CUQI

A research initiative in Computational Uncertainty Quantification for Inverse problems

Per Christian Hansen
Technical University of Denmark

With thanks to everyone in the CUQI team for all their input to this presentation.

The CUQI Project, 2019-2025

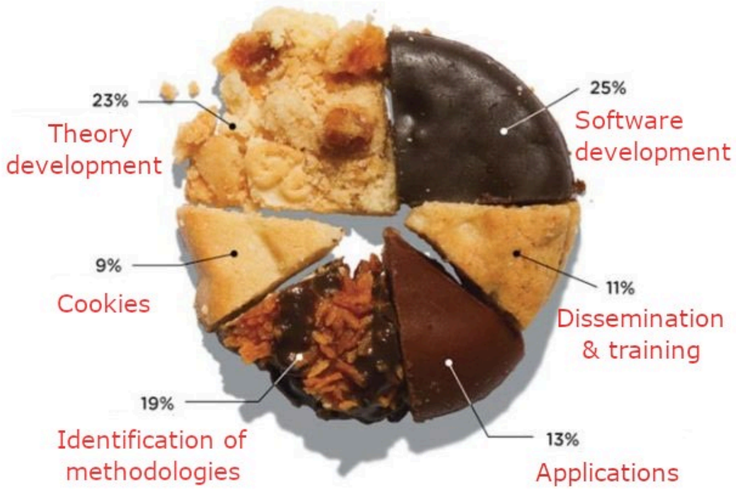
VILLUM FONDEN

A unique collaborative effort to develop a mathematical, statistical and computational **framework** for applying uncertainty quantification (UQ) to inverse problems.

We also develop a Python **software** package **CUQIpy** for modeling and computations, allowing experts as well as non-experts to apply UQ to their inverse problems.

The team, as of Nov. 2021

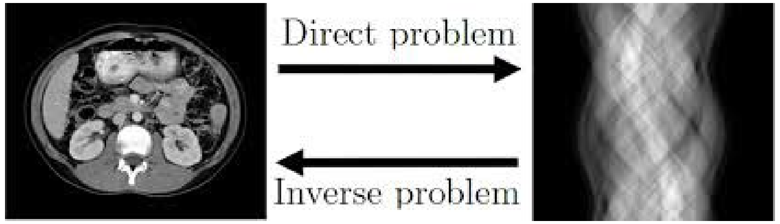
In this talk we look at some of the ingredients.



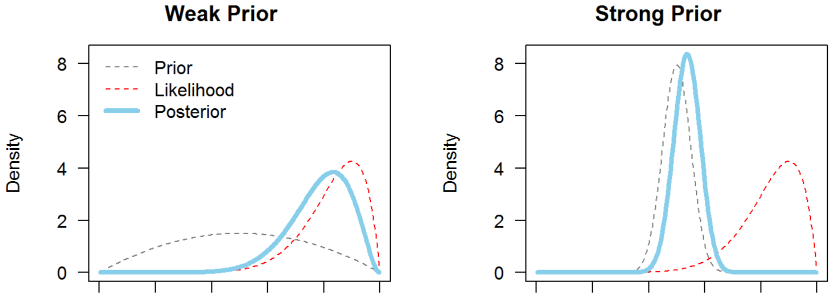
Main Steps of Bayesian Inference and UQ

$$\mathcal{R}[f](\theta,s) = \int_{L_{\theta,s}} f(\xi_1,\xi_2)\,d\ell = g(\theta,s)$$

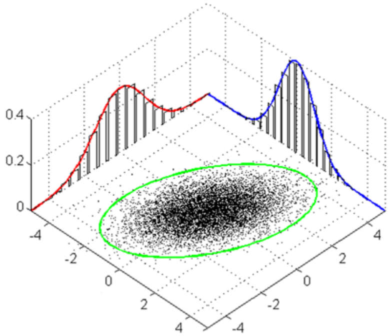
A
Define the **model**



B
Specify the **prior**



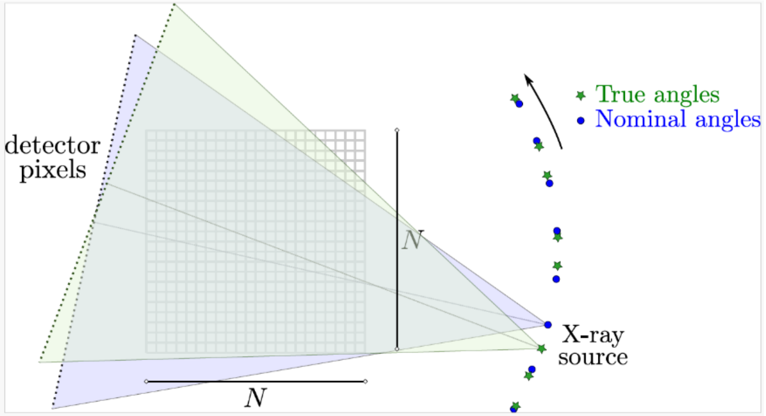
C
Choose the **sampler**



- Uncertain CT projection angles
- Structural priors for CT
- The horseshoe prior

Examples of CUQI activities

A: CT Model for Uncertain View Angles



Uribe, Bardsley, Dong, Hansen, Riis (2022).

For each position of the X-ray source, we measure a set of data = a **view**.

The **true view angles** may differ from the assumed **nominal view angles**.

- The description of the measured data uses the *unknown* true angles.
- A bad reconstruction uses the nominal angles.

What is new in this work:

- Joint computation of the image and the correct angles.
- UQ of the improved angles.

Model: $\boldsymbol{b} = \boldsymbol{A}(\boldsymbol{\theta})\,\boldsymbol{x} + \boldsymbol{e}, \qquad \boldsymbol{e}$ = noise.

$\boldsymbol{A}(\boldsymbol{\theta})$ = forward model for angles $\boldsymbol{\theta}$.

Unknowns: the image $\boldsymbol{x}$ and the true view angles $\boldsymbol{\theta}$:

$$\pi_{\text{pos}}(\boldsymbol{x},\boldsymbol{\theta}) \propto \pi_{\text{lik}}(\boldsymbol{b}\,|\,\boldsymbol{x},\boldsymbol{\theta}) \times \pi_{\text{pri}}(\boldsymbol{x}) \times \pi_{\text{pri}}(\boldsymbol{\theta}) \ .$$

- $\pi_{\text{lik}}(\boldsymbol{b}\,|\,\boldsymbol{x},\boldsymbol{\theta})$ = **Gaussian** (and approximation to the log-Poisson noise in CT).
- $\pi_{\text{pri}}(\boldsymbol{x})$ = **Laplace distribution of the differences of neighbour pixels** $\rightarrow$ sharp edges in the image.
- $\pi_{\text{pri}}(\boldsymbol{\theta})$ = **von Mises distribution** (i.e., a *periodic* normal distribution).

We introduce hyperparameters in all three distributions, and use a hybrid Gibbps sampler.

Some Details of the Algorithm

The likelihood for data $\boldsymbol{b} \in \mathbb{R}^m$ with Gaussian noise:

$$\pi_{\text{lik}} = \left(\frac{\lambda}{2\pi}\right)^{m/2} \exp\left(-\frac{\lambda}{2}\|\boldsymbol{A}(\boldsymbol{\theta})\,\boldsymbol{x} - \boldsymbol{b}\|_2^2\right), \qquad \lambda = \text{hyperparameter.}$$

The priors for $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{\theta} \in \mathbb{R}^p$:

$$\pi_{\text{pri}}(\boldsymbol{x}\,|\,\delta) = \left(\frac{\delta}{2}\right)^n \exp\big(-\delta\,(\|(\boldsymbol{I} \otimes \boldsymbol{D})\,\boldsymbol{x}\|_1 + \|(\boldsymbol{D} \otimes \boldsymbol{I})\,\boldsymbol{x}\|_1)\big)$$

$$\pi_{\text{pri}}(\boldsymbol{\theta}\,|\,\kappa) = \left(\frac{1}{2\pi I_0(\kappa)}\right)^p \exp\left(\kappa\,\mathbf{1}^{\mathrm{T}}\cos(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})\right), \qquad \delta,\ \kappa = \text{hyperparameters,}$$

