Sparse Non-Stationary Hierarchical Priors for Bayesian Inversion

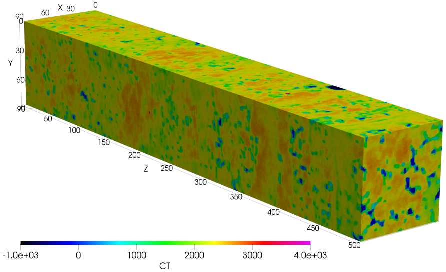
Lassi Roininen

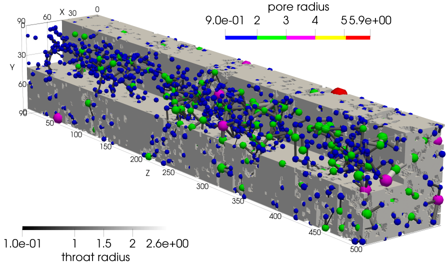
LUT University

Workshop at DTU: Uncertainty Quantification for Inverse Problems
Copenhagen – 17 December 2018

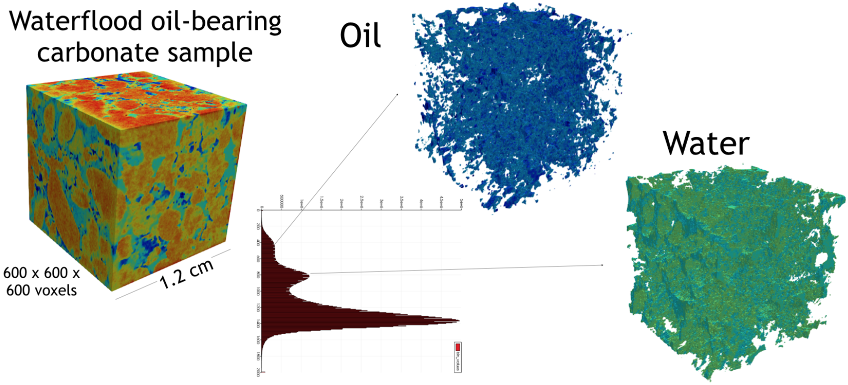
Motivation

Rock core samples and classified pore structure





Mixed-wet carbonate reservoir rocks



Seismic tomography

with Timo Lähivaara and Kenneth Muhumuza (University of Eastern Finland)

Within the CoE, a collaboration with Associate Professor Lassi Roininen from the Lappeenranta University of Technology on non-Gaussian discretisation-invariant edge-preserving prior models has been started.

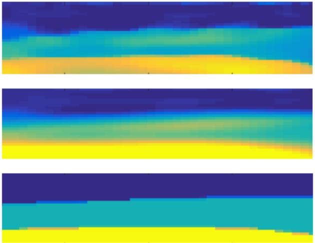
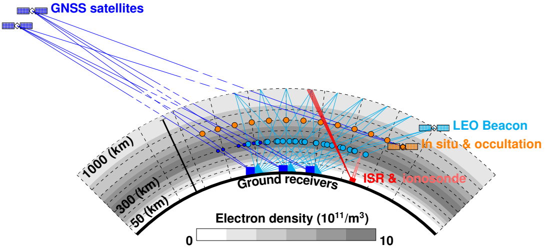
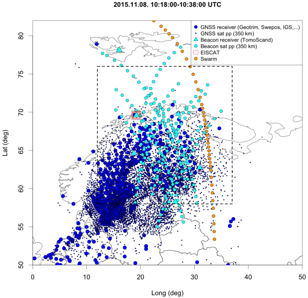


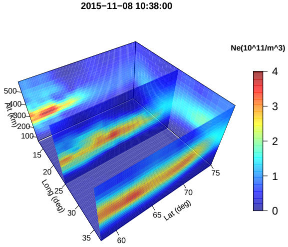
Fig: Preliminary results of the seismic tomography (conditional mean estimates) with Cauchy difference prior (top) and Gaussian smoothness prior (middle). True profile is shown at the bottom.

Fig: Measurement infrastucture at the Le Mans Universite.

