Generalised Pareto Distribution with Predictive Processes and Metropolis-Hastings MCMC updates

John O'Sullivan

November 3, 2019

Introduction

This document includes further details on the model in the code; namely, it describes in detail how we modelled a dataset of extreme observations using a generalised Pareto distribution with predictive processes and Markov chain Monte Carlo updates. Details for the Binomial component of the model are similar, and are omitted here for brevity.

The following pages include:

- the pseudocode in order to programme the algorithm;
- the model directed acyclic graph (DAG);
- details of the model notation;
- and the details of all equations needed for the MCMC updates of the parameters and the hyperparameters.

Algorithm details

Data: y_{ij}, the declustered threshold excesses at locations $i = 1 \ldots n$, with $j = 1 \ldots n_i$ excesses at location i;

X_ϕ, X_ξ, $n \times (p+1)$ and $n \times (q+1)$ matrices of p and q covariates at locations $i = 1 \ldots n$.

Result: Samples from the posterior distributions of $\phi = \log(\sigma)$ and ξ (the unknown parameters of interest), which can then be used to calculate return level estimates and other desired quantities

Initialisation;

Random starting values of ϕ^* and ξ^* ;

Projection of ϕ and ξ from ϕ^* and ξ^* respectively;

Hyper-parameter values of α_ϕ, β_ϕ, ς^2_ϕ, τ^2_ϕ, ν_ϕ, α_ξ, β_ξ, ς^2_ξ, τ^2_ξ, and ν_ξ;

Number of iterations N;

for *iterations i from 1 to N* **do**

  Generate $u \sim U(0,1)$;

  for *sub-grid locations k from 1 to m* **do**

    Simulate $\phi^*_{new,k}$;

    Project ϕ_{new} from ϕ^*_{new};

    Set l_{new} = log full conditional of new vector ϕ^*_{new};

    Set l_{old} = log full conditional of old vector ϕ^*;

    Set $a = \exp(l_{new} - l_{old})$, that is, evaluate equation (1);

    if $a > u$ **then**

      Set $\phi^* = \phi^*_{new}$;

    end

    Simulate $\xi^*_{new,k}$;

    Project ξ_{new} from ξ^*_{new};

    Set l_{new} = log full conditional of new vector ξ^*_{new};

    Set l_{old} = log full conditional of old vector ξ^*;

    Set $a = \exp(l_{new} - l_{old})$, that is, evaluate equation (2);

    if $a > u$ **then**

      Set $\xi^* = \xi^*_{new}$;

    end

  end

end

---

for *iterations i from 1 to N (continued)* **do**

  for *each element c of the vector* α_ϕ **do**

    Simulate $\alpha_{\phi_{new,c}}$;

    Set $a =$ result of equation (3);

    If $a > u$, set $\alpha_\phi = \alpha_{\phi_{new}}$;

  end

  for *each element d of the vector* α_ξ **do**

    As above, but set $a =$ result of equation (4);

  end

  Simulate $\beta_{\phi_{new}}$;

  Set $a =$ result of equation (5);

  If $a > u$, set $\beta_\phi = \beta_{\phi_{new}}$;

  Repeat for β_ξ with equation (6);

  Simulate $\nu_{\phi_{new}}$;

  Set $a =$ result of equation (7);

  If $a > u$, set $\nu_\phi = \nu_{\phi_{new}}$;

  Repeat for ν_ξ with equation (8) ;

  Simulate $\varsigma^2_{\phi_{new}}$;

  Set $a =$ result of equation (9);

  If $a > u$, set $\varsigma^2_\phi = \varsigma^2_{\phi_{new}}$;

  Repeat for ς^2_ξ with equation (10) ;

  Simulate $\tau^2_{\phi_{new}}$;

  Set $a =$ result of equation (11);

  If $a > u$, set $\tau^2_\phi = \tau^2_{\phi_{new}}$;

  Repeat for τ^2_ξ with equation (12) ;

