ORIGINAL PAPER

Bayesian interpolation of unequally spaced time series

Luis E. Nieto-Barajas · Tapen Sinha

Abstract A comparative analysis of time series is not feasible if the observation times are different. Not even a simple dispersion diagram is possible. In this article we propose a Gaussian process model to interpolate an unequally spaced time series and produce predictions for equally spaced observation times. The dependence between two observations is assumed a function of the time differences. The novelty of the proposal relies on parametrizing the correlation function in terms of Weibull and Log-logistic survival functions. We further allow the correlation to be positive or negative. Inference on the model is made under a Bayesian approach and interpolation is done via the posterior predictive conditional distributions given the closest m observed times. Performance of the model is illustrated via a simulation study as well as with real data sets of temperature and CO_2 observed over 800,000 years before the present.

Keywords Bayesian inference · EPICA · Gaussian process · Kriging · Survival functions.

1 Introduction

The majority of the literature on time series analysis assumes that the variable of interest, say X_t, is observed on a regular or equally spaced sequence of times, say $t \in \{1, 2, \ldots\}$ (e.g.

L. E. Nieto-Barajas (✉)
Department of Statistics, ITAM, Rio Hondo 1, Col. Progreso Tizapan, 01080 Mexico, D.F., Mexico
e-mail: lnieto@itam.mx

T. Sinha
Department of Actuarial Sciences, ITAM, Rio Hondo 1, Col. Progreso Tizapan, 01080 Mexico, D.F., Mexico
e-mail: tapen@itam.mx

Box et al. 2004; Chatfield 1989). Often, time series are observed at uneven times. Unequally spaced (also called unevenly or irregularly spaced) time series data occur naturally in many fields. For instance, natural disasters, such as earthquakes, floods and volcanic eruptions, occur at uneven intervals. As Eckner (2012) noted, in astronomy, medicine, finance and ecology observations are often made at uneven time intervals. For example, a patient might take a medication frequently at the onset of a disease, but the medication may be taken less frequently as the patient gets better.

The example that motivated this study consists of time series observations of temperature (Jouzel et al. 2007) and CO_2 (carbon dioxide) (Lüthi et al. 2008) which were observed over 800,000 years before the present. The European Project for Ice Coring in Antarctica (EPICA) drilled two deep ice cores at Kohnen and Concordia Stations. At the latter station, also called Dome C, the team of researchers produced climate records focusing on water isotopes, aerosol species and greenhouse gases. The temperature dataset consists of temperature anomaly with respect to the mean temperature of the last millennium, it was dated using the EDC3 timescale (Parrenin et al. 2007). Note that the actual temperature was not observed but inferred from deuterium observations. The CO_2 dataset consists of carbon dioxide concentrations, dated using EDC3_gas_a age scale (Loulergue et al. 2007).

The temperature and the CO_2 time series were observed at unequally spaced times. Any attempt of comparative or associative analysis between these two time series requires them to have both been measured at the same times. For example, regarding Figure 1.6 of the Intergovernmental Panel on Climate Change (IPCC) first assessment report (Houghton et al. 1990), it notes "over the whole period (160 thousand years before present to 1950) there is a remarkable correlation between polar temperature, as

deduced from deuterium data, and the CO_2 profile." However from the data no correlation was calculated. The reason is simple, a correlation measure is not possible unless the data from both series have the same dates. The aim of this paper is to propose a model to interpolate unequal spaced time series and produce equally spaced observations that will evaluate past climate research.

Our approach aims to produce equally spaced observations via stochastic interpolation. We propose a Gaussian process model with a novel correlation function parametrized in terms of a parametric survival function, which allows for positive or negative correlations. For the parameter estimation we follow a Bayesian approach. Once posterior inference on the model parameters is done, interpolation is carried out by using the posterior predictive conditional distributions of a new location given a subset of size m of neighbours, similar to what is done in spatial data known as Bayesian kriging (e.g. Handcock and Stein 1993, Bayraktar and Turalioglu 2005). The number of neighbours m, to be used to interpolate, is decided by the user. Renard et al. (2006) provided an excellent review of the Bayesian thinking applied to environmental statistics.

The rest of the paper is organized as follows: We first present a review of related literature in Sect. 2. In Section 3 we describe the model together with the prior to posterior analysis as well as the interpolation procedure. In Sect. 4 we compare our model with alternative models and assess the performance of our approach. Section 5 contains the data analysis of our example that motivated this study, and finally Sect. 6 contains concluding remarks.

Before we proceed we introduce some notations: N (μ, σ^2) denotes a normal density with mean μ and variance σ^2; Ga (a, b) denotes a gamma density with mean a/b; IGa (a, b) denotes an inversed gamma density with mean $b/(a-1)$; Un (a, b) denotes a continuous uniform density on the interval (a, b); Tri (a, c, b) denotes a triangular density on (a, b) and mode in c; and Ber (p) denotes a Bernoulli density with probability of success p.

2 Related literature

The study of unequally spaced time series has concentrated on two approaches: models for the unequally spaced observed data in its unaltered form, and models that reduce the irregularly observed data to equally spaced observations and apply the standard theory for equally spaced time series.

Within the former approach, Eckner (2012) defined empirical continuous time processes as piecewise constant or linear between observations. Others such as Jones (1981), Jones and Tryon (1987) and Belcher et al. (1994)

suggested an embedding into continuous time diffusions constructed by replacing the difference equation formulation by linear differential equations, defining continuous time autoregressive or moving average processes. Detailed reviews of the issue appeared in Harvey and Stock (1993) and Brockwell et al. (2007).

Within the latter approach there are deterministic interpolation methods (summarised in Adorf (1995)), or stochastic interpolation methods, the most popular method was the linear interpolation with a standard Brownian motion (e.g. Chang 2012, Eckner 2012).

In a different perspective, unevenly spaced time series have also been treated as equally spaced time series with missing observations. Aldrin et al. (1989), for instance, used a dynamic linear model and predicted the missing observations with the model. Alternatively, Friedman (1962) and more recently Kashani and Dinpashoh (2012) made use of auxiliary data to predict the missing observations. The latter one also compared the performance of several classical and artificial intelligence models. In contrast, our proposal does not rely on any auxiliary data, we make use of the past and future data of the same variable of interest.

More general spatio-temporal interpolation problems were addressed by Serre and Christakos (1999) using modern geostatistics techniques and in particular with the Bayesian maximum entropy method.

3 Model

3.1 Sampling model

Let $\{X_t\}$ be a continuous time stochastic process defined for an index set $t \in \mathcal{T} \subset \mathbb{R}$ and which takes values in a state space $\mathcal{X} \subset \mathbb{R}$. We will say that $X_{t_1}, X_{t_2}, \ldots, X_{t_n}$ is a sample path of the process observed at unequal times $t_1, t_2, \ldots, t_n$, and $n > 0$. In a time series analysis we only observe a single path and that is used to make inference about the model.

We assume that X_t follows a Gaussian process with constant mean μ and covariance function $\text{Cov}(X_s, X_t) = \Sigma(s, t)$. In notation

$$X_t \sim \text{GP}(\mu, \Sigma(s, t)). \tag{1}$$

We further assume that the covariance is a function only of the absolute times difference $|t - s|$. In this case it is said that the covariance function is isotropic (Rasmussen and Williams 2006). Typical choices of the covariance function include the exponential $\Sigma(s, t) = \sigma^2 e^{-\lambda|t-s|}$ and the squared exponential $\Sigma(s, t) = \sigma^2 e^{-\lambda(t-s)^2}$, with λ the decay parameter. These two covariance functions are particular cases of

a more general class of functions called Matérn functions (Stein 1999).