Spatiotemporal ionospheric tomography – TomoScand

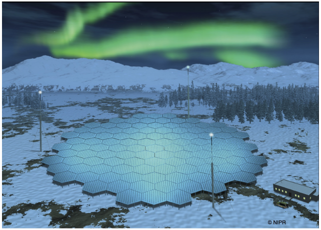




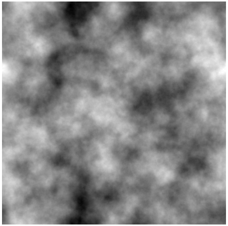


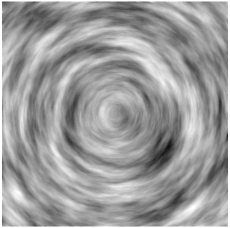
EISCAT_3D Incoherent Scatter Radar

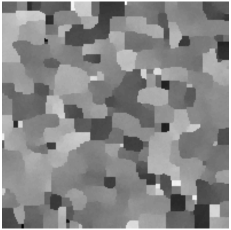
- EUR 70 million international research collaboration, on the European large-scale research infrastructure ESFRI roadmap.
- Research councils of China, Finland, Japan, Norway, Sweden and United Kingdom.
- Currently under built in arctic Finland, Norway and Sweden
- Able to monitor continuously threats due to e.g. solar storms, and detecting hard-to-find asteroids
- PB-scale annual data production
- 5D-inversion — 3 spatial, 1 temporal, and 1 spectral coordinate



Gaussian and Cauchy Priors







- Stationary Gaussian random fields
- Anisotropic and inhomogeneous Gaussian random fields
- Cauchy random fields

Hierarchical models

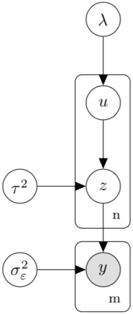
Hierarchical models

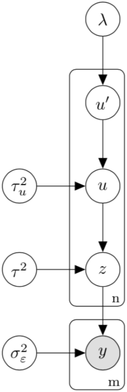
- Lassi Roininen, Mark Girolami, Sari Lasanen and Markku Markkanen, Hyperpriors for Matérn fields with applications in Bayesian inversion, Inverse Problems and Imaging, **13**:1 (2019).
- Karla Monterrubio-Gómez, Lassi Roininen, Sara Wade, Theo Damoulas and Mark Girolami, Posterior Inference for Sparse Hierarchical Non-stationary Models, ArXiv 2018.
- Neil K Chada, Marco A Iglesias, Lassi Roininen and Andrew Stuart, Parameterizations for ensemble Kalman inversion, Inverse Problems, (2018).

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

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

- 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 $N \to \infty$.

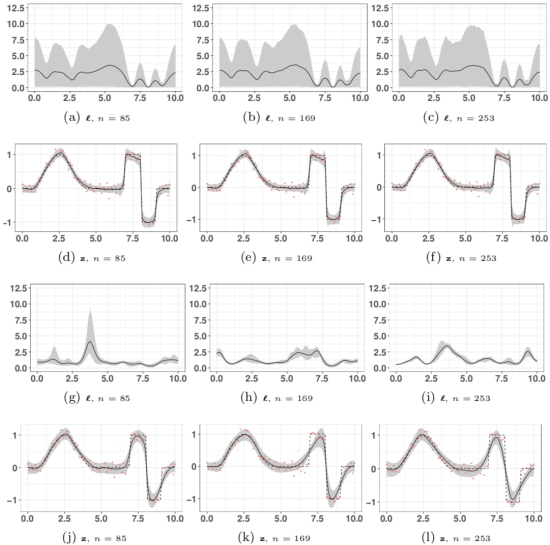
Hyperprior field & parameters and normalisation constants

- Hyperprior fields
  - Matérn covariance
  - Exponential covariance
  - Squared exponential covariance
  - Cauchy walk
- Parameters
  - $\log \sigma_\varepsilon^2 \sim \mathcal{N}(\cdot, \cdot)$ – observation noise variance
  - $\log \lambda \sim \mathcal{N}(\cdot, \cdot)$ – 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 $\mathbf{R}^d$

MCMC, VB and Optimisation

- MCMC schemes
    - Adaptive Metropolis-within-Gibbs
    - Whitened Elliptical Slice Sampling
    - Marginal Elliptical Slice Sampling
- Hierarchical Ensemble Kalman Inversion
- Variational Bayes (in progress)
- Series expansion, dimension reduction techniques (in progress, early stage)

Metropolis-within-Gibbs



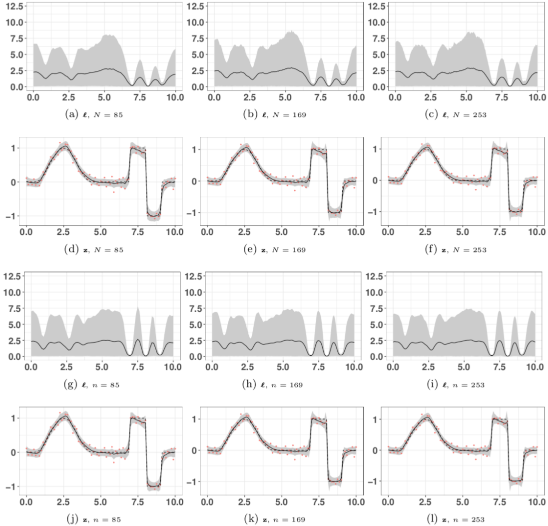
(a) $\boldsymbol{\ell}$, $n = 85$ (b) $\boldsymbol{\ell}$, $n = 169$ (c) $\boldsymbol{\ell}$, $n = 253$

