Multivariate Nearest-Neighbors Gaussian Processes with Random Covariance Matrices

Isabelle Grenier *[1], Bruno Sansó †[1], and Jessica L. Matthews ‡[2]

[1]Department of Statistics, University of California, Santa Cruz
[2]NOAA's National Centers for Environmental Information

Supplementary Materials

Multivariate NN-RCM Full Conditionals

As previously mentioned, the structure, reminiscent of multivariate linear regression models, allows us to derive the full conditionals for each observed location s_i independently. Therefore, in the implementation of the MCMC algorithm, the spatial field $w(s_i)$ can be sampled in a fully parallel fashion. The full conditional posterior distribution for s_i, where $i = 1, ..., k$ are obtained using the Bayesian normal linear model posterior inference formulas,

$$\boldsymbol{w}(s_i)|- \sim N_q(\boldsymbol{u}(s_i), S(s_i)),$$

where

$$\begin{aligned} \boldsymbol{u}(s_i) &= S(s_i)\left(\Gamma'(s_i)\Phi^{-1}(s_i)\boldsymbol{w}_{N(s_i)} + A'\left(\boldsymbol{\tau}^2 I_q\right)^{-1}\boldsymbol{y}(s_i) + \sum_{u\in N^{-1}(s_i)}\Gamma'_{u,s_i}\Phi^{-1}(u)b_{u,s_i}\right) \\ S(s_i) &= \left(\Phi^{-1}(s_i) + A'\left(\boldsymbol{\tau}^2 I_q\right)^{-1}A + \sum_{u\in N^{-1}(s_i)}\Gamma'_{u,s_i}\Phi^{-1}(u)\Gamma_{u,s_i}\right)^{-1}, \end{aligned}$$

where the set $u \in N^{-1}(s_i)$ is the group of locations which has s_i as a neighbor. The notation Γ_{u,s_i} indicates the rows of $\Gamma(u)$ that corresponds to the neighbors located at s_i and $b_{u,s_i} = \boldsymbol{w}(u) - \sum_{r\in N(u)\setminus s_i}\Gamma'(u,r)\boldsymbol{w}_r$.

*igrenier@ucsc.edu
†bruno@soe.ucsc.edu
‡jessica.matthews@noaa.gov

The full conditionals for the spatial random effects $\Gamma(s_i)$ and $\Phi(s_i)$ can also be obtained for each location separately. Similar to the spatial field, we leverage the Bayesian linear model posterior inference equations to obtained the closed-form full conditional distributions.

$$\begin{aligned}
\Gamma(s_i)|- &\sim N_{mq,q}(U(s_i), V(s_i), \Phi(s_i)),\\
U(s_i) &= V(s_i)\left(\boldsymbol{w}_{N(s_i)}\boldsymbol{w}(s_i) + (\alpha - kq - 1)\mathcal{C}_{\boldsymbol{\theta},N(s_i),s_i}\right)\\
V(s_i) &= \left(\boldsymbol{w}_{N(s_i)}\boldsymbol{w}'_{N(s_i)} + (\alpha - kq - 1)\mathcal{C}_{\boldsymbol{\theta},N(s_i)}\right)^{-1}\\
\Phi(s_i)|- &\sim IW(\alpha - kq + 2q(1+m), (\alpha - kq - 1)\mathcal{C}_{\boldsymbol{\theta},s_i|N(s_i)} + Q(s_i) + R(s_i)\\
Q(s_i) &= (\boldsymbol{w}(s_i) - \Gamma'(s_i)\boldsymbol{w}_N(s_i))(\boldsymbol{w}(s_i) - \Gamma'(s_i)\boldsymbol{w}_N(s_i))'\\
R(s_i) &= (\alpha - kq - 1)(\Gamma(s_i) - \mathcal{C}^{-1}_{\boldsymbol{\theta},N(s_i)}\mathcal{C}_{\boldsymbol{\theta},N(s_i),s})'\mathcal{C}_{\boldsymbol{\theta},N(s_i)}(\Gamma(s_i) - \mathcal{C}^{-1}_{\boldsymbol{\theta},N(s_i)}\mathcal{C}_{\boldsymbol{\theta},N(s_i),s}).
\end{aligned}$$

