

Some Statistical Issues in Palaeoclimatology[‡]

MIKE WEST

Institute of Statistics and Decision Sciences
Duke University, Durham, NC 27708-0251, USA

SUMMARY

This paper discusses some issues arising in the study of patterns of historical climate change based on geochemical measurements in deep lake sediment cores. The specific climatological focus is on patterns of variability, and possible cycles, in climatic indicators during the past two or three millenia. The statistical issues raised relate to various aspects of statistical calibration of raw data records, data quality and uncertainty propagation through data processing stages, and problems of time series analysis to assess patterns of variation in the geological records over time. Novel time series methods are introduced to address the questions of cyclicity in the data records and to allow for various sources of uncertainty in dating of the records. Dynamic models with cyclical auto-regressive components, and various aspects of their simulation based analyses, are discussed, and some exploratory analyses of currently available palaeoclimatological data are summarised. Other issues discussed include radiocarbon calibration, uncertainties in data timing, missing data, model combination, and the future prospects in the current and related palaeoclimatological contexts.

Keywords: Bayesian Spectral Analysis; Climate Change; Cyclical Time Series; Dynamic Linear Models; MCMC; Radiocarbon Dating; Random Observation Times

1 SOME SCIENTIFIC BACKGROUND AND DATA DESCRIPTION

Palaeoclimatology is *the study or investigation of the climates of periods in the geological past* (OED, 2nd Edn.). One current focus of much palaeoclimatological research is inference about global climate change based on historical records, particularly in connection with whether or not recent apparent trends in temperature or precipitation are unusual and perhaps attributable to anthropogenic cause. However, the study of past climate change on time scales as short as a few centuries is very challenging. Geochemical measurements made in sediment cores from lakes and oceans, and from ice cores and tree rings, provide the primary sources of relevant data, but variations in climate throughout the past few thousand years have been relatively mild so that signals in such proxy records are correspondingly subtle. Also, recorded data are typically very noisy,

[‡]Invited paper at the *Fifth Valencia International Meeting on Bayesian Statistics* in Alicante, Spain, June 5th to 10th 1994. Refereed paper published, with discussion, in 1996, *Bayesian Statistics 5*, (eds. J.O. Berger, J.M. Bernardo, A.P. Dawid and A.F.M. Smith), Oxford University Press, 461-486. The author won the international competition for the *1994 Mitchell Prize* for “the Bayesian analysis of a substantive and concrete problem” appearing in this paper.

and current age dating technology is of rather limited precision. Just how informative existing data may be requires assessment and evaluation of processes of data collection and calibration, the explicit tracking of uncertainties through such processes, and explicit recognition of all sources of uncertainty in formal models. This paper discusses some of these issues.

Geologists at Duke and other universities are attempting to decipher palaeoclimate signals archived in the sediments of large lakes in the East African Rift Valley, notably lakes Turkana, Tanganyika and Malawi. These lakes are unusually sensitive to minor climatic changes, and have experienced dramatic lake-level changes during the past 10–15ka (thousands of years); see Butzer (1972), Johnson *et al* (1987), and Owen *et al* (1982). Lake Turkana in Kenya, the northern-most of the large lakes, has experienced level changes of up to, and possibly more than, $\pm 60\text{m}$ from its present level. Lake level variations are accompanied by equally impressive changes in water chemistry and lake floor sediment composition (Halfman and Johnson 1988; Johnson *et al* 1991). Coupled with this, sedimentation rates in the rift lakes are fast and there is only very minor biological mixing of the sediments due to the scarcity of lake-bed organisms (Johnson 1984). Consequently, the sedimentological record from these lakes is in principle rather informative relative to other sources.

Most of our data derives from piston cores from Lake Turkana collected in 1984. These cores are each about 12 metres long, stretching from the present time at the top to around 3000 years BP (before present, with present conventionally set at 1955AD) at 12 metres down. Geochemical measurements made at intervals down the cores include per cent (by weight) calcium carbonate and stable carbon and oxygen isotope ratios, whose variations down cores may reflect lake level and incident river flow history, and hence rainfall patterns, during this period. Such data have been analysed to assess possible cyclical patterns so related to climatic changes. Here we restrict attention to carbonate abundance, based on data recently reported by Halfman, Johnson and Finney (1994). The levels of CaCO_3 present at depth in a core may depend on several factors, including influx of carbonates and non-carbonates from rivers or wind, aquatic organisms that may produce carbonate shells that become buried in the sediments, or inorganic precipitation of carbonates in the water column, driven by water chemistry, temperature and, at times, biological influences on water chemistry. In the case of Lake Turkana, carbonate abundance in the sediments has been attributed primarily to variations in rainfall and evaporation in the lake basin, with high carbonate abundance corresponding to relatively dry conditions (Halfman and Johnson, 1988; Halfman, Johnson and Finney, 1994).

Related data concerns radiocarbon dates at chosen depths in sediment cores. Even the very best methods of radiocarbon dating are, however, rather imprecise. Note also that the top of each core does not directly date to the present as the piston coring equipment typically significantly over-penetrates the lake bed, so that “core-top” is typically up to a few hundred years BP. Hence the radiocarbon estimates provide, for this and other current work, essentially the only currently available approach to age assignment on this time scale (i.e. centuries to millenia). Understanding the limitations imposed on any kind of geochemical time series analysis by the significant uncertainties about the depth-to-radiocarbon datings emerges as a major statistical issue.

This article discusses some of these issues in developing time series analyses of the CaCO_3 records. This is a rather preliminary and broad-brush exploration of a single, composite data series, focusing on various approaches to cyclical time series analysis and how such analyses may be tailored to account for the effects of timing uncertainties such as arise in this context. Simulation based exploration of very high-dimensional posterior distributions for latent time series components and their parameters, and for uncertain observation timings, are central to the development.

Carbonate data are displayed in Figure 1. Measurements are made at horizons down the cores ranging from close to core-top to about 1040cm down core; spacings between consecutive samples are variable, ranging between 0.1 and 10cm but concentrated in the 1-5cm spacing range. Note the “bunching” of observations of nominal ages 500-700 years, corresponding to more finely sampled upper core locations. In order to develop models for assessment of patterns of variation in the CaCO_3 data in actual calendar time, we need to “calibrate” depth in cms to calendar years BP; an appropriate “calibration curve” would account for the non-linearities in the sedimentation processes over time and the nature of possible distortions of the depth measurements due to the core extraction and subsequent analysis. Currently, and for our purposes here, the main source of information available is radiocarbon dating. A fairly large collection of radiocarbon dates has been made at various depth horizons in several cores (Halfman, Johnson and Finney, 1994). Sample depths within cores are measured quite accurately, to within a cm or so, but the traditional laboratory estimates of standard deviations of the radiocarbon dates are significant with typical values in the 40-70 years range; the datings are also clearly contaminated with gross outliers. Figure 6 displays, as the full line, a piece-wise linear map of depth to radiocarbon years BP derived from these data and developed in Halfman, Johnson and Finney (1994). Note that this calibration curve identifies the core top at around 670 years BP, and 12m down core at around 2600 years BP (which contrasts to previous works that had assumed 0 years BP at core top and ranging up to 4000 years BP at 12m down core). In the current paper we use this curve as the basis of a prior distribution for radiocarbon dates given depth measurements, and later assess effects of uncertainty about it; the map simply represents a consensus (the development involved several palaeoclimatologists) summary of expert opinion about the depth:radiocarbon profile.

Calibration of depth to radiocarbon years must be followed by a subsequent map to actual calendar years. Here we follow Buck, Litton and Smith (1992) who presented an appropriate Bayesian approach to computing posterior distributions for calendar dates of collections of radiocarbon dated samples; this is based on the high-precision, piece-wise linear calibration curve of Pearson and Stuiver (1986), and Stuiver and Pearson (1986). More on this in Section 3.

Carbonate data displayed in Figure 1 are plotted against a “nominal” time scale in years BP, moving *back* in time as we move along the time axis (i.e. down the core). Thus 0 years BP represents 1955AD and this paper is being presented at the point -39 years BP. This “nominal” time scale is a direct linear map from depth down core, $\text{time} = 675 + 0.7 \times \text{depth}$ (with depth in cms and time in years BP). This linear map, displayed as the dashed line in Figure 6, is a simple approximation to the crooked curve just discussed, used for initial, very exploratory investigation. This also provides contact with previous work in the area, almost all of which has been based on some kind of linear depth:time map, reading time as actual calendar time and implicitly assuming constant sedimentation rates and negligible distortion of the depth measurements due to core extraction and experimentation. Note the apparent decreasing trend in CaCO_3 up to more recent times; this is seen in cores from all over Lake Turkana and reflects wetter conditions in more recent times. Wetter conditions lead to decreased lake salinity which tends to reduce calcite precipitation from the water (the primary source of the carbonate), and also dilutes the carbonate by river-derived, non-carbonate sediment. This acyclic trend must be incorporated in any model attempting to assess trend-free cyclical variations.

Recorded CaCO_3 levels from these and other cores range from 1 to 25% of the dry sediment, and repeat geochemical analyses at selected core locations suggest approximate “reproducibility” of about 0.25%, which may be taken as a crude initial estimate of measurement error standard deviation. The data in Figure 1 are combined measurements from two cores taken at different locations in the lake, cores LT84-7P and

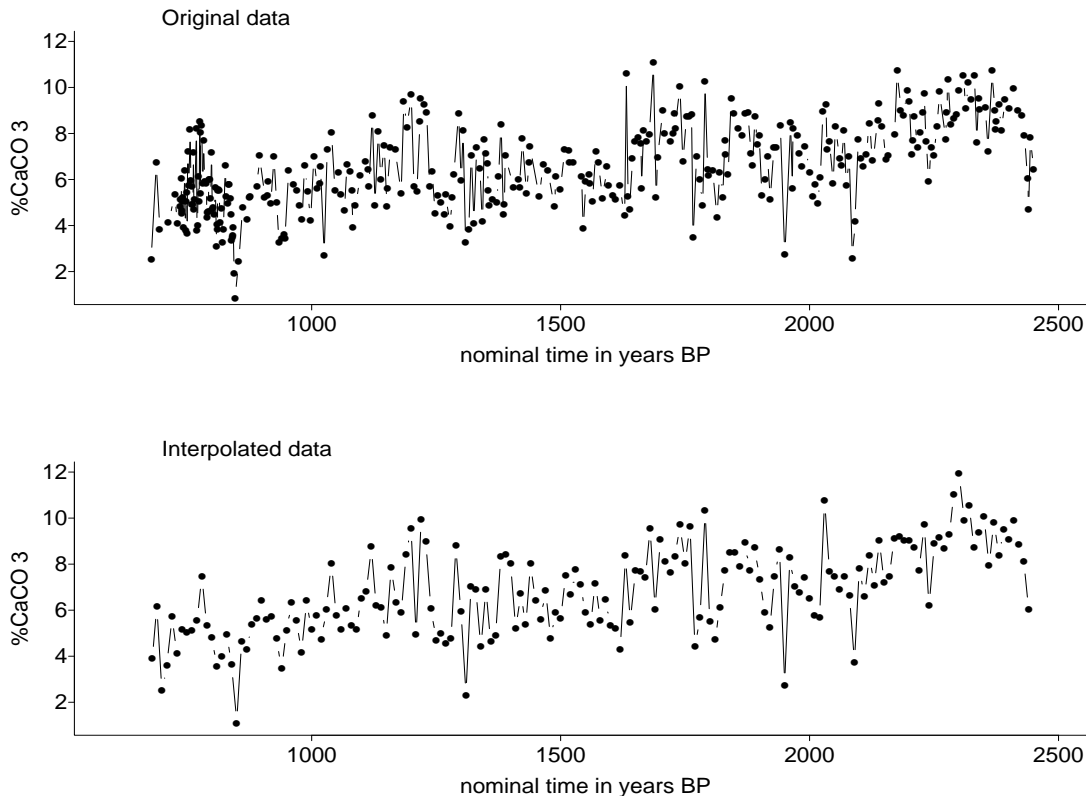


Figure 1. CaCO_3 time series plotted against nominal time (the nominal time derived from the linear map from depth to years BP discussed in text). The upper frame displays actual carbonate records. The lower frame displays 177 interpolated values obtained by fitting a cubic spline interpolant to the depth:carbonate data, and then evaluating the interpolant at equal intervals of 10cm down core. There are 350 values in the original series which, after mapping to nominal time, produce several groups of consecutive dates less than a year apart; these dates are truncated to nearest year and then common dates averaged to produce 276 observations with distinct dates prior to interpolation.

