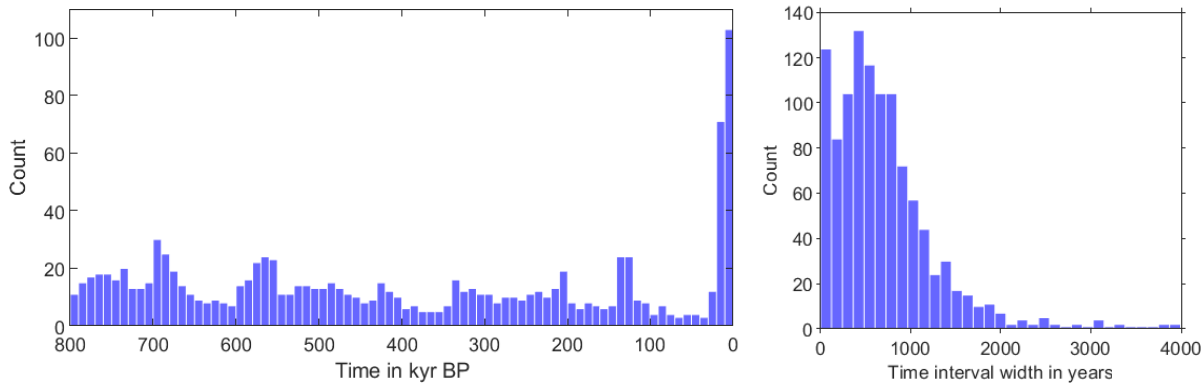

Figure 7: The distribution of sampling points over time, and of the width of the sampling intervals, for the CO_2 series.

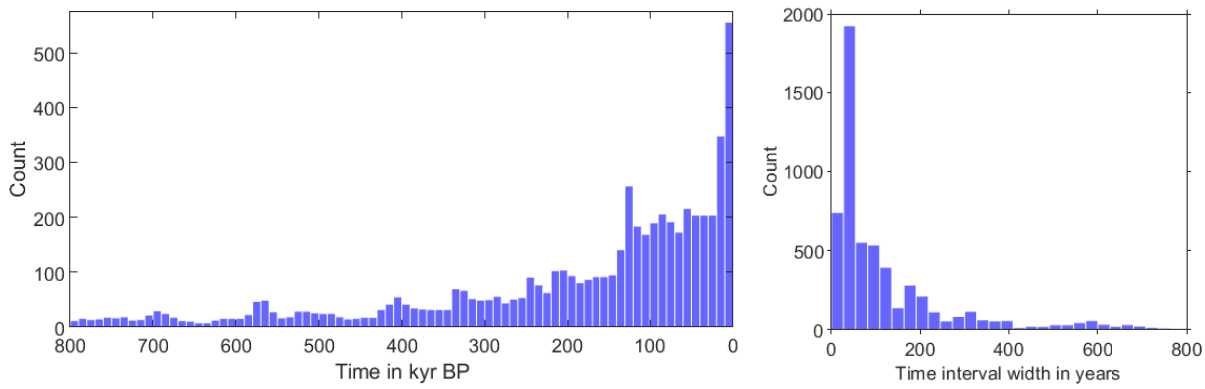

Figure 8: The distribution of sampling points over time, and of the width of the sampling intervals for the temperature series.

It is now widely considered that these cycles are substantially due to the influence on the earth's climate of variations of solar radiation (insolation) deriving from cyclical fluctuations in

the earth's orbit. This theory, due to Milutin Milankovitch, is the subject of extensive research, Short *et al.* (1991), Berger *et al.* (1984). The temperature series might therefore be considered to be a response to, or forced by, these variations in insolation. The periods of the major cyclical fluctuations are approximately 100, 41 and 21 kyr, though they are not purely sinusoidal. Ideally, perhaps, we should include these variations as a third, explanatory, series. The response is likely to be non-linear and distributed over an appreciable time lag, however, and introduces a level of complexity beyond the scope of our present models. In the supplementary document we show a simulation from our fitted model to demonstrate that it can reproduce cyclical behaviour similar to that in the observed data.