Finally, for each sampling iteration of the MCMC algorithm, we have global full conditionals for each element τ_q^2 of the observational error $\boldsymbol{\tau}^2$. The likelihood and the choice of prior for the observational errors in the hierarchical structure leads to conditional conjugacy. Therefore, the full conditional of τ_q^2 is also an inverse gamma distribution,

$$\tau_q^2|- \sim IG\left(a_\tau + \frac{k}{2}, b_\tau + \sum_{s_i \in S} \frac{(w_q(s_i) - y_q(s_i))^2}{2}\right).$$

Special case: univariate observations full conditionals

We again call attention to the special case of univariate spatial observations. Setting $q = 1$, the model simplifies to

$$y(s) = \boldsymbol{x}(s)\boldsymbol{\beta} + w(s) + \epsilon(s),$$

where

$$\begin{aligned}
y(s)|\boldsymbol{\beta}, w(s), \tau^2 &\sim N(\boldsymbol{x}(s)\boldsymbol{\beta} + w(s), \tau^2)\\
\boldsymbol{\beta} &\sim N_p(0, s_\beta^2 I)\\
w(s)|\gamma(s), \phi(s), \sigma^2 &\sim N(\gamma(s)\boldsymbol{w}_{N(s)}, \sigma^2\phi(s))\\
\gamma(s)|\phi(s) &\sim N_m(\mathcal{C}^{-1}_{\boldsymbol{\theta},N(s)}C_{\boldsymbol{\theta},N(s),s}, \frac{\phi(s)}{\alpha - k - 1}\mathcal{C}^{-1}_{\boldsymbol{\theta},N(s)})\\
\phi(s) &\sim IG(\alpha - k + (1+m), (\alpha - k - 1)\mathcal{C}_{\boldsymbol{\theta},s|N(s)}),\\
\pi(\sigma^2, \tau^2) &= IG(\tau^2|a_\tau, b_\tau) \times IG(\sigma^2|a_\sigma, b_\sigma),
\end{aligned}$$

where p is the dimension of the fixed effects and m is the number of neighbors. We recognize the framework of a multivariate linear regression model in the prior structure of the spatial random effects $w(s)$. By using the posterior equations from the Normal-Normal model, we obtain a closed-form solution for the full conditional distribution of $w(s)$:

$$w(s_i)|- \sim N(\mu(s_i), \sigma^2(s_i)),$$

where

$$\mu(s_i) = \left(\frac{1}{\sigma^2\phi(s_i)} + \frac{1}{\tau^2}\right)^{-1}\left(\frac{\gamma'(s_i)\boldsymbol{w}_{N(s_i)}}{\sigma^2\phi(s_i)} + \frac{y(s_i)}{\tau^2} + \sum_{u\in N^{-1}(s_i)}\frac{\gamma_{u,s_i}}{\sigma^2\phi(u)}b_{u,s_i}\right)$$

$$\sigma^2(s_i) = \left(\frac{1}{\sigma^2\phi(s_i)} + \frac{1}{\tau^2} + \sum_{u\in N^{-1}(s_i)}\frac{\gamma^2_{u,s_i}}{\sigma^2\phi(u)}\right)^{-1},$$

where the notation γ_{u,s_i} indicates the element of $\gamma(u)$ that corresponds to the neighbors located at s_i and $b_{u,s_i} = w(u) - \sum_{r\in N(u)\setminus s_i}\gamma_{u,r}w(r)$.

