Spatial Predictive Process Model

John Tipton

February 28, 2014

1 Full Dimensional Model Statement

1.1 Data Model

$$\boldsymbol{y}_t = \boldsymbol{H}_t \boldsymbol{X} \boldsymbol{\beta}_t + \boldsymbol{\eta}_t + \boldsymbol{\epsilon}_t$$

1.2 Process Model

$$\begin{aligned}
\boldsymbol{\beta}_t &\sim N(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta) & \boldsymbol{\Sigma}_\beta &= \sigma^2_\beta \boldsymbol{I}_t & & \\
\boldsymbol{\eta}_t &\sim N(0, \boldsymbol{\Sigma}_\eta) & \boldsymbol{\Sigma}_\eta &= \sigma^2_\eta \boldsymbol{R}(\phi) & \boldsymbol{R}(\phi) &= \exp\left(-\boldsymbol{D}_t/\phi\right) \\
\boldsymbol{\epsilon}_t &\sim N(0, \boldsymbol{\Sigma}_\epsilon) & \boldsymbol{\Sigma}_\epsilon &= \sigma^2_\epsilon \boldsymbol{I}_t & &
\end{aligned}$$

1.3 Parameter Model

$$\begin{aligned}
\boldsymbol{\mu}_\beta &\sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \\
\sigma^2_\beta &\sim IG(\alpha_\beta, \beta_\beta) \\
\sigma^2_\eta &\sim IG(\alpha_\eta, \beta_\eta) \\
\sigma^2_\epsilon &\sim IG(\alpha_\epsilon, \beta_\epsilon) \\
\phi &\sim IG(\alpha_\phi, \beta_\phi)
\end{aligned}$$

where $\boldsymbol{I}_\beta$ is the identity matrix of size $\tau \times \tau$ where τ is the number of parameters in $\boldsymbol{\beta}_t$, $\boldsymbol{I}_t$ is the identity matrix of size $n_t \times n_t$ and n_t is the number of samples of y_t at time t and $\boldsymbol{D}_t$ is the distance matrix between locations observed at time t. Define $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_\eta + \boldsymbol{\Sigma}_\epsilon$

2 Predictive Process

For large dimensional spatial processes it can be computationally expensive to invert $\boldsymbol{\Sigma}_t$ over the set of all locations $\boldsymbol{s}_t$ at time t from the set of all locations $\boldsymbol{S}$. This motivates the use of a predictive process $\tilde{\boldsymbol{\eta}}$ to approximate $\boldsymbol{\eta}_t$ over a set of knots $\boldsymbol{S}^*$ where $\tilde{\boldsymbol{\eta}} = \boldsymbol{c}_t(\boldsymbol{s}_t, \boldsymbol{s}^* | \sigma^2_\eta, \phi)^T \boldsymbol{C}^{*-1}(\boldsymbol{s}^*, \boldsymbol{s}^* | \sigma^2_\eta, \phi)\boldsymbol{\eta}*$. The covariance between the desired locations $\boldsymbol{s}_t \in \boldsymbol{S}$ at time t and the set of knots $\boldsymbol{s}^* \in \boldsymbol{S}^*$ is $\boldsymbol{c}_t(\boldsymbol{s}_t, \boldsymbol{s}^* | \sigma^2_\eta, \phi)$. The covariance matrix over the set of knots is $\boldsymbol{C}^{*-1}(\boldsymbol{s}^*, \boldsymbol{s}^* | \sigma^2_\eta, \phi)$. The lower dimensional $\boldsymbol{\eta}^* \sim \text{MVN}(\boldsymbol{0}, \boldsymbol{C}^{*-1}(\boldsymbol{s}^*, \boldsymbol{s}^* | \sigma^2_\eta, \phi))$. Equivalently, $\tilde{\boldsymbol{\eta}} \sim MVN(\boldsymbol{0}, \boldsymbol{c}_t^T \boldsymbol{C}^{*-1} \boldsymbol{c}_t)$ independent of $\boldsymbol{\epsilon}_t$. This model needs to be modified so as to not underestimate the variance. This is done by replacing $\tilde{\boldsymbol{\eta}}_t$ with $\tilde{\boldsymbol{\eta}}_t + \tilde{\boldsymbol{\epsilon}}_t$ where $\tilde{\boldsymbol{\epsilon}}_t = MVN(\boldsymbol{0}, \sigma^2_\eta \boldsymbol{I}_{n_t} - \text{diag}(\boldsymbol{c}_t^T \boldsymbol{C}^{*-1} \boldsymbol{c}_t))$

