

**SUPPLEMENT TO “RECONSTRUCTING PAST
TEMPERATURES FROM NATURAL PROXIES AND
ESTIMATED CLIMATE FORCINGS USING SHORT- AND
LONG-MEMORY MODELS”**

BY LUIS BARBOZA, BO LI, MARTIN TINGLEY AND FREDERI VIENS

ONLINE SUPPLEMENT A

A.1. Background on long memory models. As mentioned in the introduction, long-memory estimation is typically difficult both in theory and in practice; fundamental stochastic analysis research on this question is ongoing. In discrete time series, the long-memory autoregressive moving average (ARMA) and autoregressive conditional heteroskedasticity (ARCH) models that are popular in financial econometrics (see [Granger and Joyeux \(1980\)](#) and [Baillie, Bollerslev and Mikkelsen \(1996\)](#) for variations of ARMA and ARCH), cannot be adapted to our Bayesian context. These models are complex and are not explicitly specified in terms of distributions; they often rely on long data series or high-frequency in-fill data. It can be argued that the only models for which the discrete-time parametric memory-length estimation is straightforward and reliable in finite samples are those that come from continuous-time underlying models which exhibit self-similarity (identical distribution at all time scales), of which the typical example is fractional Brownian motion (fBm). Pioneering work in this direction was performed in [Coeurjolly \(2001\)](#), and further details and developments were provided in [Chronopoulou, Viens and Tudor \(2009\)](#). Long-memory structures which are slightly more complex than those studied therein are already difficult to estimate.

For instance, the so-called fractional Ornstein-Uhlenbeck (fOU) process has a well-known estimator for its drift parameter in continuous time ([Kleptsyna and Le Breton, 2002](#)), and with discrete data for increasing horizon asymptotics ([Tudor and Viens, 2007](#)), but the question of estimating its long-memory parameter remains elusive. When in-fill asymptotics can be attained in the data this can be accomplished via path regularity in continuous time (see [Istas and Lang \(1997\)](#) and its application in [Brouste and Iacus \(2012\)](#)). Neither this, nor the use of increasing-horizon asymptotics in [Tudor and Viens \(2007\)](#) are applicable in our paleoclimatology situation, because our observation frequency is not high enough and our calibration period is too short.

A.1.1. *Properties of fBm and fGn.* Fractional Gaussian noise (fGn) is defined as the first-order increment sequence of fBm. Information on fBm and fGn can be found in standard references; [Biagini et al. \(2010\)](#) provides an excellent mathematical treatment. The following is a short summary of properties relevant to our study. Fractional Brownian motion (fBm), $B^H = \{B_t^H : t \geq 0\}$ with Hurst parameter $H \in (0, 1)$, is a continuous centered Gaussian process with covariance function:

$$E[B_t^H B_s^H] = \frac{1}{2} (t^{2H} + s^{2H} - |t - s|^{2H})$$

for $t, s > 0$. From this definition we immediately deduce fundamental properties:

- $B_0^H = 0$,
- B^H has stationary increments:

$$E[(B_{t+h}^H - B_{s+h}^H)^2] = E[(B_t^H - B_s^H)^2] = |t - s|^{2H},$$

but these increments are not independent (see long-range dependence property below), except if $H = 1/2$, in which case $B_t^{1/2}$ is the standard Brownian motion.

- The increment process N^H defined by $N_t^H := B_{t+\Delta t}^H - B_t^H$ is called a fractional Gaussian noise (fGn), for fixed time step $\Delta t > 0$. N^H is a stationary, mean-zero Gaussian process, with constant variance $(\Delta t)^{2H}$ and autocovariance function given by:

$$(A.1) \quad E[N_{s+t}^H N_s^H] = \frac{1}{2} [|t + \Delta t|^{2H} + |t - \Delta t|^{2H} - 2|t|^{2H}].$$

For convenience, Δt can be interpreted as one unit of time, and t as an integer. In our study, t is measured in years.

We say that a stationary process $\{X_t, t \in \mathbb{N}\}$ is a “long-range dependent” process, or “long-memory” process, if its autocovariance function $\rho(t) = E[X_{s+t} X_s]$ satisfies:

$$\lim_{t \rightarrow \infty} \frac{\rho(t)}{t^{-M}} = c_1$$

for some constant $c_1 > 0$ and $M \in (0, 1)$, so that $\sum_{n=1}^{\infty} \rho(t) = \infty$. For a process such as fBm, which is not stationary, but has stationary increments, we say that it has long-range dependence (long memory) if its increment process has long range dependence. There is an alternative condition for

long-range dependence in terms of the spectral behavior of the stationary process X_n :

$$\lim_{\lambda \rightarrow 0} \frac{f(\lambda)}{|\lambda|^{-D}G(|\lambda|)} = c_2$$

where $c_2 > 0$, $D \in (0, 1)$ and G is a slow varying function at 0. It turns out that fGn (and thus fBm) have long-range dependence for $H > \frac{1}{2}$. Indeed, for these processes we have $M = 2 - 2H \in (0, 1)$ and $D = 2H - 1 \in (0, 1)$ in the above limits, with the constants $c_1 = H(2H - 1)$ and $c_2 = \pi / (H\Gamma(2H) \sin(\pi H))$. For further details, see Chapter 5 in [Nualart \(2010\)](#) and Chapter 7 in [Samorodnitsky and Taqqu \(1994\)](#).

A.1.2. Hypothesis testing for possible long-memory.

Robinson's test ([Robinson, 1995](#)) Define the semiparametric Gaussian estimate of H as:

$$\hat{H} = \operatorname{argmin}_H \left\{ \log \left(\frac{1}{m} \sum_{j=1}^m \lambda_j^{2H-1} I(\lambda_j) \right) - (2H - 1) \frac{1}{m} \sum_{j=1}^m \log \lambda_j \right\}$$

where $\lambda_j = 2\pi j/n$ for $j = 1, \dots, m$, $m = \frac{n}{2}$, and $I(\cdot)$ is the Periodogram estimator constructed using the observations X_1, X_2, \dots, X_n . Assume X 's spectral density satisfies $f(\lambda) \sim G\lambda^{1-2H}$ (for some fixed H). Under rather weak distributional assumptions on X , [Robinson \(1995\)](#) proves that \hat{H} is a consistent estimator of H , and is asymptotic normal in the sense that $2m^{\frac{1}{2}}(\hat{H} - H)$ converges to a standard normal distribution as $m \rightarrow \infty$. This asymptotic behavior implies that we can test the following null and alternative hypotheses (Robinson's test):

(A.2) Null Hypothesis: $H = 0.5$ (no memory)

Alternative: $H > 0.5$ (presence of memory)

Davies and Harte's test ([Davies and Harte, 1987](#)) Assume that

$X = (X_1, \dots, X_n)$ are normally distributed with a autocorrelation function given in (A.1). In order to test the same alternatives as Robinson's test (A.2), Davies and Harte derived a locally optimal test with critical region:

$$X'[c(I - 11'/n) - (I - 11'/n)T(I - 11'/n)]X < 0$$

where

$$T_{i,j} = \begin{cases} 0 & i = j, \\ \log 2 & i = j \pm 1, \\ \frac{1}{2} \left[|i - j| \log(1 - (i - j)^{-2}) + \log \frac{|i-j|+1}{|i-j|-1} \right] & \text{otherwise,} \end{cases}$$

and c is a constant chosen to attain a desired significance level in the test.

Goodness of fit test (Beran, 1992) Let X_t be a stationary Gaussian process with spectral density $f(\lambda)$ satisfying $f(\lambda) \sim G\lambda^{1-2H}$ (assume H is known). Let X_1, \dots, X_n be the n observations of X . If $f(\lambda, H)$ is the spectral density of the fGn with Hurst parameter H , then we want to test, for all λ , the Null Hypothesis $f(\lambda) = f(\lambda, H)$ against the Alternative: $f(\lambda) \neq f(\lambda, H)$. If we take \hat{H} as the Whittle estimator of H (see Beran, 1994), then Beran proposes the statistic:

$$T_n(\hat{H}) = \frac{A_n(\hat{H})}{B_n^2(\hat{H})}$$

with $A_n(\hat{H}) = \frac{4\pi}{n} \sum_{k=1}^{n^*} \left[\frac{I(s_j)}{f(s_j, \hat{H})} \right]^2$ and $B_n(\hat{H}) = \frac{4\pi}{n} \sum_{k=1}^{n^*} \frac{I(s_j)}{f(s_j, \hat{H})}$, where $s_j = 2\pi j/n$, $I(\cdot)$ is the periodogram as in Robinson's test, and n^* is the integer part of $\frac{1}{2}(n-1)$. Under mild technical assumptions on the distribution of X 's, the Null distribution T_n is asymptotically normal with mean π^{-1} and variance $2\pi^{-2}n^{-1}$, which results in Beran's test.

