Bayesian Analysis (2021) **TBA**, Number TBA, pp. 1

Contributed Discussion

Arnab Hazra* and Raphaël Huser*

It was our pleasure to read this paper and to get the opportunity to discuss it. Studying the convergence and mixing of the Markov chain Monte Carlo (MCMC) chains is often neglected. Here, the authors raise this point and obtain some theoretical results about the convergence of the Gibbs sampler for multilevel conditionally hierarchical Gaussian models using multigrid decomposition. The authors also go beyond the Gaussian case and describe an example of a Poisson crossed-effects model. Importantly, the authors also discuss two different types of parametrizations for the same models, the so-called non-centered and centered parametrizations (NCP and CP, respectively).

Here, we focus on studying the convergence properties of the Gibbs sampler under different parametrizations used in some recent papers on spatial geostatistics and spatial extreme-value analysis using hierarchical Gaussian processes (GP), where independent (temporal) replications are available. Instead of focusing on analytic expressions, we focus on simulations. In a purely spatial setting, Bass and Sahu (2017) studied the convergence rates under different choices of the spatial correlation structures for a GP.

First, we consider a simple spatial Gaussian process model (Banerjee et al., 2003, Chapter 5) defined as

$$Y_t(\boldsymbol{s}) = \mu + \varepsilon_t(\boldsymbol{s}) + \eta_t(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathcal{D} \subset \mathbb{R}^2,\ t = 1, \ldots, T, \tag{1}$$

where μ denotes the global mean, $\varepsilon_t(\cdot)$ are independent and identically distributed (IID) zero-mean GPs with spatial covariance $\mathrm{Cov}\{\varepsilon_t(\boldsymbol{s}_1), \varepsilon_t(\boldsymbol{s}_2)\} = r\exp\{-d(\boldsymbol{s}_1, \boldsymbol{s}_2)/\phi\}, r, \phi > 0$, with $d(\boldsymbol{s}_1, \boldsymbol{s}_2)$ denoting the Euclidean distance between $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$, and $\eta_t(\boldsymbol{s}) \overset{\mathrm{IID}}{\sim} \mathrm{Normal}(0, 1-r)$. We simulate $T = 100$ replications at $N = 121$ uniform spatial grid locations $\{(i,j) : i, j = 0, 0.1, \ldots, 1\}$. True parameter choices are $\mu = 5$, $\phi = 0.2$ and $r = 0.9$. Here, conjugate priors for ϕ and r are not known and hence, to stick to Gibbs sampling, we prefer to treat them as known and choose a weakly informative prior $\mu \sim \mathrm{Normal}(0, 100^2)$. Let $\boldsymbol{X}_t = [X_t(\boldsymbol{s}_1), \ldots, X_t(\boldsymbol{s}_N)]'$ be the generic notation for the spatial vectors and $\boldsymbol{\Sigma}_\phi$ be the correlation matrix obtained from $\mathrm{Cov}\{\varepsilon_t(\boldsymbol{s}_i), \varepsilon_t(\boldsymbol{s}_j)\}, i, j = 1, \ldots, N$. We fit (1) under NCP and CP. In NCP, we treat the levels as $\boldsymbol{Y}_t \overset{\mathrm{Indep}}{\sim} \mathrm{Normal}(\mu\boldsymbol{1} + \boldsymbol{\varepsilon}_t, (1-r)\boldsymbol{I}_N)$, $\boldsymbol{\varepsilon}_t \overset{\mathrm{IID}}{\sim} \mathrm{Normal}(\boldsymbol{0}, r\boldsymbol{\Sigma}_\phi)$, and $\mu \sim \mathrm{Normal}(0, 100^2)$. In CP, the levels are modified as $\boldsymbol{Y}_t \overset{\mathrm{Indep}}{\sim} \mathrm{Normal}(\tilde{\boldsymbol{\varepsilon}}_t, (1-r)\boldsymbol{I}_N)$, $\tilde{\boldsymbol{\varepsilon}}_t \overset{\mathrm{IID}}{\sim} \mathrm{Normal}(\mu\boldsymbol{1}, r\boldsymbol{\Sigma}_\phi)$, and $\mu \sim \mathrm{Normal}(0, 100^2)$. We study the trace plots and the autocorrelation function (ACF) plots of μ and $\bar{\varepsilon} = (NT)^{-1}\sum_{i=1}^N \sum_{t=1}^T \varepsilon_t(\boldsymbol{s}_i)$ under NCP, and of μ and $\bar{\tilde{\varepsilon}} = (NT)^{-1}\sum_{i=1}^N \sum_{t=1}^T \tilde{\varepsilon}_t(\boldsymbol{s}_i)$ under CP, and observe good mixing for CP while the mixing is poor for NCP. This corroborates with the results for the two-layer models

*Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia. Email: arnab.hazra@kaust.edu.sa, raphael.huser@kaust.edu.sa

 DOI: 0000

mentioned in the paper. The corresponding effective sample sizes (ESS) are presented in Table 1. Theoretical results follow from Bass and Sahu (2017).

