Available online at www.sciencedirect.com

ScienceDirect

Nuclear Instruments and Methods in Physics Research B 259 (2007) 426–431

NIM B
Beam Interactions with Materials & Atoms

www.elsevier.com/locate/nimb

Bayesian spectral analysis of raw tree-ring IntCal04 data: No continuous sinusoids – some short duration sinusoids

V. Palonen *, P. Tikkanen

Accelerator Laboratory, University of Helsinki, Finland

Available online 8 February 2007

Abstract

An improved Bayesian model for detection of periodic signals is presented and applied to IntCal04 tree-ring data. Our previous model used a discrete-time autoregressive process to model the noise and here a continuous autoregressive process is implemented. In order to take into account the temporal width of the raw tree-ring data samples, the model function has been changed to a mean of the underlying signal for temporal interval of the datapoint. A wavelet-type variant of the model is also presented.

It is shown that the presence of continuous cycles in the raw tree-ring data is doubtful. There is however evidence for wavelet-type temporally constrained high-frequency oscillations with periods in the 2–20 year range. The temporal location of these oscillations is given. It is probable that even these oscillations result from the measurement offsets between the datasets used for calibration.

PACS: 95.75.Wx; 92.70.Gt; 29.85.+c; 02.70.Hm; 05.45.Tp

Keywords: Spectral analysis; IntCal04; Bayes; Continuous autoregressive process; Fourier transform; Signal detection; Reversible jump MCMC

1. Introduction

One possible way to try to understand the complex sun-earth system is to search for certain periodic structures within different proxies. The possible existence of climatic and solar frequencies in different weather records and proxies like temperature or rainfall records, tree ring thicknesses, lake varves, isotope concentrations in icecores, etc., is therefore an important question. Several analyses have been done on these proxies and spectral peaks have been found, but the significance of the peaks has remained open. This is due to problematic or nonexistent confidence levels, different spectra of other similar proxies, and lack of known mechanism to produce the frequencies [1,2]. There is also considerable evidence for red noise in the climatic time series [3,4] as a glance at power spectra of most proxies will reveal. For atmospheric radiocarbon concentration in particular, red noise arises when the flux of radiocarbon is taken to have white noise [5].

For the atmospheric radiocarbon concentration, our previous work has shown that although there are peaks in the power spectrum, there are no believable continuous sinusoids in the IntCal98 $\Delta^{14}C$ data [6–8]. With the availability of the new and revised calibration dataset [9], another search for sinusoids is in order. First, a revised Bayesian signal-detection model for raw tree-ring radiocarbon data is presented. The model takes the temporal length of the sample into account and uses a continuous autoregressive process as the correlation structure. Furthermore, a simple variant of the continuous-sinusoid model is presented which enables a wavelet-type analysis of the data. Second, the raw tree-ring data of IntCal04 data is analysed with both models.

2. Model

Each point in the raw tree-ring data for IntCal04 is a measurement on a wood sample covering 1–31 annual

* Corresponding author. Tel.: +358 9 19150010; fax: +358 9 19150042. *E-mail address:* vesa.palonen@helsinki.fi (V. Palonen).

0168-583X/$ - see front matter
doi:10.1016/j.nimb.2007.02.007

rings. Let $\mathbf{y} = [y_1, \ldots, y_N]^t$ be a vector of the N measured data values of blocks of tree-rings with the corresponding uncertainties $\mathbf{s} = [s_1, \ldots, s_N]^t$. Let $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_N]^t$ and $\boldsymbol{\omega} = [\omega_1, \ldots, \omega_N]^t$ be the starting and ending years of the measured blocks. The centerpoint in time of a block is denoted by $t = [t_1, \ldots, t_N]^t$. We take the measurements $\mathbf{y}$ to consist of a parametric function g_i and two types of noise, i.e.

$$y_i = g_i + e_i + x(t_i), \tag{1}$$

where g_i are values for the model function, e_i is the measurement error and $x(t)$ is a first-order continuous autoregressive [CAR(1)] process. The measurement errors e_i are distributed normally with known variances $e_i \sim N(0, s_i^2)$. The model function g_i is taken to be a mean of the annual values of sinusoids for the corresponding years:

$$g_i = \frac{1}{\omega_i - \beta_i} \sum_{t=\beta_i}^{\omega_i} \left(m_{k,l} + \delta_{l1} r_{k,l}(t - t_1) + \sum_{j=1}^{k} A_{j,k,l} \sin(2\pi f_{j,k,l} t + \phi_{j,k,l}) \right), \tag{2}$$

where k and l are model indicator parameters. k denotes the number of sinusoids and l the existence of a linear trend. The linear trend is included because a possibility has been raised that the radiocarbon half-life used could be incorrect [10]. An incorrect radiocarbon half-life would result in an exponential trend, which can be approximated with a linear trend with an error less than two percent for the present data. $A_{j,k,l}$, $\phi_{j,k,l}$ and $f_{j,k,l}$ are the amplitudes, phases and frequencies of the jth sinusoid in the model $M_{k,l} \cdot m_{k,l}$ is the mean value of the data and $r_{k,l}$ the slope of the linear trend. δ_{l1} is the Kronecker delta, giving a value 1 if $l = 1$ and 0 otherwise. For brevity, we will drop the subscripts k and l from the parameters.