LT84-8P; the sample depths for the two cores have been stratigraphically correlated (by Halfman, Johnson and Finney, 1994) on the basis of down-core variability in carbonate abundance and magnetic susceptibility, two parameters that are independent of one another, at least in Lake Turkana. This provides a match on the depth axis to combine the CaCO_3 measurements. Here we assume from the outset that this process produces a perfect match, so that the raw data we begin with are just the resulting depth: CaCO_3 observations used to produce the graph in the upper frame of Figure 1.

The lower frame in Figure 1 displays interpolated CaCO_3 values obtained by fitting a cubic spline interpolant to the original depth:carbonate data, and then evaluating the interpolant at equal intervals of 10cm down core. On depth and hence the displayed nominal time scale, this therefore produces an interpolated series of equally spaced observations. Cyclical time series analysis in geology generally is predicated on the use of the fast Fourier Transform to compute periodogram estimates of spectral densities, and this requires equally spaced data. Again, we use this interpolated series in initial exploratory analysis to contact such previous work.

2 SOME TIME SERIES ISSUES, MODELS AND METHODS

Assessment of possible cyclical patterns in geochemical records involves some kind of time series analysis; some elements of spectral analysis and cyclical dynamic linear modelling are covered here. Assume that measurements y_i are available at known times t_i , ($i = 1, 2, \dots, n$); for us, the data y_i represent the measured CaCO_3 levels, and the t_i are assumed known *calendar* times (in years BP) corresponding to the measurement locations down core. Write $Y = \{y_1, \dots, y_n\}$ for the observed data.

2.1 Spectral Estimation

Most current practice is based on traditional spectrum estimation, via fast Fourier transforms, with associated (non-Bayesian) methods of periodogram smoothing. This is our starting point. In establishing the current project directions, the geologists were (rightly) concerned about the difficulties of assessing the relevance of periodogram based inferences, the extent of noise in spectral density estimates, problems of smoothing, uncertainties about how to handle trends in the data, among other things. Bretthorst (1987) has illuminated the essential correspondence between traditional, non-parametric spectral analysis and parametric inference in the simple model $y_i = a \cos(2\pi t_i/\lambda) + b \sin(2\pi t_i/\lambda) + \omega_i$ where the innovations ω_i are assumedly iid $N(\omega_i|0, w)$ with constant variance w . This single cycle model has amplitude $r = \sqrt{a^2 + b^2}$ and wavelength λ . Under a reference prior for a, b and w , the induced marginal log-likelihood function $\log(p(Y|\lambda))$ is intimately related to the traditional, periodogram based spectrum estimate. In problems with large n , $\log(p(Y|\lambda))$ is essentially an increasing function of the traditional periodogram, though more typically there are major differences (Bretthorst 1987, chapters 2 and 3). Issues of interpretation of spectral peaks, smoothing periodogram estimates, etc, are now replaced by direct inspection of this function – just the log posterior for the wavelength λ under a proper uniform prior over a finite range – as an initial, exploratory tool. West (1995) provides further discussion and examples.

The relevant calculations, laid out by Bretthorst, have been performed for both the original and interpolated CaCO_3 series of Figure 1. Results appear in Figure 2. In each case a crude trend was subtracted first (simply based on a lowess interpolation that produces a trend “estimate” closely similar to the more formal posterior estimates of acyclical trend in further analyses below). Analysis without some kind of detrending, however crude, naturally leads to “modes” in the marginal likelihood function at high values, here in several hundreds of years, that represent low frequency, acyclic trend rather than longer term cycles. Note first the close correspondence between results based on original and interpolated data. On the log scale, the oscillations of the posterior are quite apparent and dramatic for smaller wavelengths; this is theoretically predicted behaviour, and the oscillations can be expected to be more marked in cases where the sinusoidal signal is dominated by noise, i.e. when w/r^2 is large.

The posterior evidences a global mode at about 155 years BP and subsidiary minor modes near 110 years, 140 years, 90 years, 75 years and on down into the lower values where the increasingly minor peaks are noise driven. Subtracting other trend estimates shifts the modes around a little, the main effect being the enhancement of minor modes when the detrending really oversmooths; otherwise, the modes near 150 and 110 are quite stable, the indication of some action near 155 years BP reasonably strong. How strong may be evaluated via inference on the amplitude and other waveform characteristics; this is done below in the context of an elaboration of the basic models that permits more appropriate and realistic accounting for cyclical variability.

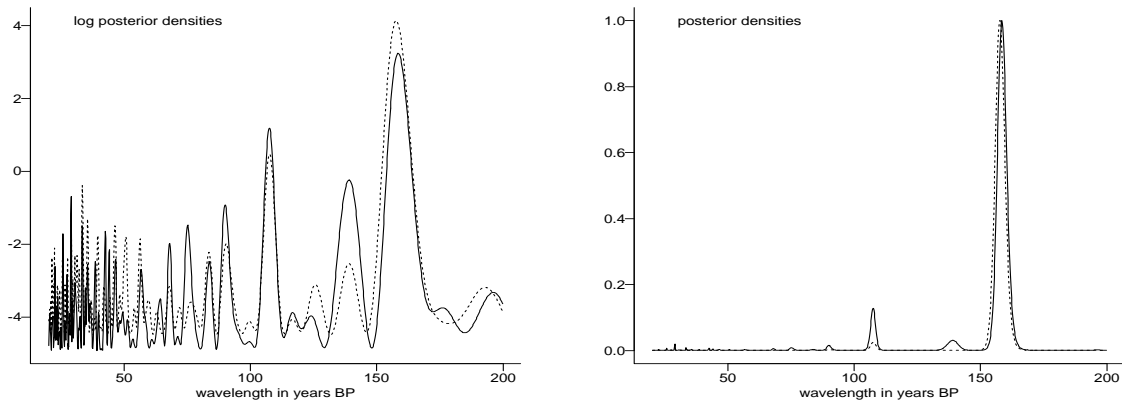


Figure 2. Reference posteriors for single wavelength in Bayesian “spectral” analysis of actual (full lines) and interpolated (broken lines) carbonate series (note that the posteriors have been standardised to unit maxima, rather than normalised) on the nominal time scale. Each series was “detrended” by subtracting a lowess smooth prior to analysis; similar results are obtained with other, perhaps more formal approaches to detrending, as in the modelling sections below.

2.2 Periodic Component Dynamic Linear Models

West (1995) has discussed using cyclical component dynamic linear models for inference about multiple frequencies in time series. Simple DLMs provide opportunity to investigate and estimate variation in amplitudes and phases of cycles of assumedly fixed (or even varying) wavelengths, and therefore to protect inference about wavelengths from biases induced in simpler, constant waveform models. DLMs can also be developed to incorporate appropriate and flexible trend descriptions (and beyond to regression and correlated noise terms, as in West and Harrison 1989, chapter 9). In addition, simulation methods of posterior analysis yield approximation to posterior distributions for all model characteristics, including latent cyclical processes, predicted behaviour, and arbitrary functions of defining model parameters. Further investigation of the CaCO_3 series is vested in these models in this paper. An initial analysis simply extends that of Section 2.1 to allow for possible time variation in the amplitude and phase characteristics of a single waveform, and models an underlying acyclic trend using a simple first-order polynomial DLM (a random walk, West and Harrison 1989, chapter 2).

The model incorporating trend plus possibly several cycles laid out in West (1995) is structured as follows. Assume an equally spaced time series x_t , $t = 1, 2, \dots$, and k distinct wavelengths $\lambda = \{\lambda_1, \dots, \lambda_k\}$; the model is

$$x_t = x_{t,0} + \sum_{j=1}^k x_{t,j} + \nu_t,$$

$$x_{t,0} = x_{t-1,0} + \omega_{t,0},$$

$$x_{t,j} = \beta_j x_{t-1,j} - x_{t-2,j} + \omega_{t,j}, \quad (j = 1, \dots, k),$$

with independent and mutually independent innovations sequences, $\omega_{t,j} \sim N(\omega_{t,j}|0, w_j)$, that are independent of the independent observation noise terms $\nu_t \sim N(\nu_t|0, v)$. The model is a component DLM with latent sub-series components:

- $x_{t,0}$, a “steady” acyclical trend modelled specifically as a first-order polynomial process (though others are possible, of course);
- $x_{t,j}$, ($j = 1, \dots, k$), a set of cyclical AR(2) components with time-varying amplitudes and phases (the

time variation “driven” by the noise “inputs” $\omega_{t,j}$) but fixed wavelengths λ_j defined by $\beta_j = 2 \cos(2\pi/\lambda_j)$ (West and Harrison 1989, section 9.4); clearly $0 < \beta_j < 2$ in this model.

The DLM representation is

$$\begin{aligned}x_t &= F' z_t + \nu_t \\z_t &= G z_{t-1} + \delta_t\end{aligned}$$

where

$$z_t = \begin{pmatrix} x_{t,0} \\ x_{t,1} \\ x_{t-1,1} \\ x_{t,2} \\ x_{t-1,2} \\ \vdots \\ x_{t,k} \\ x_{t-1,k} \end{pmatrix}, \quad F = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ \vdots \\ 1 \\ 0 \end{pmatrix}, \quad G = \begin{pmatrix} 1 & & & & & & & & \\ & \beta_1 & -1 & & & & & & \\ & 1 & 0 & & & & & & \\ & & & \beta_2 & -1 & & & & \\ & & & 1 & 0 & & & & \\ & & & & & \ddots & & & \\ & & & & & & \beta_k & -1 & \\ & & & & & & 1 & 0 \end{pmatrix}, \quad \delta_t = \begin{pmatrix} \omega_{t,0} \\ \omega_{t,1} \\ 0 \\ \omega_{t,2} \\ 0 \\ \vdots \\ \omega_{t,k} \\ 0 \end{pmatrix}.$$

Here G is block-diagonal so the “empty” off-diagonal entries are all zero; also $\delta_t \sim N(\delta_t|0, W)$ with diagonal variance matrix $W = \text{diag}(w_0; w_1, 0; w_2, 0; \dots; w_k, 0)$.