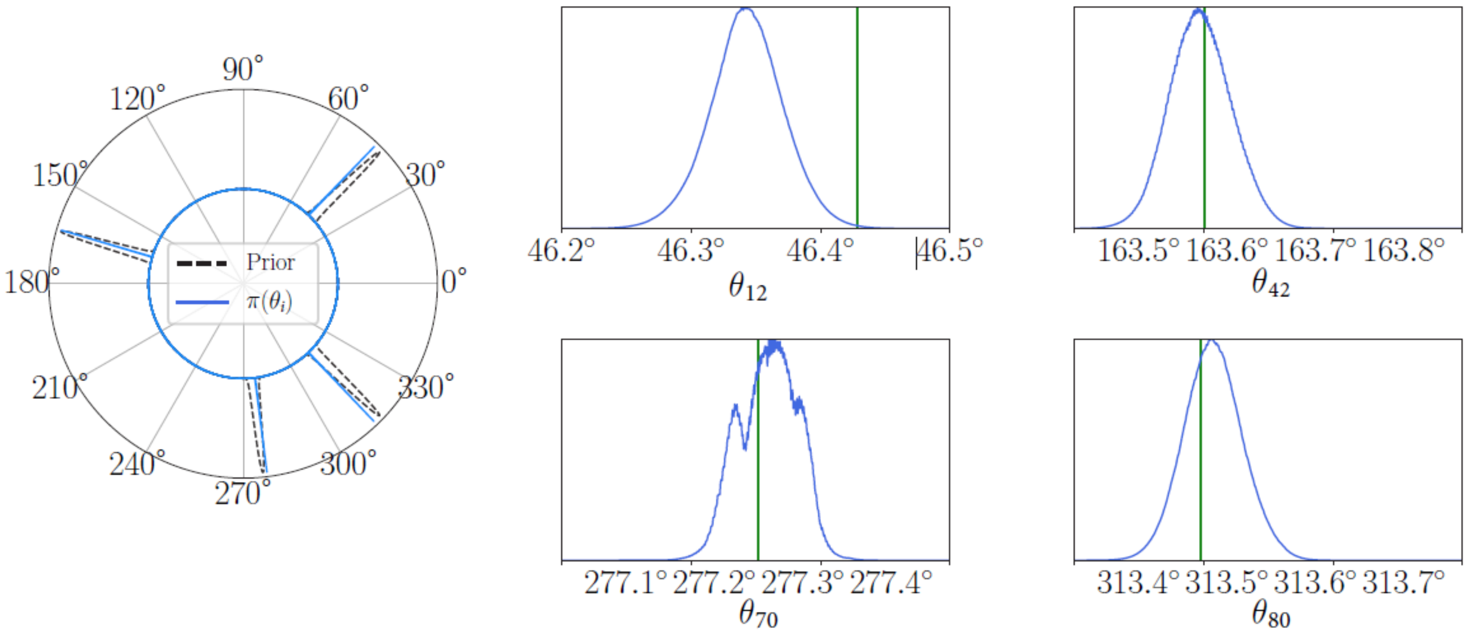
where I_0 = 0-order modified Bessel function, $\boldsymbol{D}$ = bidiag$(-1, 1)$, and $\bar{\boldsymbol{\theta}}$ = nominal angles.

Sampling the image pixels and view angles (see paper for details):

$$\pi(\boldsymbol{x}\,|\,\boldsymbol{\theta}, \lambda, \delta) \propto \exp\left(-\frac{\lambda}{2}\|\boldsymbol{A}(\boldsymbol{\theta})\,\boldsymbol{x} - \boldsymbol{b}\|_2^2 - \delta(\|(\boldsymbol{I} \otimes \boldsymbol{D})\,\boldsymbol{x}\|_1 + \|(\boldsymbol{D} \otimes \boldsymbol{I})\,\boldsymbol{x}\|_1)\right)$$

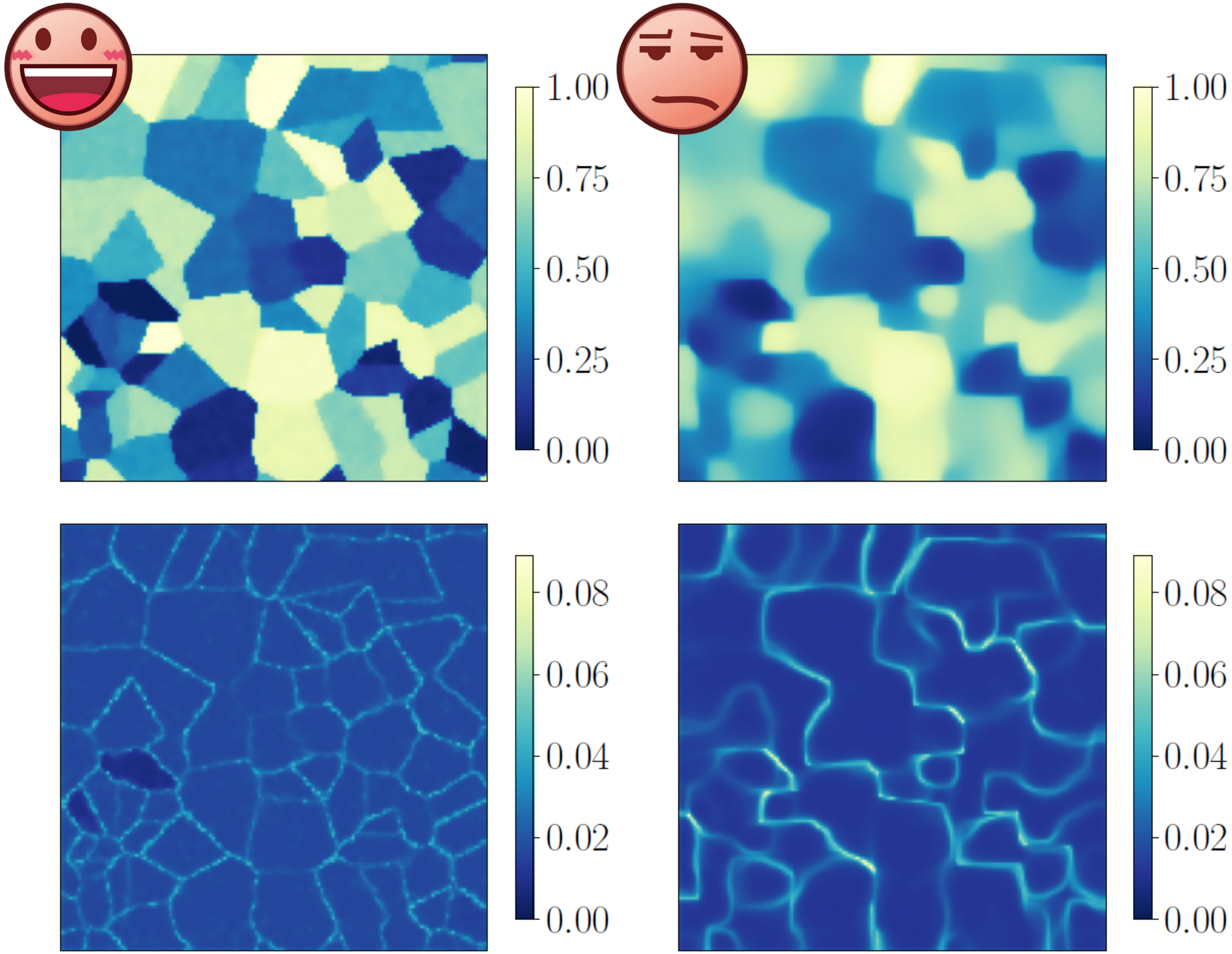
$$\pi(\boldsymbol{\theta}\,|\,\boldsymbol{x}, \lambda, \kappa) \propto \exp\left(-\frac{\lambda}{2}\|\boldsymbol{A}(\boldsymbol{\theta})\,\boldsymbol{x} - \boldsymbol{b}\|_2^2 + \kappa\mathbf{1}^{\mathrm{T}}\cos(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})\right)$$

Simulation Results – View Angles



Left: von Mises prior with the respective densities for selected angles in $\boldsymbol{\theta}$.
Right: some component densities and true angles shown as vertical green lines.

Simulation Results – Metallic Grains Phantom



Left: our method.

Right: using the incorrect nominal angles.

Top: posterior mean

Bottom: st. dev.

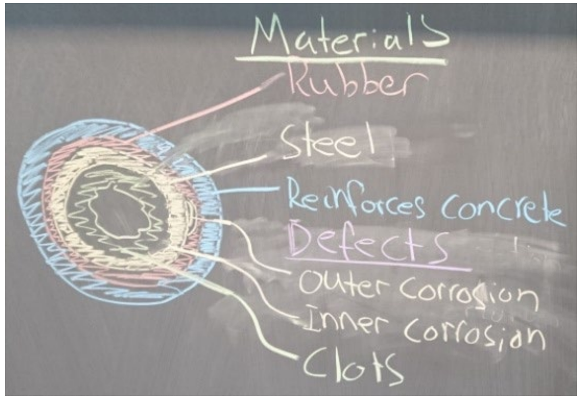
Nominal angles give a blurry image with uncertain boundaries.