end

DAG

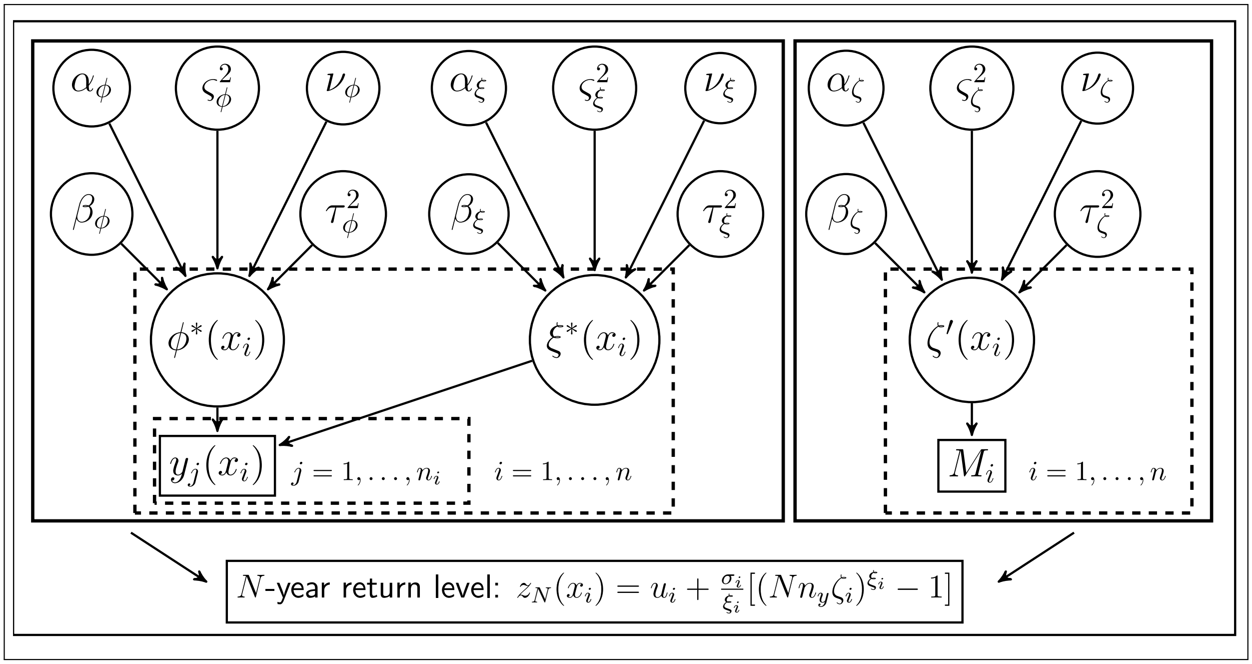

Figure 1: A directed acyclic graph (DAG) of the Bayesian hierarchical model fitted to the spatial dataset of daily maximum temperature anomalies. On the left is the model for the excesses (using the GPD) and on the right is the model for the probability of an observation being a cluster maximum (using the Binomial distribution). The parameters of the distributions are modelled on a subgrid using Gaussian predictive processes as detailed in Section **??**. These are represented as circles in the middle layer, with the hyperparameters controlling these represented in the top layer. The data is represented in the bottom layer (in rectangles). Arrows run into nodes from their direct predecessors (often called parents). Given its parents, each node is independent of all other nodes in the graph except its descendants (often called children). Posterior estimates of the parameters' distributions can be used to form quantities of interest - typically return levels, as illustrated. Though omitted for clarity of the diagram, it should be remembered that the three parameters (ϕ^*, ξ^*, and ζ') are defined on a subgrid of the domain, but give rise to the predictive processes (ϕ, ξ, and $\tilde{\zeta}$, respectively) defined on the full grid of the domain, using equation (**??**). Further details of each layer and the parameters involved may be found in the text and in this appendix.

Notation

- x_i Multivariate (bivariate or tri-variate) location values for location i, $i = 1, \ldots, n$. Write the matrix of all locations as just x
- y_{ij} Excess j for observation i, $j = 1, \ldots, n_i$ where n_i is the number of excesses at location i
- $\sigma(x_i)$ Scale parameter for location x_i
- $\phi(x_i) = \log(\sigma(x_i))$ The re-parameterised scale parameter
- $\xi(x_i)$ Shape parameter for location x_i
- z_k Sub-grid locations $k = 1, \ldots, m$. Together written as z
- A, B Projection matrices of dimension $n \times m$
- ϕ^*, ξ^* Gaussian processes of ϕ and ξ defined on sub-grid z
- $\mu_\phi(z), \mu_\xi(z)$ Means for the Gaussian processes
- Σ, Ψ Covariance matrices for the Gaussian processes
- τ_ϕ^2, τ_ξ^2 Nugget parameters for the Gaussian processes
- α_ϕ, α_ξ vectors of coefficients (including intercept) for Gaussian process means
- $X_\phi, X_\xi, Z_\sigma, Z_\xi$ Matrices of covariates (including column for intercept term) on the x grid and the z sub-grid
- ν_ϕ, ν_ξ Matern smoothness parameters for the Gaussian processes
- β_ϕ, β_ξ Matern length scale scalars or matrices for the Gaussian processes
- $\varsigma_\phi^2, \varsigma_\xi^2$ Variance parameters (partial sills) for the Gaussian processes

Model outline

In hierarchical notation:

$$
\begin{aligned}
y_{ij} &\sim GPD(\sigma(x_i), \xi(x_i)) \\
\log(\sigma(x)) = \phi(x) &= A(x,z)\Sigma^{-1}(z,z)\phi^*(z) \\
\xi(x) &= B(x,z)\Psi^{-1}(z,z)\xi^*(z) \\
A(x_i, z_k) &= \varsigma_\phi^2 \frac{2^{1-\nu_\phi}}{\Gamma(\nu_\phi)} \left(\sqrt{2\nu_\phi} \frac{\|x_i - z_k\|}{\beta_\phi} \right)^{\nu_\phi} K_{\nu_\phi} \left(\sqrt{2\nu_\phi} \frac{\|x_i - z_k\|}{\beta_\phi} \right) \\
B(x_i, z_k) &= \varsigma_\xi^2 \frac{2^{1-\nu_\xi}}{\Gamma(\nu_\xi)} \left(\sqrt{2\nu_\xi} \frac{\|x_i - z_k\|}{\beta_\xi} \right)^{\nu_\xi} K_{\nu_\xi} \left(\sqrt{2\nu_\xi} \frac{\|x_i - z_k\|}{\beta_\xi} \right) \\
\phi^* &\sim MVN_m(\mu_\phi(z), \tau_\phi^2 I_m + \Sigma(z,z)) \\
\xi^* &\sim MVN_m(\mu_\xi(z), \tau_\xi^2 I_m + \Psi(z,z)) \\
\mu_\phi(z) &= Z_\phi \alpha_\phi \\
\mu_\xi(z) &= Z_\xi \alpha_\xi \\
\mu_\phi(x) &= X_\phi \alpha_\phi \\
\mu_\xi(x) &= X_\xi \alpha_\xi \\
\Sigma(z_k, z_l) &= \varsigma_\phi^2 \frac{2^{1-\nu_\phi}}{\Gamma(\nu_\phi)} \left(\sqrt{2\nu_\phi} \frac{\|z_k - z_l\|}{\beta_\phi} \right)^{\nu_\phi} K_{\nu_\phi} \left(\sqrt{2\nu_\phi} \frac{\|z_k - z_l\|}{\beta_\phi} \right) \\
\Psi(z_k, z_l) &= \varsigma_\xi^2 \frac{2^{1-\nu_\xi}}{\Gamma(\nu_\xi)} \left(\sqrt{2\nu_\xi} \frac{\|z_k - z_l\|}{\beta_\xi} \right)^{\nu_\xi} K_{\nu_\xi} \left(\sqrt{2\nu_\xi} \frac{\|z_k - z_l\|}{\beta_\xi} \right)
\end{aligned}
$$

