Scalable Bayesian uncertainty quantification with learned convex regularisers for radio interferometric imaging

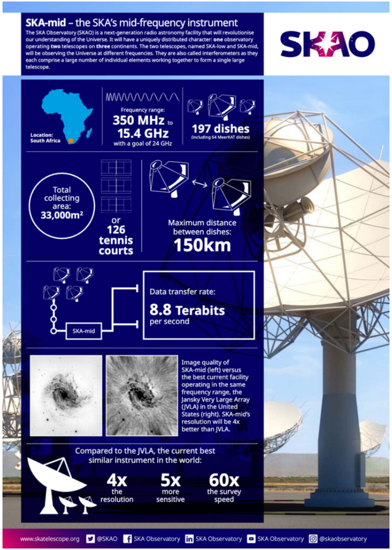
Tobías I. Liaudat
IRFU, CEA Paris-Saclay

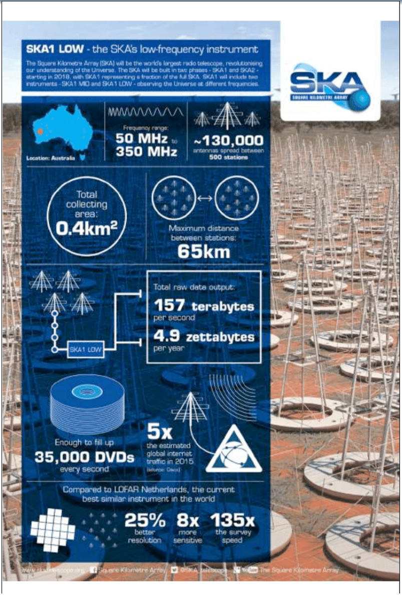
In collabortaion with Jason D. McEwen, Marcelo Pereyra and Marta Betcke

Statistical Challenges in 21st Century Cosmology, Chania

22nd May 2024

Motivation: SKA's radio interferometer





Radio interferometric imaging

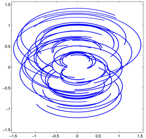
Linear observational model

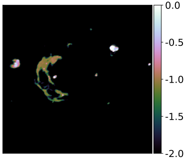
$$\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{n}$$

- $\mathbf{y} \in \mathbb{C}^M$: Observed Fourier coefficients
- $\mathbf{n} \in \mathbb{C}^M$: Observational noise (White and Gaussian)
- $\mathbf{x} \in \mathbb{R}^N$: Sky intensity image
- $\mathbf{\Phi} \in \mathbb{C}^{M \times N}$: Linear measurement operator
  - FFT and Fourier mask

Due to $\mathbf{n}$ and $\mathbf{\Phi}$ the inverse problem is ill-posed

We need to estimate $\hat{\mathbf{x}}$ from $\mathbf{y}$

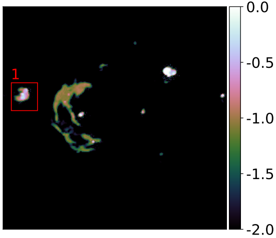




$\hat{\mathbf{x}}$

Uncertainty quantification: more than a point estimate

Image reconstruction: $\hat{\mathbf{x}}$



Is this blob *physical*?

→ Is it a reconstruction artefact?

→ Is it backed by the data?

Several reasons to develop uncertainty quantification (UQ) techniques for the reconstruction

Usual UQ techniques from the Bayesian framework rely on interrogating the posterior exploiting Bayes' theorem:

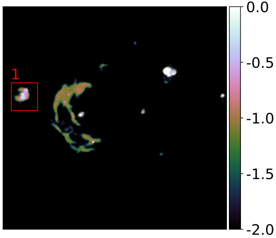
$$\underbrace{p(\mathbf{x}|\mathbf{y})}_{\text{Posterior}} \propto \underbrace{p(\mathbf{y}|\mathbf{x})}_{\text{Likelihood}} \underbrace{p(\mathbf{x})}_{\text{Prior}}$$

Represent the posterior through samples drawn from $\sim p(\mathbf{x}|\mathbf{y})$ obtained through a Markov chain Monte Carlo (MCMC) alg.

For example, Cai et al. (2018a) applies this for radio imaging

Uncertainty quantification: more than a point estimate

Image reconstruction: $\hat{\mathbf{x}}$



Is this blob *physical*?

→ Is it a reconstruction artefact?

→ Is it backed by the data?

Several reasons to develop uncertainty quantification (UQ) techniques for the reconstruction

Usual UQ techniques from the Bayesian framework rely on interrogating the posterior exploiting Bayes' theorem:

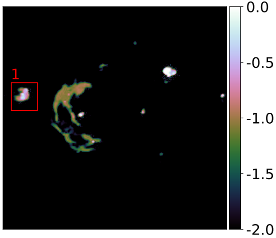
$$\underbrace{p(\mathbf{x}|\mathbf{y})}_{\text{Posterior}} \propto \underbrace{p(\mathbf{y}|\mathbf{x})}_{\text{Likelihood}} \underbrace{p(\mathbf{x})}_{\text{Prior}}$$

Represent the posterior through samples drawn from $\sim p(\mathbf{x}|\mathbf{y})$ obtained through a Markov chain Monte Carlo (MCMC) alg.

For example, Cai et al. (2018a) applies this for radio imaging

Uncertainty quantification: more than a point estimate

Image reconstruction: $\hat{\mathbf{x}}$



Is this blob *physical*?

→ Is it a reconstruction artefact?

→ Is it backed by the data?

Several reasons to develop uncertainty quantification (UQ) techniques for the reconstruction

Usual UQ techniques from the Bayesian framework rely on interrogating the posterior exploiting Bayes' theorem:

$$\underbrace{p(\mathbf{x}|\mathbf{y})}_{\text{Posterior}} \propto \underbrace{p(\mathbf{y}|\mathbf{x})}_{\text{Likelihood}} \underbrace{p(\mathbf{x})}_{\text{Prior}}$$

Represent the posterior through samples drawn from $\sim p(\mathbf{x}|\mathbf{y})$ obtained through a Markov chain Monte Carlo (MCMC) alg.

For example, Cai et al. (2018a) applies this for radio imaging

Uncertainty quantification: more than a point estimate

Cai et al. (2018a) approach:

1. Define a likelihood $p(\mathbf{y}|\mathbf{x}) = \exp[-f(\mathbf{x},\mathbf{y})]$
   - → The Gaussian likelihood $f(\mathbf{x},\mathbf{y})$ is known: $\|\mathbf{y} - \mathbf{\Phi x}\|_2^2 / 2\sigma^2$
2. Define a prior $p(\mathbf{x}) = \exp[-g(\mathbf{x})]$
   - → Solution $\mathbf{x}$ is sparse in a wavelet dictionary $\mathbf{\Psi}$. The prior $g(\mathbf{x})$ is: $\lambda\|\mathbf{\Psi}^\dagger x\|_1$
   - → CLEAN example: the prior $g_{\mathrm{CL}}(\mathbf{x})$ is $\lambda\|x\|_0$
3. Choose a point estimate
   - → Use the Maximum-a-posteriori (MAP) estimation:

$$\hat{\mathbf{x}}_{\mathrm{MAP}} = \underset{\mathbf{x}\in\mathbb{R}^N}{\arg\max}\, p(\mathbf{x}|\mathbf{y}) = \underset{\mathbf{x}\in\mathbb{R}^N}{\arg\min}\, \|\mathbf{y} - \mathbf{\Phi x}\|_2^2/2\sigma^2 + \overbrace{\lambda\|\mathbf{\Psi}^\dagger x\|_1}^{\text{CLEAN: } \lambda\|x\|_0},$$

   - → Estimate $\hat{\mathbf{x}}_{\mathrm{MAP}}$ through **convex optimisation** using a **proximal algorithm**
4. Sample from the posterior which is non-smooth to obtain $\{\mathbf{x}^{(j)}\}_{j=1}^K$, $\mathbf{x}^{(j)} \sim p(\mathbf{x}|\mathbf{y})$
   - → **Proximal MCMC algorithm** (Pereyra, 2016) following Langevin dynamics

Is the problem solved?