To obtain the full conditional distributions of the random effects $\boldsymbol{\gamma}(s_i)$ and $\phi(s_i)$, we use the normal equations from the linear models posterior inference.

$$\begin{aligned}
\boldsymbol{\gamma}(s_i)|- &\sim N_m(\boldsymbol{\mu}_{\boldsymbol{\gamma}(s_i)}, h^2_{\boldsymbol{\gamma}(s_i)}),\\
\boldsymbol{\mu}_{\boldsymbol{\gamma}(s_i)} &= \frac{h^2_{\boldsymbol{\gamma}(s_i)}}{\phi(s_i)}\left(\frac{\boldsymbol{w}_{N(s_i)}w(s_i)}{\sigma^2} + (\alpha-k-1)C_{\boldsymbol{\theta},N(s_i),s_i}\right)\\
h^2_{\boldsymbol{\gamma}(s_i)} &= \phi(s_i)\left(\frac{\boldsymbol{w}_{N(s_i)}\boldsymbol{w}'_{N(s_i)}}{\sigma^2} + (\alpha-k-1)C_{\boldsymbol{\theta},N(s_i)}\right)^{-1}\\
\phi(s_i)|- &\sim IG\left(\alpha-k+1+m+\frac{m+1}{2}, (\alpha-k-1)C_{\boldsymbol{\theta},s_i|N(s_i)} + \frac{q(s_i)}{2\sigma^2} + \frac{r(s_i)}{2}\right)\\
q(s_i) &= (w(s_i) - \boldsymbol{\gamma}'(s_i)\boldsymbol{w}_N(s_i))^2\\
r(s_i) &= (\alpha-k-1)(\boldsymbol{\gamma}(s_i) - C^{-1}_{\boldsymbol{\theta},N(s_i)}C_{\boldsymbol{\theta},N(s_i),s})'C_{\boldsymbol{\theta},N(s_i)}(\boldsymbol{\gamma}(s_i) - C^{-1}_{\boldsymbol{\theta},N(s_i)}C_{\boldsymbol{\theta},N(s_i),s})
\end{aligned}$$

Finally, we are left with the posterior inference of the observational error τ^2 and the partial sill σ^2. In contrast to the multivariate model, the partial sill is factored out of the covariance matrix. In the previously defined model, we therefore have

$$\sigma^2 C_{\boldsymbol{\theta}} = \sigma^2(C_\nu + \xi^2 I).$$

This allows us to learn the full posterior distribution for the partial sill under the hierarchical NN-RCM model. The choice of prior for the observational error and partial sill both lead to closed-form full conditional distributions:

$$\sigma^2|- \sim IG\left(a_\sigma + \frac{k}{2}, b_\sigma + \sum_{s_i\in S}\frac{(w(s_i) - \boldsymbol{\gamma}'(s_i)\boldsymbol{w}_N(s_i))^2}{2\phi(s_i)}\right)$$

$$\tau^2|- \sim IG\left(a_\tau + \frac{k}{2}, b_\tau + \sum_{s_i\in S}\frac{(w(s_i) - y(s_i))^2}{2}\right).$$

Sensitivity Analysis of Parameter Choices

A sensitivity analysis of the choice of number of neighbors, the range parameter as well as the smoothness parameter was conducted for the univariate simulated dataset setting. The goal of the analysis is to explore the impact of such choices on the inference and on the predictive power of the model.

Following the same framework as presented in section 3.1, we estimate the posterior parameters of the NNRCM model using $m = 10, 15$ and 20 neighbors, a range ν varying between 0.6 and 1, and three possible smoothness parameter $\kappa = 0.5, 1$ and 1.5. Table 1 to Table 3 summarize the posterior estimates found for each combination of smoothness and number of neighbors for two different range values ($\nu = 0.6$ and 1). Table 2 to Table 4 then look at the predictive scores under the same two simulations.

We conclude that the posterior estimates are not sensitive to the neighborhood size as they remain unchanged for a given range and smoothness parameter. As expected, the sensitivity to the smoothness parameter is such that the posterior estimate of the range becomes smaller (larger) as the smoothness parameter increases (decreased).

