Full Conditional Distributions

This document provides the full conditional distributions for the models introduced in the paper. Letting $\theta = (\mu, \beta, \tau, \omega)$, the posterior distribution is

$$[\theta, X|Y] \propto [Y|X, \theta] [X|\theta] [\theta] = [Y|X, \omega] [X|\beta, \tau] [\beta] [\omega] [\tau],$$

assuming independence of the four components within $\theta$. The distribution is not available in closed form and so we use a Markov chain Monte Carlo algorithm (hybrid Gibbs sampler with Metropolis steps within Gibbs) to sample from it.

Before providing the full conditional distributions, we first note that for a generic parameter $\psi$, the quantities $a^\psi$ and $b^\psi$ denote the hyperprior parameters for gamma prior distributions and $m^\psi$ and $V^\psi$ are the hyperprior parameters for normal prior distributions. These hyperparameter parameters were not introduced in the manuscript for the sake of brevity.

1 The local environment to biomarker (LEB) model

1. For each $j$ and $i$, we draw augmented data, $Y^M_{ji}^*$, as follows:

$$Y^M_{ji}^* \begin{cases} = Y^M_{ji}, & Z^M_{ji} = 0; \\ \sim \text{Truncated } N(X^M_{ji}, 1/\omega^M_{ji}) \text{ on } (-\infty, M^M_{ji}], & Z^M_{ji} = 1; \\ \sim N(X^M_{ji}, 1/\omega^M_{ji}), & Z^M_{ji} = 2. \end{cases}$$

2. For the process variable $X^M$, we sample each $X^M_j$, $j = 1 \ldots, N^M$, conditional on the other process variables $\{X^M_k : k \neq j\}$ and the other parameters in the model. We have that

$$[X^M_j|\{X^M_k : k \neq j\}, \mu^M, \beta^M, \tau^M, Y^M, \omega^M] \propto [Y^s_j|X^M_j, \omega^M] [X^M_j|\{X^M_k : k \neq j\}, \beta^M_j, \tau^M_j] \prod_{l \in \chi_j} [X^M_l|\{X^M_k : k \neq l\}, \beta^M_l, \tau^M_l],$$

1
where \( \chi_j \) denotes the columns of the process variable \( X^M \) that depend on \( X^M_j \) in the specification of the conditional mean \( \eta^M_{ji} \) (i.e., those \( l \) that belong to \( S^M_j \)). This conditional mean \( \eta^M_{ji} \) is defined as:

\[
\eta^M_{ji} = \mu^M_j + \sum_{k=1}^{N^M_j} \beta^M_{jk} X^M_{s^M_{jk}i},
\]

where \( s^M_j = (s^M_{j1}, \ldots, s^M_{jN^M_j}) \) is a vector of length \( N^M_j \) (the cardinality of \( S^M_j \)) containing the columns of the process variable matrix \( X^M \) that affect the mean of the process variable \( X^M_j \).

The expression (2) above is equal to

\[
\prod_{i=1}^{N^I} \left( [Y^M_{ji}|X^M_{ji}, \omega^M_j] [X^M_{ji}|\{X^M_{ki} : k \neq j\}, \beta^M_j, \tau^M_j] \prod_{l \in \chi_j} [X^M_{li}|\{X^M_{ki} : k \neq l\}, \beta^M_l, \tau^M_l] \right),
\]

from which we can see that each \( X^M_{ji} \), \( i = 1, \ldots, N^I \), can be sampled independently. Now

\[
[Y^M_{ji}|X^M_{ji}, \omega^M_j] \propto \exp \left( \frac{\omega^M_j}{2} (Y^M_{ji} - X^M_{ji})^2 \right),
\]

\[
[X^M_{ji}|\{X^M_{ki} : k \neq j\}, \beta^M_j, \tau^M_j] \propto \exp \left( \frac{\tau^M_j}{2} (X^M_{ji} - \eta^M_{ji})^2 \right),
\]

and for each \( l \in \chi_j \),

\[
[X^M_{li}|\{X^M_{ki} : k \neq l\}, \beta^M_l] \propto \exp \left( \frac{\tau^M_l}{2} (X^M_{li} - \delta^M_{li}(j) - \epsilon^M_{li}(j))^2 \right). \tag{4}
\]

In (4),

\[
\delta^M_{ji}(n) \equiv \mu^M_j + \sum_{\{k:s^M_{jk} \neq n\}} \beta^M_{jk} X^M_{s^M_{jk}i},
\]

is the conditional mean of \( X^M_{ji} \) minus the regression terms that depend on \( X^M_{ni} \), and

\[
\epsilon^M_{ji}(n) \equiv \sum_{\{k:s^M_{jk} = n\}} \beta^M_{jk} X^M_{s^M_{jk}i}, \tag{5}
\]

is the regression terms that depend on \( X^M_{ni} \). Always, \( \delta^M_{ji}(n) - \epsilon^M_{ji}(n) = \eta^M_{ji} \).

