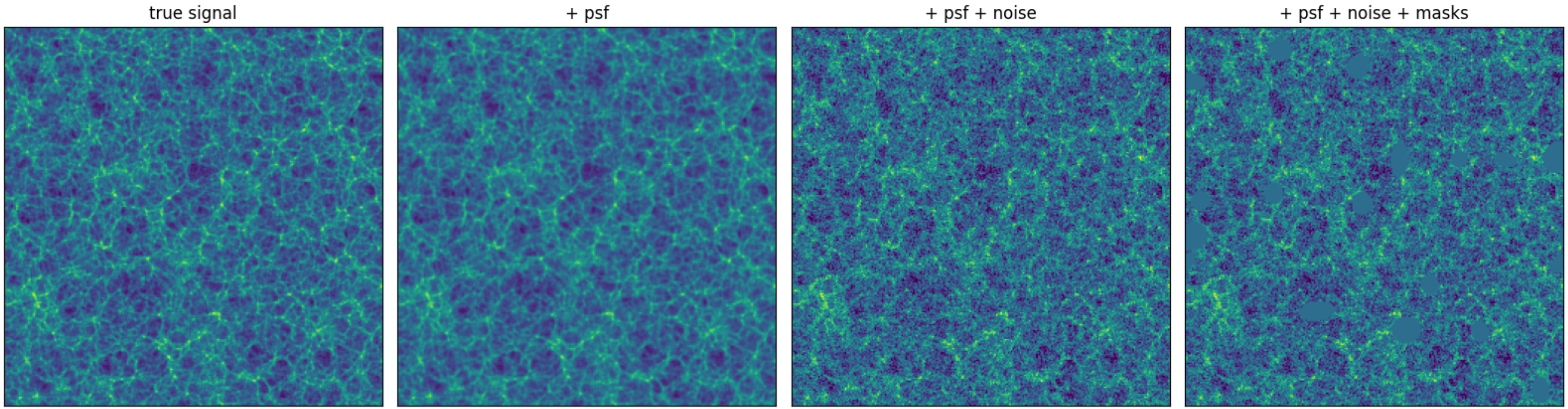
Bayesian inverse problem with scattering transform : application to instrumental decontamination

Sébastien Pierre
Erwan Allys, Alexandros Tsouros
22/05/2025

Bayesian inverse problems

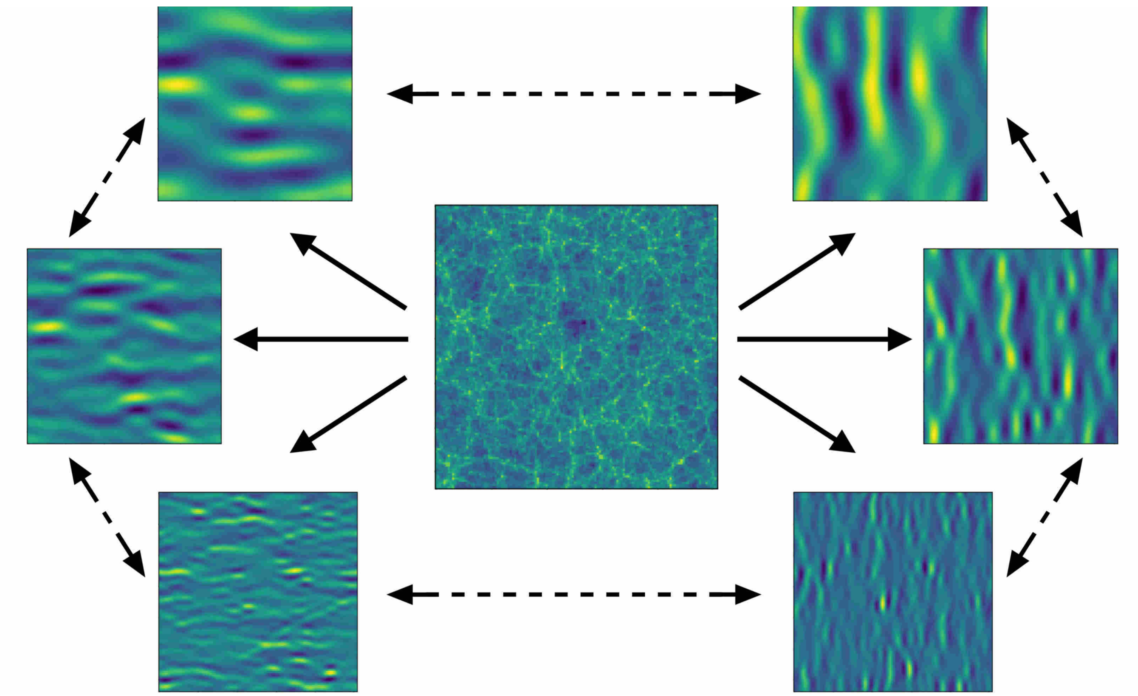
- We observe a data $d_0 = f(s_0)$ with s_0 signal of interest
- f probabilistic forward model, assumed known



- Only one observational data d_0 and no external prior model of s
- Ideal goal : find s_0 but no single solution for stochastic f
- Ill posed problem : need for a Bayesian formulation $\rightarrow p(s \mid d_{obs})$

Scattering transform statistics

- Scattering transform (ST) statistics $\phi(s)$: non-Gaussian and multi-scale statistics
- Can be efficiently estimated from a single image