By assuming a constant marginal variance for each X_t, we can express the covariance function in terms of the correlation function $R(s, t)$ as $\Sigma(s, t) = \sigma^2 R(s, t)$. We note that isotropic correlation functions behave like survival functions as a function of the absolute time difference $|t - s|$. For instance if $S_\theta(t) = e^{-\lambda t^\alpha}$, with $\theta = (\lambda, \alpha)$, which corresponds to the survival function of a Weibull distribution, we can define a covariance function in terms of S_θ as $\Sigma(s, t) = \sigma^2 S_\theta(|t - s|)$. Moreover, for $\alpha = 1$ and $\alpha = 2$ we obtain the exponential and squared exponential covariance functions, respectively, mentioned above. Alternatively, if $S_\theta(t) = 1/(1 + \lambda t^\alpha)$, where again $\theta = (\lambda, \alpha)$ (which corresponds to survival function of a Log-logistic distribution) we can define another class of covariance functions.

The correlation between any two states of the process, say X_s and X_t, can take values in the whole interval $(-1, 1)$. However, the previous covariance functions only allow the correlation to be positive. We can modify the previous construction to allow for positive or negative correlations in the process. We propose

$$\Sigma_{\sigma^2, \theta, \beta}(s, t) = \sigma^2 S_\theta(|t - s|)(-1)^{\beta|t-s|}, \qquad (2)$$

with $\beta \in \{1, 2\}$ in such a way that $\beta = 1$ implies a negative/positive correlation for odd/even time differences $|t - s|$, and it is always positive regardless $|t - s|$ being odd or even, for $\beta = 2$. Note that $|t - s|$ needs to be an integer.

For the covariance function (2) to be well defined, it needs to satisfy a positive semi-definite (psd) condition. Since the product of two psd functions is also psd (Stein 1999, p.20), we can split (2) into two factors: $\sigma^2 S_\theta(|t - s|)$ and $(-1)^{\beta|t-s|}$. The first factor is a psd covariance function since it is a monotonically decreasing positive function. The second factor defines a variance-covariance matrix for any set of n points that is psd, with the first eigenvalue being n and the rest zero. Therefore (2) is psd and (1) is a well defined model.

3.2 Prior to posterior analysis

Let $\mathbf{x} = (x_{t_1}, x_{t_2}, \ldots, x_{t_n})$ be the observed unequally spaced time series, and $\boldsymbol{\eta} = (\mu, \sigma^2, \theta, \beta)$ the vector of model parameters. The joint distribution of the data $\mathbf{x}$ induced by model (1) is a n-dimensional multivariate normal distribution of the form

$$f(\mathbf{x} \mid \boldsymbol{\eta}) = (2\pi\sigma^2)^{-n/2} |R_{\theta,\beta}|^{-1/2}$$
$$\times \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{x} - \boldsymbol{\mu})' R_{\theta,\beta}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\}, \qquad (3)$$

where $\boldsymbol{\mu} = (\mu, \ldots, \mu)$ is the vector of means, $R_{\theta,\beta} = (r_{ij}^{\theta,\beta})$ is the correlation matrix with (i, j) term $r_{ij}^{\theta,\beta} = \Sigma_{\sigma^2, \theta, \beta}(t_i, t_j)/\sigma^2$ and $\Sigma_{\sigma^2, \theta, \beta}$ is given in (2).

We assign conditionally conjugate prior distributions for $\boldsymbol{\eta}$ when possible. In particular we take independent priors of the form $\mu \sim \mathrm{N}(\mu_0, \sigma_\mu^2)$, $\sigma^2 \sim \mathrm{IGa}(a_\sigma, b_\sigma)$, $\lambda \sim \mathrm{Ga}(a_\lambda, b_\lambda)$, $\alpha \sim \mathrm{Un}(0, A_\alpha)$, and $\beta - 1 \sim \mathrm{Ber}(p_\beta)$.

The posterior distribution of $\boldsymbol{\eta}$ will be characterised via the full conditional distributions. These are given below and their derivation is included in the Appendix.

(a) The posterior conditional distribution of μ is

$$f(\mu \mid \text{rest})$$
$$= \mathrm{N}\left(\mu \,\middle|\, \frac{\frac{1}{\sigma^2}\mathbf{x}' R_{\theta,\beta}^{-1} + \frac{\mu_0}{\sigma_\mu^2}}{\frac{1}{\sigma^2}\mathbf{1}' R_{\theta,\beta}^{-1}\mathbf{1} + \frac{1}{\sigma_\mu^2}}, \left(\frac{1}{\sigma^2}\mathbf{1}' R_{\theta,\beta}^{-1}\mathbf{1} + \frac{1}{\sigma_\mu^2}\right)^{-1}\right),$$

where $\mathbf{1}' = (1, \ldots, 1)$ of dimension n.

(b) The posterior conditional distribution of σ^2 is

$$f(\sigma^2 \mid \text{rest})$$
$$= \mathrm{IGa}\left(\sigma^2 \,\middle|\, a_\sigma + \frac{n}{2}, b_\sigma + \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' R_{\theta,\beta}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)$$

(c) The posterior conditional distribution of $\beta - 1$ is $\mathrm{Ber}(p_\beta^*)$ with

$$p_\beta^* = \left\{1 + \frac{f(\mathbf{x} \mid \beta = 1)(1 - p_\beta)}{f(\mathbf{x} \mid \beta = 2)p_\beta}\right\}^{-1}$$

(d) Finally, since there are no conditionally conjugate prior distributions for $\theta = (\lambda, \alpha)$, its posterior conditional distribution is simply $f(\theta \mid \text{rest}) \propto f(\mathbf{x} \mid \boldsymbol{\eta})f(\theta)$, where $f(\mathbf{x} \mid \boldsymbol{\eta})$ is given in (3).

Therefore, posterior inference will be obtained by implementing a Gibbs sampler (Smith and Roberts 1993). Sampling from conditionals (a), (b) and (c) is straightforward since they are of standard form. However, to sample from (d) we will require Metropolis-Hastings (MH) steps (Tierney 1994). If $\theta = (\lambda, \alpha)$, we propose to sample conditionally from $f(\lambda \mid \alpha, \text{rest})$ and $f(\alpha \mid \lambda, \text{rest})$.

For λ, at iteration $r + 1$ we take random walk MH steps by sampling proposal points $\lambda^* \sim \mathrm{Ga}(\kappa, \kappa/\lambda^{(r)})$. This proposal distribution has mean $\lambda^{(r)}$ and coefficient of variation $1/\sqrt{\kappa}$. By controlling κ we can tune the MH acceptance rate. For the examples below we set $\kappa = 10$ inducing an acceptance rate of around 30%. Such gamma random walk proposals for variables on the positive real line has been used successfully by other authors (e.g. Barrios et al. 2013) showing good behaviour. For α, we also take random walk steps from a triangular distribution with mode in $\alpha^{(r)}$, that is $\alpha^* \sim \mathrm{Tri}(\max\{0, \alpha^{(r)} - \kappa\}, \alpha^{(r)}, \min\{A_\alpha, \alpha^{(r)} + \kappa\})$. With $\kappa = 0.1$ the acceptance

rate induced in this case is around 32% for the examples considered below. In both cases, the acceptance rates are adequate according to (Robert and Casella 2010, ch.6).