Uncertainty quantification: more than a point estimate

Cai et al. (2018a) approach:

1. Define a likelihood $p(\mathbf{y}|\mathbf{x}) = \exp[-f(\mathbf{x}, \mathbf{y})]$
   → The Gaussian likelihood $f(\mathbf{x}, \mathbf{y})$ is known: $\|\mathbf{y} - \mathbf{\Phi x}\|_2^2 / 2\sigma^2$
2. Define a prior $p(\mathbf{x}) = \exp[-g(\mathbf{x})]$
   → Solution $\mathbf{x}$ is sparse in a wavelet dictionary $\mathbf{\Psi}$. The prior $g(\mathbf{x})$ is: $\lambda\|\mathbf{\Psi}^\dagger x\|_1$
   → CLEAN example: the prior $g_{\mathrm{CL}}(\mathbf{x})$ is $\lambda\|x\|_0$
3. Choose a point estimate
   → Use the Maximum-a-posteriori (MAP) estimation:

$$\hat{\mathbf{x}}_{\mathrm{MAP}} = \underset{\mathbf{x}\in\mathbb{R}^N}{\arg\max}\, p(\mathbf{x}|\mathbf{y}) = \underset{\mathbf{x}\in\mathbb{R}^N}{\arg\min}\, \|\mathbf{y} - \mathbf{\Phi x}\|_2^2/2\sigma^2 + \overbrace{\lambda\|\mathbf{\Psi}^\dagger x\|_1}^{\text{CLEAN: } \lambda\|x\|_0},$$

   → Estimate $\hat{\mathbf{x}}_{\mathrm{MAP}}$ through **convex optimisation** using a **proximal algorithm**
4. Sample from the posterior which is non-smooth to obtain $\{\mathbf{x}^{(j)}\}_{j=1}^{K}$, $\mathbf{x}^{(j)} \sim p(\mathbf{x}|\mathbf{y})$
   → **Proximal MCMC algorithm** (Pereyra, 2016) following Langevin dynamics

Is the problem solved?

Uncertainty quantification: more than a point estimate

Cai et al. (2018a) approach:

1. Define a likelihood $p(\mathbf{y}|\mathbf{x}) = \exp[-f(\mathbf{x}, \mathbf{y})]$
   → The Gaussian likelihood $f(\mathbf{x}, \mathbf{y})$ is known: $\|\mathbf{y} - \mathbf{\Phi x}\|_2^2/2\sigma^2$
2. Define a prior $p(\mathbf{x}) = \exp[-g(\mathbf{x})]$
   → Solution $\mathbf{x}$ is sparse in a wavelet dictionary $\mathbf{\Psi}$. The prior $g(\mathbf{x})$ is: $\lambda\|\mathbf{\Psi}^\dagger x\|_1$
   → CLEAN example: the prior $g_{\mathrm{CL}}(\mathbf{x})$ is $\lambda\|x\|_0$
3. Choose a point estimate
   → Use the Maximum-a-posteriori (MAP) estimation:

$$\hat{\mathbf{x}}_{\mathrm{MAP}} = \underset{\mathbf{x}\in\mathbb{R}^N}{\arg\max}\, p(\mathbf{x}|\mathbf{y}) = \underset{\mathbf{x}\in\mathbb{R}^N}{\arg\min}\, \|\mathbf{y} - \mathbf{\Phi x}\|_2^2/2\sigma^2 + \overbrace{\lambda\|\mathbf{\Psi}^\dagger x\|_1}^{\text{CLEAN: } \lambda\|x\|_0},$$

   → Estimate $\hat{\mathbf{x}}_{\mathrm{MAP}}$ through **convex optimisation** using a **proximal algorithm**
4. Sample from the posterior which is non-smooth to obtain $\{\mathbf{x}^{(j)}\}_{j=1}^K$, $\mathbf{x}^{(j)} \sim p(\mathbf{x}|\mathbf{y})$
   → **Proximal MCMC algorithm** (Pereyra, 2016) following Langevin dynamics

Is the problem solved?

Uncertainty quantification: more than a point estimate

Cai et al. (2018a) approach:

1. Define a likelihood $p(\mathbf{y}|\mathbf{x}) = \exp[-f(\mathbf{x}, \mathbf{y})]$

   $\rightarrow$ The Gaussian likelihood $f(\mathbf{x}, \mathbf{y})$ is known: $\|\mathbf{y} - \mathbf{\Phi x}\|_2^2 / 2\sigma^2$

2. Define a prior $p(\mathbf{x}) = \exp[-g(\mathbf{x})]$

   $\rightarrow$ Solution $\mathbf{x}$ is sparse in a wavelet dictionary $\mathbf{\Psi}$. The prior $g(\mathbf{x})$ is: $\lambda \|\mathbf{\Psi}^\dagger x\|_1$

   $\rightarrow$ CLEAN example: the prior $g_{\mathrm{CL}}(\mathbf{x})$ is $\lambda \|x\|_0$

3. Choose a point estimate

   $\rightarrow$ Use the Maximum-a-posteriori (MAP) estimation:

$$\hat{\mathbf{x}}_{\mathrm{MAP}} = \underset{\mathbf{x} \in \mathbb{R}^N}{\arg\max}\, p(\mathbf{x}|\mathbf{y}) = \underset{\mathbf{x} \in \mathbb{R}^N}{\arg\min}\, \|\mathbf{y} - \mathbf{\Phi x}\|_2^2 / 2\sigma^2 + \overbrace{\lambda \|\mathbf{\Psi}^\dagger x\|_1}^{\text{CLEAN: } \lambda \|x\|_0} \,,$$

   $\rightarrow$ Estimate $\hat{\mathbf{x}}_{\mathrm{MAP}}$ through **convex optimisation** using a **proximal algorithm**

4. Sample from the posterior which is non-smooth to obtain $\{\mathbf{x}^{(j)}\}_{j=1}^{K}$, $\mathbf{x}^{(j)} \sim p(\mathbf{x}|\mathbf{y})$

   $\rightarrow$ **Proximal MCMC algorithm** (Pereyra, 2016) following Langevin dynamics

Is the problem solved?

Uncertainty quantification: more than a point estimate

Cai et al. (2018a) approach:

1. Define a likelihood $p(\mathbf{y}|\mathbf{x}) = \exp[-f(\mathbf{x},\mathbf{y})]$

   $\rightarrow$ The Gaussian likelihood $f(\mathbf{x},\mathbf{y})$ is known: $\|\mathbf{y} - \mathbf{\Phi x}\|_2^2/2\sigma^2$

2. Define a prior $p(\mathbf{x}) = \exp[-g(\mathbf{x})]$

   $\rightarrow$ Solution $\mathbf{x}$ is sparse in a wavelet dictionary $\mathbf{\Psi}$. The prior $g(\mathbf{x})$ is: $\lambda\|\mathbf{\Psi}^\dagger x\|_1$
   $\rightarrow$ CLEAN example: the prior $g_{\mathrm{CL}}(\mathbf{x})$ is $\lambda\|x\|_0$

3. Choose a point estimate

   $\rightarrow$ Use the Maximum-a-posteriori (MAP) estimation:

$$\hat{\mathbf{x}}_{\mathrm{MAP}} = \underset{\mathbf{x}\in\mathbb{R}^N}{\arg\max}\, p(\mathbf{x}|\mathbf{y}) = \underset{\mathbf{x}\in\mathbb{R}^N}{\arg\min}\, \|\mathbf{y} - \mathbf{\Phi x}\|_2^2/2\sigma^2 + \overbrace{\lambda\|\mathbf{\Psi}^\dagger x\|_1}^{\text{CLEAN: } \lambda\|x\|_0} ,$$

   $\rightarrow$ Estimate $\hat{\mathbf{x}}_{\mathrm{MAP}}$ through **convex optimisation** using a **proximal algorithm**

4. Sample from the posterior which is non-smooth to obtain $\{\mathbf{x}^{(j)}\}_{j=1}^{K}$, $\mathbf{x}^{(j)} \sim p(\mathbf{x}|\mathbf{y})$

   $\rightarrow$ **Proximal MCMC algorithm** (Pereyra, 2016) following Langevin dynamics

Is the problem solved?