	$m = 10$			$m = 15$			$m = 20$		
	$\kappa = 0.5$	$\kappa = 1$	$\kappa = 1.5$	$\kappa = 0.5$	$\kappa = 1$	$\kappa = 1.5$	$\kappa = 0.5$	$\kappa = 1$	$\kappa = 1.5$
α	2534	2536	2535	2590	2590	2582	2532	2544	2540
σ^2	0.79	0.75	0.73	0.8	0.75	0.72	0.81	0.76	0.73
τ	0.33	0.41	0.47	0.3	0.40	0.47	0.29	0.39	0.46
ν	0.68	0.47	0.38	0.72	0.49	0.4	0.72	0.5	0.4

Table 1: Estimated parameters for varying number of nearest neighbors and smoothness parameter (data simulated with $\nu = 1$)

	$m = 10$			$m = 15$			$m = 20$		
	$\kappa = 0.5$	1	1.5	$\kappa = 0.5$	1	1.5	$\kappa = 0.5$	1	1.5
PMSE	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26
CRPS	0.30	0.29	0.29	0.30	0.29	0.29	0.30	0.29	0.29
PPLC	845	759	749	808	740	738	797	734	735
Coverage	1	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99

Table 2: Posterior predictive assessment using test set for varying number of nearest neighbors and smoothness parameter (data simulated with $\nu = 1$)

	$m = 10$			$m = 15$			$m = 20$		
	$\kappa = 0.5$	$\kappa = 1$	$\kappa = 1.5$	$\kappa = 0.5$	$\kappa = 1$	$\kappa = 1.5$	$\kappa = 0.5$	$\kappa = 1$	$\kappa = 1.5$
α	6739	6751	6709	6499	6494	6507	6686	6645	6600
σ^2	0.81	0.72	0.69	0.8	0.70	0.67	0.79	0.68	0.65
τ	0.25	0.42	0.48	0.27	0.47	0.53	0.28	0.49	0.56
ν	1.23	0.87	0.77	1.29	0.91	0.8	1.30	0.93	0.82

Table 3: Estimated parameters for varying number of nearest neighbors and smoothness parameter (data simulated with $\nu = 0.6$)

	$m = 10$			$m = 15$			$m = 20$		
	$\kappa = 0.5$	1	1.5	$\kappa = 0.5$	1	1.5	$\kappa = 0.5$	1	1.5
PMSE	0.13	0.13	0.13	0.12	0.13	0.13	0.12	0.13	0.14
CRPS	0.21	0.22	0.22	0.21	0.22	0.23	0.21	0.22	0.23
PPLC	1234	1353	1430	1263	1410	1481	1269	1439	1508
Coverage	1	1	1	1	1	1	1	1	1

Table 4: Posterior predictive assessment using test set for varying number of nearest neighbors and smoothness parameter (data simulated with $\nu = 0.6$)

Sensitivity Analysis of Prior Distributions

A sensitivity analysis of the prior distributions for the partial sill σ^2 and the nugget parameter ξ^2 was conducted for a univariate simulated dataset.

For this simulation study, 10,000 observations were generated with a range parameter $\nu = 1$, partial sill $\sigma^2 = 1$ and using an exponential covariance function. The dataset was evenly assigned to a training to fit the model and a testing set to assess predictive performance.

We evaluate the performance of the model using five different priors for σ^2 and four different priors for ξ^2. The prior distributions are shown in Figure 1 to illustrate the variability of their shape. The prior for the degrees of freedom α is kept constant, with the choice of a Pareto distribution with shape parameter equal to 1. Note that for the reported results below, the choice of neighborhood size is $m = 15$ and the smoothness parameter is kept constant at $\kappa = 0.5$.

For Table 5 and Table 6, we note that the posterior estimates for σ^2 and ξ^2 are more sensitive to the choice of prior for ξ^2 than the one for σ^2. The estimate of the range ν remains constant under any prior specification. In terms of the predictive power, the model performs similarly under any prior specification. Thus, we conclude that the model produces reliable estimates and is not sensitive to the choice of prior.