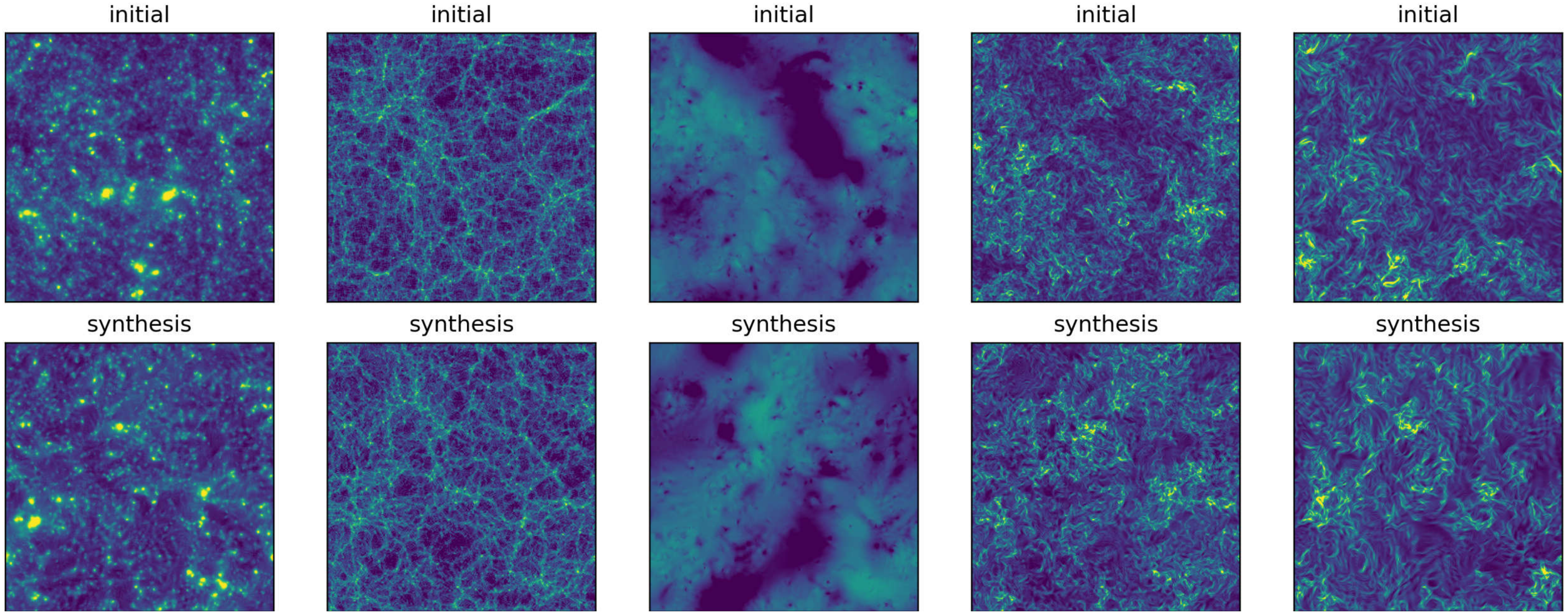


- Wavelet filters separating the different scales
- Coupling between scales with non-linearities

Generative models from ST statistics

- Ability to construct and sample maximum entropy models from ST statistics
- Parametrised by $\mu_s = \mathbb{E}_{s \sim p(s)}[\phi(s)]$

$$p(s) \rightarrow s_0 \xrightarrow{\text{estimation}} \mu_s \simeq \phi(s_0) \rightarrow p^{m.e.}_{\phi(s_0)} \xrightarrow{\text{sample}} \tilde{s}$$