(d) $\mathbf{z}$, $n = 85$ (e) $\mathbf{z}$, $n = 169$ (f) $\mathbf{z}$, $n = 253$

(g) $\boldsymbol{\ell}$, $n = 85$ (h) $\boldsymbol{\ell}$, $n = 169$ (i) $\boldsymbol{\ell}$, $n = 253$

(j) $\mathbf{z}$, $n = 85$ (k) $\mathbf{z}$, $n = 169$ (l) $\mathbf{z}$, $n = 253$

Same, but with marginal elliptical slice sampling



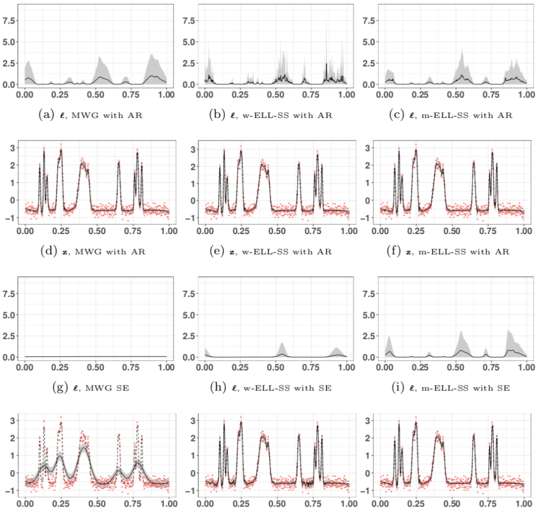
(a) $\boldsymbol{\ell}$, $N = 85$ (b) $\boldsymbol{\ell}$, $N = 169$ (c) $\boldsymbol{\ell}$, $N = 253$

(d) $\mathbf{z}$, $N = 85$ (e) $\mathbf{z}$, $N = 169$ (f) $\mathbf{z}$, $N = 253$

(g) $\boldsymbol{\ell}$, $n = 85$ (h) $\boldsymbol{\ell}$, $n = 169$ (i) $\boldsymbol{\ell}$, $n = 253$

(j) $\mathbf{z}$, $n = 85$ (k) $\mathbf{z}$, $n = 169$ (l) $\mathbf{z}$, $n = 253$

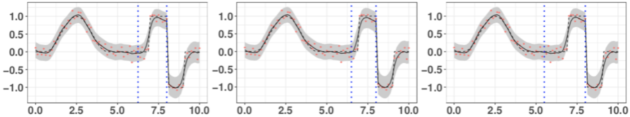
Highly non-stationary synthetic case



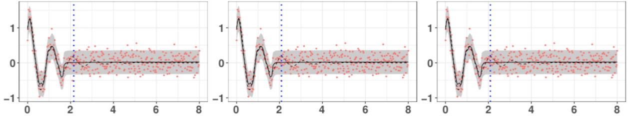
(a) $\boldsymbol{\ell}$, MWG with AR
(b) $\boldsymbol{\ell}$, w-ELL-SS with AR
(c) $\boldsymbol{\ell}$, m-ELL-SS with AR
(d) $\mathbf{z}$, MWG with AR
(e) $\mathbf{z}$, w-ELL-SS with AR
(f) $\mathbf{z}$, m-ELL-SS with AR
(g) $\boldsymbol{\ell}$, MWG SE
(h) $\boldsymbol{\ell}$, w-ELL-SS with SE
(i) $\boldsymbol{\ell}$, m-ELL-SS with SE
(j) $\mathbf{z}$, MWG with SE
(k) $\mathbf{z}$, w-ELL-SS with SE
(l) $\mathbf{z}$, m-ELL-SS with SE

Comparative Evaluation — R TGP package (CRAN)

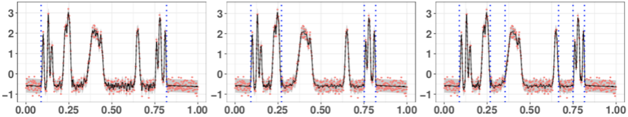
(a) 10^5 iterations, $2 \cdot 10^4$ burn-in,
(b) $2 \cdot 10^5$ iterations, $5 \cdot 10^4$ burn-in,
(c) $5 \cdot 10^5$ iterations, 10^5 burn-in



