# Hamiltonian Monte Carlo computation in spatial statistics

DEBASHIS MONDAL

*Department of Mathematics and Statistics,*

*Washington University in St Louis*

joint work with former PhD student Chunxiao Wang

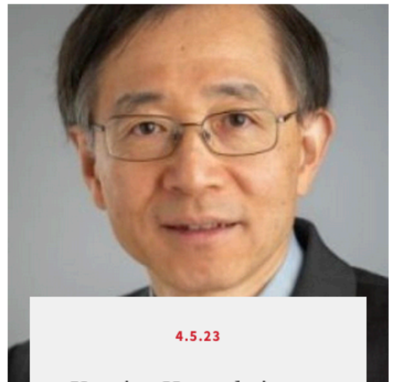
*supported by NSF award DMS 2153669*

ISNET-9, May 22 – 26, 2023

# Some news from Washington University

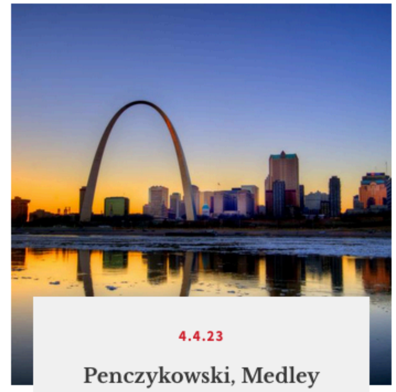
4.7.23

A cat's-eye view of one of the most beloved pets



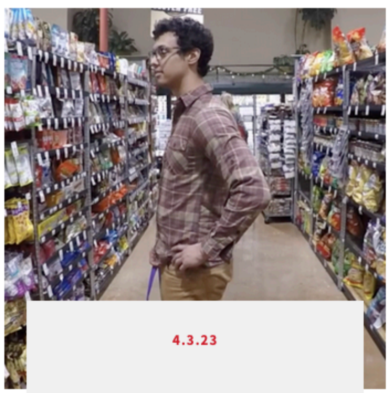
4.5.23

Xuming He to chair new Department of Statistics and Data Science



4.4.23

Penczykowski, Medley share seed grant to precisely measure St. Louis climate

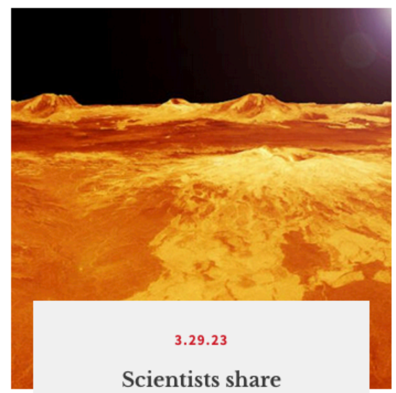


4.3.23

This is your brain on everyday life

3.29.23

WashU to host StanCon 2023



3.29.23

Scientists share 'comprehensive' map of volcanoes on Venus — all 85,000 of them

## Stan Conference 2023

StanCon 2023

**StanCon 2023**

CONVENTION ON STAN PROGRAMMING AND BAYESIAN MODELING

WASHINGTON UNIVERSITY IN ST. LOUIS · JUNE 20-23, 2023

Buy tickets for StanCon 2023

---

**Sponsors**

We thank our sponsors who both support conference costs, scholarships, and Stan as a whole. If you're interested in sponsoring StanCon, please email stancon2023@mc-stan.org.

Washington University in St. Louis

METRUM RESEARCH GROUP

NUMFOCUS OPEN CODE = BETTER SCIENCE

## Stan Conference 2023

### Tutorials

*Listed below are confirmed tutorials. Proposals for tutorials are reviewed and accepted on a rolling basis throughout April 30th.*

**Fundamentals of Stan**
**Instructor:** Charles Margossian (Flatiron Institute). *This course serves as an introduction to Stan and may be used as a stepping stone before taking more advanced tutorials.* Course description.

**Introduction to Bayesian hierarchical modeling using Stan and brms**
**Instructor:** Mitzi Morris (Columbia University) and Mike Lawrence (Axem Neurotechnology)

**Ordinary differential equation (ODE) models in Stan**
**Instructor:** Daniel Lee.

**Cognitive diagnostic models in R and Stan**
**Instructor:** Jake Thompson (University of Kansas). Course description.

**Advances of model assessment, selection, and inference after model selection**
**Instructor:** Andrew Johnson (Aalto University)

# Stan Conference 2023

## Scholarships

The purpose of the StanCon scholarship is to make StanCon a more accessible and inclusive event.

Participants who require financial assistance to attend the conference may apply for a scholarship by filling out **this form**. **The StanCon scholarship covers registration for the tutorial and the main conference, as well as local lodging**. Scholarships are awarded on a need-base, and prioritize early career scientists, including students and post-docs, and members of underrepresented groups in STEM.

Applications are reviewed on a rolling basis, and scholarships are awarded based on available funds.

## Organizers

- Charles Margossian (Flatiron Institute)
- Debashis Mondal (Washington University in St. Louis)
- Eric Ward (NOAA & University of Washington)
- Vianey Leos Barajas (University of Toronto)
- Yi Zhang (Sage Therapeutics, Inc)

## Outline for the talk

- An **overview**
- Spatial models that
  - combine **Markov random fields** and **geostatistics**
  - give rise to **scalable, matrix-free** computation
- Hamiltonian Monte Carlo sampling
  - Inverse **mass matrix** calculations
  - Leapfrog integration
- Further **challenges** and future directions

## Overview

Spatial variables are often observed indirectly, via treatments, covariates, blur, noise, ...

$$
\begin{aligned}
\text{Data } y &= \text{modeled as response to linear predictor } \eta \\
\eta &= Z\beta + Fx + \epsilon \\
\beta &= \text{treatment / variety / covariate effects} \\
Z &= \text{design matrix (covariate information)} \\
x &= \text{random spatial effects} \\
F &= \text{linear operator (typically an identity/ incidence/ averaging matrix)} \\
\epsilon &= \text{residual effects}
\end{aligned}
$$

Gaussian priors on $\beta$ and $\epsilon$.

