The Value of Multi-proxy Reconstruction of Past Climate – Supplement

Outline

1. **Synthetic proxies generation**

2. **Justification for the linear model between temperature and forcings and AR(2) errors**

3. **MCMC sampling**

4. **Model comparison**

5. **Posterior predictive check**

All the data and R codes for this paper are posted at http://www.image.ucar.edu/Data/.

1 Synthetic proxies generation

Here we show the details in generating synthetic tree-rings, pollen and borehole.

Synthetic tree-ring observations

We first selected 15 local temperature series with length of 1500 years from the CCSM output, and then for each time series we apply a rectangular sliding-average smooth over 11 adjacent points to obtain the approximate 10 year smoothing average. Specifically, let T_i be the temperature at year i, $i = 1, \ldots, 1500$, and let S_i be the smoothed temperature also at year i,

$$S_i = \frac{T_{i-5} + T_{i-4} + T_{i-3} + T_{i-2} + T_{i-1} + T_i + T_{i+1} + T_{i+2} + T_{i+3} + T_{i+4} + T_{i+5}}{11}.$$

Then $(T_1, \ldots, T_{1500})' - (S_1, \ldots, S_{1500})'$ gives one synthetic tree-ring series. We repeat this procedure for all 15 local temperature series to generate 15 synthetic tree-rings. Figure 1 compares a random sample of our synthetic tree-rings with nine tree-ring

width or density from Mann *et al.* (1999). All the series are scaled for a easy comparison. It can be seen that our synthetic tree-rings look similar to the real tree-ring series, in particular the real tree-ring series numbered with 3, 4, 5 and 9.
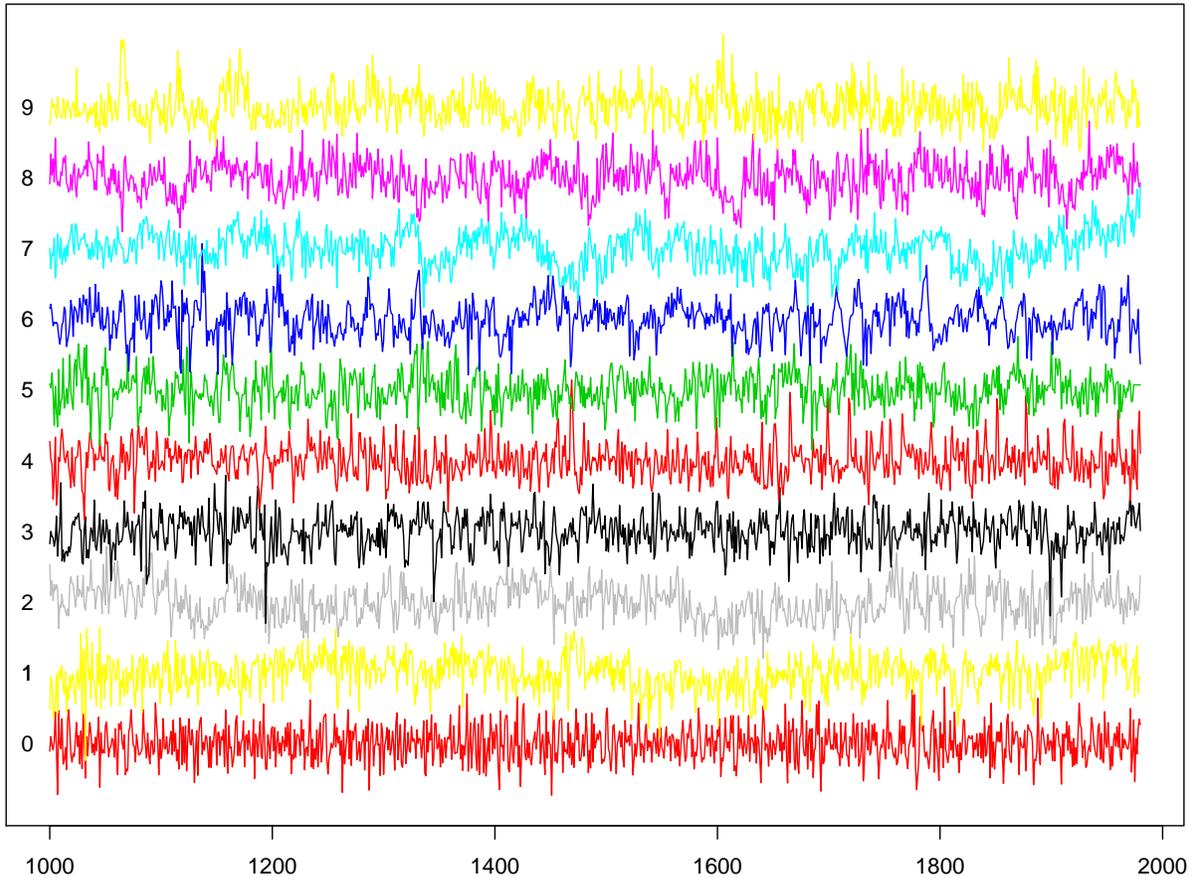
Figure 1: Comparison between synthetic tree-rings and observed tree-ring width or density. 0 - a random sample of our synthetic tree-rings. $1, \ldots, 9$ - tree-ring observations from Mann *et al.* (1999).

Synthetic borehole observations

We generate borehole data based on five regional composite temperature series which are the local average of model temperature output over five $20^0 \times 20^0$ squares. We follow the pre-observation mean-surface air temperature (POM-SAT) model to generate the borehole temperature profiles. Let Z_i, $i = 1, \ldots, k$ be the depth, and τ_i,

$i = 1, \ldots, n$ be the difference between the reference year and the i-th year. Define two matrices

$$\mathbf{Z} = \begin{pmatrix} Z_1 & \ldots & Z_1 \\ \vdots & \vdots & \vdots \\ Z_k & \ldots & Z_k \end{pmatrix} \text{ and } \boldsymbol{\tau} = \begin{pmatrix} \tau_1 & \ldots & \tau_n \\ \vdots & \vdots & \vdots \\ \tau_1 & \ldots & \tau_n \end{pmatrix}.$$

Let α be the thermal diffusivity and POM be the pre-observation mean, then the depth profile T_{Z_i} is

$$\begin{pmatrix} T_{Z_1} \\ T_{Z_2} \\ \vdots \\ T_{Z_n} \end{pmatrix} = \left(erfc(\frac{\mathbf{Z}}{\sqrt{4\alpha\boldsymbol{\tau}}}) \right) \begin{pmatrix} -1 & 1 & 0 & \ldots & 0 \\ 0 & -1 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} POM \\ T_1 \\ \vdots \\ T_n \end{pmatrix},$$

where $erfc(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt$ is the complementary error function. In our data generation procedures, we choose $\alpha = 10e^{-6} m^2/s$, $POM = 0$ without loss of generality, $Z_i = \{0, 5m, 10m, \ldots, 500m\}$. We choose the reference year as 2000, so τ_i for the i-th year is $\tau_i = 2000 - i$.

Synthetic pollen observations

We select 10 regional composite temperature series as the local average of $7.5^0 \times 7.5^0$ squares to generate pollen data. For each composite temperature series $\mathbf{T} = (T_1, \ldots, T_{1500})'$, we again apply the rectangular sliding-average smooth over 11 adjacent points to obtain the approximate 10 year smoothing average $\mathbf{S} = (S_1, \ldots, S_{1500})'$. Then we extract points from $\mathbf{S}$ at every 30 years and consider those points as the synthetic pollen observations.

2 Justification for the linear model between temperature and forcings and AR(2) errors

Linear process model with AR(2) error structure

We assess the linear model between temperatures and forcings as well as the AR(2)

error structure using the HadCRUT3v temperature series that has been used in Li et al. (2007) and is available at *http://www.image.ucar.edu/~boli/research.html.* This is because our models are developed eventually for real data. The overlap period for the HadCRUT3v temperature series and three forcings is 1850-1999. Except for the zero-inflated volcanism, both the scatter plots between temperatures and solar irradiance and between temperatures and CO_2 concentrations in Figure 2 display a clearly linear relationship. This suggests that the choice of a linear model relating temperatures and forcings is reasonable. Furthermore, we examine residuals from the simple linear model of

$$T = \beta_0 + \beta_1 S + \beta_2 V + \beta_3 C + \epsilon,$$

where T is temperature, S, V and C are solar irradiance, volcanism and greenhouse gases, respectively. The residuals and their normal probability plot in Figure 3 show that residuals are well behaved and follow a perfect normal distribution. Figure 4 checks the autocorrelation function (acf) and partial autocorrelation function (pacf) of the residuals. The acf and pacf suggest an AR(1) model for the error structure. In summary, the diagnostics indicate that a linear model with an AR(1) error structure seems sufficient to link the real temperatures and forcings between 1850-1999. Nevertheless, we choose a more conservative AR(2) error structure in our model to account for any extra correlations potentially for the much longer period of 850-1999.