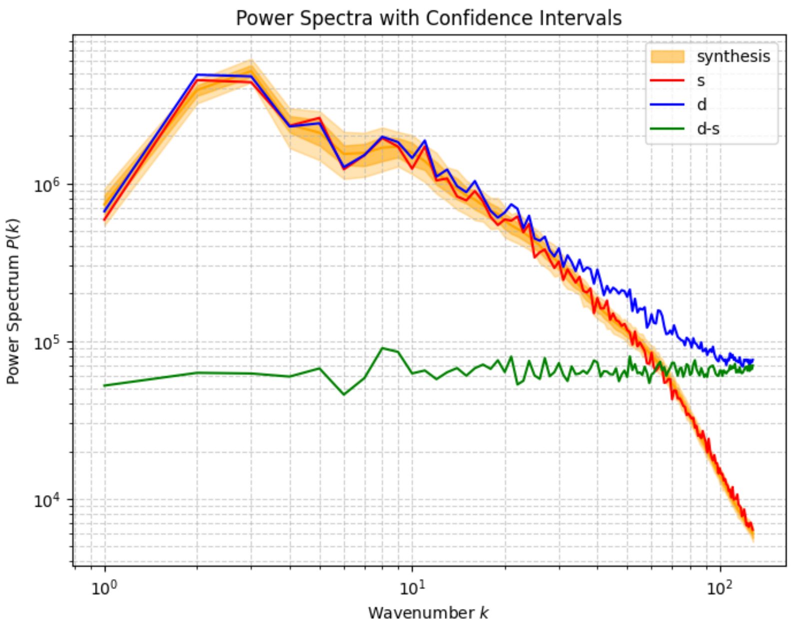


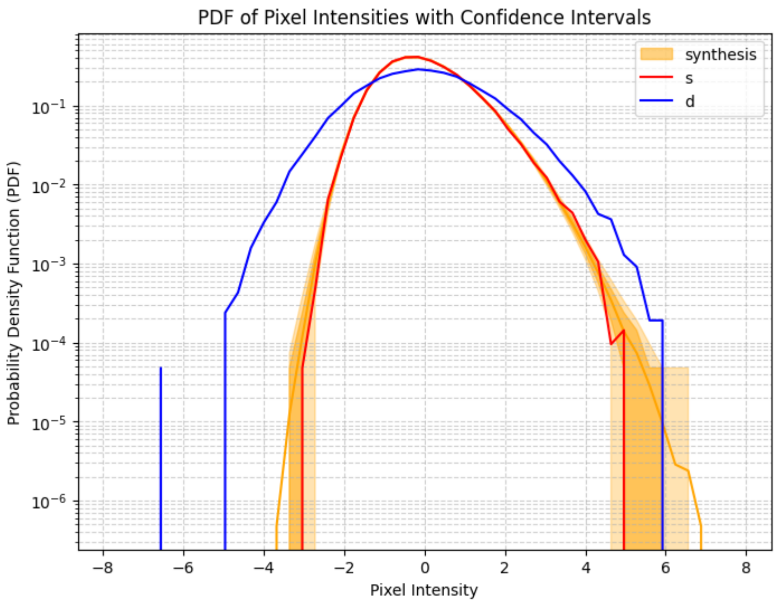
Realistic non-Gaussian models from $O(10^2)$ coefficients

Generative models from ST statistics

- Ability to construct and sample maximum entropy models from ST statistics
- Parametrised by $\mu_s = \mathbb{E}_{s \sim p(s)}[\phi(s)]$

$$p(s) \rightarrow s_0 \xrightarrow{\text{estimation}} \mu_s \simeq \phi(s_0) \rightarrow p^{m.e.}_{\phi(s_0)} \xrightarrow{\text{sample}} \tilde{s}$$



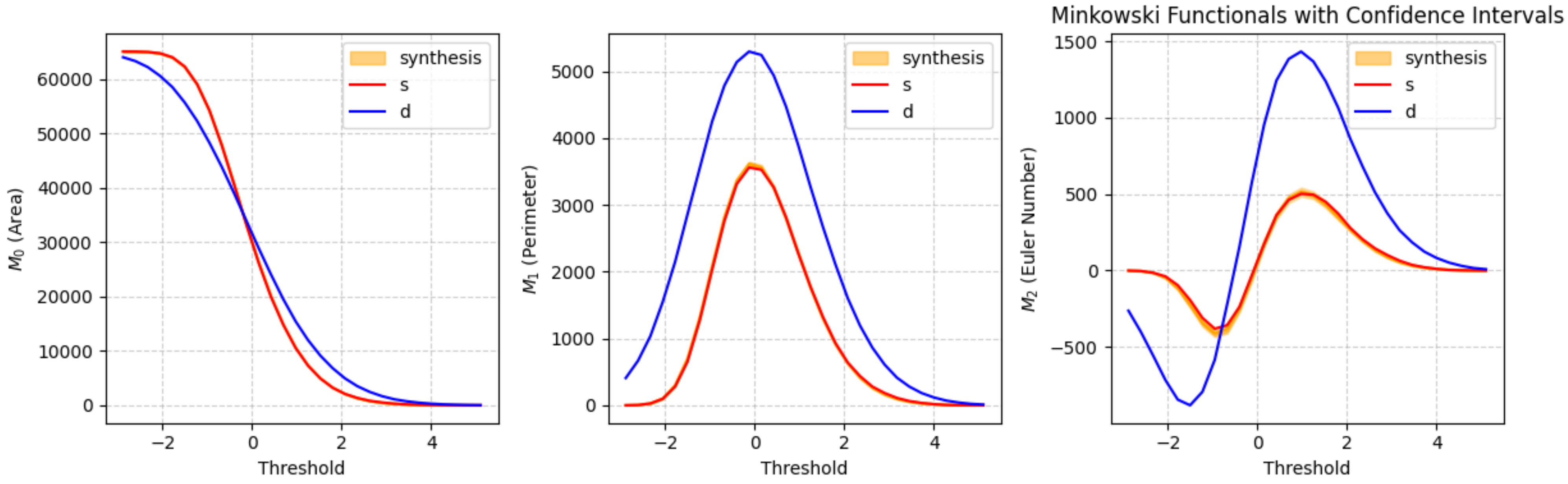


Recovers power spectrum and PDF

Generative models from ST statistics

- Ability to construct and sample maximum entropy models from ST statistics
- Parametrised by $\mu_s = \mathbb{E}_{s\sim p(s)}[\phi(s)]$

$$p(s) \to s_0 \xrightarrow{\text{estimation}} \mu_s \simeq \phi(s_0) \to p^{m.e.}_{\phi(s_0)} \xrightarrow{\text{sample}} \tilde{s}$$



