Matrix–free High Dimensional Gaussian Simulation

Somak Dutta

Iowa State University, Ames, IA

Collaborators



Debashis Mondal (WashU)



Subrata Pal (Mayo)



Hongyuan Wang (ISU)



Ani Pathak (ISU)

Table of Contents

- **Motivation**
- **Proposed algorithm**
- **Other spatial examples**
- **Applications**

High-dimensional Gaussian simulation

Example 1: Gaussian mixed linear models (Henderson, 1950; Lee and Nelder, 1996)

$$
\begin{aligned}
Y &= X\beta + F\psi + \epsilon \\
\beta &\sim N(0, R_1^{-1}) \\
\psi &\sim N(0, R_2^{-1}) \\
\epsilon &\sim N(0, (1/\lambda_1) I_n)
\end{aligned}
$$

Then

$$
\begin{pmatrix} \beta \\ \psi \end{pmatrix} \mid Y \sim N\left(A^{-1} b, A^{-1}\right),
$$

where

$$
b = \begin{pmatrix} \lambda_1 X^\top Y \\ \lambda_1 F^\top Y \end{pmatrix} \qquad A = \left[\lambda_1 \begin{pmatrix} X^\top X & X^\top F \\ F^\top X & F^\top F \end{pmatrix} + \begin{pmatrix} R_1 & 0 \\ 0 & R_2 \end{pmatrix} \right].
$$

Often X, F, R_1, R_2 are high-dimensional but sparse matrices.

High-dimensional Gaussian simulation

Example 2: Logistic regression using Pólya–Gamma augmentation (Polson et al., 2013)

$$Y_i \sim \text{Bernoulli}(\pi_i), \quad \text{logit}(\pi_i) = x_i^\top \beta + f_i^\top \psi,$$
$$\beta \sim N(0, R_1^{-1}) \quad \psi \sim N(0, R_2^{-1}).$$

Using Pólya–Gamma augmentation,

$$\omega_i \mid \beta, \psi, Y_i \sim PG(1, \eta_i), \qquad \eta_i = x_i^\top \beta + f_i^\top \psi.$$

Let $\Omega = \text{diag}(\omega_1, \ldots, \omega_n)$ and $\kappa_i = Y_i - 1/2$. Then

$$\begin{pmatrix} \beta \\ \psi \end{pmatrix} \mid Y, \omega \sim N\left(A^{-1} b, A^{-1}\right),$$

where

$$b = \begin{pmatrix} X^\top \kappa \\ F^\top \kappa \end{pmatrix} \qquad A = \left[\begin{pmatrix} X^\top \Omega X & X^\top \Omega F \\ F^\top \Omega X & F^\top \Omega F \end{pmatrix} + \begin{pmatrix} R_1 & 0 \\ 0 & R_2 \end{pmatrix} \right].$$

Table of Contents

- Motivation
- **Proposed algorithm**
- Other spatial examples
- Applications

Proposed algorithm

Multivariate statistics 1010 (Undergraduate course, 2006):

- If A_1 and A_2 both have n rows and v_1 and v_2 are standard multivariate Gaussian random vectors of appropriate dimensions then

$$\eta = b + A_1 v_1 + A_2 v_2 \sim N(b, A)$$

  where $A = A_1 A_1^\top + A_2 A_2^\top$ and
- Solution to $Ax = \eta$ follows $N(A^{-1} b, A^{-1})$

In many problems: A, A_1 and A_2 are matrix–free operators, that is, we can compute Av, $A_1 v_1$ and $A_2 v_2$ using

- at most $O(n)$ storage in RAM
- in less than $O(n \log n)$ computational complexity
- and solve $Ax = \eta$ in less than $O(n^3)$ computational complexity.