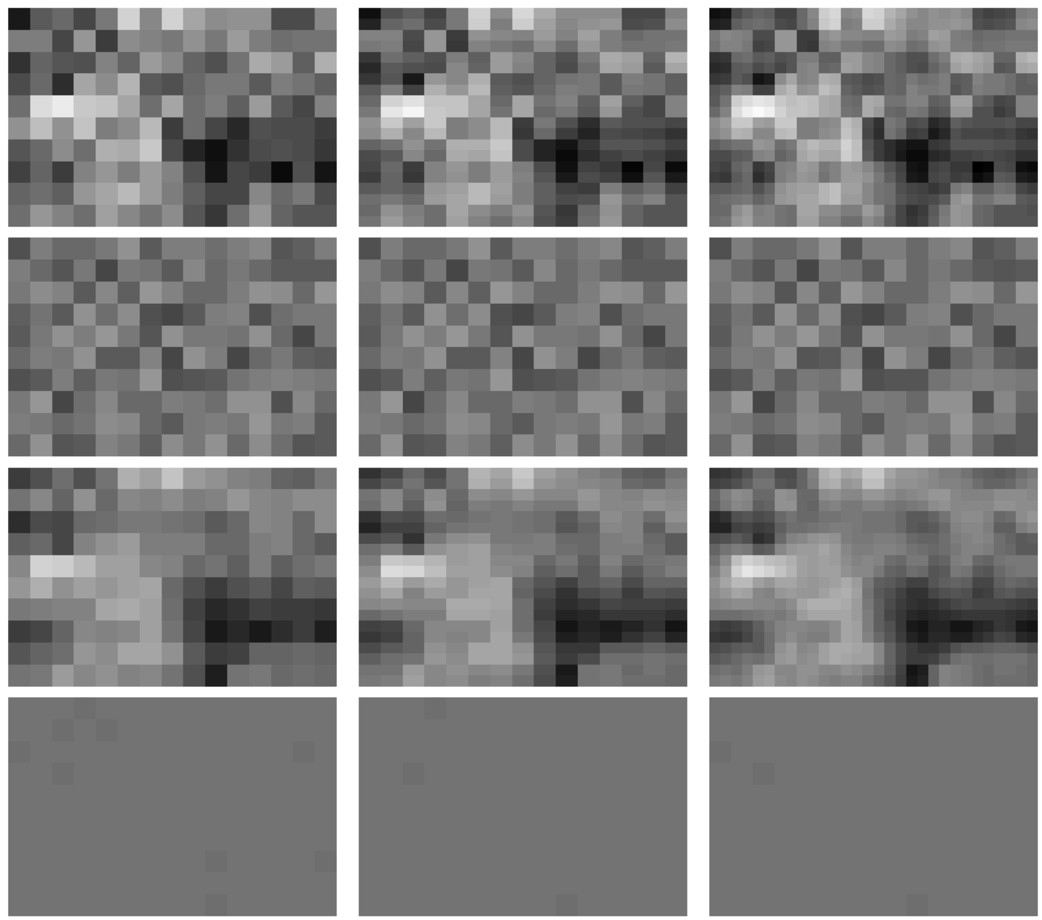
Usually, goal is to make probabilistic inferences about $\beta$ and $x$ (MCMC or ...).

Stochastic representation of $x$ via geostatistical or MRF approach.

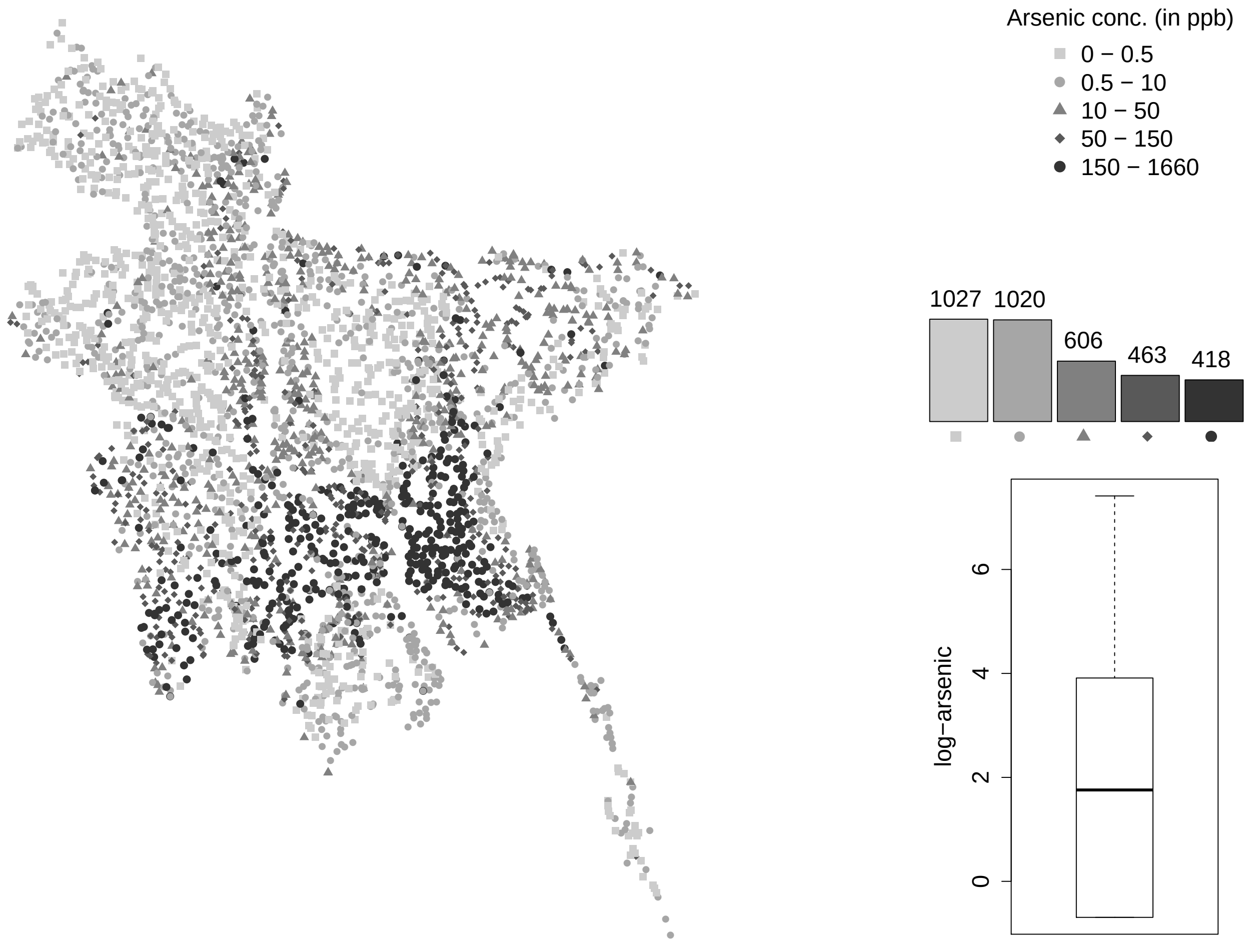
# Agricultural variety trials


Goal is to pick best few varieties, but need to take account spatial effects.

