

Spatial Predictive Process Model

John Tipton

April 14, 2014

1 Full Dimensional Model Statement

1.1 Data Model

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

1.2 Process Model

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

1.3 Parameter Model

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

where \mathbf{I}_β is the identity matrix of size $\tau \times \tau$ where τ is the number of parameters in $\boldsymbol{\beta}_t$, \mathbf{I}_t is the identity matrix of size $n_t \times n_t$ and n_t is the number of samples of \mathbf{y}_t at time t and \mathbf{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 \mathbf{s}_t at time t from the set of all locations \mathbf{S} . This motivates the use of a predictive process $\tilde{\boldsymbol{\eta}}$ to approximate $\boldsymbol{\eta}_t$ over a set of knots \mathbf{S}^* where $\tilde{\boldsymbol{\eta}} = \mathbf{c}_t(\mathbf{s}_t, \mathbf{s}^* | \sigma_\eta^2, \phi)^T \mathbf{C}^{*-1}(\mathbf{s}^*, \mathbf{s}^* | \sigma_\eta^2, \phi) \boldsymbol{\eta}^*$. The covariance between the desired locations $\mathbf{s}_t \in \mathbf{S}$ at time t and the set of knots $\mathbf{s}^* \in \mathbf{S}^*$ is $\mathbf{c}_t(\mathbf{s}_t, \mathbf{s}^* | \sigma_\eta^2, \phi)$. The covariance matrix over the set of knots is $\mathbf{C}^{*-1}(\mathbf{s}^*, \mathbf{s}^* | \sigma_\eta^2, \phi)$. The lower dimensional $\boldsymbol{\eta}^* \sim \text{MVN}(\mathbf{0}, \mathbf{C}^{*-1}(\mathbf{s}^*, \mathbf{s}^* | \sigma_\eta^2, \phi))$. Equivalently, $\tilde{\boldsymbol{\eta}} \sim \text{MVN}(\mathbf{0}, \mathbf{c}_t^T \mathbf{C}^{*-1} \mathbf{c}_t)$ 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 $\tilde{\boldsymbol{\epsilon}}_t = \text{MVN}(\mathbf{0}, \sigma_\eta^2 \mathbf{I}_{n_t} - \text{diag}(\mathbf{c}_t^T \mathbf{C}^{*-1} \mathbf{c}_t))$

2.1 Data Model

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

2.2 Process Model

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

2.3 Parameter Model

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

where \mathbf{I}_β is the identity matrix of size $\tau \times \tau$ where τ is the number of parameters in $\boldsymbol{\beta}_t$, \mathbf{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 \mathbf{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

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

and the determinant

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

3 Posterior

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

4 Full Conditionals

4.1 Full Conditional for β_t

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

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

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

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

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

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

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

4.3 Full Conditional for σ_β^2

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

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

4.4 Full Conditional for σ_η^2

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

which can be sampled using a Metropolis-Hastings step

4.5 Full Conditional for σ_ϵ^2

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

which can be sampled using a Metropolis-Hastings step

4.6 Full Conditional for ϕ

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

which can be sampled using a Metropolis-Hastings step

5 Posterior Predictive Distribution

The posterior predictive distribution for \mathbf{y}_t is sampled at each MCMC iteration k by

$$\mathbf{y}_t^{(k)} \sim N(\mathbf{H}_t \mathbf{X} \beta_t^{(k)}, \Sigma^{(k)})$$