Conjugate Gradient algorithm for solving $Ax = \eta$

Fast forward to 2011, I learned about the conjugate gradient algorithm (Stiefel, 1951):

Set $x = x_0$ and $\phi = \eta - Ax$
For $i = 1, 2, \ldots$ until convergence

- $\rho_{.1} = \phi^\top v$
- $\beta = \rho_{.1} / \rho_{.2}$
- $u = v$ if $i = 1$, $v + \beta u$ otherwise.
- $\xi = Au$
- $\alpha = \rho_{.1} / (u^\top \xi)$
- $x = x + \alpha u$
- $\phi = \phi - \alpha \xi$
- $\rho_{.2} = \rho_{.1}$
- Test convergence: $\|\phi\| < \mathsf{tol}$

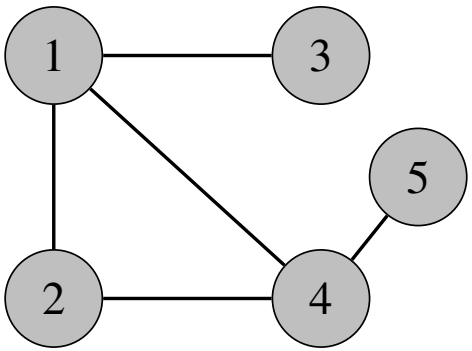
End for.

In practice, we solve $P^{-1}Ax = P^{-1}\eta$ where P is a preconditioner so that $P^{-1}\eta$ is easy to compute and $P^{-1}A$ has smaller condition number than A.

Example: Gaussian Markov random fields

$$Y = X\beta + F\psi + \epsilon$$

- Precision matrix R_2 of ψ with off-diagonal entries $-w_{ij}$.
- $w_{ij} = 0$ if i and j are not neighbors.
- Diagonal entries w_{ii}.



	$\{1,2\}$	$\{1,3\}$	$\{1,4\}$	$\{2,4\}$	$\{4,5\}$
1	$\sqrt{w_{12}}$	$\sqrt{w_{13}}$	$\sqrt{w_{14}}$		
2	$-\sqrt{w_{12}}$			$\sqrt{w_{24}}$	
3		$-\sqrt{w_{13}}$			
4			$-\sqrt{w_{14}}$	$-\sqrt{w_{24}}$	$\sqrt{w_{45}}$
5					$-\sqrt{w_{45}}$

Figure: Illustration of A_2 for a graph. Only the non-zero entries are shown in the matrix. $A_2 A_2^\top = R_2$.

Assuming $w_{ii} + \sum_{j \neq i} w_{ij} = 0$ (intrinsic case); there would be another diagonal matrix for the stationary case.

Table of Contents

- Motivation
- Proposed algorithm
- **Other spatial examples**
- Applications

First order (intrinsic) autoregressions on regular lattice

On an $r \times c$ lattice

$$R_2 = \alpha I_c \otimes W_r + (1-\alpha) W_c \otimes I_r$$

then

$$A_2 = \left[\sqrt{\alpha} I_c \otimes L_r \quad : \quad \sqrt{1-\alpha} L_c \otimes I_r \right]$$

where

$$W_k = \begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & \ddots & \ddots & \\ & & \ddots & 2 & -1 \\ & & & -1 & 1 \end{bmatrix}_{k \times k} \qquad L_k = \begin{bmatrix} 1 & & & & \\ -1 & 1 & & & \\ & -1 & \ddots & & \\ & & \ddots & 1 & \\ & & & -1 & 1 \end{bmatrix}$$

so that $L_k L_k^\top = W_k$.

Fractional Gaussian fields on a regular lattice

Fractional Gaussian fields (e.g., Matérn) arise as solutions to SPDE:

$$\left(\kappa - \alpha\frac{\partial^2}{\partial u^2} - (1-\alpha)\frac{\partial^2}{\partial v^2}\right)^{\nu/2} \psi(u,v) = \xi(u,v),$$

where $\nu \in (0,\infty)$, $\alpha \in (0,1)$, $\kappa \geq 0$; $\xi(u,v)$, for $(u,v) \in \mathbb{R}^2$ represents Gaussian white noise.
On an $r \times c$ lattice

$$R_2 = M^\top D M$$

- D is diagonal with analytically known entries (Dutta and Mondal, 2016).
- M is the two-dimensional discrete cosine transformation.
- Mv and $M^\top v$ are computed via FFT, without storing M, using $O(rc\log rc)$ FLOPs.
- $A_2 = M^\top D^{1/2}$ is a matrix–free root of R_2.