A.2. Multitaper Estimator. The general theory of spectral estimation, including the multitaper estimator, can be found in Thomson (1982) and Percival and Walden (1993) among others. Here we summarize relevant results concerning multitaper estimation. Suppose we have observations X_1, X_2, \dots, X_N of a stationary process $\{X_t\}$ with mean 0, variance σ^2 and spectral density function $S(\cdot)$. The observations are separated in time by Δt . The direct spectral estimator is defined as:

$$\hat{S}^{(d)}(f) = \Delta t \left| \sum_{t=1}^N h_t X_t e^{-i2\pi f t \Delta t} \right|^2$$

where $\{h_t\}$ is a sequence of real-valued constants (tapers) and $f \in [-\frac{1}{2\Delta t}, \frac{1}{2\Delta t}]$. The Fourier transform of $\{h_t\}$ will be denoted as $\mathcal{H}(f)$. It is not hard to conclude the following (defining $f_N = \frac{1}{2\Delta t}$):

$$E[\hat{S}^{(d)}(f)] = \int_{-\frac{1}{2\Delta t}}^{\frac{1}{2\Delta t}} \mathcal{H}(f-f') S(f') df'$$

where $\mathcal{H}(f) = \frac{|H(f)|^2}{\Delta t}$. If we ask h_t to be normalized ($\sum_{t=1}^N h_t^2 = 1$), then in the case of a white noise with variance σ^2 and $S(t) = \sigma^2 \Delta t$ we have:

$$E[\hat{S}^{(d)}(f)] = \sigma^2 \Delta t \int_{-f_N}^{f_N} \mathcal{H}(f) df = \sigma^2 \Delta t = S(t).$$

Hence the direct spectral estimator is unbiased. Moreover, in the case of $\{h_t = \frac{1}{\sqrt{N}}\}$, we have the usual Periodogram estimator with:

$$\begin{aligned} \mathcal{H}(f) &= N \Delta t \mathcal{D}_N^2(f \Delta t) \\ &= \frac{\Delta t \sin^2(N \pi f \Delta t)}{N \sin^2(\pi f \Delta t)} \end{aligned}$$

where $\mathcal{D}_N(\cdot)$ is the Dirichlet's kernel. In this case $\mathcal{H}(f)$ is usually denoted as $\mathcal{F}(f)$. $\mathcal{F}(f)$ has interesting properties, one of them is that it acts like a Dirac delta function with an infinite spike at $f = 0$ when $N \rightarrow \infty$. This implies:

$$\lim_{N \rightarrow \infty} E[\hat{S}^{(p)}(f)] = S(f).$$

One of the main problems of the periodogram is that series with large spectral range¹ have a large bias due to the existence of significant sidelobes on $\mathcal{F}(f)$. This kind of bias is known as spectral leakage. The main idea of tapering is to provide smaller sidelobes in \mathcal{H} than \mathcal{F} has. A possible way to solve this problem is by means of the Slepian tapers. The main idea is to solve the concentration problem:

$$\max_{\{h_t\}} \beta^2(W)$$

where $W > 0$ is fixed, and assuming $\Delta t = 1$,

$$\beta^2(W) = \int_{-W}^W |\mathcal{H}(f)|^2 df / \int_{-1/2}^{1/2} |\mathcal{H}(f)|^2 df.$$

This problem leads to the solution:

$$h_{t,0} = \frac{1}{\lambda_0(N, W)} \int_{-W}^W U_0(f, N, W) e^{i2\pi f[t - \frac{N-1}{2}]} df$$

¹the spectral range can be defined as $\log \frac{\max S(f)}{\min S(f)}$

for $t = 0, \pm 1, \dots$; where $U_0(\cdot, N, W)$ and $\lambda_0(\cdot, N, W)$ are the first eigenvector and eigenvalue of the system:

$$\int_{-W}^W N \mathcal{D}_N(f' - f) U(f) df = \lambda U(f)$$

In general $U_k(\cdot, N, W)$ is called the k -th order discrete prolate spheroidal wave function and h_k is the Slepian sequence of k -th order:

$$h_{t,k} = \frac{(-1)^k}{\epsilon_k \lambda_k(N, W)} \int U_k(f, N, W) e^{i2\pi f[t - \frac{N-1}{2}]} df$$

We then define the multitaper spectral estimator by:

$$\hat{S}^{(mt)}(f) = \frac{1}{K} \sum_{k=0}^{K-1} \hat{S}_k^{(mt)}(f)$$

where

$$\hat{S}_k^{(mt)}(f) = \Delta t \left| \sum_{t=1}^N h_{t,k} X_t e^{-i2\pi f t \Delta t} \right|^2$$

The multi taper estimator can be improved by considering a set of weights $b_k(f)$ such that the spectral leakage is minimized. It is possible to define such weights by considering an approximation of the Fourier transform of the time series; see [Thomson \(1982\)](#) and [Percival and Walden \(1993\)](#) for further details. The weights are then computed using an iterative procedure based on minimization of the spectral leakage. The algorithm generally converges after a few iterations (see [Percival and Walden, 1993](#)) and finally it is possible to define the Adaptive Multitaper spectral estimator as:

$$S^{(amt)}(f) = \frac{\sum_{k=0}^{K-1} b_k^2(f) \lambda_k \hat{S}_k^{mt}(f)}{\sum_{k=0}^{K-1} b_k^2(f) \lambda_k}$$

In order to calculate the multitaper estimators in [Figure 2](#) and [3](#) in the main document, we employed an estimator with $K = 10$ (number of tapers) and $NW = 5$. Hence, the design bandwidth (see [Walden, McCoy and Percival \(1995\)](#)) is approximately $2 \times 0.0602 = 0.1204$, which is a good approximation of the effective bandwidth of the multitaper estimation in the long-memory case.

A.3. Scoring Rules. [Gneiting and Raftery \(2007\)](#) define scoring rules as “summary measures for the evaluation of probability forecasts, by assigning a numerical score based on the predictive distribution and on the event or value that materializes”. Scoring rules are designed to reward both calibration quality and the sharpness of predictive distributions. Sharpness is understood as the concentration of the predictive distribution and is related to forecast precision (see [Gneiting, Balabdaoui and Raftery \(2007\)](#)).

“Proper scoring rules” are those with expected value that is minimized when the predictive distribution is equal to the true distribution, and are considered optimal (for more technical details, see [Gneiting and Raftery, 2007](#)). There are several examples of proper rules, and they depend on the type of data analyzed: for categorical data, the Brier, Spherical and Logarithmic scores; for continuous variables, the Continuous Ranked Probability score and Energy score; and for quantiles, the Interval score. See (see [Gneiting and Raftery, 2007](#)) for details and additional proper scoring rules.

In the current analysis we obtain samples from the posterior distribution of the temperature anomalies, and consider two proper scoring rules to assess these predictive distributions.

A.3.1. Continuous Ranked Probability score (CRPS). This score is defined in terms of the predictive distribution function (F) directly (see [Gneiting, Balabdaoui and Raftery, 2007](#)):

$$CRPS(F, x) = \int_{-\infty}^{\infty} \{F(y) - \mathbf{1}(y \geq x)\}^2 dy$$

where x is the observed realization of the true distribution. [Gneiting and Raftery \(2007\)](#) give an alternative representation as:

$$CRPS(F, x) = E_F |X - x| - \frac{1}{2} E_F |X - X'|$$

where E_F is the expectation with respect to the predictive distribution F and X' is an independent copy of X with the same law F . The main advantage of this representation is that under R MCMC samples of the predictive F (y_1, \dots, y_R) and an independent set of R MCMC samples from the same distribution ($\tilde{y}_1, \dots, \tilde{y}_R$), we can approximate the $CRPS$ as ([Gschlöbl and Czado \(2007\)](#)):

$$\frac{1}{R} \sum_{i=1}^R \left[|y_i - x| - \frac{|y_i - \tilde{y}_i|}{2} \right].$$

If the realization x_t depends on time t (as well as the predictive distribution F_t), we can compute the mean CRPS score as:

$$(A.3) \quad CRPS = \frac{1}{M} \sum_{t=1}^M CRPS(F_t, x_t).$$

A.3.2. Interval score. The interval score is a special case of the quantile score (See equation 42 in [Gneiting and Raftery, 2007](#)). These scores are useful when we are able to compute exact or approximate quantiles of the predictive distribution F . This is the main reason why we were able to calculate this score under the normality assumption of Mann et al. (2008), as well as under our MCMC posterior calculations for comparison purposes. In the MCMC case, we compute the empirical $\alpha/2$ and $1 - \alpha/2$ quantiles of the posterior distribution F_t , let us call them l_t and u_t respectively. Hence the interval score at level $1 - \alpha$ is defined as ([Gschlößl and Czado \(2007\)](#)):

$$IS_{(1-\alpha)\%}(l_t, u_t, x_t) = \begin{cases} 2\alpha(u_t - l_t) + 4(l_t - x_t) & \text{if } x_t \leq l_t \\ 2\alpha(u_t - l_t) & \text{if } l_t \leq x_t \leq u_t \\ 2\alpha(u_t - l_t) + 4(x_t - u_t) & \text{if } x_t \geq u_t. \end{cases}$$

and we can compute the mean IS at level $1 - \alpha$ using the same formula as [\(A.3\)](#).