We compute a sharper image with uncertainty confined to pixels on grain boundaries.

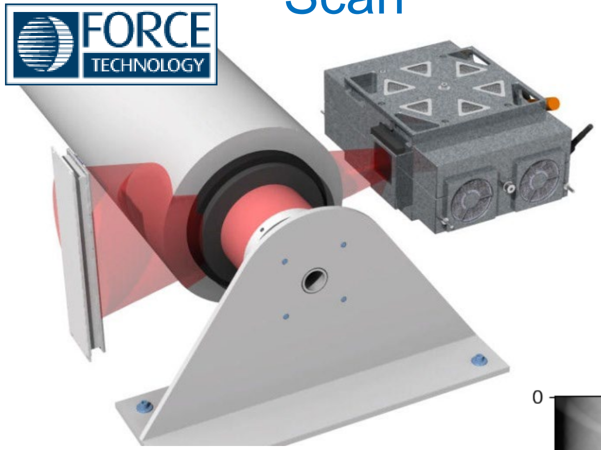
B: Structural Priors for Oil/Gas Pipes

Christensen, Riis, Pereyra, Jørgensen (2023)

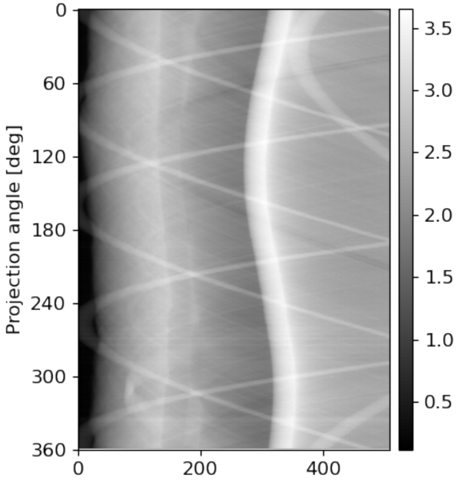
X-ray CT → cross-sectional images of oil/gas pipes on the seabed.
Detect *defects*, *cracks*, etc. in the pipe (expensive to repair).



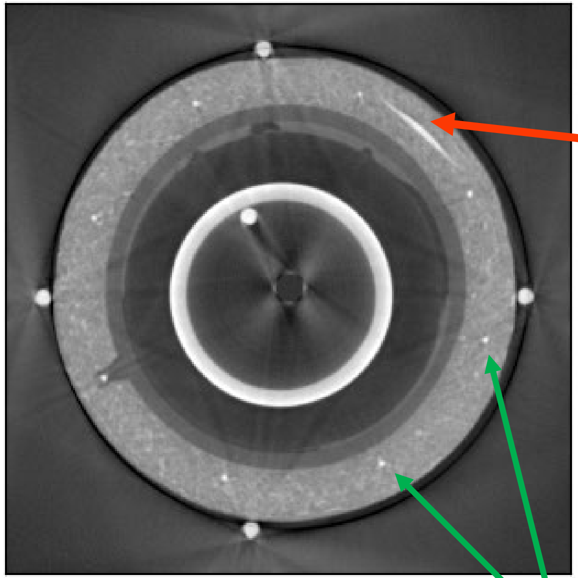
Scan



Data



Reconstruction



Defect! *How much can we trust the size and location?*

Reinforcing bars

Structural Prior for the Pipes

Model:
$\boldsymbol{y} = \boldsymbol{A}\,(\boldsymbol{z}+\boldsymbol{d}) + \text{noise}$
and $\boldsymbol{x} = \boldsymbol{z}+\boldsymbol{d}$.

Prior for $\boldsymbol{z}$ represents the layered, circular structure.

Prior for $\boldsymbol{d}$ represents small "spots" of random shape.

What is new in this work:

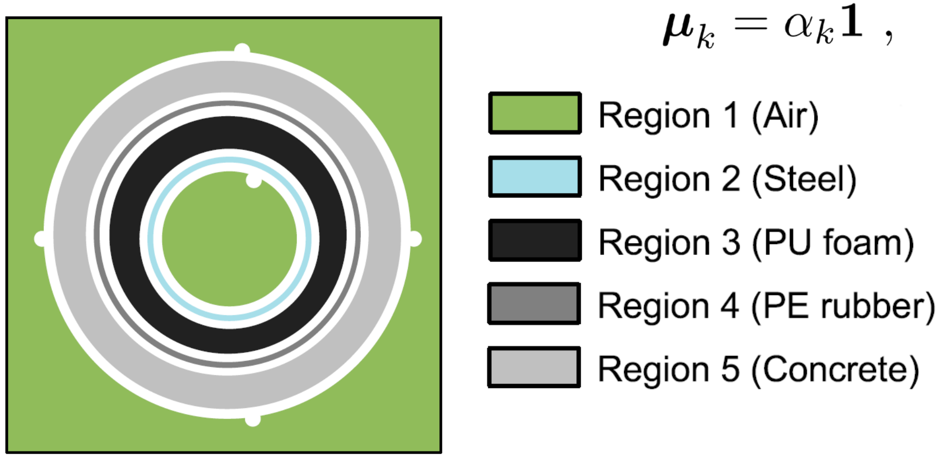
- Priors that capture completely different geometric features.
- A prior especially suited for sparse solutions with structure.

The structural prior captures the annular structure of the pipe. It is Gaussian:

$$\boldsymbol{z} \sim \mathcal{N}(\,\boldsymbol{\mu}\,,\,\boldsymbol{C}\,)\ , \qquad \boldsymbol{\mu} = \boldsymbol{C}\sum_{k=1}^{5}\boldsymbol{M}_k\,\boldsymbol{\mu}_k\ , \qquad \boldsymbol{C} = \left(\sum_{k=1}^{5}\boldsymbol{M}_k\right)^{-1}$$

in which

$$\boldsymbol{\mu}_k = \alpha_k \mathbf{1}\ , \qquad \boldsymbol{M}_k = \rho_k\,\mathrm{diag}(\boldsymbol{m}_k)\ , \qquad [\boldsymbol{m}_k]_j = \begin{cases} 1 & \text{if Pixel } j \in \text{Region } k \\ 0 & \text{otherwise} \end{cases}$$



Here, in region k: α_k is the unknown attenuation coefficient,
ρ_k^{-1} is the variance,
$\boldsymbol{m}_k$ defines the region's "mask.".

Prior for the Defects (thanks, Marcelo)

The key idea is to use a prior that promotes a defect image $\boldsymbol{d}$ that is sparse with small and spatially coherent structures (it has small "lumps" of nonzeros).

This is achieved with a *hidden gamma Markov random field* (Altmann, Pereyra, McLaughlin 2015) in the form of a Gaussian distribution with zero mean and a spatially varying variance that

1. pushes pixel values of $\boldsymbol{d}$ towards zero, and at the same time
2. has regions where the variance is large and where the posterior does not "feel" the prior.

The defects can occur in the regions with large variance. The details:

$$\boldsymbol{d} = \mathrm{vec}(\boldsymbol{\Delta}) \ , \qquad \boldsymbol{\Delta} = \{\delta_{ij}\} \ , \qquad \delta_{ij} \,|\, s_{ij} \sim \mathcal{N}(0, s_{ji})$$

$$\boldsymbol{S} = \{s_{ij}\} \ , \qquad s_{ij} \,|\, \boldsymbol{W} \sim \mathcal{IG}\big(\omega \, , \, \omega g_{ij}(\boldsymbol{W})\big) \qquad \text{(inverse Gamma distrib.)}$$