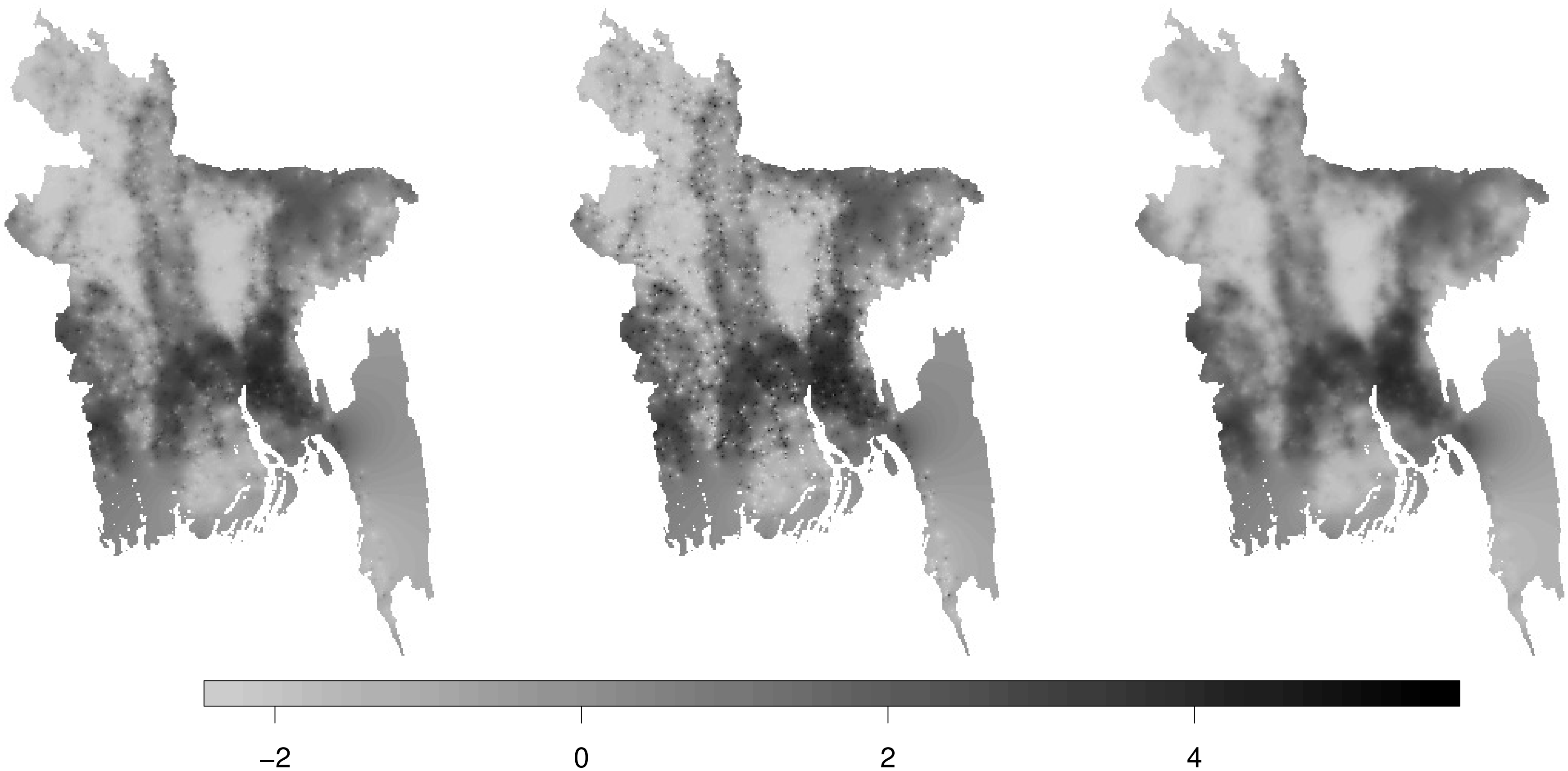
## Bayesian spatial analysis: effect of scale



# Groundwater arsenic contamination in Bangladesh



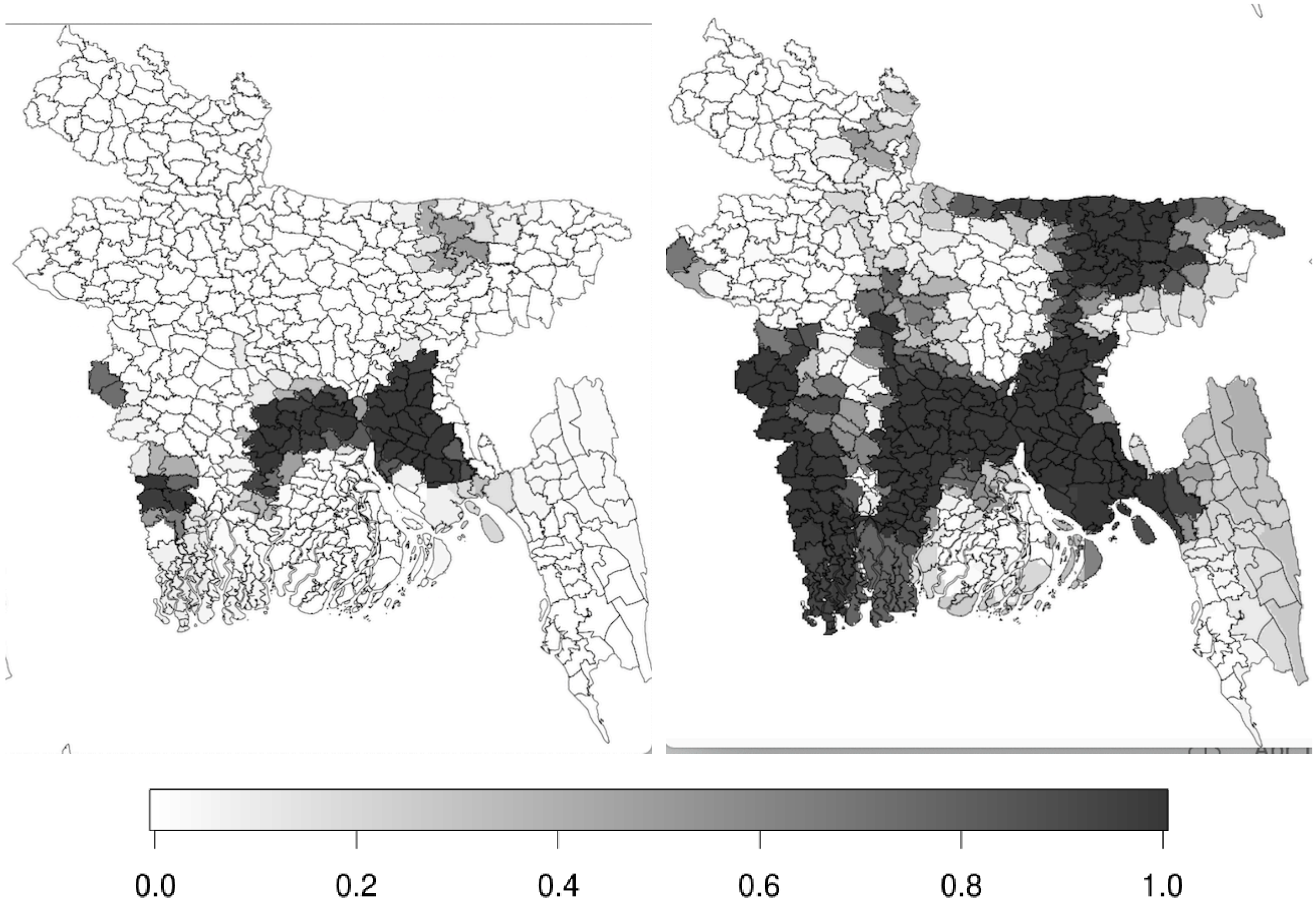
## Predictions for log arsenic contamination in Bangladesh




About 3000 observations. We embedded the data on a $500 \times 300$ grid.

Predictions from 3 different models.

