2024 International Society for Bayesian Analysis (ISBA) World Meeting — Venice, Italy

Accelerated Meta-Kriging for Massive Spatial Dataset

Luca Presicce [1] Sudipto Banerjee [2]

[1]Department of Economics, Management and Statistics, University of Milano - Bicocca

[2]Department of Biostatistics, University of California Los Angeles

HIT ME UP!

1. Introduction

Geostatistical modeling is afflicted by onerous computational effort when the number of locations is vast (the so-called "Big-n" problem).

- Notwithstanding burgeoning literature, spatial inference remains unfeasible for moderate data sets on modest computing environments without access to high-powered architectures.
- Our efforts fall into "meta-" approaches: a massive data set is split into smaller sets, analyzed independently, and local results combined to approximate full Bayesian inference.
- We introduce Bayesian predictive stacking (BPS) in spatial meta-analysis, providing feasible uncertainty quantification without demanding hardware or slowpoke run-time.

2. Conjugate spatial regression - Latent model

Let's consider the following hierarchical model

$$\begin{aligned} y \mid \omega, \beta, \sigma^2 &\sim N(X\beta + \omega, \delta^2\sigma^2\mathbb{I}_n) \\ \omega \mid \sigma^2 &\sim N\left(0, \sigma^2\rho_\phi(\mathcal{S},\mathcal{S})\right) \\ \beta \mid \sigma^2 &\sim N(\mu_\beta, \sigma^2 V_\beta) \\ \sigma^2 &\sim IG(a_\sigma, b_\sigma). \end{aligned} \tag{1}$$

- $\mathcal{S} = \{s_1, \dots, s_n\} \in \mathbb{R}^2$ be a set of n locations,
- $y = [y(s_i)]^\top$ is $n \times 1$ vector $(i = 1, \dots, n)$,
- $X = [x(s_i)^\top]$ is $n \times p$ matrix full rank p $(i = 1, \dots, n)$,
- $\mathscr{D} = \{y, X\}$ be the observed dataset over $\mathcal{S}$.
- $\delta^2 := \tau^2/\sigma^2 \in [0, 1]$ is the noise-to-spatial variance ratio,
- $\omega = [\omega(s_i)]^\top$ is $n \times 1$ latent process $(i = 1, \dots, n)$,
- $\rho_\phi(\mathcal{S},\mathcal{S})$ be the $n \times n$ spatial correlation matrix,
- $\phi \in \mathbb{R}^+$ index spatial correlation function $\rho_\phi(\cdot,\cdot)$.

We cast the model in Equation (1) into an augmented linear system

$$\underbrace{\begin{bmatrix} y \\ \mu_\beta \\ 0 \end{bmatrix}}_{y_\star} = \underbrace{\begin{bmatrix} X & \mathbb{I}_n \\ \mathbb{I}_p & 0 \\ 0 & \mathbb{I}_n \end{bmatrix}}_{X_\star} \underbrace{\begin{bmatrix} \beta \\ \omega \end{bmatrix}}_{\gamma} + \underbrace{\begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix}}_{\eta_\star}, \quad \eta_\star \sim N(0, \sigma^2 V_\star), \quad V_\star = \begin{bmatrix} \delta^2\mathbb{I}_n & 0 & 0 \\ 0 & V_\beta & 0 \\ 0 & 0 & \rho_\phi(\mathcal{S},\mathcal{S}) \end{bmatrix}.$$

For any fixed $\{\phi, \delta^2\}$ we have conjugacy, leading to the posterior density

$$\begin{aligned} p\left(\gamma, \sigma^2 \mid \mathscr{D}\right) &= p\left(\sigma^2 \mid \mathscr{D}\right) p\left(\gamma \mid \sigma^2, \mathscr{D}\right) \\ &= IG\left(\sigma^2 \mid a_\sigma^\star, b_\sigma^\star\right) N\left(\gamma \mid \hat{\gamma}, \sigma^2 M_\star\right). \end{aligned} \tag{2}$$

- $a_\sigma^\star = a_\sigma + n/2$,
- $b_\sigma^\star = b_\sigma + 1/2\,(y_\star - X_\star\hat{\gamma})^\top V_\star^{-1}(y_\star - X_\star\hat{\gamma})$,
- $M_\star^{-1} = X_\star^\top V_\star^{-1} X_\star$,
- $\hat{\gamma} = M_\star X_\star^\top V_\star^{-1} y_\star$.