Recovers Minkowski functionals

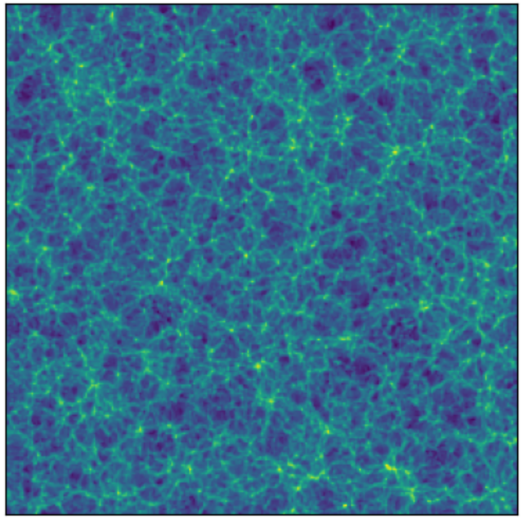
Forward model

- Idea : target μ_{s_0} instead of $s_0 \rightarrow p(\mu_s \mid d_0)$
- Also describe d_0 by its ST statistics $\rightarrow$ ***Goal*** : $p(\mu_s \mid \phi(d_0))$
- Full forward model :

$$\mu_s \rightarrow p_{\mu_s} \rightarrow s \rightarrow f(s) = d \rightarrow \phi(d)$$

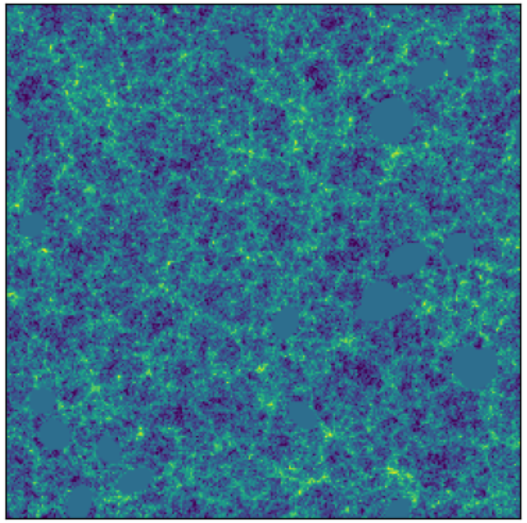
$$\mu_s = \begin{bmatrix} \mu_{s_1} \\ \mu_{s_2} \\ \mu_{s_3} \\ \vdots \\ \mu_{s_n} \end{bmatrix} \xrightarrow{\text{Synthesis}}$$

sample from p_{μ_s}



$\xrightarrow{\text{Contamination}}$

after contamination



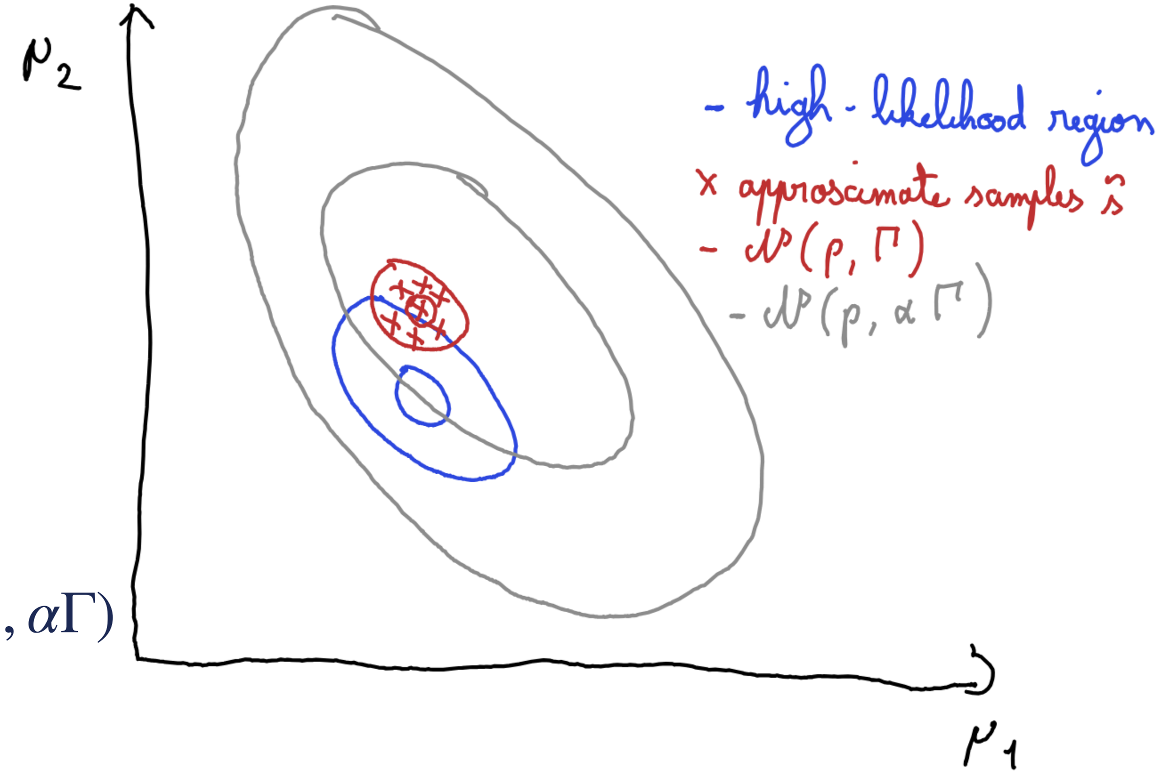
$$\xrightarrow{\text{Compression}} \begin{bmatrix} \phi_1(d) \\ \phi_2(d) \\ \phi_3(d) \\ \vdots \\ \phi_n(d) \end{bmatrix} = \phi(d)$$