A.4. MCMC Posterior Distributions. In what follows, the notation $[X|Y]$ indicates the conditional probability of the random variable X given Y . Suppose we have the following prior distributions:

- $\alpha \sim N_2(\hat{\alpha}, \hat{D}_\alpha)$
- $\beta \sim N_4(\hat{\beta}, \hat{D}_\beta)$
- $\sigma_P^2 \sim IG(q_1, r_1)$
- $\sigma_T^2 \sim IG(q_2, r_2)$
- $H \sim \text{Unif}(0, 1)$
- $K \sim \text{Unif}(0, 1)$

where all the prior parameters are treated as known constants. For the temperature reconstructions in this work, we set $\hat{\alpha} = (0, 1)$, $\hat{\beta} = (0, 1, 1, 1)$, $\hat{D}_\alpha = I_2$, $\hat{D}_\beta = I_4$, $q_1 = q_2 = 2$, $r_1 = r_2 = 0.1$.

We first write the model of Eq. [\(3.4\)](#) in matrix form:

$$\begin{cases} \mathbf{RP} & = \alpha_0 + \alpha_1 \mathbf{T} + \sigma_P \eta(H) = \tilde{\mathbf{T}}\alpha + \sigma_P \eta(H) \\ \mathbf{T} & = \beta_0 + \beta_1 \mathbf{S} + \beta_2 \tilde{\mathbf{V}} + \beta_3 \tilde{\mathbf{C}} + \sigma_T \epsilon(K) = \mathbf{F}\beta + \sigma_T \epsilon(K) \end{cases}$$

for vectors \mathbf{RP} and \mathbf{T} of size $M \times 1$. Also assume that $\mathbf{T} = [\mathbf{T}_1, \mathbf{T}_2]$ where \mathbf{T}_1 is a vector of size $M - L \times 1$ and \mathbf{T}_2 has size $L \times 1$ where $L < M$. Let Σ_H and Σ_K be the covariance matrices of $\eta(H)$ and $\epsilon(K)$ respectively.

After some calculations we can deduce the following posterior distributions:

Distribution of $\mathbf{T}|\mathbf{RP}, \theta$.

$$\begin{aligned} [\mathbf{T}|\mathbf{RP}, \theta] &\propto \exp \left[-\frac{1}{2} \left[\mathbf{T}^T \left(\frac{\alpha_1^2}{\sigma_P^2} \Sigma_H^{-1} + \frac{\Sigma_K^{-1}}{\sigma_T^2} \right) \mathbf{T} - 2 \left(\frac{\alpha_1}{\sigma_P^2} (\mathbf{RP} - \alpha_0)^T \Sigma_H^{-1} + \frac{\beta^T \mathbf{F}^T \Sigma_K^{-1}}{\sigma_T^2} \right) \mathbf{T} \right] \right] \\ &\propto \exp \left[-\frac{1}{2} [T^T \Omega^{-1} T - 2\Delta^T \Omega \Omega^{-1} T + \Delta \Omega \Omega^{-1} \Omega \Delta^T] \right] \end{aligned}$$

where $\Omega^{-1} = \frac{\alpha_1^2}{\sigma_P^2} \Sigma_H^{-1} + \frac{\Sigma_K^{-1}}{\sigma_T^2}$ and $\Delta = \frac{\alpha_1}{\sigma_P^2} (\mathbf{RP} - \alpha_0)^T \Sigma_H^{-1} + \frac{\beta^T \mathbf{F}^T \Sigma_K^{-1}}{\sigma_T^2}$. Finally:

$$\mathbf{T}|\mathbf{RP}, \theta \sim N_M(\Delta^T \Omega, \Omega).$$

Distribution of $\alpha|\mathbf{RP}, \mathbf{T}, \sigma_P^2, H$. For ease of calculation, assume that α is independent of \mathbf{T} and σ_P^2 and H at the prior level. Then:

$$\begin{aligned} [\alpha|\mathbf{RP}, \mathbf{T}, \sigma_P^2, H] &\propto \exp \left[-\frac{1}{2} \left[\alpha^T \left(\frac{1}{\sigma_P^2} \tilde{\mathbf{T}}^T \Sigma_H^{-1} \tilde{\mathbf{T}} + \hat{D}_\alpha^{-1} \right) \alpha - 2 \left(\frac{\mathbf{RP}^T \Sigma_H^{-1} \tilde{\mathbf{T}}}{\sigma_P^2} + \hat{\alpha}^T \hat{D}_\alpha^{-1} \right) \alpha \right] \right] \\ &\propto \exp \left[-\frac{1}{2} [\alpha^T \Omega_\alpha^{-1} \alpha - 2\Delta_\alpha \Omega_\alpha^{-1} \alpha + \Delta_\alpha \Omega_\alpha \Omega_\alpha^{-1} \Omega_\alpha \Delta_\alpha^T] \right] \end{aligned}$$

where $\Omega_\alpha^{-1} = \frac{1}{\sigma_P^2} \tilde{\mathbf{T}}^T \Sigma_H^{-1} \tilde{\mathbf{T}} + \hat{D}_\alpha^{-1}$ and $\Delta_\alpha = \frac{\mathbf{RP}^T \Sigma_H^{-1} \tilde{\mathbf{T}}}{\sigma_P^2} + \hat{\alpha}^T \hat{D}_\alpha^{-1}$. Hence:

$$\alpha|\mathbf{RP}, \mathbf{T}, \sigma_P^2, H \sim N_2(\Delta_\alpha \Omega_\alpha, \Omega_\alpha).$$

Distribution of $\beta|\mathbf{T}, \sigma_T^2, K$. For ease of calculation, let us assume that β is independent of σ_T^2 and K at the prior level. Then:

$$\begin{aligned} [\beta|\mathbf{T}, \sigma_T^2, K] &\propto \exp \left[-\frac{1}{2} \left[\beta^T \left(\frac{1}{\sigma_T^2} \mathbf{F}^T \Sigma_K^{-1} \mathbf{F} + \hat{D}_\beta^{-1} \right) \beta - 2 \left(\frac{\mathbf{T}^T \Sigma_K^{-1} \mathbf{F}}{\sigma_T^2} + \hat{\beta} \hat{D}_\beta^{-1} \right) \beta \right] \right] \\ &\propto \exp \left[-\frac{1}{2} [\beta^T \Omega_\beta^{-1} \beta - 2\Delta_\beta \Omega_\beta^{-1} \beta + \Delta_\beta \Omega_\beta \Omega_\beta^{-1} \Omega_\beta \Delta_\beta^T] \right] \end{aligned}$$

where

$$(A.4) \quad \Omega_\beta^{-1} = \frac{1}{\sigma_T^2} \mathbf{F}^T \Sigma_K^{-1} \mathbf{F} + \hat{D}_\beta^{-1}$$

and $\Delta_\beta = \frac{\mathbf{T}^T \Sigma_K^{-1} \mathbf{F}}{\sigma_T^2} + \hat{\beta} \hat{D}_\beta^{-1}$. Therefore:

$$\beta | \mathbf{T}, \sigma_T^2, K \sim N_4(\Delta_\beta \Omega_\beta, \Omega_\beta).$$

Distribution of $\sigma_P^2 | \mathbf{R}\mathbf{P}, \mathbf{T}, \alpha, H$.

$$\begin{aligned} [\sigma_P^2 | \mathbf{R}\mathbf{P}, \mathbf{T}, \alpha, H] &= \frac{\exp \left[-\frac{1}{2\sigma_P^2} (\mathbf{R}\mathbf{P} - \tilde{\mathbf{T}}\alpha)^T \Sigma_H^{-1} (\mathbf{R}\mathbf{P} - \tilde{\mathbf{T}}\alpha) \right]}{\sqrt{(2\pi)^M |\Sigma_H \sigma_P^2|}} (\sigma_P^2)^{-q_1-1} \exp \left[-\frac{r_1}{\sigma_P^2} \right] \\ &\propto (\sigma_P^2)^{-(q_1+1)} \exp \left[-\frac{r'_1}{\sigma_P^2} \right] \end{aligned}$$

where $q'_1 = q_1 + \frac{M}{2}$ and $r'_1 = r_1 + \frac{1}{2} (\mathbf{R}\mathbf{P} - \tilde{\mathbf{T}}\alpha)^T \Sigma_H^{-1} (\mathbf{R}\mathbf{P} - \tilde{\mathbf{T}}\alpha)$. Then:

$$\sigma_P^2 | \mathbf{R}\mathbf{P}, \mathbf{T}, \beta, H \sim IG(q'_1, r'_1).$$

Distribution of $\sigma_T^2 | \mathbf{T}, \beta, K$.