Uncertainty quantification: more than a point estimate

Cai et al. (2018a) approach:

1. Define a likelihood $p(\mathbf{y}|\mathbf{x}) = \exp[-f(\mathbf{x}, \mathbf{y})]$
   $\rightarrow$ The Gaussian likelihood $f(\mathbf{x}, \mathbf{y})$ is known: $\|\mathbf{y} - \mathbf{\Phi x}\|_2^2 / 2\sigma^2$
2. Define a prior $p(\mathbf{x}) = \exp[-g(\mathbf{x})]$
   $\rightarrow$ Solution $\mathbf{x}$ is sparse in a wavelet dictionary $\mathbf{\Psi}$. The prior $g(\mathbf{x})$ is: $\lambda \|\mathbf{\Psi}^\dagger x\|_1$
   $\rightarrow$ CLEAN example: the prior $g_{\mathrm{CL}}(\mathbf{x})$ is $\lambda \|x\|_0$
3. Choose a point estimate
   $\rightarrow$ Use the Maximum-a-posteriori (MAP) estimation:

$$\hat{\mathbf{x}}_{\mathrm{MAP}} = \underset{\mathbf{x} \in \mathbb{R}^N}{\arg\max}\, p(\mathbf{x}|\mathbf{y}) = \underset{\mathbf{x} \in \mathbb{R}^N}{\arg\min}\, \|\mathbf{y} - \mathbf{\Phi x}\|_2^2 / 2\sigma^2 + \overbrace{\lambda \|\mathbf{\Psi}^\dagger x\|_1}^{\text{CLEAN: } \lambda \|x\|_0} ,$$

   $\rightarrow$ Estimate $\hat{\mathbf{x}}_{\mathrm{MAP}}$ through **convex optimisation** using a **proximal algorithm**
4. Sample from the posterior which is non-smooth to obtain $\{\mathbf{x}^{(j)}\}_{j=1}^{K}$, $\mathbf{x}^{(j)} \sim p(\mathbf{x}|\mathbf{y})$
   $\rightarrow$ **Proximal MCMC algorithm** (Pereyra, 2016) following Langevin dynamics

Is the problem solved?

The problem is not solved

Difficulties in the high-dimensional setting:

1. Even if we know the likelihood, applying $\boldsymbol{\Phi}$ is **computationally expensive**
2. Handcrafted priors like wavelets are **not expressive enough**
3. Sampling-based techniques are **prohibitively expensive** in this setting

How can we obtain information from the high-dimensional posterior $p(\mathbf{x}|\mathbf{y})$ without sampling from it?

If we restrict to log-concave posteriors something beautiful happens!
$\rightarrow$ **A concentration phenomenom** (Pereyra, 2017)

log-concave posterior $p(\mathbf{x}|\mathbf{y}) = \exp[-f(\mathbf{x}) - g(\mathbf{x})]/Z \rightarrow$ convex potential $f(\mathbf{x}) + g(\mathbf{x})$

The problem is not solved

Difficulties in the high-dimensional setting:

1. Even if we know the likelihood, applying $\boldsymbol{\Phi}$ is **computationally expensive**
2. Handcrafted priors like wavelets are **not expressive enough**
3. Sampling-based techniques are **prohibitively expensive** in this setting

How can we obtain information from the high-dimensional posterior $p(\mathbf{x}|\mathbf{y})$ without sampling from it?

If we restrict to log-concave posteriors something beautiful happens!
$\rightarrow$ **A concentration phenomenom** (Pereyra, 2017)

log-concave posterior $p(\mathbf{x}|\mathbf{y}) = \exp[-f(\mathbf{x}) - g(\mathbf{x})]/Z \rightarrow$ convex potential $f(\mathbf{x}) + g(\mathbf{x})$

The problem is not solved

Difficulties in the high-dimensional setting:

1. Even if we know the likelihood, applying $\mathbf{\Phi}$ is **computationally expensive**
2. Handcrafted priors like wavelets are **not expressive enough**
3. Sampling-based techniques are **prohibitively expensive** in this setting

How can we obtain information from the high-dimensional posterior $p(\mathbf{x}|\mathbf{y})$ without sampling from it?

If we restrict to log-concave posteriors something beautiful happens!
→ **A concentration phenomenom** (Pereyra, 2017)

log-concave posterior $p(\mathbf{x}|\mathbf{y}) = \exp[-f(\mathbf{x}) - g(\mathbf{x})]/Z \rightarrow$ convex potential $f(\mathbf{x}) + g(\mathbf{x})$

The problem is not solved

Difficulties in the high-dimensional setting:

1. Even if we know the likelihood, applying $\mathbf{\Phi}$ is **computationally expensive**
2. Handcrafted priors like wavelets are **not expressive enough**
3. Sampling-based techniques are **prohibitively expensive** in this setting

How can we obtain information from the high-dimensional posterior $p(\mathbf{x}|\mathbf{y})$ without sampling from it?

If we restrict to log-concave posteriors something beautiful happens!
$\rightarrow$ **A concentration phenomenom** (Pereyra, 2017)

log-concave posterior $p(\mathbf{x}|\mathbf{y}) = \exp[-f(\mathbf{x}) - g(\mathbf{x})]/Z \rightarrow$ convex potential $f(\mathbf{x}) + g(\mathbf{x})$

The problem is not solved

Difficulties in the high-dimensional setting:

1. Even if we know the likelihood, applying $\mathbf{\Phi}$ is **computationally expensive**
2. Handcrafted priors like wavelets are **not expressive enough**
3. Sampling-based techniques are **prohibitively expensive** in this setting

How can we obtain information from the high-dimensional posterior $p(\mathbf{x}|\mathbf{y})$ without sampling from it?

If we restrict to log-concave posteriors something beautiful happens!
$\rightarrow$ **A concentration phenomenom** (Pereyra, 2017)

log-concave posterior $p(\mathbf{x}|\mathbf{y}) = \exp[-f(\mathbf{x}) - g(\mathbf{x})]/Z \rightarrow$ convex potential $f(\mathbf{x}) + g(\mathbf{x})$

The problem is not solved

Difficulties in the high-dimensional setting:

1. Even if we know the likelihood, applying $\mathbf{\Phi}$ is **computationally expensive**
2. Handcrafted priors like wavelets are **not expressive enough**
3. Sampling-based techniques are **prohibitively expensive** in this setting

How can we obtain information from the high-dimensional posterior $p(\mathbf{x}|\mathbf{y})$ without sampling from it?

If we restrict to log-concave posteriors something beautiful happens!
$\rightarrow$ **A concentration phenomenom** (Pereyra, 2017)

log-concave posterior $p(\mathbf{x}|\mathbf{y}) = \exp[-f(\mathbf{x}) - g(\mathbf{x})]/Z \rightarrow$ convex potential $f(\mathbf{x}) + g(\mathbf{x})$

Highest posterior density region

Posterior credible region:

$$p(\mathbf{x} \in C_\alpha | \mathbf{y}) = \int_{\mathbf{x} \in \mathbb{R}^N} p(\mathbf{x}|\mathbf{y}) \mathbb{1}_{C_\alpha} \mathrm{d}\mathbf{x} = 1 - \alpha,$$

We consider the highest posterior density (HPD) region

$$C^*_\alpha = \{\mathbf{x} : \underbrace{f(\mathbf{x}) + g(\mathbf{x})}_{\text{potential}} \leq \gamma_\alpha\}, \quad \text{with } \gamma_\alpha \in \mathbb{R}, \quad \text{and } p(\mathbf{x} \in C^*_\alpha | \mathbf{y}) = 1 - \alpha \text{ holds},$$

Theorem 3.1 (Pereyra, 2017)

Suppose the posterior $p(\mathbf{x}|\mathbf{y}) = \exp[-f(\mathbf{x}) - g(\mathbf{x})]/Z$ is log-concave on $\mathbb{R}^N$. Then, for any $\alpha \in (4\exp[(-N/3)], 1)$, the HPD region C^*_α is contained by

$$\hat{C}_\alpha = \left\{\mathbf{x} : f(\mathbf{x}) + g(\mathbf{x}) \leq \hat{\gamma}_\alpha = f(\hat{\mathbf{x}}_{\text{MAP}}) + g(\hat{\mathbf{x}}_{\text{MAP}}) + \sqrt{N}\tau_\alpha + N\right\},$$

with a positive constant $\tau_\alpha = \sqrt{16\log(3/\alpha)}$ independent of $p(\mathbf{x}|\mathbf{y})$.