In the above formulae for $A, B, \Sigma, \text{and} \Psi$, β_ϕ and β_ξ are treated as scalars. When considered as matrices, where d is the distance between two gridpoints, $\frac{d}{\beta}$ is replaced by $\sqrt{d^T \beta^{-1} d}$.

Hyperparameter prior distributions (justifications for these values can be found in Section **??**):

$$
\begin{aligned}
\log(\varsigma_\phi^2), \log(\varsigma_\xi^2) &\sim N(0,1)\\
\log(\tau_\phi^2), \log(\tau_\xi^2) &\sim N(-2.3,1)\\
\beta_\phi, \beta_\xi &\sim DU(0, 0.05, 1, 10)\\
\nu_\phi, \nu_\xi &\sim DU(0.5, 2.5)\\
\alpha_\phi &\sim MVN(\vec{\eta}_\phi, H_\phi)\\
\alpha_\xi &\sim MVN(\vec{\eta}_\xi, H_\xi)
\end{aligned}
$$

$\vec{\eta}_\phi$ and $\vec{\eta}_\xi$ are vectors of the appropriate length with all entries equal to 0. H_ϕ and H_ξ are the relevant covariance matrices - to begin, these are diagonal matrices with 2 on the diagonal. The code is designed to allow the relationship between the coefficients to be modelled (if desired) by adjusting this covariance matrix.

Posterior distribution

The full posterior distribution is:

$$
\begin{aligned}
p(\alpha_\phi, \alpha_\xi, \beta_\phi, \beta_\xi, \varsigma_\phi^2, \varsigma_\xi^2, \tau_\phi^2, \tau_\xi^2, \nu_\phi, \nu_\xi, \phi^*, \xi^* | y, x, z, X_\phi, X_\xi, Z_\phi, Z_\xi) \propto&\\
\left[\prod_{i=1}^{n}\prod_{j=1}^{n_i} p(y_{ij}|\sigma(x_i) = \exp(\phi(x_i)), \xi(x_i))\right] \times&\\
p(\phi^*(x)|\mu_\phi, \tau_\phi^2, \Sigma)p(\xi^*(x)|\mu_\xi, \tau_\xi^2, \Psi)\times&\\
p(\alpha_\phi)p(\alpha_\xi)p(\beta_\phi)p(\beta_\xi)p(\nu_\phi)p(\nu_\xi)\times&\\
p(\tau_\phi^2)p(\tau_\xi^2)p(\varsigma_\phi^2)p(\varsigma_\xi^2)&
\end{aligned}
$$

Conditional posterior distributions: Layer 2

Updating the second layer of the DAG (parameters $\phi^* = \log(\sigma^*)$ and ξ^*): The conditional posterior distribution of ϕ^* is given by:

$$
\begin{aligned}
&\pi(\phi^*|y,x,z,X_\phi,X_\xi,Z_\phi,Z_\xi,\alpha_\phi,\alpha_\xi,\beta_\phi,\beta_\xi,\varsigma_\phi^2,\varsigma_\xi^2,\tau_\phi^2,\tau_\xi^2,\nu_\phi,\nu_\xi,\xi^*) \propto \\
&p(y|\phi^*,x,z,X_\phi,X_\xi,Z_\phi,Z_\xi,\alpha_\phi,\alpha_\xi,\beta_\phi,\beta_\xi,\varsigma_\phi^2,\varsigma_\xi^2,\tau_\phi^2,\tau_\xi^2,\nu_\phi,\nu_\xi,\xi^*) \times \\
&p(\phi^*|x,z,X_\phi,X_\xi,Z_\phi,Z_\xi,\alpha_\phi,\alpha_\xi,\beta_\phi,\beta_\xi,\varsigma_\phi^2,\varsigma_\xi^2,\tau_\phi^2,\tau_\xi^2,\nu_\phi,\nu_\xi,\xi^*) \propto \\
&p(y|\sigma,\xi)p(\phi^*|\mu_\phi,\tau_\phi^2,\Sigma)
\end{aligned}
$$

That is, the conditional posterior of ϕ^* is proportional to the product of the likelihood of the data y given ϕ^* (and its associated projection on the full grid, ϕ) and all other parameters, and the prior probability density of ϕ^* given all of the other parameters. In the final line, most parameters have dropped out from the right-hand side, as the densities are independent of these, given the remaining terms (see the DAG in Figure 1 for this). Remember that $\phi = \log(\sigma)$ throughout this. Of the remaining terms, ξ, μ_ϕ and Σ are deterministic given the other parameters. These will have been calculated using the formulae in the model outline (above).