$$\begin{aligned} [\sigma_T^2 | \mathbf{T}, \beta, H] &= \frac{\exp \left[-\frac{1}{2\sigma_T^2} (\mathbf{T} - \mathbf{F}\beta)^T \Sigma_K^{-1} (\mathbf{T} - \mathbf{F}\beta) \right]}{\sqrt{(2\pi)^M |\Sigma_K \sigma_T^2|}} \cdot (\sigma_T^2)^{-q_2-1} \exp \left[-\frac{r_2}{\sigma_T^2} \right] \\ \text{(A.5)} \quad &\propto (\sigma_T^2)^{-(q_2+1)} \exp \left[-\frac{r'_2}{\sigma_T^2} \right] \end{aligned}$$

where $q'_2 = q_2 + \frac{M}{2}$ and $r'_2 = r_2 + \frac{1}{2} (\mathbf{T} - \mathbf{F}\beta)^T \Sigma_K^{-1} (\mathbf{T} - \mathbf{F}\beta)$. Therefore:

$$\sigma_T^2 | \mathbf{T}, \beta, K \sim IG(q'_2, r'_2).$$

Distribution of $\mathbf{T}_1 | \mathbf{T}_2, \theta$. Let $\Delta^T \Omega = [\mu_{1T}, \mu_{2T}]$ and $\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}$ the corresponding decomposition of Ω and $\Delta^T \Omega$ according to the dimensions of $\mathbf{T} = [\mathbf{T}_1, \mathbf{T}_2]$. If $\tilde{\mathbf{T}}_2$ is a realization (or observations) of the random variable \mathbf{T}_2 then:

$$\mathbf{T}_1 | \mathbf{T}_2, \theta \sim N(\bar{\mu}, \bar{\Omega})$$

where $\bar{\mu} = \mu_{1T} + \Omega_{12} \Omega_{22}^{-1} (\tilde{\mathbf{T}}_2 - \mu_{2T})$ and $\bar{\Omega} = \Omega_{11} - \Omega_{12} \Omega_{22}^{-1} \Omega_{21}$.

Distribution of $H|\mathbf{RP}, \mathbf{T}, \alpha, \sigma_P^2$. After similar calculations as the previous cases, the posterior distribution of H turns out to be proportional to:

$$[H|\mathbf{RP}, \mathbf{T}, \alpha, \sigma_P^2] \propto [\det(\Sigma_H)]^{-1/2} \exp \left[-\frac{1}{2\sigma_P^2} (\mathbf{RP} - \tilde{\mathbf{T}}\alpha)^T \Sigma_H^{-1} (\mathbf{RP} - \tilde{\mathbf{T}}\alpha) \right]$$

where we are assuming that the prior of H does not depend on α and σ_P^2 . It is immediate that the posterior does not have a known distribution, then a Metropolis-Hastings is preferable for the sampling process. We assume the proposal distribution at repetition i ($1 < i \leq R$, R : total number of repetitions of the MCMC):

$$p(H|H^{(i-1)}) = TN(H^{(i-1)}, S_a(i), 0, K^{(i)})$$

where $H^{(i)}$ is the i -th sample of H , $K^{(i)}$ is the i -th sample of K , $S_a(i) = (D_0 - D_1) \cdot AR_i + D_1$ (AR_i is the acceptance ratio of H at time i and $D_0 > D_1 > 0$) and $TN(\mu, \sigma, a, b)$ stands for a Truncated Normal distribution with mean μ , standard deviation σ and lower and upper limit a and b respectively.

Distribution of $K|\mathbf{T}, \beta, \sigma_T^2$. After some calculations the posterior distribution of K is proportional to:

$$[K|\mathbf{T}, \beta, \sigma_T^2] \propto [\det(\Sigma_K)]^{-1/2} \exp \left[-\frac{1}{2\sigma_T^2} (\mathbf{T} - \mathbf{F}\beta)^T \Sigma_K^{-1} (\mathbf{T} - \mathbf{F}\beta) \right]$$

The Metropolis-Hastings algorithm is employed here using the proposal distribution:

$$p(K|K^{(i-1)}) = TN(K^{(i-1)}, S_a(i), 0, 1)$$

where as before $1 < i \leq R$ and $K^{(i)}$ is the i -th member of the simulated Metropolis-Hastings sample. $S_a(i)$ is given as the previous section, but it depends now on the acceptance ratio of the current sample of K .

REFERENCES

- BAILLIE, R. T., BOLLERSLEV, T. and MIKKELSEN, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **74** 3–30.
- BERAN, J. (1992). A Goodness-of-Fit Test for Time Series with Long Range Dependence. *Journal of the Royal Statistical Society. Series B (Methodological)* **54**.
- BERAN, J. (1994). *Statistics for long-memory processes. Monographs on Statistics and Applied Probability*, 61. Chapman & Hall.
- BIAGINI, F., HU, Y., ØKSENDAL, B. and ZHANG, T. (2010). *Stochastic Calculus for Fractional Brownian Motion and Applications (Probability and Its Applications)*, 1st ed. Springer.

- BROUSTE, A. and IACUS, S. M. (2012). Parameter estimation for the discretely observed fractional Ornstein-Uhlenbeck process and the Yuima R package. *Computational Statistics* **28** 1529–1547.
- CHRONOPOULOU, A., VIENS, F. G. and TUDOR, C. A. (2009). Variations and Hurst index estimation for a Rosenblatt process using longer filters. *Electronic Journal of Statistics* **3** 1393–1435.
- COEURJOLLY, J.-F. (2001). Estimating the Parameters of a Fractional Brownian Motion by Discrete Variations of its Sample Paths. *Statistical Inference for Stochastic Processes* **4** 199–227.
- DAVIES, R. B. and HARTE, D. S. (1987). Tests for Hurst Effect. *Biometrika* **74** 95–101.
- GNEITING, T., BALABDAOUI, F. and RAFTERY, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69** 243–268.
- GNEITING, T. and RAFTERY, A. E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association* **102** 359–378.
- GRANGER, C. W. J. and JOYEUX, R. (1980). An Introduction to Long-Memory Time Series Models and Fractional Differencing. *Journal of Time Series Analysis* **1** 15–29.
- GSCHLÖSSL, S. and CZADO, C. (2007). Spatial modelling of claim frequency and claim size in non-life insurance. *Scandinavian Actuarial Journal* **2007** 202–225.
- ISTAS, J. and LANG, G. (1997). Quadratic variations and estimation of the local Hölder index of a Gaussian process. *Annales de l'Institut Henri Poincaré (B) Probability and Statistics* **33** 407–436.
- KLEPTSYNA, M. L. and LE BRETON, A. (2002). Statistical Analysis of the Fractional Ornstein-Uhlenbeck Type Process. *Statistical Inference for Stochastic Processes* **5** 229–248.
- NUALART, D. (2010). *The Malliavin Calculus and Related Topics (Probability and Its Applications)*, Second ed. Springer.
- PERCIVAL, D. B. and WALDEN, A. T. (1993). *Spectral Analysis for Physical Applications*, First ed. Cambridge University Press.
- ROBINSON, P. M. (1995). Gaussian Semiparametric Estimation of Long Range Dependence. *The Annals of Statistics* **23** 1630–1661.
- SAMORODNITSKY, G. and TAQQU, M. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, First ed. *Stochastic modeling*. Chapman and Hall/CRC.
- THOMSON, D. J. (1982). Spectrum estimation and harmonic analysis. *Proceedings of the IEEE* **70** 1055–1096.
- TUDOR, C. A. and VIENS, F. G. (2007). Statistical Aspects of the Fractional Stochastic Calculus. *The Annals of Statistics* **35** 1183–1212.
- WALDEN, A. T., MCCOY, E. J. and PERCIVAL, D. B. (1995). The Effective Bandwidth of a Multitaper Spectral Estimator. *Biometrika* **82** 201–214.

ONLINE SUPPLEMENT B: ADDITIONAL PLOTS AND TABLES

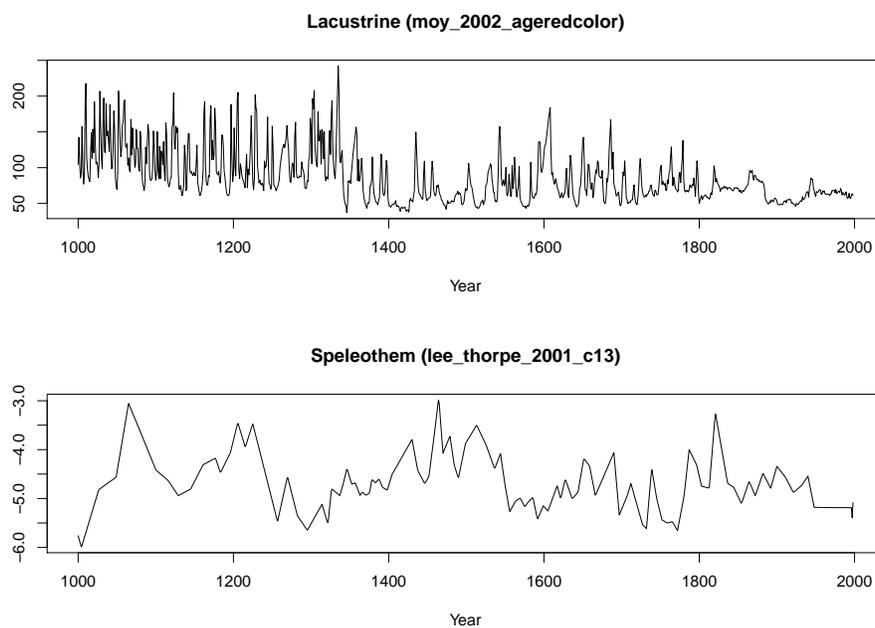


Fig B.1: Two examples of individual proxy series.