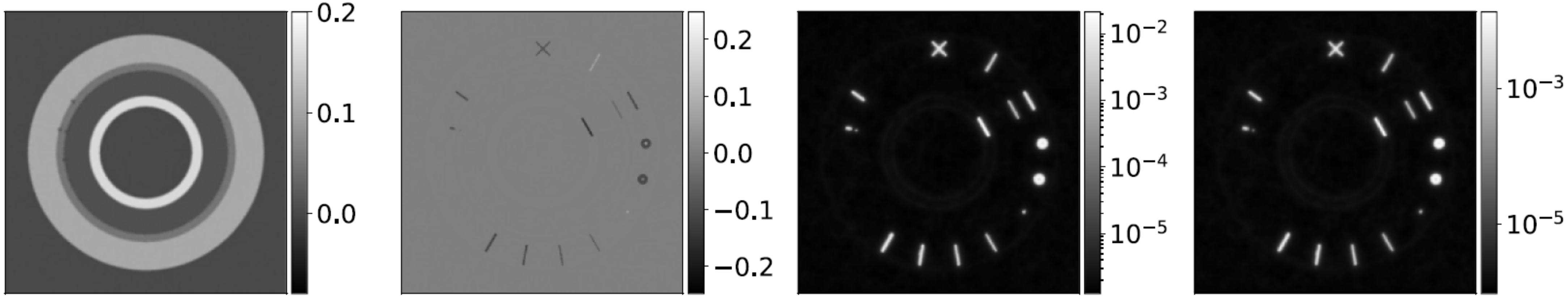
$$\boldsymbol{W} = \{w_{ij}\} \ , \qquad w_{ij} \,|\, \boldsymbol{S} \sim \mathcal{G}\big(\omega \, , \, (\omega h_{ij}(\boldsymbol{S}))^{-1}\big) \qquad \text{(Gamma distrib.)}$$

$$g_{ij} = {}^1\!/_4 \, (w_{ij} + w_{i+1,j} + w_{i,j+1} + w_{i+1,j+1})$$

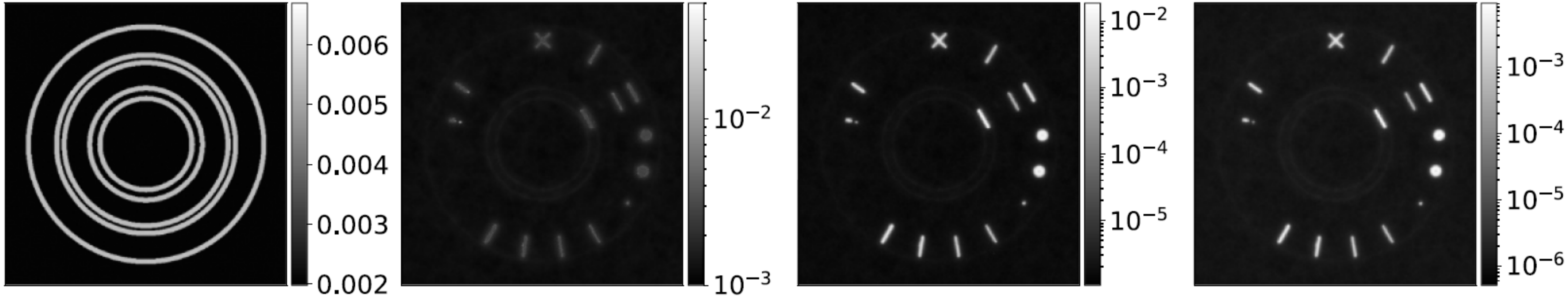
$$h_{ij} = {}^1\!/_4 \, (s_{ij}^{-1} + s_{i-1,j}^{-1} + s_{i,j-1}^{-1} + s_{i-1,j-1}^{-1})$$

Note: $\boldsymbol{S}$ is heavy tailed for small ω, with correlation between neighbour elements controlled by $\boldsymbol{W}$.

Simulation Results, 360 View Angles

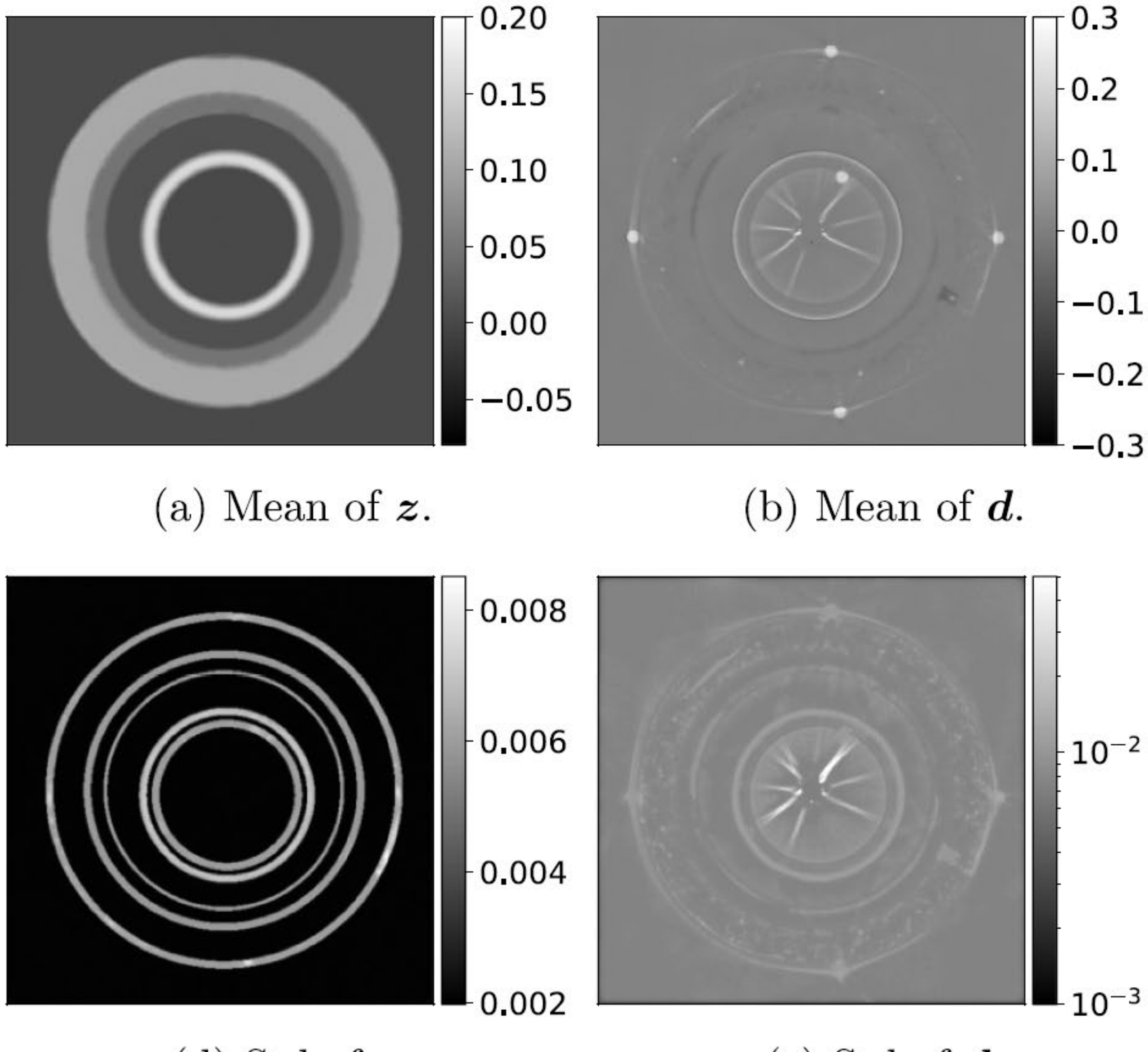


(a) Mean of $\boldsymbol{z}$. (b) Mean of $\boldsymbol{d}$. (c) Mean of $\boldsymbol{S}$. (d) Mean of $\boldsymbol{W}$



(e) Std of $\boldsymbol{z}$. (f) Std of $\boldsymbol{d}$. (g) Std of $\boldsymbol{S}$. (h) Std of $\boldsymbol{W}$

Real Data, 360 View Angles



(a) Mean of $\boldsymbol{z}$.

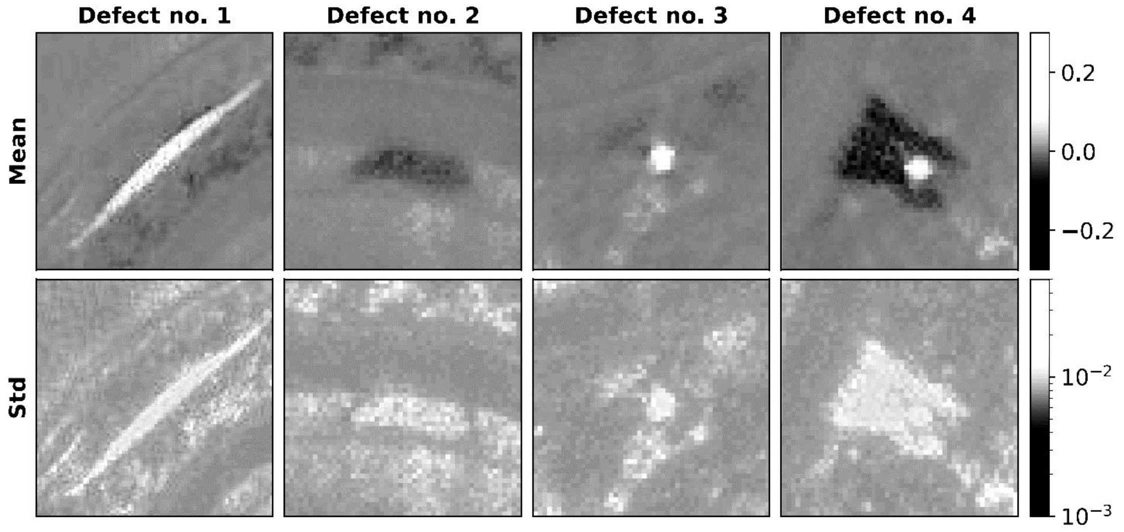
(b) Mean of $\boldsymbol{d}$.