2.1 Data Model

$$\boldsymbol{y}_t = \boldsymbol{H}_t \boldsymbol{X} \boldsymbol{\beta}_t + \tilde{\boldsymbol{\eta}}_t + \tilde{\boldsymbol{\epsilon}}_t + \boldsymbol{\epsilon}_t$$

2.2 Process Model

$$\begin{aligned}
\boldsymbol{\beta}_t &\sim N(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta) & \boldsymbol{\Sigma}_\beta &= \sigma^2_\beta \boldsymbol{I}_t \\
\tilde{\boldsymbol{\eta}}_t &\sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_{\tilde{\eta}}) & \boldsymbol{\Sigma}_{\tilde{\eta}} &= \boldsymbol{c}^T \boldsymbol{C}^{*-1} \boldsymbol{c} \\
\tilde{\boldsymbol{\epsilon}}_t &\sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_{\tilde{\epsilon}}) & \boldsymbol{\Sigma}_{\tilde{\epsilon}} &= \sigma^2_\eta \boldsymbol{I}_t - diag(\boldsymbol{c}^T \boldsymbol{C}^{*-1} \boldsymbol{c}) \\
\boldsymbol{\epsilon}_t &\sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_\epsilon) & \boldsymbol{\Sigma}_\epsilon &= \sigma^2_\epsilon \boldsymbol{I}_t
\end{aligned}$$

2.3 Parameter Model

$$\begin{aligned}
\boldsymbol{\mu}_\beta &\sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \\
\sigma^2_\beta &\sim IG(\alpha_\beta, \beta_\beta) \\
\sigma^2_\eta &\sim IG(\alpha_\eta, \beta_\eta) \\
\sigma^2_\epsilon &\sim IG(\alpha_\epsilon, \beta_\epsilon) \\
\phi &\sim IG(\alpha_\phi, \beta_\phi)
\end{aligned}$$

where $\boldsymbol{I}_\beta$ is the identity matrix of size $\tau \times \tau$ where τ is the number of parameters in $\boldsymbol{\beta}_t$, $\boldsymbol{I}_t$ is the identity matrix of size $n_t \times n_t$ and n_t is the number of samples of y_t at time t and $\boldsymbol{D}_t$ is the distance matrix between locations observed at time t. Define $\boldsymbol{\Sigma}_{\tilde{\epsilon}+\epsilon} = \boldsymbol{\Sigma}_{\tilde{\epsilon}} + \boldsymbol{\Sigma}_\epsilon$ and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\tilde{\eta}} + \boldsymbol{\Sigma}_{\tilde{\epsilon}+\epsilon}$. Each MCMC iteration requires evaluation of the inverse and determinant of $\boldsymbol{\Sigma}$. This is accomplished through the use of the Sherman-Woodbury-Morrison equations for the inverse

$$\begin{aligned}
\boldsymbol{\Sigma}^{-1} &= (\boldsymbol{\Sigma}_{\tilde{\eta}} + \boldsymbol{\Sigma}_{\tilde{\epsilon}+\epsilon})^{-1} \\
&= (\boldsymbol{c}^T {\boldsymbol{C}^*}^{-1} \boldsymbol{c} + \boldsymbol{\Sigma}_{\tilde{\epsilon}+\epsilon})^{-1} \\
&= \boldsymbol{\Sigma}^{-1}_{\tilde{\epsilon}+\epsilon} + \boldsymbol{\Sigma}^{-1}_{\tilde{\epsilon}+\epsilon} \boldsymbol{c}^T \left(\boldsymbol{C}^* + \boldsymbol{c} \boldsymbol{\Sigma}^{-1}_{\tilde{\epsilon}+\epsilon} \boldsymbol{c}^T \right)^{-1} \boldsymbol{c} \boldsymbol{\Sigma}^{-1}_{\tilde{\epsilon}+\epsilon}
\end{aligned}$$