Implementing these two MH steps for λ and α is a challenge since the probability of accepting the proposal draw requires to evaluate the n-multivariate normal distribution (3) in two different values of the parameter. When n is large, as is the case for the real data at hand, the computational time required to invert an n-dimensional matrix and to compute the determinant value increases at a rate $O(n^3)$. To reduce the computational time, we suggest that a subset of the time series data is to take, say 500, and use the subset as a training sample to estimate the parameters of the model. We observed that the loss in precision in the parameter estimation is minimal, and the computational time is highly reduced. (Rasmussen and Williams (2006), ch.8) provided a thorough discussion of different approximation methods.

3.3 Interpolation

Once the parameter estimation is done, that is, we have a sample from the posterior distribution of $\boldsymbol{\eta} = (\mu, \sigma^2, \theta, \beta)$, we can proceed with the interpolation.

Within the Bayesian paradigm, interpolation of the unequally spaced series, to produce an equally spaced series, is done via the posterior predictive distribution. We propose to interpolate using the posterior predictive conditional distribution given a subset of neighbours. Let $\mathbf{x_s} = (x_{s_1}, \ldots, x_{s_m})$ be a set of size m of observed points, such that $\mathbf{s} = (s_1, \ldots, s_m)$ are the m observed times nearest to time t, with $s_j \in \{t_1, \ldots, t_n\}$. If $m = n$, $\mathbf{x_s} = \mathbf{x}$ the whole observed time series. Therefore, the conditional distribution of the unobserved data point X_t given its closest m observations is given by

$$f(x_t \mid \mathbf{x_s}, \boldsymbol{\eta}) = \mathrm{N}\left(x_t \mid \mu_t, \sigma_t^2\right), \tag{4}$$

with

$$\mu_t = \mu + \Sigma(t, \mathbf{s})\Sigma(\mathbf{s}, \mathbf{s})^{-1}(\mathbf{x}_m - \boldsymbol{\mu}) \text{ and}$$
$$\sigma_t^2 = \sigma^2 - \Sigma(t, \mathbf{s})\Sigma(\mathbf{s}, \mathbf{s})^{-1}\Sigma(\mathbf{s}, t),$$

where, as before, $\Sigma(t, \mathbf{s}) = \Sigma(\mathbf{s}, t)' = \mathrm{Cov}(X_t, \mathbf{X_s})$ and $\Sigma(\mathbf{s}, \mathbf{s}) = \mathrm{Cov}(\mathbf{X_s}, \mathbf{X_s})$.

As in standard linear regression models, we assume that $(X_t, \mathbf{X_s})$ are a new set of variables (conditionally) independent of the observed sample $\mathbf{x}$. In this case, the posterior predictive conditional distribution of X_t given $\mathbf{x_s}$ is obtained by integrating out the parameter vector $\boldsymbol{\eta}$ from (4) using its posterior distribution, that is, $f(x_t \mid \mathbf{x_s}, \mathbf{x}) = \int f(x_t \mid \mathbf{x_s}, \boldsymbol{\eta}) f(\boldsymbol{\eta} \mid \mathbf{x}) \, \mathrm{d}\boldsymbol{\eta}$. This marginalization process is usually done numerically via Monte Carlo.

To have an idea how the interpolation process works, let us consider the specific case when $m = 2$, so that $\mathbf{x_s} = (x_{s_1}, x_{s_2})$ consists of the two closest observations to time t. Then, from (4) we obtain that $\mathrm{E}(X_t \mid \mathbf{x_s}, \boldsymbol{\eta})$ becomes

$$\mu_t = \mu + \frac{(\rho_{t,s_1} - \rho_{t,s_2}\rho_{s_1,s_2})(x_{s_1} - \mu) + (\rho_{t,s_2} - \rho_{t,s_1}\rho_{s_1,s_2})(x_{s_2} - \mu)}{1 - \rho_{s_1,s_2}^2}, \tag{5}$$

where $\rho_{t,s} = \mathrm{Corr}(X_t, X_s) = \Sigma(t, s)/\sigma^2$. The conditional expected value (5) is a linear function obtained as a weighted average of the neighbour observations $\mathbf{x_s}$ where the weights are given by the correlations among x_t, x_{s_1} and x_{s_2}. The higher the correlation between x_t and x_{s_j} the more important the observed value x_{s_j} is, in determining the conditional expected value of X_t, for $j = 1, 2$. Finally, the posterior predictive conditional expected value $\mathrm{E}(X_t \mid \mathbf{x_s}, \mathbf{x})$, which can be taken as the estimated interpolated point, is a mixture of expression (5) with respect to the posterior distribution of $\boldsymbol{\eta}$.

In general, for any $m > 0$, the estimated interpolated point at time t will be a weighted average of the closest m observed data points, as sliding windows, with weights determined by their respective correlations with the unobserved X_t.

In the literature, what is known as linear interpolation, arises when assuming a standard Brownian (or Wiener) process for X_t, i.e., X_t is a Gaussian process with mean 0 and variance-covariance function $\Sigma(s, t) = \min\{s, t\}$. This implies that the conditional distribution of X_t given its two closest observations $\mathbf{x_s} = (x_{s_1}, x_{s_2})$ is a normal distribution with mean μ_t and variance σ_t^2, where: $\mu_t = \{(s_2 - t)/(s_2 - s_1)\}x_{s_1} + \{(t - s_1)/(s_2 - s_1)\}x_{s_2}$ and $\sigma_t^2 = (t - s_1)(s_2 - t)/(s_2 - s_1)$, if $s_1 < t < s_2$; $\mu_t = (t/s_1)x_{s_1}$ and $\sigma_t^2 = (t/s_1)(s_1 - t)$, if $t < s_1 < s_2$; and $\mu_t = x_{s_2}$ and $\sigma_t^2 = t - s_2$, if $s_1 < s_2 < t$.

Note that the linear interpolation model just described does not rely on any parameter, so the estimated interpolated point μ_t is a weighted average of the two closest neighbours with weights given by the time distances, whereas our proposal relies on parameters that need to be estimated with the data and the estimated interpolated point is a weighted average with weights given by the correlations. Moreover, the linear interpolation only interpolates using the two closest neighbours while our proposal considers using the closest $m > 0$ neighbours.