(c) Mean of $\boldsymbol{z} + \boldsymbol{d}$.

(d) Std of $\boldsymbol{z}$.

(e) Std of $\boldsymbol{d}$.

ZOOM



C: The Horseshoe Prior for Edge-Preservation

Uribe, Dong, Hansen (2023).

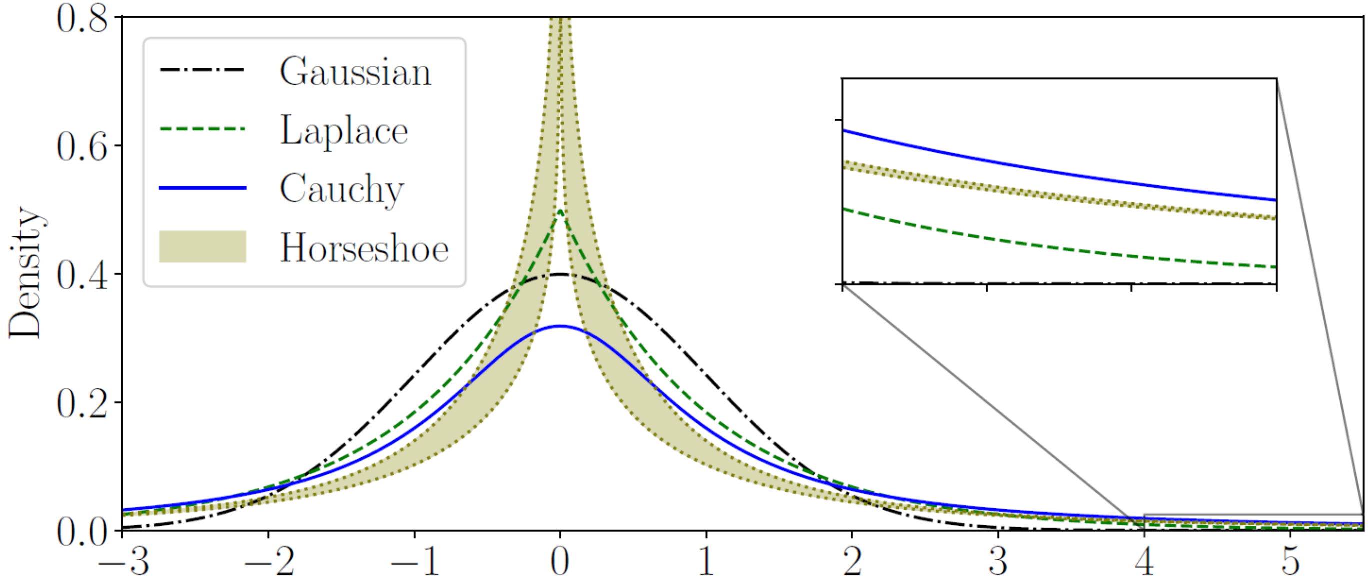
We often prefer heavy-tailed priors that promote *sharp edges*, such as

- the Cauchy or Laplace distribution of the difference between neighbor pixels.

Unfortunately, these priors are computationally demanding.

The *horseshoe* prior, which resembles the Cauchy and Laplace priors, is a computationally attractive alternative.

Why the name "horseshoe"?
→ Appendix.



The horseshoe prior's density is guaranteed to lie in the shaded band.

Defining The Horseshoe Prior

The standard horseshoe prior is *conditionally Gaussian*:

$$\pi(\boldsymbol{x}) \propto \exp\left(-{}^1\!/_2\, \boldsymbol{x}^T \boldsymbol{\Sigma}(\tau,\boldsymbol{\sigma})\, \boldsymbol{x}\right) \;, \qquad \boldsymbol{\Sigma}(\tau,\boldsymbol{\sigma}) = \tau^2\, \mathrm{diag}(\boldsymbol{\sigma}^2)$$

with hyperparameters τ (global shrinkage) and $\boldsymbol{\sigma}$ (local shrinkage):

$$\pi(\tau) \propto \frac{1}{\tau_0 \left(1 + \tau^2/\tau_0^2\right)} \;, \qquad \pi(\boldsymbol{\sigma}) \propto \prod_{i=1}^{n} \frac{1}{1+\sigma_i^2} \;, \qquad \tau_0 = \text{scale parameter.}$$

The horseshoe prior on pixel differences, with $\boldsymbol{x} = \mathrm{vec}(N \times N \text{ image})$:

$$\pi(\boldsymbol{x})_{\mathrm{dif}} \propto \exp\left(-{}^1\!/_2\, \boldsymbol{x}^T \boldsymbol{\Lambda}(\tau,\boldsymbol{w})\, \boldsymbol{x}\right)$$

where

$$\boldsymbol{\Lambda}(\tau,\boldsymbol{w}) = \boldsymbol{D}^T \begin{pmatrix} \boldsymbol{W}(\tau,\boldsymbol{w}) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{W}(\tau,\boldsymbol{w}) \end{pmatrix} \boldsymbol{D} \;, \qquad \boldsymbol{W}(\tau,\boldsymbol{w}) = \mathrm{diag}(\tau^2 \boldsymbol{w}^2)^{-1}$$

$$\boldsymbol{D} = \begin{pmatrix} \boldsymbol{I}_N \otimes \boldsymbol{D}_N \\ \boldsymbol{D}_N \otimes \boldsymbol{I}_N \end{pmatrix}, \qquad \boldsymbol{D}_N = \mathrm{bidiag}(-1,1) \;.$$

Next step: make τ and $\boldsymbol{w}$ hyperparameters $\rightarrow$ next slide.

And Now: Horseshoe With Hyperparameters

In the precision matrix $\boldsymbol{\Lambda}(\tau, \boldsymbol{w})$, we make τ and $\boldsymbol{w}$ hyperparameters ($\mathcal{IG}$ = inverse Gamma):

$$\pi(\tau^2 \,|\, \gamma) = \mathcal{IG}\big(1/2, 1/\gamma\big) \ , \quad \pi(\gamma) = \mathcal{IG}\big(1/2, 1/\tau_0\big) \ , \qquad \pi(w_i^2 \,|\, \xi_i) = \mathcal{IG}\big(1/2, 1/\xi_i\big) \ , \quad \pi(\xi_i) = \mathcal{IG}\big(1/2, 1\big)$$

With this formulation, the horseshow prior yields a Cauchy prior on $\boldsymbol{x}$ (see paper).

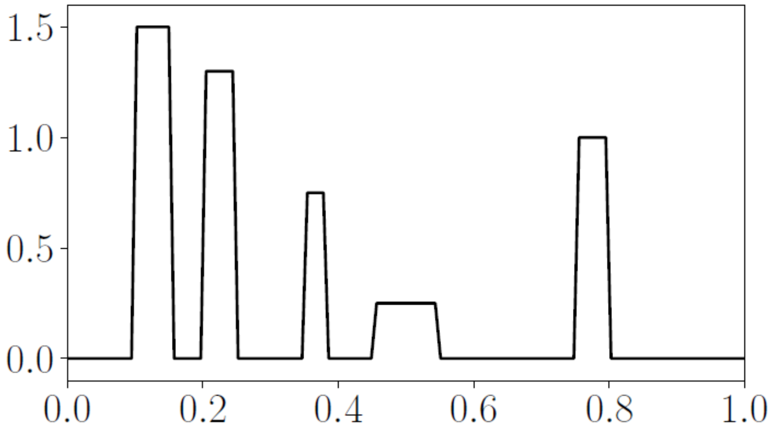
We use a *Gibbs sampler* with these steps:

