Hierarchical Stochastic Partial Differential Equations for Bayesian Inverse Problems

Lassi Roininen
University of Oulu, Finland
Karla Monterrubio-Gómez, Sari Lasanen, Alberto Mendoza

SIAM UQ 2018

Hierarchical models

Results presented are in:

- Lassi Roininen, Mark Girolami, Sari Lasanen and Markku Markkanen, Hyperpriors for Matérn fields with applications in Bayesian inversion, ArXiv 2016
- Karla Monterrubio-Gómez, Lassi Roininen, Sara Wade, Theo Damoulas and Mark Girolami, Posterior Inference for Sparse Hierarchical Non-stationary Models, ArXiv 2018

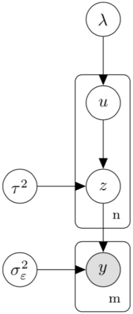
These studies are closely linked to eg:

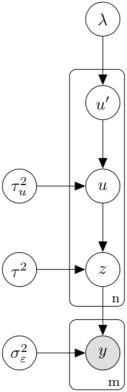
- Finn Lindgren, Håvard Rue and Johan Lindström, An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach, J. R. Statist. Soc. B 2011
- Matt Dunlop, Mark Girolami, Andrew Stuart and Aretha Teckentrup, How deep are deep Gaussian processes?, ArXiv 2017

Plate diagram for a non-stationary hierarchical model

3-level GP model

2-level GP model





y: observed data
σ_ε^2: noise variance
z: non-stationary process
τ^2: variance of z
u: length-scale process
τ_u^2: variance of u
u': length-scale process
λ: length-scale

Hierarchical model

- Hierarchical formulation for a spatial interpolation problem

$$\begin{aligned} y_i &\sim \mathcal{N}(z(x_i), \sigma_\varepsilon^2), \quad i = 1, \ldots, m, \\ z(\cdot) &\sim \mathcal{GP}\left(0, C_\phi^{\mathrm{NS}}(\cdot,\cdot)\right), \\ u(\cdot) := \log \ell(\cdot) &\sim \mathcal{GP}\left(0, C_\varphi^{\mathrm{S}}(\cdot,\cdot)\right), \\ (\tau^2, \varphi, \sigma_\varepsilon^2) &\sim \pi(\tau^2)\pi(\varphi)\pi(\sigma_\varepsilon^2), \end{aligned} \tag{1}$$

- Performing inference under this model amounts to exploring the posterior

$$\pi(\mathbf{z}, \mathbf{u}, \tau^2, \varphi, \sigma_\varepsilon^2 \mid \mathbf{y}) \propto \mathcal{N}(\mathbf{y} \mid \mathbf{z}, \sigma_\varepsilon^2 I_m)\mathcal{N}(\mathbf{z} \mid 0, C_\phi^{\mathrm{NS}})\mathcal{N}(\mathbf{u} \mid 0, C_\varphi^{\mathrm{S}})\pi(\tau^2)\pi(\varphi)\pi(\sigma_\varepsilon^2)$$

- . . . and using sparse presentations – and fixing τ^2

$$\pi(\mathbf{z}, \mathbf{u}, \lambda, \sigma_\varepsilon^2 \mid \mathbf{y}) \propto \mathcal{N}(\mathbf{y} \mid A\mathbf{z}, \sigma_\varepsilon^2 I_m)\mathcal{N}(\mathbf{z} \mid 0, Q_{\mathbf{u}}^{-1})\mathcal{N}(\mathbf{u} \mid 0, Q_\lambda^{-1})\pi(\lambda)\pi(\sigma_\varepsilon^2).$$

Prior: Gaussian Markov random fields

- Matérn fields are often defined as stationary Gaussian random field with a covariance function

$$\mathrm{Cov}(x, x') = \mathrm{Cov}(x - x') = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{|x - x'|}{\ell} \right)^{\nu} K_\nu \left(\frac{|x - x'|}{\ell} \right) \tag{2}$$

  where $x, x' \in \mathbb{R}^d$, $\nu > 0$ is the smoothness parameter, and K_ν is modified Bessel function of the second kind or order ν.