Table of Contents

- Motivation
- Proposed algorithm
- Other spatial examples
- **Applications**

A simulation study

1024×1024 grid: Data generated from exponential covariance model with 50% of the grid cells observed.

Table: Summary of computational times in the numerical experiments. Numbers are averaged over 10 replications.

Method	#CPU	Conditional simulation time (s)
GpGp (Guiness, 2018; Vecchia approximation)	32	106.8 / draw
NNGP (Datta et al., 2016; 15 nearest neighbors)	32	8035/1000 draws
Proposed (Dutta, 2015)	32	103.2 /1000 draws

Groundwater arsenic contamination in Bangladesh

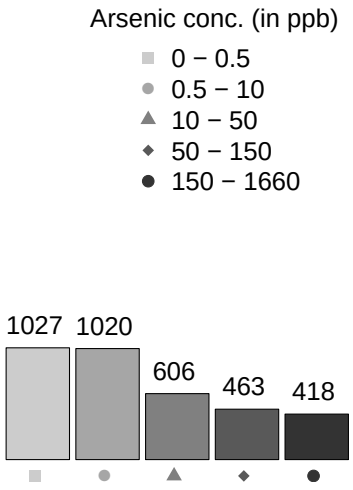


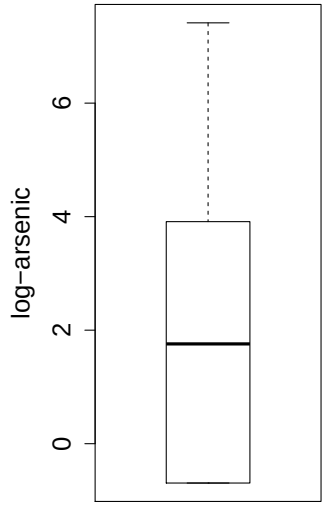
Source: Human Rights Watch.

Arsenic in 42 districts above the maximum limit of the WHO of **10 mg / L**.
More than **20 million people** are affected.
An estimated **43,000** people die each year from arsenic poisoning.