4.2 The initial model selection

We fit the structural bivariate form of the model (5) including the coefficient ξ_0 of $dx(t)$ with unit diagonals. We select, without loss of generality, the off-diagonal element to be that which allows dependence of temperature upon CO_2. This does not pre-judge the nature of any causal dependence, and in fact this coefficient proves to be small and not statistically significant, with t value -0.52, so is removed in the final model.

The mean sampling rate of the CO_2 series is 0.0014 and that of the temperature series is 0.0072, though using the median interval length these rates would be, respectively 0.0017 and 0.0172. The suggested upper limit on the value of κ to be considered in the model, is therefore determined by the CO_2 series as 0.0022, or 0.0027 using the median based value. We limit the upper range of κ to 0.0020 and display the results for $\kappa \geq 0.0003$.

We now show plots, in Figure 9 on which we base our consideration of two models. The graph on the left shows plots of the HQIC, Hannan & Quinn (1979), over κ for models of order $p = 14$ and 19. The important point here is that the minima of the two graphs are close, at respectively $\kappa = 0.0008, 0.0010$. These locations do not change with the choice of HQIC rather than AIC, only the offset of the plots. Our reason for choosing the HQIC is in the graph

on the right which shows both the AIC and the HQIC. These are for $\kappa = 0.0008$ but the plot, not shown, for $\kappa = 0.0010$ is very similar. The overall minimum of the AIC occurs, in fact, at $\kappa = 0.0010$ and $p = 19$ and for the HQIC at $\kappa = 0.0008$ and $p = 14$.

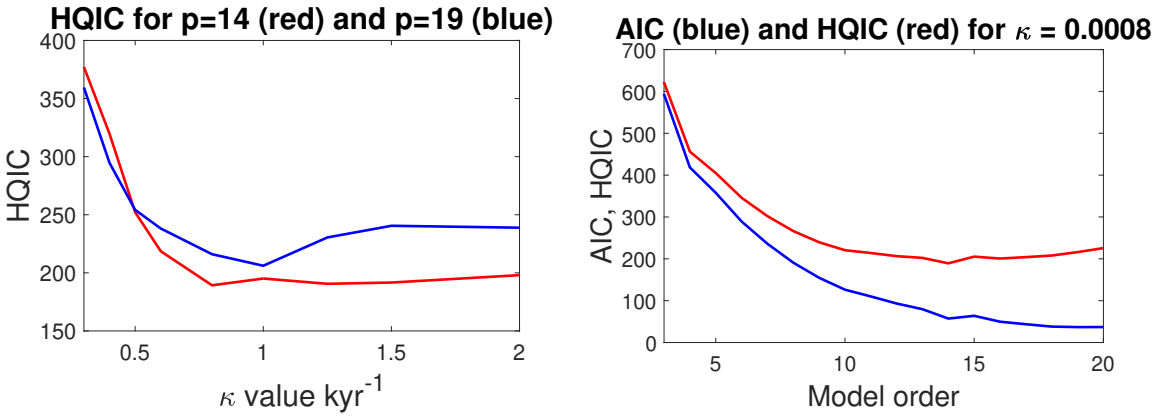

Figure 9: On the left, plots of the HQIC over κ for models of order $p = 14$ in red and $p = 19$ in blue. On the right, plots of the AIC in blue and the HQIC in red for models of order up to 20 and $\kappa = 0.0008$.

Our reason for preferring the lower order model, selected by the HQIC, is the evidence shown in Figure 10 of the estimated model autocovariances at very low lags. The sharp peaks around lag zero are characteristic of observation noise in the series. In the corresponding graphs, not shown, for the model of order 19, the peaks are very much narrower. The improvement in fit in the higher order model is due to further coefficients approximating this noise structure covariance. The use of 20 additional coefficients is not efficient for modelling these autocovariance peaks which can be represented by just two noise variance terms. By extrapolating the covariance curves back to lag zero beneath the peak, we obtain an approximate value of 4.0 for the observation noise variance of CO_2, giving a standard deviation of approximately 2 ppmv, and similarly a variance of 0.3 for the temperature, giving a standard deviation of 0.55 $^{\circ}$C.