Gibbs sampling methods for DLMs, as introduced by Carter and Kohn (1994), and Frühwirth-Schnatter (1994), are extended in West (1995) to provide posterior distributions for the parameters of matrices such as G and the variances of all error terms, together with samples from posteriors of the latent time series components in z_t for all time t . Note that we are working in a fairly high dimensional space here as the entire collection of component time series $x_{t,j}$ are “latent” or uncertain. Full details, with examples, are given in West (1995). Note further that this represents an example of inference about parameters of system evolution matrices and variance components in a much larger class of models.

This model is fitted to the equally spaced, interpolated carbonate series in a first analysis here assuming just $k = 1$ cycle. Prior distributions in this analysis assume independence, zero means and large variance (10^4) for the components of the initial state vector z_0 , and a uniform prior for the transformed wavelength parameter β_1 over a range transforming to 10-500 years for $10\lambda_1$ (recall that, as the interpolated data are in decades, so is λ_1 .) The benchmark s.d. estimate of 0.25 for CaCO_3 readings is coupled with a low degrees of freedom, just 2 here, to define an inverse gamma prior for the observational variance v ; this determines a gamma prior for precision $1/v$, with shape 2.5 and scale 0.0625. Incremental changes in the acyclic trend and AR(2) cycle terms are expected to be much smaller by comparison, modelled initially via independent inverse gamma priors for the (assumedly constant) variance components w_j ($j = 0, 1$); the precision in each case has a gamma prior with shape 2.5 and scale 0.0125.

The simulation scheme of West (1995) was run to burn-in for (a wasteful) 20,000 initial iterations, then a sample of size 16,000 posterior draws saved, skipping 50 iterations between draws to break correlations (confirmed by subsequent data analysis). Saving draws produces posterior samples of the entire sequences $x_{t,j}$, ($j = 0, 1$), for all $t = 0, 1, \dots, n = 177$ in this case, plus samples of λ_1 and each of the three variance components. Figure 3 is the resulting histogram approximation to the posterior for the single wavelength λ_1 (after multiplying by 10 to map to years) in this first analysis of the interpolated data. The modes near 110 years and 150-160 years correspond closely with the exploratory plots in Figure 2, though the posterior here is clearly rather more diffuse. The mode near 110 years is also much more prominent here. Several reanalyses using varying sample sizes and starting values produce the same picture.

As an aside here, note that starting iterations with very small values of λ_1 has and continues to lead to convergence problems with the simulation analysis; this is to be expected, as the lower tail of the posterior is

evidently highly multimodal, with many very minor modes of little mass, comparable to the oscillatory form in the “spectrum estimate” in Figure 2. The simulation run here did not wander down to that low density region, where it would have become “trapped” in oscillating between what are negligible modes. This is an area of current investigation from the viewpoint of refining sampling schemes to cope with such irregular behaviour, particularly since it arises more acutely in models with several cyclical components.

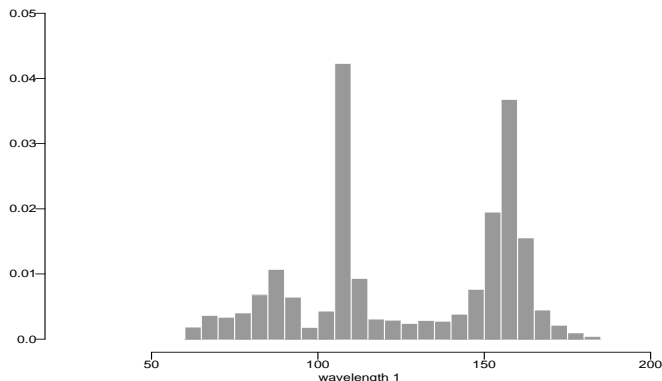


Figure 3. Posterior for nominal wavelength in reported analysis of the dynamic model with a single cycle plus smooth trend.

The indication of two wavelengths is quite apparent, so the next analysis extends to incorporate a second cyclical component, again following the developments in West (1995) to the letter. The same model applies but with $k = 2$, having similar priors extended in two ways: the same inverse gamma prior applies to the variance component w_2 of the second latent sub-series, and the prior for the two wavelengths is uniform on the β_j scale subject to $10 < 10\lambda_1 < 10\lambda_2 < 500$. Some summaries of the interpolated CaCO_3 analysis appear in Figures 4 and 5. Figure 5 plots posteriors for the two wavelengths (again after multiplying by 10 to map to years) confirming λ_2 in the region of 150-160 years and λ_1 near 100-110 years. Note the additional mass between 50-100 for the first wavelength; here the posterior has several minor modes and appears very “jagged” around smaller values, again consistent with the oscillatory nature of the tail of the “spectrum estimate” in Figure 2. Figure 4 presents some posterior trajectories—simply approximate posterior means of the three latent components $x_{t,j}$, ($j = 0, 1, 2$), over time—with bands representing point-wise single standard deviation limits and hence indicating uncertainty about the form (amplitude and phase) of the cyclical components. Note the clear non-linearity of the trend, the indication that the wavelength λ_1 contributes in only minor ways to the overall pattern of the data, and that the extent of contribution of the second, dominant cycle is rather small though apparently persistent. Note that the scale of the amplitude axes span the same range as the data for comparison.

Before proceeding further to assess the effects of timing uncertainties on these kinds of analyses, note that further exploration with $k = 3$ cycles produces an estimated component of lower wavelength, but the trajectory plot indicates the term is completely negligible, sitting around zero across time with appreciable uncertainty bands.

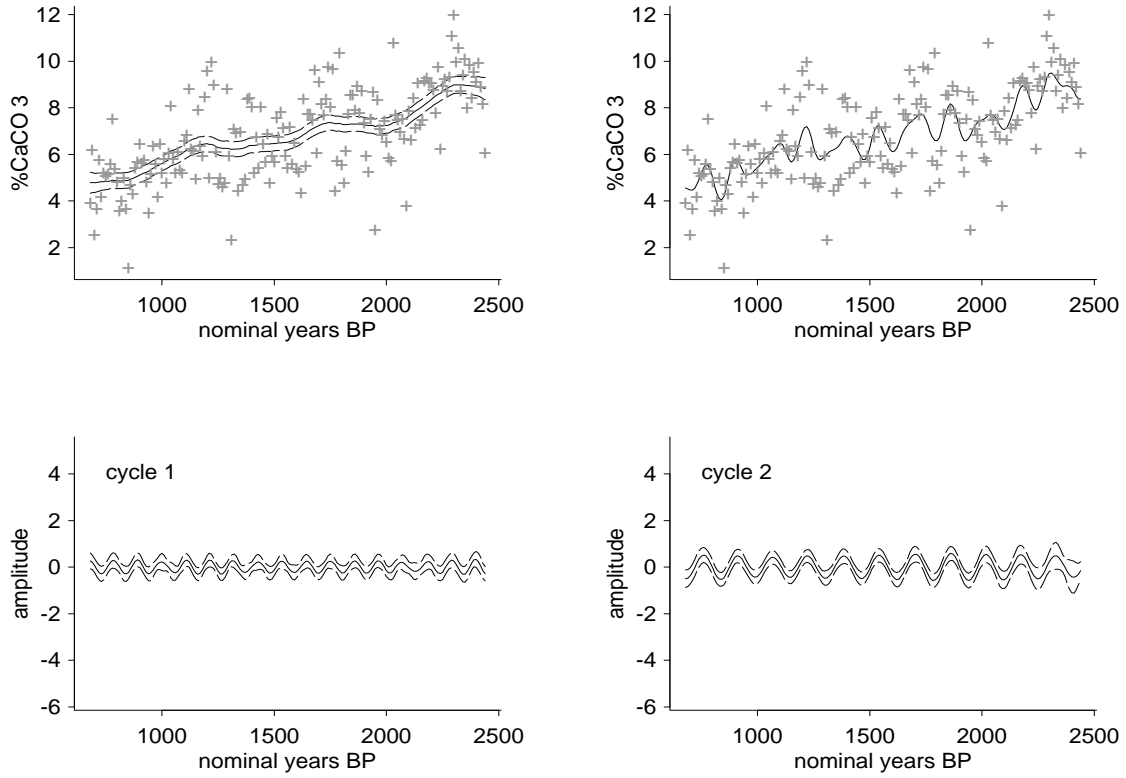


Figure 4. Some outputs from analysis of interpolated data with a two wavelength model. The upper frames scatter plot the data on posterior estimates of the underlying trend (first frame) and trend plus cyclical components (second frame). The trend plot also has bands denoting one posterior standard deviation limits. The lower frames plot trajectories, with one standard deviation limits, of the cyclical components over nominal time.

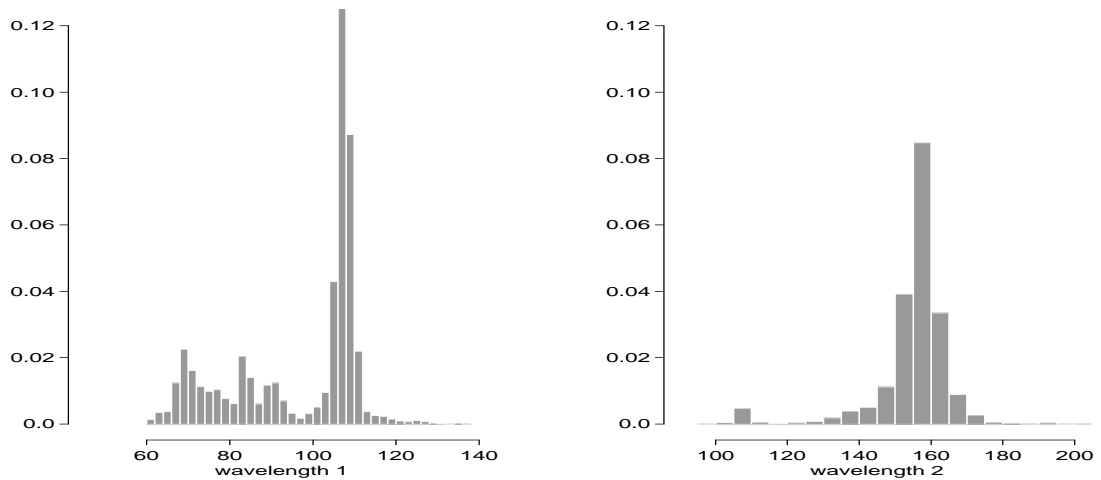


Figure 5. Margins of $p(\lambda_1, \lambda_2|Y)$ in analysis of interpolated CaCO_3 series assuming $k = 2$.

3 PALAEOCLIMATOLOGICAL TIMING AND CALIBRATION QUESTIONS

To move beyond interpolated data and “nominal” calendar time scales requires, among other things, more formal consideration of two processes of calibration of depth/time measurements. Effectively, calendar time is an “independent variable” subject to uncertainty, so we’re working in an elaborate errors-in-variables context in usual statistical parlance. Dealing with these “errors” in the context of the time series models of Section 2, and appropriately integrating expert palaeoclimatological opinions about calibration curves, is dealt with in Sections 4 and 5. This relies on developments of the two critical calibration activities, discussed here.