and the determinant

$$\begin{aligned}
|\boldsymbol{\Sigma}| &= |\boldsymbol{\Sigma}_{\tilde{\eta}} + \boldsymbol{\Sigma}_{\tilde{\epsilon}+\epsilon}| \\
&= |\boldsymbol{c}^T {\boldsymbol{C}^*}^{-1} \boldsymbol{c} + \boldsymbol{\Sigma}_{\tilde{\epsilon}+\epsilon}| \\
&= |\boldsymbol{C}^* + \boldsymbol{c} \boldsymbol{\Sigma}^{-1}_{\tilde{\epsilon}+\epsilon} \boldsymbol{c}^T| |{\boldsymbol{C}^*}^{-1}| |\boldsymbol{\Sigma}_{\tilde{\epsilon}+\epsilon}|
\end{aligned}$$

3 Posterior

$$\prod_{t=1}^T [\boldsymbol{\beta}_t, \boldsymbol{\mu}_\beta, \sigma^2_\beta, \sigma^2_\eta, \sigma^2_\epsilon, \phi | \boldsymbol{y}_t] \propto \prod_{t=1}^T [\boldsymbol{y}_t | \boldsymbol{\beta}_t, \sigma^2_\eta, \phi, \sigma^2_\epsilon][\boldsymbol{\beta}_t | \boldsymbol{\mu}_\beta, \sigma^2_\beta][\boldsymbol{\mu}_\beta][\sigma^2_\beta][\sigma^2_\eta][\sigma^2_\epsilon][\phi]$$

4 Full Conditionals

4.1 Full Conditional for $\boldsymbol{\beta}_t$

For $t = 1, \ldots, T$,

$$
\begin{aligned}
[\boldsymbol{\beta}_t|\cdot] &\propto [\boldsymbol{y}_t|\boldsymbol{\beta}_t, \sigma^2_\eta, \sigma^2_\epsilon, \phi][\boldsymbol{\beta}_t|\boldsymbol{\mu}_\beta, \sigma^2_\beta] \\
&\propto e^{-\frac{1}{2}(\boldsymbol{y}_t - \boldsymbol{H}_t\boldsymbol{X}\boldsymbol{\beta}_t)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_t - \boldsymbol{H}_t\boldsymbol{X}\boldsymbol{\beta}_t)} e^{-\frac{1}{2}(\boldsymbol{\beta}_t - \boldsymbol{\mu}_\beta)^T\boldsymbol{\Sigma}_\beta^{-1}(\boldsymbol{\beta}_t - \boldsymbol{\mu}_\beta)} \\
&\propto e^{-\frac{1}{2}\{\boldsymbol{\beta}_t^T(\boldsymbol{X}^T\boldsymbol{H}_t^T\boldsymbol{\Sigma}^{-1}\boldsymbol{H}_t\boldsymbol{X} + \boldsymbol{\Sigma}_\beta^{-1})\boldsymbol{\beta}_t - 2\boldsymbol{\beta}_t^T(\boldsymbol{X}^T\boldsymbol{H}_t^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_t + \boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\mu}_\beta)\}}
\end{aligned}
$$

which is Normal with mean $\boldsymbol{A}^{-1}\boldsymbol{b}$ and variance $\boldsymbol{A}^{-1}$ where

$$
\begin{aligned}
\boldsymbol{A}^{-1} &= (\boldsymbol{X}^T\boldsymbol{H}_t^T\boldsymbol{\Sigma}^{-1}\boldsymbol{H}_t\boldsymbol{X} + \boldsymbol{\Sigma}_\beta^{-1})^{-1} \\
\boldsymbol{b} &= (\boldsymbol{X}^T\boldsymbol{H}_t^T\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_t + \boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\mu}_\beta)
\end{aligned}
$$