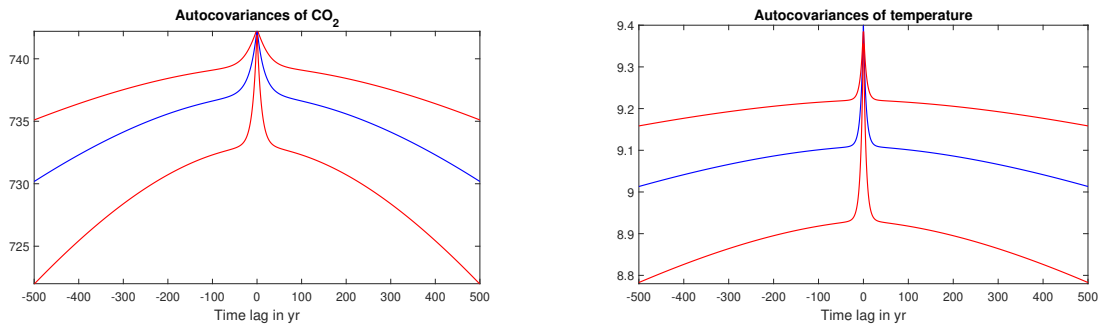

Figure 10: Estimated autocovariances of the CO_2 and temperature series at low lags to reveal evidence of observation noise.

4.3 A final model with observation noise terms and its spectral properties

Including observation noise terms in the model introduces non-linearity and possible singularities into the likelihood, so that fitting a model of the form (6) with these terms ab-initio commonly results in numerical estimation difficulties, particularly when fitting a range of model orders p and values of κ. These difficulties can be overcome by using an initial model without these terms, as in the previous section, to investigate the possible presence of observation noise and obtain good initial values of their variances. Using these, model estimation is much improved. We have considered a more direct method using the series variogram for detecting the presence of, and estimating the magnitude of, the observation noise. This would avoid the need for the initial modelling step. We investigate this approach in the supplementary document, but find that it is not effective for this data set.

We therefore re-fitted the model of order 14 starting with the same coefficients plus the noise terms of specified variances 3.60 for the CO_2 and 0.279 for the temperature series. On re-estimating this model these were little changed, with respective, highly significant, t values of 6.40 and 39.2. These terms do, however, improve the modelling efficiency in the sense that fewer coefficients are needed to achieve the same fit. Consequently, we found that a much improved AIC could be achieved using a final model of lower order $p = 10$ and $\kappa = 0.0003$.

The spectral properties of the noise-free structure of this model are shown, on an untrans-

formed frequency scale, in Figure 11. The spectrum of the temperature series, on the left, has pronounced peaks at frequencies of 0.010, 0.0244 and a less clear peak at 0.041 cycles/kyr. These correspond to respective periods of 100, 41 and 24 kyr which agree well with those predicted by the astronomical forcing theory of temperature variations. The spectrum of the CO_2 shares the prominent peak corresponding to the 100 kyr period cycle. The estimated coherency between the series is noteworthy because it remains very high throughout the displayed frequency range, and indeed remains significant, though much lower, up to very much higher frequencies, corresponding to periods less than 1000 years. We also show here the low lag cross-correlation between the series which peaks at a lag close to 700 years; the model estimates that on average the CO_2 series follows the temperature series by this lag. However, this must not be taken to imply that CO_2 is simply dependent on temperature, as we show when considering issues of causality in the next section.