Groundwater arsenic contamination in Bangladesh

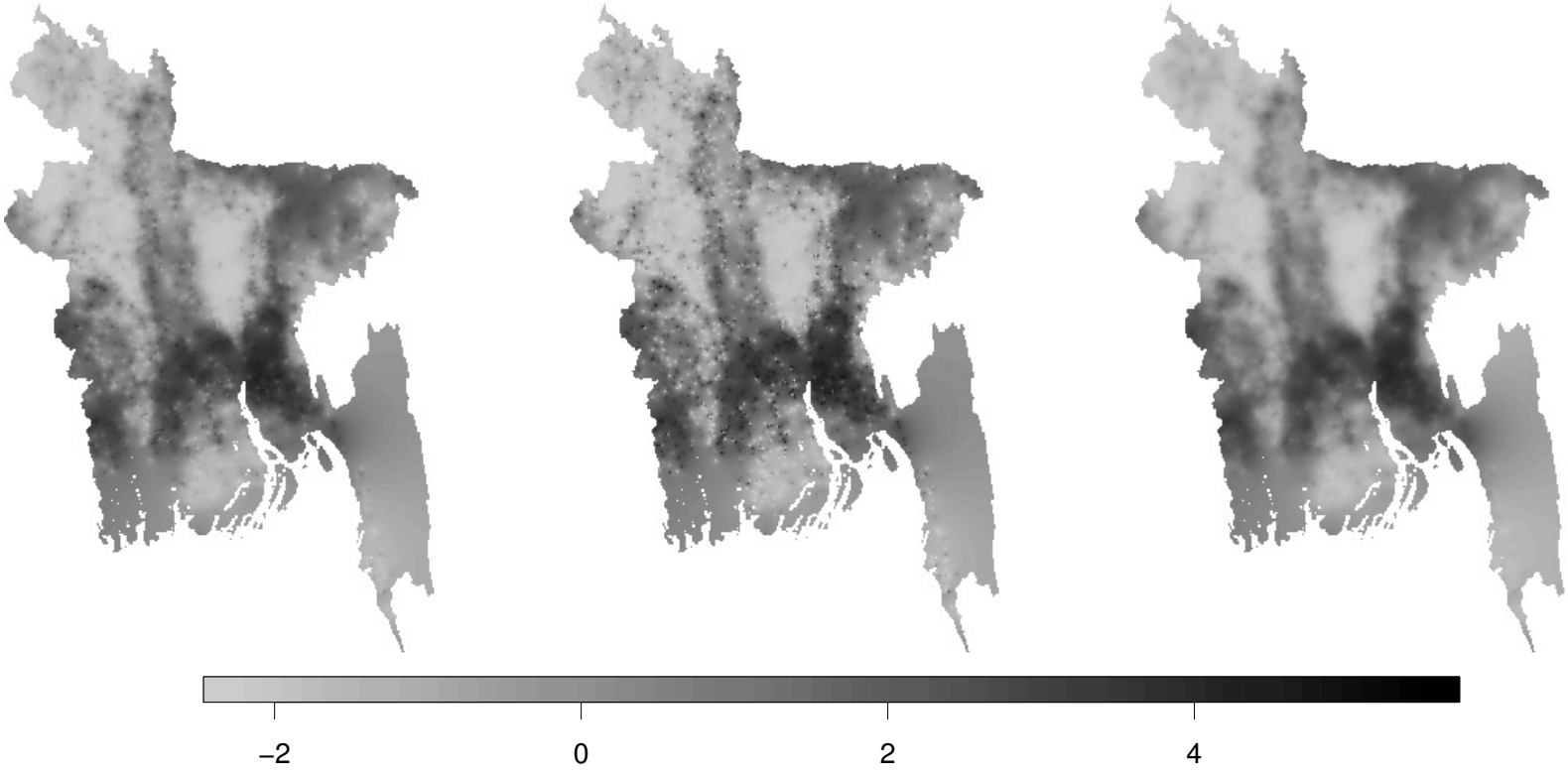






Data at 3500 locations from British Geological Survey, 1999–2000.