GPs have been criticized for modeling spatial extremes, due to their light tails and inability to capture strong tail dependence. To extend this class, while retaining the computational attractiveness of GPs, several authors have proposed some location and/or scale mixture models (e.g., Huser et al., 2017; Morris et al., 2017; Krupskii et al., 2018; Hazra et al., 2020) and Huser and Wadsworth (2020) reviewed some of them. Here, we focus on some of them which allow Gibbs sampling for the higher level random effects.

We next consider a simplified location-mixture model (Krupskii et al., 2018)

$$Y_t(\boldsymbol{s}) = E_t + \varepsilon_t(\boldsymbol{s}) + \eta_t(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathcal{D} \subset \mathbb{R}^2, \ t = 1, \ldots, T, \tag{2}$$

where $E_t \overset{\text{IID}}{\sim} \text{Exponential}(\lambda)$ and the specifications for $\varepsilon_t(\cdot)$ and $\eta_t(\cdot)$ are the same as in (1); here, $\lambda > 0$ is the rate parameter. We choose a weakly informative prior $\lambda \sim \text{Gamma}(0.01, 0.01)$. The model has three layers and allows Gibbs sampling for the unknown parameters and the latent variables. The simulation design is the same as before and we choose the true value $\lambda = 1$. We fit (2) under NCP and CP. In NCP, we treat the levels as $\boldsymbol{Y}_t \overset{\text{Indep}}{\sim} \text{Normal}(E_t\boldsymbol{1} + \boldsymbol{\varepsilon}_t, (1-r)\boldsymbol{I}_N)$, $\boldsymbol{\varepsilon}_t \overset{\text{IID}}{\sim} \text{Normal}(\boldsymbol{0}, r\boldsymbol{\Sigma}_\phi)$, $E_t \overset{\text{IID}}{\sim} \text{Exponential}(\lambda)$, and $\lambda \sim \text{Gamma}(0.01, 0.01)$. In CP, we replace the first two levels of NCP by $\boldsymbol{Y}_t \overset{\text{Indep}}{\sim} \text{Normal}(\tilde{\boldsymbol{\varepsilon}}_t, (1-r)\boldsymbol{I}_N)$ and $\tilde{\boldsymbol{\varepsilon}}_t \overset{\text{IID}}{\sim} \text{Normal}(E_t\boldsymbol{1}, r\boldsymbol{\Sigma}_\phi)$, respectively. The trace plots and ACF plots of $\bar{\varepsilon}$ and $\bar{\tilde{\varepsilon}}$ show a similar pattern as that for (1). The trace plots and ACF plots of $\bar{E} = T^{-1}\sum_{t=1}^T E_t$ and λ show good mixing under CP while it is poor for $\bar{E}$ under NCP. The corresponding ESS are presented in Table 1.

Finally, we consider a scale-mixture model (Morris et al., 2017; Hazra et al., 2020)

$$Y_t(\boldsymbol{s}) = \sqrt{b\tau_t}\{\varepsilon_t(\boldsymbol{s}) + \eta_t(\boldsymbol{s})\}, \quad \boldsymbol{s} \in \mathcal{D} \subset \mathbb{R}^2, \ t = 1, \ldots, T, \tag{3}$$