(a) MAE=0.056 (b) MAE=.057 (c) MAE=.057

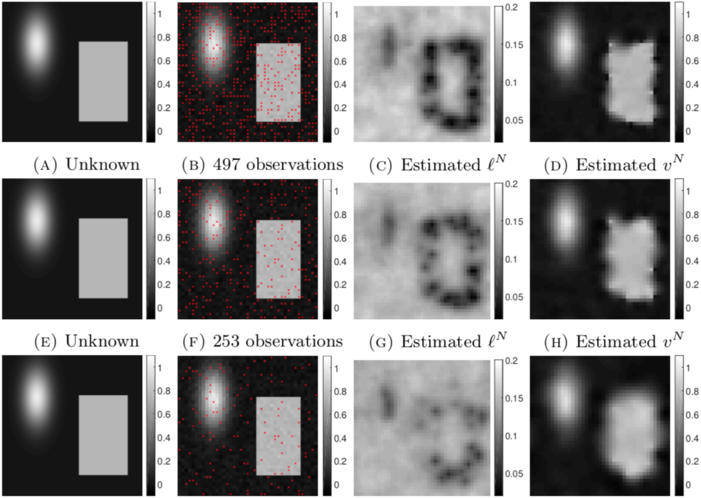


(a) MAE=0.043 (b) MAE=0.043 (c) MAE=.043



(a) MAE=0.079 (b) MAE=.067 (c) MAE=.065

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

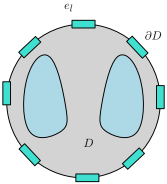
(I) Unknown (J) 119 observations (K) Estimated ℓ^N (L) Estimated v^N

Electrical Impedance Tomography (EIT)

Forward Problem

Given $(\kappa, I) \in L^\infty(D; \mathbb{R}^+) \times \mathbb{R}^m$ find $(\nu, V) \in H^1(D) \times \mathbb{R}^m$:

$$
\begin{aligned}
-\nabla \cdot (\kappa \nabla \nu) &= 0 \quad \in \; D, \\
\nu + z_\ell \kappa \nabla \nu \cdot n &= V_\ell \;\; \in \; e_\ell, \quad \ell = 1, \dots, m, \\
\nabla \nu \cdot n &= 0 \quad \in \; \partial D \setminus \cup_{\ell=1}^m e_\ell, \\
\int \kappa \nabla \nu \cdot n \, ds &= I_\ell \quad \in \; e_\ell, \quad \ell = 1, \dots, m.
\end{aligned}
$$



Ohm's Law: $V = R(\kappa) \times I$.

Inverse Problem

Set $\kappa = \exp(u)$. Given a set of K noisy measurements of voltage $V(k)$ from currents $I(k)$, and $\mathcal{G}_k(u) = R\big(\exp(u)\big) \times I(k)$, find u from y where:

$$
y(k) = \mathcal{G}_k(u) + \eta, \quad \eta \sim \mathrm{N}(0, \gamma^2), \quad k = 1, \dots, K.
$$

Basic EnKF Inversion (Iglesias et al (2013))

- Initial Ensemble $\{u_0^{(j)}\}_{j=1}^J \subset X$.
- Ensemble First and Second Order Moments

  Means:

$$\overline{u}_n = \frac{1}{J}\sum_{k=1}^J u_n^{(k)}, \quad \overline{w}_n = \frac{1}{J}\sum_{k=1}^J \mathcal{G}(u_n^{(k)}).$$

  Covariances:

$$C_n^{ww} = \frac{1}{J}\sum_{k=1}^J \big(\mathcal{G}(u_n^{(k)}) - \overline{w}_n\big) \otimes \big(\mathcal{G}(u_n^{(k)}) - \overline{w}_n\big),$$

$$C_n^{uw} = \frac{1}{J}\sum_{k=1}^J \big(u_n^{(k)} - \overline{u}_n\big) \otimes \big(\mathcal{G}(u_n^{(k)}) - \overline{w}_n\big).$$

- Update step $n \mapsto n+1$:

$$u_{n+1}^{(j)} = u_n^{(j)} + C_n^{uw}\big(C_n^{ww} + \Gamma\big)^{-1}\big(y - \mathcal{G}(u_n^{(j)})\big).$$

Whittle-Matérn Initial Ensembles

- Create initial ensemble of functions via Gaussian random fields.
- Common choice: Whittle-Matérn family

$$c_{\sigma,\nu,\tau}(x,x') := \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\tau|x-x'|\right)^{\nu} K_\nu\left(\tau|x-x'|\right).$$

  - Smoothness parameter: $\nu \in \mathbb{R}^+$.
  - Inverse length-scale parameter: $\tau \in \mathbb{R}^+$.
  - Amplitude parameter: $\sigma \in \mathbb{R}$.