- The Fourier transform of the covariance function gives a power spectrum

$$S(\xi) = \frac{2^d \pi^{d/2} \Gamma(\nu + d/2)}{\Gamma(\nu) \ell^{2\nu}} \left(\frac{1}{\ell^2} + |\xi|^2 \right)^{-(\nu + d/2)} .$$

- Rozanov 1977: only fields with spectral density given by the reciprocal of a polynomial have a Markov representation.

Prior: Stochastic Partial Differential Equation

- Let w be white noise. We may define the basic Matérn field z via $\widehat{z} = \sigma\sqrt{S(\xi)}\widehat{w}$ in the sense of distributions.
- By using inverse Fourier transforms, write SPDE

$$\left(1 - \ell^2\Delta\right) z = \sigma\sqrt{\ell^d} w.$$

  The field z is isotropic.
- Inhomogeneous field by allowing a spatially variable length-scaling field $\ell(x)$

$$\left(1 - \ell(x)^2\Delta\right) z = \sigma\sqrt{\ell(x)^d} w.$$

Convergence of the discretised prior $h \to 0$

Theorem

Let $z(x;u)$ satisfy

$$\left(1 - \ell(x;u)^2 \Delta\right) z = \sigma_0 \sqrt{\ell(x;u)^d} w \text{ in } D \tag{3}$$

with the periodic boundary condition where $\ell(x;u) = g(u(x))$ and

$$g(s) = \exp(s)$$

Let $z^N(x;u^N)$ satisfy

$$\left(1 - \ell(x;u^N)^2 \Delta_N\right) z^N(x;u^N) = \sigma_0 \sqrt{\ell(x;u^N)^d} w^N,$$

on $h\mathbb{Z}^d \cap D$, with the periodic boundary.
Then $z^N(\cdot;u^N)$ converges to z in $L^2(L^2(D_h), P)$ as $N \to \infty$.

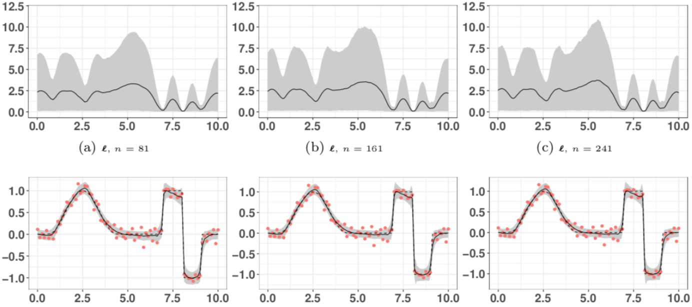
Hyperprior field & parameters and normalisation constants

- Hyperprior fields
  - Matérn covariance
  - Exponential covariance
  - Squared exponential covariance
  - Cauchy walk
- Parameters
  - $\log \sigma_\varepsilon^2 \sim \mathcal{N}(\cdot,\cdot)$ – observation moise
  - $\log \lambda \sim \mathcal{N}(\cdot,\cdot)$ – hyperprior length-scaling
- Normalisation constants
  - $\log\det \sigma_\varepsilon^2 I$ – Easy
  - $\log\det Q_{\mathbf{u}}^{-1}$ – Utilise sparsity of $Q_{\mathbf{u}}$
  - $\log\det Q_\lambda^{-1}$ – Easy for 1D exponential covariance, difficult generally in $\mathbf{R}^d$

Three MCMC sampling approaches

- Implemented:
  - Adaptive Metropolis-within-Gibbs MwG: draws samples from the multidimensional vector u
  - Whitened Elliptical Slice Sampling w-ELL-SS: ancillary augmentation over z and u and uses elliptical slice sampling
  - Marginal Elliptical Slice Sampling m-ELL-SS: integrates out the non-stationary process, resulting in a marginal sampler that draws from u by combining ancillary augmentation and ELL-SS to break the correlation between u and λ.
- Under development:
  - Variational Bayes (due to be submitted by 30 April)