AR(2) for proxies given temperatures

We still use the HadCRUT3v temperature series and the nine tree-rings in Mann *et al.* (1999) to explore the model for errors. The overlap period for the temperatures and those nine tree-rings is 1850-1980. After we filter the temperature series by removing the 10 year smoothing average, we fit a simple linear model between each tree-ring series and the filtered temperatures with tree-rings as the response variable. The pattern of acf and pacf from all the nine sets of residuals is hardly differentiable, thus we only show one such example in Figure 5. This plot may indicate that iid errors are probably ok despite a significant correlation at lag 3. Again, we made a more
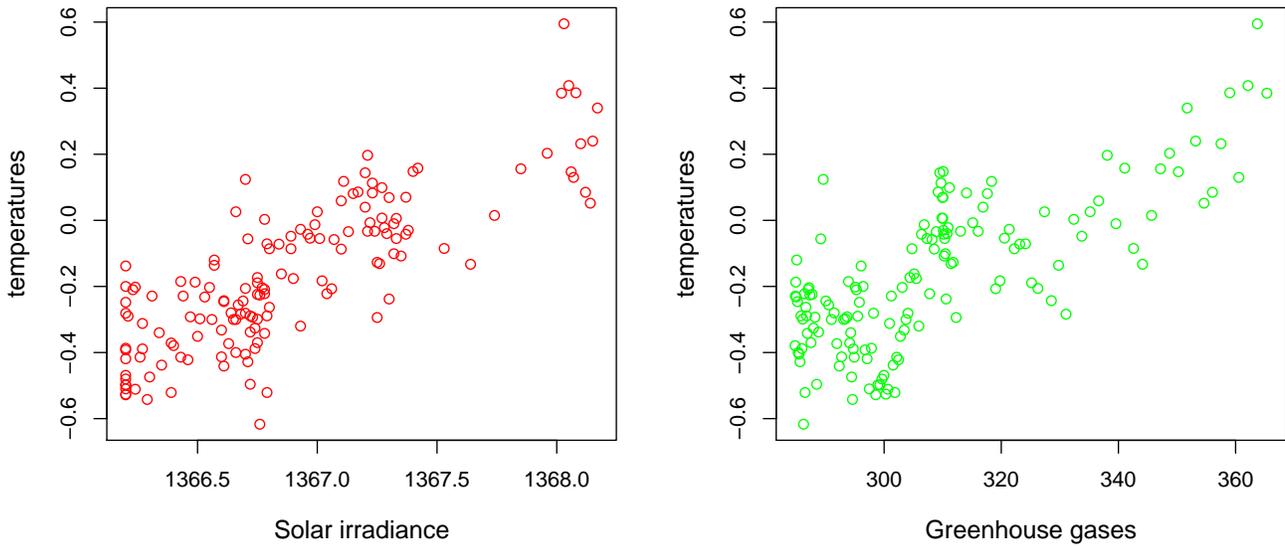
Figure 2: The scatter plots between temperatures and solar irradiance and between temperatures and CO_2.
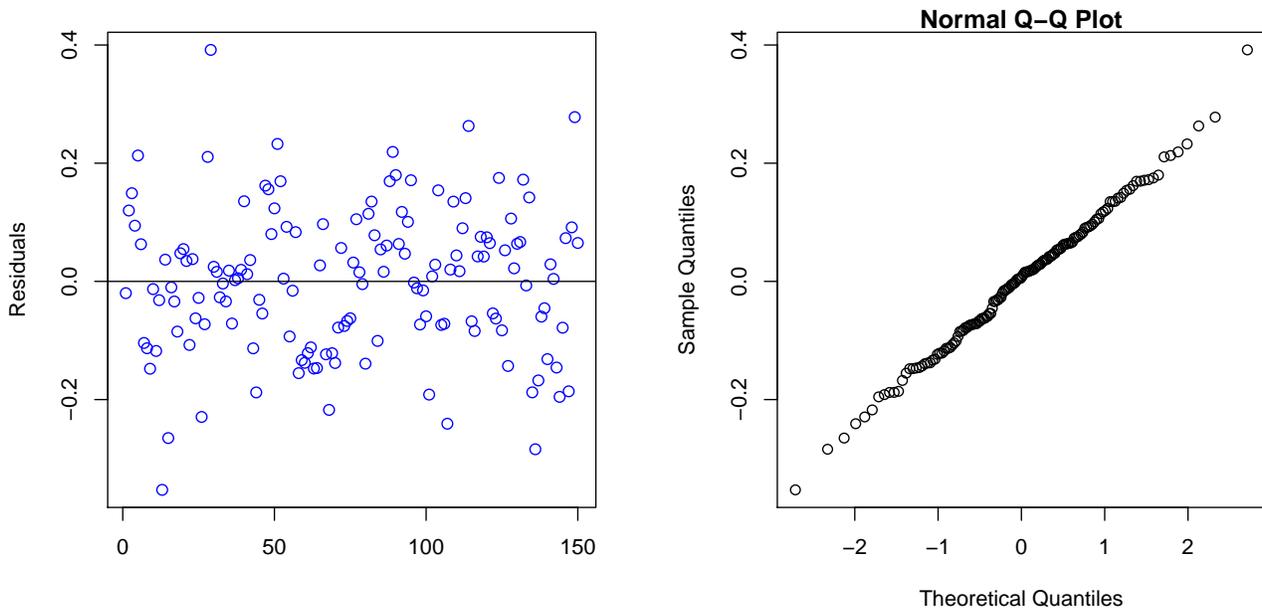
Figure 3: The residual plot and the normal probability plot of residuals

conservative choice because the setting of this experiment is not exactly the same as our models. For instance, the time period for our model to hold is 850-1999 which is
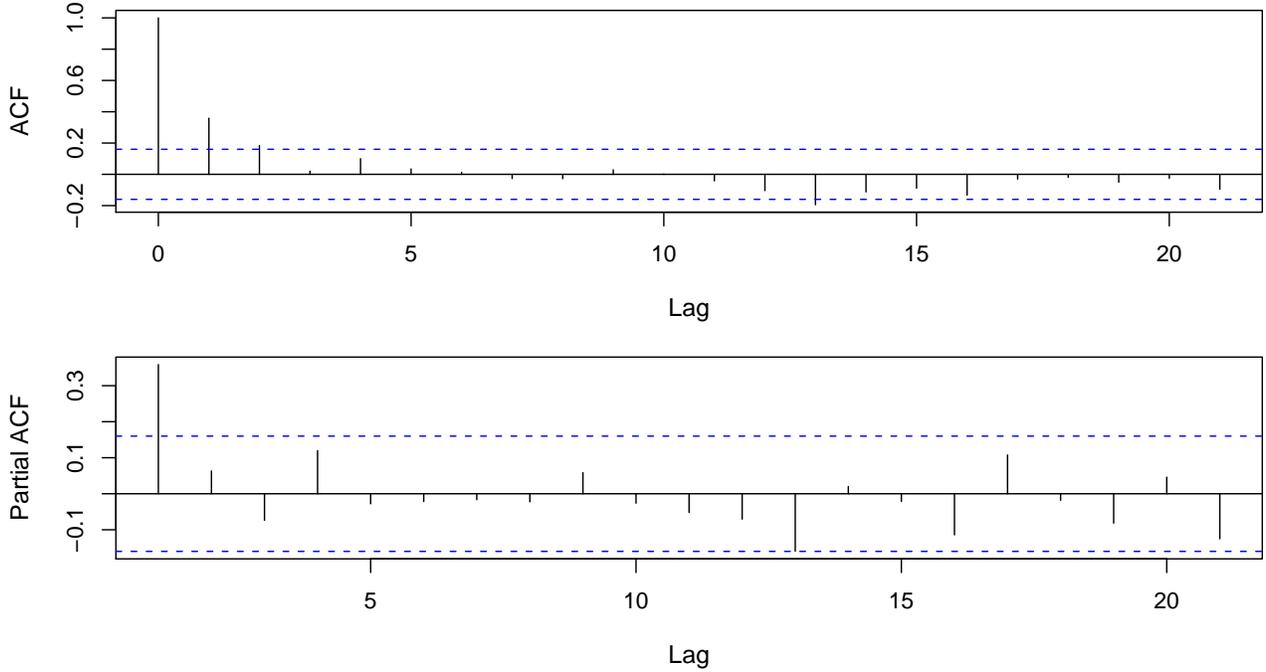
Figure 4: The autocorrelations and partial autocorrelations of residuals obtained from temperatures given forcings

much longer than 1850-1980, and those nine tree-rings are only samples of tree-ring observations. In case that AR(2) structure is unnecessary, the estimates of the two time lag coefficients in the error model will be simply close to zero.