In briefer notation (to be used from now on):

$$
\begin{aligned}
\pi(\phi^*|\dots) &\propto p(y|\sigma,\xi)p(\phi^*|\mu_\phi,\tau_\phi^2,\Sigma) \\
&= \left[\prod_{i=1}^{n}\prod_{j=1}^{n_i} p(y_{ij}|\sigma(x_i),\xi(x_i))\right] p(\phi^*|\mu_\phi,\tau_\phi^2,\Sigma) \\
&= \left[\prod_{i=1}^{n}\prod_{j=1}^{n_i} \frac{1}{\sigma(x_i)}\left(1+\xi(x_i)\frac{y_{ij}}{\sigma(x_i)}\right)^{-(1/\xi(x_i)+1)}\right] \times \\
&\quad \frac{1}{\sqrt{\det(2\pi(\Sigma+\tau_\phi^2 I_m))}} e^{-\frac{1}{2}(\phi^*-\mu_\phi)'(\Sigma+\tau_\phi^2 I_m)^{-1}(\phi^*-\mu_\phi)}
\end{aligned}
$$

Similarly:

$$
\begin{aligned}
\pi(\xi^*|\dots) \propto & p(y|\sigma,\xi)p(\xi^*|\mu_\xi,\tau_\xi^2,\Psi) \\
= & \left[\prod_{i=1}^{n}\prod_{j=1}^{n_i} p(y_{ij}|\sigma(x_i),\xi(x_i))\right] p(\xi^*|\mu_\xi,\tau_\xi^2,\Psi) \\
= & \left[\prod_{i=1}^{n}\prod_{j=1}^{n_i} \frac{1}{\sigma(x_i)}\left(1+\xi(x_i)\frac{y_{ij}}{\sigma(x_i)}\right)^{-(1/\xi(x_i)+1)}\right] \times \\
& \frac{1}{\sqrt{\det(2\pi(\Psi+\tau_\xi^2 I_m))}} e^{-\frac{1}{2}(\xi^*-\mu_\xi)'(\Psi+\tau_\xi^2 I_m)^{-1}(\xi^*-\mu_\xi)}
\end{aligned}
$$

Metropolis-Hastings Markov chain Monte Carlo sampling

In order to sample from these conditional posterior distributions, we use the Metropolis-Hastings Markov chain Monte Carlo (MCMC) algorithm. New samples are accepted or rejected at random according to the algorithm outlined below.

A new value of ϕ_k^* is suggested: $\phi_k^{*'}$, where k is a location on the subgrid. The new vector ϕ' is then calculated on the full grid using the projection formula.

Though $\phi^{*'}$ differs from ϕ^* in just one location k, ϕ and ϕ' can be different from each other in many locations i due to this projection.

Suggested updates are drawn from a Normal distribution centred on the old value and with a variance of a manually set tuning parameter used to control the size of the proposed steps.

We calculate:

$$
\rho(\phi_k^*,\phi_k^{*'}) = \min\left(1, \frac{\pi(\phi^{*'}|\dots)q_t(\phi_k^{*'}\to\phi_k^*)}{\pi(\phi^*|\dots)q_t(\phi_k^*\to\phi_k^{*'})}\right)
$$

where $\pi(\phi^*|\dots)$ is as defined above, and $q_t(a\to b)$ is the transition probability of proposing value b given value a. Since updates are proposed using a Normal distribution, these transition probabilites above and below the line will always cancel, so the above simplifies to:

$$\rho(\phi_k^*, \phi_k^{*'}) = \min\left(1, \frac{\pi(\phi^{*'}|\dots)}{\pi(\phi^*|\dots)}\right)$$

Following this calculation, we always accept proposed value $\phi_k^{*'}$ when $\rho(\phi_k^*, \phi_k^{*'})$ equals 1 and we reject accordingly when the ratio is smaller than 1 by simulating a random variable $u \sim U[0,1]$ and accepting proposed value $\phi_k^{*'}$ when $u \leq \rho(\phi_k^*, \phi_k^{*'})$.

Evaluating $\rho(\phi_k^*, \phi_k^{*'})$ typically involves products and quotients of many terms which may be close to 0. In order to work with something far more computationally stable, we use the property that $x = \exp(\log(x))$.

Following this observation we need to evaluate:

$$\begin{aligned}
&\exp\left(\log\left(\frac{\pi(\phi^{*'}|\dots)}{\pi(\phi^*|\dots)}\right)\right) \\
&= \exp\left[\log(\pi(\phi^{*'}|\dots)) - \log(\pi(\phi^*|\dots))\right] \\
&= \exp\Big[\log\Big(\Big[\prod_{i=1}^{n}\prod_{j=1}^{n_i} p(y_{ij}|\sigma'(x_i), \xi(x_i))\Big] p(\phi^{*'}|\mu_\phi, \tau_\phi^2, \Sigma)\Big) - \\
&\qquad \log\Big(\Big[\prod_{i=1}^{n}\prod_{j=1}^{n_i} p(y_{ij}|\sigma(x_i), \xi(x_i))\Big] p(\phi^*|\mu_\phi, \tau_\phi^2, \Sigma)\Big)\Big]
\end{aligned}$$