We only need to evaluate $f + g$ on the MAP estimation $\hat{\mathbf{x}}_{\text{MAP}}$!

Highest posterior density region

Posterior credible region:

$$p(\mathbf{x} \in C_\alpha | \mathbf{y}) = \int_{\mathbf{x} \in \mathbb{R}^N} p(\mathbf{x}|\mathbf{y}) \mathbb{1}_{C_\alpha} \mathrm{d}\mathbf{x} = 1 - \alpha,$$

We consider the highest posterior density (HPD) region

$$C^*_\alpha = \{\mathbf{x} : \underbrace{f(\mathbf{x}) + g(\mathbf{x})}_{\text{potential}} \leq \gamma_\alpha\}, \quad \text{with } \gamma_\alpha \in \mathbb{R}, \quad \text{and } p(\mathbf{x} \in C^*_\alpha | \mathbf{y}) = 1 - \alpha \text{ holds},$$

Theorem 3.1 (Pereyra, 2017)

Suppose the posterior $p(\mathbf{x}|\mathbf{y}) = \exp[-f(\mathbf{x}) - g(\mathbf{x})]/Z$ is log-concave on $\mathbb{R}^N$. Then, for any $\alpha \in (4\exp[(-N/3)], 1)$, the HPD region C^*_α is contained by

$$\hat{C}_\alpha = \left\{\mathbf{x} : f(\mathbf{x}) + g(\mathbf{x}) \leq \hat{\gamma}_\alpha = f(\hat{\mathbf{x}}_{\text{MAP}}) + g(\hat{\mathbf{x}}_{\text{MAP}}) + \sqrt{N}\tau_\alpha + N\right\},$$

with a positive constant $\tau_\alpha = \sqrt{16 \log(3/\alpha)}$ independent of $p(\mathbf{x}|\mathbf{y})$.

We only need to evaluate $f + g$ on the MAP estimation $\hat{\mathbf{x}}_{\text{MAP}}$!

Highest posterior density region

Posterior credible region:

$$p(\mathbf{x} \in C_\alpha|\mathbf{y}) = \int_{\mathbf{x}\in\mathbb{R}^N} p(\mathbf{x}|\mathbf{y})\mathbb{1}_{C_\alpha}\mathrm{d}\mathbf{x} = 1-\alpha,$$

We consider the highest posterior density (HPD) region

$$C^*_\alpha = \{\mathbf{x} : \underbrace{f(\mathbf{x}) + g(\mathbf{x})}_{\text{potential}} \leq \gamma_\alpha\}, \quad \text{with } \gamma_\alpha \in \mathbb{R}, \quad \text{and } p(\mathbf{x} \in C^*_\alpha|\mathbf{y}) = 1-\alpha \text{ holds},$$

Theorem 3.1 (Pereyra, 2017)

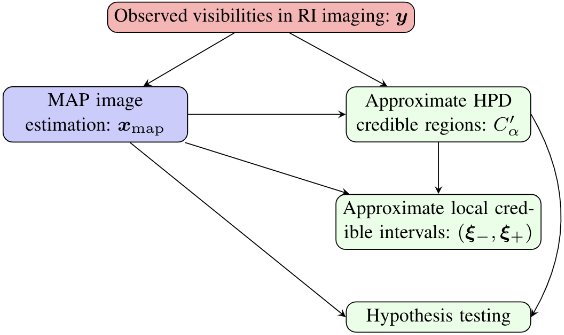
Suppose the posterior $p(\mathbf{x}|\mathbf{y}) = \exp[-f(\mathbf{x}) - g(\mathbf{x})]/Z$ is log-concave on $\mathbb{R}^N$. Then, for any $\alpha \in (4\exp[(-N/3)], 1)$, the HPD region C^*_α is contained by

$$\hat{C}_\alpha = \left\{\mathbf{x} : f(\mathbf{x}) + g(\mathbf{x}) \leq \hat{\gamma}_\alpha = f(\hat{\mathbf{x}}_{\text{MAP}}) + g(\hat{\mathbf{x}}_{\text{MAP}}) + \sqrt{N}\tau_\alpha + N\right\},$$

with a positive constant $\tau_\alpha = \sqrt{16\log(3/\alpha)}$ independent of $p(\mathbf{x}|\mathbf{y})$.

We only need to evaluate $f+g$ on the MAP estimation $\hat{\mathbf{x}}_{\text{MAP}}$!

MAP-based uncertainty quantification



Hypothesis test with significance α:

1. Calculate the MAP: $\mathbf{x}_{\mathrm{MAP}}$
2. Compute HPD region threshold $\hat{\gamma}_\alpha$
3. Construct a surrogate image $\mathbf{x}_{\mathrm{sgt}}$
4. Compute $\mathcal{E} = f(\mathbf{x}_{\mathrm{sgt}}) + g(\mathbf{x}_{\mathrm{sgt}})$
5. If $\mathcal{E} \leq \hat{\gamma}_\alpha \rightarrow$ inconclusive test
6. If $\mathcal{E} > \hat{\gamma}_\alpha \rightarrow$ reject hypothesis

Image credit: Cai et al. (2018b)

MAP-based uncertainty quantification: Pixel-level UQ visualisation

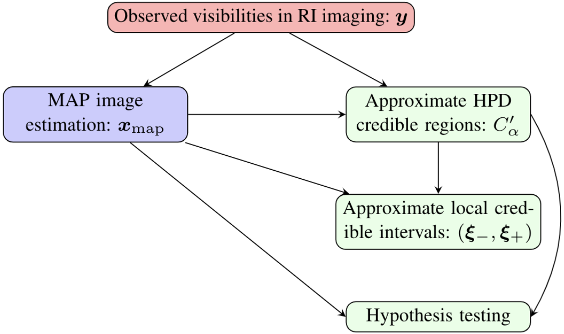


Image credit: Cai et al. (2018b)

Local credible intervals (LCI)

- Set of superpixels $\{\Omega_i\}_{i=1}^{M}$ such that $\Omega_i \cap \Omega_j = \emptyset, \forall i \neq j$ and $\Omega = \cup_i \Omega_i$ with Ω the set of all the image pixels
- One superpixel region: $\mathbf{x}_{i,\xi} = \hat{\mathbf{x}}_{\text{MAP}}(\boldsymbol{I} - \zeta_{\Omega_i}) + (\xi + \bar{\mathbf{x}}_{\text{MAP},\Omega_i})\zeta_{\Omega_i}$,
- Compute each bound (root-finding alg.): $\xi_{+,\Omega_i} = \max_\xi \{\xi \,|\, f(\mathbf{x}_{i,\xi}, \mathbf{y}) + g(\mathbf{x}_{i,\xi}) \leq \hat{\gamma}_\alpha,\, \xi \in [0, +\infty)\}$, $\xi_{-,\Omega_i} = \min_\xi \{\xi \,|\, f(\mathbf{x}_{i,\xi}, \mathbf{y}) + g(\mathbf{x}_{i,\xi}) \leq \hat{\gamma}_\alpha,\, \xi \in (-\infty, 0]\}$,
- Display the length of the intervals: $\boldsymbol{\xi} = \sum_i (\xi_{+,\Omega_i} - \xi_{-,\Omega_i})\, \zeta_{\Omega_i}$