TABLE B.1
Proxies not included in RP_t

Name	Type	Country
burns_2003_socotrad13c	Speleothem	Yemen
burns_2003_socotrad18o	Speleothem	Yemen
burns_nicoya_d18o	Speleothem	Costa Rica
curtis_1996_d13cpyro	Lacustrine	Mexico
curtis_1996_d18o	Lacustrine	Mexico
dongge	Speleothem	China
fisher_1994_agassiz	Ice core	Canada
ge_2002_10yeartemp	Other	China
lee_thorpe_2001_c13	Speleothem	South Africa
lee_thorpe_2001_o18	Speleothem	South Africa
moy_2002_ageredcolor	Lacustrine	Ecuador
tiljander_2003_lightsum	Lacustrine	Finland
tiljander_2003_thicknessmm	Lacustrine	Finland

TABLE B.2
Least Squares estimates of model (3.1). HAD and CRU temperature datasets.

Proxy	Type	HAD	CRU
ak046	Tree Ring	0.034	0.021
arge091	Tree Ring	0.008	0.021
az510	Tree Ring	0.015	0.045
baker_2002_su.96.7	Speleothem	0.035	-0.003
ca528	Tree Ring	0.002	-0.023
ca529	Tree Ring	0.019	0.026
ca630	Tree Ring	0.030	0.037
ca631	Tree Ring	-0.024	-0.020
co522	Tree Ring	-0.003	-0.028
fisher_1996_cgreenland	Ice core	-0.007	-0.012
gisp2o18	Ice core	0.014	0.030
mongolia.darrigo	Other	0.038	0.058
mt110	Tree Ring	-0.013	-0.009
norw010	Tree Ring	0.022	0.030
nv512	Tree Ring	0.033	0.015
nv513	Tree Ring	-0.008	0.004
nv518	Tree Ring	0.025	0.043
nv519	Tree Ring	-0.014	-0.020
orokonztr	Other	0.014	0.014
pola006	Tree Ring	0.001	0.010
qian_2003_yriver	Other	-0.015	-0.029
tasmania_recon_orig	Other	0.013	0.027
thompson_1992_quelccao18	Ice core	0.024	0.017
tiljander_2003_darksum	Lacustrine	0.036	0.038
tornetrask	Tree Ring	0.020	0.032
R^2		77.48%	58.25%

TABLE B.3
Relative magnitude of each LS coefficient, classified by country and proxy type (HAD)

	Tree Ring	Ice core	Lacustrine	Speleothem	Other	Total
Argentina	1.77%					1.77%
China					3.18%	3.18%
Finland			7.64%			7.64%
Greenland		4.47%				4.47%
Mongolia					8.22%	8.22%
New Zealand					3.06%	3.06%
Norway	4.76%					4.76%
Peru		5.05%				5.05%
Poland	0.16%					0.16%
Scotland				7.54%		7.54%
Sweden	4.27%					4.27%
Tasmania					2.88%	2.88%
USA	46.99%					46.99%
Total	57.96%	9.52%	7.64%	7.54%	17.34%	100.00%

TABLE B.4
Relative magnitude of each LS coefficient, classified by country and proxy type (CRU)

	Tree Ring	Ice core	Lacustrine	Speleothem	Other	Total
Argentina	3.50%					3.50%
China					4.70%	4.70%
Finland			6.28%			6.28%
Greenland		6.84%				6.84%
Mongolia					9.46%	9.46%
New Zealand					2.27%	2.27%
Norway	4.98%					4.98%
Peru		2.79%				2.79%
Poland	1.63%					1.63%
Scotland				0.46%		0.46%
Sweden	5.27%					5.27%
Tasmania					4.34%	4.34%
USA	47.47%					47.47%
Total	62.87%	9.63%	6.28%	0.46%	20.76%	100.00%

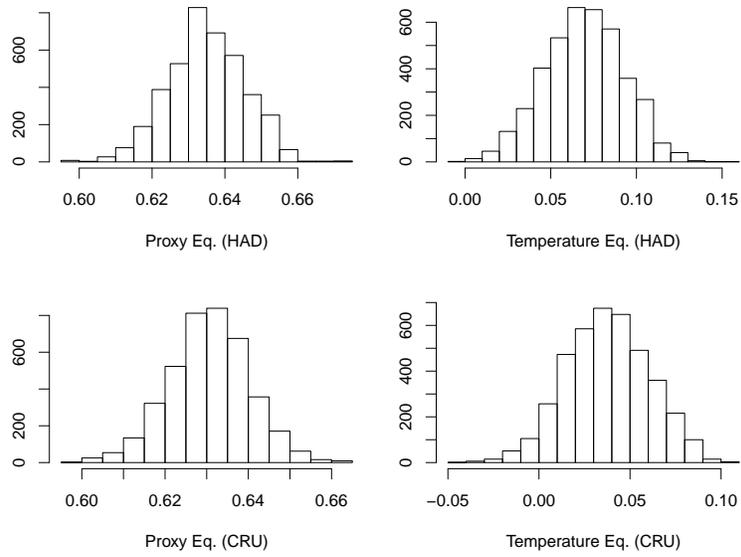


Fig B.2: Bayesian Estimation of Error parameters. Scenario B.

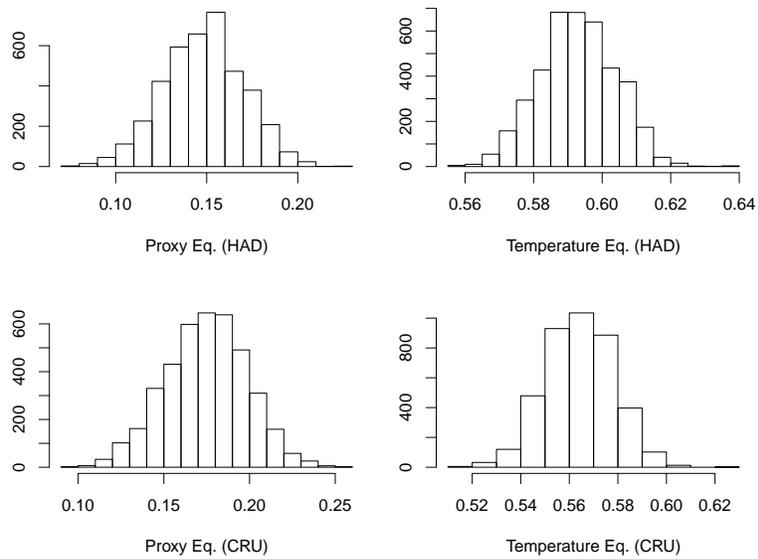


Fig B.3: Bayesian Estimation of Error parameters. Scenario C.

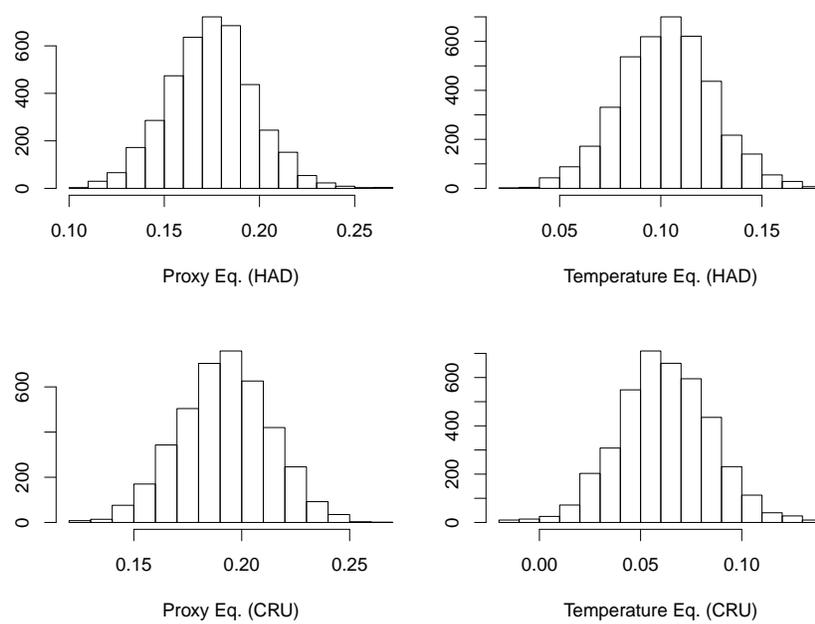


Fig B.4: Bayesian Estimation of Error parameters. Scenario D.