A simple variant of the model above is a model where the sinusoid is constrained to a temporal interval. This is done by introducing two time markers τ and τ', the smaller of which represents the start, and the bigger the stop of a signal. This will enable us to see which part of the data the power spectral peaks are coming from and whether there are any plausible finite-duration sinusoid in the data. In this case the model function is

$$g_i = \frac{1}{\omega_i - \beta_i} \sum_{t=\beta_i}^{\omega_i} \left(m + \delta_{l1} r(t - t_1) + \sum_{j=1}^{k} \mathbb{I}_{\mathbf{D}}(t) A_j \sin(2\pi f_j t + \phi_j) \right), \tag{3}$$

where $\mathbf{D} = [\min(\tau_j, \tau_j'), \max(\tau_j, \tau_j')]$ and the indicator function $\mathbb{I}_{\mathbf{D}}(t) = 1$ if $t \in \mathbf{D}$ and 0 otherwise.

The CAR(1) process is a generalization of the discrete time AR(1) process to continuous time domain. It is a solution to the stochastic differential equation [11,12]

$$\mathrm{d}x(t) + \alpha x(t)\mathrm{d}t = \sigma \,\mathrm{d}W(t), \tag{4}$$

where $W(t)$ is a continuous random walk process (Wiener process) with $W(t_1) - W(t_2) \sim N(0, |t_1 - t_2|)$. The correlation coefficient $\alpha \in [0, \infty[$. The non-negative parameter σ is related to the variance of the CAR(1) process. The conditional moments of the CAR(1) process are given by [12,13]

$$E[x(t)|x(0)] = \mathrm{e}^{-\alpha t} x(0), \tag{5}$$

$$\mathrm{var}[x(t)|x(0)] = \frac{\sigma^2}{2\alpha}(1 - \mathrm{e}^{-2\alpha t}). \tag{6}$$

In our previous model [8] the measurement errors e_i were taken into account numerically by using a model parameter for each error. However, it is possible to calculate the effect of measurement error analytically as has been done in a bit different context in [14,5]. Computation is facilitated by the use of the differences

$$f_i = g_i - \mathrm{e}^{-\alpha \Delta t_i} g_{i-1} \tag{7}$$

$$z_i = y_i - \mathrm{e}^{-\alpha \Delta t_i} y_{i-1} = f_i + e_i - \mathrm{e}^{-\alpha \Delta t_i} e_{i-1} + v_i, \tag{8}$$

where $\Delta t_i = t_i - t_{i-1}$ and $v_i \sim N(0, \frac{1 - \mathrm{e}^{-2\alpha \Delta t_i}}{2\alpha} \sigma^2)$. In the following, we will use the vectors $\mathbf{z} = [z_2, \ldots, z_N]^t$ and $\mathbf{f} = [f_2, \ldots, f_N]^t$.

The variances and covariances of this Gaussian process are

$$\mathrm{var}(z_i|\boldsymbol{\Theta}) = s_i^2 + \mathrm{e}^{-2\alpha \Delta t_i} s_{i-1}^2 + \frac{1 - \mathrm{e}^{-2\alpha \Delta t_i}}{2\alpha} \sigma^2, \text{ and} \tag{9}$$

$$\mathrm{cov}(z_i, z_{i+1}|\boldsymbol{\Theta}) = -\mathrm{e}^{-\alpha \Delta t_{i+1}} s_i^2, \tag{10}$$

where $\boldsymbol{\Theta} = [k, l, \alpha, \sigma, m, r, \{A_1, \ldots, A_k\}, \{\phi_1, \ldots, \phi_k\}, \{f_1, \ldots, f_k\}]^t$. The covariances between the z_i's whose indices differ by more than one are zero. Let $\mathbf{C}$ be the $(N-1) \times (N-1)$ covariance matrix of the z_i's, given by Eqs. (9) and (10). The likelihood is then

$$p(\mathbf{z}|\boldsymbol{\Theta}) = (2\pi)^{N-1} |\mathbf{C}|^{-1/2} \exp\left[-\frac{1}{2} (\mathbf{z} - \mathbf{f})^t \mathbf{C}^{-1} (\mathbf{z} - \mathbf{f}) \right], \tag{11}$$

where $|\mathbf{C}|$ is the determinant of the matrix $\mathbf{C}$. Table 1 gives the priors used in the model. Obtaining the posterior is now straightforward.

Table 1
Priors for the model parameters

Parameter	Symbol	Prior type	Prior range
Number of sinusoids	k	Uniform	$[0, 1, \ldots, 50]$
Existence of linear trend	l	Uniform	0 or 1
CAR correlation coefficient	α	Uniform	$[0, 100]$
CAR standard deviation	σ	σ^{-1}	$[0.01, 100]$
Data mean	m	Uniform	$[-100, 300]$
Trend slope	r	Cauchy (0, 0.02)	
Amplitude	A_j	Γ (1.25, 15)	
Phase	ϕ_j	Uniform	$[0, 2\pi]$
Frequency	f_j	f_j^{-1}	$[1/12500, 1/2]$
Time marker 1	τ_j	Uniform	$[-10415, 1945.5]$
Time marker 2	τ_j'	Uniform	$[-10415, 1945.5]$