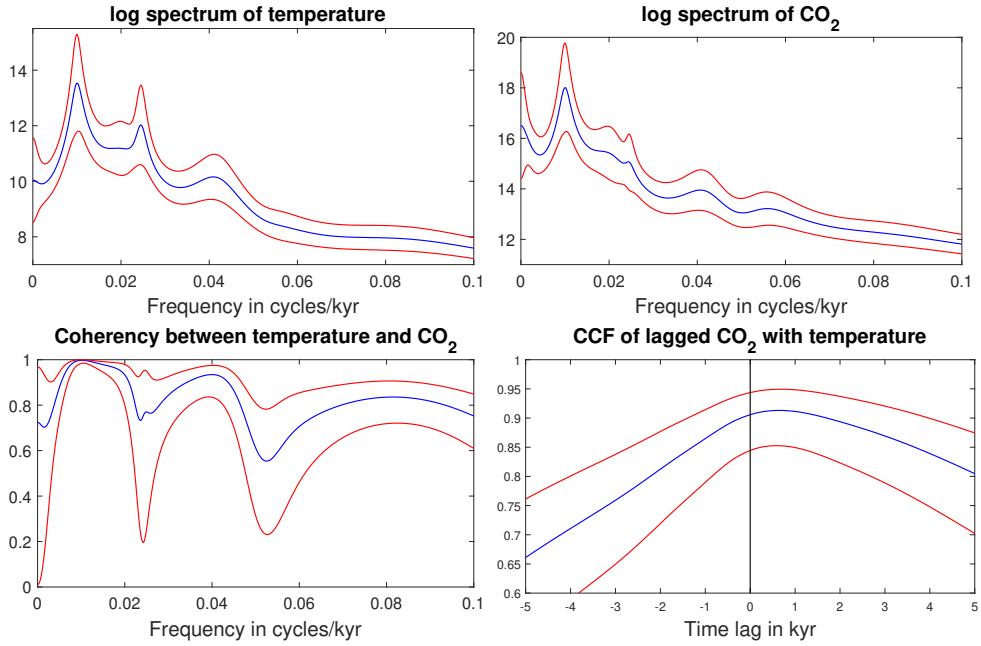

Figure 11: Estimated spectra of the noise free temperature and CO_2 series, their estimated coherency and low lag crosscorrelation, for the model fitted with observation noise terms.

5 Causal and predictive properties of the model

5.1 Causality testing

Using the prediction error decomposition of the likelihood, the model (6) has been selected and fitted so as best to predict future measurements from past observations. In consequence of this it is a stochastic model which well represents the evolution of the series over time, and in which the random disturbance term $dW(t)$ represents the effect of unpredictable fluctuations in the system on its future evolution. Without such disturbances the model would predict that the states of CO_2 and temperature would eventually settle to long term equilibria. Their presence allows us to interpret any significant future responses of each series to past perturbations of itself and the other series as causal, at least in a predictive sense. This is the sense in which the term is used in Granger causality, Granger (1980). This was developed in the context of vector autoregressive modelling of discrete time series. In the bivariate context it is not restricted to showing that one series is causal for the other. It allows the possibility that each series is causal for the other. It is based on testing whether past values of a second series can reduce the prediction error variance of the first, beyond that which can be achieved when past values of the second series are excluded as a predictor of the first.

In our context, we therefore test for causal dependence of CO_2 on temperature by refitting (6) omitting the 10 cross terms in the noise free form (5) which express the dependence of CO_2 on temperature. The difference in deviance between the unrestricted and restricted models was 159.3, which when referred to the chi-squared distribution on 10 degrees of freedom gives a tail probability of zero. The dependence is very highly significant. The corresponding exercise for the dependence of temperature on CO_2 gave a lower deviance difference of 40.5 but still with a very small tail probability of 0.000014. This is again highly significant and supports mechanisms which explain the causality of CO_2 for temperature given, for example, in Shakun *et al.* (2012). These refute the simple notion that temperature is solely causal for CO_2, which

has been suggested by the widely observed fact that CO_2 lags temperature, as reflected in the lagged cross-correlation of our Figure 11. Evidence from, and modelling of, the last glaciation shows global warming driven by increasing CO_2 concentrations.