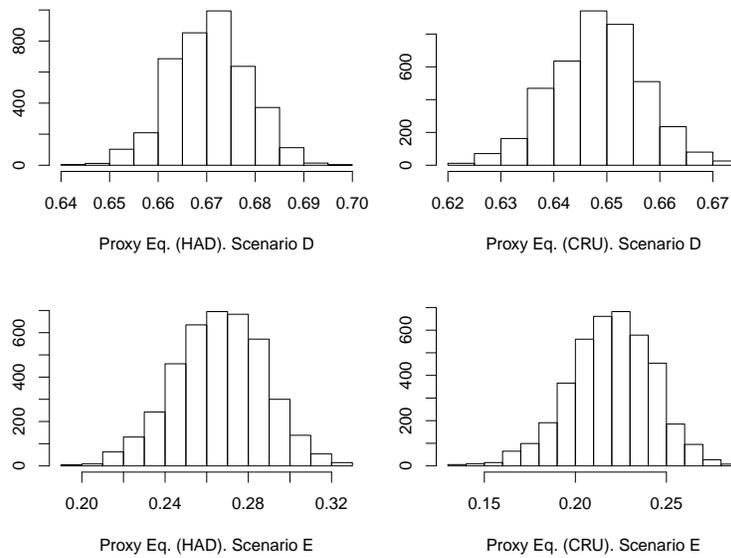
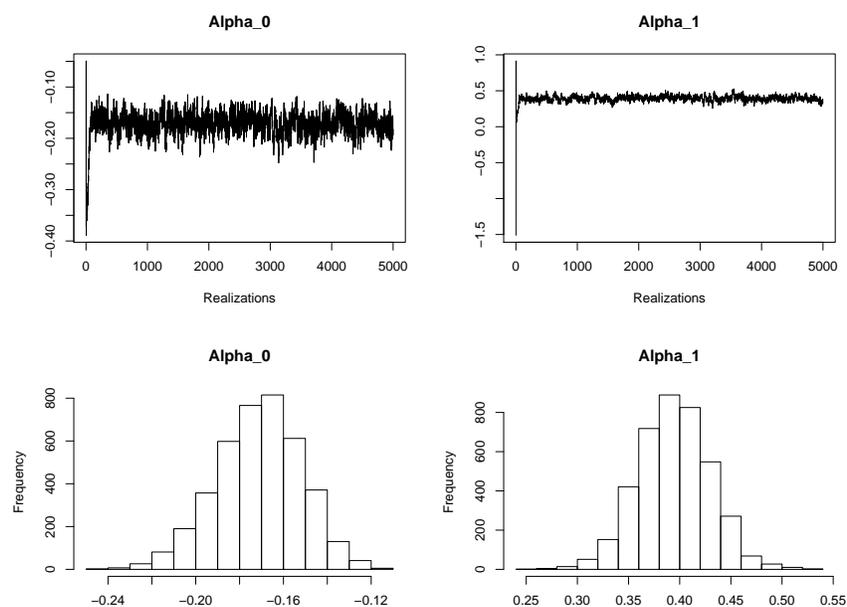
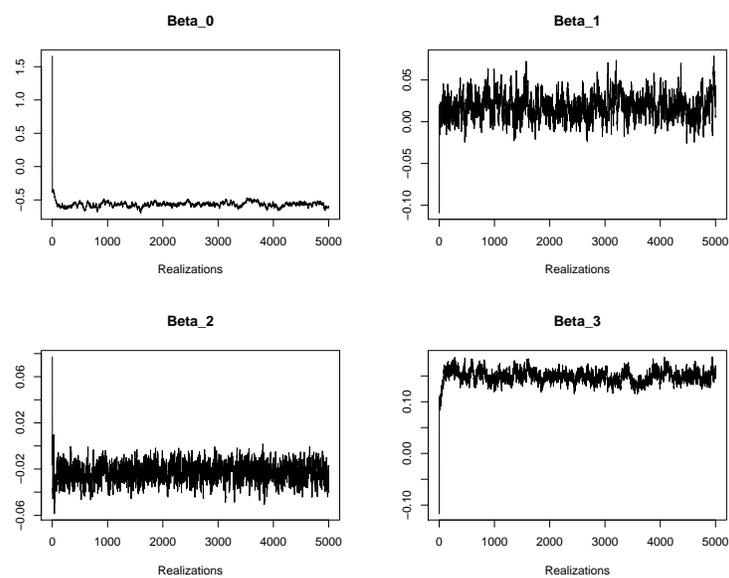


Fig B.5: Bayesian Estimation of Error parameters in the Proxy equation. Scenarios F and G.

Fig B.6: Bayesian Estimation of α . (HAD, Scenario A)Fig B.7: Bayesian Estimation of β . (Realizations) (HAD, Scenario A)

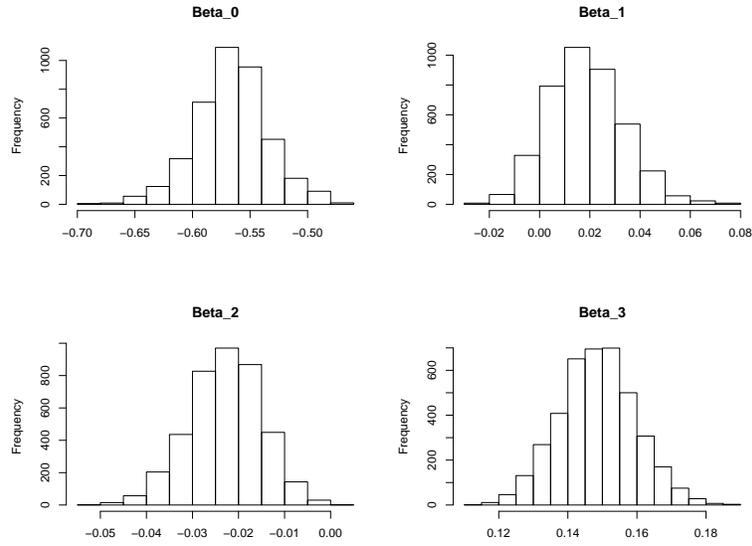


Fig B.8: Bayesian Estimation of β . (Histograms) (HAD, Scenario A)

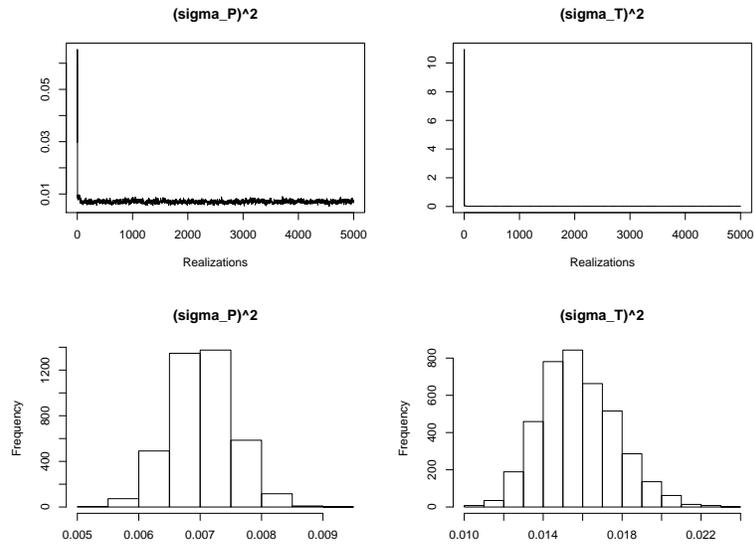


Fig B.9: Bayesian Estimation of σ_P^2 and σ_T^2 . (HAD, Scenario A)

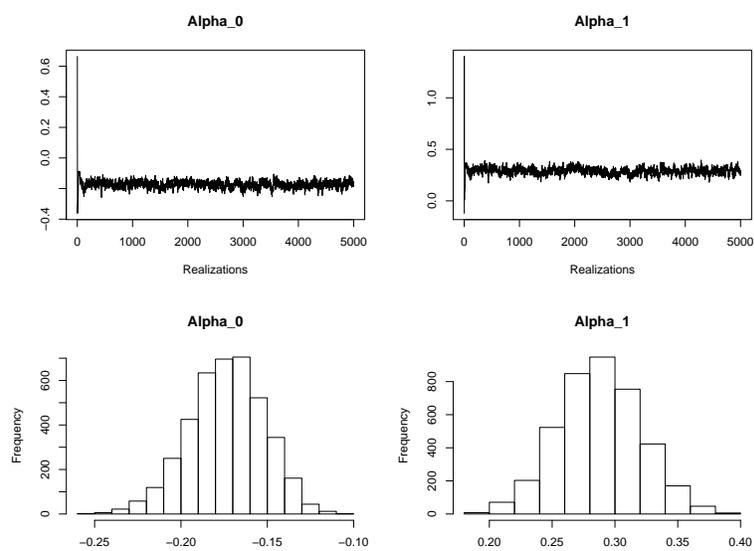


Fig B.10: Bayesian Estimation of α . (CRU, Scenario A)