1. Sample $\boldsymbol{x} \sim \exp\left(-\frac{1}{2}\left(\|\boldsymbol{A}\,\boldsymbol{x} - \boldsymbol{b}\|_2^2 + \|\boldsymbol{\Lambda}^{1/2}\boldsymbol{x}\|_2^2\right)\right)$ using the iterative least-squares solver CGLS.
2. Sample $\tau \sim \mathcal{IG}\left(\frac{n+1}{2}, \frac{1}{2}\sum_{i=1}^{n} \frac{[\boldsymbol{D}\boldsymbol{x}]_i^2}{w_i^2} + \frac{1}{\gamma}\right)$
3. Sample $w_i \sim \mathcal{IG}\left(1, \frac{1}{2}\sum_{i=1}^{n} \frac{[\boldsymbol{D}\boldsymbol{x}]_i^2}{2\tau^2} + \frac{1}{\xi_i}\right), \quad i = 1, 2, \ldots, n$
4. Sample $\gamma \sim \mathcal{IG}\left(1, \frac{1}{\tau_0^2} + \frac{1}{\tau^2}\right)$
5. Sample $\xi_i \sim \mathcal{IG}\left(1, 1 + \frac{1}{w_i^2}\right), \quad i = 1, 2, \ldots, n$

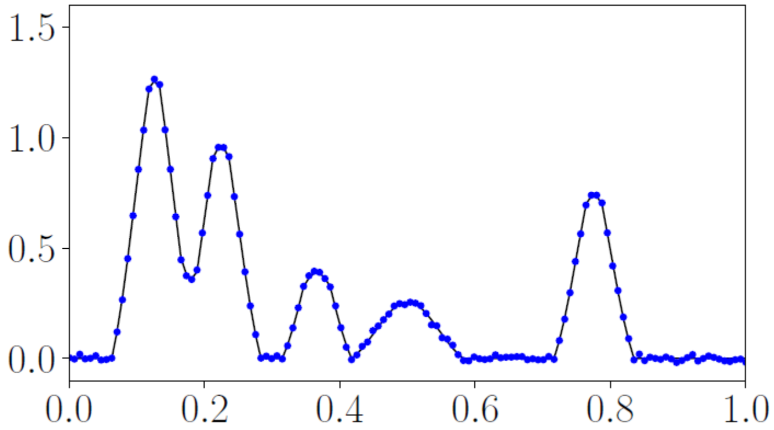
} In closed form

Example: 1D Deconvolution

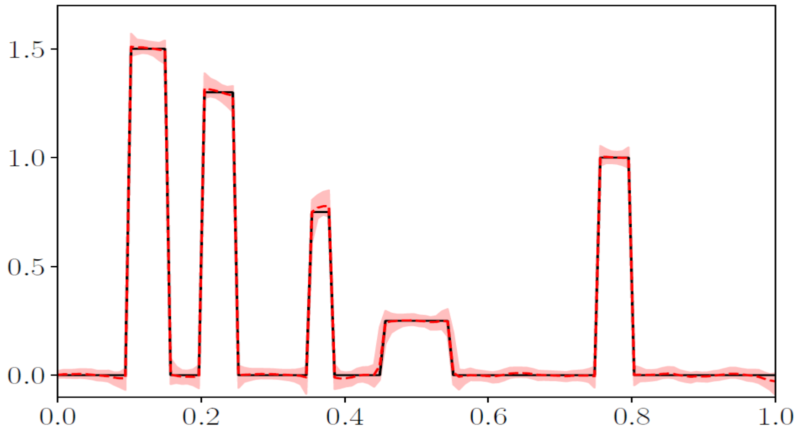
n = 128 and 2% Gaussian noise.



Ground truth



Noisy data

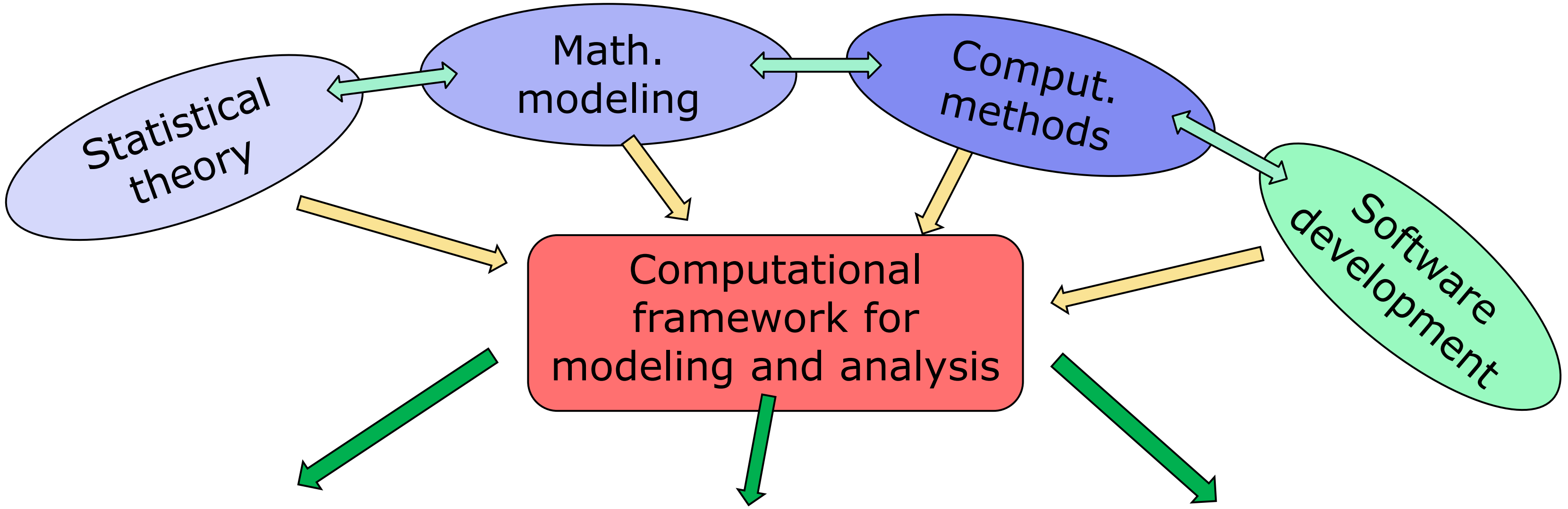


Posterior median (dashed) and 95% CI.

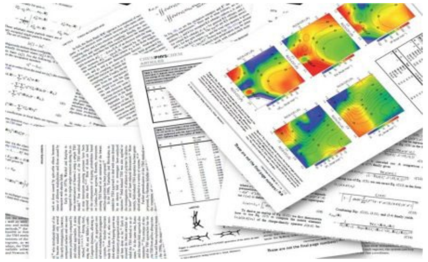
What is new in this work:

- Utilization of Gaussian prior sampled via least squares methods.
- Hyperpriors that can be sampled analytically.
- This allows us to use a Gibbs sampler.

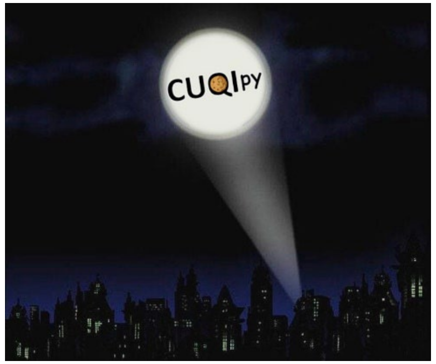
Putting It All Together



Papers



Software



Talks and courses

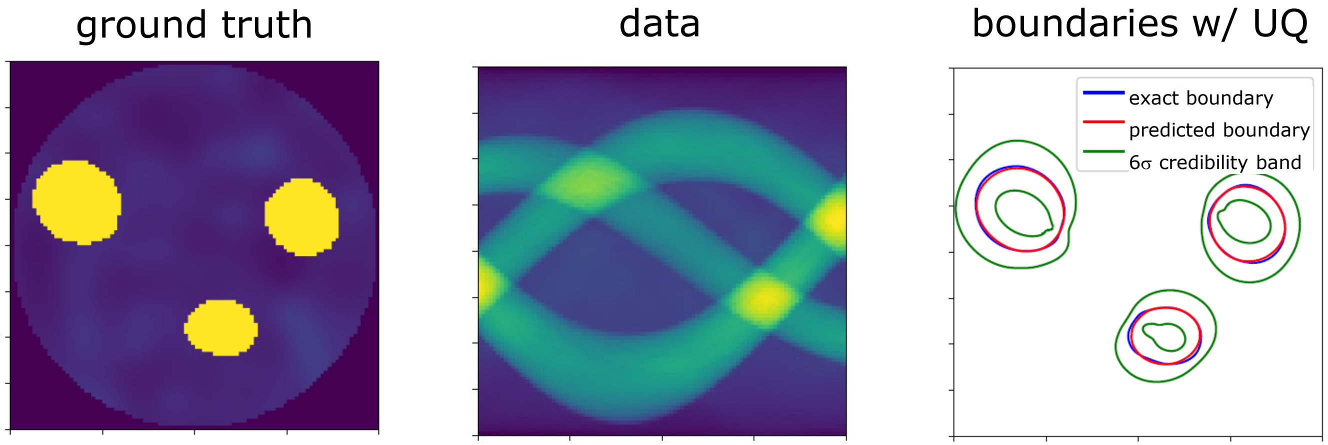
Goal-Oriented UQ in X-Ray CT

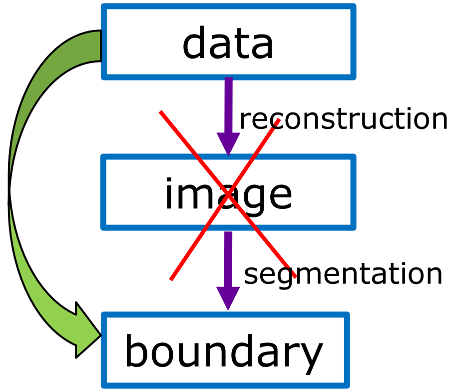
Afkham, Dong, Hansen (2023).
Afkham, Riis, Dong, Hansen (2024).