3.1 Mapping Core Depth To Radiocarbon Time

Refer again to Figure 6. The nominal calibration “curve” and its linear approximation have already been discussed and used in creating the earlier analyses. The former is now taken as the benchmark depth-to-radiocarbon years mapping, as in Halfman, Johnson and Finney (1994). Each radiocarbon measurement used in constructing this rather subjective prior estimate is subject to laboratory estimates of uncertainty based on traditional carbon dating techniques, represented by a nominal standard error. At a rather formal level, the reported radiocarbon dates and standard errors should be combined in consultation with expert palaeoclimatologists in order to arrive at a formal model for the uncertain calibration *function*, together with appropriate and adequate descriptions of uncertainties about the function. One can imagine, for example, developing classes of Gaussian process priors to model the mapping from depth to radiocarbon years, and incorporating the dating measurements and error assessments in a formal analysis intended to produce a predictive distribution for collections of radiocarbon dates at to-be-specified depths. This may be for the future. Here, the direct, piece-wise linear map in Figure 6 is used to define the mean of such a distribution. Writing $r = r(d)$ for the radiocarbon date associated with a specific depth d down core, we simply take $E(r|d)$ to be the crooked line displayed, and here (in following sections) investigate some elements of the sensitivity of some resulting CaCO_3 analyses to variations in measures of uncertainties about this prior expectation. Very simply, the baseline model for imputing radiocarbon dates r_i at given depths d_i assumes the dates to be conditionally independent and normally distributed about the specified mean $E(r_i|d_i)$; further, the variance is assumed constant across the length of the core, denoted by v_d . Experimental errors are likely dominant factors in determining uncertainties in laboratory measurements of radiocarbon dates, hence the independence assumption is a reasonable starting point, though a plausible later extension might introduce a positive correlation function to allow for overlapping material in neighbouring core locations. The constancy of variance assumption may also be refined at a later stage. So, for our purposes here, radiocarbon dates r_i at depths d_i are assumed drawn from independent normal distributions $(r_i|d_i) \sim N(r_i|E(r(d_i)), v_d)$ where v_d is to be specified. Note that there is no restriction to monotonicity; this prior simply predicts to-be-measured radiocarbon dates at future chosen core locations, and such outcomes are subject to the experimental sources of error that may well lead to non-monotonicity for what may be physically closely located core samples.

In the radiocarbon calibration data sets reported in Halfman, Johnson and Finney (1994), quoted nominal standard errors for radiocarbon dates vary typically in the 40-70 year range, with one or two much larger values (and it should be noted that several reported dates lie well away from the nominal curve $E(r|d)$ for their reported depths, and have been discarded as gross outliers in producing the curve.) In further CaCO_3 time series analyses here, we begin by assuming negligible uncertainty, with v_d very small, to compare with our above analyses using nominal years and interpolated CaCO_3 data. We then contrast with a more appropriate analysis with a larger and appreciable variance.

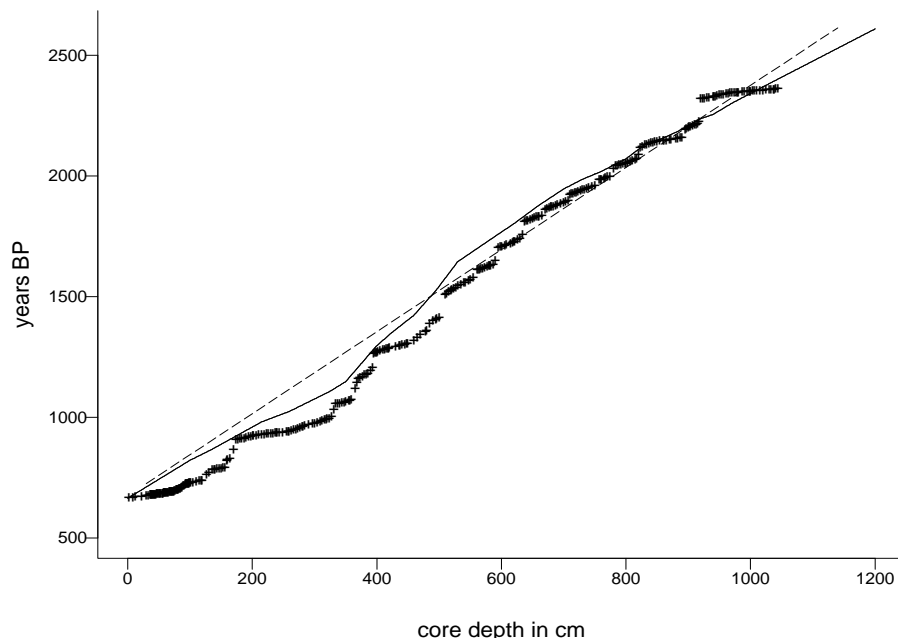


Figure 6. Approximate calibration curves: the full line gives the (current version of) the consensus map from depth to C14 years BP based on radiocarbon measurements from aligned cores LT84-7P and LT84-7P. The dashed straight line is the direct linear approximation used to provide the nominal dates used in the analyses of section 2. The crosses represent the directly interpolated *calendar* dates associated with points down core at which the carbonate values are measured; these dates arise by mapping from depth to C14 years through the full line, and then mapping (with an assumedly negligible associated s.d.) to calendar years via the radiocarbon calibration process described in Sections 3.2 and 3.3.

3.2 Mapping Radiocarbon To Calendar Time

Bayesian inference about calendar ages of archaeological specimens based on laboratory estimates of radiocarbon ages is developed in Buck, Litton and Smith (1992) with a view to dealing with collections of assumedly time-ordered samples; see also Buck and Litton (1991), and Litton and Leese (1991), and the earlier foundational work of Naylor and Smith (1989). This is essentially the framework here, and we are able to build on these previous approaches in developing prior distributions for calendar dates of collections of specified sediment core locations; this is done by linking the depth:radiocarbon model to a radiocarbon:calendar time model drawn from this previous work. Some brief review of the developments in Buck, Litton and Smith (1992) are covered here.

At the centre of radiocarbon calibration lie the high precision calibration curves of Stuiver and Pearson (1986) and Pearson and Stuiver (1986); we are interested in those sections spanning the last three or four thousand years. This calibration data base provides calendar time “knots” (at spacings of 10-20 years) at which rather precise determinations of associated radiocarbon dates have been experimentally derived and confirmed by other methods. A piecewise linear interpolant through the resulting points provides a map from calendar years to radiocarbon years BP. The map is non-linear, and quite non-monotonic, assumedly reflecting now well-established patterns of secular variation in atmospheric carbon-14. Again write r for radiocarbon years BP and now take t for true calendar years. It is assumed, following the above authors, that a laboratory determination of radiocarbon age of a sediment core sample provides a normally distributed measurement with a variance determined by the standard error quoted by the laboratory. If $m(t)$ represents the piecewise linear radiocarbon calibration curve, then the sampling model for a set of radiocarbon dates

r_i given corresponding true calendar dates t_i and quoted standard errors s_i is simply one of conditionally independent normal observations, $(r_i|t_i, s_i) \sim N(m(t_i), s_i^2)$. Consider now a collection of n samples that may be radiocarbon dated. This will, at least in principle, provide pairs (r_i, s_i) to be used in inferring calendar dates t_1, \dots, t_n . Constrained uniform priors for calendar dates, as in Buck *et al* (1992), assume the samples are known to be, or supposed to be ordered in calendar time. That is, the prior for dates is uniform over a finite interval with appropriate endpoints and in which $t_1 \leq t_2 \leq \dots \leq t_n$. These authors describe and illustrate the computations required to evaluate and sample from the resulting posterior distribution for $(T|R)$ where $T = \{t_1, \dots, t_n\}$ and $R = \{r_1, s_1; \dots, r_n, s_n\}$. The posterior for a single calendar date, $p(t_i|r_i, s_i)$, is trivially computed as a truncated piecewise normal distribution. Similar forms arise for conditional distributions such as $p(t_i|T^{(i)}, R)$ where $T^{(i)}$ is, in common notation, just the set T with t_i removed. Hence marginal posterior densities for single dates given only their radiocarbon measurements may be computed and plotted, and full joint posteriors for collections of dates may be simulated via embedding in a Markov chain sampler; that is, $(T|R)$ represents the stationary distribution of a Markov Chain sampler that iteratively samples through sequences of conditional distributions $p(t_i|T^{(i)}, R)$.

It is appropriate in our sediment core analyses to adopt such uniform priors for calendar dates, but constrained so that points ordered in depth are correspondingly ordered in time (though note, again, there is no necessary ordering in radiocarbon time). Under this assumption, the technology laid out by Buck *et al* applies directly. Figure 7 displays resulting posterior densities for calendar dates at four selected radiocarbon dates, each computed separately as if each were a single radiocarbon determination. The radiocarbon values chosen (and indicated on the graph) roughly span the range of dates of our CaCO_3 samples. In each case, the error standard deviation is taken as $s_i = 50$, in agreement with the ranges of errors quoted in the depth-to-radiocarbon calibration data set. The multimodalities appear as a result of the “saw-toothing” non-monotonocities in the high precision calibration curve. Note both the biases and the wide ranges of supported values in each case. These four densities are computed separately, and actually overlap very little; larger overlap occurs with larger values of s_i , and this impacts the joint posterior under which the times are ordered.

3.3 Mapping Core Depth to Calendar Time

Write $D = \{d_1, \dots, d_n\}$ for sediment core depths at which CaCO_3 values are measured; assume depths ordered so that $d_1 < d_2 < \dots < d_n$. We require inference about true calendar dates at the known core locations, so are explicitly interested in combining what we have in terms of the two calibration priors—namely depth-to-radiocarbon and radiocarbon-to-calendar time—into a resulting prior that describes the map from depth down core to calendar time. Formally, mathematically, we know that the required density for the collection of calendar dates T at the new depths D is simply $p(T|D) = \int p(T|R, D)p(R|D)dR$. The “standard” radiocarbon dating analysis of Section 3.2 delivers the component $p(T|R) = p(T|R, D)$ under the additional assumption that, given R , the depth information serves only to induce the time ordering of calendar dates. The nominal depth-to-radiocarbon prior in Section 3.1 is just the other component, $p(R|D)$. Combining the two components gives, in principle, the complete prior $p(T|D)$. In practice, this is not an easy function to evaluate or summarise, but is trivially simulated. To see this, note that we can draw directly from the specific prior $p(R|D)$ discussed in Section 3.1 (actually, only the radiocarbon dates r_i are drawn, the standard errors $s_i = v_d$ assumed fixed, known and constant). Then a draw from $p(T|D)$ results by sampling $p(T|R)$ with R set at the values just sampled.