Bayesian formulation

- Parameter space : μ_s — data space : $\phi(d)$
- $p(\mu_s \mid \phi(d)) \propto p(\phi(d) \mid \mu_s)p(\mu_s)$
- Uninformative prior in scattering space
- SBI for the likelihood ?
  - Unstructured space of dimension $O(10^2)$
  - Challenging setting to train a Normalising Flow
- Postulate a closed form for the likelihood $p(\phi(d) \mid \mu_s) = \mathcal{N}(A\mu_s + b, \Sigma)$
  - Linked to a Taylor expansion
  - Need to estimate A, b and Σ
  - Use a proposal distribution $q(\mu_s)$

Proposal distribution

- Requirement for $q(\mu_s)$:
  - Localised in scattering space
  - Take into account the correlation between ST coefficients
- Previous approximate solutions based on GD
  - Biased
  - Not enough variability
- Fit a gaussian to the $\tilde{s}$ and rescale it $q(\mu_s) = \mathcal{N}(\rho, \alpha\Gamma)$
  - Need to cover the high-likelihood region
  - Check that inference is not driven by α

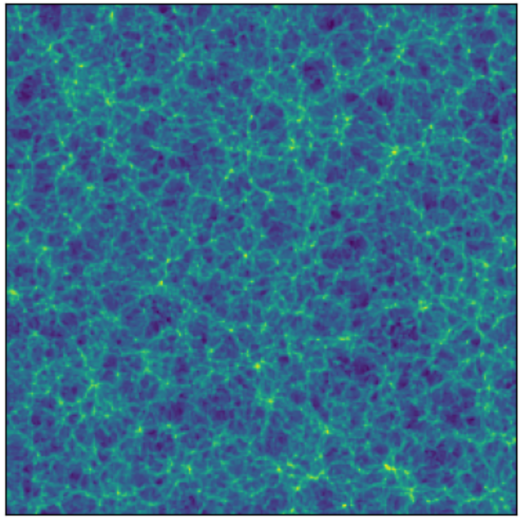


Posterior distribution $p(\mu_s \mid \phi(d))$

- With the proposal distribution we are able to estimate A, b and Σ

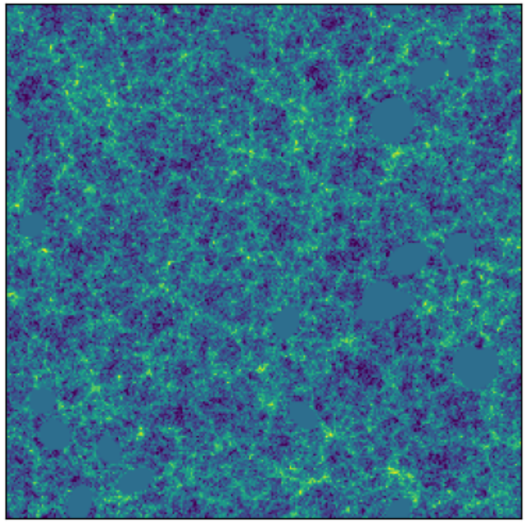
$$\mu_s = \begin{bmatrix} \mu_{s_1} \\ \mu_{s_2} \\ \mu_{s_3} \\ \vdots \\ \mu_{s_n} \end{bmatrix} \sim q(\mu_s) \xrightarrow{\text{Synthesis}}$$

sample from p_{μ_s}



$$\xrightarrow{\text{Contamination}}$$

after contamination



$$\xrightarrow{\text{Compression}} \begin{bmatrix} \phi_1(d) \\ \phi_2(d) \\ \phi_3(d) \\ \vdots \\ \phi_n(d) \end{bmatrix} = \phi(d)$$

- Likelihood $p(\phi(d) \mid \mu_s) \sim \mathcal{N}(A\mu_s + b, \Sigma)$
- Posterior distribution $p(\mu_s \mid \phi(d)) \propto \mathcal{N}(A\mu_s + b, \Sigma)p(\mu_s) = q(\mu_s)\mathcal{N}(A\mu_s + b, \Sigma)\dfrac{p(\mu_s)}{q(\mu_s)}$
- Multi-round estimation of A, b and Σ

Results

- Setting :
  - Only one observed data d_0
  - No prior model for s
  - Know the forward function f (and quick to evaluate)
- Modelling hypothesis :
  - s well described by a maximum entropy model conditioned by ST statistics
  - Likelihood function modelled as $p(\phi(d) \mid \mu_s) = \mathcal{N}(A\mu_s + b \mid \Sigma)$