Filling in the distribution details, this becomes:

$$\exp\Bigg[\log\Big(\Big[\prod_{i=1}^{n}\prod_{j=1}^{n_i}\frac{1}{\sigma'(x_i)}\Big(1+\xi(x_i)\frac{y_{ij}}{\sigma'(x_i)}\Big)^{-(1/\xi(x_i)+1)}\Big]\times$$

$$\frac{1}{\sqrt{\det(2\pi(\Sigma+\tau_\phi^2 I_m))}}e^{-\frac{1}{2}(\phi^{*'}-\mu_\phi)'(\Sigma+\tau_\phi^2 I_m)^{-1}(\phi^{*'}-\mu_\phi)}\Big)$$

$$-\log\Big(\Big[\prod_{i=1}^{n}\prod_{j=1}^{n_i}\frac{1}{\sigma(x_i)}\Big(1+\xi(x_i)\frac{y_{ij}}{\sigma(x_i)}\Big)^{-(1/\xi(x_i)+1)}\Big]\times$$

$$\frac{1}{\sqrt{\det(2\pi(\Sigma+\tau_\phi^2 I_m))}}e^{-\frac{1}{2}(\phi^{*}-\mu_\phi)'(\Sigma+\tau_\phi^2 I_m)^{-1}(\phi^{*}-\mu_\phi)}\Big)\Bigg]$$

$$=\exp\Bigg[\sum_{i=1}^{n}\sum_{j=1}^{n_i}\log\Big(\Big[\frac{1}{\sigma'(x_i)}\Big(1+\xi(x_i)\frac{y_{ij}}{\sigma'(x_i)}\Big)^{-(1/\xi(x_i)+1)}\Big]\Big)+$$

$$\log\left(\frac{1}{\sqrt{\det(2\pi(\Sigma+\tau_\phi^2 I_m))}}\right)-\frac{1}{2}(\phi^{*'}-\mu_\phi)'(\Sigma+\tau_\phi^2 I_m)^{-1}(\phi^{*'}-\mu_\phi)$$

$$-\sum_{i=1}^{n}\sum_{j=1}^{n_i}\log\Big(\Big[\frac{1}{\sigma(x_i)}\Big(1+\xi(x_i)\frac{y_{ij}}{\sigma(x_i)}\Big)^{-(1/\xi(x_i)+1)}\Big]\Big)-$$

$$\log\left(\frac{1}{\sqrt{\det(2\pi(\Sigma+\tau_\phi^2 I_m))}}\right)+\frac{1}{2}(\phi^{*}-\mu_\phi)'(\Sigma+\tau_\phi^2 I_m)^{-1}(\phi^{*}-\mu_\phi)\Bigg]$$

$$=\exp\Bigg[\sum_{i=1}^{n}\sum_{j=1}^{n_i}\log\Big(\Big[\frac{1}{\sigma'(x_i)}\Big(1+\xi(x_i)\frac{y_{ij}}{\sigma'(x_i)}\Big)^{-(1/\xi(x_i)+1)}\Big]\Big)-$$

$$\frac{1}{2}(\phi^{*'}-\mu_\phi)'(\Sigma+\tau_\phi^2 I_m)^{-1}(\phi^{*'}-\mu_\phi)-$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n_i}\log\Big(\Big[\frac{1}{\sigma(x_i)}\Big(1+\xi(x_i)\frac{y_{ij}}{\sigma(x_i)}\Big)^{-(1/\xi(x_i)+1)}\Big]\Big)+$$

$$\frac{1}{2}(\phi^{*}-\mu_\phi)'(\Sigma+\tau_\phi^2 I_m)^{-1}(\phi^{*}-\mu_\phi)\Bigg]$$

The common term $\log(1/\sqrt{\det(2\pi(\Sigma+\tau_\phi^2 I_m))})$ in both numerator and denominator above dropped out as it will be equal in both (it has no dependence on ϕ^*).

The GPD can clearly be simplified further using the properties of logs. Just taking one of the functions on its own for clarity:

$$
\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{n_i}\log\left(\left[\frac{1}{\sigma'(x_i)}\left(1+\xi(x_i)\frac{y_{ij}}{\sigma'(x_i)}\right)^{-(1/\xi(x_i)+1)}\right]\right) &= \\
\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left[\log\left(\frac{1}{\sigma'(x_i)}\right)+\log\left(1+\xi(x_i)\frac{y_{ij}}{\sigma'(x_i)}\right)^{-(1/\xi(x_i)+1)}\right] &= \\
\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left[-\log(\sigma'(x_i))-\left(\frac{1}{\xi(x_i)}+1\right)\log\left(1+\xi(x_i)\frac{y_{ij}}{\sigma'(x_i)}\right)\right]
\end{aligned}
$$

Using this form, the full term we need to evaluate in order to update ϕ^* is:

$$
\begin{aligned}
= \exp\Bigg[\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left[-\log(\sigma'(x_i))-\left(\frac{1}{\xi(x_i)}+1\right)\log\left(1+\xi(x_i)\frac{y_{ij}}{\sigma'(x_i)}\right)\right]- \\
\frac{1}{2}(\phi^{*'}-\mu_\phi)'(\Sigma+\tau_\phi^2 I_m)^{-1}(\phi^{*'}-\mu_\phi)- \\
\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left[-\log(\sigma(x_i))-\left(\frac{1}{\xi(x_i)}+1\right)\log\left(1+\xi(x_i)\frac{y_{ij}}{\sigma(x_i)}\right)\right]+ \\
\frac{1}{2}(\phi^{*}-\mu_\phi)'(\Sigma+\tau_\phi^2 I_m)^{-1}(\phi^{*}-\mu_\phi)\Bigg]
\end{aligned}
\tag{1}
$$