3 MCMC sampling

Let $\mathbf{D}_i$, $\mathbf{P}_j$ and $\mathbf{B}_k$ be vectors of synthetic tree-ring (**D**endrochronology), **P**ollen and **B**orehole data indexed by their various locations with $i = 1, .., 15$, $j = 1, ..., 10$ and $k = 1, ..., 5$. Note that these groups of proxy vectors will have different lengths due to their sampling. Moreover, each tree-ring and pollen vectors are indexed with respect to time, and the borehole vectors are indexed by depth. Also let $\mathbf{S}$, $\mathbf{V}_0$, and $\mathbf{C}$ be the time series vectors of solar irradiance, volcanism and greenhouse gases, and let $\mathbf{V}$ denote the volcanic series with error. Let $\mathbf{M}_D$, $\mathbf{M}_P$ and $\mathbf{M}_B$ be the three transformation matrices to link temperature series to tree-ring, pollen and borehole,

6
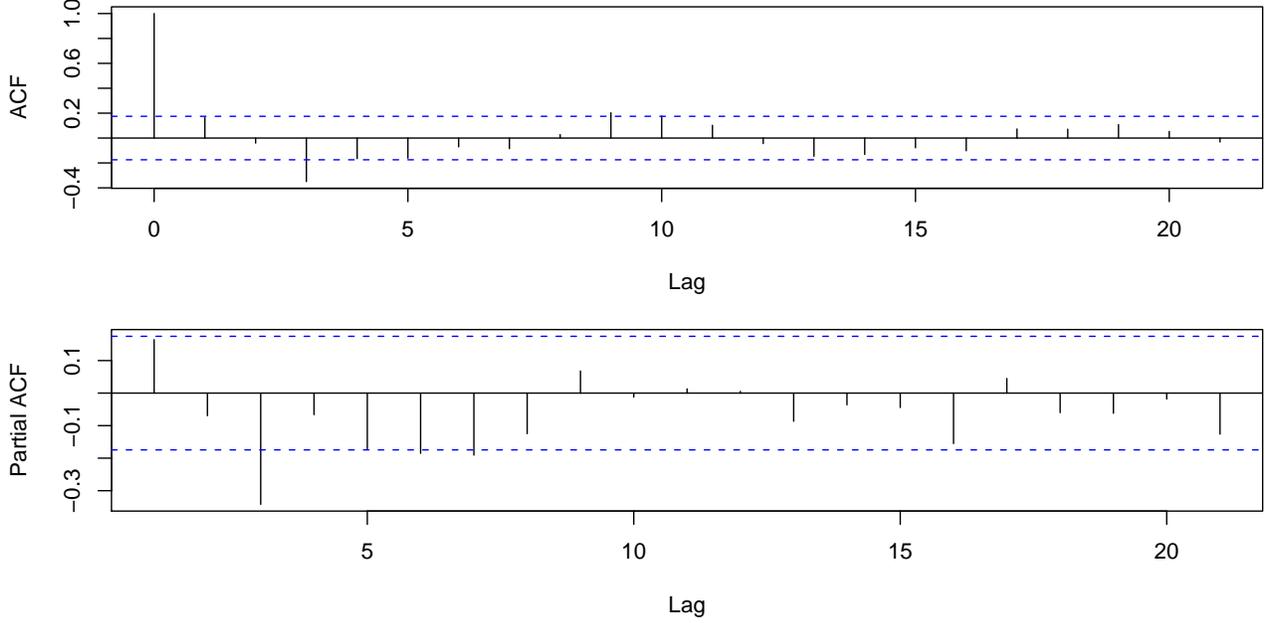
Figure 5: The autocorrelations and partial autocorrelations of residuals obtained from tree-rings given the filtered temperatures

respectively. More specifically, they represent the linear filters used to generate the corresponding pseudo proxies from the model temperature series. Finally it is useful to partition the full length temperature process $\mathbf{T}$ into the unknown temperatures $\mathbf{T}_1$ requiring reconstruction over the time span of available proxy data, and the observed instrumental temperatures $\mathbf{T}_2$ (1850-present), i.e., $\mathbf{T} = (\mathbf{T}_1', \mathbf{T}_2')'$. Then we have the three hierarchies below:

(i) Data stage:

$$\mathbf{D}_i|(\mathbf{T}_1', \mathbf{T}_2')' = \mu_{iD} + \beta_{iD}\mathbf{M}_D(\mathbf{T}_1', \mathbf{T}_2')' + \boldsymbol{\epsilon}_{iD}, \quad \boldsymbol{\epsilon}_{iD} \sim \mathrm{AR}(2)(\sigma_D^2, \phi_{1D}, \phi_{2D}), \quad (3.1)$$

$$\mathbf{P}_j|(\mathbf{T}_1', \mathbf{T}_2')' = \mu_{jP} + \beta_{jP}\mathbf{M}_P(\mathbf{T}_1', \mathbf{T}_2')' + \boldsymbol{\epsilon}_{jP}, \quad \boldsymbol{\epsilon}_{jP} \sim \mathrm{AR}(2)(\sigma_P^2, \phi_{1P}, \phi_{2P}), \quad (3.2)$$

$$\mathbf{B}_k|(\mathbf{T}_1', \mathbf{T}_2')' = \mathbf{M}_B\{\mu_{kB} + \beta_{kB}(\mathbf{T}_1', \mathbf{T}_2')' + \boldsymbol{\epsilon}_{kB}\}, \quad \boldsymbol{\epsilon}_{kB} \sim \mathrm{iid}\ N(0, \sigma_B^2), \quad (3.3)$$

$$\mathbf{V}|\mathbf{V}_0 = (1 + \epsilon_V)\mathbf{V}_0, \quad \epsilon_V \sim \mathrm{iid}\ N(0, 1/64); \quad (3.4)$$

7

(ii) Process stage:

$$(\mathbf{T}_1', \mathbf{T}_2')'|(\mathbf{S}, \mathbf{V}_0, \mathbf{C}) = \beta_0 + \beta_1\mathbf{S} + \beta_2\mathbf{V}_0 + \beta_3\mathbf{C} + \boldsymbol{\epsilon}_T, \quad \boldsymbol{\epsilon}_T \sim \text{AR}(2)(\sigma_T^2, \phi_{1T}, \phi_{2T});$$

$$(3.5)$$

(iii) Priors:

$\mu_{iL} \sim N(\widetilde{\mu}_{iL}, \widetilde{\sigma}_{iL}^2), \beta_{iL} \sim N(\widetilde{\mu'}_{iL}, \widetilde{\sigma'}_{iL}^2), L \in \{D, P, B\};$

$\beta_i \sim N(\widetilde{\mu}_i, \widetilde{\sigma}_i^2), i = 0, 1, 2, 3;$

$\phi_{2L} \sim \text{unif}(-1, 1), \phi_{1L}|\phi_{2L} \sim (1 - \phi_{2L}) \times \text{unif}(-1, 1), L \in \{D, P, T\};$

$\sigma_L^2 \sim IG(\widetilde{q}_L, \widetilde{r}_L), L \in \{D, P, B, T\}.$

An AR(2)$(\sigma^2, \phi_1, \phi_2)$ process is defined as $e_t = \phi_1 e_{t-1} + \phi_2 e_{t-2} + \epsilon_t$, $\epsilon_t \sim$ iid Normal$(0, \sigma^2)$. The specific choice of priors for time lag coefficients (ϕ_{1L}, ϕ_{2L}) guarantees their corresponding AR(2) process to be stationary and causal (Shumway and Stoffer, 2006, ch. 3), and the conjugate priors for all the other parameters allow for an explicit full conditional posterior distribution for those parameters and $\mathbf{T}_1$. There is no closed form for the posterior distribution of time lag coefficients. Thus the posterior is sampled by alternating the Gibbs sampler, which is used for updating $\mathbf{T}_1$ and parameters with explicit full conditional distribution, and the Metropolis-Hasting (M-H) algorithm, which is used for updating autoregressive parameters. More specifically, we generate posteriors in the following order with the indicated sampling method: $\mathbf{T}_1$ (Gibbs), $\mathbf{V}_0$ (Gibbs), (μ_{iL}, β_{iL}) with $L \in \{D, P, B\}$ and β_i with $i = 0, 1, 2, 3$ (Gibbs), (ϕ_{1L}, ϕ_{2L}), $L \in \{D, P, T\}$ (M-H Algorithm), σ_L^2, $L \in \{D, P, B, T\}$ (Gibbs).

Sample $\mathbf{T}_1$

Let $[X|\cdot]$ be the conditional distribution of X. Let $\boldsymbol{\Sigma}_L$, $L \in \{D, P, B, T\}$ be the covariance matrix of $\boldsymbol{\epsilon}_{iD}$, $\boldsymbol{\epsilon}_{jP}$, $\boldsymbol{\epsilon}_{kB}$ and $\boldsymbol{\epsilon}_T$, respectively. Let $\mathbf{H}_D = \mathbf{M}_D'\boldsymbol{\Sigma}_D^{-1}\mathbf{M}_D$, $\mathbf{H}_P = \mathbf{M}_P'\boldsymbol{\Sigma}_P^{-1}\mathbf{M}_P$ and $\mathbf{H}_T = \boldsymbol{\Sigma}_T^{-1}$. Let $\mathbf{E}$ be the matrix consisting of three rows as the first three eigenvectors of $\mathbf{M}_B\mathbf{M}_B'$ and let $\mathbf{M}_{BR} = \mathbf{E}\mathbf{M}_B$, we define $\mathbf{H}_B = \mathbf{M}_{BR}'(\mathbf{M}_{BR}\boldsymbol{\Sigma}_B\mathbf{M}_{BR}')^{-1}\mathbf{M}_{BR}$. Let $\mathbf{G}_D = \mathbf{M}_L'\boldsymbol{\Sigma}_L^{-1}$, $\mathbf{G}_P = \mathbf{M}_P'\boldsymbol{\Sigma}_P^{-1}$, $\mathbf{G}_B = \mathbf{M}_{BR}'(\mathbf{M}_{BR}\boldsymbol{\Sigma}_B\mathbf{M}_{BR}')^{-1}$. The size of $\mathbf{H}_L$, $L \in \{D, P, B, T\}$ is 1150×1150.