Spatial predictive inference follows in closed form from the posterior distribution (Zhang et al., 2021; Banerjee, 2020)

$$\begin{aligned} p(\omega_\mathcal{U}, y_\mathcal{U} \mid \mathscr{D}) &= \int p\left(y_\mathcal{U} \mid \omega_\mathcal{U}, \beta, \sigma^2\right) p\left(\omega_\mathcal{U} \mid \omega, \sigma^2\right) p\left(\gamma, \sigma^2 \mid \mathscr{D}\right) d\gamma d\sigma^2 \\ &= t_{2a_\sigma^\star}\left(\tilde{\mu}, \left(b_\sigma^\star/a_\sigma^\star\right)\tilde{M}\right). \end{aligned} \tag{3}$$

- $\mathcal{U} = \{u_1, \dots, u_{n'}\} \in \mathbb{R}^2$ be a set of n' unknown locations,
- $y_\mathcal{U} = [y(u_i)]^\top$ is $n' \times 1$ vector $(i = 1, \dots, n')$,
- $X_\mathcal{U} = [x(u_i)^\top]$ is $n' \times p$ matrix $(i = 1, \dots, n')$,
- $\omega_\mathcal{U} = [\omega(u_i)]^\top$ is $n' \times 1$ latent process $(i = 1, \dots, n')$,
- $\rho_\phi(\mathcal{U},\mathcal{U})$ be the $n' \times n'$ spatial correlation at $\mathcal{U}$.
- $\tilde{\mu} = W\hat{\gamma}, \quad \tilde{M} = WM_\star W^\top + V_e$,
- $V_\omega = \rho_\phi(\mathcal{U},\mathcal{U}) - \rho_\phi(\mathcal{U},\mathcal{S})\rho_\phi^{-1}(\mathcal{S},\mathcal{S})\rho_\phi^\top(\mathcal{U},\mathcal{S})$,
- $M_\omega = \rho_\phi(\mathcal{U},\mathcal{S})\rho_\phi^{-1}(\mathcal{S},\mathcal{S})$,
- $W = \begin{bmatrix} 0 & M_\omega \\ X_\mathcal{U} & M_\omega \end{bmatrix}, \quad V_e = \begin{bmatrix} V_\omega & V_\omega \\ V_\omega & V_\omega + \delta^2\mathbb{I}_{n'} \end{bmatrix}$.