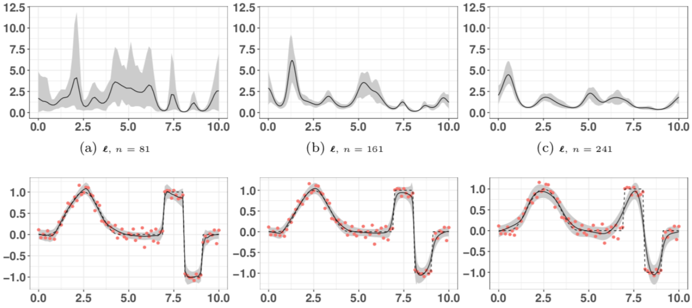
AR(1) hyperpior with the MwG algorithm



(a) $\boldsymbol{\ell}$, $n = 81$ (b) $\boldsymbol{\ell}$, $n = 161$ (c) $\boldsymbol{\ell}$, $n = 241$

(d) $\mathbf{z}$, $n = 81$ (e) $\mathbf{z}$, $n = 161$ (f) $\mathbf{z}$, $n = 241$

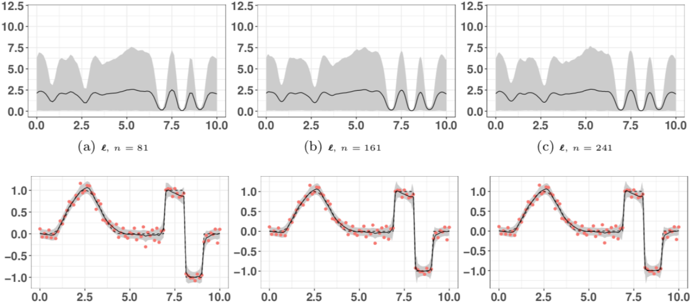
SE hyperpior with the MwG algorithm



(a) $\boldsymbol{\ell}$, $n = 81$ (b) $\boldsymbol{\ell}$, $n = 161$ (c) $\boldsymbol{\ell}$, $n = 241$

(d) $\mathbf{z}$, $n = 81$ (e) $\mathbf{z}$, $n = 161$ (f) $\mathbf{z}$, $n = 241$

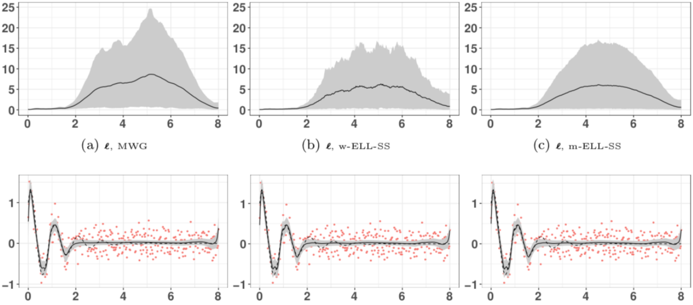
SE hyperpior model with m-ELL-SS algorithm



(a) $\boldsymbol{\ell}$, $n = 81$ (b) $\boldsymbol{\ell}$, $n = 161$ (c) $\boldsymbol{\ell}$, $n = 241$

(d) $\mathbf{z}$, $n = 81$ (e) $\mathbf{z}$, $n = 161$ (f) $\mathbf{z}$, $n = 241$