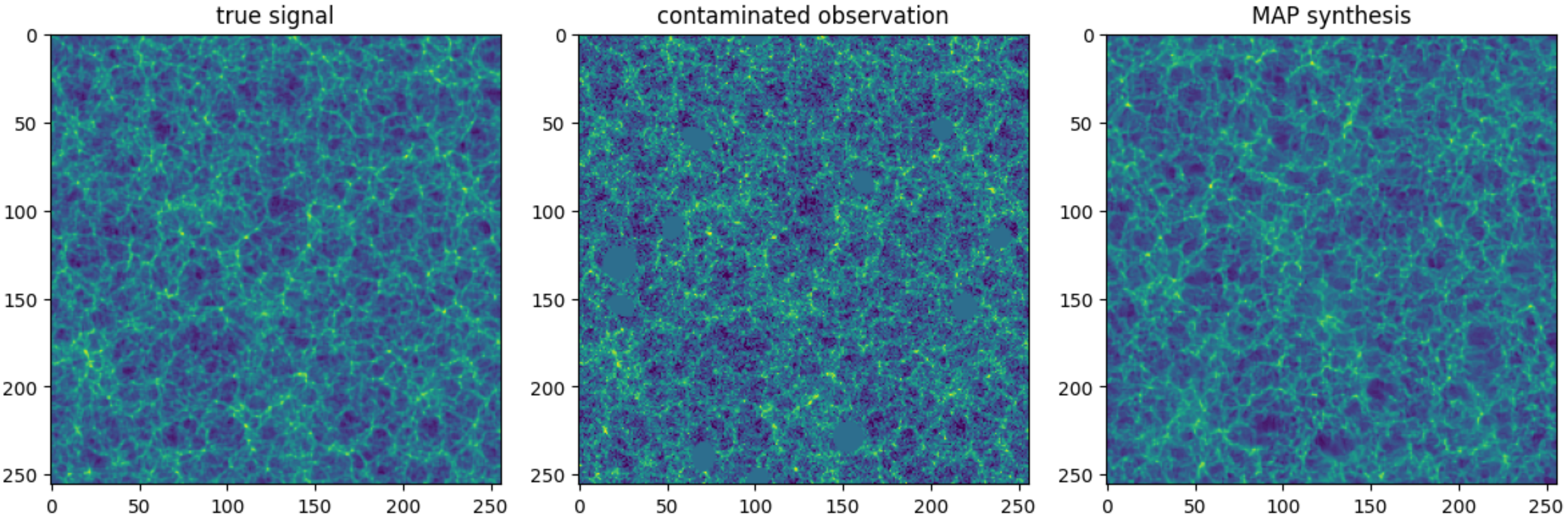
→ Posterior distribution $p(\mu_s \mid \phi(d_0))$ of scattering statistics of s