## Sub-district-wise predictions




Left: Pr( average As level > 50ppb | data); right: Pr( max As level > 50ppb | data)

## **Disease mapping** (Besag et al., 1991, Rue and Held, 2005, Mondal 2023)

For any region $A$,

$$y(A) = \text{ conditionally Poisson } \big(e(A)R(A)\big), \qquad \sum_A e(A) = \sum_A y(A),$$

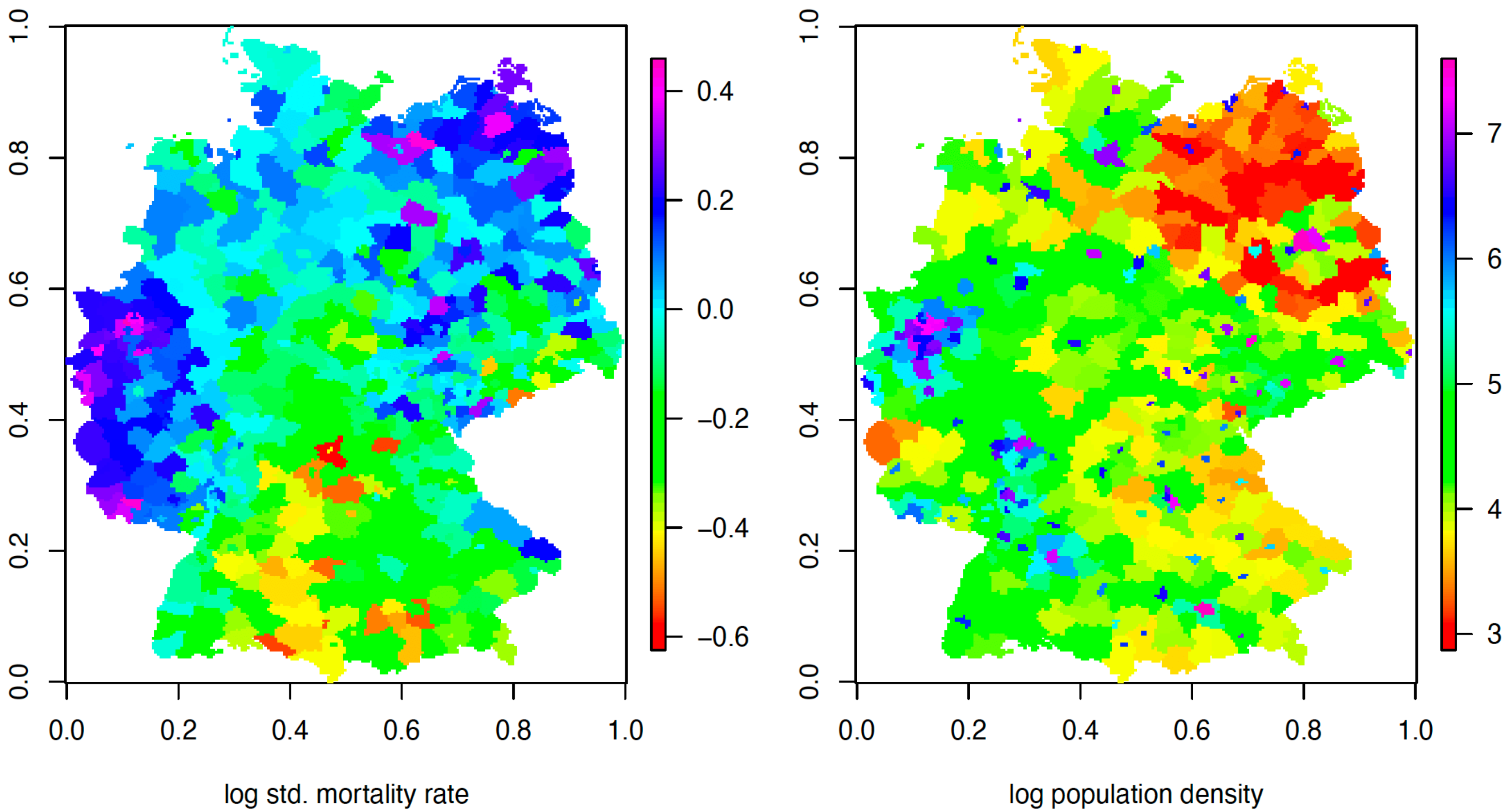
where

$$R(A) = \sum_{(u,v)\in A} f(u,v)e^{Z(u,v)\beta + x_{u,v} + \epsilon_{u,v}}. \tag{1}$$

$$\begin{aligned}
y(A) &= \# \text{ of cases of disease in region (e.g., county) } A \\
e(A) &= \text{expected number of cases of disease in region } A \\
R(A) &= \text{relative risk} \\
x_{u,v} &= q = q_1 \times q_2 \text{ spatial random effects on a very fine grid} \\
\epsilon &= \text{residual effects} \\
f(u,v) &= \text{population density,}
\end{aligned}$$

Goal is to predict spatial risk.