- Corresponding covariance operator

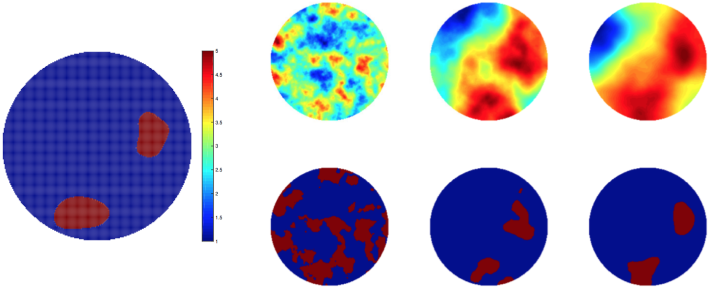
$$\mathcal{C}_{\sigma,\nu,\tau} \propto \sigma^2 \tau^{2\nu} (\tau^2 I - \triangle)^{-\nu-\frac{d}{2}}.$$

  - $\nu = \alpha - \frac{d}{2}$.

- Hierarchical: invert for parameters such as σ, ν, τ as well as field itself.

Electrical Impedance Tomography

- Non-centred Hierarchical.
- Reconstructed level set function at three different iteration steps.
- Reconstructed conductivity at three different iteration steps.



Cauchy Priors

Total variation prior is Gaussian when $h \to 0$

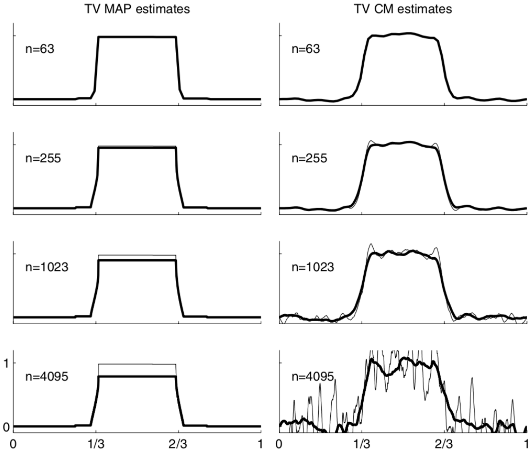


Figure 4. In all the plots in this figure, the coordinate axis limits are the same to allow easy comparison. Left column: MAP estimates for the TV prior with parameter $\alpha_n = 135$ (thin line) and $\alpha_n = 16.875\sqrt{n+1}$ (thick line). Right column: CM estimates for the TV prior with parameter $\alpha_n = 135$ (thin line) and $\alpha_n = 16.875\sqrt{n+1}$ (thick line).

TV and Besov space priors

- Lassas and Siltanen 2004 showed that TV are not discretisation-invariant
- Lassas, Saksman and Siltanen 2009 constructed Besov space priors
  - Often defined via wavelet expansions.
  - For edge-preserving inversion the Haar wavelet basis is often used
  - However due to the structure of the Haar basis, discontinuities are preferred on an underlying dyadic grid given by the discontinuities of the basis functions. For example, on the domain $(0, 1)$, discontinuity is vastly preferred at $x = 1/4$ over $x = 1/3$.
  - Thus Besov priors make, in most practical cases, both a strong and unrealistic assumption.

Non-Gaussian models – Cauchy priors

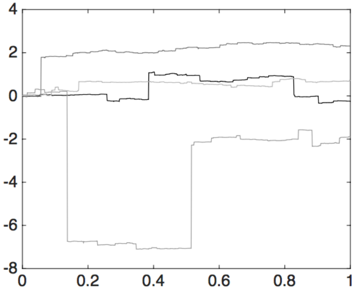
- Markku Markkanen, Lassi Roininen, Janne M J Huttunen and Sari Lasanen, Cauchy difference priors for edge-preserving Bayesian inversion with an application to X-ray tomography, ArXiv 2016.
- A. Mendoza, L. Roininen, M. Girolami, J. Heikkinen, and H. Haario, *Statistical Methods To Enable Practical On-Site Tomographic Imaging of Whole-Core Samples*. SPWLA 59th Annual Logging Symposium (2018), final version submitted to Geophysics.
- A. Mendoza, L. Roininen, M. Girolami, J. Heikkinen, and H. Haario, *Accelerated whole core analysis optimization with wellsite tomography instrumentation and Bayesian inversion.*, submitted to Petrophysics.