8

Now we partition $\mathbf{H}_L$ into

$$\mathbf{H}_L = \begin{pmatrix} \mathbf{H}_{L11} & \mathbf{H}_{L12} \\ \mathbf{H}_{L21} & \mathbf{H}_{L22} \end{pmatrix},$$

where the size of $\mathbf{H}_{L11}$ is 1000×1000. We also partition $\mathbf{G}_L$ into

$$\mathbf{G}_L = \begin{pmatrix} \mathbf{G}_{L1} \\ \mathbf{G}_{L2} \end{pmatrix},$$

where $\mathbf{G}_{L1}$ has 1000 rows. By completing square, we can get:

$$[\mathbf{T}_1|\cdot] = N(\mathbf{Ab}, \mathbf{A}),$$

where

$$\mathbf{A}^{-1} = \sum_{i=1}^{15} \beta_{iD}^2 \mathbf{H}_{D11} + \sum_{j=1}^{10} \beta_{jP}^2 \mathbf{H}_{P11} + \sum_{k=1}^{5} \beta_{kD}^2 \mathbf{H}_{B11} + \mathbf{H}_{T11}$$

$$\begin{aligned}
\mathbf{b} = \; & \mathbf{G}_{D1} \sum_{i=1}^{15} \beta_{iD}(\mathbf{D}_i - \mu_{iD}) - \sum_{i=1}^{15} \beta_{iD}^2 \mathbf{H}_{D12}\mathbf{T}_2 \\
& + \mathbf{G}_{P1} \sum_{j=1}^{10} \beta_{jP}(\mathbf{P}_j - \mu_{jP}) - \sum_{j=1}^{10} \beta_{jP}^2 \mathbf{H}_{P12}\mathbf{T}_2 \\
& + \mathbf{G}_{B1} \sum_{k=1}^{5} \beta_{kB}(\mathbf{B}_{kR} - \mu_{kB}\mathbf{M}_{BR}\mathbf{1}) - \sum_{k=1}^{5} \beta_{kB}^2 \mathbf{H}_{B12}\mathbf{T}_2,
\end{aligned}$$

where $\mathbf{B}_{kR} = \mathbf{E}\mathbf{B}_k$, and $\mathbf{1}$ is a vector consisting of 1.

Sample $\mathbf{V}_0$ and linear model coefficients

Given $\mathbf{T} = (\mathbf{T}_1', \mathbf{T}_2')'$, $[\mathbf{V}_0|\cdot]$, (μ_{iL}, β_{iL}) with $L \in \{D, P, B\}$ and β_i with $i = 0, 1, 2, 3$ are straightforward, so we skip them.

Sample (ϕ_{1L}, ϕ_{2L}) using Metroplis Hasting algorithm

(1) Example of (ϕ_{1D}, ϕ_{2D})

let $\phi_{2D}^*|\phi_{2D}^t \sim N(\phi_{2D}^t, \xi_2)I_{(|\phi_{2D}^*|<1)}$, $\phi_{1D}^*|(\phi_{1D}^t, \phi_{2D}^*) \sim N(\phi_{1D}^t, \xi_1)I_{(|\phi_{1D}^*|<1-\phi_{2D}^*)}$

$$(\phi_{1D}, \phi_{2D})|\cdot \propto [\prod_{i=1}^{15} \mathbf{D}_i|(\mathbf{T}_1', \mathbf{T}_2')', \mu_{iD}, \beta_{iD}, \phi_{1D}, \phi_{2D}, \sigma_P^2][\phi_{2D}][\phi_{1D}|\phi_{2D}]$$

$$r = \frac{\prod_{i=1}^{15}[\mathbf{D}_i|(\mathbf{T}_1', \mathbf{T}_2')', \mu_{iD}, \beta_{iD}, \phi_{1D}^*, \phi_{2D}^*, \sigma_P^2][\phi_{2D}^*][\phi_{1D}^*|\phi_{2D}^*][(\phi_{1D}^*, \phi_{2D}^*)|(\phi_{1D}^t, \phi_{2D}^t)]}{\prod_{i=1}^{15}[\mathbf{D}_i|(\mathbf{T}_1', \mathbf{T}_2')', \mu_{iD}, \beta_{iD}, \phi_{1D}^t, \phi_{1D}^t, \phi_{2D}^t, \sigma_P^2][\phi_{2D}^t][\phi_{1D}^t|\phi_{2D}^t][(\phi_{1D}^t, \phi_{2D}^t)|(\phi_{1D}^*, \phi_{2D}^*)]}$$

since $\phi_{2D} \sim \text{unif}(-1, 1)$, $[\phi_{2D}^*] = [\phi_{2D}^t]$;

$$[(\phi_{1D}^*, \phi_{2D}^*)|(\phi_{1D}^t, \phi_{2D}^t)] = [\phi_{2D}^*|(\phi_{1D}^t, \phi_{2D}^t)][\phi_{1D}^*|(\phi_{2D}^*, \phi_{1D}^t, \phi_{2D}^t)]$$

$$[(\phi_{1D}^t, \phi_{2D}^t)|(\phi_{1D}^*, \phi_{2D}^*)] = [\phi_{2D}^t|(\phi_{1D}^*, \phi_{2D}^*)][\phi_{1D}^t|(\phi_{2D}^t, \phi_{1D}^*, \phi_{2D}^*)]$$

Let Φ_1^* be the pdf for $N(\phi_{1D}^t, \xi_1)$ and Φ_1^t be the pdf for $N(\phi_{1D}^*, \xi_1)$; and Φ_2^* be the pdf

for $N(\phi_{2D}^t, \xi_2)$ and Φ_2^t be the pdf for $N(\phi_{2D}^*, \xi_2)$.

$$\frac{[(\phi_{1D}^*, \phi_{2D}^*)|(\phi_{1D}^t, \phi_{2D}^t)]}{[(\phi_{1D}^t, \phi_{2D}^t)|(\phi_{1D}^*, \phi_{2D}^*)]} = \frac{(\Phi_2^t(1) - \Phi_2^t(-1))|\Phi_1^t(1 - \phi_2^t) - \Phi_1^t(\phi_2^t - 1)|}{(\Phi_2^*(1) - \Phi_2^*(-1))|\Phi_1^*(1 - \phi_2^*) - \Phi_1^*(\phi_2^* - 1)|}$$

$$r = \frac{\prod_{i=1}^{15}[\mathbf{D}_i|(\mathbf{T}_1', \mathbf{T}_2')', \mu_{iD}, \beta_{iD}, \phi_{1D}^*, \phi_{2D}^*, \sigma_P^2](1 - \phi_{2D}^t)(\Phi_2^t(1) - \Phi_2^t(-1))|\Phi_1^t(1 - \phi_{2D}^t) - \Phi_1^t(\phi_{2D}^t - 1)|}{\prod_{i=1}^{15}[\mathbf{D}_i|(\mathbf{T}_1', \mathbf{T}_2')', \mu_{iD}, \beta_{iD}, \phi_{1D}^t, \phi_{2D}^t, \sigma_P^2](1 - \phi_{2D}^*)(\Phi_2^*(1) - \Phi_2^*(-1))|\Phi_1^*(1 - \phi_{2D}^*) - \Phi_1^*(\phi_{2D}^* - 1)|}$$

$$\begin{aligned}
\log(r) &= -0.5 * \{15\log(|\boldsymbol{\Sigma}_D^*|) + \sum_{i=1}^{15}(\mathbf{D}_i - \mu_{iD} - \beta_{iD}\mathbf{M}_D\mathbf{T})'\boldsymbol{\Sigma}_D^{*-1}(\mathbf{D}_i - \mu_{iD} - \beta_{iD}\mathbf{M}_D\mathbf{T})\} \\
&+ \log\{(1 - \phi_{2D}^t)(\Phi_2^t(1) - \Phi_2^t(-1))|\Phi_1^t(1 - \phi_{2D}^t) - \Phi_1^t(\phi_{2D}^t - 1)|\} \\
&+ 0.5 * \{15\log(|\Sigma_D^t|) + \sum_{i=1}^{15}(\mathbf{D}_i - \mu_{iD} - \beta_{iD}\mathbf{M}_D\mathbf{T})'\Sigma_D^{t-1}(\mathbf{D}_i - \mu_{iD} - \beta_{iD}\mathbf{M}_D\mathbf{T})\} \\
&- \log\{(1 - \phi_{2D}^*)(\Phi_2^*(1) - \Phi_2^*(-1))|\Phi_1^*(1 - \phi_{2D}^*) - \Phi_1^*(\phi_{2D}^* - 1)|\}
\end{aligned}$$

(2) Example of (ϕ_{1T}, ϕ_{2T})

Let $\boldsymbol{\mu} = \beta_0 + \beta_1\mathbf{S} + \beta_2\mathbf{V}_0 + \beta_3\mathbf{C}$. Let $\phi_{2T}^*|\phi_{2T}^t \sim N(\phi_{2T}^t, \xi_2)I_{(|\phi_{2T}^*|<1)}$, $\phi_{1T}^*|(\phi_{1T}^t, \phi_{2T}^*) \sim$

$N(\phi_{1T}^t, \xi_1)I_{(|\phi_{1T}^*|<1-\phi_{2T}^*)}$

$$(\phi_{1T}, \phi_{2T})|\cdot \propto [(T_1^T, T_2^T)|\boldsymbol{\mu}, \phi_{1T}, \phi_{2T}, \sigma_T^2][\phi_{2T}][\phi_{1T}|\phi_{2T}]$$

$$r = \frac{[\mathbf{T}'|\boldsymbol{\mu}, \phi_{1T}^*, \phi_{2T}^*, \sigma_T^2][\phi_{2T}^*][\phi_{1T}^*|\phi_{2T}^*][(\phi_{1T}^*, \phi_{2T}^*)|(\phi_{1T}^t, \phi_{2T}^t)]}{[T^T|\boldsymbol{\mu}, \phi_{1T}^t, \phi_{2T}^t, \sigma_T^2][\phi_{2T}^t][\phi_{1T}^t|\phi_{2T}^t][(\phi_{1T}^t, \phi_{2T}^t)|(\phi_{1T}^*, \phi_{2T}^*)]}$$