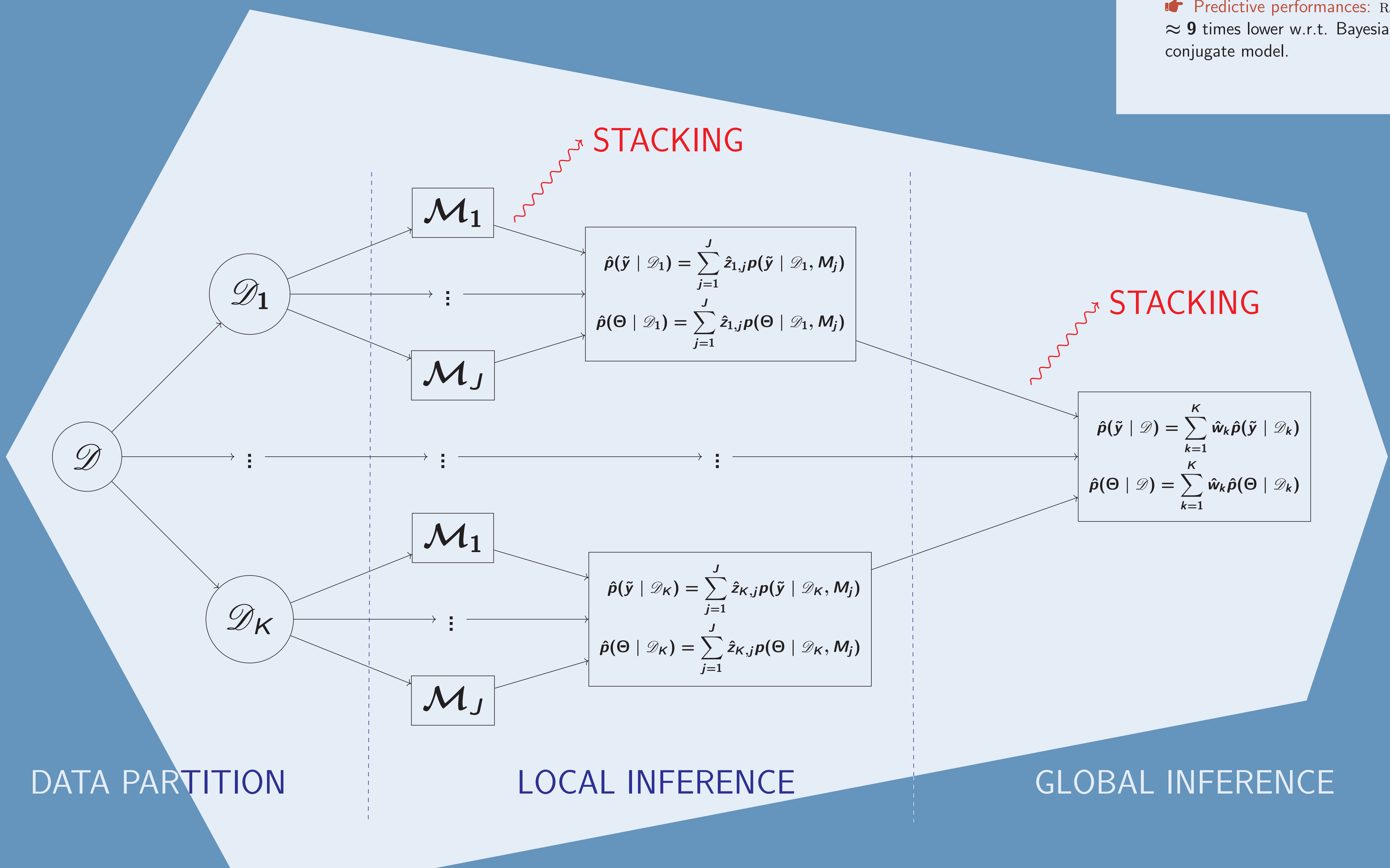

Figure: Double Bayesian predictive stacking approach representation

3. Bayesian Stacking of Predictive Densities

Bayesian predictive stacking (BPS) of predictive densities, assimilates different models defining the stacking weights as the following convex optimization problem (Yao et al., 2018; Gneiting and Raftery, 2007):

$$\max_{w \in \mathcal{S}_1^J} \frac{1}{n} \sum_{i=1}^n \log \sum_{j=1}^J w_j p\left(y_i \mid \mathscr{D}_{-i}, M_j\right), \tag{4}$$

- $\mathcal{M} = \{M_1, \dots, M_J\}$ competitive models specified with a collection of values for $\{\delta^2, \phi\}$,
- Equation (4) minimizes KL divergence from the (unknown) true predictive distribution,
- we use the K-fold cross-validation estimate of the expected value of the logarithm score.

4. Accelerated Learning for Spatial Random Fields

In order to accelerate the learning for spatial modeling by BPS

- Data partition: we divide the full data into K subsets: $\mathscr{D} = \{\mathscr{D}_1, \dots, \mathscr{D}_K\}$.
- Local inference: Within each partition, we obtain stacked estimations of posterior distributions over the J competitive models

$$\begin{aligned} &\hat{p}(\cdot \mid \mathscr{D}_k) = \sum_{j=1}^J \hat{z}_{k,j} p(\cdot \mid \mathscr{D}_k, M_j), \\ &\text{solving for: } \hat{z}_k = \max_{z_k \in \mathcal{S}_1^J} \frac{1}{n_k} \sum_{i=1}^{n_k} \log \sum_{j=1}^J z_{k,j}\, p\left(y_{k,i} \mid \mathscr{D}_{k,[-l]}, M_j\right). \end{aligned} \tag{5}$$

- Global inference: To achieve inferences on $\mathscr{D}$, we stack local posterior distribution estimates between partitions

$$\begin{aligned} &\hat{p}(\cdot \mid \mathscr{D}) = \sum_{k=1}^K \hat{w}_k \hat{p}(\cdot \mid \mathscr{D}_k), \\ &\text{solving for: } \hat{w} = \max_{w \in \mathcal{S}_1^K} \frac{1}{n} \sum_{i=1}^n \log \sum_{k=1}^K w_k \sum_{j=1}^J \hat{z}_{k,j}\, p\left(y_{k,i} \mid \mathscr{D}_{k,[-l]}, M_j\right) \end{aligned} \tag{6}$$

5. Sea Surface Temperature data

Data application focus on Sea Surface Temperature (SST) collected in June 2022

- Trainin data: $n = 1,000,000$ locations spread all over the world's ocean surface,
- Test data: $n' = 2,500$ holdout locations to evaluate predictive performances,
- Source: National Oceanic and Atmospheric Administration (NOAA).