Statistical theory — Math. modeling — Comput. methods — Software development

Reconstruct the desired quantity directly from data and perform UQ on this quantity.

Example in CT:
Compute tumor boundaries and their regularity.

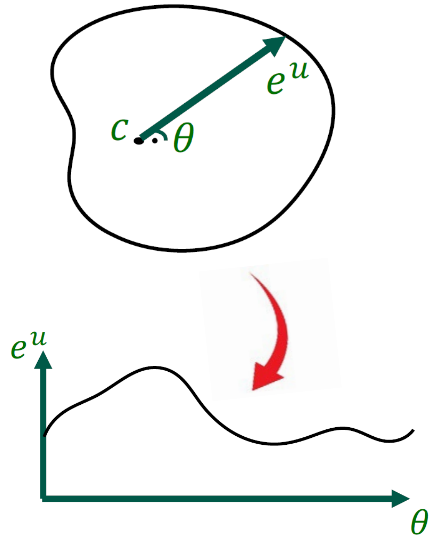




What is new in this work:

- *2D → 1D* computational problem, no pixels, no error accumulation.
- Represent the inclusion boundaries as *random-field functions*.
- Assign a hyper-parameter that controls the boundary's *regularity*.
- Perform UQ by assigning *probabilities* to the functions and their regularity.

Some Details



Model. Define a center for the inclusion, and define its boundary by a *radial function*

$$\text{radius}(\theta) = r_0 \exp\big(u(\theta)\big) \ , \qquad \theta \in [0, 2\pi) \ , \qquad u = \text{periodic function}$$

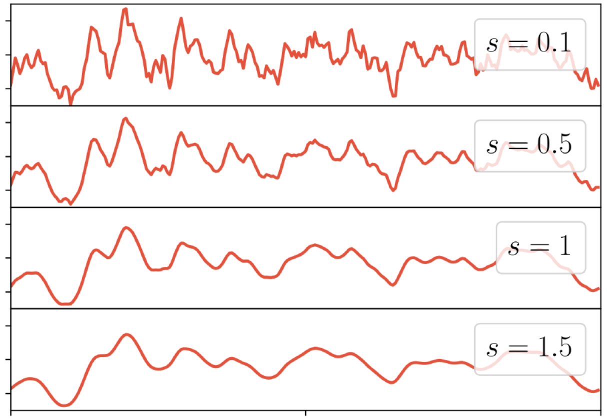
$$u(\theta) = \sum_{j=1}^{k} \alpha_j \sin(j\theta) + \beta_j \cos(j\theta) \ .$$

Prior. Whittle-Matérn prior for u → write the expansion coefficients as:

$$\alpha_j = v_{2j-1} \, (\sigma + j^2)^{(s+1/2)} \ , \quad \beta_j = v_{2j} \, (\sigma + j^2)^{(s+1/2)} \ , \qquad \sigma = \text{length scale.}$$

The coefficients v_i are zero-mean random Gaussian.

The parameter s characterizes the *roughness* of the function.



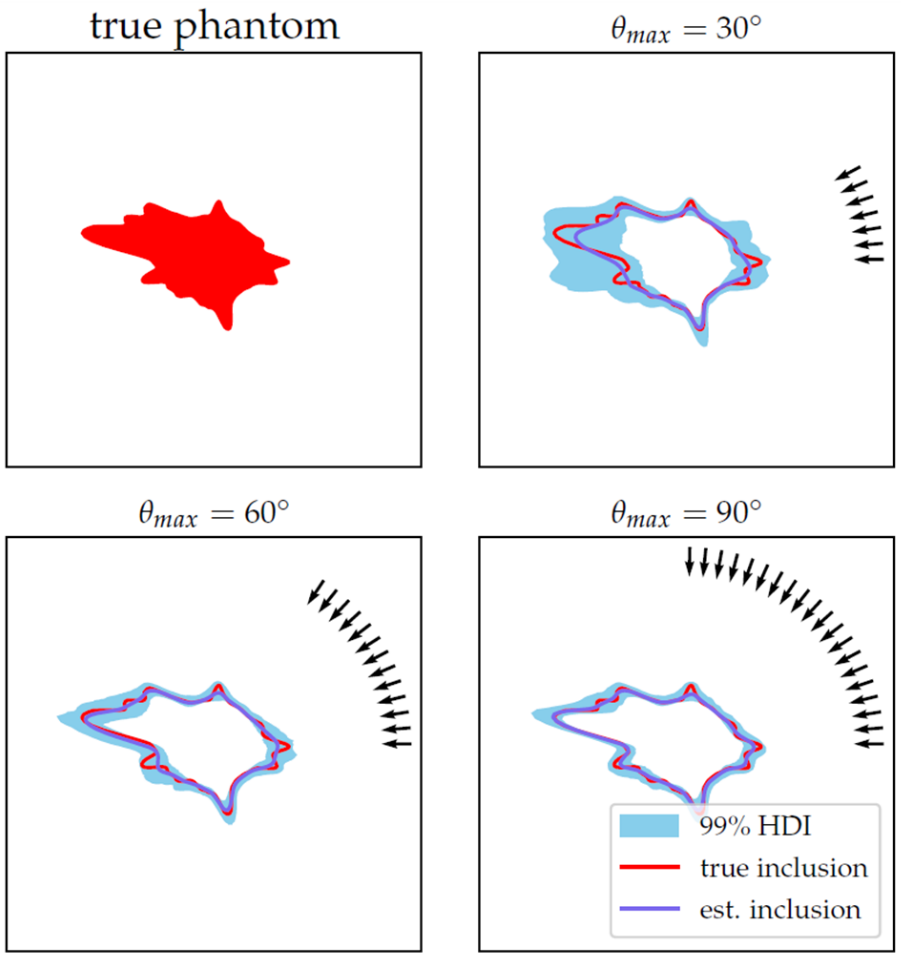
Sampling. Use a Gibbs sampler for the posterior $\pi(\boldsymbol{v}, s \,|\, \boldsymbol{y})$, with $\boldsymbol{y}$ = sinogram.