4.2 Full Conditional for $\boldsymbol{\mu}_\beta$

$$
\begin{aligned}
[\boldsymbol{\mu}_\beta|\cdot] &\propto \prod_{t=1}^T [\boldsymbol{\beta}_t|\boldsymbol{\mu}_\beta, \sigma^2_\beta][\boldsymbol{\mu}_\beta] \\
&\propto e^{-\frac{1}{2}\sum_{t=1}^T(\boldsymbol{\beta}_t - \boldsymbol{\mu}_\beta)^T\boldsymbol{\Sigma}_\beta^{-1}(\boldsymbol{\beta}_t - \boldsymbol{\mu}_\beta)} e^{-\frac{1}{2}(\boldsymbol{\mu}_\beta - \boldsymbol{\mu}_0)^T\Sigma_0^{-1}(\boldsymbol{\mu}_\beta - \boldsymbol{\mu}_0)} \\
&\propto e^{-\frac{1}{2}(\boldsymbol{\mu}_\beta^T(T\boldsymbol{\Sigma}_\beta^{-1} + \boldsymbol{\Sigma}_0^{-1})\boldsymbol{\mu}_\beta - 2\boldsymbol{\mu}_\beta^T(\sum_{t=1}^T\boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\beta}_t + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0))}
\end{aligned}
$$

which is multivariate normal with mean $(T\boldsymbol{\Sigma}_\beta^{-1} + \boldsymbol{\Sigma}_0^{-1})^{-1}(\sum_{t=1}^T\boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\beta}_t + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0)$ and variance $(T\boldsymbol{\Sigma}_\beta^{-1} + \boldsymbol{\Sigma}_0^{-1})^{-1}$

4.3 Full Conditional for σ^2_β

$$
\begin{aligned}
[\sigma^2_\beta|\cdot] &\propto \prod_{t=1}^T [\boldsymbol{\beta}_t|\boldsymbol{\mu}_\beta, \sigma^2_\beta][\sigma^2_\beta] \\
&\propto (\prod_{t=1}^T |\boldsymbol{\Sigma}_\beta|^{-\frac{1}{2}}) e^{-\frac{1}{2}\sum_{t=1}^T(\boldsymbol{\beta}_t - \boldsymbol{\mu}_\beta)^T\boldsymbol{\Sigma}_\beta^{-1}(\boldsymbol{\beta}_t - \boldsymbol{\mu}_\beta)} (\sigma^2_\beta)^{-(\alpha_\beta+1)} e^{-\frac{\beta_\beta}{\sigma^2_\beta}} \\
&\propto (\sigma^2_\beta)^{-(\alpha_\beta + \frac{t\tau}{2} + 1)} e^{-\frac{1}{\sigma^2_\beta}(\frac{1}{2}\sum_{t=1}^T(\boldsymbol{\beta}_t - \boldsymbol{\mu}_\beta)^T(\boldsymbol{\beta}_t - \boldsymbol{\mu}_\beta) + \beta_\beta)}
\end{aligned}
$$

which is IG$(\alpha_\beta + \frac{t\tau}{2}, \frac{1}{2}\sum_{t=1}^T(\boldsymbol{\beta}_t - \boldsymbol{\mu}_\beta)^T(\boldsymbol{\beta}_t - \boldsymbol{\mu}_\beta) + \beta_\beta)$ since the determinant $|\boldsymbol{\Sigma}_\beta| = (\sigma^2_\beta)^\tau$ and $\boldsymbol{\Sigma}_\beta^{-1} = \frac{1}{\sigma^2_\beta}\boldsymbol{I}_t$