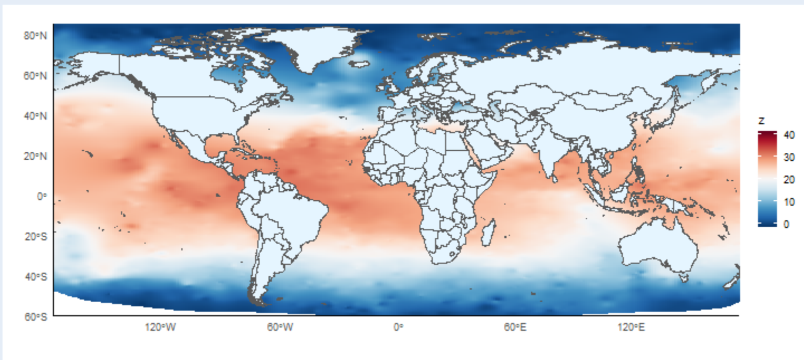

Figure: Holdout data surface interpolation for SST data analysis.

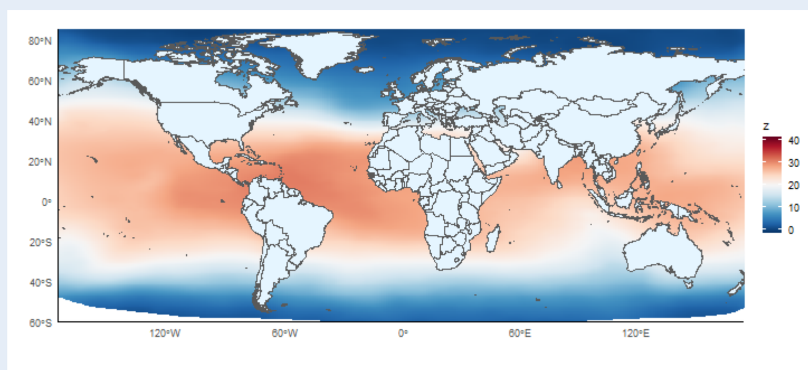

Figure: Predicted (MAP) surface interpolation for SST data analysis.

- Number of subsets: $K = 2,000$ random partitions (**500** locations each),
- Competetive models: $J = 3$ collections of values for $\{\delta^2, \phi\}$,
- Computing environment: Intel Core I7-**8750**H CPU with **5** physical cores.

- Timing: $\approx$ **63** minutes (1 hour) for full Bayesian analysis (including model fitting, posterior sampling, and prediction),
- Predictive performances: RMSPE $\approx$ **9** times lower w.r.t. Bayesian conjugate model.

	Conjugate linear model	Bayesian predictive stacking
β_0	17.890 (17.87, 17.91)	8.846 (3.382, 12.627)
β_{long}	0.010 (0.01, 0.02)	-0.040 (-0.347, 0.221)
β_{lat}	-0.540 (-0.550, -0.540)	-0.102 (-0.427, 0.412)
σ^2	96.470 (96.21, 96.73)	25.607 (18.032, 38.050)
RMSPE	9.855	1.165
Time	–	63

Table: Sea Surface Temperature data analysis parameter estimates, RMSPE, and computing time in minutes for candidate models. Parameter posterior summary 50 (2.5, 97.5) percentiles.

6. Statistical software

A novel optimized R package was created, named `spBPS`:

- Introduce the BPS framework for univariate, and multivariate, geostatistical modeling,
- Use Rcpp/C++ -based code, allowing faster and more scalable parallel computations,
- Available @lucapresicce/spBPS on GitHub (soon on CRAN).

7. References

Banerjee, S. (2017). High-dimensional Bayesian geostatistics. *Bayesian Analysis 12*(2), 583–614.

Banerjee, S. (2020). Modeling massive spatial datasets using a conjugate Bayesian linear modeling framework. *Spatial Statistics 37*, 100417. Frontiers in Spatial and Spatio-temporal Research.

Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association 102*(477), 359–378.

Guhaniyogi, R. and S. Banerjee (2018). Meta-kriging: Scalable Bayesian modeling and inference for massive spatial datasets. *Technometrics 60*(4), 430–444.

Yao, Y., A. Vehtari, D. Simpson, and A. Gelman (2018). Using stacking to average Bayesian predictive distributions (with discussion). *Bayesian Analysis 13*(3), 917–1007.

Zhang, L., S. Banerjee, and A. O. Finley (2021). High-dimensional multivariate geostatistics: A Bayesian matrix-normal approach. *Environmetrics 32*(4), e2675.

Zhang, L., W. Tang, and S. Banerjee (2023). Exact Bayesian geostatistics using predictive stacking. *arXiv preprint*, arXiv:2304.12414.