Supplemental material for

"Space-time modeling of trends in temperature series"

Peter F. Craigmile

Department of Statistics, The Ohio State University, Columbus, OH 43210, USA

pfc@stat.osu.edu

Peter Guttorp

Department of Statistics, University of Washington, Seattle, WA 98105, USA

Norwegian Computing Center, Oslo, Norway

peter@stat.washington.edu

A. The full vectorized wavelet-based model

Let $\boldsymbol{s}_k$ $(k = 1, \ldots, K)$ denote the locations of the $K = 17$ temperature series. At each location k we observe $M(\boldsymbol{s}_k)$ temperature values $\boldsymbol{z}(\boldsymbol{s}_k) = (z_1(\boldsymbol{s}_k), \ldots, z_{M(\boldsymbol{s}_k)}(\boldsymbol{s}_k))^T$ drawn from the $M(\boldsymbol{s}_k)$-variate random vector (RV) $\boldsymbol{Z}(\boldsymbol{s})$. Suppose at location k that these observations correspond to times $T_t(\boldsymbol{s}_k)$, for $t = 1, \ldots, M(\boldsymbol{s}_k)$.

Since all sampling in the MCMC algorithm is performed on the wavelet domain, for each k let $\boldsymbol{W}_Z(\boldsymbol{s}_k)$ denote the vector of wavelet and scaling coefficients of the wavelet transform of $\boldsymbol{Z}(\boldsymbol{s}_k)$ analyzed to level J_μ, and partitioned as (5) of the main article. Similarly at each location k let $\boldsymbol{W}_\mu(\boldsymbol{s}_k)$ denote the coefficients of the wavelet transform of the trend $\{\mu_t(\boldsymbol{s}_k)\}$, $\boldsymbol{W}_\psi(\boldsymbol{s}_k)$ denote the coefficients of the wavelet transform of the seasonality $\{\psi_t(\boldsymbol{s}_k)\}$, and $\boldsymbol{W}_\zeta(\boldsymbol{s}_k)$ denote the coefficients of the wavelet transform of the noise $\{\zeta_t(\boldsymbol{s}_k)\}$. In our model the latent temperature process is $\{Y_t(\boldsymbol{s})\}$. Let $\boldsymbol{Y}(\boldsymbol{s}_k)$ denote this process at location $\boldsymbol{s}_k$ evaluated at the same $M(\boldsymbol{s}_k)$ times point as $\boldsymbol{Z}(\boldsymbol{s}_k)$, and let $\boldsymbol{W}_Y(\boldsymbol{s}_k)$ denote the coefficients of its wavelet transform.

Transformed into the wavelet domain, our model for $\{\boldsymbol{W}_Z(\boldsymbol{s}_k) : k = 1, \ldots, K\}$ is that

$$\boldsymbol{W}_Z(\boldsymbol{s}_k)|\boldsymbol{W}_Y(\boldsymbol{s}_k), \sigma_\epsilon^2 \;\; \sim \;\; N_{M(\boldsymbol{s}_k)}(\boldsymbol{W}_Y(\boldsymbol{s}_k), \sigma_\epsilon^2 \boldsymbol{I})$$

independently for $k = 1, \ldots, K$, where $\boldsymbol{I}$ is the identity matrix. For each k, let $\boldsymbol{D}(\boldsymbol{s}_k) = \boldsymbol{D}(\delta(\boldsymbol{s}_k), \{\alpha_l(\boldsymbol{s}_k)\}, \{\theta_l(\boldsymbol{s}_k)\})$ be the diagonal matrix of wavelet variances of the wavelet and scaling coefficients, as discussed in Section 5.1 of the main article. Then our model for $\{\boldsymbol{W}_Y(\boldsymbol{s}_k) : k = 1, \ldots, K\}$ is

$$\boldsymbol{W}_Y(\boldsymbol{s}_k)|\boldsymbol{W}_\mu(\boldsymbol{s}_k), \boldsymbol{W}_\psi(\boldsymbol{s}_k), \boldsymbol{D}(\boldsymbol{s}_k) \;\; \sim \;\; N_{M(\boldsymbol{s}_k)}(\boldsymbol{W}_\mu(\boldsymbol{s}_k) + \boldsymbol{W}_\psi(\boldsymbol{s}_k), \boldsymbol{D}(\boldsymbol{s}_k))$$

independently for $k = 1, \ldots, K$.