5.2 Impulse response dependence

The mechanism of the dependence of each series on past values of itself and other series may be examined by an extension of the idea, widely used in the context of econometric modelling of discrete data, of plotting the model impulse response sequences. In that context the random disturbances are replaced by inputs of unit magnitude for just one time step, to generate the response sequences. In our continuous time context we will spread the impulse uniformly over a specified period, which we choose here to be 10 kyr. We will then examine the response over 200 kyr, so that it covers the length of about two of the main cycles. We will also choose an integrated input of 100 ppmv for the CO_2 disturbance and 1°C for the temperature series. The responses are generated using the state space form of the model.

Shown first, in Figure 12, are the responses of the two series to the temperature input. All the responses are those inferred by the model for the assumed underlying or hidden series before they are affected by the observation noise.

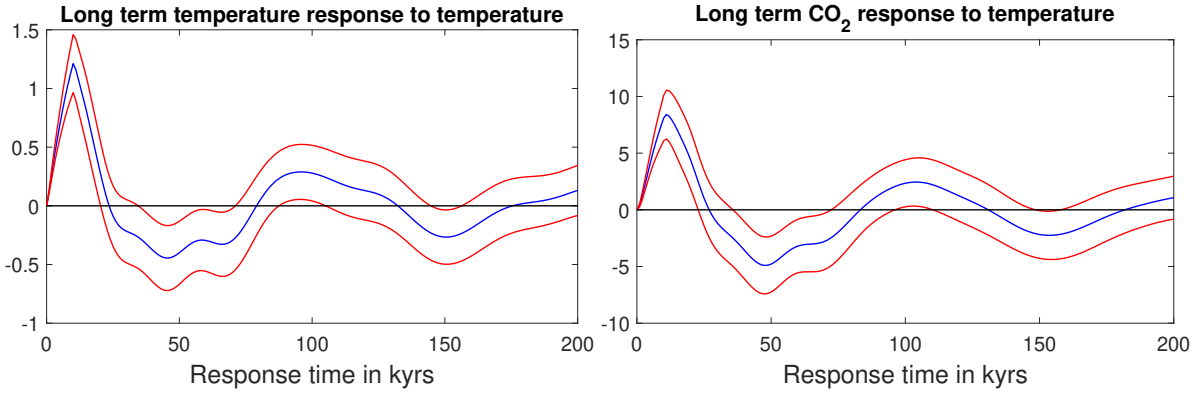

Figure 12: The response of the full model to a temperature disturbance input lasting 10 kyr. On the left the response for the temperature series and on the right for the CO_2 series.

The temperature response peaks at just over 1°C at the end of the 10 kyr input, slightly amplified by the system from the total input. It is highly significant statistically, falling away over a similar period corresponding to dissipation of heat from the system. The longer term pattern of response is not so highly significant, but reflects the features of the model which extrapolate the 100 kyr cycle in the series. The response of CO_2 to the temperature series is again significant, peaking about 700 years after the 10 kyr input, before falling away.