where $\tau_t \overset{\text{IID}}{\sim} \text{Inverse-gamma}(a/2, a/2)$ and the other terms are as before. We choose the prior $a \sim \text{Discrete-uniform}(0.1, 0.2, \ldots, 50)$ similar to Hazra et al. (2020) and Hazra and Huser (2021), and a flat prior for b over $\mathbb{R}_+$. While different representations of the same model are possible, not all of them allow Gibbs sampling for τ_t, and thus, we skip them here. The model (3) has three layers and allows Gibbs sampling for the unknown parameters (probability proportional to size sampling for a) and the latent variables. The simulation design is the same as before and we choose the true values $a = 5$ and $b = 1$. We fit (3) under non-scaled and scaled parametrizations (NSP and SP, respectively). In NSP, we treat the levels as $\boldsymbol{Y}_t \overset{\text{Indep}}{\sim} \text{Normal}(\tilde{\boldsymbol{\varepsilon}}_t, b(1-r)\tau_t\boldsymbol{I}_N)$, $\tilde{\boldsymbol{\varepsilon}}_t \overset{\text{Indep}}{\sim} \text{Normal}(\boldsymbol{0}, br\tau_t\boldsymbol{\Sigma}_\phi)$, $\tau_t \overset{\text{IID}}{\sim} \text{Inverse-gamma}(a/2, a/2)$, and the priors for a and b. In SP, we replace the first three levels of NSP by $\boldsymbol{Y}_t \overset{\text{Indep}}{\sim} \text{Normal}(\tilde{\boldsymbol{\varepsilon}}_t^*, (1-r)\tilde{\tau}_t\boldsymbol{I}_N)$, $\tilde{\boldsymbol{\varepsilon}}_t^* \overset{\text{Indep}}{\sim} \text{Normal}(\boldsymbol{0}, r\tilde{\tau}_t\boldsymbol{\Sigma}_\phi)$, and $\tilde{\tau}_t \overset{\text{IID}}{\sim} \text{Inverse-gamma}(a/2, ab/2)$, respectively. The trace plots and ACF plots of $\bar{\tilde{\varepsilon}}$ and $\bar{\tilde{\varepsilon}}^*$ (notation as before) and a show good mixing under both NSP and SP. The trace plots and ACF plots of $\bar{\tau} = T^{-1}\sum_{t=1}^T \tau_t$, $\bar{\tilde{\tau}} = T^{-1}\sum_{t=1}^T \tilde{\tau}_t$, and b show a good mixing behavior under SP (after thinning by keeping one per four/five

Table 1: Effective sample sizes (ESS) for the Gibbs samplers for models (1), (2), and (3), under different parametrizations. ESS values correspond to 10^4 iterations, starting from true parameter choices.

Parametrization	GP $(\mu, \bar{\varepsilon}/\tilde{\bar{\varepsilon}})$	Location-mixture $(\bar{\varepsilon}/\tilde{\bar{\varepsilon}}, E, \lambda)$	Scale-mixture$(\tilde{\bar{\varepsilon}}/\tilde{\bar{\varepsilon}}^*, \bar{\tau}/\tilde{\bar{\tau}}, a, b)$
NCP/NSP	(48,47)	(55, 53, 902)	(10^4, 95, 6940, 94)
CP/SP	(10^4, 10^4)	(10^4, 8531, 8896)	(10^4, 3256, 7292, 8146)

samples, for $\tilde{\bar{\tau}}$) but not under NSP. The corresponding ESS are presented in Table 1. Specifying $\tau_t \sim$ Inverse-gamma$(a/2, b/2)$ as in Morris et al. (2017) show long-range dependence in the ACF plots and using $\tau_t \sim$ Inverse-gamma$(a/2, ab/2)$, as in Hazra et al. (2020), is recommended.

Overall, through simulation studies, we have illustrated the mixing of a Gibbs sampler under different parametrizations for some popular models for spatial geostatistics and spatial extrems in the light of multigrid decomposition proposed in this paper, which would help practitioners to design the MCMC effectively. Theoretical derivations of the convergence rates in these spatial settings is a possible future endeavor.

References

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2003). *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC. 1

Bass, M. R. and Sahu, S. K. (2017). "A comparison of centring parameterisations of Gaussian process-based models for Bayesian computation using MCMC." *Statistics and Computing*, 27(6): 1491–1512. 1, 2

Hazra, A. and Huser, R. (2021). "Estimating high-resolution Red Sea surface temperature hotspots, using a low-rank semiparametric spatial model." *The Annals of Applied Statistics*, 15(2): 572 – 596. 2

Hazra, A., Reich, B. J., and Staicu, A.-M. (2020). "A multivariate spatial skew-t process for joint modeling of extreme precipitation indexes." *Environmetrics*, 31(3): e2602. 2, 3

Huser, R., Opitz, T., and Thibaud, E. (2017). "Bridging asymptotic independence and dependence in spatial extremes using Gaussian scale mixtures." *Spatial Statistics*, 21: 166–186. 2

Huser, R. and Wadsworth, J. L. (2020). "Advances in statistical modeling of spatial extremes." *Wiley Interdisciplinary Reviews: Computational Statistics*, e1537. 2

Krupskii, P., Huser, R., and Genton, M. G. (2018). "Factor copula models for replicated spatial data." *Journal of the American Statistical Association*, 113(521): 467–479. 2

Morris, S. A., Reich, B. J., Thibaud, E., and Cooley, D. (2017). "A space-time skew-t model for threshold exceedances." *Biometrics*, 73(3): 749–758. 2, 3