4.4 Full Conditional for σ^2_η

$$
\begin{aligned}
[\sigma^2_\eta|\cdot] &\propto \prod_{t=1}^T [\boldsymbol{y}_t|\boldsymbol{\beta}_t, \sigma^2_\eta, \phi, \sigma^2_\epsilon][\sigma^2_\eta] \\
&\propto (\prod_{t=1}^T |\boldsymbol{\Sigma}|^{-\frac{1}{2}}) e^{-\frac{1}{2}\sum_{t=1}^T(\boldsymbol{y}_t - \boldsymbol{H}_t\boldsymbol{X}\boldsymbol{\beta}_t)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_t - \boldsymbol{H}_t\boldsymbol{X}\boldsymbol{\beta}_t)} (\sigma^2_\eta)^{-\alpha_\eta+1} e^{-\frac{\beta_\eta}{\sigma^2_\eta}} \\
&\propto |\boldsymbol{\Sigma}|^{-\frac{T}{2}} e^{-\frac{1}{2}\sum_{t=1}^T(\boldsymbol{y}_t - \boldsymbol{H}_t\boldsymbol{X}\boldsymbol{\beta}_t)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_t - \boldsymbol{H}_t\boldsymbol{X}\boldsymbol{\beta}_t)} (\sigma^2_\eta)^{-\alpha_\eta+1} e^{-\frac{\beta_\eta}{\sigma^2_\eta}}
\end{aligned}
$$

which can be sampled using a Metropolis-Hastings step

4.5 Full Conditional for σ_ϵ^2

$$\begin{aligned}
[\sigma_\epsilon^2|\cdot] &\propto \prod_{t=1}^{T} [\boldsymbol{y}_t|\boldsymbol{\beta}_t, \sigma_\eta^2, \phi, \sigma_\epsilon^2][\sigma_\epsilon^2] \\
&\propto (\prod_{t=1}^{T} |\boldsymbol{\Sigma}|^{-\frac{1}{2}}) e^{-\frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{y}_t - \boldsymbol{H}_t\boldsymbol{X}\boldsymbol{\beta}_t)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_t - \boldsymbol{H}_t\boldsymbol{X}\boldsymbol{\beta}_t)} (\sigma_\epsilon^2)^{-\alpha_\epsilon+1} e^{-\frac{\beta_\epsilon}{\sigma_\epsilon^2}} \\
&\propto |\boldsymbol{\Sigma}|^{-\frac{T}{2}} e^{-\frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{y}_t - \boldsymbol{H}_t\boldsymbol{X}\boldsymbol{\beta}_t)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_t - \boldsymbol{H}_t\boldsymbol{X}\boldsymbol{\beta}_t)} (\sigma_\epsilon^2)^{-\alpha_\epsilon+1} e^{-\frac{\beta_\epsilon}{\sigma_\epsilon^2}}
\end{aligned}$$

which can be sampled using a Metropolis-Hastings step

4.6 Full Conditional for ϕ

$$\begin{aligned}
[\phi|\cdot] &\propto \prod_{t=1}^{T} [y_t|\beta_t, \sigma_\eta^2, \phi, \sigma_\epsilon^2][\phi] \\
&\propto \prod_{t=1}^{T} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} e^{-\frac{1}{2}(\boldsymbol{y}_t - \boldsymbol{H}_t\boldsymbol{X}\boldsymbol{\beta}_t)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_t - \boldsymbol{H}_t\boldsymbol{X}\boldsymbol{\beta}_t)} \phi^{-\alpha_\phi+1} e^{-\frac{\beta_\phi}{\phi}} \\
&\propto |\boldsymbol{\Sigma}|^{-\frac{T}{2}} e^{-\frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{y}_t - \boldsymbol{H}_t\boldsymbol{X}\boldsymbol{\beta}_t)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_t - \boldsymbol{H}_t\boldsymbol{X}\boldsymbol{\beta}_t)} \phi^{-\alpha_\phi+1} e^{-\frac{\beta_\phi}{\phi}}
\end{aligned}$$

which can be sampled using a Metropolis-Hastings step

5 Posterior Predictive Distribution

The posterior predictive distribution for $\boldsymbol{y}_t$ is sampled a each MCMC iteration k by

$$\boldsymbol{y}_t^{(k)} \sim N(\boldsymbol{H_t}\boldsymbol{X}\boldsymbol{\beta}_t^{(k)}, \boldsymbol{\Sigma}^{(k)})$$