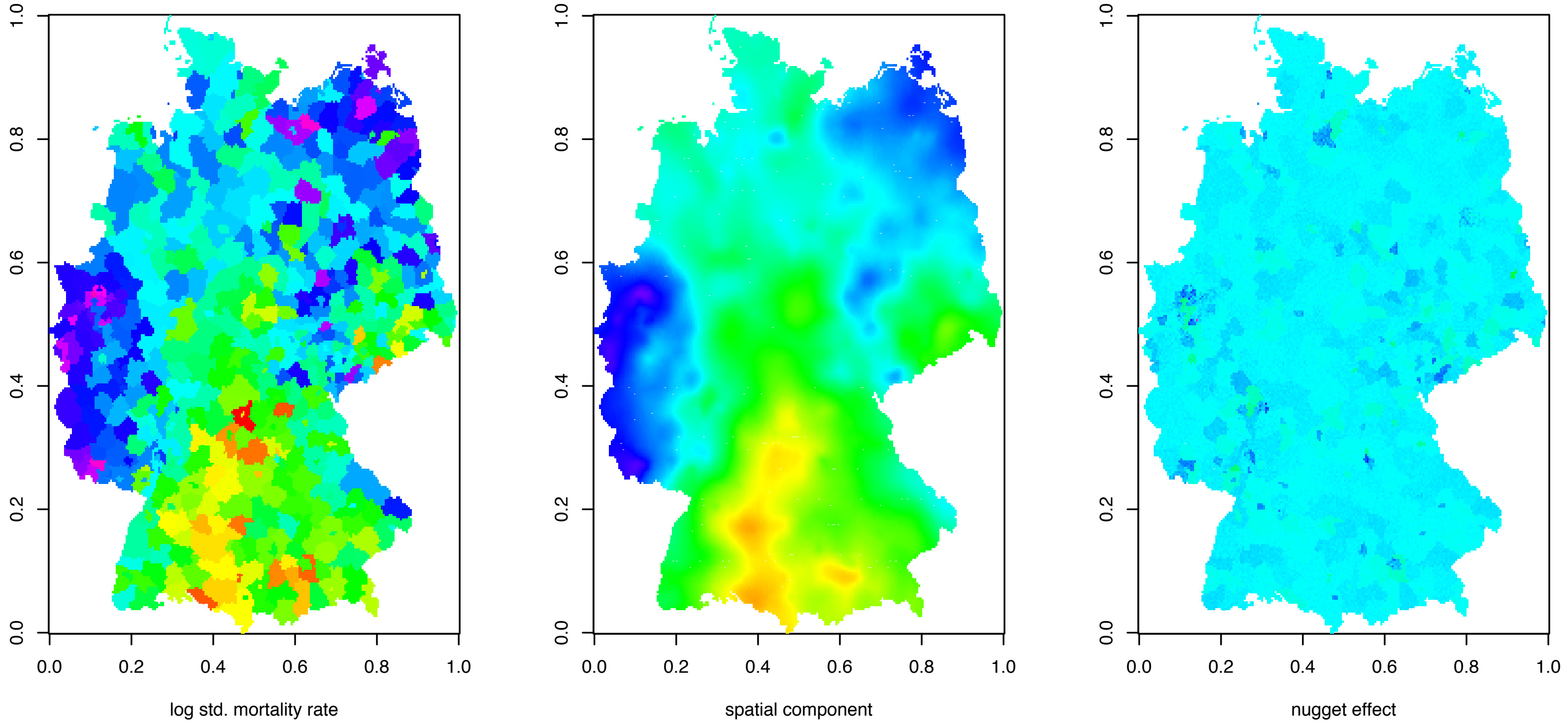
## Application for German cancer data




Lung Cancer data in the period 1986-1990. 544 districts.

**log standardized mortality ratios** (left) and **log population density** (right).

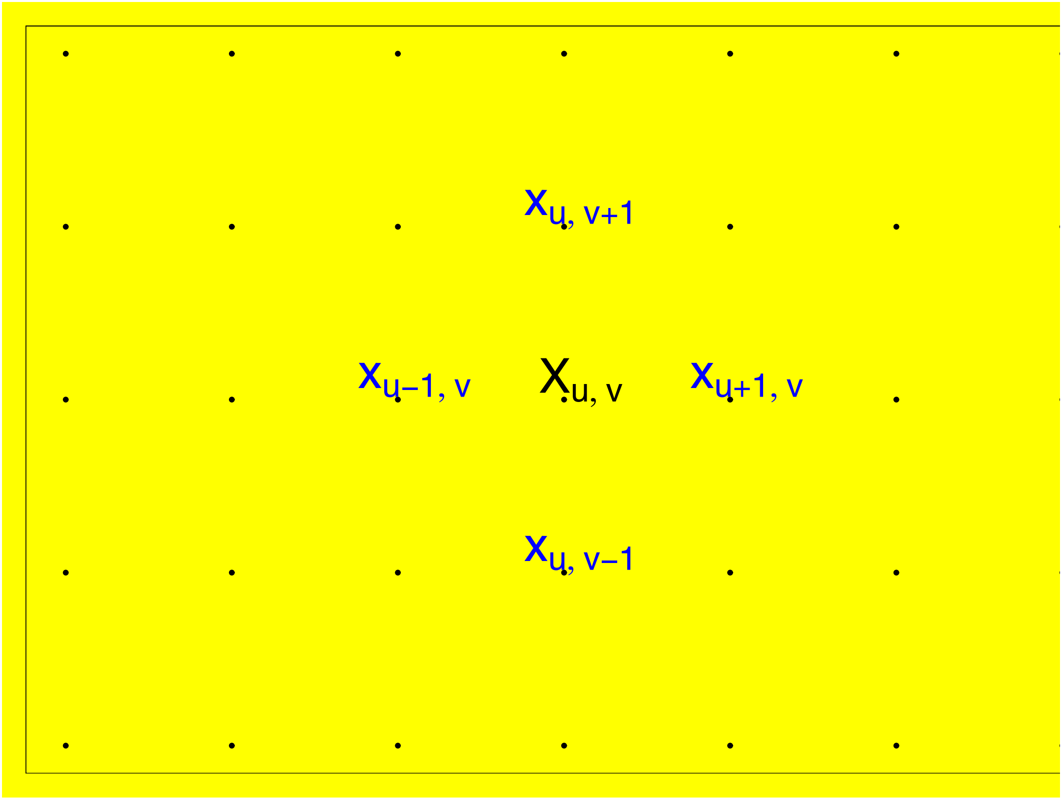
# Lung cancer in Germany




Data in the period 1986-1990. 544 districts embedded in 289 $\times$ 214 girds.

## Conditional auto-regressions on regular lattice

... has an extensive literature following Besag, Künsch, Cressie, and others

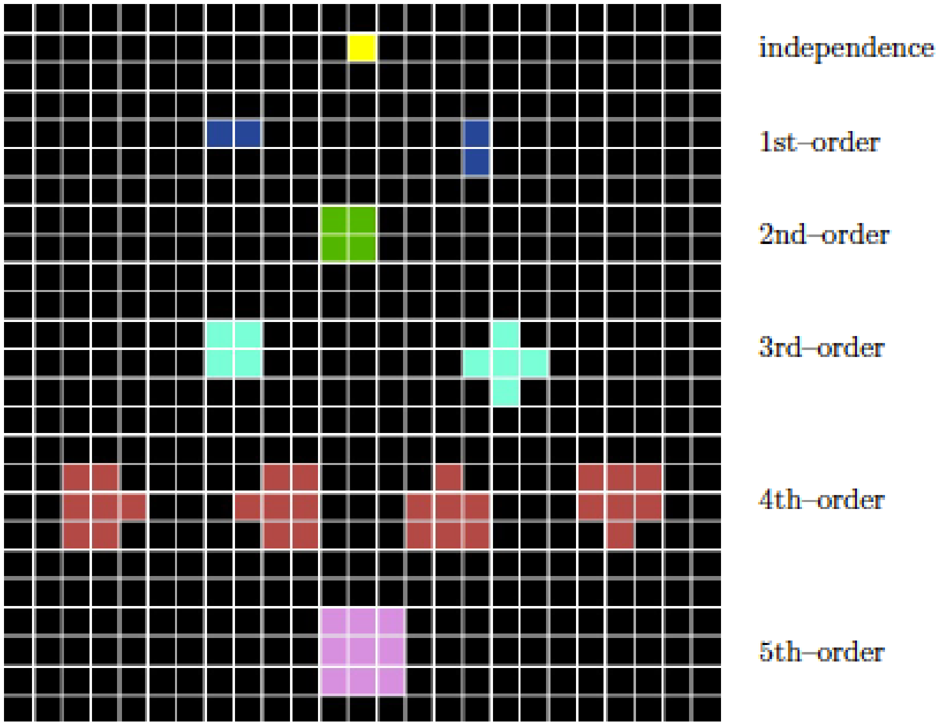



In the simplest case:

$$\mathrm{E}\,(x_{u,v} \mid \ldots) = \gamma_{10}\,(x_{u-1,v} + x_{u+1,v}) + \gamma_{01}\,(x_{u,v-1} + x_{u,v+1}), \quad \mathrm{var}\,(x_{u,v} \mid \ldots) = \sigma^2.$$

Conditional on other values, $x_{u,v}$ is **Gaussian** and $\gamma_{01}, \gamma_{10} \geq 0$ & $\gamma_{01} + \gamma_{10} < \frac{1}{2}$.