- Theoretical foundations: Sari Lasanen, Matt Dunlop (Helsinki/NYU), Tim Sullivan (Berlin), Neil Chada (NUS)
- Computational aspects: Simo Särkkä (Aalto), Matt Moores (Wollongong)
- Industrial applications: Finnos Oy, Imperial College London, Alan Turing Institute (Mark Girolami and Alberto Mendoza)

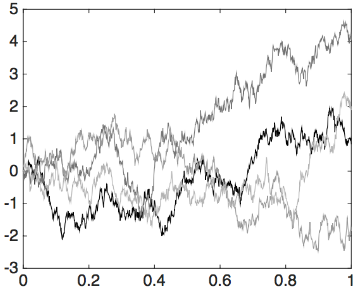
Stable random walks

- 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



(a) Cauchy random walk



(b) Gaussian random walk

Figure 1. Realizations of Cauchy and Gaussian random walks.

Edge-preserving property – Cauchy walk

The conditional density for Cauchy difference prior for X_j can then be written as a probability density

$$D(X_j|X_{j-1} = -a, X_{j+1} = a) \propto \frac{1}{1+(X_j-a)^2}\frac{1}{1+(X_j+a)^2}. \tag{5}$$

This is simply a product of two Cauchy probability density functions, one from each neighbor of X_j. In order to see the properties of these functions, consider the second derivative of the probability density function with respect to X_j at zero:

$$\begin{aligned} D''(0) < 0, \text{when, } |a| < 1 &: \text{ maximum at } 0, \text{ unimodal} \\ D''(0) = 0, \text{when, } |a| = 1 &: \text{ as flat as possible at } 0 \\ D''(0) > 0, \text{when, } |a| > 1 &: \text{ minimum at } 0, \text{ bimodal.} \end{aligned} \tag{6}$$

Edge-preserving property – Cauchy walk

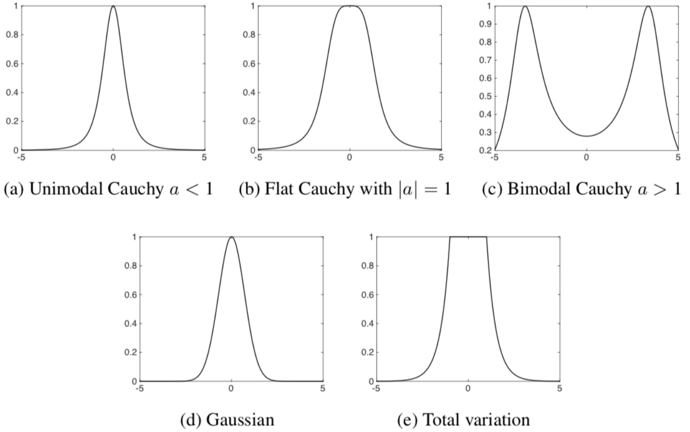


Figure 3. Upper panel: Cauchy probability density function for X_2, given fixed $X_1 = -a$ and $X_3 = a$. Bottom panel: The same case for Gaussian difference prior and total variation prior.

2D Cauchy vs Gaussian priors – Estimators and UQ

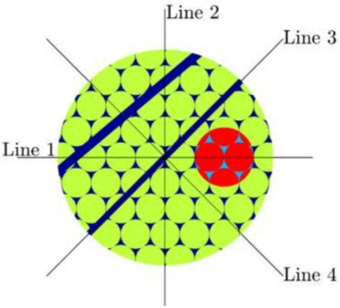
$$X_{j,k} - X_{j-1,k} \sim \mathrm{Cauchy}(\lambda h_1) \qquad X_{j,k} - X_{j-1,k} \sim \mathcal{N}(0, \sigma^2 h_1/h_2)$$

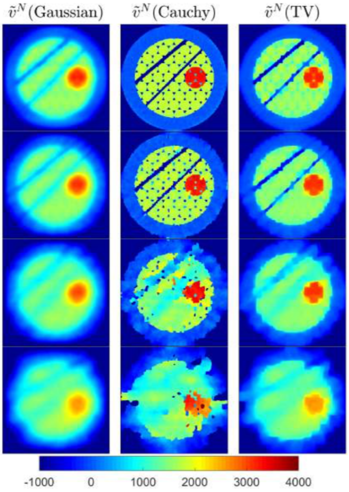
$$X_{j,k} - X_{j,k-1} \sim \mathrm{Cauchy}(\lambda h_2) \qquad X_{j,k} - X_{j,k-1} \sim \mathcal{N}(0, \sigma^2 h_2/h_1)$$