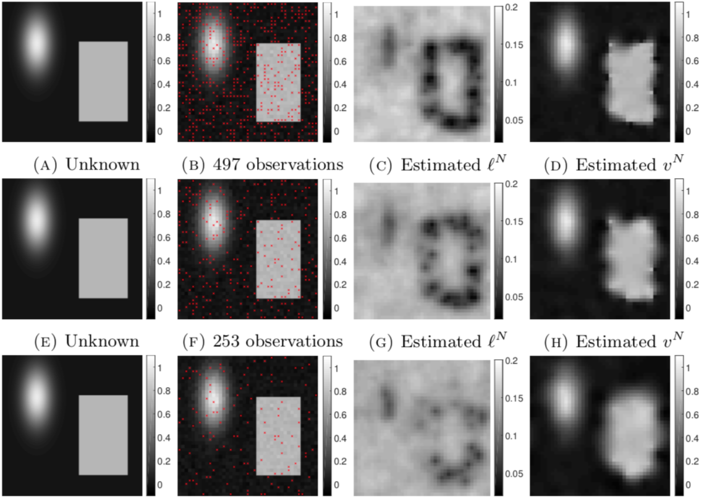
AR(1) hyperpior with MWG, w-ELL-SS and m-ELL-SS



(a) $\boldsymbol{\ell}$, MWG (b) $\boldsymbol{\ell}$, w-ELL-SS (c) $\boldsymbol{\ell}$, m-ELL-SS

(d) $\mathbf{z}$, MWG (e) $\mathbf{z}$, w-ELL-SS (f) $\mathbf{z}$, m-ELL-SS

MwG for 2D interpolation



(A) Unknown (B) 497 observations (C) Estimated ℓ^N (D) Estimated v^N

(E) Unknown (F) 253 observations (G) Estimated ℓ^N (H) Estimated v^N

(I) Unknown (J) 119 observations (K) Estimated ℓ^N (L) Estimated v^N

Non-Gaussian models – Cauchy priors

- Markku Markkanen, Lassi Roininen, Janne M J Huttunen and Sari Lasanen, Cauchy difference priors for edge-preserving Bayesian inversion with an application to X-ray tomography, ArXiv 2016.
- Alberto Mendoza, Lassi Roininen, Mark Girolami, Jere Heikkinen, and Heikki Haario, Statistical methods to enable practical on-site tomographic imaging of whole-core samples, SPWLA 2018.

Stable random walks

- Let $\{\mathcal{X}(t), t \in \mathbb{I} \subset \mathbb{R}^+\}$ be a stochastic process. We call it a Lévy α-stable process starting from zero, or simply as stable process, if $\mathcal{X}(0) = 0$, $\mathcal{X}$ has independent increments and
$$\mathcal{X}(t) - \mathcal{X}(s) \sim S_\alpha\left((t-s)^{1/\alpha}, \beta, 0\right) \tag{4}$$
for any $0 \leq s < t < \infty$ and for some $0 < \alpha \leq 2, -1 \leq \beta \leq 1$.
- For the continuous limit of the Cauchy walk, we apply independently scattered measures. We obtain random walk approximation
$$X_{t_i} - X_{t_{i-1}} \sim S_\alpha(h^{\frac{1}{\alpha}}, \beta, 0)$$
where $t_i - t_{i-1} =: h$. It is easy to see that such random walk approximations converge to the α-stable Lévy motion as $h \to 0$ in distribution on the Skorokhod space of functions that are right-continuous and have left limits.

Cauchy and Gaussian random walk realisations

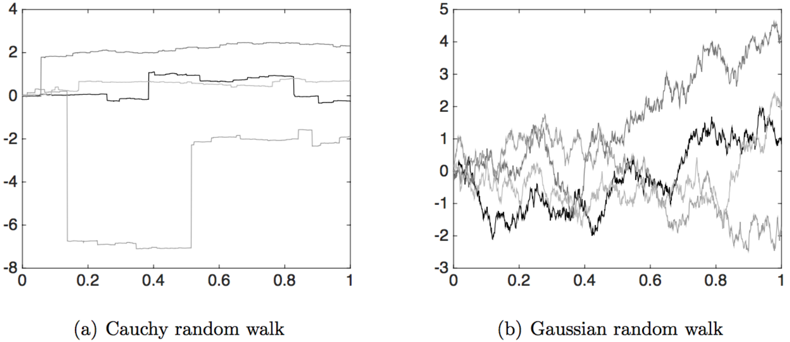
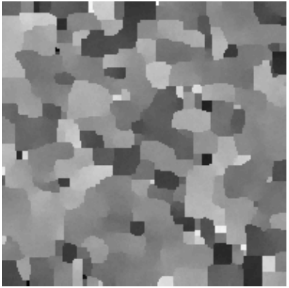


Figure 1. Realizations of Cauchy and Gaussian random walks.