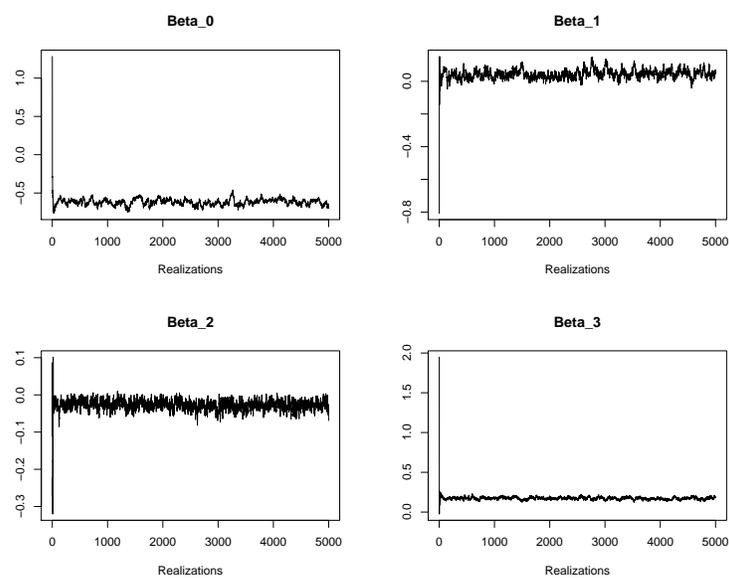


Fig B.11: Bayesian Estimation of β . (Realizations) (CRU, Scenario A)

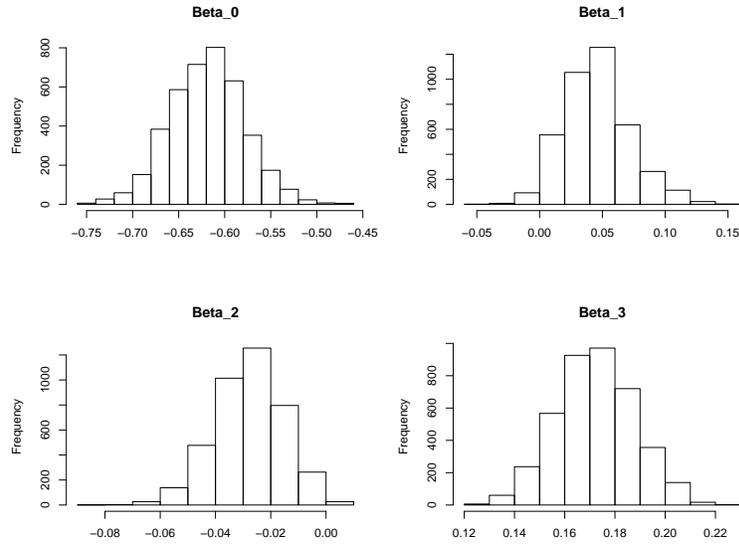


Fig B.12: Bayesian Estimation of β . (Histograms) (CRU, Scenario A)

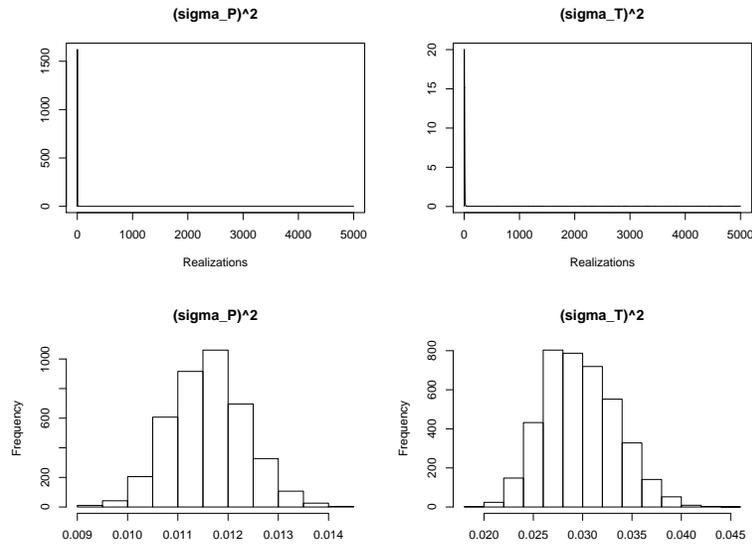


Fig B.13: Bayesian Estimation of σ_P^2 and σ_T^2 . (CRU, Scenario A)

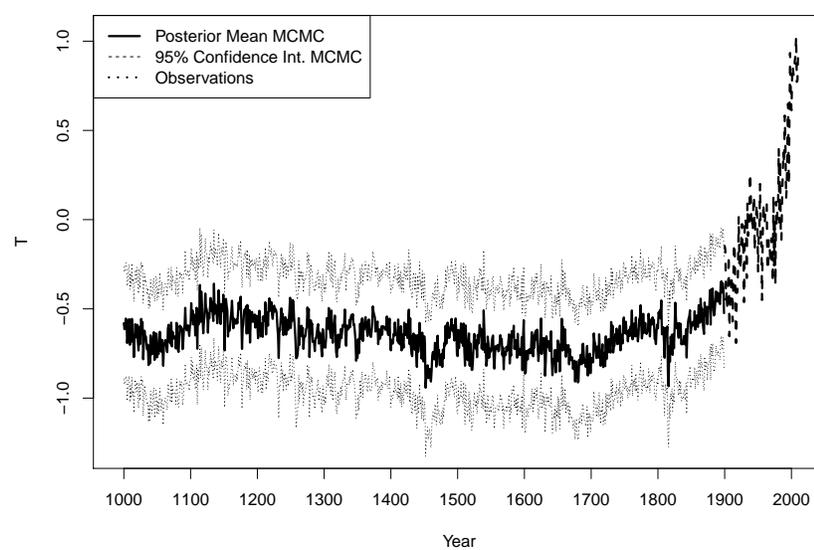


Fig B.14: Temperature reconstruction (1000-1899) using the CRU dataset, Scenario B.

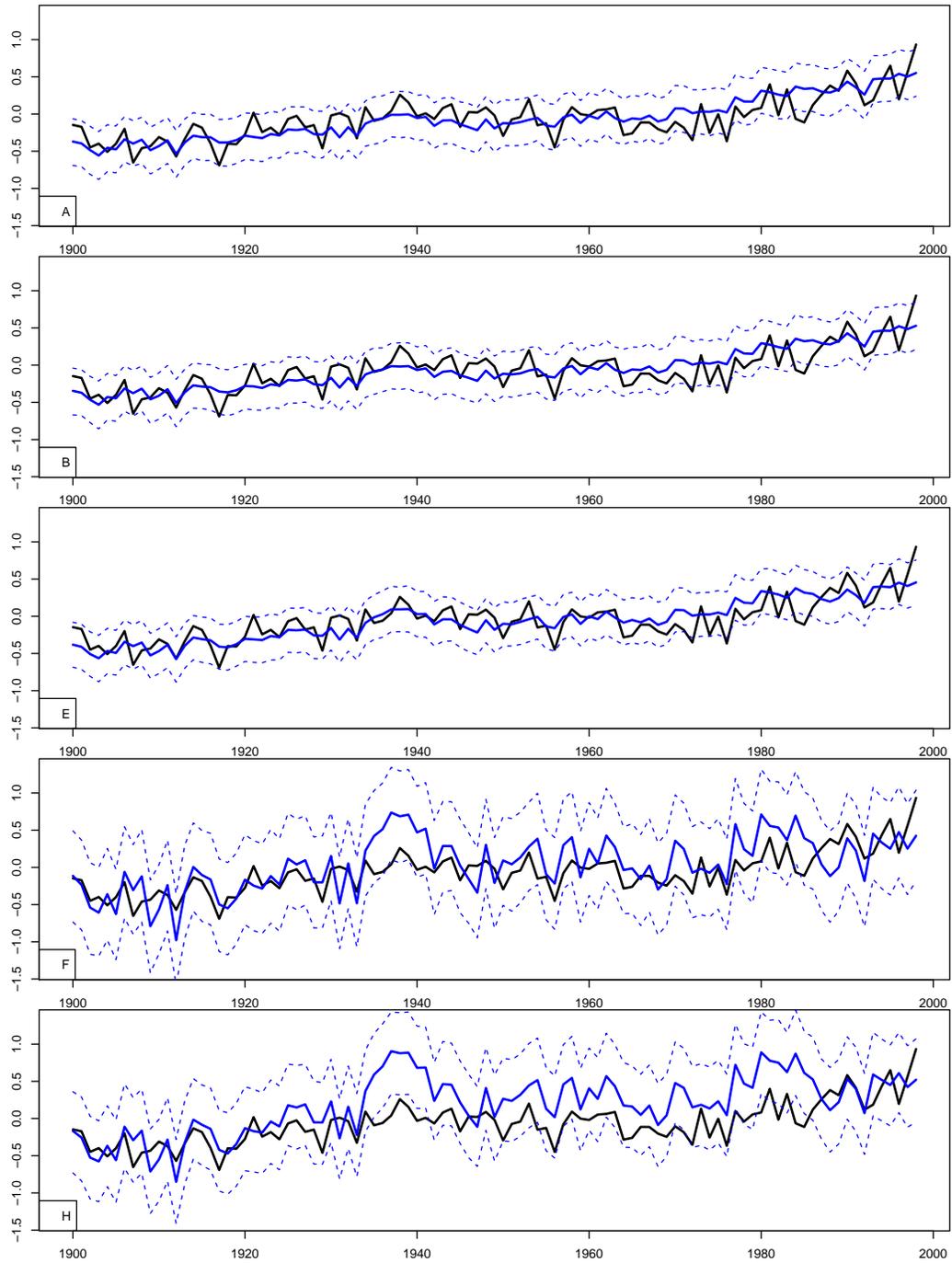


Fig B.15: Temperature reconstruction (1900-1998) using the CRU dataset under Scenarios A, B, E, F and H. Black: Observations; Purple: 95% credible intervals MCMC; Dark Purple line: posterior mean MCMC.