- Metropolis-within-Gibbs — Operational
- Optimisation — Local minima, artifacts
- Hamiltonian Monte Carlo - No U-Turn Samples — Recent implementation, under development

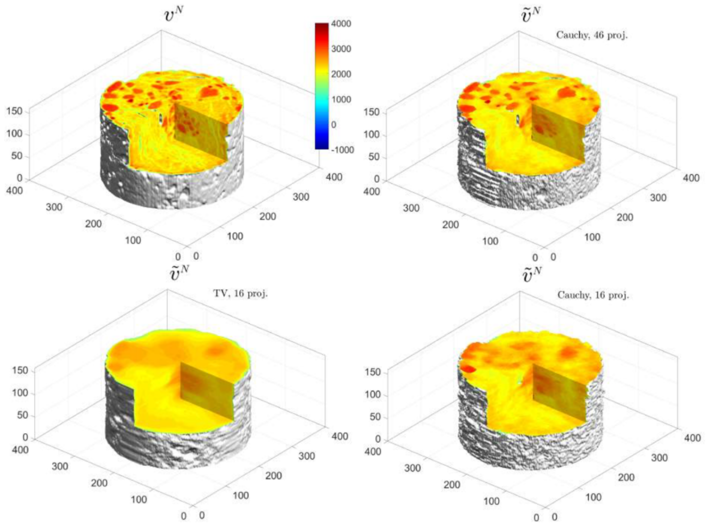
Tomographic imaging of whole-core samples

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





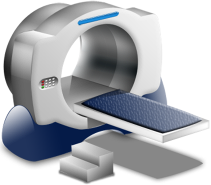
Sandstone 3D tomography with 10% noise

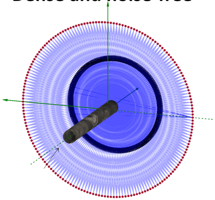


On-site whole-core X-ray CT

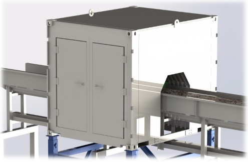
Measurement principle

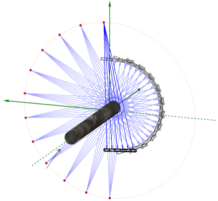
Laboratory (Medical) | **Dense and noise-free**





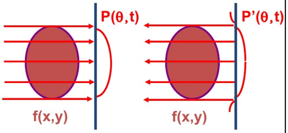
On-site (Industrial) | **Sparse + noise-contamination**





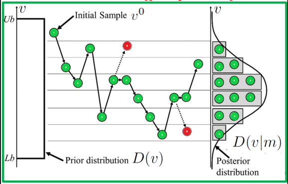
Imaging method

Direct (standard)



From Patel, G. (2009)

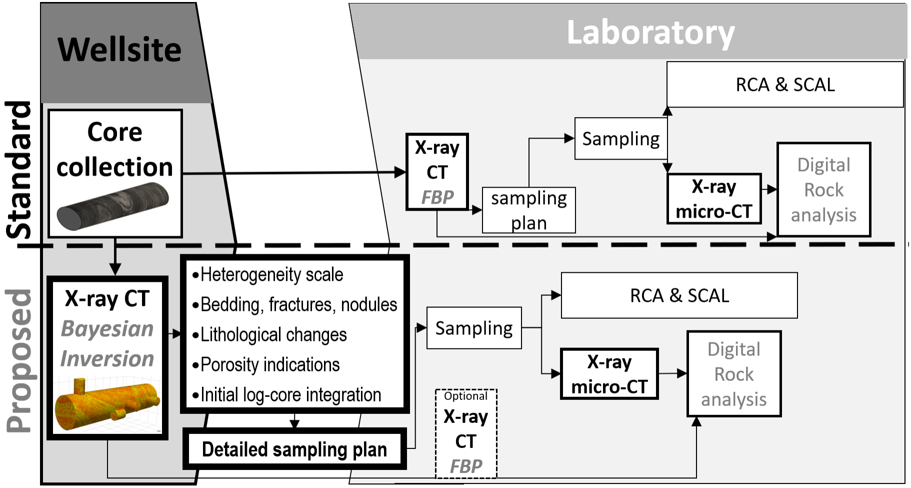
Statistical (proposed)



Modified from: Lee et al. (2015)

wellsite versatility

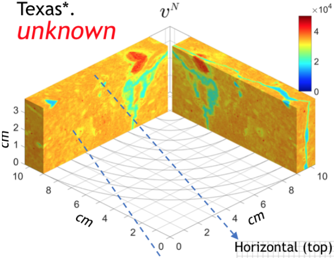
Core analysis workflow

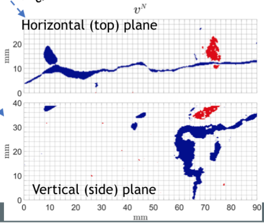


Identifying fractures

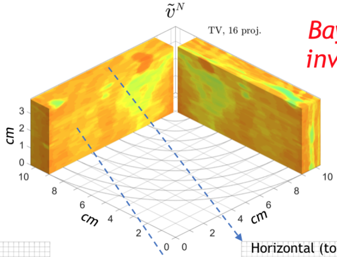
Example 2: Fractured welded tuff from the Santana tuff from Closed Canyon, Big Bend National Park, Texas*.