MAP-based uncertainty quantification: Pixel-level UQ visualisation

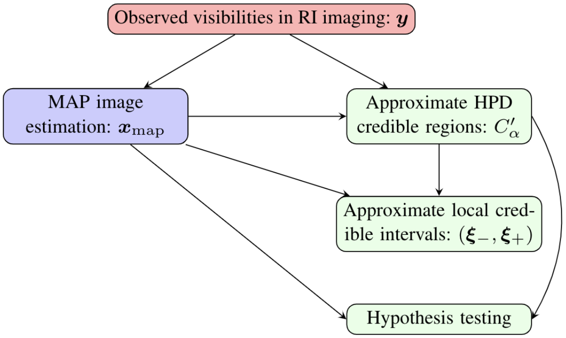


Image credit: Cai et al. (2018b)

Fast pixel-wise errors at different scales

- Multiscale wavelet decomposition of level J:
  $\hat{\mathbf{x}}_{\mathrm{MAP}} = \boldsymbol{\Psi}\, \hat{\mathbf{a}}_{\mathrm{MAP}} = \sum_{l=0}^{J} \boldsymbol{\Psi}_l\, \hat{\mathbf{a}}_{\mathrm{MAP},l}\,,$
- Threshold that saturates the HPD region:

$$\hat{\xi}_{\mathrm{th}} = \max_{\xi_{\mathrm{th}}} \{ \xi_{\mathrm{th}} \mid f(\hat{\mathbf{x}}_{\mathrm{MAP},\,\xi_{\mathrm{th}}}, \mathbf{y}) + g(\hat{\mathbf{x}}_{\mathrm{MAP},\,\xi_{\mathrm{th}}}) \leq \hat{\gamma}_\alpha,$$

$$\hat{\mathbf{x}}_{\mathrm{MAP},\,\xi_{\mathrm{th}}} = \boldsymbol{\Psi}\, S_{\mathrm{hard},\,\xi_{\mathrm{th}}}(\hat{\mathbf{a}}_{\mathrm{MAP}}),\ \xi \in [0, +\infty) \}\,.$$

  Compute using a root-finding algorithm

- Thresholded surrogate image at level j:
  $\hat{\mathbf{x}}_{\mathrm{MAP},\,\hat{\xi}_{\mathrm{th}},\,j} = \sum_{l=0,\, l\neq j}^{J} \boldsymbol{\Psi}_l\, \hat{\mathbf{a}}_{\mathrm{MAP},l} + \boldsymbol{\Psi}_j \hat{\mathbf{a}}_{\mathrm{MAP},\,\hat{\xi}_{\mathrm{th}},\,j}\,,$
- Approximated error at level e_j:
  $e_j = \hat{\mathbf{x}}_{\mathrm{MAP}} - \hat{\mathbf{x}}_{\mathrm{MAP},\,\hat{\xi}_{\mathrm{th}},\,j}$

Scalable Bayesian uncertainty quantification

1. **Scalability** → Need to rely on optimisation ~~sampling~~, use the MAP estimator
2. **Uncertainty quantification** → Need the potential to be convex and explicit
3. **Good reconstruction** → Need to use data-driven (learned) approaches

The approach requires our prior to be convex and with an explicit potential

We constrain our prior to be convex, but we gain an effortless UQ!

1. **Scalability** → Need to rely on optimisation ~~sampling~~, use the MAP estimator
2. **Uncertainty quantification** → Need the potential to be convex and explicit
3. **Good reconstruction** → Need to use data-driven (learned) approaches

The approach requires our prior to be convex and with an explicit potential

We constrain our prior to be convex, but we gain an effortless UQ!

Scalable Bayesian uncertainty quantification

1. **Scalability** → Need to rely on optimisation ~~sampling~~, use the MAP estimator
2. **Uncertainty quantification** → Need the potential to be convex and explicit
3. **Good reconstruction** → Need to use data-driven (learned) approaches

The approach requires our prior to be convex and with an explicit potential

We constrain our prior to be convex, but we gain an effortless UQ!

Scalable Bayesian uncertainty quantification

1. **Scalability** → Need to rely on optimisation ~~sampling~~, use the MAP estimator
2. **Uncertainty quantification** → Need the potential to be convex and explicit
3. **Good reconstruction** → Need to use data-driven (learned) approaches

The approach requires our prior to be convex and with an explicit potential

We constrain our prior to be convex, but we gain an effortless UQ!

Scalable Bayesian uncertainty quantification

1. **Scalability** $\rightarrow$ Need to rely on optimisation ~~sampling~~, use the MAP estimator
2. **Uncertainty quantification** $\rightarrow$ Need the potential to be convex and explicit
3. **Good reconstruction** $\rightarrow$ Need to use data-driven (learned) approaches

The approach requires our prior to be convex and with an explicit potential

We constrain our prior to be convex, but we gain an effortless UQ!

Learned convex regulariser

We use the neural-network-based convex regulariser R from Goujon et al. (2023), where

$$R : \mathbb{R}^N \mapsto \mathbb{R}, \quad R(\mathbf{x}) = \sum_{n=1}^{N_C} \sum_{k} \psi_n \left(\left(\mathbf{h}_n * \mathbf{x} \right) [k] \right),$$

- ψ_n are learned convex profile functions with Lipschitz continuous derivate
  - Learnable 2nd degree splines
- There are N_C learned convolutional filters $\mathbf{h}_n$
- R is trained as a (multi-)gradient step denoiser

Properties:

1. **Explicit cost**
2. **Convex**
3. **Smooth regulariser with known Lipschitz constant**

Learned convex regulariser

We use the neural-network-based convex regulariser R from Goujon et al. (2023), where

$$R : \mathbb{R}^N \mapsto \mathbb{R}, \quad R(\mathbf{x}) = \sum_{n=1}^{N_C} \sum_k \psi_n \left((\mathbf{h}_n * \mathbf{x}) [k] \right),$$

- ψ_n are learned convex profile functions with Lipschitz continuous derivate
  - Learnable 2nd degree splines
- There are N_C learned convolutional filters $\mathbf{h}_n$
- R is trained as a (multi-)gradient step denoiser

Properties:

1. **Explicit cost**
2. **Convex**
3. **Smooth regulariser with known Lipschitz constant**

Numerical experiments

RI imaging models:

Model from Cai et al. (2018a): $\hat{\mathbf{x}}_{\text{MAP}} = \arg\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{y} - \mathbf{\Phi x}\|_2^2/2\sigma^2 + \lambda_1 \|\mathbf{\Psi}^\dagger \mathbf{x}\|_1 + \iota_{\mathbb{R}^N}(\mathbf{x})$,

Proposed model: $\hat{\mathbf{x}}_{\text{MAP}} = \arg\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{y} - \mathbf{\Phi x}\|_2^2/2\sigma^2 + \lambda_2/\mu R_{\boldsymbol{\theta}}(\mu \mathbf{x}) + \iota_{\mathbb{R}^N}(\mathbf{x})$,

MAP estimations are computed using the FISTA algorithm

Validation of the UQ is done by sampling both posterior distributions using a proximal MCMC algorithm, SK-ROCK (Pereyra et al., 2020)

Experiment settings:

- Image size 256×256
- Input SNR of 30dB
- Gridded Fourier sampling: 10% coverage from a Gaussian distribution ($M \approx 6.5 \times 10^3$)
- Wavelets used: Daubechies 8

Numerical experiments

RI imaging models:

Model from Cai et al. (2018a): $\hat{\mathbf{x}}_{\text{MAP}} = \arg\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2/2\sigma^2 + \lambda_1\|\mathbf{\Psi}^\dagger\mathbf{x}\|_1 + \iota_{\mathbb{R}^N}(\mathbf{x})$,

Proposed model: $\hat{\mathbf{x}}_{\text{MAP}} = \arg\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2/2\sigma^2 + \lambda_2/\mu R_{\boldsymbol{\theta}}(\mu\mathbf{x}) + \iota_{\mathbb{R}^N}(\mathbf{x})$,