The model for the trend component assumes a space-time process $\{V_\tau(\boldsymbol{s})\}$ for the scaling coefficients of the trend. Letting $\boldsymbol{V}(\boldsymbol{s}_k)$ denote this process collected in a $M_{J_\mu}(\boldsymbol{s}_k)$-variate random vector at location k this implies that

$$\boldsymbol{W}_\mu(\boldsymbol{s}_k) \;\; = \;\; (\boldsymbol{0}_{M_1(\boldsymbol{s}_k)}, \ldots, \boldsymbol{0}_{M_{J_\mu(\boldsymbol{s}_k)}}, \boldsymbol{V}(\boldsymbol{s}_k))^T,$$

where $M_j(\boldsymbol{s}_k) = M(\boldsymbol{s}_k)/2^j$ denotes the number of wavelet coefficients at level j and location k. Letting $\mathbf{1}$ denote a vector of ones, we assume that $\boldsymbol{V} = (\boldsymbol{V}(\boldsymbol{s}_1), \ldots, \boldsymbol{V}(\boldsymbol{s}_K))$ is jointly Gaussian with mean $\mathbf{1}\lambda$ and space-time covariance matrix $\boldsymbol{\Psi}(\omega_V, \theta_V, \phi_v)$ with the elements of the matrix defined by equation (9) of the main article.

At each location k we let $\boldsymbol{X}_\psi(\boldsymbol{s}_k)$ denote the $M(\boldsymbol{s}_k) \times 2$ design matrix that has a first column that is the wavelet transform of $\{\sin(2\pi T_t(\boldsymbol{s}_k)/365.25)\}$, and a second column that is the wavelet transform of $\{\cos(2\pi T_t(\boldsymbol{s}_k)/365.25)\}$. Then

$$\boldsymbol{W}_\psi(\boldsymbol{s}_k) = \boldsymbol{X}_\psi(\boldsymbol{s}_k)\, \boldsymbol{\gamma}(\boldsymbol{s}_k),$$

where $\boldsymbol{\gamma}(\boldsymbol{s}_k) = (\gamma_1(\boldsymbol{s}_k), \gamma_2(\boldsymbol{s}_k))^T$. Let $\boldsymbol{E}(\omega_{\gamma_1}, \theta_{\gamma_1})$ denote the exponential covariance matrix implied by equation (9) over the K spatial locations. We assume mutual independence between $\boldsymbol{\gamma}_j \;=\; (\gamma_j(\boldsymbol{s}_1), \ldots, \gamma_j(\boldsymbol{s}_K))^T \;\sim\; N_K(\mathbf{1}\mu_{\gamma_j}, \boldsymbol{E}(\omega_{\gamma_j}, \theta_{\gamma_j}))$, for $j = 1$ and $j = 2$.

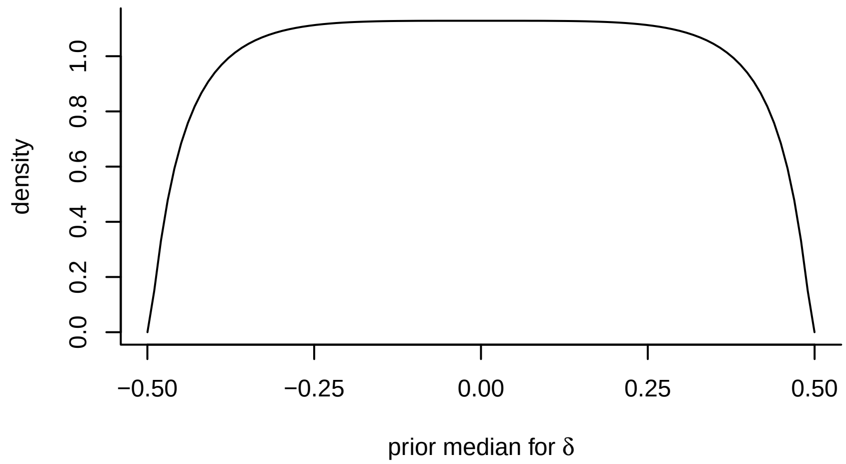

Figure 1: The prior probability density function for the median of $\{\delta(\boldsymbol{s})\}$.