Rearranging the terms of \( X^M_{ji} \), the distribution of \( X^M_{ji} \) conditional on all other parameters is \( N(q/p, 1/p) \), where

\[
p = \omega^M_j + \tau^M_j + \sum_{l \in \chi_j} \tau^M_l \left( \sum_{\{k:s^M_{jk} = n\}} \beta^M_{jk} \right)^2,
\]
and
\[ q = \omega_j M Y_{ji}^{M*} + \tau_j M \eta_{ji} M + \sum_{l \in \chi_j} \tau_j M (X_{li} M - \delta_{li}(j)) \left( \sum_{\{k:s_{jk}^M = n\}} \beta_{jk}^M \right). \]

3. For the regression parameters \( \beta^M \), we sample each \( \beta_j^M \) conditional on the other parameters in the model. Letting
\[
A_j = \begin{bmatrix}
1 & \{X_{s_{jk1}}^M\} \\
1 & \{X_{s_{jk2}}^M\} \\
\vdots & \vdots \\
1 & \{X_{s_{jkN_j}}^M\}
\end{bmatrix},
\]
the full conditional distribution of \( \beta_j^M \), is
\[
N_{r_j} \left( R_j^{-1} q_j, R_j^{-1} \right),
\]
where \( R_j = \tau_j^M A_j^T A_j + \Sigma_j^{-1} \) and \( q_j = \tau_j^M A_j^T X_j^M + \Sigma_j^{-1} m_j \).

4. For each \( j \), the conditional distribution of the process precision \( \tau_j^M \), given the other parameters, is the gamma distribution,
\[
\text{Ga} \left( a_{\tau_j^M} + N^I / 2, b_{\tau_j^M} + \sum_{i=1}^{N_j^I} (X_{ji}^M - \eta_{ji}^M)^2 / 2 \right),
\]
where \( \eta_{ji}^M \) is defined in (3) above.

2 The global to local environment (GLE) model

In this section, we provide the full conditional distributions for the parameters in the GLE models. Linking global environmental-media variables to local environmental-media variables will change the full conditional distributions for some of the latent processes and the associated regression coefficients that were given in the previous section. Specifically, now the full conditional distributions of \( X_{m(W)}^M \) depends on \( X^W \), and \( X_{m(S)}^M \) depends on \( X^T \) and \( X^H \). The full conditional distributions for other media in the LEB model remain unchanged.
2.1 Global water model

1. With the addition of the global water model, we need to change some of the full conditional
distributions in the LEB model for water. For the Water medium we replace equation (1)
with:

\[
Y^M_{i,m(W)} \begin{cases} = Y^M_{i,m(W)}, & \text{if } Z^M_{i,m(W)} = 0; \\
\sim \text{Truncated } N(X^M_{i,m(W)}, 1/\omega^M_{m(W)}) \text{ on } (-\infty, M^M_{i,m(W)}], & \text{if } Z^M_{i,m(W)} = 1; \\
\sim N(X^M_{i,m(W)}, 1/\omega^M_{m(W)}), & \text{if } Z^M_{i,m(W)} = 2.
\end{cases}
\]

To sample from the conditional distribution of the water process, \(\{X^M_{i,m(W)} : i = 1, \ldots, N^I\}\),
we let \(\mu^M_{i,m(W)} = I(\zeta^W_{ij} = 1) \mu^W_j\). Then \(X^M_{i,m(W)}\) is \(N(\mu^M_{i,m(W)}, 1/p^M_{i,m(W)})\),
for each individual \(i\).