4 Simulation study

To assess how good our model behaves we carried out a simulation study. Time series of length 1000 were generated under six different scenarios. We considered trended
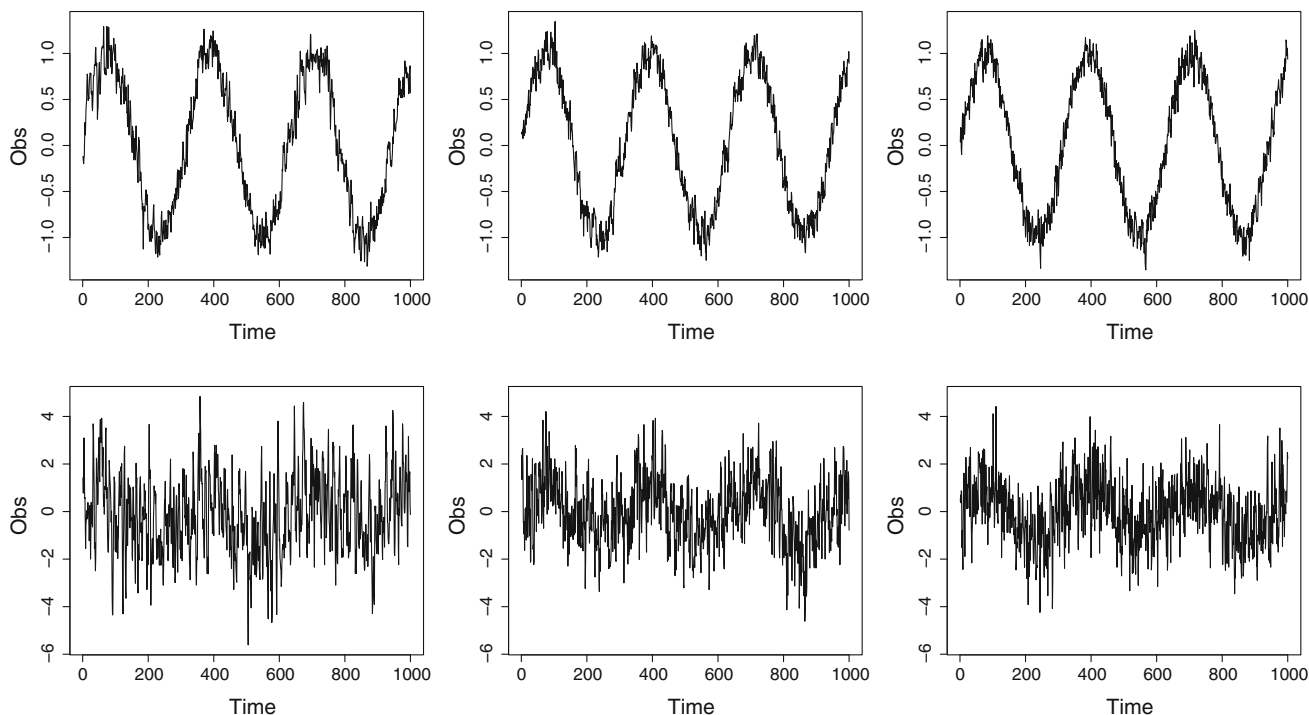
Fig. 1 Simulated data. 1,000 observations under 6 scenarios. In the columns $(p, q) \in \{(1, 1), (1, 0), (0, 1)\}$ and in the rows $\sigma_0^2 \in \{0.1, 1\}$

time series of the form $X_t = \sin(t/50) + \epsilon_t$, for $t = 1, \ldots, 1000$, where ϵ_t follows an ARMA(p, q) processes with N $(0, \sigma_0^2)$ errors and with autoregressive and moving average coefficients equal to 0.5. The six scenarios are the result of combining $(p, q) \in \{(1, 1), (1, 0), (0, 1)\}$ and $\sigma_0^2 \in \{0.1, 1\}$. Scenarios 1 to 3 correspond to the low variance and scenarios 4 to 6 to the high variance. The simulated data are shown in Fig. 1.

For each of the scenarios, we randomly chose a proportion $\pi 100\%$ of the observations to be the "observed data" and predicted the rest $(1 - \pi)100\%$ of the data with our Bayesian interpolation model. We varied $\pi \in \{0.1, 0.2, 0.5\}$.

The effectiveness of the interpolation is quantified via the L-measure (Ibrahim and Laud 1994), which is a function of the variance and bias of the predictive distribution. For the removed observations $x_1, \ldots, x_r$ we computed

$$L(v) = \frac{1}{r} \sum_{i=1}^{r} \text{Var}(x_i^F | \mathbf{x}) + \frac{v}{r} \sum_{i=1}^{r} \{ \text{E}(x_i^F | \mathbf{x}) - x_i \}^2,$$

where x_i^F is the predictive value of x_i and $v \in (0, 1)$ is a weighting term which determines a trade-off between variance and bias. The bias (B) itself can be obtained from the L-measure through $B = L(1) - L(0)$. Smaller values of the L-measure and the bias indicate better fit.

We compare the performance of our model with two other alternatives. One is the linear interpolation, which is the result of assuming a standard Brownian motion process in the data as described in the interpolation part of Sect. 3. The

other approach is to consider a random walk dynamic linear model (Harrison and Stevens 1976) and treating the unevenly observed data as an equally spaced time series with missing observations. The missing observations are to be predicted with the model. To be specific, the model has the form $x_t = \mu_t + v_t$ and $\mu_t = \mu_{t-1} + \omega_t$ with $v_t \sim \text{N}(0, \sigma_v^2)$ and $\omega_t \sim \text{N}(0, \sigma_\omega^2)$. This approach has been successfully used by Aldrin et al. (1989) to interpolate copper contamination levels in the river Gaula in Norway. We note that this latter model can also be seen as a Gaussian process model for integer times. Therefore the three competitors are Gaussian process models with different specifications.

To specify the dynamic linear model we took independent priors $\mu_1 \sim \text{N}(0, 1000)$, $\sigma_v^{-2} \sim \text{Ga}(0.001, 0.001)$ and $\sigma_\omega^{-2} \sim \text{Ga}(0.001, 0.001)$. To implement our model we specified the variance-covariance function (2) in terms of a Log-logistic survival function. For defining the priors we took $\mu_0 = 0$, $\sigma_\mu^2 = 100$, $a_\sigma = 2$, $b_\sigma = 1$, $a_\lambda = b_\lambda = 1$, $A_\alpha = 2$ and $p_\beta = 0.5$. Interpolations with our model were based on the $m = 10$ closest observed data points. In all cases we ran Gibbs samplers for 10,000 iterations with a burn-in of 1,000, keeping one iteration of every third for computing the estimates. The results are summarised in Table 1.

From Table 1 we can see that our model is the one that produces estimates with the smallest variance $(L(0))$ among the three methods, for all scenarios, and when the percentage of observed data is smaller (e.g. 10 and 20 %). As the percentage of observed data increases (e.g. 50 %)

Table 1 L-measures for $v \in \{0, 0.5, 1\}$ and bias (B) for the three competing models, for the six scenarios (S), and for three poportions of observed data π