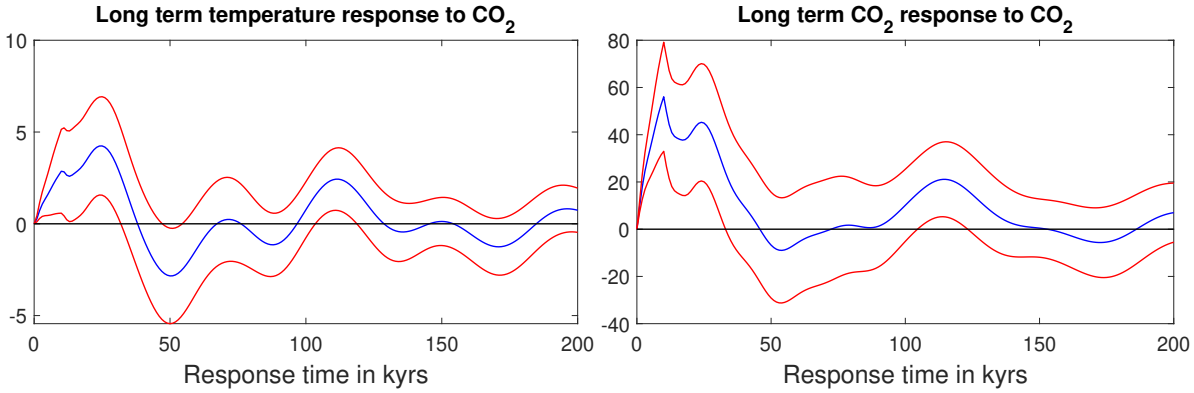

Figure 13: The response of the full model to a CO_2 disturbance input lasting 10 kyr. On the left the response for the temperature series and on the right for the CO_2 series.

Note that in Figure 13 the temperature response to the CO_2 input, though not so highly significant, is more extended. Following the 10 kyr input there is a slight peak then a rise to a higher peak after about 25kyr, before the response falls away fairly quickly. This shape of response may be due to the full system response to the further feedback from temperature. The 2.5°C rise in temperature after 10kyr might be expected, from Figure 12, to trigger a further substantial rise in CO_2, feeding back into the temperature response. At greater time lags the response is barely significant but in this case appears to reflect mainly the 41 kyr cyclical component of the series. The response of CO_2 to the total of 100 units of input CO_2 rises to somewhat less than 100, suggesting that a fraction of it is absorbed over the 10 kyr period.

In summary, all the responses are initially significant and positive and support the mechanisms proposed in Shakun *et al.* (2012). The stability of the whole system in the presence

of such positive feedback is due only to the natural decay of the responses. Heat is ultimately radiated and CO_2 absorbed by the land and oceans.

5.3 Application to recent and current time scales

We complete our investigation of the model by examining the pattern of temperature which it implies in response to the observed increase in CO_2 series from 1750 to 2019, chosen as beginning when CO_2 first appeared to be increasing beyond its previous general level. This response is projected to the year 2500 on the assumption of a hypothetical pattern of CO_2 levels beyond 2019. There is much discussion of possible future scenarios for CO_2 concentrations, and the one we show in the plot on the left of Figure 14 is an optimistic one, based on simple autoregressive modelling of earlier series values. It shows CO_2 returning to former levels in about 300 years. Such a rapid reduction requires net zero anthropogenic CO_2 emissions in the next few years to achieve a return to the historic equilibrium level of emission and re-absorption of CO_2 by the land and seas. Even so, the model predicts further rising temperature for well over a century and only a slow decline thereafter.

The projections are constructed by using the state space model form to simulate the process while constraining the CO_2 to follow its observed pattern from 1750. The response, as shown on the right of Figure 14, is obtained by omitting the disturbances affecting the temperature series. The error limits again derive from the uncertainties in the estimated model parameters, but indicate only one standard error. The observed CO_2, in black, and the possible future pattern, in red, are shown on the left. In Figure 15 we show, on the left, the graph of the measured actual global temperature series over the period 1850 to 2018 for comparison, as compiled by the Met Office Hadley Centre and Climatic Research Unit. On the right we show a typical sample from our model over this period, in which simulated disturbances affecting the temperature series are added in to the expected response.

The magnitude of the temperature rise predicted from the model is less than half that of the