With the addition of the global water model, the full conditional distribution of the NHEXAS
Water model precision, \(\tau^M_m(W)\), is now

\[
\text{Ga} \left( \frac{N^I}{2} + \omega^m(W), \frac{1}{2} \sum_{i=1}^{N^I} \left( X^M_{i,m(W)} - \mu^M_{i,m(W)} \right)^2 + b^M_{m(W)} \right).
\]

2. The full conditional distributions of the remaining parameters in the global water model de-
PENDS on the augmented data \(Y^{W*}_{j,k}\), defined as follows:

\[
Y^{W*}_{j,k} \begin{cases} = Y^W_{j,k}, & \text{if } Z^W_{j,k} = 0; \\
\sim \text{Truncated } N(X^W_{j,k}, 1/\omega^W) \text{ on } (-\infty, M^W_{j,k}], & \text{if } Z^W_{j,k} = 1; \\
\sim N(X^W_{j,k}, 1/\omega^W), & \text{if } Z^W_{j,k} = 2.
\end{cases}
\]

The full conditional distribution of the PWS measurement-error precision, \(\omega^W\), is then

\[
\text{Ga} \left( \frac{\sum_{j=1}^{N^W} N^W_j}{2} + a^{\omega W}, \frac{1}{2} \sum_{j=1}^{N^W} \sum_{k=1}^{N^W_j} (Y^{W*}_{j,k} - X^W_{j,k})^2 + b^{\omega W} \right).
\]

3. For the global water model, the full conditional distribution of \(X^W_{j,k}\) for the \(k\)th observation
\((k = 1, \ldots, N^W_j)\) of \(j\)th PWS \((j = 1, \ldots, N^W)\) is \(N(q_{j,k}/p_{j,k}, 1/p_{j,k})\), where

\[
p_{j,k} = \omega^W + \tau^W \quad \text{and} \quad q_{j,k} = \omega^W Y^{W*}_{j,k} + \tau^W \mu^W_j.
\]
Next, define $X_j^W = \{X_{i,m(W)}^M: \zeta_{ij} = 1\}$ and $l_j$ to be the number of elements in $X_j^W$. Then, for each PWS $j = 1, \ldots, N^W$, the full conditional distribution of the PWS-specific mean $\mu_j^W$ is $N(q_j/p_j, 1/p_j)$, with

$$p_j = N_j^W \tau^W + C^W + l_j M_{M(W)}$$

and

$$q_j = \tau^W \sum_{k=1}^{N^W} X_{jk}^W + C^W \alpha^W + \tau_{M(W)}^M 1_{l_j} X_j^W.$$ 

The full conditional distribution of the process precision, $\tau^W$, is

$$\text{Ga} \left( \frac{\sum_{j=1}^{N^W} N_j^W}{2} + a \tau^W, \frac{1}{2} \sum_{j=1}^{N^W} \sum_{k=1}^{N^W} (X_{jk}^W - \mu_j^W)^2 + b \tau^W \right).$$

4. For the parameters in the global water-LEB linking model, recall that $\zeta_{ij}^W = 1$ implies that individual $i$ is served by PWS $j$ and $\zeta_{ij}^W = 0$ implies that individual $i$ is not served by water-system $j$. For each individual $i = 1, \ldots, N^I$, the full conditional distribution of $\zeta_i^W = (\zeta_{i1}^W, \ldots, \zeta_{iN^W}^W)$ is Multinomial$(1, (\lambda^*_{c(i),1}, \ldots, \lambda^*_{c(i),N^W}))$ where the probability that individual $i$ in country $c(i)$ is served by PWS $j$, is

$$\lambda^*_{c(i),j} = \frac{\lambda_{c(i),j} n \left(X_{i,m(W)}^M; \mu_j^W, \tau_{M(W)}^M\right)}{\sum_{r=1}^{N^W} \lambda_{c(i),r} n \left(X_{i,m(W)}^M; \mu_r^W, \tau_{M(W)}^M\right)}.$$ 

Here, $n \left(X_{i,m(W)}^M; \mu_j^W, \tau_{M(W)}^M\right)$ is a normal pdf (with mean $\mu_j^W$ and variance $1/\tau_{M(W)}^M$) evaluated at the value of the water process, $X_{i,m(W)}^M$, for each individual $i$.