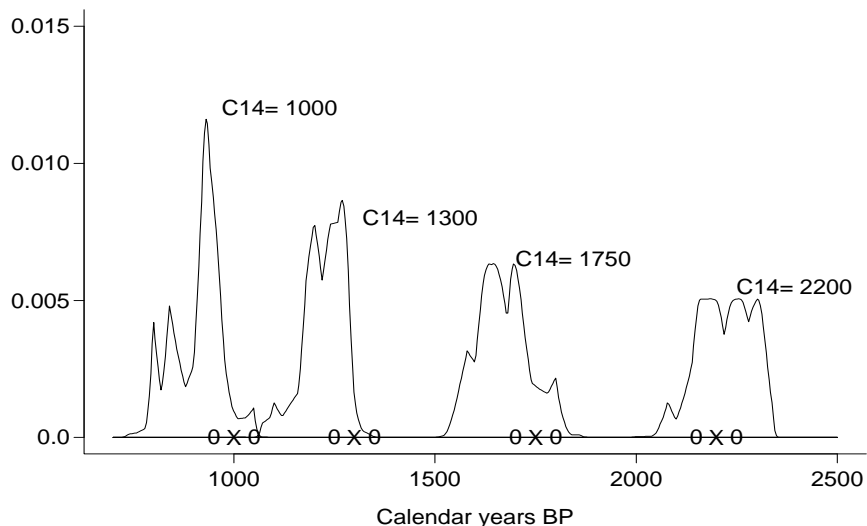


Figure 7. Posteriors for true calendar dates of core samples at four possible C14 dates each with s.d. of fifty years (following Buck, Litton and Smith 1992). The X symbols mark the C14 observations with one s.d. either side indicated by the O symbols.

To give the flavour of this completed calibration step, single draws of calendar dates have been made for our CaCO_3 data series with negligible uncertainty. The normal standard deviations $s_i = v_d$ are set to 0.01 so that sampled radiocarbon dates corresponding to the actual depth locations of our CaCO_3 data series are essentially just read off the piecewise linear function in Figure 6, and the sampled calendar dates represent essentially direct interpolation from the radiocarbon dates via the high-precision calibration curve. The dates so evaluated appear as small crosses in Figure 6. Several features emerge here. One general point is that, up to about 2300 years BP the radiocarbon dates exceed calendar dates, a feature clear from the standard high precision curves. A second important feature is that there are periods during which there is rather little variation in imputed calendar dates across stretches of the core, notably around 700 and again around 900 years BP, and further back in time around 2300 years BP. These suggest periods of extremely rapid sedimentation rates and (at least in principle) offer potential for exploration of variations in geochemical indicators, such as the CaCO_3 data, on very short time scales (though we have yet to focus closely on these periods). These “directly imputed” calendar dates provide the times for plotting the CaCO_3 response series in Figure 8; the nominal time span of the $n = 350$ observations is now about 660–2440, a total of about 1780 years and so an average spacing of about 5 years per observation. After truncating to nearest year, only 302 of the observations are timed in distinct years, 94 consecutive observations are timed in consecutive years, a further 67 are just two years apart, 44 are just three years apart, 53 more are between 4 and 5 years apart, and the inter-observation spacings of the remaining 44 observations range up to an isolated extreme of 95 years; this variation in spacings is clearly apparent in Figure 7.

4 UNEQUALLY SPACED DATA AND UNCERTAIN TIMINGS

With a facility for imputing the actual calendar dates t_i for the CaCO_3 values y_i , it now becomes necessary to extend the time series modelling to cater for the resulting unequally spaced observations, and also to admit and allow for the uncertainties about actual timings. In our context, though calendar time is effectively

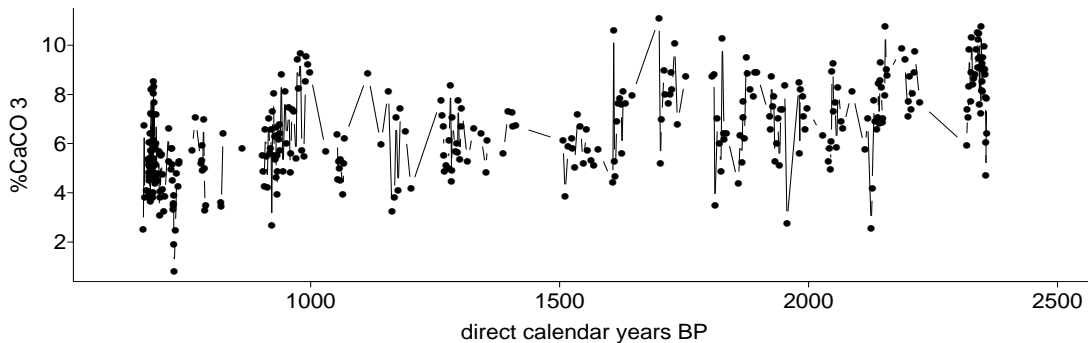


Figure 8. CaCO_3 time series plotted against approximate calendar times induced by direct maps of depth measurements to C14 years BP and then to the calendar time scale, as in Figure 6.

continuous, the scales on which we are working and the concomitant uncertainties imply that discretisation to years, or even small integer multiples of years, is appropriate (recall the close agreement between Bayesian spectrum estimates, on the nominal time scale, based on original data and on decadal interpolations, in Figure 2; the agreement provides one perspective on the negligible loss of information through the data aggregation there). So we work with calendar dates truncated to the nearest year throughout this section, then aggregate to the decadal level in subsequent analyses for comparison with the analyses in Section 2.

4.1 Aggregation Questions and Unequal Spacings

As at the outset of Section 2, CaCO_3 measurements y_i are assumed to correspond to actual calendar dates t_i in years, ($i = 1, 2, \dots, n$). First suppose that $T = \{t_1, \dots, t_n\}$ is known, so we have yearly data that is unequally spaced. The final paragraph of Section 3 discussed the variability in spacings apparent in the “directly imputed” benchmark calendar dates; the $n = 350$ observations are there timed at just 302 distinct calendar years that range between about 660-2440 years BP, so that, for a model built on an annual basis, the 302 distinct time points represents a “missing data” rate of about 83%. So there are two details to consider in dealing with the CaCO_3 series conditional on a set of times T ; concurrent times (to the nearest year) of neighbouring observations, and allowing for (lots of) missing data between consecutive observations more than a single year apart.

Dealing with concurrent observations is trivial. If a sequence of some h consecutive observation times are equal, say $t_j = t_{j+1} = \dots = t_{j+h-1}$ for some $h > 0$ and some j , then the CaCO_3 value to be used for the common year t_j is simply the arithmetic mean of the corresponding h values y_j, \dots, y_{j+h-1} , and the observational variance for this case j is weighted from v to v/h . So generally, time series analysis will be modified to incorporate observational variance weights. It is assumed from here on that this is automatically done, so that any observation y_j in year t_j may actually represent the average of several observations with imputed time t_j to the nearest year. The variance weighted modifications to a DLM analysis are obviously trivial.

Dealing with unequal spacing of observations, by treating multiple years with no observations as representing “missing data”, is also straightforward (West and Harrison, chapter 10). In the time series model displayed in Section 2.2, let time index t represent years so that x_t represents the CaCO_3 value in year t ; then $y_i \equiv x_{t_i}$ are the observed data, the remaining values being unobserved. In analysis of the observed data using this model framework, missing data are trivially subsumed as noted in West (1995). In the DLM

updating computations at the heart of the analysis, a missing observation simply means that steps involving Bayes' theorem are vacuous; in a general sense, missing data implies the posterior distribution of interest is just the prior. Simulation analysis directly caters for unequally spaced data, and delivers the same kinds of draws from full posteriors of all model parameters and the latent sub-series components over time as in the canonical, full but equally spaced data case of Section 2. Also, by appropriately extending the time scale endpoints beyond those of the data, these outputs automatically include samples from the posterior *predictive* distributions of the component series. This is illustrated below.

4.2 Uncertain Timings: Combining Time Series and Calibration Models

So far we have a class of time series models that, at a general level, specifies $p(Y|T, \theta)$, where Y is the (to be) observed data, T is the set of observation times for Y , and θ represents all model parameters and latent sub-series; so inference about θ is the primary goal, and the simulation analyses above allow us to explore $p(\theta|T, Y)$ for any specified set of times T . Now we must more realistically address the fact that T is (very) uncertain, subject to the uncertainties deriving from the various calibration processes in Section 3. Note that these processes were explored to develop a prior distribution for these times, $p(T|D)$, where D is the set of sediment core depths and this prior is implicitly also conditional on the radiocarbon calibration data set as well as, to a great extent, less specific expert opinion. To simplify notation in the current, rather general setting, we drop the conditioning symbol D though of course bear in mind the dependence on depth information.

Formal incorporation of uncertainty about T in the inferences about θ requires evaluation of the full joint posterior $p(T, \theta|Y)$ and then integration to marginalise with respect to T . Obviously, we do this via simulation. The existing Markov chain algorithms to (eventually) simulate from $p(\theta|T, Y)$ can be coupled with some appropriate algorithm for making draws from the corresponding conditional posterior $p(T|\theta, Y) \propto p(T)p(Y|T, \theta)$ (implicitly assuming prior independence of T and θ .) Embedding draws of new calendar times within each iteration of the existing sampling scheme for θ extends the state-space of the Markov Chain to include uncertain timings. It is immediately clear that direct sampling from the conditional for T is infeasible, and so we approach this using Metropolis/independence chain sampling approaches (Müller 1991, Tierney 1991). Note that, under our time series models, knowledge of θ and T implies the CaCO_3 data are conditionally independent and normally distributed with known moments; as a result, the likelihood function $p(Y|T, \theta)$ factors into products of terms $p(y_i|t_i, \theta)$. This opens the way to targeting draws from univariate conditional posteriors

$$p(t_i|T^{(i)}, \theta, Y) \propto p(t_i|T^{(i)})p(y_i|t_i, \theta),$$

in which the conditional priors $p(t_i|T^{(i)})$ are used as (tentative) proposal distributions (Müller 1991, Tierney 1991). In more detail, at a given stage in the sampling iterations with a given θ , the current set of calendar times $T^{(i)}$ determines a specific (and in our case, truncated piecewise normal) conditional prior $p(t_i|T^{(i)})$; we may easily sample from this, as described in Section 3, to deliver a ‘‘candidate’’ value t_i^* for the i^{th} calendar time. To proceed, we either save and record this candidate value in T , or reject it and revert to the previous value t_i . The Metropolis acceptance probability is simply

$$\min\{1, p(y_i|t_i^*, \theta)/p(y_i|t_i, \theta)\},$$

which is trivially computed based on the conditional normal densities determining the involved likelihood ratio. Sequencing through from $i = 1, \dots, n$ will produce an updated set of times T , among which there will typically be some times that are unchanged, all truncated to the nearest year.

As with much of the earlier work in this paper, this development represents an initial direction to exploring rather complex and high-dimensional posterior distributions. Refined approaches that involve candidate times drawn from a proposal distribution that may, in some sense, represent a better approximation to the posterior than does the prior, is one direction of future interest.

5 TWO FURTHER DATA ANALYSES

Two further explorations of the CaCO_3 data are briefly summarised, in which the previous analyses are extended to the actual rather than interpolated data, and the above development is followed to include calendar times of observations as uncertain quantities. The first analysis assumes very little variation in the depth-to-calendar time calibration setting $s_i^2 = v_d = 0.5$ in sampling the calendar times. The second analysis admits more realistic degrees of uncertainty in calibration via $v_d = 30$. In all other respects, in terms of model and prior specifications, the set-up replicates that in Section 2; each analysis is based on the DLM with a first-order polynomial trend plus two cyclical AR(2) components, as in the analysis of the nominally timed, interpolated data analysis summarised in Figures 4 and 5.