## Conditional auto-regressions on regular lattice



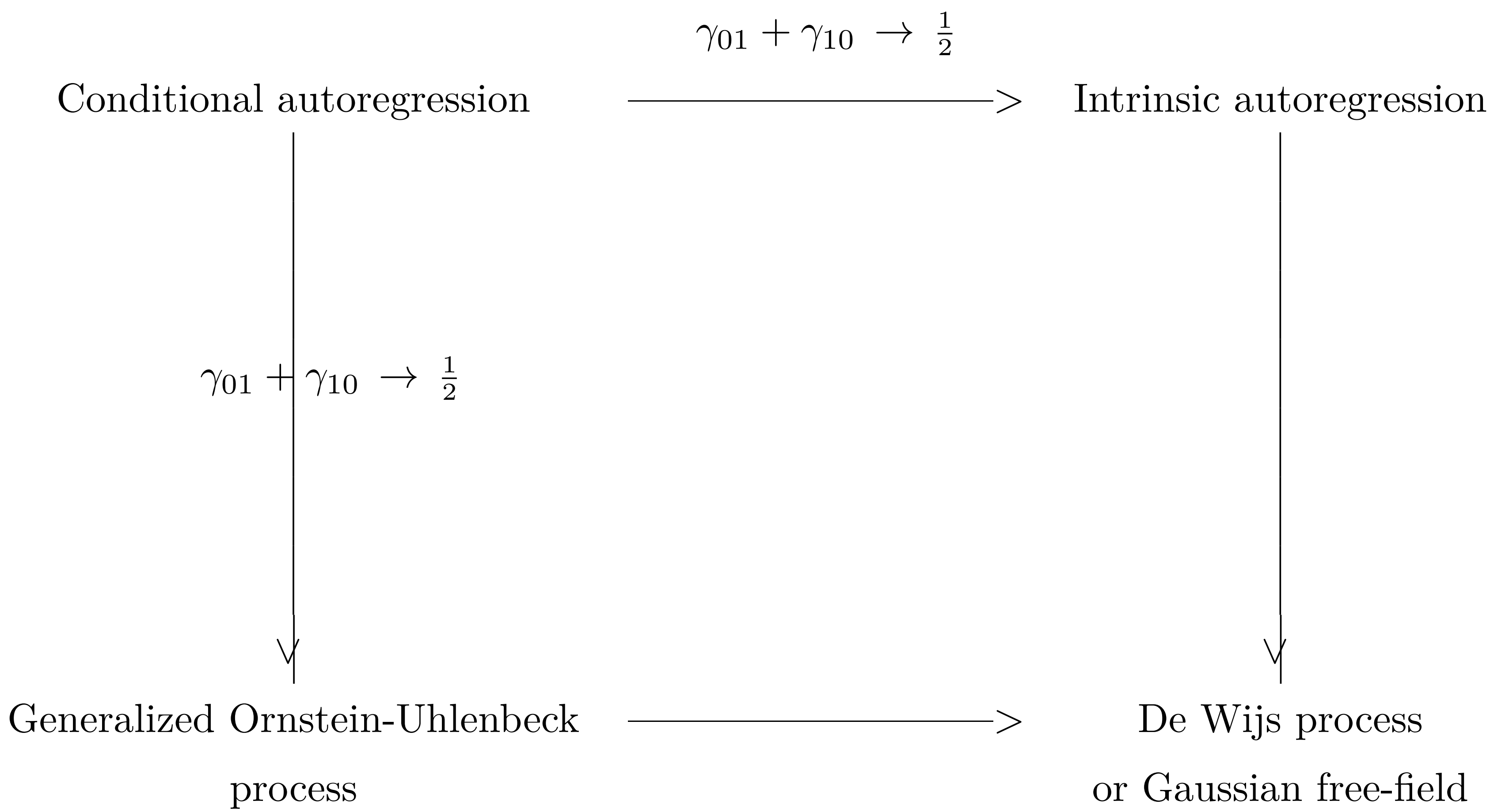

Extends to higher-neighborhood-order models with

$$\mathrm{E}\,(x_{u,v}\mid\ldots)\;=\;\sum_{k,l}\gamma_{k,l}\,x_{u-k,v-l},\quad \mathrm{var}\,(x_{u,v}\mid\ldots)\;=\;\sigma^2,$$

$\gamma_0 = 0$, $\gamma_{k,l} = \gamma_{-k,-l}$, and $\sum_{k,l}\gamma_{k,l}\cos(\omega_1 k + \omega_2 l) < 1$, $\omega_1, \omega_2 \in (-\pi, \pi]$.

- In general, density factorized over the cliques (or complete subgraphs) of the graph.

# Geostatistical limits of conditional autoregressions




- Similar limit diagram for higher-neighborhood order models.

## Spatial models on finite arrays

- Intrinsic auto-regressions typically have a distribution of the form

$$|W|_+^{\frac{1}{2}} \exp\left\{-\tfrac{1}{2} x^T W x\right\}$$

  where $W$ is proportional to the graph-Laplacian matrix.

- On a finite rectangular array, $W$ has a spectral decomposition

$$W = M \Lambda M^T.$$

  $M$ is the two dimensional discrete cosine transformation (DCT).

- To approximate fractional Gaussian fields, replace $W$ by $W^\alpha$, $\alpha > 0$.

- For other higher neighborhood-order models, replace $W$ by $\sigma(W)$, $\sigma()$ a positive polynomial/ function

## Hamiltonian Monte Carlo sampling

To sample from posterior $\pi(\theta \mid x)$, set up the Hamiltonian system as follows ...

- Consider a fictitious particle. Take

$$U(\theta) = -\log \pi(\theta \mid x)$$

as its potential energy at position $\theta$ as define its kinetic energy by

$$K(p) = \tfrac{1}{2} p^T M^{-1} p.$$

- Define total energy or the Hamiltonian as

$$H(\theta, p) = U(\theta) + K(p).$$

- The dynamical system then follows the differential equations

$$\begin{aligned} \frac{d\theta}{dt} &= \frac{\partial H}{\partial p} = M^{-1} p, \\ \frac{dp}{dt} &= -\frac{\partial H}{\partial \theta} = -\frac{\partial U(\theta)}{\partial \theta} = \frac{\nabla \pi(\theta \mid x)}{\pi(\theta \mid x)}. \end{aligned}$$

## HMC with leapfrog approximations

Typically, Hamiltonian dynamics have no analytic solutions ...