MAP estimations are computed using the FISTA algorithm

Validation of the UQ is done by sampling both posterior distributions using a proximal MCMC algorithm, SK-ROCK (Pereyra et al., 2020)

Experiment settings:

- Image size 256×256
- Input SNR of 30dB
- Gridded Fourier sampling: 10% coverage from a Gaussian distribution ($M \approx 6.5 \times 10^3$)
- Wavelets used: Daubechies 8

Numerical experiments

RI imaging models:

Model from Cai et al. (2018a): $\hat{\mathbf{x}}_{\mathrm{MAP}} = \arg\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2/2\sigma^2 + \lambda_1\|\mathbf{\Psi}^\dagger\mathbf{x}\|_1 + \iota_{\mathbb{R}^N}(\mathbf{x})\,,$

Proposed model: $\hat{\mathbf{x}}_{\mathrm{MAP}} = \arg\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2/2\sigma^2 + \lambda_2/\mu R_{\boldsymbol{\theta}}(\mu\mathbf{x}) + \iota_{\mathbb{R}^N}(\mathbf{x})\,,$

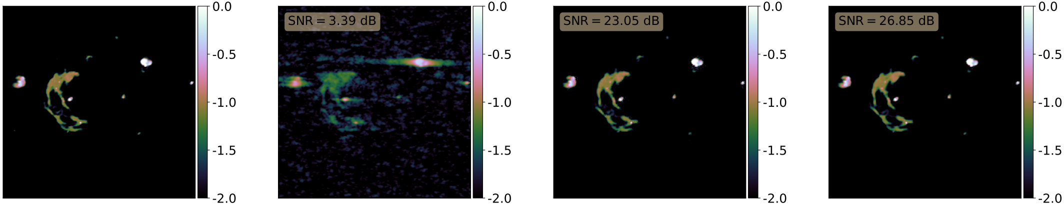
MAP estimations are computed using the FISTA algorithm

Validation of the UQ is done by sampling both posterior distributions using a proximal MCMC algorithm, SK-ROCK (Pereyra et al., 2020)

Experiment settings:

- Image size 256×256
- Input SNR of 30dB
- Gridded Fourier sampling: 10% coverage from a Gaussian distribution ($M \approx 6.5 \times 10^3$)
- Wavelets used: Daubechies 8

MAP reconstructions

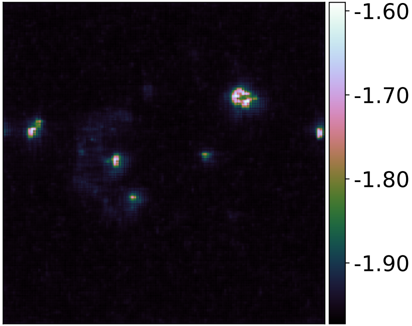


Ground truth	Dirty image	Wavelet prior	Learned prior
	SNR $=$ 3.39 dB	SNR $=$ 23.05 dB	**SNR $=$ 26.85 dB**

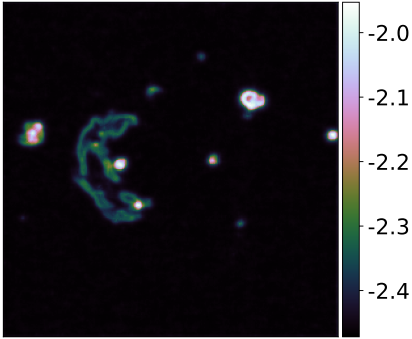
Improved the reconstruction by 3.8 dB

Posterior standard deviation

Computed using 10^4 samples obtained from the sampling algorithm SK-ROCK (Pereyra et al., 2020)



Wavelet



Learned regulariser

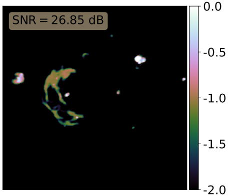
More meaningful uncertainties in the posterior Std Dev

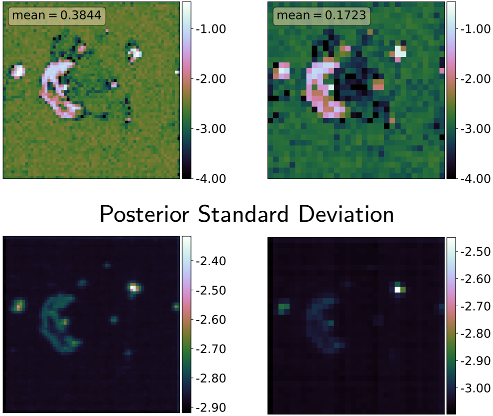
The learned convex regulariser was trained on natural images, not RI images

Pixel-based uncertainty quantification

The local credible intervals (LCI) give a local measure of uncertainty

LCI − < LCI >





Pixel size 4×4 Pixel size 8×8

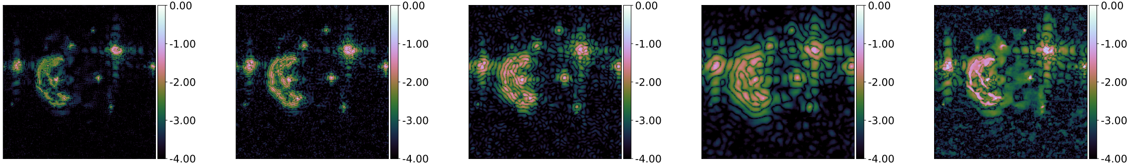
Computation time reduced by a factor of 10^3 wrt sampling

Fast pixel-wise errors at different scales

We can compute pixel-wise error at different scales using:

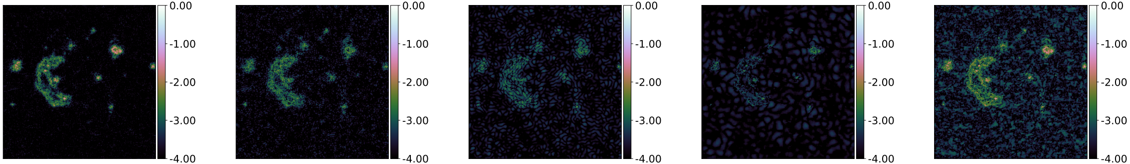
- A wavelet dictionary (e.g. Carrillo et al. (2012))
- The HPD region computed

Error prediction



Level 4 | Level 3 | Level 2 | Level 1 | All levels

True error



Level 4 | Level 3 | Level 2 | Level 1 | All levels

Minimizing the number of likelihood evaluations

Computation time reduced by a factor of 10^5 wrt sampling

Computing time and likelihood evaluations

Table: Computation wall-clock times for the W28 image in seconds.

Models	MAP optim.	Posterior sampling	LCIs 8×8	Fast pixel UQ
Wavelet-based	0.94	36.0×10^3	149.7	—
QUANTIFAI	0.64	6.44×10^3	108.2	**0.17**

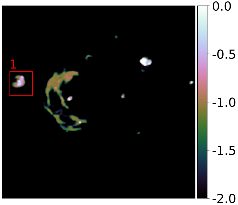
Table: The number of measurement operator evaluations used by the QUANTIFAI for the W28 image.

MCMC sampling	LCIs 8×8	LCIs 16×16	Fast pixel UQ
11×10^6	81.5×10^3	21.2×10^3	**28**

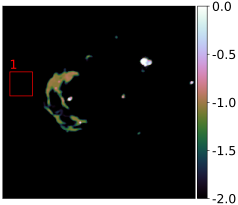
The fast pixel UQ is 10^6 and 10^3 times faster than the MCMC sampling and LCIs, respectively.