The first analysis, with negligible variation about the radiocarbon estimates, is summarised in Figures 9 and 10, to be compared with Figures 4 and 5. The trend estimate is similar, though the refined analysis exhibits more uncertainty about both the trend and the estimated trajectories of cyclical components. There is more variability apparent in the estimated cyclical components, with increased amplitude around 900-1100 years BP and again back beyond 2100 years BP. Otherwise, the figures broadly agree with the previous analysis. One additional feature in Figure 9 is the predictive extrapolation from the most recent observations around 660 years BP up to more recent times; the graphs of predicted trend and cyclical components between 350 and 660 years BP. In terms of forecasting future behaviour, the sum of trend and cycles in upper right frame is relevant. Figure 10 parallels Figure 5 in displaying posteriors for the two wavelengths; in addition to histograms of sampled values, the full lines in each case represent “improved” approximations to the posterior densities constructed as Monte Carlo averages of the theoretically exact conditional posteriors for the wavelength parameters available at each iteration of the simulation analysis; in this model, the averaged conditionals are, on the β_j scales, simply truncated normals. Again there are strong similarities between these pictures and those in Figure 5, the main differences being, firstly, the shift of the peak for the lower wavelength to around 85-95 from the higher region around 110 based on the nominal time scale, and, secondly, increased mass at lower values for this same wavelength. Again, if there are cyclical components with wavelengths in the 40-70 year range, their amplitudes over time are insignificant, as further analysis with $k = 3$ components confirms. This multimodality around small values is noise driven; the increased mass in the region in this analysis is in part a reflection of the fact that this component of lower wavelength is itself verging on insignificance, as the uncertainty bands in the lower left frame of Figure 9 indicate.

The second analysis, with $s_i^2 = v_d = 30$, introduces much more uncertainty into the model through variations on the calendar time scale, and can be expected to impact on inferences at least in terms of increased posterior uncertainties (though even 30 years as s.d. under-represents realistic uncertainties). Before briefly commenting on implications of the the analysis, it should be noted that the Metropolis steps to resample calendar times achieved quite high acceptance rates, saving in the range of 70-90% of the candidate draws for the $n = 350$ observations at each simulation iteration, with an average rate of about 79%; this gives some reason to believe that sampling the conditional priors in the Metropolis sub-analysis, as implemented, does not unduly restrict movement around the calendar time space, and that the posterior for times is being reasonably well investigated. That said, much remains to be explored in this area.

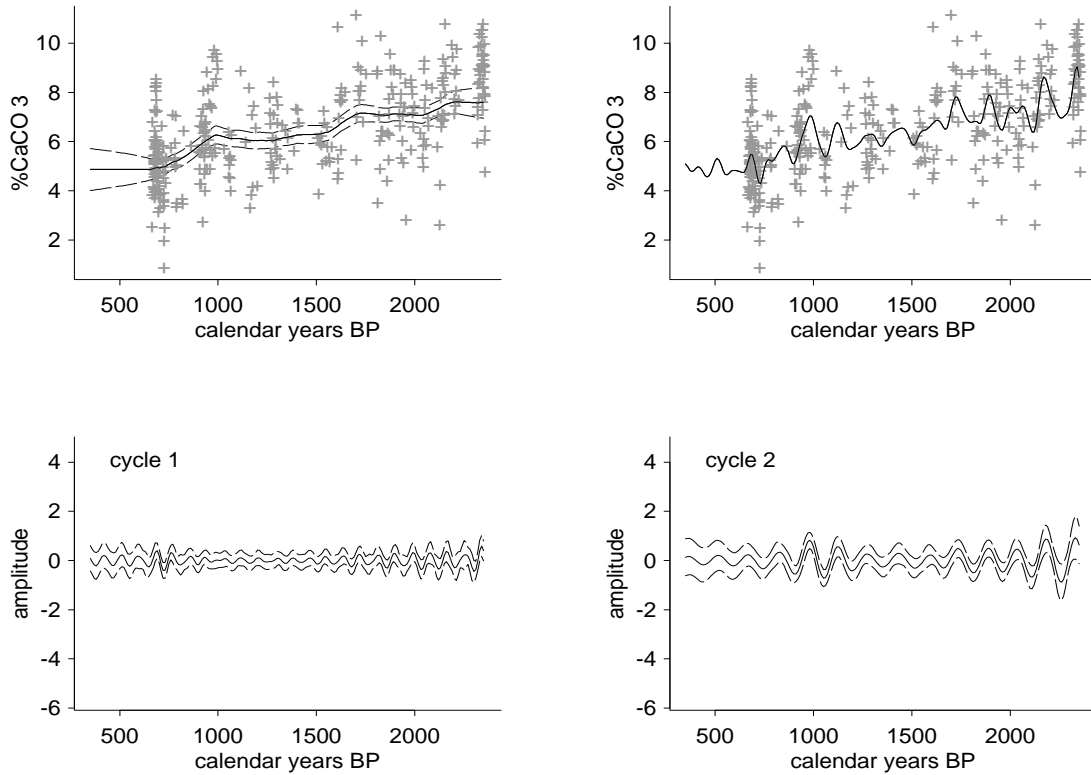


Figure 9. Some outputs from analysis of original CaCO_3 data with a two wavelength model and with minimum prior variation, $v_d = 0.5$, in the depth to radiocarbon calibration. The upper frames scatter plot the data on posterior estimates of the underlying trend (first frame) and trend plus cyclical components (second frame). The trend plot also has bands denoting one posterior standard deviation limits. The lower frames plot trajectories, with one standard deviation limits, of the cyclical components over nominal time.

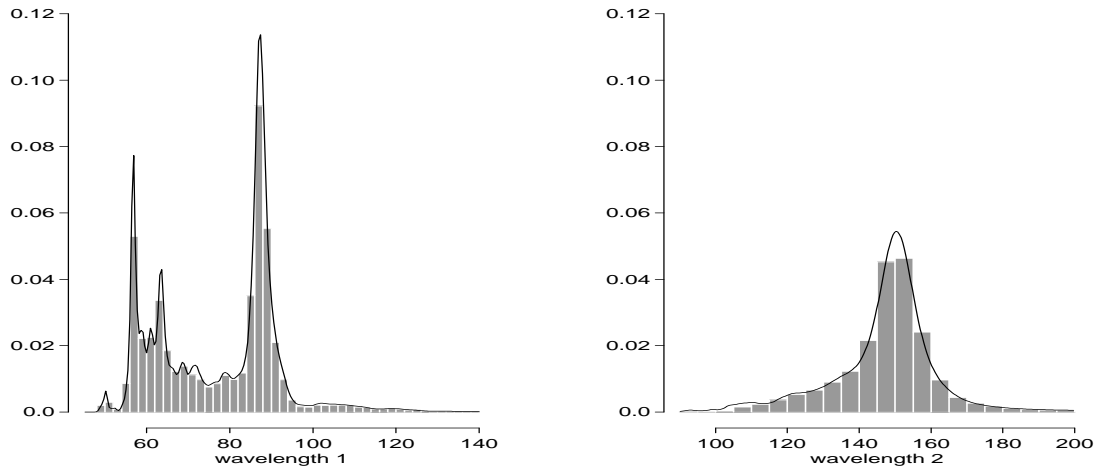


Figure 10. Posteriors for nominal wavelengths λ_1 and λ_2 from the analysis of the original CaCO_3 series using a two wavelength model with minimum prior variation, $v_d = 0.5$, in the depth to radiocarbon calibration.

This analysis is summarised in Figures 11 and 12 in what is now a familiar format. The posteriors for wavelengths favour similar regions, though are now much more diffuse, as expected. More importantly, the major changes in estimated trajectories of the two cyclical AR terms are edifying; uncertainty now dominates, the components being apparently insignificant in describing variation in the data in the context of such uncertainty about observation times. The message is salutary, particularly since even higher degrees of uncertainty about times seem appropriate, and strongly highlight the limitations imposed on the palaeoclimatologists by the limited precision associated with carbon dating methods. Additional approaches to refined and more precise dating of core locations is radically needed.

6 CONCLUDING COMMENTS

At the outset of this still infant study, we were interested in developing a deeper appreciation of the limitations of existing sedimentary records, in improving communication with geologists, and in developing and refining time series methods for palaeoclimatological data analyses. We have made a small start on all counts. On the specific issue of patterns in the CaCO_3 series, these initial investigations lead to negative conclusions. The various DLM analyses essentially confirm that, if there are short-term cyclical patterns in the CaCO_3 series, they are extremely subtle and almost wholly dominated by noise; taken at face value, the primary wavelength is in the 140-160 year range, but the predictive value of models admitting cyclical patterns is negligible. Admitting even small degrees of uncertainty about the calendar time scale really buries these conclusions in noise; the data is quite limited by this together with the experimental and environmental noise in the CaCO_3 records.

Previous analyses of similar data (Halfman, Johnson and Finney 1994, and references therein) have focused on high frequency behaviour and tentatively identified cycles of wavelengths down to ten or eleven years; the current paper clearly invalidates these kinds of conclusions, attributing any such high frequency characteristics to noise. Additional analyses (not reported) that extend the models here to include three or more cyclical components, some of possibly very low wavelengths, confirm their insignificance.

Several dimensions of the scientific context and experimental framework that have not been touched on here are evidently open for investigation and may lead to refined analyses of greater precision. These include taking a closer statistical look at the processes by which the CaCO_3 data from two separate cores were combined via subjective stratigraphic correlations of independent parameters, and investigating combining data across several more cores. Spatial variation in CaCO_3 across the lake bed will be an issue here. Related questions have to do with investigations of other geochemical measurements in sediment, including, notably, oxygen and carbon isotope ratios (Johnson *et al* 1991), and the possible combination of such records with CaCO_3 and other data. Additional investigation of the original radiocarbon calibration data set may also be pursued, though additional approaches to refining the accuracy of calendar datings are evidently critically needed; some prospects exist in studying various “interesting” patterns of dark:light “laminations” that are apparent down several cores and that might relate to identifiable interannual variations in weather patterns (Halfman, Johnson and Finney 1994).

There are other, likely subsidiary issues that have not been covered, including, for example, sources of error in locating samples down core and problems of remixing of sediments during core extraction and subsequent laboratory handling. These kinds of details might be woven into the analysis, extending models to include additional components that might be directly incorporated in posterior simulations.