Let Φ_{1T}^* be the pdf for $N(\phi_{1T}^t, \xi_1)$ and Φ_{1T}^t be the pdf for $N(\phi_{1T}^*, \xi_1)$; and Φ_{2T}^* be the

pdf for $N(\phi_{2T}^t, \xi_2)$ and Φ_{2T}^t be the pdf for $N(\phi_{2T}^*, \xi_2)$.

$$r = \frac{[\mathbf{T}'|\boldsymbol{\mu}, \phi_{1T}^*, \phi_{2T}^*, \sigma_T^2](1 - \phi_{2T}^t)(\Phi_{2T}^t(1) - \Phi_{2T}^t(-1))|\Phi_{1T}^t(1 - \phi_{2T}^t) - \Phi_{1T}^t(\phi_{2T}^t - 1)|}{[T^T|\boldsymbol{\mu}, \phi_{1T}^t, \phi_{2T}^t, \sigma_T^2](1 - \phi_{2T}^*)(\Phi_{2T}^*(1) - \Phi_{2T}^*(-1))|\Phi_{1T}^*(1 - \phi_{2T}^*) - \Phi_{1T}^*(\phi_{2T}^* - 1)|}$$

$$\begin{aligned}
\log(r) &= -0.5 * \{\log(|\boldsymbol{\Sigma}_T^*|) + (\mathbf{T} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}_T^{*-1}(\mathbf{T} - \boldsymbol{\mu})\} \\
&+ \log\{(1 - \phi_{2T}^t)(\Phi_{2T}^t(1) - \Phi_{2T}^t(-1))|\Phi_{1T}^t(1 - \phi_{2T}^t) - \Phi_{1T}^t(\phi_{2T}^t - 1)|\} \\
&+ 0.5 * \{\log(|\boldsymbol{\Sigma}_T^t|) + (\mathbf{T} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}_T^{t-1}(\mathbf{T} - \boldsymbol{\mu})\} \\
&- \log\{(1 - \phi_{2T}^*)(\Phi_{2T}^*(1) - \Phi_{2T}^*(-1))|\Phi_{1T}^*(1 - \phi_{2T}^*) - \Phi_{1T}^*(\phi_{2T}^* - 1)|\}
\end{aligned}$$

Sample σ_L^2

(1) Example of σ_D^2

Let N_D be the number of tree-rings and N_T be the length of **T**. Here $N_D = 15$ and $N_T = 1150$. Let $\widetilde{\mathbf{\Sigma}}_D^{-1} = \sigma_D^2 \mathbf{\Sigma}_D^{-1}$.

$$\sigma_D^2 |\cdot \propto \prod_{i=1}^{15} \{[\mathbf{D}_i | \mathbf{T}, \mu_{iD}, \beta_{iD}, \phi_{1D}, \phi_{2D}, \sigma_D^2]\}[\sigma_D^2]$$

$[\sigma_D^2 |\cdot] \sim IG(q, r)$, where $q = \widetilde{q}_P + N_D * N_T / 2$, and

$$r = \{\frac{1}{\widetilde{r}_D} + 0.5 \sum_{i=1}^{15} (\mathbf{D}_i - \mu_{iD} - \beta_{iD}\mathbf{T})'\widetilde{\mathbf{\Sigma}}_D^{-1}(\mathbf{D}_i - \mu_{iD} - \beta_{iD}\mathbf{T})\}^{-1}$$

(2) Example of σ_T^2

$$\sigma_T^2 |\cdot \propto [\mathbf{T}'|\boldsymbol{\mu}, \phi_{1T}, \phi_{2T}, \sigma_T^2][\sigma_T^2]$$

$[\sigma_T^2 |\cdot] \sim IG(q, r)$, where $q = \widetilde{q}_T + N_T / 2$, $t_1 = 854, t_2 = 127$ and

$$r = \{1/\widetilde{r}_T + 0.5(\mathbf{T} - \boldsymbol{\mu})'\widetilde{\mathbf{\Sigma}}_T^{-1}(\mathbf{T} - \boldsymbol{\mu})\}^{-1},$$

where $\widetilde{\mathbf{\Sigma}}_T^{-1} = \sigma_T^2 \mathbf{\Sigma}_T^{-1}$

Normal and Inverse Gamma

This is to verify the posterior distribution of the variance parameter given the conjugate prior.

Inverse Gamma:

$$f(x; \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{-(\alpha+1)} e^{-\frac{1}{x\beta}}$$

$$E(x) = \frac{1}{(\alpha - 1)\beta}$$

$$\text{var}(x) = \frac{1}{(\alpha - 1)^2(\alpha - 2)\beta^2}$$

Multivariate Normal:

$$f(\mathbf{y}; \boldsymbol{\mu}, \mathbf{\Sigma}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{\Sigma}|}} e^{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^{\mathrm{T}}\mathbf{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})} \tag{3.6}$$

11

If $\boldsymbol{\Sigma}$ can be written as $\sigma^2\widetilde{\boldsymbol{\Sigma}}$, i.e., a product of constant variance σ^2 and correlation matrix $\widetilde{\boldsymbol{\Sigma}}$, (3.6) becomes into:

$$f(\mathbf{y};\boldsymbol{\mu},\sigma^2,\widetilde{\boldsymbol{\Sigma}}) = \frac{1}{\sqrt{(2\pi)^n\sigma^{2n}|\widetilde{\boldsymbol{\Sigma}}|}}e^{-\frac{1}{2\sigma^2}(\mathbf{y}-\boldsymbol{\mu})^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}^{-1}(\mathbf{y}-\boldsymbol{\mu})}.$$

Suppose $\sigma^2 \sim IG(q,r)$,

$$[\mathbf{y}|\boldsymbol{\mu},\sigma^2,\widetilde{\boldsymbol{\Sigma}}][\sigma^2] \propto (\sigma^2)^{-n/2}e^{-\frac{1}{2\sigma^2}(\mathbf{y}-\boldsymbol{\mu})^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}^{-1}(\mathbf{y}-\boldsymbol{\mu})}(\sigma^2)^{-(q+1)}e^{-\frac{1}{r\sigma^2}}$$

$$= (\sigma^2)^{-(q+n/2+1)}e^{-\frac{1}{\sigma^2}\{\frac{1}{r}+\frac{(\mathbf{y}-\boldsymbol{\mu})^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}^{-1}(\mathbf{y}-\boldsymbol{\mu})}{2}\}}$$

So the posterior $[\sigma^2|\cdot]$ is $IG(q + n/2, \{1/r + 0.5(\mathbf{y} - \boldsymbol{\mu})^{\mathrm{T}}\widetilde{\boldsymbol{\Sigma}}^{-1}(\mathbf{y} - \boldsymbol{\mu})\}^{-1}$

AR(2) process

Below is to show a small trick for obtaining the inverse of an AR(2) covariance matrix without physically inverting the large covariance matrix. For an AR(2) process

$$e_t = \phi_1 e_{t-1} + \phi_2 e_{t-2} + \epsilon_t, \ \epsilon \sim N(0,\sigma^2),$$

we can get:

$$e_3 - \phi_1 e_2 - \phi_2 e_1 = \epsilon_3,$$

$$e_4 - \phi_1 e_3 - \phi_2 e_2 = \epsilon_4,$$

$$\vdots$$

Or equivalently,

$$\begin{pmatrix} -\phi_2 & -\phi_1 & 1 & & & & \\ & -\phi_2 & -\phi_1 & 1 & & & \\ & & -\phi_2 & -\phi_1 & 1 & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & -\phi_2 & -\phi_1 & 1 \\ & & & & & -\phi_2 & -\phi_1 \\ & & & & & & -\phi_2 \end{pmatrix}\begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_{n-2} \\ e_{n-1} \\ e_n \end{pmatrix} = \begin{pmatrix} \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \vdots \\ \epsilon_n \\ -\phi_1 e_n - \phi_2 e_{n-1} \\ -\phi_2 e_n \end{pmatrix}$$

12

Denote the left matrix as $\mathbf{A}$. Let $\mathbf{\Sigma} = \text{var}(\mathbf{e})$, $\text{var}(\mathbf{Ae}) = \mathbf{A}\mathbf{\Sigma}\mathbf{A}^{\text{T}} \approx \sigma^2\mathbf{I}$. Then $\mathbf{\Sigma} = \sigma^2\mathbf{A}^{-1}(\mathbf{A}^{\text{T}})^{-1}$ and $\mathbf{\Sigma}^{-1} = \frac{1}{\sigma^2}\mathbf{A}^{\text{T}}\mathbf{A}$. To avoid the edge problem at the end, we usually take a longer vector of $\mathbf{e}$ than needed.

4   Model comparison

In order to more formally compare different reconstructions, we employ the posterior predictive loss criterion proposed in Gelfand and Ghosh (1998) that considers predictive biases penalized by predictive variances. Specifically, we consider the loss function $D_k(m)$ for $k = 1, 3, 9$ and ∞ as defined in the paper and then compute $D_k(m)$ for each model (reconstruction) m, $m = 1, 2, \ldots, 40$. The results are reported in Figure 6. Those plots show that the ordering of models under $D_k(m)$ is insensitive to the choice of k, and moreover, the pattern of $D(m)$ highly matches the pattern of rmse in Figure 5(c) in the paper. Therefore, for convenience we simply examine rmse for model comparison.