Similar reasoning leads to the update for ξ^*. We need to evaluate:

$$
\rho(\xi_k^*,\xi_k^{*'}) = \min\left(1, \frac{\pi(\xi^{*'}|\dots)q_t(\xi_k^{*'}\to\xi_k^*)}{\pi(\xi^{*}|\dots)q_t(\xi_k^{*}\to\xi_k^{*'})}\right)
$$

The full term we need to evaluate is:

$$
\begin{aligned}
= \exp\Bigg[\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left[-\log(\sigma(x_i))-\left(\frac{1}{\xi'(x_i)}+1\right)\log\left(1+\xi'(x_i)\frac{y_{ij}}{\sigma(x_i)}\right)\right]- \\
\frac{1}{2}(\xi^{*'}-\mu_\xi)'(\Psi+\tau_\xi^2 I_m)^{-1}(\xi^{*'}-\mu_\xi)- \\
\sum_{i=1}^{n}\sum_{j=1}^{n_i}\left[-\log(\sigma(x_i))-\left(\frac{1}{\xi(x_i)}+1\right)\log\left(1+\xi(x_i)\frac{y_{ij}}{\sigma(x_i)}\right)\right]+ \\
\frac{1}{2}(\xi^{*}-\mu_\xi)'(\Psi+\tau_\xi^2 I_m)^{-1}(\xi^{*}-\mu_\xi)\Bigg]
\end{aligned}
\tag{2}
$$

Conditional posterior distributions: Layer 3

$\boldsymbol{\alpha}$

Updating the third layer of the DAG (that is, all hyperparameters) starting with α_ϕ:

$$\begin{aligned}\pi(\alpha_\phi|y,\dots) &\propto p(y|\alpha_\phi,\dots)p(\alpha_\phi|\dots)\\ &\propto p(y|\sigma,\xi)p(\phi^*|\mu_\phi,\tau_\phi^2,\Sigma)p(\alpha_\phi)\\ &\propto p(\phi^*|\mu_\phi,\tau_\phi^2,\Sigma)p(\alpha_\phi)\end{aligned}$$

Essentially the GPD component is independent of α_ϕ once the other parameters are known, and so can be absorbed into the constant of proportionality. What we're left with is the MVN piece (since α_ϕ features in the calculation of μ_ϕ) and the prior on α_ϕ.

Then we have:

$$\begin{aligned}\pi(\alpha_\phi|\dots) &\propto p(\phi^*|\mu_\phi,\tau_\phi^2,\Sigma)p(\alpha_\phi)\\ &\propto \frac{1}{\sqrt{\det(2\pi(\Sigma+\tau_\phi^2 I_m))}}e^{-\frac{1}{2}(\phi^*-\mu_\phi)'(\Sigma+\tau_\phi^2 I_m)^{-1}(\phi^*-\mu_\phi)}\times\\ &\quad\frac{1}{\sqrt{\det(2\pi H_\phi)}}e^{-\frac{1}{2}(\alpha_\phi-\eta_\phi)'H_\phi^{-1}(\alpha_\phi-\eta_\phi)}\end{aligned}$$

where H_ϕ is the covariance matrix for the prior distribution of α_ϕ and η_ϕ is the prior mean.
Similarly:

$$\begin{aligned}\pi(\alpha_\xi|\dots) &\propto p(\xi^*|\mu_\xi,\tau_\xi^2,\Psi)p(\alpha_\xi)\\ &\propto \frac{1}{\sqrt{\det(2\pi(\Psi+\tau_\xi^2 I_m))}}e^{-\frac{1}{2}(\xi^*-\mu_\xi)'(\Psi+\tau_\xi^2 I_m)^{-1}(\xi^*-\mu_\xi)}\times\\ &\quad\frac{1}{\sqrt{\det(2\pi H_\xi)}}e^{-\frac{1}{2}(\alpha_\xi-\eta_\xi)'H_\xi^{-1}(\alpha_\xi-\eta_\xi)}\end{aligned}$$

where H_ξ is the covariance matrix for the prior distribution of α_ξ and η_ξ is the prior mean.

As with the parameters ϕ^* and ξ^*, updates for α_ϕ are suggested element-wise: $\alpha_{\phi,\kappa} \to \alpha'_{\phi,\kappa}$, where the index κ runs over the vector of coefficients. The ratio $\rho(\alpha_{\phi,\kappa}, \alpha'_{\phi,\kappa})$ is then calculated. Remembering that μ_ϕ is calculated from α_ϕ (so that μ'_ϕ is the updated value, given α'_ϕ), then similar manipulations to those used in the previous section lead to the following calculation needed to update α_ϕ:

$$
\begin{aligned}
\exp\Big[&-\frac{1}{2}\log(\det(2\pi(\Sigma+\tau_\phi^2 I_m))) - \frac{1}{2}(\phi^* - \mu'_\phi)'(\Sigma+\tau_\phi^2 I_m)^{-1}(\phi^* - \mu'_\phi) \\
&\quad -\frac{1}{2}\log(\det(2\pi H_\phi)) - \frac{1}{2}(\alpha'_\phi - \eta_\phi)' H_\phi^{-1}(\alpha'_\phi - \eta_\phi)) \\
&\quad +\frac{1}{2}\log(\det(2\pi(\Sigma+\tau_\phi^2 I_m))) + \frac{1}{2}(\phi^* - \mu_\phi)'(\Sigma+\tau_\phi^2 I_m)^{-1}(\phi^* - \mu_\phi) \\
&\quad +\frac{1}{2}\log(\det(2\pi H_\phi)) + \frac{1}{2}(\alpha_\phi - \eta_\phi)' H_\phi^{-1}(\alpha_\phi - \eta_\phi))\Big] = \\
\exp\Big[&-\frac{1}{2}(\phi^* - \mu'_\phi)'(\Sigma+\tau_\phi^2 I_m)^{-1}(\phi^* - \mu'_\phi) - \frac{1}{2}(\alpha'_\phi - \eta_\phi)' H_\xi^{-1}(\alpha'_\phi - \eta_\phi)) \\
&\quad +\frac{1}{2}(\phi^* - \mu_\phi)'(\Sigma+\tau_\phi^2 I_m)^{-1}(\phi^* - \mu_\phi) + \frac{1}{2}(\alpha_\phi - \eta_\phi)' H_\phi^{-1}(\alpha_\phi - \eta_\phi))\Big]
\end{aligned} \tag{3}
$$