We complete the model by assuming that the following are all mutually independent:

$$\begin{aligned}
\boldsymbol{\alpha}_j &= (\alpha_j(\boldsymbol{s}_1), \ldots, \alpha_j(\boldsymbol{s}_K))^T &\sim N_K(\mathbf{1}\mu_{\alpha_j}, \boldsymbol{E}(\omega_{\alpha_j}, \theta_{\alpha_j})), \qquad j = 0, \ldots, 4; \\
\boldsymbol{\kappa} &= (\kappa(\boldsymbol{s}_1), \ldots, \kappa(\boldsymbol{s}_K))^T &\sim N_K(\mathbf{1}\mu_{\kappa}, \boldsymbol{E}(\omega_{\kappa}, \theta_{\kappa})); \\
\boldsymbol{\theta}_j &= (\theta_j(\boldsymbol{s}_1), \ldots, \theta_j(\boldsymbol{s}_K))^T &\sim N_K(\mathbf{1}\mu_{\theta_j}, \boldsymbol{E}(\omega_{\theta_j}, \theta_{\theta_j})), \qquad j = 1, \ldots, p,
\end{aligned}$$

where $\boldsymbol{E}(\omega_{\kappa}, \theta_{\kappa})$, $\{\boldsymbol{E}(\omega_{\alpha_j}, \theta_{\alpha_j})\}$, and $\{\boldsymbol{E}(\omega_{\theta_j}, \theta_{\theta_j})\}$ are all exponential covariance matrices parameterized in a similar fashion to (9) of the main article, but evaluated at different hyperparameter values.

All remaining priors are mutually independent and defined in Table 1 of the main article. To illustrate the prior weight that we place on the long memory spatial process, Figure 1 displays the prior probability density function for the median of δ (the transformation of μ_{κ} from the $\{\kappa(\boldsymbol{s})\}$ scale to the $\{\delta(\boldsymbol{s})\}$ scale).

B. The MCMC algorithm

Where appropriate we use sparse matrix calculations in practice.

Sampling the latent temperature process

At each location k, we draw $\boldsymbol{W}_Y(\boldsymbol{s}_k)$ from $N_{M(\boldsymbol{s}_k)}(\boldsymbol{P}^{-1}\boldsymbol{m}, \boldsymbol{P}^{-1})$, with

$$\boldsymbol{m} \;=\; \boldsymbol{w}_z(\boldsymbol{s}_k)/\sigma_\epsilon^2 + \boldsymbol{D}(\boldsymbol{s}_k)^{-1}(\boldsymbol{W}_\mu(\boldsymbol{s}_k) + \boldsymbol{W}_\psi(\boldsymbol{s}_k)), \quad \text{and} \quad \boldsymbol{P} \;=\; \boldsymbol{D}(\boldsymbol{s}_k)^{-1} + \sigma_\epsilon^{-2}\boldsymbol{I}.$$

Sampling the parameters characterizing the trend

Let $M_\mu = \sum_{k=1}^{K} M_{J_\mu}(\boldsymbol{s}_k)$ denote the total number of scaling coefficients observed at all the spatial locations. We sample $\boldsymbol{V}$ from $N_{M_\mu}(\boldsymbol{P}^{-1}\boldsymbol{m}, \boldsymbol{P}^{-1})$, with

$$\begin{aligned}
\boldsymbol{m} &= \text{diag}\{\boldsymbol{D}_V(\boldsymbol{s}_1)^{-1}, \ldots, \boldsymbol{D}_V(\boldsymbol{s}_K)^{-1}\}(\boldsymbol{V}_{DS}(\boldsymbol{s}_1), \ldots, \boldsymbol{V}_{DS}(\boldsymbol{s}_1)) + \lambda\boldsymbol{\Psi}(\omega_V, \theta_V, \phi_v)^{-1}\mathbf{1} \text{ and} \\
\boldsymbol{P} &= \text{diag}\{\boldsymbol{D}_V(\boldsymbol{s}_1)^{-1}, \ldots, \boldsymbol{D}_V(\boldsymbol{s}_K)^{-1}\} + \boldsymbol{\Psi}(\omega_V, \theta_V, \phi_v)^{-1},
\end{aligned}$$

where $\boldsymbol{V}_{DS}(\boldsymbol{s}_k)$ denotes the vector of scaling coefficients of the wavelet transform of deseasonalized data for each k (i.e., the last $M_{J_\mu}(\boldsymbol{s}_k)$ values of $\boldsymbol{W}_Y(\boldsymbol{s}_k) - \boldsymbol{W}_\psi(\boldsymbol{s}_k)$), and $\boldsymbol{D}_V(\boldsymbol{s}_k)$ is the diagonal covariance matrix of the scaling coefficients at location $\boldsymbol{s}_k$ as defined at the bottom of page 19 of the main article.