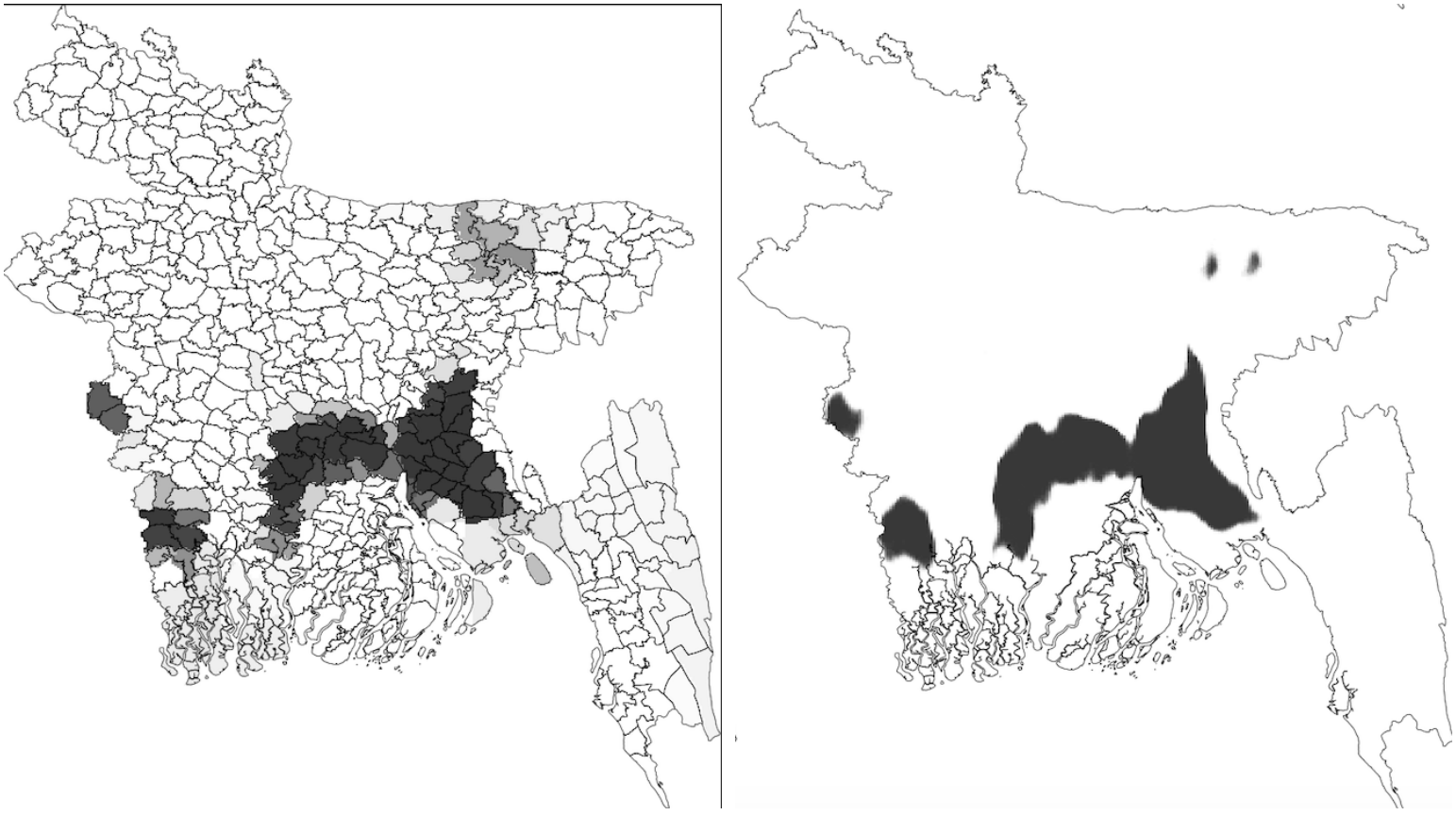
BLUPs for arsenic (log) concentration



Embed data on a 500×300 grid. Images are BLUPs for 3 different fractional Gaussian models: Left: $\nu = 1$, Center: $\nu = 0.858$ (no nugget), Right: REML estimate $\nu = 1.240$.

BLUPs are **starting points**; we also care about **uncertainty**.

Non-linear functions of random fields



Left: Median arsenic in sub-counties exceeding 50ppb.
Right: Maximal region with joint exceedance probability 90%.
Need **conditional simulation** to compute non-linear functions of random fields.

Summary

- Conditional simulation for a large class of important mixed models.
- The method is matrix–free: needs $O(n)$ RAM (read MegaBytes) and $O(n \log n)$ computation time (read minutes).
- Uses rectangular roots of precision matrices and preconditioned conjugate gradient algorithm.

Acknowledgments: USDA NIFA 2023-70412-41087 and Plant Science Institute, Iowa State University.

Key references

- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B*, 36, 192–225.
- Besag, J. and Green, P. J. (1993). Spatial statistics and Bayesian computation. *Journal of the Royal Statistical Society Series B: Statistical*, 55, 25–37.
- Dutta, S. (2015) Residual Likelihood Analysis for Spatial Mixed Linear Models. *Ph.D. Thesis*, University of Chicago.
- Henderson, C. R. (1950). Estimation of genetic parameters. Ann. Math. Statist., 21, 309—310.
- Lee, Y., Nelder, J. A. and Pawitan, Y. (2006). *Generalized Linear Models with Random Effects Unified Analysis via h–likelihood*. Boca Raton: Chapman and Hall.
- Polson, N. G., Scott, J. G., and Windle, J. (2013). Bayesian inference for logistic models using Pólya–Gamma latent variables. *Journal of the American statistical Association*, 108(504), 1339-1349.

Thank you!