And by sampling $\tilde{s}$ from μ_s of any other statistics

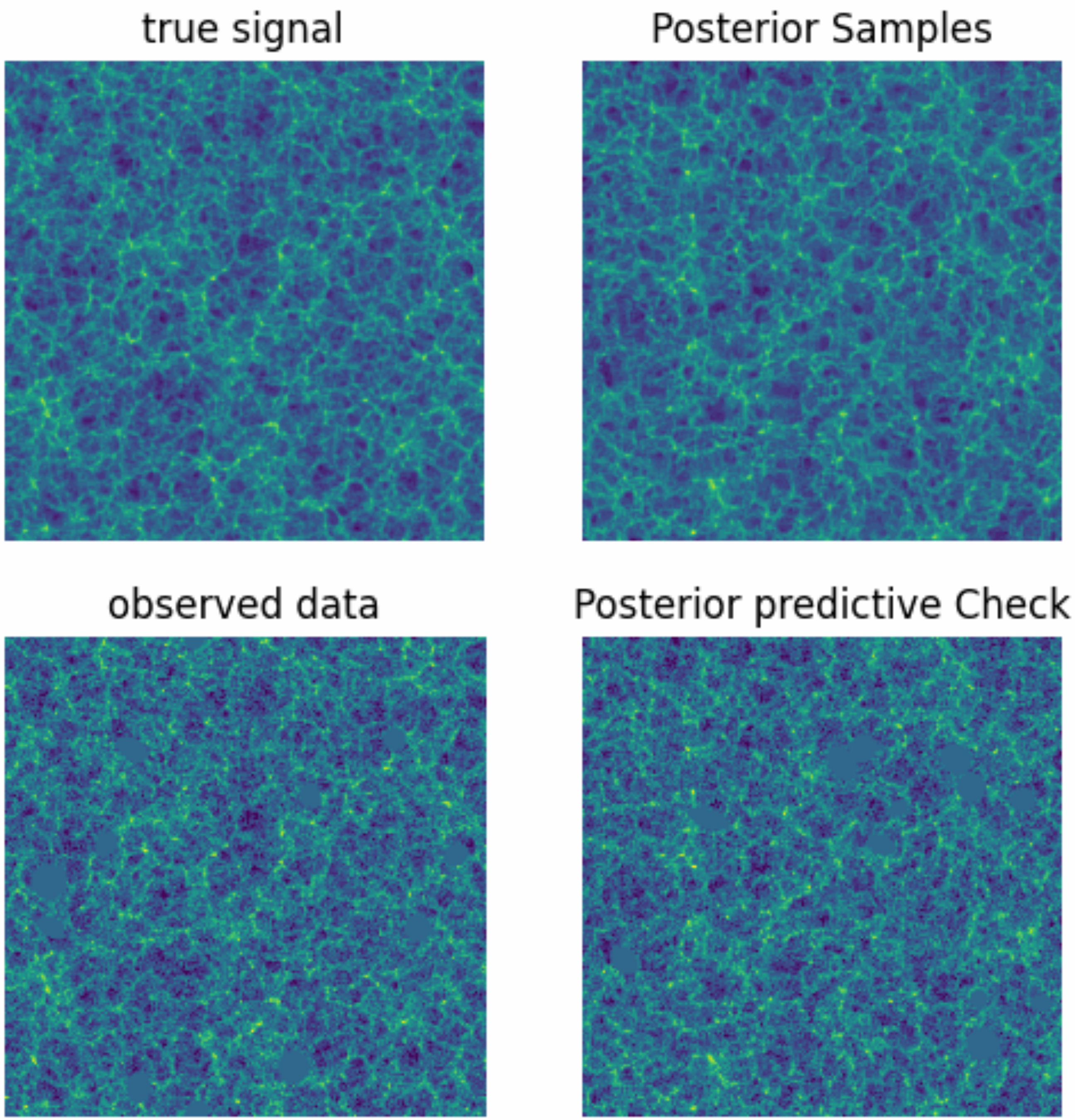
Results

PRELIMINARY

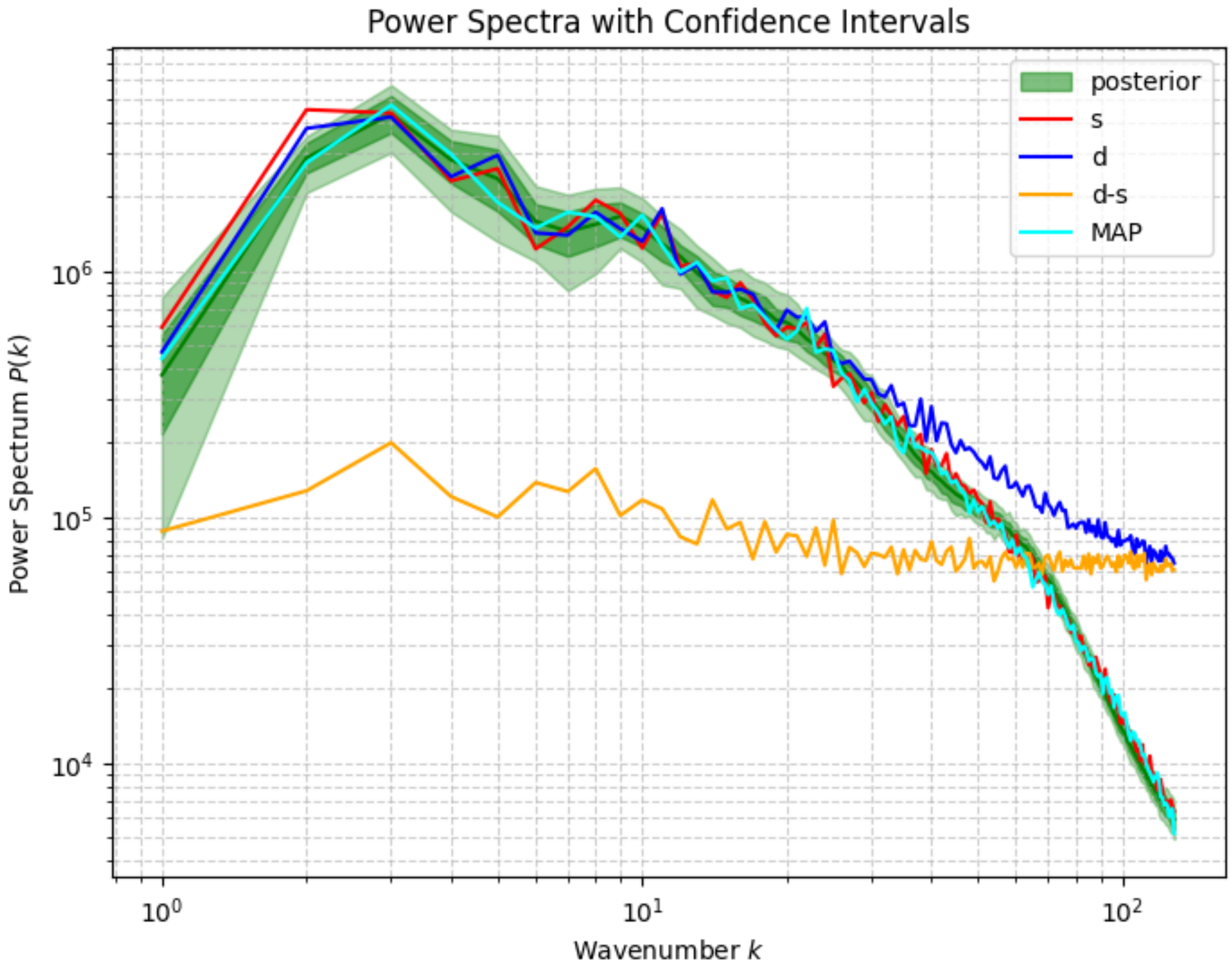
- Proof of concept on 256*256 LSS maps and instrumental contamination