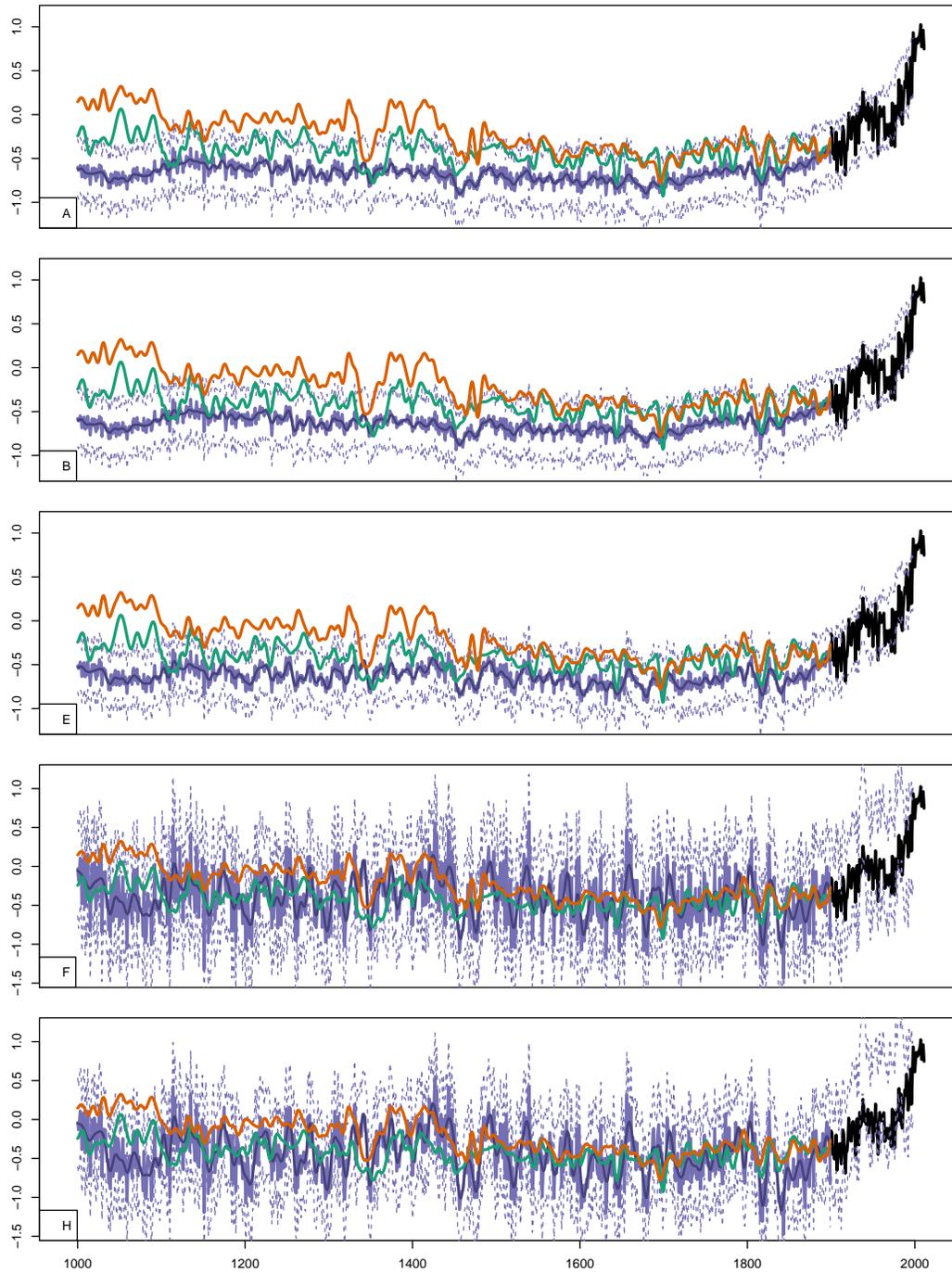


Fig B.16: Comparison of Scenarios A, B, E, F, H with previous reconstructions (CRU). Black: Observations, Purple: MCMC posterior mean with 95% credible intervals, Gold: EIV (Mann et al., 2008), Green: CPS (Mann et al., 2008). Dark Purple line: mean of smoothed posteriors.

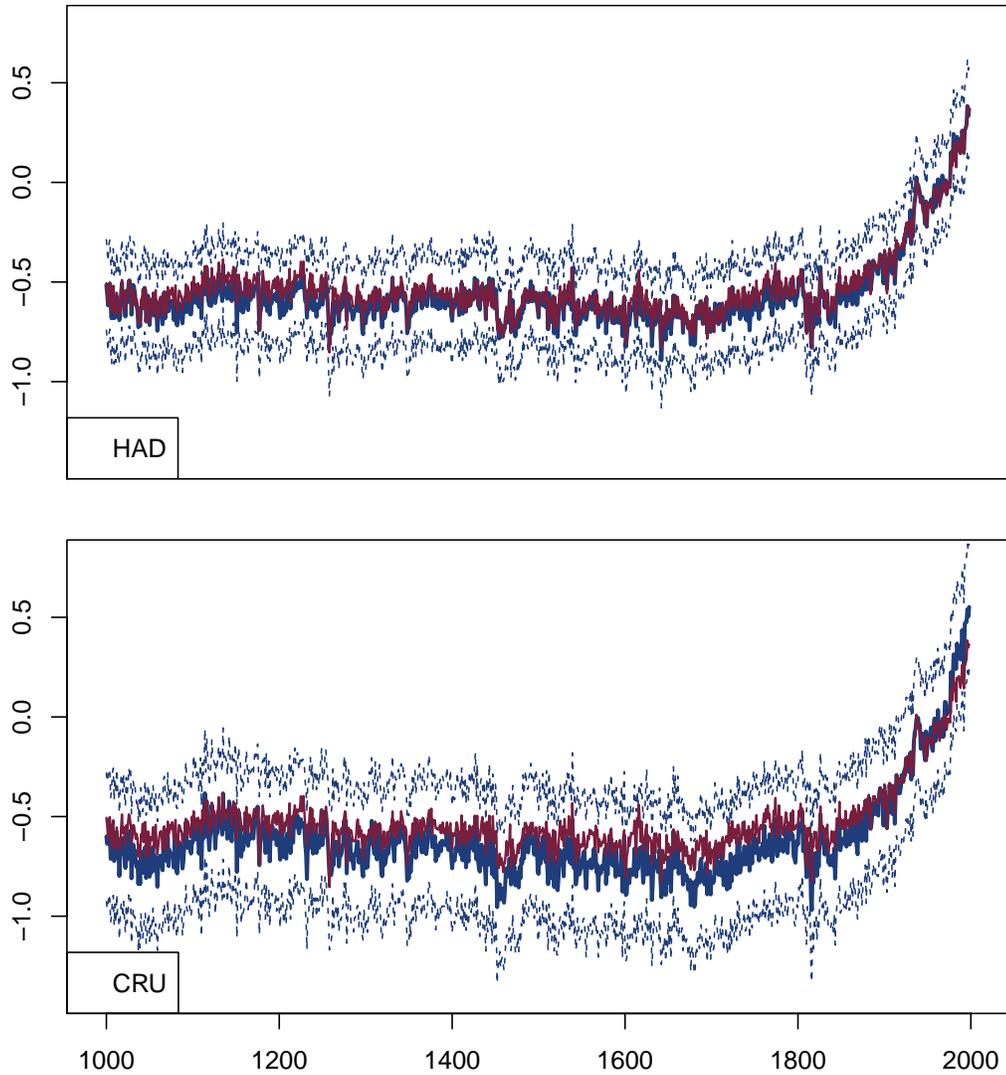


Fig B.17: Comparison of HAD and CRU reconstructions with and without the single lacustrine series in RP_t (Scenario A). Blue line: posterior mean of the original MCMC reconstruction, blue dotted line: 95% prediction intervals of the original MCMC reconstruction, red line: posterior mean of the MCMC reconstruction without the lacustrine series.