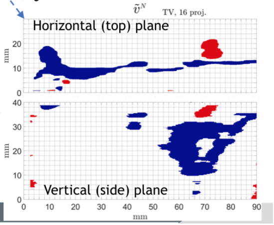
unknown





Bayesian inversion





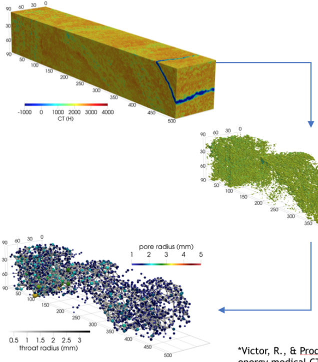
*Ketcham, R. A., (2016), Fracture in tuff. *Digital Rocks Portal*. DOI: 10.17612/P7PP4B. http://www.digitalrocksportal.org/projects/45.

7 11/29/2018

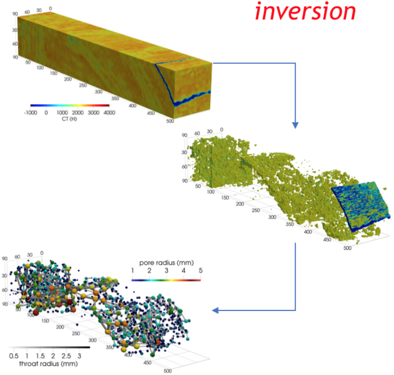
Pore network analysis

Example 3: Complex pore topology, oil-bearing carbonate from offshore south-eastern Brazil*.

unknown



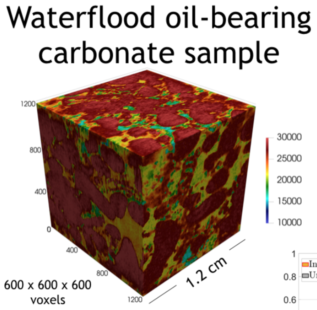
Bayesian inversion

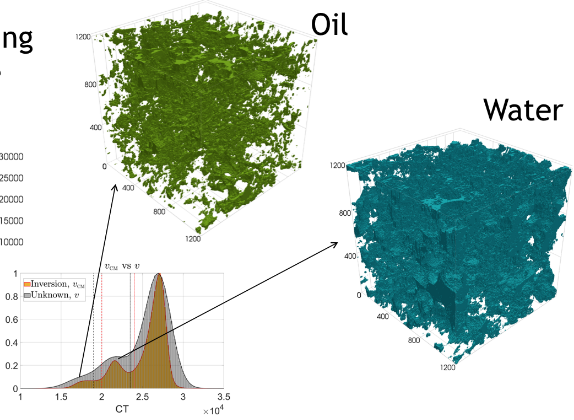


*Victor, R., & Prodanovic, M. (2017, August 7). Dual-energy medical CT in carbonate rocks. Retrieved August 14, 2018, from www.digitalrocksportal.org.

Micro-CT

Example 4: mixed-wet carbonate reservoir rocks from the Middle-East*.





*Alhammadi, A., Alratrout, A., Bijeljic, B., & Blunt, M. (2018, May 17). X-ray micro-tomography datasets of mixed-wet carbonate reservoir rocks for in situ effective contact angle measurements. Retrieved August 16, 2018, from www.digitalrocksportal.org

Conclusion & Future Prospects

- Methods & Applications
  - Hierarchical and Cauchy priors, convergence
  - MCMC (MwG, elliptical), hierarchical EnKF
  - Numerous application domains, including, near-space radar and tomography, and industrial tomography
- Future research
  - Non-Gaussian priors — Student's t, Cauchy, Lévy α-stable random fields (iid models).
  - Joint inversion-classification schemes

$$\begin{aligned} x(\mathbf{r}) &\sim \mathcal{GP}(\cdot) \\ y &= Ax + \epsilon \\ c &= \Phi(w^\top x + b), \end{aligned} \tag{7}$$