With the prior $\lambda \sim N(m_\lambda, v_\lambda)$, we sample λ conditional on all other parameters and the data using a $N(m_\lambda^P, v_\lambda^P)$ distribution where

$$m_\lambda^P \;=\; \mathbf{1}^T\boldsymbol{\Psi}(\omega_V, \theta_V, \phi_v)^{-1}\boldsymbol{V} + v_\lambda^{-1}m_\lambda \quad \text{and} \quad v_\lambda^P \;=\; \mathbf{1}^T\boldsymbol{\Psi}(\omega_V, \theta_V, \phi_v)^{-1}\mathbf{1} + v_\lambda^{-1}.$$

With the prior $1/\omega_V \sim \text{Gamma}(s_{\omega_V}, r_{\omega_V})$ we draw $1/\omega_V$ from a Gamma($s^P_{\omega_V}$, $r^P_{\omega_V}$) distribution with

$$s^P_{\omega_V} \;=\; s_{\omega_V} + 0.5M_\mu \quad \text{and} \quad r^P_{\omega_V} \;=\; r_{\omega_V} + 0.5(\boldsymbol{V} - \mathbf{1}\lambda)^T \boldsymbol{\Psi}_R(\theta_V, \phi_v)^{-1}(\boldsymbol{V} - \mathbf{1}\lambda).$$

where $\boldsymbol{\Psi}_R(\theta_V, \phi_v)$ is the correlation matrix for the space-time process (the covariance matrix $\boldsymbol{\Psi}(\omega_V, \theta_V, \phi_v)$ scaled by ω_V).

We use a Metropolis-Hastings algorithm to jointly sample the parameters θ_V and ϕ_V. We assume a prior for θ_V of Gamma($s_{\theta_V}, r_{\theta_V}$), and a prior for ϕ_V of $N_{[-1,1]}(0.4, 0.2^2)$ (a normal distribution truncated to the interval $[-1, 1]$). Given that we are currently at $\log \theta_V^{curr}$, we propose $\log \theta_V^{new}$ from a N($\log \theta_V^{curr}$, v_{prop}) distribution where v_{prop} is set equal to half the estimated variance of the MLE of $\log \theta_V^{new}$ based on a preliminary exploratory data analysis. Given a current value of ϕ_V^{curr} we propose ϕ_V^{new} from a $N_{[-1,1]}(\phi_V^{curr}, 0.05^2)$. We accept these two proposals jointly with probability

$$\min\{\exp(lp(\theta_V^{new}, \phi_V^{new}|\theta_V^{curr}, \phi_V^{curr}) - lp(\theta_V^{curr}, \phi_V^{curr}|\theta_V^{new}, \phi_V^{new})), 1\}$$

or remain at the current values, where for example

$$\begin{aligned} lp(\theta_V^{new}, \phi_V^{new}|\theta_V^{curr}, \phi_V^{curr}) \;&=\; \log f(\boldsymbol{V}|\mathbf{1}\lambda, \boldsymbol{\Psi}(\omega_V, \theta_V^{new}, \phi_v^{new})) + \log f(\theta_V^{new}|s_{\theta_V}, r_{\theta_V}) + \log \theta_V^{new} + \\ &\quad \log f_{[-1,1]}(\phi_V^{new}|0.4, 0.2^2) - \log f_{[-1,1]}(\phi_V^{new}|\phi_V^{curr}, 0.05^2), \end{aligned}$$

a log multivariate normal density, plus a log gamma density, plus a log Jacobian term, plus a truncated normal log density, minus the truncated normal log density for the proposal for ϕ_V.