One big question has to do with the extent to which CaCO_3 and other geochemical indicators really do stand as proxy for local climatic conditions. Here, as in most existing work, simple extrapolations of estimated

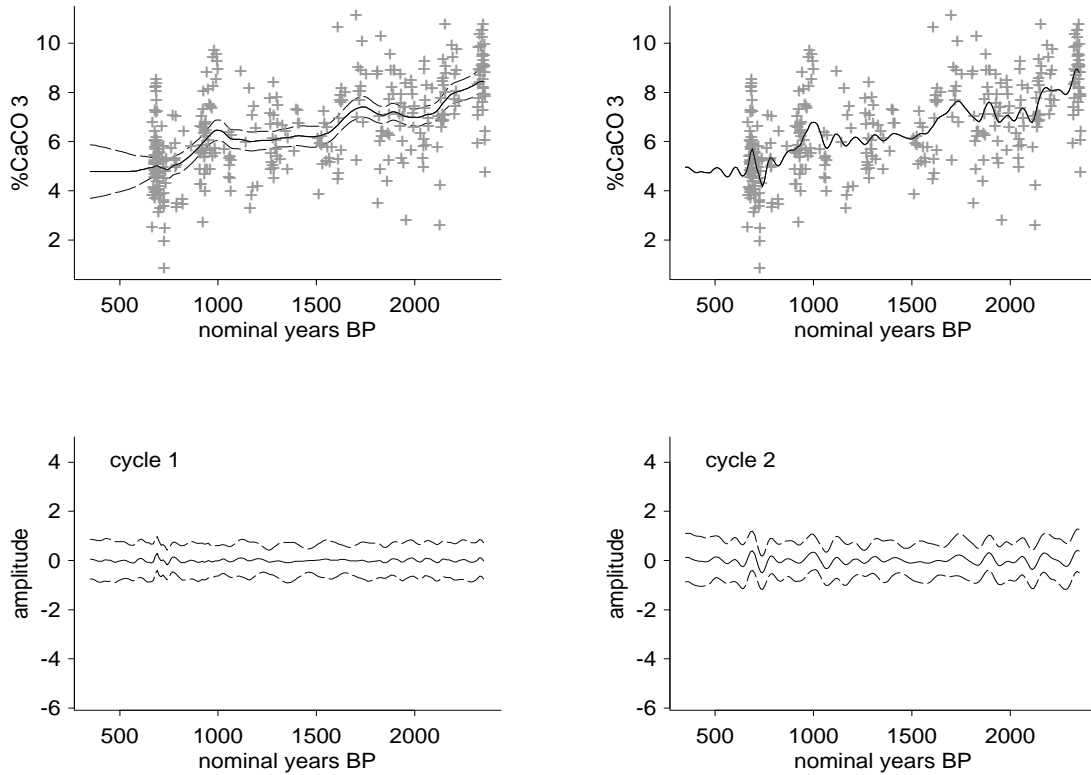


Figure 11. Some outputs from analysis of carbonate data with a two wavelength model and with realistic prior variation, $v_d = 30$, in the depth to radiocarbon calibration. The upper frames scatter plot the data on posterior estimates of the underlying trend (first frame) and trend plus cyclical components (second frame). The trend plot also has bands denoting one posterior standard deviation limits. The lower frames plot trajectories, with one standard deviation limits, of the cyclical components over nominal time.

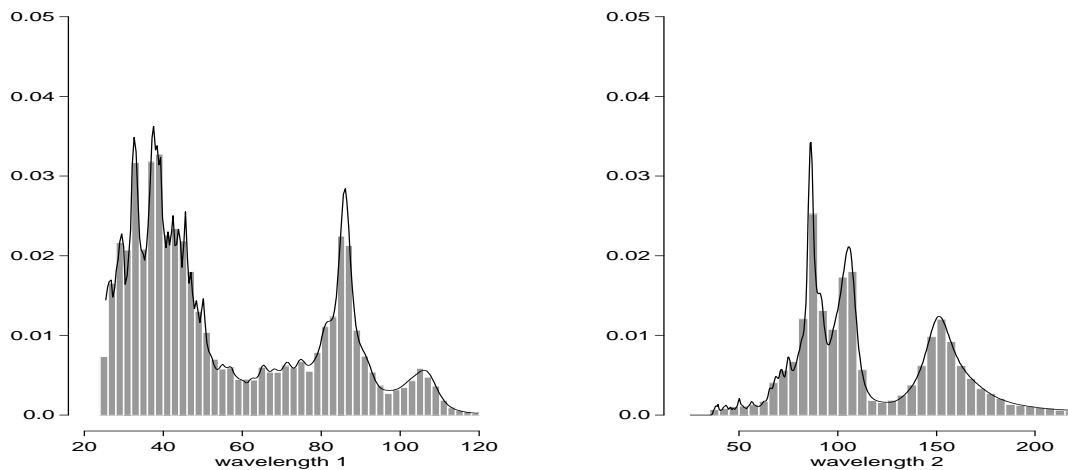


Figure 12. Posteriors for nominal wavelengths λ_1 and λ_2 from the analysis of the two wavelength model with realistic prior variation, $v_d = 30$, in the depth to radiocarbon calibration.

levels are typically used in informal inference about the climatic variables of real significance in assessing climate changes—notably historical lake levels and water temperature. With respect to the specific study of Lake Turkana, variations in CaCO_3 and isotope records are believed to be affected primarily by variations in lake water evaporation relative to freshwater input from inflowing rivers, and so are indicative of local climatic conditions in the drainage basin. Inferences drawn on this basis are informal and cautious, recognising that, in addition to assuming secondary effects are negligible, the relationships between measurement levels and the lake levels are not realistically as simple as direct linear extrapolations imply. Geophysical theory may eventually provide guidance to model non-linear distortions and assess uncertainties in the extrapolations; another calibration component.

From the viewpoint of statistical research, this paper illustrates novel time series modelling developments and the integration of several, rather complex component models, that clearly have much wider utility, though have been stimulated by this beginning collaboration. The linking of several complex component models is enabled, primarily, by the use of Markov chain simulation methods of analysis, being essentially unmanageable otherwise. There are generic technical issues to be addressed in the context of the specific time series models, having to do with questions of convergence of simulations in high dimensional parameter spaces and with what are expected to be “erratically” behaved posteriors; work on these issues, including extensions and modifications of the simulation methods illustrated, will apply to other kinds of time series models, including DLMs with quasi-cyclical and stationary components. Also, the modelling and simulation methodology clearly has applications in other areas of time series analysis. The simulation techniques have opened up inference about parameters of dynamic linear models in important ways. The developments to allow for uncertainties, and prior models, for observations times in essentially linear time series models relate directly to wider issues of non-linear modelling via stochastic variations in times and time axis deformations; some applications are currently being investigated in a couple of exciting areas.

ACKNOWLEDGEMENTS

This work arose from discussions with Tom Johnson, of the Department of Geology at Duke (now at the University of Minnesota), with whom further collaborative research is in progress. He, and John Halfman of the University of Notre Dame, Indiana, provided the data explored in this paper. Discussions with graduate students Alyson Wilson and Yang Chen of ISDS on some of these issues have been most useful. This research was partially supported by the NSF under grants DMS-9024793, DMS-9305699 and DMS-9311071.

REFERENCES

- Bretthorst, L.G. (1988) *Bayesian Spectrum Analysis and Parameter Estimation*, Lecture Notes in Statistics, Springer-Verlag, New York.
- Bowman, S. (1990) *Interpreting the Past: Radiocarbon Dating*, British Museum Publications.
- Buck, C.E. and Litton, C.D. (1991) A computational Bayes approach to some common archaeological problems, in *Computer Applications and Quantitative Methods in Archaeology 1990*, (S.P.Q. Rahtz, ed.). BAR International Series, 565, 93-99.
- Buck, C.E., Litton, C.D. and Smith, A.F.M. (1992) Calibration of radiocarbon results pertaining to related archaeological events, *J. Archaeological Science*, 19, 497-512.
- Butzer, K.W., Isaac, G.L., Richardson, J.L. and Washbourn-Kamau, C. (1972) Radiocarbon dating of east African lake levels, *Science*, 175, 1069-1076.
- Carter, C.K., and Kohn, R., (1994) On Gibbs sampling for state space models. *Biometrika* **81** 541-553.

- Frühwirth-Schnatter, S. (1994) Data augmentation and dynamic linear models. *J. Time Series Analysis* **15**, 183-102.
- Halfman, J.D. and Johnson, T.C. (1988) High-resolution record of cyclic climate change during the past 4ka from Lake Turkana, Kenya, *Geology*, **16**, 496-500.
- Halfman, J.D., Johnson, T.C. and Finney, B.P. (1994) New AMS dates, stratigraphic correlations and decadal climatic cycles for the past 4ka at Lake Turkana, Kenya. *Paleogeography, Paleoclimatology, Paleoecology*, **111**, 83-98.
- Johnson, T.C. (1984) Sedimentation in large lakes. *Ann. Rev. Earth Planet Sci.*, **12**, 179-204.
- Johnson, T.C., Halfman, J.D., Rosendahl, B.R. and Lister, G.S. (1987) Climatic and tectonic effects on sedimentation in a rift-valley lake: Evidence from Lake Turkana, Kenya. *Bull. Geol. Soc. Amer.*, **98**, 439-447.
- Johnson, T.C., Halfman, J.D. and Showers, W.J. (1991) Paleoclimate of the past 4000 years at Lake Turkana, Kenya, based on the isotopic composition of authigenic calcite, *Paleogeography, Paleoclimatology, Paleoecology*, **85**, 189-198.
- Litton, C.D. and Leese, M.N. (1991) Some statistical problems arising in radiocarbon calibration, in *Computer Applications and Quantitative Methods in Archaeology 1990*, (S.P.Q. Rahtz, ed.). BAR International Series, **565**, 101-109.
- Müller, P. (1991) Metropolis based posterior integration schemes. Invited revision for *J. Amer. Statist. Ass.*
- Naylor, J.C. and Smith, A.F.M. (1989) An archaeological inference problem, *J. Amer. Statist. Ass.*, **83**, 588-595.
- Owen, R.B. Barthelme, J.W., Renault, R.W. and Vincens, A. (1982) Paleolimnology and archaeology of Holocene deposits north-east of Lake Turkana, Kenya. *Nature*, **198**, 523-529.
- Pearson, G.W. and Stuiver, M. (1986) High-precision calibration of the radiocarbon time scale, 500-2500 BC, *Radiocarbon*, **28(2B)**, 839-862.
- Pole, A., West, M. and Harrison, P.J. (1994) *Applied Bayesian Forecasting and Time Series Analysis*, Chapman-Hall, New York.
- Stuiver, M. (1970) Oxygen and carbon isotope ratios of freshwater carbonates as climatic indicators. *J. Geophys. Res.*, **75**, 5247-5257.
- Stuiver, M. and Pearson, G.W. (1986) High-precision calibration of the radiocarbon time scale, AD 1950-500 BC, *Radiocarbon*, **28(2B)**, 805-838.
- Talbot, M.R. (1990) A review of the paleohydrological interpretation of carbon and oxygen isotope ratios in primary lacustrine carbonates. *Chem. Geol.*, **80**, 2610279.
- Tierney, L. (1991) Exploring posterior distributions using Markov chains. In *Computer Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, (ed. E. M. Keramidas), 563-570, Interface Foundation, Fairfax Station.
- West, M. and Harrison, P.J. (1989) *Bayesian Forecasting and Dynamic Models*, Springer-Verlag, New York.
- West, M. (1995) Bayesian inference in cyclical component dynamic linear models. *J. Amer. Statist. Ass.* (to appear, December 1995 issue).

DISCUSSION BY DANI GAMERMAN, UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

This discussion is roughly divided into 3 parts. The first one with comments about a selection of good things of the paper. Then some comments about parts of the paper that could have been developed more or at least more clearly described. These have mostly to do with the application of the models to the data. Finally, I turn to the related topics of reversed time and various cyclical series that are touched by parts of the paper.

The paper provides a comprehensive analysis of geological data including models for: D (core depths) \rightarrow R (radiocarbon times); R \rightarrow T (calendar times) and T \rightarrow Y ($CaCO_3$ levels) where an \rightarrow indicates the specification of a conditional distribution. The combination of these submodels leads to the model D \rightarrow R \rightarrow T \rightarrow Y. The absence of other possible arrows is mostly justified in terms of the application.