5. In the parameters in the global water prior distributions, $\alpha^W$ conditional on all the other parameters is $N(q/p, p)$, where

$$p = N^W C^W + V_{\alpha^W} \text{ and } q = \left( C^W \sum_{j=1}^{N^W} \mu_j^W + V_{\alpha^W} m_{\alpha^W} \right).$$ 

The full conditional distribution of $C^W$ is

$$\text{Ga} \left( \frac{N^W}{2} + a C^W, \frac{1}{2} \sum_{j=1}^{N^W} \left( \mu_j^W - m_{\alpha^W} \right)^2 + b C^W \right).$$
2.2 Global soil model

1. With the addition of the global topsoil/stream-sediment model, we need to change some of the conditional distributions in the LEB model for soil. For the soil medium, we replace equation (1) with:

\[
Y_{i,m(s)}^{M} = \begin{cases} 
Y_{i,m(s)}^{M}, & \text{if } Z_{i,m(s)}^{M} = 0; \\
\text{Truncated } N(X_{i,m(s)}^{M}, 1/\omega_{m(s)}^{M}) \text{ on } (-\infty, M_{i,m(s)}^{M}]), & \text{if } Z_{i,m(s)}^{M} = 1; \\
N(X_{i,m(s)}^{M}, 1/\omega_{m(s)}^{M}), & \text{if } Z_{i,m(s)}^{M} = 2.
\end{cases}
\]

2. For parameters appearing in the global topsoil/stream-sediment model, for the topsoil observation at the \(s\)th location:

\[
Y_{T}^{*}(s) = \begin{cases} 
Y_{T}(s), & \text{if } Z_{T}(s) = 0; \\
N(X_{T}(s), 1/\omega_{T}), & \text{if } Z_{T}(s) = 2
\end{cases}
\]

(there is no censored topsoil data).

The full conditional distribution of the topsoil measurement-error precision, \(\omega_{T}\), is

\[
\text{Ga} \left( \frac{N_{T}}{2} + a_{\omega_{T}}, \frac{1}{2} \sum_{i=1}^{N_{T}} (Y_{T}^{*}(s_{i}) - X_{T}(s_{i}))^{2} + b_{\omega_{T}} \right).
\]

3. For parameters appearing in the stream-sediment data model, for the \(k\)th observation in \(j\)th HUC8:

\[
Y_{jk}^{H} = \begin{cases} 
Y_{jk}^{H}, & \text{if } Z_{jk}^{H} = 0; \\
\text{Truncated } N(X_{j}^{H}, 1/\omega^{H}) \text{ on } (-\infty, M_{jk}^{H}), & \text{if } Z_{jk}^{H} = 1; \\
N(X_{j}^{H}, 1/\omega^{H}), & \text{if } Z_{jk}^{H} = 2.
\end{cases}
\]

The full conditional distribution of the stream-sediment measurement-error precision, \(\omega^{H}\), is

\[
\text{Ga} \left( \sum_{j=1}^{N_{H}} \frac{N_{j}^{H}}{2} + a_{\omega^{H}}, \frac{1}{2} \sum_{j=1}^{N_{H}} \sum_{k=1}^{N_{j}^{H}} (Y_{jk}^{H} - X_{j}^{H})^{2} + b_{\omega^{H}} \right).
\]

4. The full conditional distribution of the topsoil process \(X_{T}(s)\) is \(N(q_{s}/p_{s}, 1/p_{s})\), where

\[
p_{s} = \omega_{T} + \tau_{T} \quad \text{and} \quad q_{s} = \tau_{T}(\beta_{0}^{T} + \beta_{1}^{T} X_{h(s)}^{H}) + \omega_{T} Y_{T}^{*}(s),
\]

for \(s \in D\).
The full conditional distribution of the topsoil model precision, $\tau^T$, is
\[
\text{Ga} \left( \frac{|D|}{2} + a\tau^T, \frac{1}{2} \sum_{s \in D} (X^T(s) - \beta_0^T - \beta_1^T X_{h(s)})^2 + b\tau^T \right).
\]

The parameters that appear in the mean function of the topsoil model are $\beta^T = (\beta_0^T, \beta_1^T)'$.