5   Posterior predictive check

We select one reconstruction as shown in Figure 7b from the 40 in total to perform the posterior predictive check. This selected one is reconstructed by modeling only $\mathbf{T}_1$ in the process stage and uses information from both error-free tree-rings and pollen together with forcings.

Model Check

Since the NH global temperature is simply the weighted average of all local temperatures and our "oracle" proxies are either the original or the average of some selected local temperatures, the NH temperature and those "oracle" proxies certainly follow a linear relationship. Then the discrepancy between our reconstruction based on "oracle" proxies and the NH temperature is mainly due to the sampling errors of those selected locations. In another word, the reconstruction using "oracle proxies" reaches
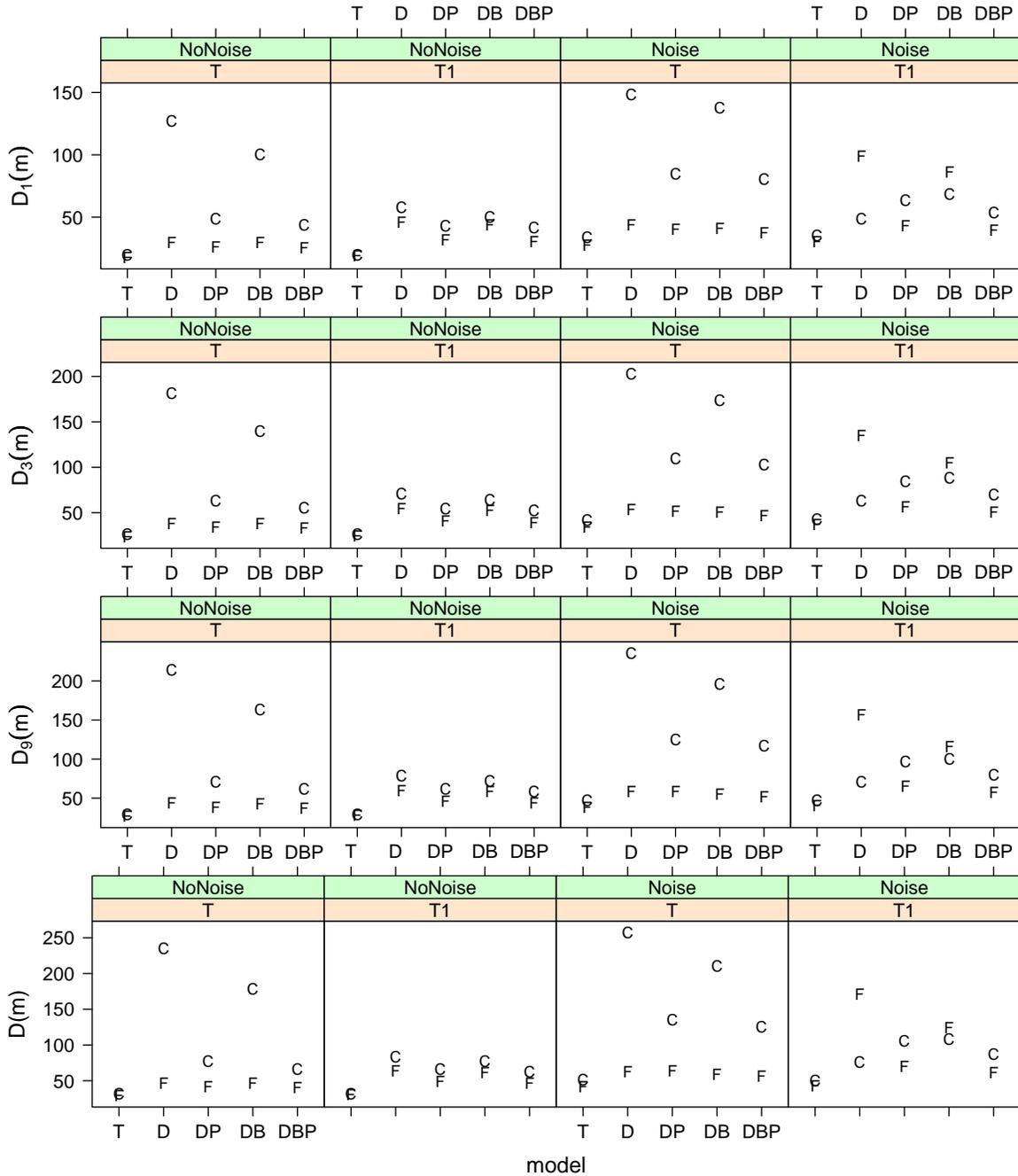
Figure 6: $D_1(m)$, $D_3(m)$, $D_9(m)$ and $D(m)$ of the reconstructions for five data models and 2^3 scenarios that are combinations of with/without forcings, with/without noise and modeling $\mathbf{T}_1/\mathbf{T}$ in the process stage. "C" and "F" are the reconstructions without forcings (with constant mean function) and with forcings incorporated, respectively.

the capacity of those selected proxy data in recovering the NH temperature, thus it provides a reference to evaluate the performance of other reconstructions based on synthetic proxies generated from the "oracle" proxies. We assess the models using two criterion as in Gel, Raftery and Geneiting (2004), the verification rank histogram (Hamill, 2001) and the coverage probability of posterior distributions.

Figure 7 compares the rank histogram of the reconstruction based on "oracle" proxies and the histogram based on tree-rings and pollen. Both have $\mathbf{T}_1$ modeled in the process stage. The two histograms share a very similar pattern that a high peak occurs at the lowest rank despite the near uniform pattern over the rest ranks. This is not extremely surprising as both reconstructions exhibit positive bias in Figure 5. We show the reconstruction based on "oracle" proxies in Figure 8. From this plot it seems that those selected local temperatures cannot recover some of the global temperatures very well. However, the similarity between the two histograms implies that our models for synthetic proxies should be appropriate, because the non-uniform distribution of the histogram for reconstructions using synthetic proxies is very likely caused by the sampling errors of the proxy data.
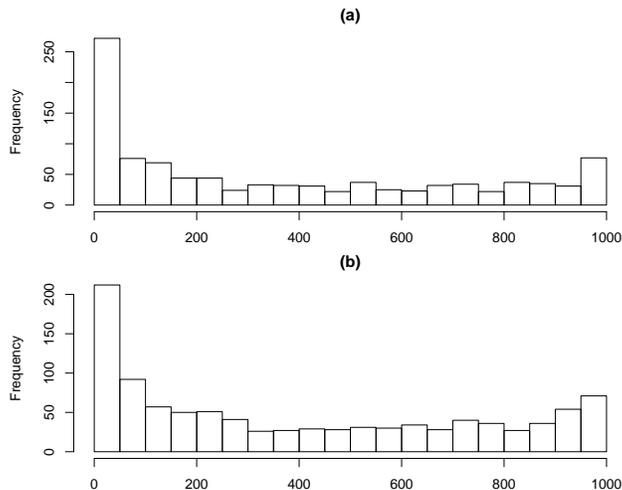
Figure 7: Verification rank histograms for (a) reconstructions using "oracle" proxy and (b) reconstructions using synthetic tree-ring and pollen

To examine the coverage probability of posterior distributions, we first choose
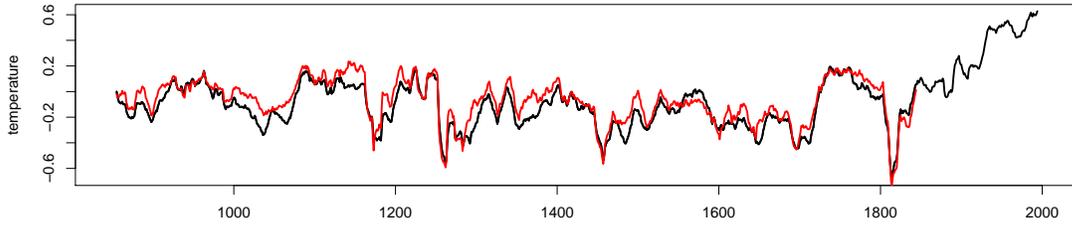
Figure 8: The reconstruction using "oracle" proxy (red) and the target temperature (black).

a sequence of probabilities as confidence levels. Then for each confidence level, we find the upper and lower bounds for each predicted temperature by identifying the corresponding quantiles from the posterior samples. After that, we calculate the proportion of the target temperatures that are covered by those confidence intervals. The results for both the reference and selected reconstructions are reported in Table 1. We see that the coverage probabilities for the two reconstructions follow the same trend and they are all lower than their corresponding nominal levels. Again, because the low coverage appears in the reference reconstruction, it is a sign that this is due to the limited capacity of the proxy data. Thus there is no evidence towards inadequacy of our models for synthetic proxies.