Sampling the parameters characterizing the seasonality

Let $\boldsymbol{X}_{j,s}(\boldsymbol{s}_k)$ denote the jth column of $\boldsymbol{X}_\psi(\boldsymbol{s}_k)$, and let $\boldsymbol{X}_\psi$ be the column bind of

$$\text{diag}\{X_{1,s}(\boldsymbol{s}_1),\ldots,X_{1,s}(\boldsymbol{s}_K)\} \quad \text{and} \quad \text{diag}\{X_{2,s}(\boldsymbol{s}_1),\ldots,X_{2,s}(\boldsymbol{s}_K)\}.$$

Then we sample $\boldsymbol{\gamma}=(\gamma_1(\boldsymbol{s}_1),\ldots,\gamma_1(\boldsymbol{s}_K),\gamma_2(\boldsymbol{s}_2),\ldots,\gamma_1(\boldsymbol{s}_K))$ from $N_{2K}(\boldsymbol{P}^{-1}\boldsymbol{m},\boldsymbol{P}^{-1})$, with

$$\begin{aligned}\boldsymbol{m} &= \boldsymbol{X}_\psi^T\text{diag}\{\boldsymbol{D}(\boldsymbol{s}_1)^{-1},\ldots,\boldsymbol{D}(\boldsymbol{s}_K)^{-1}\}\left[(\boldsymbol{W}_Y(\boldsymbol{s}_1),\ldots,\boldsymbol{W}_Y(\boldsymbol{s}_K))-(\boldsymbol{W}_\mu(\boldsymbol{s}_1),\ldots,\boldsymbol{W}_\mu(\boldsymbol{s}_K))\right]+\\ &\quad \text{diag}\{\boldsymbol{E}(\omega_{\gamma_1},\theta_{\gamma_1})^{-1},\boldsymbol{E}(\omega_{\gamma_2},\theta_{\gamma_2}))^{-1}\kappa, \text{ and}\\ \boldsymbol{P} &= \boldsymbol{X}_\psi^T\text{diag}\{\boldsymbol{D}(\boldsymbol{s}_1)^{-1},\ldots,\boldsymbol{D}(\boldsymbol{s}_K)^{-1}\}\boldsymbol{X}_\psi+\text{diag}\{\boldsymbol{E}(\omega_{\gamma_1},\theta_{\gamma_1})^{-1},\boldsymbol{E}(\omega_{\gamma_2},\theta_{\gamma_2}))^{-1},\end{aligned}$$

where κ is a vector of K copies of μ_{γ_1} followed by K copies of μ_{γ_2}.

For $j=1,2$, suppose that μ_{γ_j} has a prior $\text{N}(m_{\gamma_j},\, v_{\gamma_j})$ distribution. For each j we update μ_{γ_j} with a draw from a $\text{N}(m^P_{\gamma_j},\, v^P_{\gamma_j})$ distribution, with

$$m^P_{\gamma_j} \;=\; \mathbf{1}^T\boldsymbol{E}(\omega_{\gamma_j},\theta_{\gamma_j})^{-1}\boldsymbol{\gamma}_j+v^{-1}_{\gamma_j}m_{\gamma_j} \quad \text{and} \quad v^P_{\gamma_j} \;=\; \mathbf{1}^T\boldsymbol{E}(\omega_{\gamma_j},\theta_{\gamma_j})^{-1}\mathbf{1}+v^{-1}_{\gamma_j}.$$

For $j=1,2$, assuming the prior $1/\omega_{\gamma_j} \sim \text{Gamma}(s_{\omega_{\gamma_j}},r_{\omega_{\gamma_j}})$ we sample $1/\omega_{\gamma_j}$ from a $\text{Gamma}(s^P_{\omega_{\gamma_j}},\, r^P_{\omega_{\gamma_j}})$ distribution with

$$s^P_{\omega_{\gamma_j}} \;=\; s_{\omega_{\gamma_j}}+0.5K, \quad \text{and} \quad r^P_{\omega_{\gamma_j}} \;=\; r_{\omega_{\gamma_j}}+0.5(\boldsymbol{\gamma}_j-\mathbf{1}\mu_{\gamma_j})^T\boldsymbol{E}_R(\theta_{\gamma_j})^{-1}(\boldsymbol{\gamma}_j-\mathbf{1}\mu_{\gamma_j}),$$

where $\boldsymbol{E}_R(\theta_{\gamma_j})$ is the exponential correlation matrix.