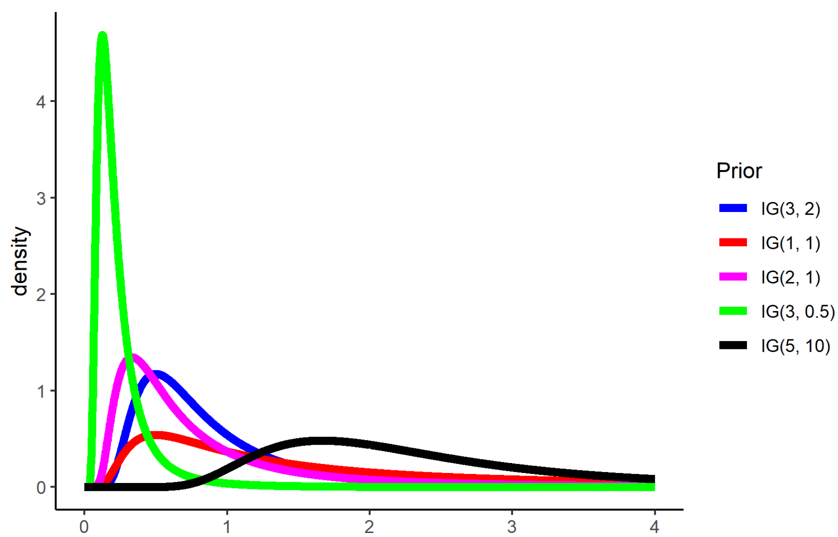

Figure 1: Prior choices for σ^2 and ξ^2 for the sensitivity analysis.

Prior for ξ^2	Prior for σ^2	σ^2	ν	ξ^2
IG(3,2)	IG(3, 2)	0.838	0.761	0.251
	IG(1, 1)	0.84	0.759	0.249
	IG(2, 1)	0.838	0.761	0.252
	IG(3, 0.5)	0.835	0.762	0.255
	IG(5, 10)	0.854	0.754	0.234
IG(1, 1)	IG(3, 2)	0.859	0.747	0.222
	IG(1, 1)	0.861	0.746	0.22
	IG(2, 1)	0.859	0.747	0.223
	IG(3, 0.5)	0.855	0.748	0.228
	IG(5, 10)	0.876	0.739	0.204
IG(3, 0.5)	IG(3, 2)	0.916	0.709	0.151
	IG(1, 1)	0.918	0.708	0.148
	IG(2, 1)	0.916	0.709	0.151
	IG(3, 0.5)	0.912	0.711	0.155
	IG(5, 10)	0.931	0.703	0.137
IG(5, 10)	IG(3, 2)	0.71	0.846	0.465
	IG(1, 1)	0.711	0.846	0.463
	IG(2, 1)	0.709	0.847	0.467
	IG(3, 0.5)	0.706	0.849	0.473
	IG(5, 10)	0.727	0.839	0.437

Table 5: Comparison of parameter estimates under different prior specification for σ^2 and ξ^2. The prior for ν is Gamma(2, $2/d$), where d is the minimum distance between two observations)

Prior for ξ^2	Prior for σ^2	PMSE	CRPS	PPLC
	IG(3, 2)	0.148	0.230	2167
	IG(1, 1)	0.148	0.230	2158
IG(3,2)	IG(2, 1)	0.148	0.230	2164
	IG(3, 0.5)	0.148	0.230	2173
	IG(5, 10)	0.148	0.230	2168
	IG(3, 2)	0.148	0.229	2077
	IG(1, 1)	0.148	0.228	2069
IG(1, 1)	IG(2, 1)	0.148	0.229	2082
	IG(3, 0.5)	0.148	0.229	2096
	IG(5, 10)	0.148	0.228	2031
	IG(3, 2)	0.150	0.226	1854
	IG(1, 1)	0.150	0.226	1846
IG(3, 0.5)	IG(2, 1)	0.150	0.226	1853
	IG(3, 0.5)	0.150	0.226	1864
	IG(5, 10)	0.150	0.225	1807
	IG(3, 2)	0.151	0.242	2689
	IG(1, 1)	0.151	0.241	2682
IG(5, 10)	IG(2, 1)	0.152	0.242	2691
	IG(3, 0.5)	0.151	0.242	2701
	IG(5, 10)	0.151	0.240	2634

Table 6: Posterior predictive assessment under different prior specification for σ^2 and ξ^2. The prior for ν is Gamma(2, $2/d$), where d is the minimum distance between two observations)