- Leapfrog approximation for time step-size $\delta$ gives

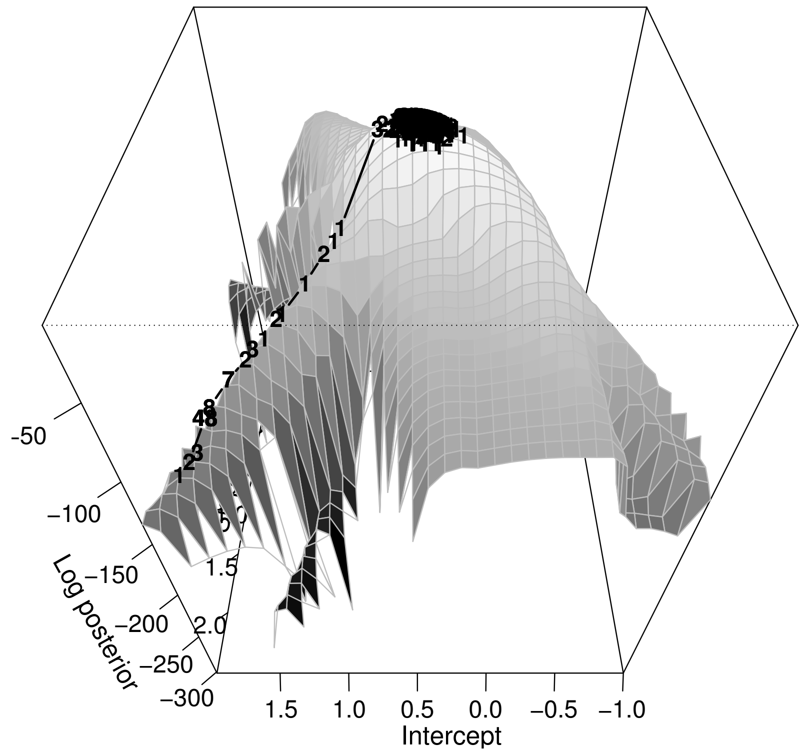
$$
\begin{aligned}
p(t+\tfrac{1}{2}\delta) &= p(t) - \tfrac{1}{2}\delta \frac{\partial U(\theta(t))}{\partial \theta}, \\
\theta(t+\delta) &= \theta(t) + \delta\, M^{-1} p(t+\tfrac{1}{2}\delta), \\
p(t+\delta) &= p(t+\tfrac{1}{2}\delta) - \tfrac{1}{2}\delta \frac{\partial U(\theta(t+\delta))}{\partial \theta}.
\end{aligned}
$$

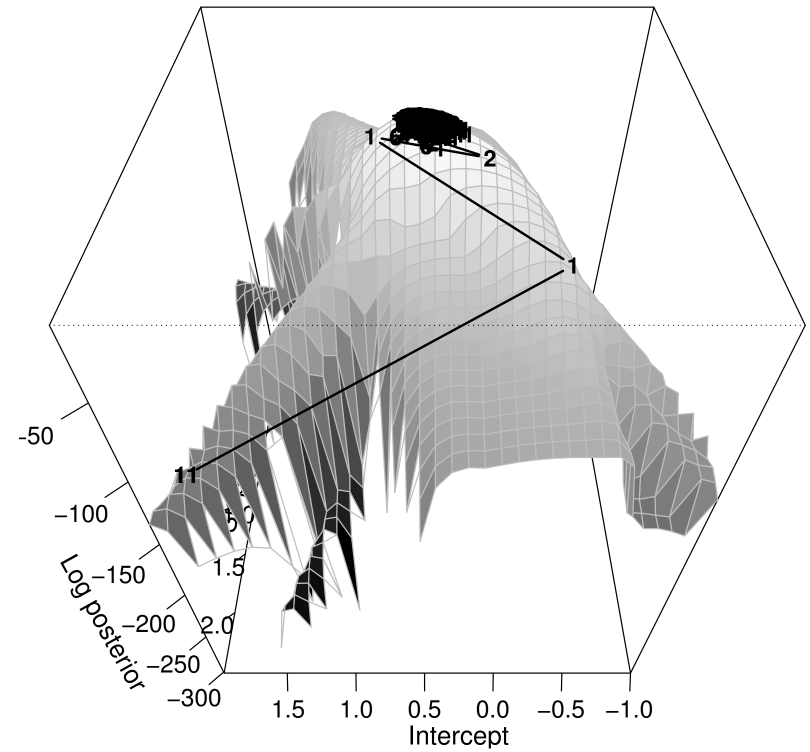
Hamiltonian Monte Carlo proceeds as follows:

1. At each iteration momentum $p$ is proposed with a $N(0, M)$
2. Leapfrog is used to move $T$ step forward and get $(\theta^*, p^*)$
3. The proposed state $(\theta^*, p^*)$ is then accepted with probability

$$\min\Big\{1, \exp(-H(\theta^*, p^*) + H(\theta, p))\Big\}.$$

# MCMC from Hamiltonian dynamics




HMC path can climb or jump a ridge ...

## HMC for spatial logistic regressions

Consider

$$\Pr\Big(y_i = 1\Big) = p_i, \qquad \eta_i = \log\{p_i/(1-p_i)\} = z_i\beta + x_i + \epsilon_i, \qquad i = 1, \ldots, n$$

It follows that

$$\pi(\theta|y,\lambda) \propto \Big[\prod_{i=1}^{n} \frac{\exp\{(z_i\beta + x_i + \epsilon_i)y_i\}}{1+\exp(z_i\beta + x_i + \epsilon_i)}\Big] \exp\left(-\lambda_3\beta^T Z^T Z\beta/2 - \lambda_2 x^T W x/2 - \lambda_1 \epsilon^T \epsilon/2\right),$$

The inverse mass matrix takes the form

$$M^{-1} = \begin{pmatrix} Z^T DZ + \lambda_3 Z^T Z & Z^T D & Z^T D \\ DZ & D + \lambda_2 W & D \\ DZ & D & D + \lambda_1 I \end{pmatrix},$$

where $D$ is a diagonal matrix with

$$D_{ii} = E[\exp\left(z_i\beta + x_i + \epsilon_i\right)/\{1 + \exp\left(z_i\beta + x_i + \epsilon_i\right)\}^2].$$

Let

$$\sigma_i^2 = \mathrm{var}\left(z_i\beta + x_i + \epsilon_i\right)$$

## HMC for spatial logistic regressions

$\sigma_i^2$ $i$th diagonal element of $\lambda_3^{-1}Z(Z'Z)^{-1}Z' + \lambda_2^{-1}W^{-1} + \lambda_1^{-1}I$. Since

$$\frac{e^z}{1+e^z} \approx \sum_{j=1}^{k} q_{k,j}\Phi(zs_{k,j}), \quad \frac{e^z}{(1+e^z)^2} \approx \sum_{j=1}^{k} q_{k,j}s_{k,j}\phi(zs_{k,j}), \quad D_{ii} \approx \sum_{j=1}^{k} \frac{\sigma_i q_{k,j} s_{k,j}}{\sqrt{2\pi(1+\sigma_i^2 s_{k,j}^2)}}.$$

| $q_{8,j}$ | $s_{8,j}$ |
|---|---|
| 0.003246343272134 | 1.365340806296348 |
| 0.051517477033972 | 1.059523971016916 |
| 0.195077912673858 | 0.830791313765644 |
| 0.315569823632818 | 0.650732166639391 |
| 0.274149576158423 | 0.508135425366489 |
| 0.131076880695470 | 0.396313345166341 |
| 0.027912418727972 | 0.308904252267995 |
| 0.001449567805354 | 0.238212616409306 |