Letting $U = [1_{|D|}, G_D X^H]$ and $W = [1_{N^H}, \zeta^T X^H]$ we sample $\beta^T$ from $N_2(P^{-1}Q, P^{-1})$, where
\[
P = \tau^T U'U + \tau^T_{m(S)} W'W + V^T M^T \beta^T
\]
and
\[
Q = \tau^T U'X^T + \tau^T_{m(S)} W'M^T + V^T M^T \beta^T.
\]

5. The full conditional distribution of the mean of the stream-sediment model, $\mu^H$, is $N(q/p, 1/p)$, where
\[
p = \tau^H 1_{N^H}(I_{N^H} - \gamma^H A)1_{N^H} + V^H \mu^H,
\]
and
\[
q = \tau^H 1_{N^H}(I_{N^H} - \gamma^H A)X^H + V^H m^H \mu^H.
\]

6. Recall that for the unknown parameter $\zeta^T_{ij}$, $\zeta^T_{ij} = 1$ if NHEXAS individual $i$ is located in HUC8 region $j$, and 0 otherwise. These $\zeta^T_{ij}$'s are updated during our MCMC algorithm using the following scheme:

Sample $\zeta^T_i = (\zeta^T_{i1}, \cdots, \zeta^T_{iN^H}) \sim \text{Multinomial} \left(1, (\lambda_{c(i),1}, \cdots, \lambda_{c(i),N^H})' \right)$, for $i = 1, \cdots, N^I$, where
\[
\lambda_{c(i),j} = \frac{\lambda_{c(i),j} n(X^M_{i,m(S)}; \beta_0^T + \beta_1^T X_j^H, \tau^T_{m(S)})}{\sum_{r=1}^{N^H} \lambda_{c(i),r} n(X^M_{i,m(S)}; \beta_0^T + \beta_1^T X_r^H, \tau^T_{m(S)})},
\]
and $n(X^M_{i,m(S)}; \beta_0^T + \beta_1^T X_j^H, \tau^T_{m(S)})$ is a normal pdf (with mean $\beta_0^T + \beta_1^T X_j^H$ and variance $1/\tau^T_{m(S)}$) evaluated at $X^M_{i,m(S)}$.

Now define the $|D| \times N^H$ matrix $G_D$ such that the $(i,j)$th entry $G_D(i,j)$ is 1 if topsoil location $s_i$ falls in $j$th HUC8, and 0 otherwise. Sample $X^H$ from $N(P^{-1}Q, P^{-1})$, where
\[
P = \tau^H (I_{N^H} - \gamma^H A) + \tau^T G_D^r G_D + \tau^T_{m(S)} \zeta^T \zeta^T,
\]
and
\[
Q = \tau^H (I_{N^H} - \gamma^H A)\mu^H 1_{N^H} + \tau^T G_D^r (X^T - \beta_0^T 1_{|D|}) + \tau^T_{m(S)} \zeta^T (X^M_{m(S)} - \beta_0^T 1_{N^I}).
\]
7. The full conditional distribution of the stream-sediment precision, $\tau^H$, is
\[
\text{Ga}
\left(\frac{NH}{2} + a^H, \frac{1}{2}(X^H - \mu^H 1_{NH}')(I_{NH} - \gamma^H A)(X^H - \mu^H 1_{NH}) + b^H \right).
\]
The full conditional distribution of the mean of the stream-sediment model, $\mu^H$, is $N(q/p, 1/p)$, where
\[
p = \tau^H 1_{NH}'(I_{NH} - \gamma^H A)\mu^H + V^H,
\]
and
\[
q = \tau^H 1_{NH}'(I_{NH} - \gamma^H A)X^H + V^H \mu^H.
\]
The full conditional distribution of the spatial-dependence parameter in the stream-sediment model, $\gamma^H$, is proportional to
\[
\frac{1}{|\lambda_{\max} - \lambda_{\min}|} I(\gamma^H \in (\lambda_{\min}, \lambda_{\max})) \times
\end{equation}
where $\lambda_{\min}, \lambda_{\max}$ are the smallest and largest eigenvalues of $A$.

8. With the addition of the global-topsoil/stream-sediment model, the full conditional distribution of the LEB Soil log As process $X^M_{m(S)}$, is modified. Now, this distribution is $N_{N^I}(P^{-1}Q, P^{-1})$, where
\[
P = \omega^M_{m(S)} I_{N^I} + \tau^M_{m(S)} I_{N^I},
\]
and
\[
Q = \omega^M_{m(S)} Y^M_{m(S)} + \tau^M_{m(S)}(\beta^T_0 1_{N^I} + \beta^T_1 \zeta^T X^H).
\]
In addition, the full conditional distribution of the local soil process precision, $\tau^M_{m(S)}$, is now
\[
\text{Ga}
\left(\frac{N^I}{2} + d^M_{m(W)}, \frac{1}{2}(X^M_{m(S)} - W\beta^T)'(X^M_{m(S)} - W\beta^T) + b^M_{m(S)} \right).
\]