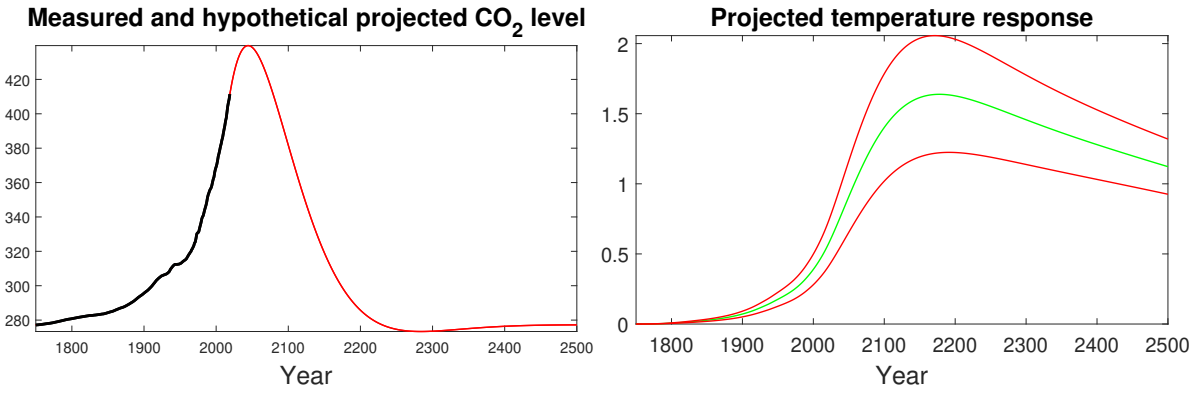

Figure 14: On the left, in black the measured CO_2 concentration from the year 1750 to 2019, extended in red with a hypothetical future pattern of the concentration. On the right the projected temperature response of the model to this rise and fall in CO_2 concentration, with one standard error limits, and relative to the level at 1750.

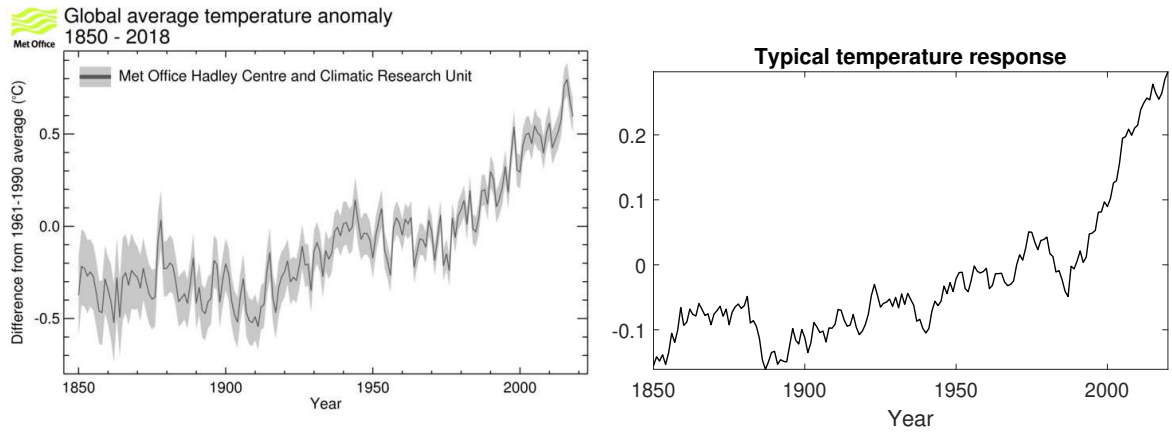

Figure 15: On the left, the measured temperatures from the year 1850 to 2018 © British Crown copyright, Met Office. On the right the temperature response of the model including typical disturbances, similarly shown as a departure from its mean level over the years 1961 to 1990.

rise in the actual global temperature series. It should be recalled that the model is now being applied to CO_2 concentrations well above those in the data to which the model was fitted. Also, our temperature data correspond to the Antarctic and rarely exceed 0°C. Furthermore, there are other greenhouse gases the effects of which have not been taken into account in our

model, and which may be more important in the present day. Different simulations of the typical response to that shown on the right of Figure 15 will show a variety of surges and pauses in growth similar to those evident in the observed series, which would appear to have two surges, besides two pauses, or even slight down-turns, in the late 1800s and around the 1950s. Very large volcanic eruptions close to the equator, which scatter persistent aerosols into the high atmosphere over the whole world are believed to be one cause of this. The eruption of Krakatoa in 1883 was massive and considered to have a global impact on temperature. The simulation shown exhibits a similar pattern and scale of variation, though more extensive simulations reveal rather greater variability is a property of our fitted model.