For $i = 1, 2, \ldots$

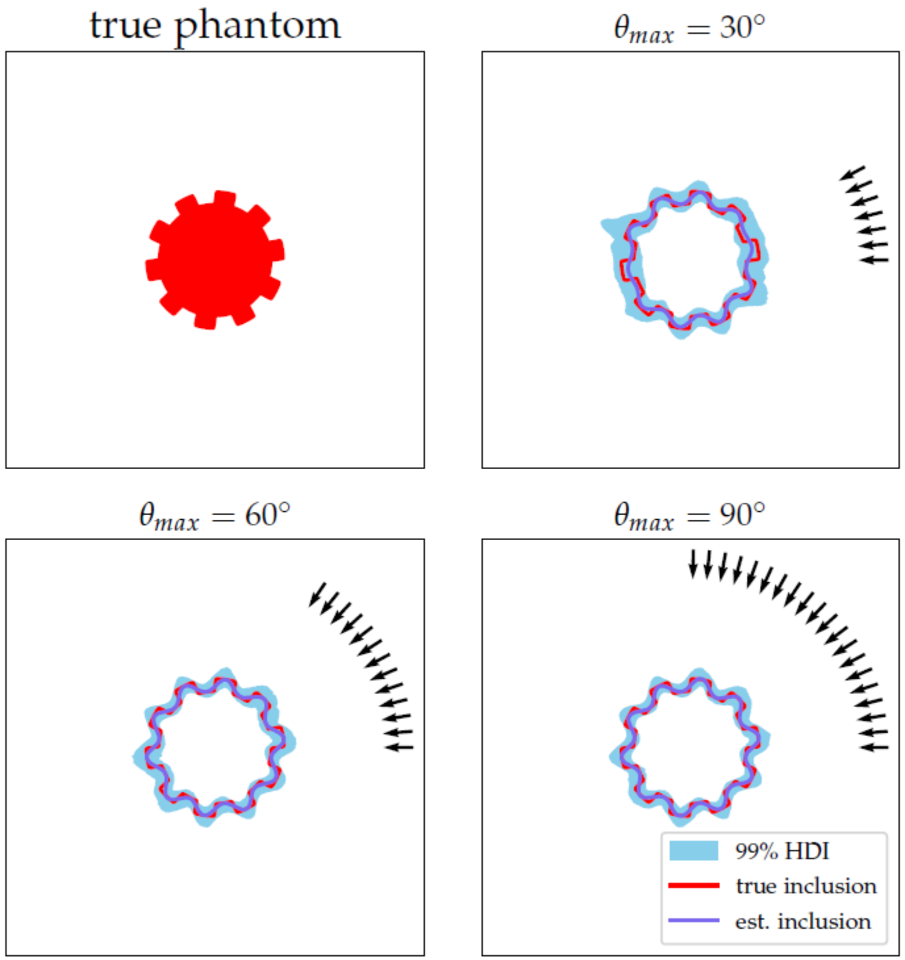
1. Use preconditioned Crank-Nicolson to sample $\pi(\boldsymbol{v} \,|\, \boldsymbol{y}, s_i) \to \boldsymbol{v}_{i+1}$
2. Use Metropolis-Hasting to sample $\pi(s \,|\, \boldsymbol{y}, \boldsymbol{v}_{i+1}) \to s_{i+1}$

Simulation Example

The boundary's credible interval depends on the angular span of the X-rays.



This phantom belongs to the prior.



This phantom does not belong to the prior.

Signing Out

Talks/posters about stuff not covered in this talk:

- Implicit priors and their interpretation → next talk (Jasper).
- UQ for a PDE-based inverse problem → next next talk (Amal).
- Comparison of RTO and Langevin sampling → talk Wednesday (Rémi).
- CUQIpy → posters Tuesday (Jakob) and Thursday (Andreea & Naoki).
- Random media and passive measurements → talk Thursday (Faouzi Triki, incl. work by Kristoffer).

More CUQI stuff not presented here:

- CT with uncertain geometry or uncertain flat field (Frederik, Jakob, Katrine, Martin)
- UQ in EIT, MREIT, and acousto-electric tomography (Aksel, Amal, Kim)
- Steerable photonic nanojet design with UQ (Amal, Mirza)
- Bayesian approach to inverse Robin problems (Aksel)
- Sampling conditioned on functionals (Lara, Mirza)
- Regularized system identification (Martin)
- Dimensionality challenges (Rafael, Yiqiu)
- Large-scale computational UQ (Charlie)
- Machine learning and UQ (Babak)
- Besov priors (Andreas, Yiqiu)

Why "Horseshoe"?

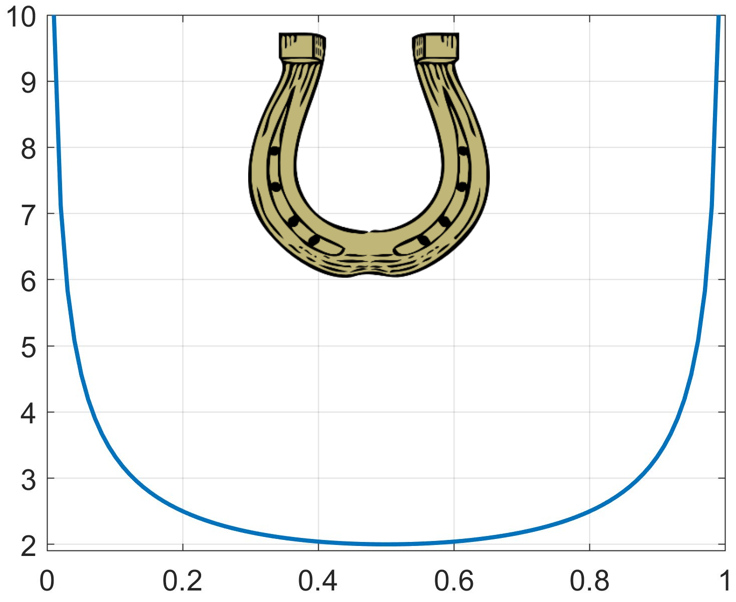
Recall that

$$\pi(\sigma_i) \propto \frac{1}{1+\sigma_i^2} \, .$$

If we introduce the *shrinkage parameter* $\kappa_i = 1/(1+\sigma_i^2)$ then we have the density

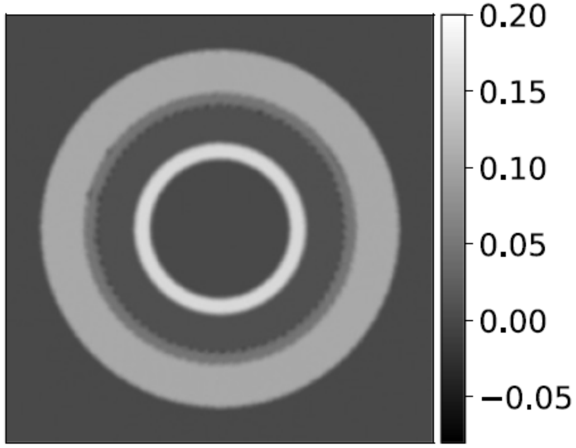
$$\pi(\kappa_i) \propto \frac{1}{\sqrt{\kappa_i(1-\kappa_i)}}$$

which (perhaps) has a horseshoe shape.

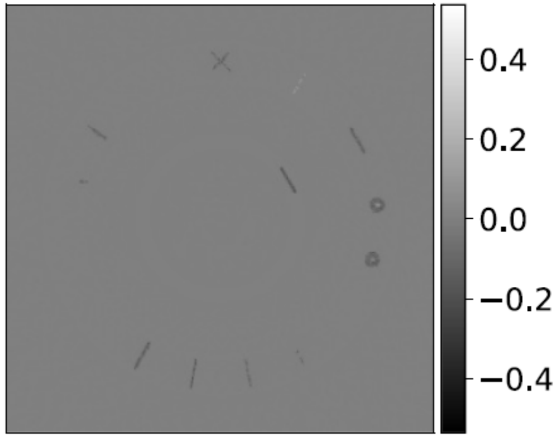


Note that $\pi(\kappa_i)$ is large for $\kappa_i \approx 1$ which corresponds to $\sigma_i \to 0$ which provides a lot of shrinkage.

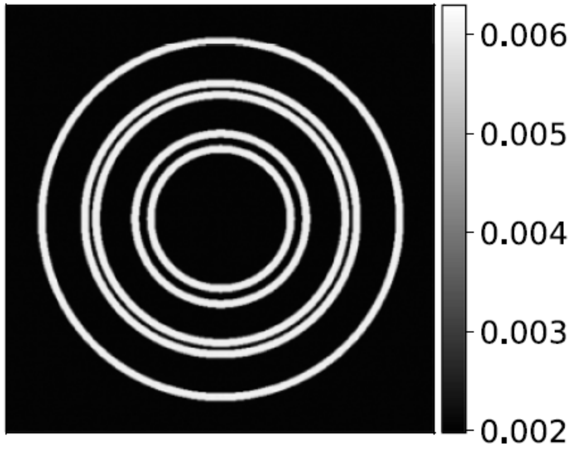
Appendix: Simulation Results, 72 View Angles



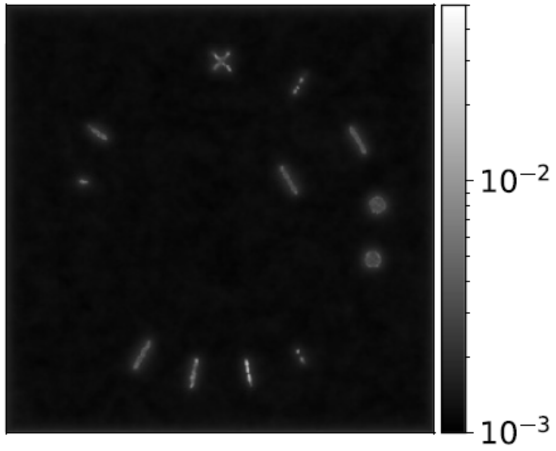
(a) Mean of $\boldsymbol{z}$.



(b) Mean of $\boldsymbol{d}$.



(d) Std of $\boldsymbol{z}$.



(e) Std of $\boldsymbol{d}$.