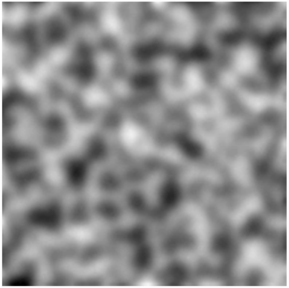
2D Cauchy vs Gaussian priors

$$X_{j,k} - X_{j-1,k} \sim \mathrm{Cauchy}(\lambda h_1)$$
$$X_{j,k} - X_{j,k-1} \sim \mathrm{Cauchy}(\lambda h_2)$$

$$X_{j,k} - X_{j-1,k} \sim \mathcal{N}(0, \sigma^2 h_1/h_2)$$
$$X_{j,k} - X_{j,k-1} \sim \mathcal{N}(0, \sigma^2 h_2/h_1)$$



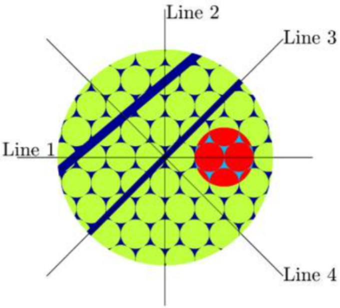
(a) Realisation of 2D Cauchy field

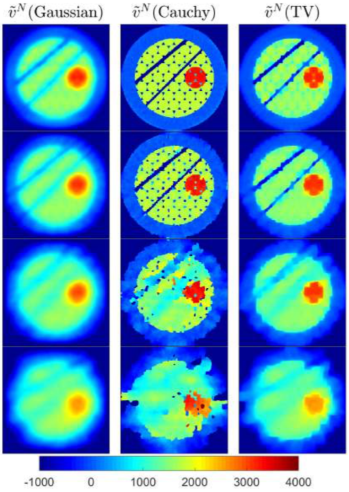


(b) Realisation of 2D Gaussian field

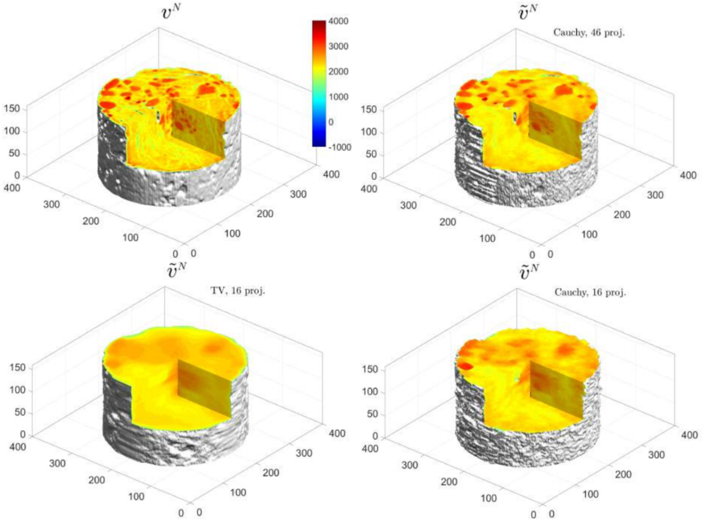
Tomographic imaging of whole-core samples

- 46, 23 ,12, 6 projections with 10% noise





Sandstone 3D tomography with 10% noise



Conclusion

- 2–4 layer spatially scalable hierarchical models that can be efficiently handled fully probabilistically.
- Gaussian and non-Gaussian iid random fields.
- This leads to computational advantages and clear statistical interpretation of all the random fields and parameters.