6 Conclusions

We have presented a class of continuous time autoregressive models and showed how they can be used to represent irregularly sampled time series, both univariate and bivariate. The models quantify the spectral properties of the observed series and the dependence between them. In our univariate example of the dust flux series we find evidence in the model spectrum that the three clear peaks in the series arose periodically rather than sporadically, and also found evidence of a multidecadal oscillation. In our bivariate example the model spectra revealed clearly the cyclical components with period approximately 100 kyr. Using observation noise terms in the model, we confirm causal dependence of each series upon the other and illustrate the mechanism of their dependence. The model qualitatively supports recent indications that raised levels of atmospheric carbon dioxide are driving a general increase in global temperature.

Acknowledgments

We thank the UK Met Office for granting permission for the re-use of the Global average temperature anomaly graph in Figure 15, under the terms of the Open Government Licence.

Data availability statement

The time series data modelled in this study are listed, together with their sources, in the supporting information for this article.

References

Akaike, H. 1973. A new look at statistical model identification. *IEEE Transactions on Automatic Control*, **AC–19**(2), 716–723.

Berger, A. L., Imbrie, J., Hays, J., Kukla, G., & Saltzman, B. (eds). 1984. *Milankovitch and Climate: Understanding the Response to Astronomical Forcing.* Netherlands: Springer.

Box, G. E. P., & Jenkins, G. M. 1970. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.

Granger, C. W. J. 1980. Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control*, **2**, 329–352.

Hannan, E. J., & Quinn, B. G. 1979. The determination of the order of an autoregression. *Journal of the Royal Statistical Society: Series B*, **41**, 190–195.

Kirchner, J. W., & Weil, A. 2000. Delayed biological recovery from extinctions throughout the fossil record. *Nature*, **404**, 177–180.

Lomb, N. R. 1976. Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, **39**, 447–462.

Lüthi, D., Le Floch, M., Bereiter, B., *et al.* . 2008. High-resolution carbon dioxide concentration record 650,000–800,000 years before present. *Nature*, **453**, 379–382.

Mantua, N. J., & Hare, S. R. 2002. The Pacific Decadal Oscillation. *Journal of Oceanography*, **58**, 35–44.

Petherick, L., McGowan, H., & Moss, P. 2008. Climate variability during the Last Glacial Maximum in eastern Australia: evidence of two stadials? *Journal of Quaternary Science*, **23**(8), 787–802.

Scargle, J. D. 1982. Studies in astronomical times series analysis II: statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, **263**, 835–853.

Scargle, J. D. 1989. Studies in astronomical time series analysis III: Fourier transforms, auto-correlation functions, and cross-correlation functions of unevenly spaced data. *The Astrophysical Journal*, **343**, 874–887.

Shakun, J., Clark, P., He, F., Marcott, S. A., Mix, A. C., Liu, Z., Otto-Bliesner, B., Schmittner, A., & Bard, E. 2012. Global warming preceded by increasing carbon dioxide concentrations during the last deglaciation. *Nature*, **484**, 49–54.

Shapiro, H. S., & Silverman, R. A. 1960. Alias-free sampling of random noise. *Journal of the Society for Industrial and Applied Mathematics*, **8**, 225–248.

Shibata, R. 1976. Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika*, **63**, 117–126.

Shibata, R. 1980. Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *The Annals of Statistics*, **8**, 147–164.

Shibata, R. 1981. An optimal autoregressive spectral estimate. *The Annals of Statistics*, **9**, 300–306.

Short, D., Mengel, J., Crowley, T., Hyde, W., & North, G. 1991. Filtering of Milankovitch cycles by Earth's geography. *Quaternary Research*, **35**(2), 157–173.

Tunnicliffe Wilson, G., Reale, M., & Haywood, J. 2015. *Models for Dependent Time Series*. New York, CRC Press.