S	$\pi = 10\%$				$\pi = 20\%$				$\pi = 50\%$			
	$L(0)$	$L(0.5)$	$L(1)$	B	$L(0)$	$L(0.5)$	$L(1)$	B	$L(0)$	$L(0.5)$	$L(1)$	B
Gaussian process model												
1	0.03	0.04	0.06	0.03	0.02	0.03	0.04	0.02	0.02	0.02	0.03	0.01
2	0.03	0.04	0.05	0.03	0.02	0.03	0.04	0.02	0.01	0.02	0.03	0.02
3	0.03	0.04	0.05	0.02	0.02	0.03	0.04	0.03	0.01	0.02	0.03	0.02
4	2.00	3.26	4.51	2.51	1.98	3.18	4.37	2.39	1.06	1.79	2.52	1.46
5	1.28	2.06	2.83	1.55	1.32	2.01	2.70	1.38	1.18	1.90	2.61	1.43
6	1.00	1.66	2.31	1.31	1.39	2.03	2.68	1.29	1.21	1.86	2.51	1.30
Standard Brownian process model												
1	3.25	3.26	3.28	0.03	1.74	1.75	1.76	0.02	0.82	0.83	0.83	0.01
2	3.25	3.26	3.27	0.02	1.74	1.75	1.76	0.02	0.82	0.83	0.83	0.01
3	3.25	3.26	3.27	0.02	1.74	1.75	1.76	0.02	0.82	0.83	0.83	0.01
4	3.26	4.81	6.37	3.12	1.74	2.96	4.19	2.45	0.82	1.48	2.14	1.31
5	3.25	4.22	5.18	1.93	1.74	2.48	3.22	1.47	0.82	1.42	2.01	1.19
6	3.25	4.30	5.34	2.09	1.74	2.56	3.38	1.64	0.82	1.47	2.11	1.29
Dynamic linear model												
1	0.03	0.04	0.06	0.03	0.02	0.03	0.04	0.02	0.01	0.02	0.03	0.01
2	0.02	0.03	0.04	0.02	0.02	0.02	0.03	0.02	0.01	0.02	0.02	0.01
3	0.03	0.03	0.04	0.02	0.02	0.02	0.03	0.02	0.01	0.02	0.02	0.01
4	2.22	3.46	4.71	2.49	2.42	3.60	4.78	2.36	1.07	1.74	2.41	1.34
5	1.37	2.13	2.89	1.52	1.35	2.02	2.69	1.34	1.27	1.89	2.51	1.24
6	1.06	1.87	2.68	1.62	1.40	2.04	2.69	1.29	1.29	1.93	2.56	1.27

our proposal produces estimates with slightly higher bias (B), but in general the bias is almost the same for the three methods. Surprisingly, the linear interpolation, based on a parameterless standard Brownian process, produced the best estimates for the high error variance scenarios 4, 5 and 6, and when the proportion of observed data is 50 %. For the other cases, the dynamic linear model is the second best model.

In summary, we can say that our method is superior for sparser datasets with large spacing in the observations, and for the other cases our method produced comparable results to the other two approaches.

5 Data analysis

We now apply our stochastic interpolation method to the dataset that motivated this study. The temperature dataset consists of 5,788 observations, whereas the CO2 dataset contains 1,095 data points spread in an observations range of 800,000 years before present. The data are shown in the two top panels of Fig. 2. The two series show similar patterns that visually seem to be highly correlated. However, a simple Pearson's correlation coefficient is impossible to compute since the observation times are different. To see the observation time differences between the two

series, we computed the observed times lags, i.e., $t_i - t_{i-1}$ and plotted them versus t_i, for $i = 1, \ldots, n$. These are shown in the two bottom panels in Fig. 2, respectively.

As can be seen from Fig. 2, the temperature series was measured very frequently close to the present time and becomes less frequently observed as we move away from the present. On the other hand, the CO_2 series shows an inverted behaviour, being measured less frequently close to the present and measured relatively more frequently in older years. The temperature series was observed with a minimum of 8 years of difference, a maximum of 1364 years and a median time of 58 years, whereas the CO_2 series was observed every 9 years as a minimum, 6029 years as a maximum and with a median observation time of 586 years.

To compare, we implemented our model with the covariance function defined in terms of both, the Weibull and Log-logistic survival functions. We took the same prior specifications as in the simulation study, i.e., $\mu_0 = 0$, $\sigma_\mu^2 = 100$, $a_\sigma = 2$, $b_\sigma = 1$, $a_\lambda = b_\lambda = 1$, $A_\alpha = 2$ and $p_\beta = 0.5$. Posterior inference is not sensitive to these choices because the amount of data is so large that it overrides any prior information. Since we do not have the real values we want to predict as in the simulation study, we computed the L-measure for the observed data points instead. Additionally, we vary the number of neighbour
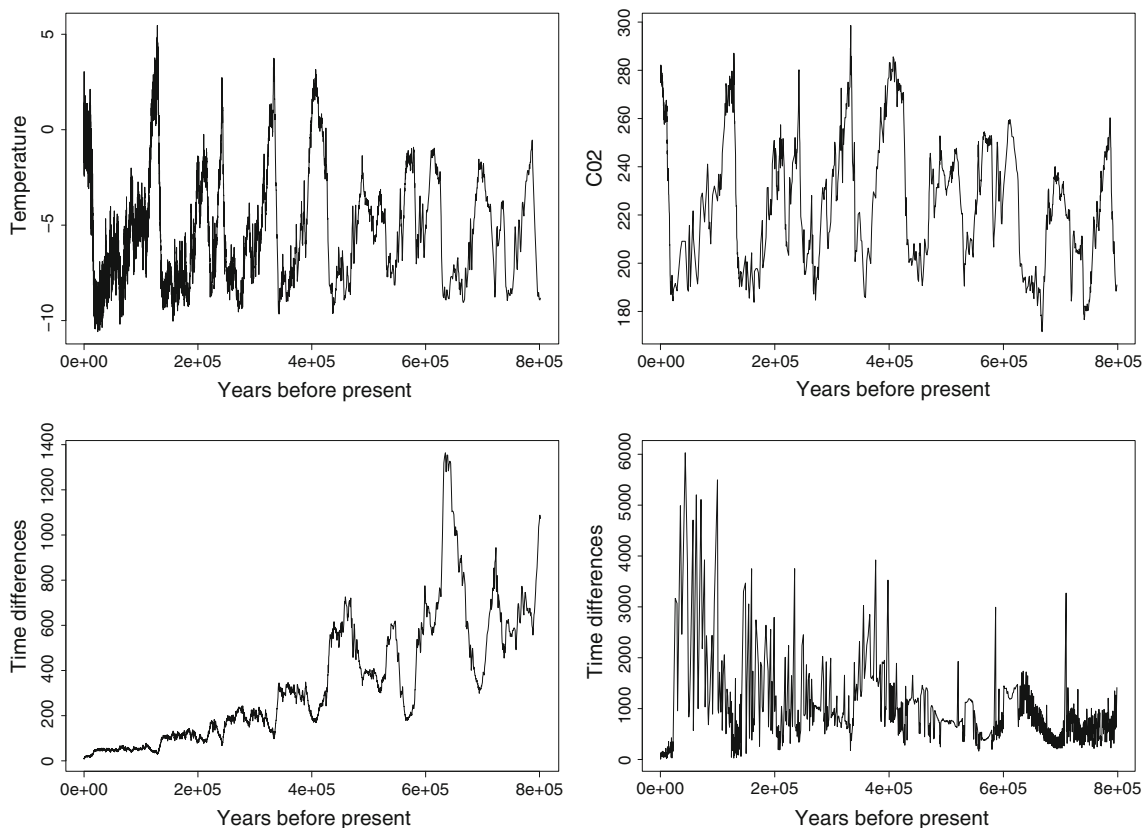
Fig. 2 Time series plots. Temperature (*top left*), CO_2 (*top right*); observed time differences, in the temperature series (*bottom left*), and in the CO_2 series (*bottom right*)