Hypothesis test

Scalable hypothesis testing for structure in the reconstruction

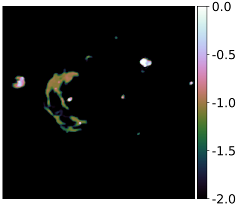


MAP reconstruction

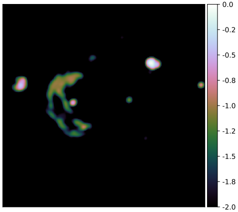


Inpainted surrogate

Is the blob physical? $\rightarrow$ Yes



MAP reconstruction



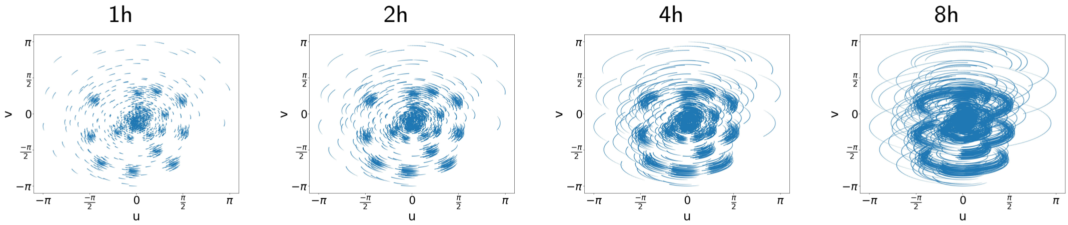
Blurred substructure

Is the substructure physical? $\rightarrow$ Yes

A more realistic experiment

Simulate single frequency MeerKAT ungridded visibility patterns

- Start frequency of 1400MHz with a channel width of 10MHz
- Pointing: J2000, RA=13h18m54.86s, DEC=-15d36m04.25s

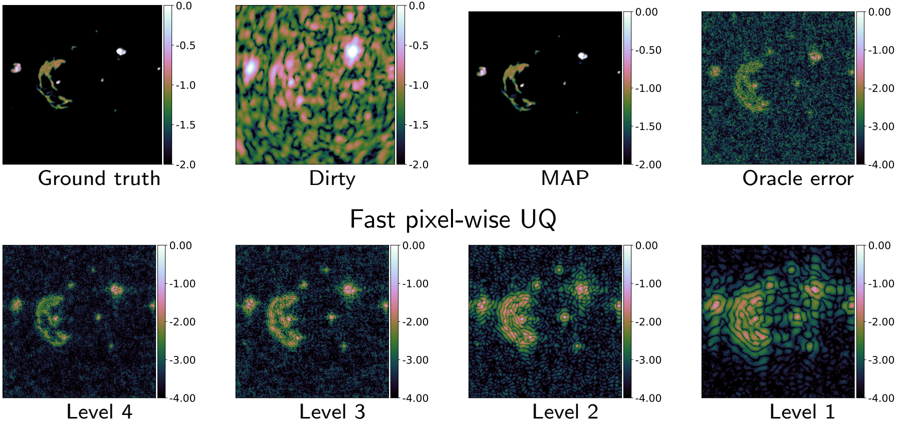


$M \approx 3 \times 10^4$ $\quad$ $M \approx 6 \times 10^4$ $\quad$ $M \approx 1.2 \times 10^5$ $\quad$ $M \approx 2.4 \times 10^5$

We use forward operator based on a torch-based 2D NUFFT with Kaisser-Bessel gridding.

A more realistic experiment

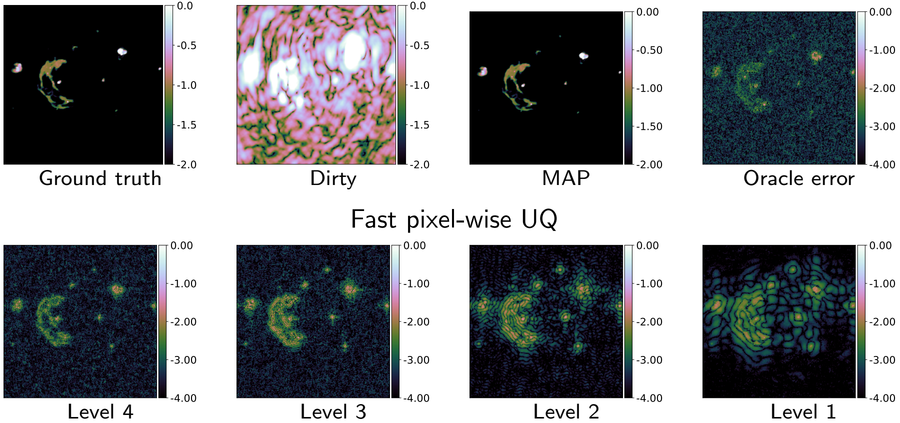
Results for **1h of observation time** ($M \approx 3 \times 10^4$). MAP reconstruction SNR: **23.88dB**



Computation wall-clock time: MAP estimation $\rightarrow$ 34.96s, **fast pixel UQ $\rightarrow$ 0.86s**

A more realistic experiment

Results for **8h of observation time** ($M \approx 2.4 \times 10^5$). MAP reconstruction SNR: **28.56dB**



Computation wall-clock time: MAP estimation $\rightarrow$ 137.0s, **fast pixel UQ $\rightarrow$ 1.84s**

Conclusions

- **Scalable uncertainty quantification**
  - We exploit a concentration phenomenon of log-concave posteriors
  - Focus on hypothesis test and pixel-wise errors at different scales

- Only rely on **optimisation** to compute the MAP and avoid sampling

- We used **learned convex regularisers**
  - Considerably decreased reconstruction errors
  - Improved quality of the posterior St Dev

Perspectives:
- Implement & benchmark QuantifAI on a massively parallelised computing env. (ongoing work)
- Bayesian model comparison of models with data-driven priors, (See Henry Aldridge's poster)

Publication: Liaudat et al. (2023), arXiv:2312.00125
Code: github.com/astro-informatics/quantifai

Conclusions

- **Scalable uncertainty quantification**
  - We exploit a concentration phenomenon of log-concave posteriors
  - Focus on hypothesis test and pixel-wise errors at different scales

- Only rely on **optimisation** to compute the MAP and avoid sampling

- We used **learned convex regularisers**
  - Considerably decreased reconstruction errors
  - Improved quality of the posterior St Dev

Perspectives:
- Implement & benchmark QuantifAI on a massively parallelised computing env. (ongoing work)
- Bayesian model comparison of models with data-driven priors, (See Henry Aldridge's poster)

Publication: Liaudat et al. (2023), arXiv:2312.00125
Code: github.com/astro-informatics/quantifai

Conclusions

- **Scalable uncertainty quantification**
  - We exploit a concentration phenomenon of log-concave posteriors
  - Focus on hypothesis test and pixel-wise errors at different scales

- Only rely on **optimisation** to compute the MAP and avoid sampling

- We used **learned convex regularisers**
  - Considerably decreased reconstruction errors
  - Improved quality of the posterior St Dev

Perspectives:
- Implement & benchmark QUANTIFAI on a massively parallelised computing env. (ongoing work)
- Bayesian model comparison of models with data-driven priors, (See Henry Aldridge's poster)

Publication: Liaudat et al. (2023), arXiv:2312.00125
Code: github.com/astro-informatics/quantifai

Conclusions

- **Scalable uncertainty quantification**
  - We exploit a concentration phenomenon of log-concave posteriors
  - Focus on hypothesis test and pixel-wise errors at different scales

- Only rely on **optimisation** to compute the MAP and avoid sampling

- We used **learned convex regularisers**
  - Considerably decreased reconstruction errors
  - Improved quality of the posterior St Dev

Perspectives:
- Implement & benchmark QUANTIFAI on a massively parallelised computing env. (ongoing work)
- Bayesian model comparison of models with data-driven priors, (See Henry Aldridge's poster)

Publication: Liaudat et al. (2023), arXiv:2312.00125
Code: github.com/astro-informatics/quantifai

Conclusions

- **Scalable uncertainty quantification**
  - We exploit a concentration phenomenon of log-concave posteriors
  - Focus on hypothesis test and pixel-wise errors at different scales

- Only rely on **optimisation** to compute the MAP and avoid sampling

- We used **learned convex regularisers**
  - Considerably decreased reconstruction errors
  - Improved quality of the posterior St Dev