Table 1: Posterior predictive coverage probability for reconstructions based on synthetic proxies and "oracle" proxies

	Probability							
Confidence level	0.5	0.6	0.7	0.8	0.9	0.95	0.98	1.00
"Oracle" proxy	0.290	0.357	0.440	0.542	0.648	0.738	0.817	0.937
Synthetic proxy	0.313	0.400	0.477	0.569	0.716	0.809	0.878	0.965

Posterior density plot

Figure 9 shows the box plots of intercepts and slopes for 15 tree-ring series and Figure

16

10 shows for 10 pollen series and three components of forcings. Figure 11 shows the histograms for the time lag coefficients in AR(2) model and the variance parameters.
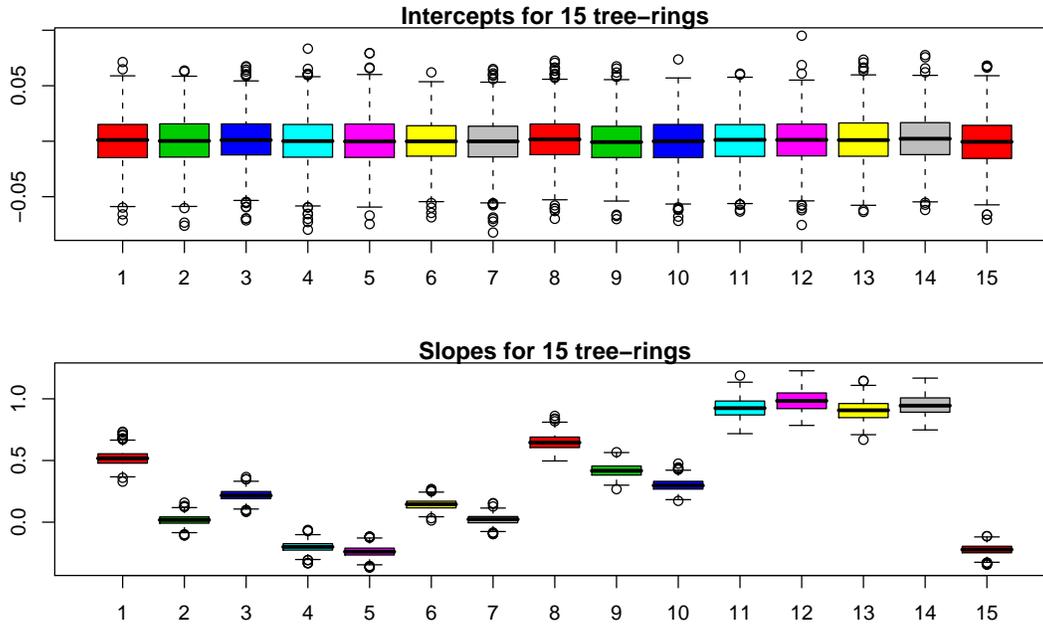
Figure 9: Box plots of μ_{iD} and β_{iD} for 15 tree-ring series

Convergence diagnostic

To check the convergence of MCMC chains, we run the MCMC starting from seven different sets of initial values and then compare the traces of parameter estimates from those different initial values. We choose μ_{5D}, β_{5D}, μ_{5P} and β_{5P} as examples for regression coefficients for tree-rings and pollen respectively to show their convergence. The other regression coefficients behave similarly. Figure 12 shows that the traces mixed very well, albeit the pollen is worse than tree-rings. This is not surprising since the pollen only has around 30 data points while the tree-ring data is annual. Figure 13 shows the regression coefficients for forcings. β_0 and β_2 that corresponds to volcanism do not seem to converge as well as the other parameters. The zero-inflated volcanism data should possibly be blamed for that. Figure 14 shows the traces plots for the time lag coefficients and variance parameters. σ_T^2 converges less impressively as the others. Again this might be caused by the zero-inflated volcanism data. Although

17
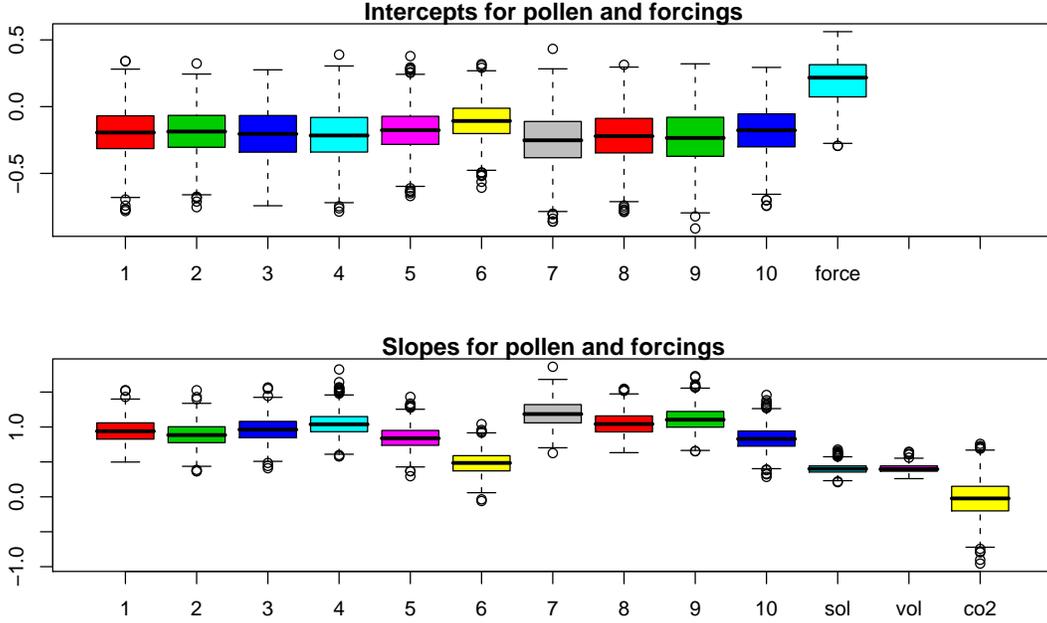
Figure 10: Box plots of μ_{jP} and β_{jP} for 10 pollen series and β's for forcings

some parameters such as β_0, β_2 and σ_{2T} do not seem to converge as well as the other parameters, our main interest, the temperature reconstructions nevertheless appear to be very insensitive to different initial values. This can be seen in Figure 15 which shows the high agreement between the seven series of posterior mean of temperature reconstructions.

References

Gel, Y., Raftery, A. E. and Gneiting, T. (2004) Calibrated probabilitic mesoscale wather field forecasting: the geostatistical output perturbation method. *Journal of the American Statistical Association*, 99, 575-583.

Gelfand, A. E. and Ghosh, S. K. (1998) Model choice: a minimum posterior predictive loss approach. *Biometrika*, 85, 1-11.

Hamill, T.M. (2001) Interpretation of rank histograms for verifying ensemble forecasts. *Monthly Weather Review*, 129, 550-560.
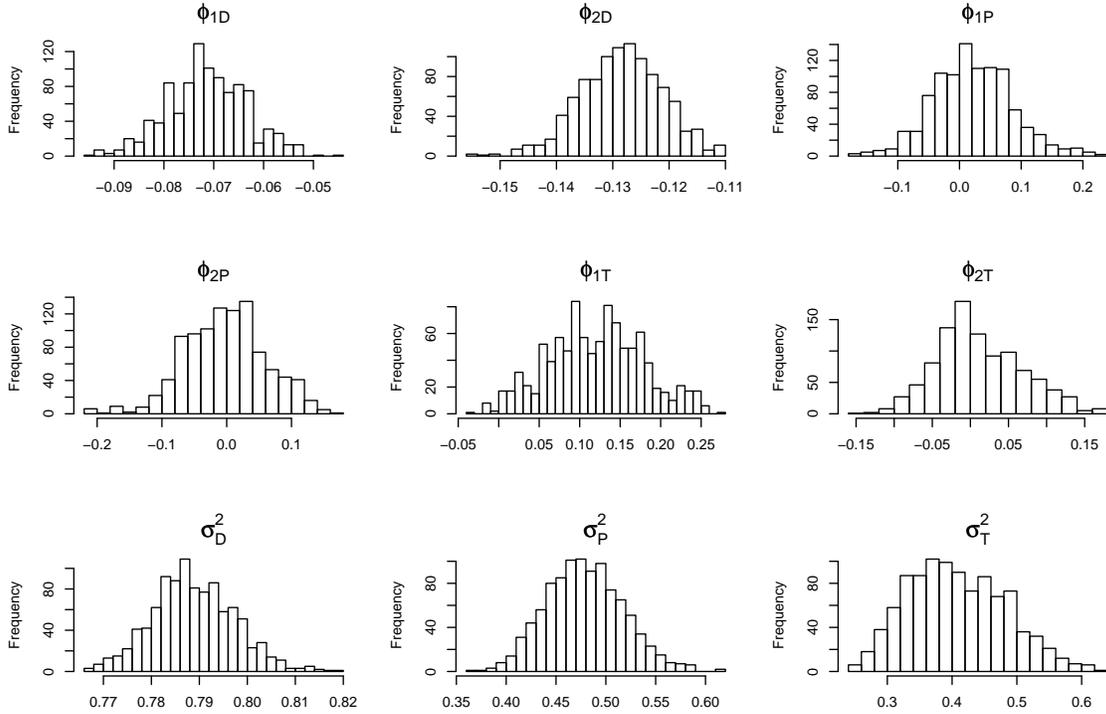
Figure 11: Histograms of time lag coefficients and variance parameters

Li, B., Nychka, W. D. and Ammann, C. M. (2007) The "Hockey Stick" and the 1990s: A statistical perspective on reconstructing hemispheric temperatures, *Tellus*, 59, 591-598.

Mann, M. E., Bradley, R. S. and Hughes, M. K. (1999) Northern Hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations. *Geophys. Res. Lett.*, 26, 759-762.