And for α_ξ:

$$
\begin{aligned}
\exp\Big[&-\frac{1}{2}(\xi^* - \mu'_\xi)'(\Psi+\tau_\xi^2 I_m)^{-1}(\xi^* - \mu'_\xi) - \frac{1}{2}(\alpha'_\xi - \eta_\xi)' H_\xi^{-1}(\alpha'_\xi - \eta_\xi)) \\
&\quad +\frac{1}{2}(\xi^* - \mu_\xi)'(\Psi+\tau_\xi^2 I_m)^{-1}(\xi^* - \mu_\xi) + \frac{1}{2}(\alpha_\xi - \eta_\xi)' H_\xi^{-1}(\alpha_\xi - \eta_\xi))\Big]
\end{aligned} \tag{4}
$$

Updates for the other hyperparameters are very similar - although none feature the MVN prior that α_ϕ and α_ξ do. In addition, the part involving the log determinant can't be dropped from the m-dim MVN, since the covariance matrix changes with any change in β, ν, ς^2 or τ^2. The remaining hyperparameters either have a discrete update (in which case the prior probability will be $^1/_{\#discrete.values}$ and will cancel above and below the line, or have a univariate Normal prior on them (or their log).

β

For β_ϕ we have:

$$\pi(\beta_\phi|\dots) \propto p(\phi^*|\mu_\phi, \tau_\phi^2, \Sigma)p(\beta_\phi)$$

Since β_ϕ has a discrete prior, to update β_ϕ we need to evaluate:

$$\begin{aligned}\exp\Big[&-\frac{1}{2}\log(\det(2\pi(\Sigma' + \tau_\phi^2 I_m))) - \frac{1}{2}(\phi^* - \mu_\phi)'(\Sigma' + \tau_\phi^2 I_m)^{-1}(\phi^* - \mu_\phi) \\ &+ \frac{1}{2}\log(\det(2\pi(\Sigma + \tau_\phi^2 I_m))) + \frac{1}{2}(\phi^* - \mu_\phi)'(\Sigma + \tau_\phi^2 I_m)^{-1}(\phi^* - \mu_\phi)\Big]\end{aligned} \tag{5}$$

where Σ' has been formed using β'_ϕ, the new proposal value.

Then for β_ξ we have:

$$\pi(\beta_\xi|\dots) \propto p(\xi^*|\mu_\xi, \tau_\xi^2, \Psi)p(\beta_\xi)$$

To update β_ξ we need to evaluate:

$$\begin{aligned}\exp\Big[&-\frac{1}{2}\log(\det(2\pi(\Psi' + \tau_\xi^2 I_m))) - \frac{1}{2}(\xi^* - \mu_\xi)'(\Psi' + \tau_\xi^2 I_m)^{-1}(\xi^* - \mu_\xi) \\ &+ \frac{1}{2}\log(\det(2\pi(\Psi + \tau_\xi^2 I_m))) + \frac{1}{2}(\xi^* - \mu_\xi)'(\Psi + \tau_\xi^2 I_m)^{-1}(\xi^* - \mu_\xi)\Big]\end{aligned} \tag{6}$$

where Ψ' has been formed using β'_ξ, the new proposal value.

$\boldsymbol{\nu}$

ν_ϕ also has a discrete update and so will look identical to the update for β_ϕ. It has conditional posterior distribution of:

$$\pi(\nu_\phi|\dots) \propto p(\phi^*|\mu_\phi, \tau_\phi^2, \Sigma)p(\nu_\phi)$$

We need to evaluate:

$$\exp\Big[-\frac{1}{2}\log(\det(2\pi(\Sigma' + \tau_\phi^2 I_m))) - \frac{1}{2}(\phi^* - \mu_\phi)'(\Sigma' + \tau_\phi^2 I_m)^{-1}(\phi^* - \mu_\phi) + \frac{1}{2}\log(\det(2\pi(\Sigma + \tau_\phi^2 I_m))) + \frac{1}{2}(\phi^* - \mu_\phi)'(\Sigma + \tau_\phi^2 I_m)^{-1}(\phi^* - \mu_\phi)\Big] \tag{7}$$

where Σ' has been formed using ν'_ϕ, the new proposal value.

In order to update ν_ξ we have:

$$\pi(\nu_\xi|\dots) \propto p(\xi^*|\mu_\xi, \tau_\xi^2, \Psi)p(\nu_\xi)$$

We therefore need to evaluate:

$$\exp\Big[-\frac{1}{2}\log(\det(2\pi(\Psi' + \tau_\xi^2 I_m))) - \frac{1}{2}(\xi^* - \mu_\xi)'(\Psi' + \tau_\xi^2 I_m)^{-1}(\xi^* - \mu_\xi) + \frac{1}{2}\log(\det(2\pi(\Psi + \tau_\xi^2 I_m))) + \frac{1}{2}(\xi^* - \mu_\xi)'(\Psi + \tau_\xi^2 I_m)^{-1}(\xi^* - \mu_\xi)\Big] \tag{8}$$

where Ψ' has been formed using ν'_ξ, the new proposal value.