I would like to emphasise that this paper is a very healthy sign of a recent trend in (Bayesian) statistics towards modelling an area rather than stylised problems within it. The machinery available now to statisticians enables this combination of different submodels into an unified framework and the Bayesian paradigm is particularly suited for that task by its probabilistic language. This trend in applied work is moving alongside the developments in expert systems which aims the same encompassing objectives.

Having said that, it is worth noting that the main purpose of the study was to detect climatic changes and, if possible, cycles. It is rather unfortunate that an important final arrow relating Y to climate or variables (such as temperature and humidity) measuring it are missing from this study. It is possible that the specifics of the area will imply this relation only through intermediate variables thus requiring extra arrows.

The main strengths of the paper to me lie in the particularly interesting uses made of simulation techniques in the elimination of R and the treatment of uncertainty about observation times. The elimination of R via

$$p(T|D) = \int p(T|R)p(R|D)dR$$

is only feasible approximately through samples t_j from $[T|D]$ obtained by successively sampling r_j from $[R|d_j]$ and then sampling from $[T|r_j]$.

The most interesting aspect of the paper however is the novel incorporation of uncertain times. This involves replacing the posterior $p(\theta|y)$ by $p(\theta, T|y)$. Again, realistic model assumptions prevent a closed form analysis and simulation must be used. Samples (θ_j, t_j) from $[\theta, T|y]$ can be obtained by cycling through $[\theta|t_j, y]$ and $[T|\theta_j, y]$. Note that the first distribution is the (standard) posterior given knowledge of observation times and the second one can be sampled by adding a Metropolis step to correct the values generated from $[T|D]$ for the likelihood.

The application to the data is the only part that I would have liked to see more exploration. The model for $[R|D]$ in section 3.1 assumed independent errors based on an anticipated dominance of experimental error (large v_d) as a prime source of uncertainty. The results of the first analysis of section 5 with small v_d are therefore questionable. More to the point however was the second analysis with variances $v_d = Var(R|D)$ and $s^2 = Var(R|T)$ fixed at 30. This analysis is to be opposed to the first one with small values for v_d and s^2 . A more comprehensive analysis should certainly assume these variance components to be unknown and estimate them. This step is very easily implemented and no reason was given for not doing it.

Also, the role of evolution variance $W = Var(z_t|z_{t-1})$ in the data analysis was not clarified. It provides information on the stability of the cyclical representation. It would be interesting to compare their posterior

distributions from the two analysis of section 5. Comparison of outputs suggests that increased values of s^2 and v_d seem also to point to a more unstable cyclical representation. A more complete analysis would allow determination of the correlation between the different variance components or the extent to which one can identify their individual influences over the model.

We can now turn to related topics, the first of which being reversed time. Time was probably used in reversed order given the retrospective nature of the data. Even though sine waves remain unchanged after reflection, one should be careful when dealing with dynamic models. Assume the notation of the paper with

$$[z_t|z_{t-1}, D_{t-1}] = N(G_t z_{t-1}, W_t), \quad (1)$$

and set also $[z_{t-1}|D_{t-1}] = N(m_{t-1}, C_{t-1})$ and $Cov(z_t, z_{t-1}|D_{t-1}) = S_t = G_t C_{t-1}$. This implies that

$$z_{t-1}|z_t, D_{t-1} \sim N(m_{t-1} + S'_t R_t^{-1}(z_t - a_t), C_{t-1} - S'_t R_t^{-1} S_t) \quad (2)$$

where a_t and R_t are the first moments of $[z_t|D_{t-1}]$. Equation (2) is not useful as an evolution equation in reversed time because it depends on *future* information D_{t-1} instead of the required *past* information $D_n \setminus D_{t-1}$. If however, (1) does not depend on D_{t-1} , remaining specifications above can be marginalised with respect to D_{t-1} , so that result (2) becomes also unconditional on observed information. Namely, $[z_{t-1}|z_t, D_{t-1}] = [z_{t-1}|z_t] = [z_{t-1}|z_t, D_n \setminus D_{t-1}]$.

The evolution equation of the paper is clearly independent of data information. So, applying (2) to the cycle model of the paper with a single wavelength $x_t = \beta x_{t-1} - x_{t-2} + w_t$ leads to

$$x_{t-2} = p(\beta x_{t-1} - x_t) + (1-p)[E(x_{t-2}) + \delta\{x_{t-1} - E(x_{t-1})\}] + \sqrt{p}w_{t-2}^*$$

where $[w_{t-2}^*] = [w_t] = N(0, w)$ and $p^{-1} = 1 + [\{Var(x_{t-2})/w\}(1 - Corr^2(x_{t-1}, x_{t-2}))]^{-1}$ for some constant δ . The evolution model is then a weighted (by p) average between 2 components, one of which is the evolution in original time scale. It is the same as in reversed time only if $p = 1$. Note, however, that if $Var(x_{t-2}) \gg w$ as is the case of the vague priors used in the paper, $p \approx 1$. By superposition of components and under some unimportant assumptions, the above result can be generalised to the model with various wavelengths and a trend component.

Finally, I would briefly like to address the problem of dealing with various series. In the paper, this issue is resolved by a subjective synchronisation performed by geologists which lead to a single series. Given the careful incorporation of uncertainties in other stages, one should allow for the uncertainty involved here. One possible model for series y_{t1}, \dots, y_{tm} could be

$$\begin{aligned} y_{ti} &= F' z_{ti} + \nu_{ti} \\ z_{ti} &= G(\lambda_i) z_{t-1,i} + \delta_{ti} \\ g(\lambda_i) &= \eta_i \\ \eta_i &= X'_i \beta + v_i \end{aligned}$$

where the dependence on the wavelengths by the evolution matrix is now made explicit. One simple example is the case of exchangeable wavelengths $\log \lambda_i = \log \lambda_0 + v_i^*$, $v_i^* \sim N(0, V)$. The analysis is virtually the same but for the introduction of λ_0 and V . If assumed priors are log-normal for $[\lambda_0]$ and IW for $[V]$, full conditionals are in conjugate form. Same ideas can be used in case hypercovariates X_i are deemed necessary.

REPLY TO COMMENTS OF D GAMERMAN

Dani discusses several directions for further development and extension of my work, most of which are (briefly) touched on in the paper, and raises questions/criticisms on several aspects of the presented material; I treat these points of contention in the order presented in the discussion.

Dani is concerned about the omission of a “final” modelling stage, representing a further “arrow” linking geochemical data records Y to interpretable measures of climatic conditions, such as ambient temperature and humidity. I explicitly address this issue in the penultimate paragraph of the paper; there are indeed prospects for adding further arrows, though it must be said that there are very many additional complications and uncertainties involved. Partly due to the complexity, these and other proxy variables routinely analysed in the palaeoclimatological literature do seem to have perhaps overly “exalted” status as real measures of climatic conditions. In the context of the reported study, however, the “main purpose” was not, as Dani suggests, to detect (cycles) in climatic change, but rather to take a formal and well-founded, model-based look at the kinds of data series that had previously been analysed by geologists, to reassess conclusions of earlier studies in the context of more adequate accounting for sources of experimental uncertainties. The fact that we are dealing with proxy climatic indicators is largely irrelevant in this context.

A second set of issues is raised in connection with the possible development of more “sophisticated” models for the calibration curve mapping core depth to radiocarbon dates. Again as anticipated and touched on in the paper, more “appropriate” models are certainly accessible and implementable without adding unduly to modelling or computational complexity. Indeed earlier work did explore analyses using simple Gaussian process priors for the curve ($R|D$) to induce smoothness, priors for variance components and variations on the normality assumptions to include heavier-tailed T distributions and outlier components. These are certainly methodologically interesting and important directions for the future and for related applications. In the current context, however, these kinds of modelling elaborations do not materially affect the major development and conclusions about the limitations of the existing calibration methods; the choice was therefore made to report summary conclusions based on rather simple models, for clarity.

Some questions are raised about the role of evolution variance components. Certainly inferences for the elements of W are accessible as posteriors are simulated in the core analysis. Extensions to include non-zero off-diagonal terms in the evolution variance matrix are feasible, though I have not implemented this nor experimented with these kinds of extensions. The practical relevance of such extensions is, *a priori*, questionable, though some experimentation is certainly needed. Dani states that “comparisons of outputs suggests that increased values of s^2 (and) v_d seem to point to more unstable cyclical representations.” One has to be careful here; the additional degrees of variation evident in posterior trajectories essentially reflects increased posterior uncertainties induced by allowing for uncertainties in observation times, rather than necessarily isolating real changes in model components.

The issues associated with the notion of time reversal are interesting and intriguing, and Dani’s development bears further study. However, I question the relevance of this thinking to the kinds of analyses reported in my paper. Dani’s point seems to be that various conditional distributions arising in DLM analyses (*all* DLM analyses, not just the specific kinds of models in my paper) should (or could?) satisfy various symmetry conditions under reversal of the direction of the time arrow. This seems to have been prompted by the fact that the palaeoclimatological data are presented and analysed with the time arrow actually reversed, as is typical in the geological community. I see no compelling reason for this suggestion; these conditional distributions change as their conditioning variates change, and that is naturally dependent on both structural

model form and assumed direction of the time arrow. On the other hand, it reasonable to expect and indeed demand that the structural components of the time series *model* itself be symmetric with respect to reversal of the time arrow, the traditional concept of reversibility in stationary time series. This is easily seen in the simplest case of a single cycle $x_t = \beta x_{t-1} - x_{t-2} + \omega_t$. In state space form, this is $z_t = Gz_{t-1} + \delta_t$ where

$$z_t = \begin{pmatrix} x_t \\ x_{t-1} \end{pmatrix}, G = \begin{pmatrix} \beta & -1 \\ 1 & 0 \end{pmatrix}, \delta_t = \begin{pmatrix} \omega_t \\ 0 \end{pmatrix}.$$

Inverting G gives $z_{t-1} = G^{-1}(z_t - \delta_t)$ and, after easy calculations, reading off the second element of z_{t-1} in this equation gives $x_{t-2} = \beta x_{t-1} - x_t + \omega_t$. Thus the model is explicitly time reversible, as required.

Dani provides final comments on potential extensions involving multivariate time series for several core records, a point raised in the concluding section of the paper. This is certainly going to be a lively area for further work, and most likely one of the most fruitful next steps. Core synchronisation difficulties can be somewhat ameliorated by simply modelling series from several cores (conditionally) independently, but with careful thought given to the structure and dependencies in prior distributions for timing patterns across cores. Dani's suggested hierarchical/dynamic model form has attractive features, though note that, in our context, wavelengths will be constrained to be common across series; if the models capture critical aspects of patterns of climatic change, then the wavelengths relate to driving climatic variables and so will appear as constant across series. Involvements of covariates, either via the kinds of model extensions Dani suggests or others, may well be of interest in allowing for spatial heterogeneity in relating cores sampled from dispersed sites in a lake or from more than one lake.

Finally let me thank Dani for his thoughtful contributions. If nothing else several directions for developments are identified here. In closing, let me echo Dani's remarks that the kinds of analyses presented illustrate context-driven modelling that combines component submodels for individual features of the problem, and resulting analyses are accessible through simulation. Future developments, anticipated or otherwise, to address the kinds of issues raised in discussion will similarly depend critically on model extension via addition of further components and their combination in simulation-based analysis.