Perspectives:

- Implement & benchmark QuantifAI on a massively parallelised computing env. (ongoing work)
- Bayesian model comparison of models with data-driven priors, (See Henry Aldridge's poster)

Publication: Liaudat et al. (2023), arXiv:2312.00125
Code: github.com/astro-informatics/quantifai

Conclusions

- **Scalable uncertainty quantification**
  - We exploit a concentration phenomenon of log-concave posteriors
  - Focus on hypothesis test and pixel-wise errors at different scales

- Only rely on **optimisation** to compute the MAP and avoid sampling

- We used **learned convex regularisers**
  - Considerably decreased reconstruction errors
  - Improved quality of the posterior St Dev

Perspectives:
- Implement & benchmark QuantifAI on a massively parallelised computing env. (ongoing work)
- Bayesian model comparison of models with data-driven priors, (See Henry Aldridge's poster)

Publication: Liaudat et al. (2023), arXiv:2312.00125
Code: github.com/astro-informatics/quantifai

Conclusions

- **Scalable uncertainty quantification**
  - We exploit a concentration phenomenon of log-concave posteriors
  - Focus on hypothesis test and pixel-wise errors at different scales

- Only rely on **optimisation** to compute the MAP and avoid sampling

- We used **learned convex regularisers**
  - Considerably decreased reconstruction errors
  - Improved quality of the posterior St Dev

Perspectives:
- Implement & benchmark QuantifAI on a massively parallelised computing env. (ongoing work)
- Bayesian model comparison of models with data-driven priors, (See Henry Aldridge's poster)

Publication: Liaudat et al. (2023), arXiv:2312.00125
Code: github.com/astro-informatics/quantifai

References

Cai, X., Pereyra, M., & McEwen, J. D. (2018a). Uncertainty quantification for radio interferometric imaging – I. Proximal MCMC methods. Monthly Notices of the Royal Astronomical Society, 480(3), 4154–4169.

Cai, X., Pereyra, M., & McEwen, J. D. (2018b). Uncertainty quantification for radio interferometric imaging: II. MAP estimation. Monthly Notices of the Royal Astronomical Society, 480(3), 4170–4182.

Carrillo, R. E., McEwen, J. D., & Wiaux, Y. (2012). Sparsity Averaging Reweighted Analysis (SARA): a novel algorithm for radio-interferometric imaging. Monthly Notices of the Royal Astronomical Society, 426(2), 1223–1234.

Goujon, A., Neumayer, S., Bohra, P., Ducotterd, S., & Unser, M. (2023). A neural-network-based convex regularizer for inverse problems. IEEE Transactions on Computational Imaging, 9, 781–795.

Liaudat, T. I., Mars, M., Price, M. A., Pereyra, M., Betcke, M. M., & McEwen, J. D. (2023). Scalable Bayesian uncertainty quantification with data-driven priors for radio interferometric imaging. arXiv e-prints, Article arXiv:2312.00125, arXiv:2312.00125.

Pereyra, M. (2016). Proximal markov chain monte carlo algorithms. Statistics and Computing, 26(4), 745–760.

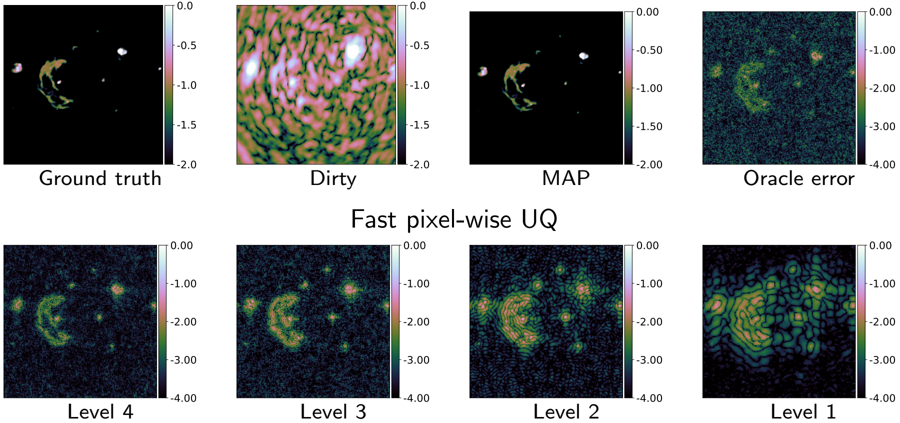
Pereyra, M. (2017). Maximum-a-posteriori estimation with bayesian confidence regions. SIAM Journal on Imaging Sciences, 10(1), 285–302.

Pereyra, M., Mieles, L. V., & Zygalakis, K. C. (2020). Accelerating proximal markov chain monte carlo by using an explicit stabilized method. SIAM Journal on Imaging Sciences, 13(2), 905–935.

Questions?

A more realistic experiment

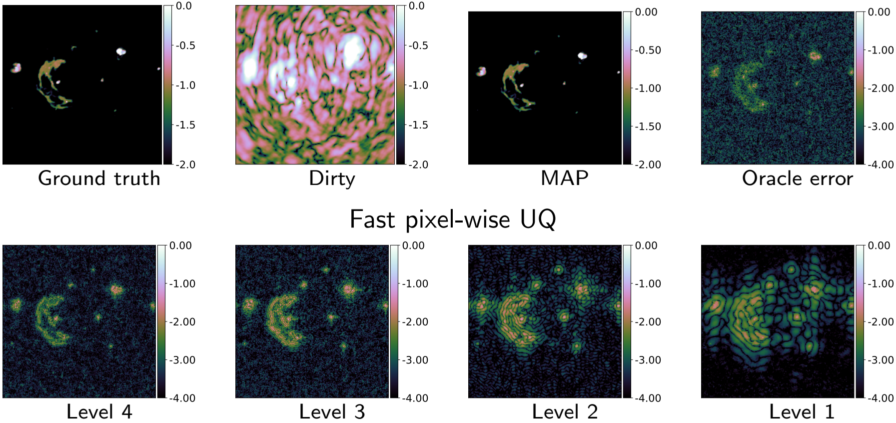
Results for **2h of observation time** ($M \approx 6 \times 10^4$). MAP reconstruction SNR: **25.89dB**



Computation wall-clock time: MAP estimation $\rightarrow$ 58.85s, **fast pixel UQ $\rightarrow$ 0.99s**

A more realistic experiment

Results for **4h of observation time** ($M \approx 1.2 \times 10^5$). MAP reconstruction SNR: **27.4dB**



Computation wall-clck time: MAP estimation $\rightarrow$ 93.63s, **fast pixel UQ $\rightarrow$ 1.32s**

Learned convex regulariser

The regulariser is trained to solve the denoising task

$$\mathbf{x}^* = \underset{\mathbf{x}\in\mathbb{R}^N}{\arg\min}\ \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|_2^2 + \lambda R_{\boldsymbol{\theta}}(\mathbf{x})\,, \tag{1}$$

The denoising problem is tackled through the unique fixed point of (1) by the gradient step:

$$T_{R_{\boldsymbol{\theta}},\lambda,\alpha}(\mathbf{x}) = \mathbf{x} - \alpha((\mathbf{x}-\mathbf{y}) + \lambda\nabla R_{\boldsymbol{\theta}}(\mathbf{x}))\,,$$

Doing a t-fold composition we obtain

$$T^t_{R_{\boldsymbol{\theta}},\lambda,\alpha}(\mathbf{y}^{(m)}) \approx \mathbf{x}^{(m)}$$

The training objective with training loss $\mathcal{L}$, e.g., ℓ_1, reads

$$\boldsymbol{\theta}^*_t, \lambda^*_t \in \underset{\boldsymbol{\theta},\lambda}{\arg\min} \sum_{m=1}^{M} \mathcal{L}\left(T^t_{R_{\boldsymbol{\theta}},\lambda,\alpha}(\mathbf{y}^{(m)}), \mathbf{x}^{(m)}\right)\,,$$

For more info: Goujon et al. (2023).