Table 2 L-measures for $v \in \{0, 0.5, 1\}$ and Bias (B) for the three competing models and for the six scenarios (S)

m	Weibull				Log-logistic			
	$L(0)$	$L(0.5)$	$L(1)$	B	$L(0)$	$L(0.5)$	$L(1)$	B
Temperature								
2	1.479	1.991	2.504	1.025	1.476	1.990	2.505	1.029
4	1.450	1.974	2.498	1.045	1.446	1.975	2.505	1.059
10	1.442	1.964	2.486	1.043	1.427	1.959	2.491	1.064
20	1.439	1.969	2.498	1.059	1.424	1.960	2.496	1.072
CO_2								
2	60.572	87.100	113.628	53.056	59.651	86.232	112.812	53.161
4	60.395	87.069	113.744	53.349	59.907	86.519	113.130	53.223
10	60.353	87.353	113.554	53.202	59.134	85.664	112.195	53.061
20	60.422	86.537	113.252	52.830	59.740	86.149	112.558	52.818

observations $m \in \{2, 4, 10, 20\}$ to compare among the different predictions obtained. The numbers are reported in Table 2. In all cases Gibbs samplers were run for 20,000 iterations with a burn-in of 2,000 and keeping one draw every 10th iteration to reduce autocorrelation in the chain. Convergence of the chain was assessed by comparing the ergodic means plots of model parameters for multiple chains, which justified that the number of iterations was

adequate (Cowles and Carlin 1996). The numbers below are based on 1,800 draws.

Several conclusions can be derived from Table 2. In all cases, the variance ($L(0)$) attains a minimal value for $m = 10$. As a function of m, the bias (B) steadily increases for the temperature dataset but decreases for the CO_2 dataset. This opposite tendency can be due to the fact that the temperature data have more data points than the CO_2
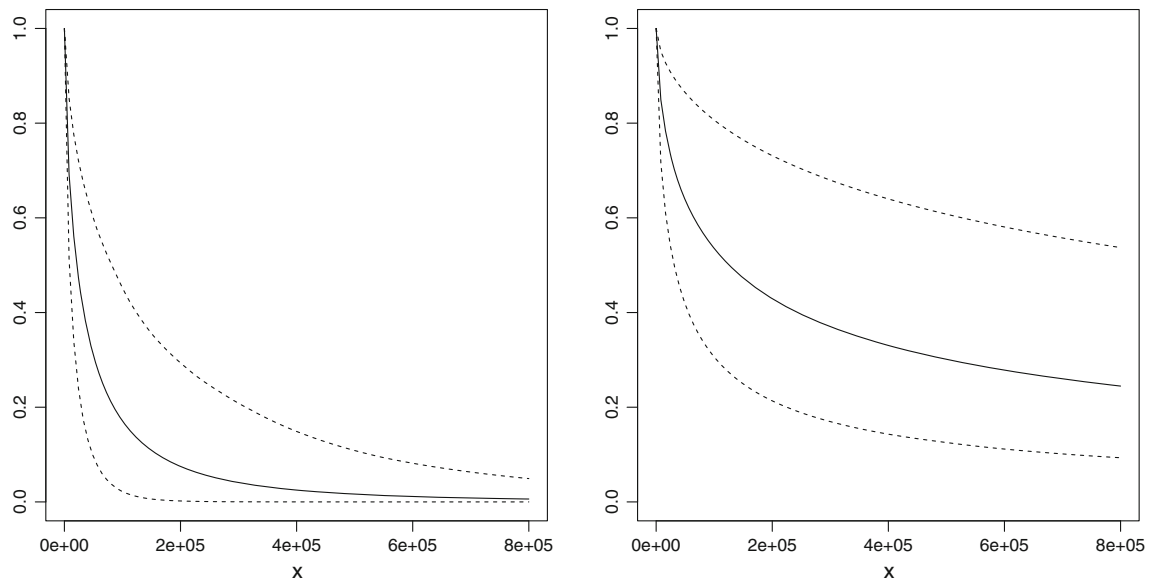
Fig. 3 Correlation function estimates in the temperature dataset. Weibull (*left*) and Log-logistic (*right*)
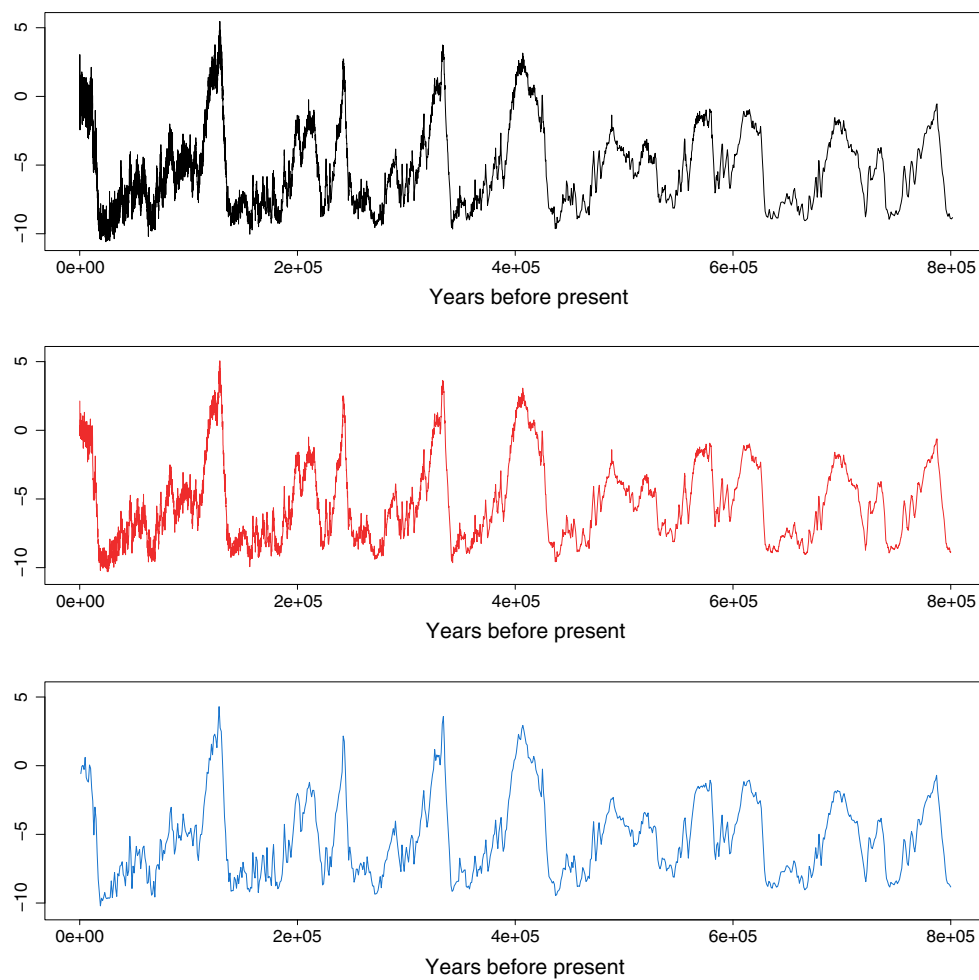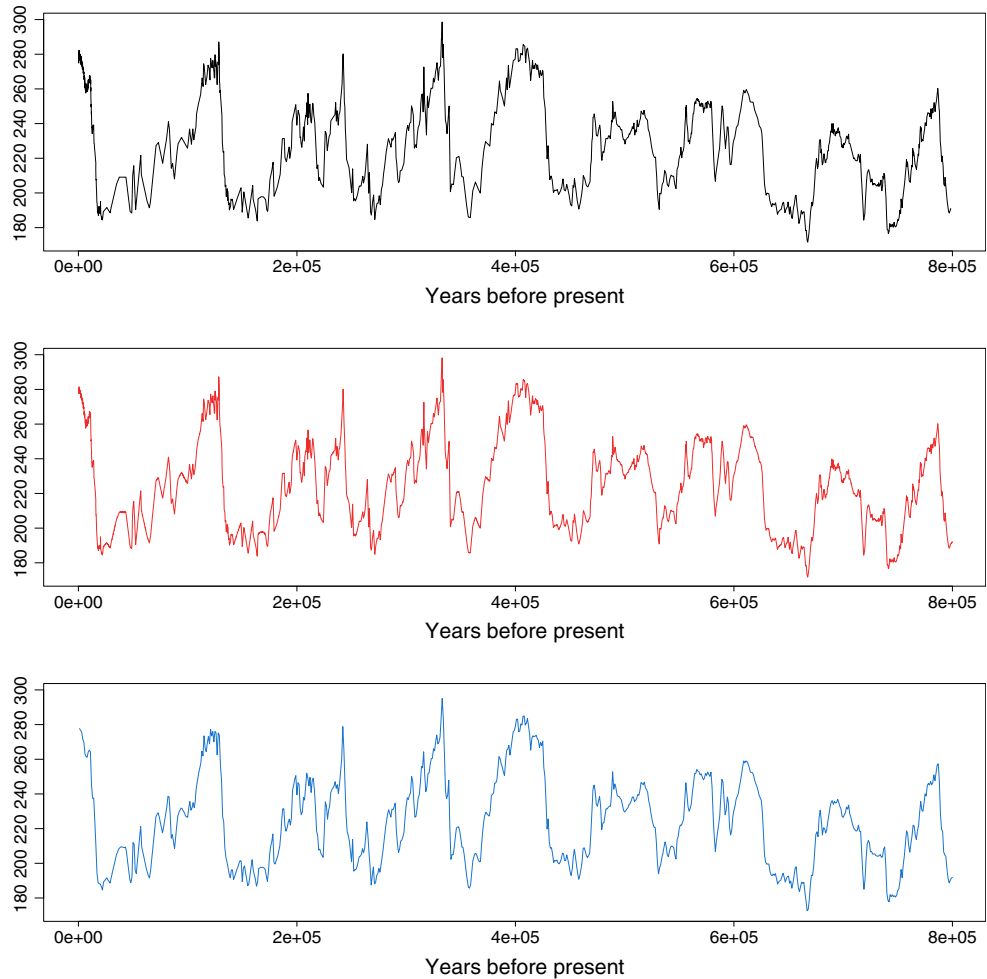