## HMC for spatial Poisson regressions

Consider

$$y_i \sim Poisson(\mu_i), \qquad \eta_i = \log\{\mu_i\} = z_i\beta + x_i + \epsilon_i, \qquad i = 1, \ldots, n$$

It follows that

$$\pi(\theta|y,\lambda) \qquad \propto \Big[\prod_{i=1}^{n} \exp\{(z_i\beta + x_i + \epsilon_i)y_i - \exp\left(z_i\beta + x_i + \epsilon_i\right)\}\Big]$$
$$\exp\left(-\lambda_3\beta^T Z^T Z\beta/2 - \lambda_2 x^T W x/2 - \lambda_1 \epsilon^T \epsilon/2\right),$$

The inverse mass matrix again takes the form

$$M^{-1} = \begin{pmatrix} Z^T D Z + \lambda_3 Z^T Z & Z^T D & Z^T D \\ DZ & D + \lambda_2 W & D \\ DZ & D & D + \lambda_1 I \end{pmatrix},$$

where $D$ is a diagonal matrix with

$$D_{ii} = E[\exp\left(z_i\beta + x_i + \epsilon_i\right)] = \exp\Big(\sigma_i^2/2\Big), \qquad \sigma_i^2 = \mathrm{var}\left(z_i\beta + x_i + \epsilon_i\right).$$

## HMC for spatial GLMM

Consider

$$y_i \sim \exp[\{y_i\nu_i - b(\nu_i)\}/a_i(\phi) + c_i(y_i,\phi)], \qquad \nu_i = z_i^T\beta + x_i + \epsilon_i \qquad i = 1,\cdots,n,$$

It follows that

$$\pi(\theta \mid y,\lambda) \propto \prod_{i=1}^{n} \exp[\{y_i\nu_i - b(\nu_i)\}/a(\phi) + c(y_i,\phi)]\exp\{-\tfrac{1}{2}\lambda_1\beta^T Z^T Z\beta - \tfrac{1}{2}x^T W x - \tfrac{1}{2}\lambda_3\epsilon^T\epsilon\}.$$

The inverse mass matrix again takes the form

$$M^{-1} = -E_{\theta|\cdot}[\partial^2\log\{f(\theta\mid\cdot)\}/\partial\theta^2] = \begin{pmatrix} Z^T DZ + \lambda_3 Z^T Z & Z^T D & Z^T D \\ DZ & D + \lambda_2 W & D \\ DZ & D & D + \lambda_1 I \end{pmatrix},$$

where $D$ is a diagonal matrix with

$$D_{ii} = E\Big(b''(\nu_i)/a_i(\phi)\Big) = h(\sigma_i^2), \qquad \sigma_i^2 = \mathrm{var}\,(z_i\beta + x_i + \epsilon_i).$$

For non-canonical link, the formula is slightly complicated ...

## HMC for spatial GLMM

Consider

$$y_i \sim \exp[\{y_i\nu_i - b(\nu_i)\}/a_i(\phi) + c_i(y_i,\phi)], \qquad \nu_i = z_i^T\beta + f_i^T x_i + \epsilon_i \qquad i = 1,\cdots,n,$$

It follows that

$$\pi(\theta \mid y, \lambda) \propto \prod_{i=1}^{n} \exp[\{y_i\nu_i - b(\nu_i)\}/a(\phi) + c(y_i,\phi)] \exp\{-\tfrac{1}{2}\lambda_1\beta^T Z^T Z\beta - \tfrac{1}{2}x^T W x - \tfrac{1}{2}\lambda_3\epsilon^T\epsilon\}.$$

The inverse mass matrix again takes the form

$$M^{-1} = -E_{\theta|\cdot}[\partial^2 \log\{f(\theta \mid \cdot)\}/\partial\theta^2] = \begin{pmatrix} Z^T DZ + \lambda_3 Z^T Z & Z^T DF & Z^T D \\ F^T DZ & F^T DF + \lambda_2 W & F^T D \\ DZ & DF & D + \lambda_1 I \end{pmatrix},$$

where $D$ is a diagonal matrix with

$$D_{ii} = E\Big(b''(\nu_i)/a_i(\phi)\Big) = h(\sigma_i^2), \qquad \sigma_i^2 = \mathrm{var}\,(z_i\beta + f_i^T x_i + \epsilon_i).$$

Slightly different formula non-canonical link. Computation can be challenging!

## Computing the diagonal of the inverse

- If $A$ is $n \times n$, sparse and positive-definite, consider

$$A = LDU$$

and

$$A^{-1} = U^{-1}D^{-1} + A^{-1}(I - L), \qquad A^{-1} = D^{-1}L^{-1} + (I - U)A^{-1}.$$

Enable us to obtain entries of $A^{-1}$ belonging to the sparsity pattern of $L$ and $U$.

- Order of complexity $O(n^{3/2})$
- $A$ singular $\Rightarrow$ order of complexity increases. In general, order of complexity is $O(n^3)$
- Probing or other approximations does not work that well.
- *Theorem* (Mondal, 2023) : For rectangular arrays, can compute diagonals of $W^{-1}$ or $\sigma(W)^{-1}$ in $O(n \log(n))$ steps using two dimensional DCT.

## Further challenges and future directions

- Scalable computation for the diagonal of inverse?
- HMC steps for dispersion parameters?
- Other tuning parameters?
- Mixing rate?
- Mass matrix from other complicated models?
- spatial analysis with and empirical likelihood (EL) methods

I am particularly interested in spatial statistics and EL methods ...

# References


Besag, J., York, J. and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. Annals of the institute of statistical mathematics, 43, 1-20.

Besag, J. and Higdon, D. (1999). Bayesian analysis of agricultural field experiments. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 61(4), 691-746.

Besag, J. and Mondal, D. (2005). First-order intrinsic autoregressions and the de Wijs process. Biometrika, 92(4), 909-920.

Chaudhuri, S., Mondal, D. and Yin, T. (2017). Hamiltonian Monte Carlo sampling in Bayesian empirical likelihood computation. Journal of the Royal Statistical Society. Series B (Statistical Methodology), 293-320.

Dutta, S. and Mondal, D. (2015). An h-likelihood method for spatial mixed linear models based on intrinsic auto-regressions. Journal of the Royal Statistical Society: Series B: Statistical Methodology, 699-726.

Erisman, A. M. and Tinney, W. F. (1975). On computing certain elements of the inverse of a sparse matrix. Communications of the ACM, 18(3), 177-179.

Girolami, M. and Calderhead, B. (2011). Riemann manifold langevin and hamiltonian monte carlo methods. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73(2), 123-214.

Mondal, D. (2018). On edge correction of conditional and intrinsic autoregressions. Biometrika, 105(2), 447-454.

Mondal (2023). Generalized Gaussian random fields and disease mapping. Under revision

Mondal and Wang (2023). Hamiltonian Monte Carlo computation in spatial statistics. In preparation.

Neal, R. (2011). MCMC for using Hamiltonian dynamics. In Handbook of Markov Chain Monte Carlo, pp. 113–162. Taylor & Francis.

Rue, H. and Held, L. (2005). Gaussian Markov random fields: theory and applications. CRC press.