We update θ_{γ_j} $(j=1,2)$ using a Metropolis-Hastings step on the log-transformed scale. We assume a prior for θ_{γ_j} of $\text{Gamma}(s_{\theta_{\gamma_j}},r_{\theta_{\gamma_j}})$. Given that we are currently at $\log\theta^{curr}_{\gamma_j}$, we propose $\log\theta^{new}_{\gamma_j}$ from a $\text{N}(\log\theta^{curr}_{\gamma_j},\, v_{prop})$ distribution where v_{prop} is set equal to half the estimated variance of the MLE of $\log\theta^{new}_{\gamma_j}$ based on a preliminary data analysis. We accept $\log\theta^{new}_{\gamma_j}$ with probability $\min\{\exp(lp(\theta^{new}_{\gamma_j})-lp(\theta^{curr}_{\gamma_j})),1\}$ or stay at $\log\theta^{curr}_{\gamma_j}$, where

$$lp(\theta_{\gamma_j}) \;=\; \log f(\boldsymbol{\gamma}_j|\mathbf{1}\mu_{\gamma_j},\boldsymbol{E}(\omega_{\gamma_j},\theta_{\gamma_j}))+\log f(\theta_{\gamma_j}|s_{\theta_{\gamma_j}},r_{\theta_{\gamma_j}})+\log\theta_{\gamma_j},$$

a log multivariate normal density, plus a log gamma density, plus a log Jacobian term.

Sampling the parameters characterizing the noise

We update the parameters $(\kappa(\boldsymbol{s}_k), \{\alpha_l(\boldsymbol{s}_k)\}, \{\theta_l(\boldsymbol{s}_k)\})$ simultaneously using a single Metropolis-Hastings step. Given we are at $\{\boldsymbol{\alpha}_j^{curr}\}$, $\boldsymbol{\kappa}^{curr}$, and $\{\boldsymbol{\theta}_j^{curr}\}$ we propose $\{\boldsymbol{\alpha}_j^{new}\}$, $\boldsymbol{\kappa}^{new}$, and $\{\boldsymbol{\theta}_j^{new}\}$ independently using a random walk multivariate normal proposal. Let

$$\boldsymbol{D}^{new} = \mathrm{diag}\{\boldsymbol{D}_V^{new}(\boldsymbol{s}_1), \ldots, \boldsymbol{D}_V^{new}(\boldsymbol{s}_K)\}$$

denote the joint covariance matrix of the wavelet transform of the noise term evaluated at the proposed parameter values, and $\boldsymbol{D}^{curr}$ the current joint covariance matrix. We accept the proposed parameters with probability $\min\{\exp(lp(new) - lp(curr)), 1\}$ or otherwise we stay at the current values, where for example

$$\begin{aligned} lp(new) = {} & \log f(\boldsymbol{W}_Y - \boldsymbol{W}_\mu - \boldsymbol{W}_\psi | \boldsymbol{0}, \boldsymbol{D}^{new}) + \log f(\boldsymbol{\kappa}^{new} | \boldsymbol{1}\mu_\kappa, \boldsymbol{E}(\omega_\kappa, \theta_\kappa)) + \\ & \sum_{j=0}^{4} \log f(\boldsymbol{\alpha}_j^{new} | \boldsymbol{1}\mu_{\alpha_j}, \boldsymbol{E}(\omega_{\alpha_j}, \theta_{\alpha_j})) + \sum_{l=1}^{p} \log f(\boldsymbol{\theta}_l^{new} | \boldsymbol{1}\mu_{\theta_l}, \boldsymbol{E}(\omega_{\theta_l}, \theta_{\theta_l})) - \\ & \sum_{k=1}^{K} \left\{\log(0.5 + \delta^{new}(\boldsymbol{s}_k)) + \log(0.5 - \delta^{new}(\boldsymbol{s}_k))\right\}, \end{aligned}$$

are sums of log multivariate normal densities plus the Jacobian for transforming between $\kappa(\boldsymbol{s})$ and $\delta(\boldsymbol{s})$. Here $\delta^{new}(\boldsymbol{s}_k))$ is the transformation of $\kappa^{new}(\boldsymbol{s}_n k))$.

We update the hyperparameters of each set of spatially-varying parameters using similar updates as for the hyperparameters of $\boldsymbol{\gamma}_j$. We update the parameters characterizing each spatial field separately.