Fig. 4 Temperature interpolation. Observed data (*top*), interpolation every 1,000 years (*middle*) and every 100 years (*bottom*)

Fig. 5 CO_2 interpolation. Observed data (*top*), interpolation every 1,000 years (*middle*) and every 100 years (*bottom*)

data so prediction is based on observations that are farther apart in the CO_2 data.

Comparing the Weibull versus the Log-logistic choices, we can say that they provide almost the same predictions. The bias is slightly smaller for the Weibull, but the variance is slightly larger. In Fig. 3 we present posterior estimates for the two correlation functions induced by these two alternatives. The Log-logistic function (right panel) shows a slower rate of decay than the Weibull function (left panel), allowing for higher correlation in observations further apart. A similar result occurs for the CO_2 dataset. Although our model allows for positive and negative covariance, in both datasets posterior estimates of the parameter β is 2 which implies a positive correlation within observations. We decided to choose the Log-logistic model for producing interpolations with $m = 10$ because it achieves the smallest variance.

We computed interpolated time series with a regular spacing of 100 and 1000 years. They are included in Figs. 4 and 5 for temperature and CO_2, respectively. In each figure, the top graph corresponds to the original observed data, the middle graph to the interpolation with 100 years

spacing and the bottom graph to the interpolation of 1,000 years gap. Visually we can see that the interpolation with 100 years and the original data are practically the same, whereas for the higher spacing (1,000 years) the interpolation is smoother, losing some extreme (peak) values. We refrained from including prediction intervals in the figures to show the path of the series better. Posterior intervals are so narrow they make the graph difficult to read. In fact, the 95 % probability intervals have an average length of 1.90 for the temperature, and 12.90 for the CO_2 datasets, with negligible differences between interpolations at 100 and 1,000 years.

Finally, we present an empirical association analysis between the two interpolated series at 100 years spacing. For each time t we have 1,800 simulated values from the predictive distribution for both series, temperature and CO_2. For each draw we computed three association coefficients: Pearson's correlation, Kendall's tau and Spearman's rho (e.g. Nelsen 2006). The sample means are 0.86, 0.66 and 0.85, with 95% probability intervals, based on quantiles, (0.858, 0.872), (0.647, 0.667) and (0.843, 0.859), respectively. The three measures can take values in the interval

$[-1, 1]$, so they all imply a positive moderate to strong dependence between the two time series. Further studies of causality are needed to see whether the temperature changes in the planet are caused by the CO_2 levels.

6 Conclusions

We have developed a new stochastic interpolation method based on a Gaussian process model with the correlation function defined in terms of a parametric survival function. We allow for positive or negative auto correlation to better capture the nature of time dependence of the data. Inference of the model is Bayesian. Therefore all uncertainties are taken into account in the prediction (interpolation) process. Within the Bayesian paradigm, interpolation is done by posterior predictive conditional distributions. This approach allows us to produce inference beyond point interpolation. Our proposal for interpolation is so flexible that the user can decide the number of neighbouring observations m to be used to interpolate. We have shown that our model is a good competitor against the other two alternatives, producing smaller variance without loss in bias.

We illustrated our method with two real sets of data. The first is the temperature and the second is the CO_2 datasets, both from the Dome C in Antarctica. As we noted in the introduction, the IPCC report of 1990 talked about the "correlation" between the temperature and the CO2 data for the past 160,000 years without ever actually calculating the correlation itself. The same state of affairs continues to the latest report of the IPCC. For example, Masson-Delmonte et al. (2013) simply report the Dome C data separately for the temperature and the CO_2 data but never together as the data does not allow for any proper bivariate analysis. Our method provides a way out of this dilemma. Not only will this method of reconstruction allow us to calculate simple correlation it will also allow us to do time series analysis in the time domain. This means, we will be able to work out stationary of the dataset for the whole period or any subperiod we choose to study. This will allow us to analyse the data to deduce if there is causality using Granger's casualty test. The latter methodology has been successfully applied in economics. For example, Sims (1972) applied the method to ask whether changes in money supply cause inflation to rise. For his work, Sims won the Nobel Memorial Prize in Economic Sciences in 2011. A similar analysis could provide a method to calculate the short run and the long run impact of CO_2 (or any other green house gas) on surface temperatures.