Shumway, R. H. and Stoffer, D. S. (2006) *Time Series Analysis and Its Applications: With R Examples*, Springer, New York.
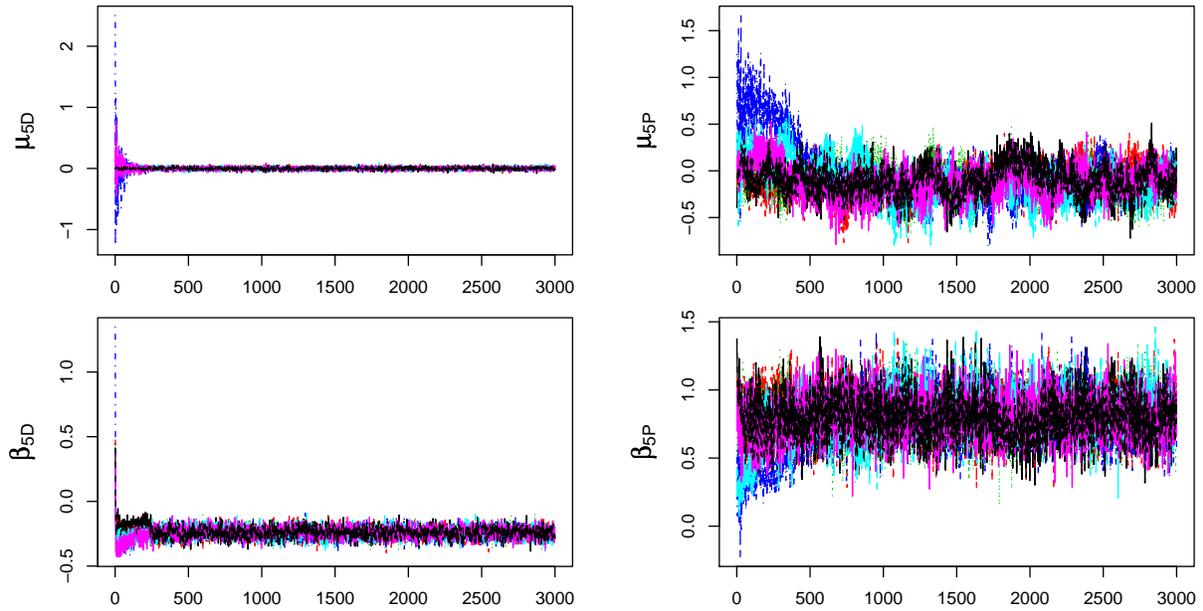
Figure 12: Trace plots of regression coefficients for tree-ring and pollen for different initial values
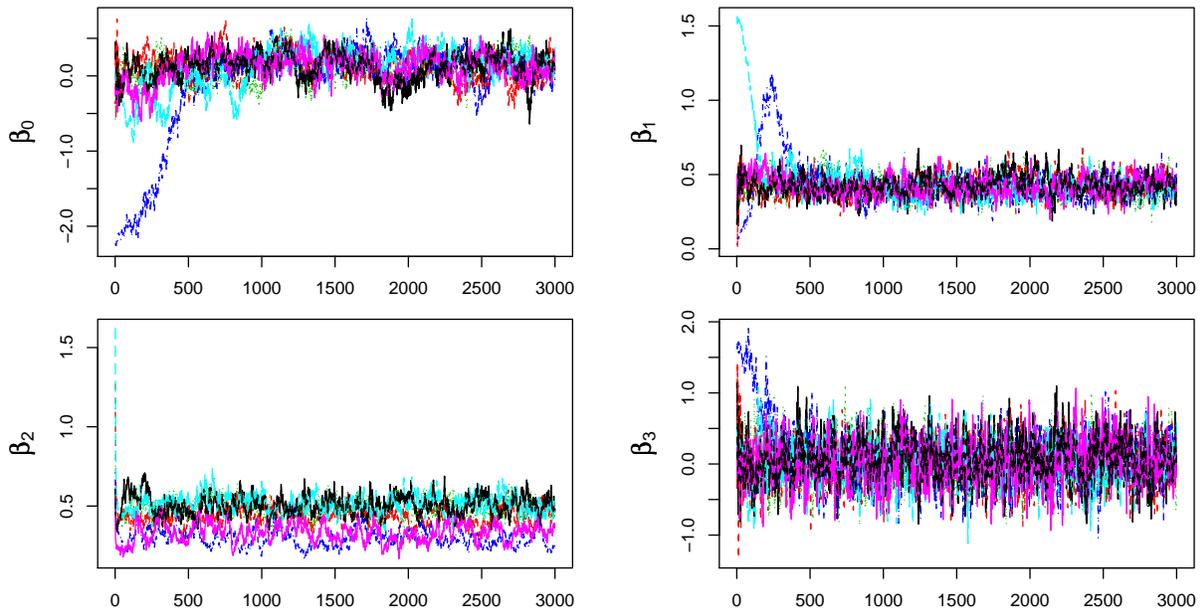
Figure 13: Trace plots of regression coefficients for forcings for different initial values
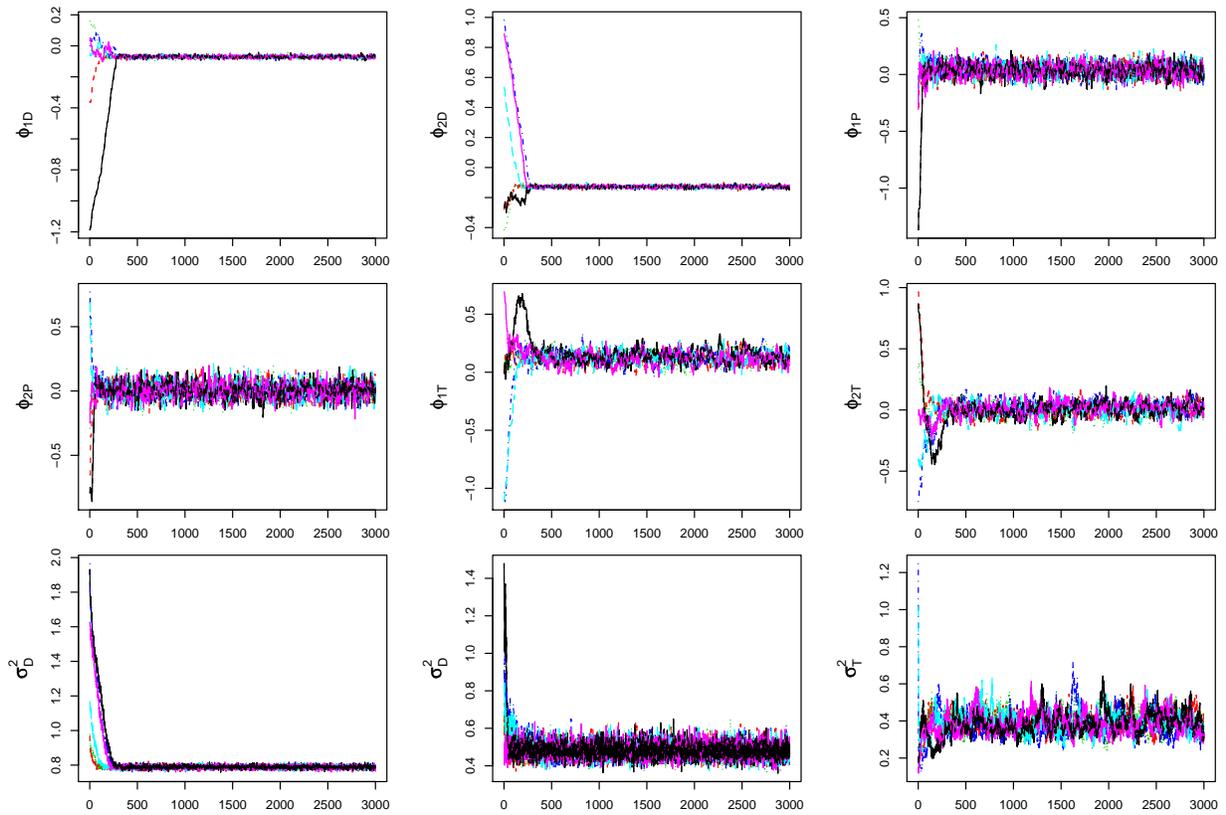
Figure 14: Trace plots of time lag coefficients and variance parameters for different initial values
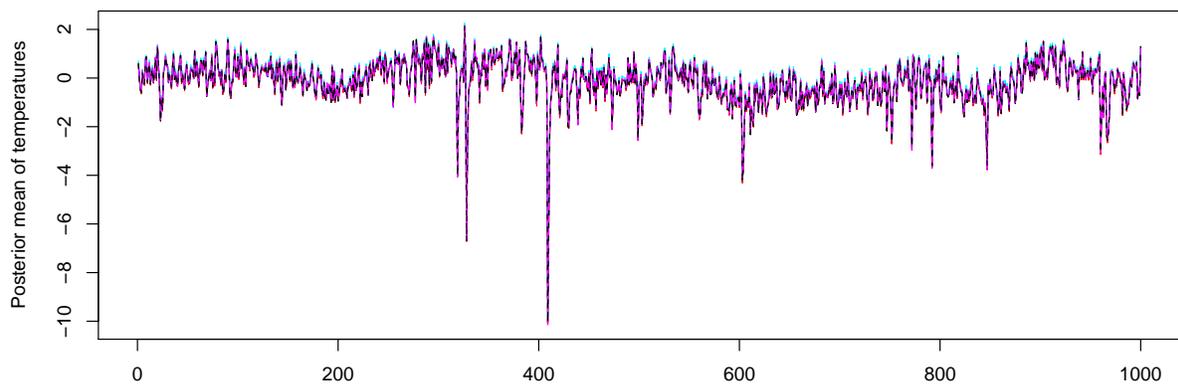
Figure 15: Posterior mean of reconstructed temperatures for seven different sets of initial values

21