ς^2

ς_ϕ^2 has a conditional posterior distribution of:

$$\pi(\varsigma_\phi^2|\dots) \propto p(\phi^*|\mu_\phi, \tau_\phi^2, \Sigma)p(\varsigma_\phi^2)$$

The prior distribution of the log of ς_ϕ^2 is a univariate Normal. So to update ς_ϕ^2 we need to evaluate:

$$
\begin{aligned}
\exp\Bigg[& -\frac{1}{2}\log(\det(2\pi(\Sigma' + \tau_\phi^2 I_m)))- \\
& \frac{1}{2}(\phi^* - \mu_\phi)'(\Sigma' + \tau_\phi^2 I_m)^{-1}(\phi^* - \mu_\phi) - \frac{(\log(\varsigma_\phi^{2'}) - m)^2}{2s^2} + \\
& \frac{1}{2}\log(\det(2\pi(\Sigma + \tau_\phi^2 I_m)))+ \\
& \frac{1}{2}(\phi^* - \mu_\phi)'(\Sigma + \tau_\phi^2 I_m)^{-1}(\phi^* - \mu_\phi) + \frac{(\log(\varsigma_\phi^{2}) - m)^2}{2s^2}\Bigg]
\end{aligned} \tag{9}
$$

where Σ' has been formed using $\varsigma_\phi^{2'}$, the new proposal value, and m and s are the prior mean and standard deviation repectively.

ς_ξ^2 has a conditional posterior distribution of:

$$
\pi(\varsigma_\xi^2 | \dots) \propto p(\xi^* | \mu_\xi, \tau_\xi^2, \Psi) p(\varsigma_\xi^2)
$$

To update ς_ξ^2 we need to evaluate:

$$
\begin{aligned}
\exp\Bigg[& -\frac{1}{2}\log(\det(2\pi(\Psi' + \tau_\xi^2 I_m)))- \\
& \frac{1}{2}(\xi^* - \mu_\xi)'(\Psi' + \tau_\xi^2 I_m)^{-1}(\xi^* - \mu_\xi) - \frac{(\log(\varsigma_\xi^{2'}) - m)^2}{2s^2} + \\
& \frac{1}{2}\log(\det(2\pi(\Psi + \tau_\xi^2 I_m)))+ \\
& \frac{1}{2}(\xi^* - \mu_\xi)'(\Psi + \tau_\xi^2 I_m)^{-1}(\xi^* - \mu_\xi) + \frac{(\log(\varsigma_\xi^{2}) - m)^2}{2s^2}\Bigg]
\end{aligned} \tag{10}
$$

where Ψ' has been formed using $\varsigma_\xi^{2'}$, the new proposal value, and m and s are the prior mean and standard deviation repectively.

$\boldsymbol{\tau^2}$

τ_ϕ^2 has a conditional posterior distribution of:

$$\pi(\tau_\phi^2|\dots) \propto p(\phi^*|\mu_\phi, \tau_\phi^2, \Sigma)p(\tau_\phi^2)$$

The prior distribution of the log of τ_ϕ^2 is a univariate Normal. So to update τ_ϕ^2 we need to evaluate:

$$\begin{aligned}\exp\Bigg[& -\frac{1}{2}\log(\det(2\pi(\Sigma+\tau_\phi^{2'}I_m)))- \\ & \frac{1}{2}(\phi^*-\mu_\phi)'(\Sigma+\tau_\phi^{2'}I_m)^{-1}(\phi^*-\mu_\phi)-\frac{(\log(\tau_\phi^{2'})-m)^2}{2s^2}+ \\ & \frac{1}{2}\log(\det(2\pi(\Sigma+\tau_\phi^{2}I_m)))+ \\ & \frac{1}{2}(\phi^*-\mu_\phi)'(\Sigma+\tau_\phi^{2}I_m)^{-1}(\phi^*-\mu_\phi)+\frac{(\log(\tau_\phi^{2})-m)^2}{2s^2}\Bigg]\end{aligned} \tag{11}$$

where m and s are the prior mean and standard deviation repectively.

τ_ξ^2 has a conditional posterior distribution of:

$$\pi(\tau_\xi^2|\dots) \propto p(\xi^*|\mu_\xi, \tau_\xi^2, \Psi)p(\tau_\xi^2)$$

To update τ_ξ^2 we need to evaluate:

$$\begin{aligned}\exp\Bigg[& -\frac{1}{2}\log(\det(2\pi(\Psi+\tau_\xi^{2'}I_m)))- \\ & \frac{1}{2}(\xi^*-\mu_\xi)'(\Psi+\tau_\xi^{2'}I_m)^{-1}(\xi^*-\mu_\xi)-\frac{(\log(\tau_\xi^{2'})-m)^2}{2s^2}+ \\ & \frac{1}{2}\log(\det(2\pi(\Psi+\tau_\xi^{2}I_m)))+ \\ & \frac{1}{2}(\xi^*-\mu_\xi)'(\Psi+\tau_\xi^{2}I_m)^{-1}(\xi^*-\mu_\xi)+\frac{(\log(\tau_\xi^{2})-m)^2}{2s^2}\Bigg]\end{aligned} \tag{12}$$

where m and s are the prior mean and standard deviation repectively.