Our method can also be used for unequally spaced time series in financial data. For example, in illiquid stock markets, many stocks do not trade daily. Thus, calculating "beta" of a stock when it is not traded regularly is a problem. There are "standard" methods like that proposed by Scholes and Williams (1977). That essentially uses a simple average between three observations with an assumed autocorrelation structure. Our procedure is more general and, in addition to a point estimate, it also provides interval estimation. There are other areas such as astronomy and medical sciences where the method described could be applied when unequally spaced time series naturally occur.

Appendix: Derivation of the full conditional distributions

The basic idea is to write the likelihood (3) as a function of the parameter in turn.

(a) For μ: The likelihood function is proportional to

$$\text{lik}(\mu \mid \text{rest}) \propto \exp\left\{-\frac{1}{2\sigma^2}\left(\mu^2 \mathbf{1}' R_{\theta,\beta}^{-1}\mathbf{1} - 2\mu \mathbf{x}' R_{\theta,\beta}^{-1}\right)\right\}.$$

The prior distribution is proportional to

$$f(\mu) \propto \exp\left\{-\frac{1}{2\sigma_\mu^2}\left(\mu^2 - 2\mu\mu_0\right)\right\}.$$

Multiplying these two factors we get that the conditional posterior is proportional to

$$f(\mu \mid \text{rest}) \propto \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2}\mathbf{1}' R_{\theta,\beta}^{-1}\mathbf{1} + \frac{1}{\sigma_\mu^2}\right)\right.$$
$$\left. \times\left(\mu^2 - 2\mu\frac{\mathbf{x}' R_{\theta,\beta}^{-1}/\sigma^2 + \mu_o/\sigma_\mu^2}{\mathbf{1}' R_{\theta,\beta}^{-1}\mathbf{1}/\sigma^2 + 1/\sigma_\mu^2}\right)\right\}.$$

Completing the perfect square trinomial we identify the kernel as a normal density. Filling in the normalising constants we obtain the full conditional.

(b) For σ: Removing the constants $(2\pi)^{-n/2}|R_{\theta,\beta}|$ from likelihood (3) and by noting that the prior is proportional to

$$f(\sigma^2) \propto (\sigma^2)^{-(a_\sigma+1)}e^{-b_\sigma/\sigma^2},$$

we then combine the likelihood with the prior and identify the kernel of an inverse gamma density. Filling in the normalising constants we get the full conditional.

(c) For β: Noting that β can only take two values, $\{1, 2\}$, we evaluate the likelihood in each of them and

multiply by the prior probabilities $1 - p_\beta$ and p_β, respectively. Renormalising the posterior probabilities we obtain the full conditional.

References

Adorf H-M (1995) Interpolation of irregularly sampled data series - a survey. In: Shaw RA, Payne HE, Hayes JJE (eds) Proceedings of the astronomical data analysis software and systems IV, Conference series 77

Aldrin M, Damsleth E, Saebo HV (1989) Time series analysis of unequally spaced observations with applications to copper contamination of the river Gaula in central Norway. Environ Monit Assess 13:227–243

Bayraktar H, Turalioglu FS (2005) A Kriging-based approach for locating a sampling sitein the assessment of air quality. Stoch Environ Res Risk Assess 19:301–305

Barrios E, Lijoi A, Nieto-Barajas LE, Prünster I (2013) Modeling with normalized random measure mixture models. Stat Sci 28:313–334

Belcher J, Hampton JS (1994) Parameterization of continuous time autoregressive models for irregularly sampled time series data. J R Stat Soc Ser B 56:141–155

Brockwell P, Davis R, Yang Y (2007) Continuous-time Gaussian autoregression. Stat Sin 17:63–80

Box GEP, Jenkins GM, Reinsel GC (2004) Time series analysis: forecasting and control. Wiley, New York

Cowles MK, Carlin BP (1996) Markov chain Monte Carlo convergence diagnostics: a comparative review. J Am Stat Assoc 91:883–904

Chang JT (2012) Stochastic Processes. Technical report, Department of Statistics, Yale University

Chatfield C (1989) The analysis of time series: an introduction. Chapman and Hall, London

Eckner A (2012) A framework for the analysis of unevenly spaced time series data. Preprint. Available at: http://www.eckner.com/papers/unevenly_spaced_time_series_analysis.

Friedman M (1962) The interpolation of time series by related series. J Am Stat Assoc 57:729–757

Handcock MS, Stein ML (1993) A Bayesian analysis of kriging. Technometrics 35:403–410

Harrison PJ, Stevens PF (1976) Bayesian forecasting. J R Stat Soc Ser B 38:205–247

Harvey AC, Stock JH (1993) Estimation, smoothing, interpolation, and distribution for structural time-series models in discrete and continuous time. In Models, Methods and Applications of Econometrics PCB Phillips (ed.) Cambridge, Blackwell: 55–70.

Houghton JT, Jenkins GJ, Ephraums JJ (1990) Climate change: The IPCC scientific assessment. Report prepared for Intergovernmental Panel on Climate Change by working group I. Cambridge University Press, Cambridge.

Ibrahim JG, Laud PW (1994) A Predictive approach to the analysis of designed experiments. J Am Stat Assoc 89:309–319

Jones RH (1981) Fitting a continuous time autoregression to dicrete data. In: Findley DF (ed) Applied time series analysis II. Academic Press, New York, pp 651–682

Jones RH, Tryon PV (1987) Continuous time series models for unequally spaced data applied to modelling atomic clocks. SIAM J Sci Stat Comput 8:71–81

Jouzel J et al (2007) Orbital and Millennial Antartic climate variability over the past 800,000 years. Science 317:793–796

Kashani MH, Dinpashoh Y (2012) Evaluation of efficiency of different estimation methods for missing climatological data. Stoch Environ Res Risk Assess 26:59–71

Loulergue L et al (2007) New constraints on the gas age—ice age difference along the EPICA ice cores, 0–50 kyr. Clim Past 3:527–540

Lüthi D et al (2008) High resolution carbon dioxide concentration record 650,000–800,000 years before present. Nature 453:379–381

Masson-Delmotte VM et al (2013) Information from paleoclimate archives. In: Stocker TF et al. (eds) Climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, pp 383–464

Nelsen RB (2006) An introduction to copulas. Springer, New York

Parrenin F et al (2007) The EDC3 chronology for the EPICA Dome C ice core. Clim Past 3:485–497

Rasmussen CE, Williams CKI (2006) Gaussian processes for machine learning. MIT press, Massachusetts

Renard B, Lang M, Bois P (2006) Statistical analysis of extreme events in a non-stationary context via a Bayesian framework. Stoch Environ Res Risk Assess 21:97–112

Robert C, Casella G (2010) Introducing Monte Carlo methods with R. Springer, New York

Scholes M, Williams J (1977) Estimating betas from nonsynchronous data. J Financ Econ 5:309–328

Serre ML, Christakos G (1999) Modern geostatistics: computational BME analysis in the light of uncertain physical knowledge—the Equus Beds study. Stoch Environ Res Risk Assess 13:1–26

Sims CA (1972) Money, income and causality. Am Econ Rev 62:540–552

Smith A, Roberts G (1993) Bayesian computations via the Gibbs sampler and related Markov chain Monte Carlo methods. J R Stat Soc Ser B 55:3–23

Stein ML (1999) Interpolation of spatial data: Some theory for kriging. Springer, New York

Tierney L (1994) Markov chains for exploring posterior distributions. Ann Stat 22:1701–1722