Results

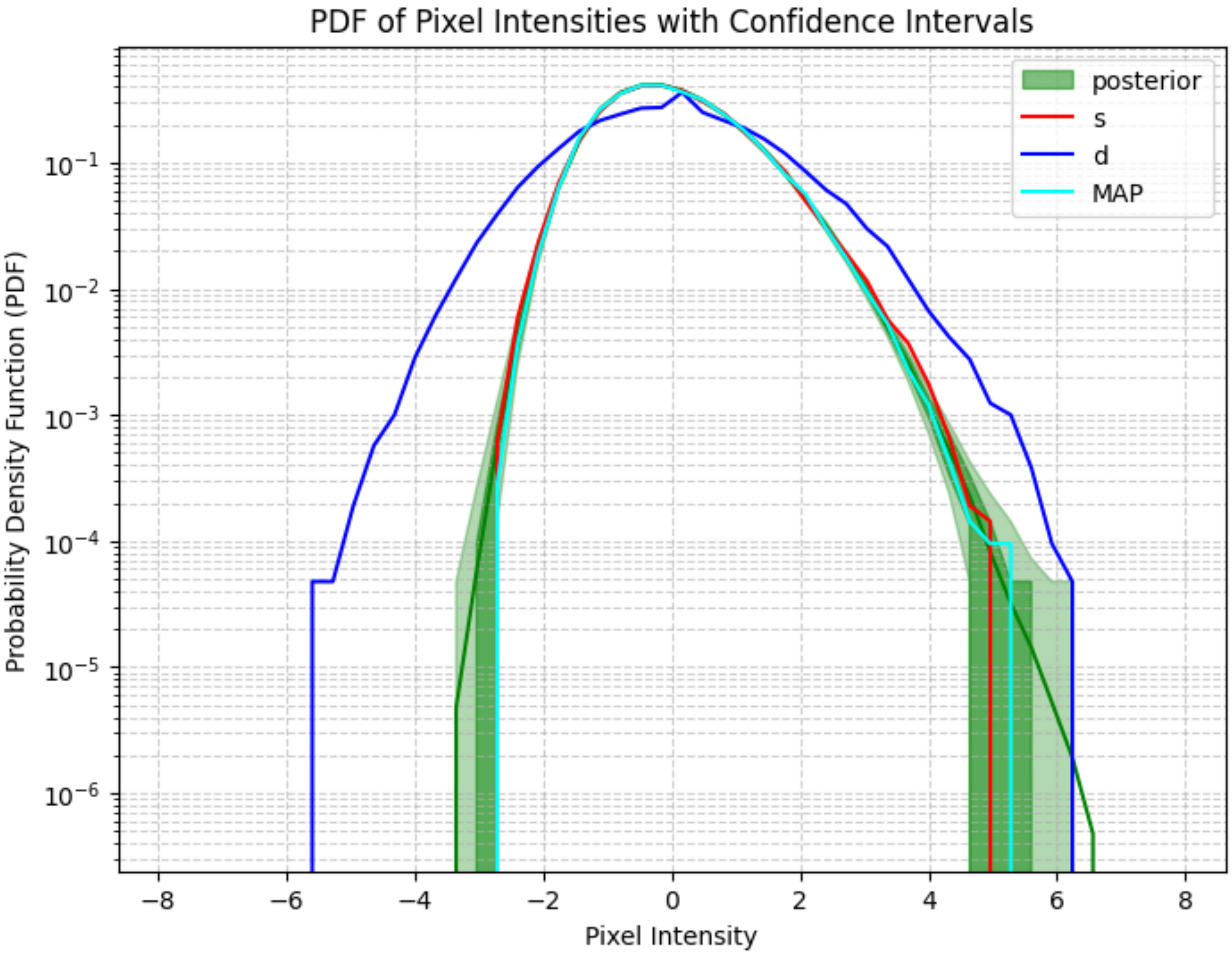

true signal
Posterior Samples
observed data
Posterior predictive Check

Results



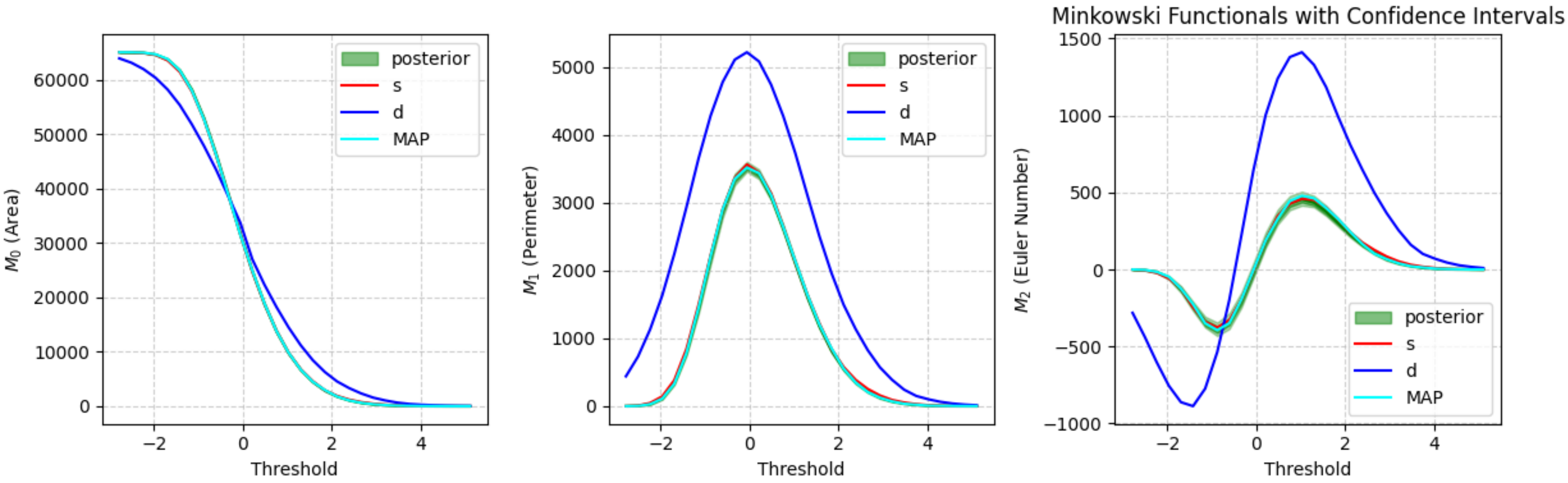
Power Spectra with Confidence Intervals
posterior
s
d
d-s
MAP
Power Spectrum $P(k)$
Wavenumber k
10^4
10^5
10^6
10^0
10^1
10^2

Results



PDF of Pixel Intensities with Confidence Intervals
posterior
s
d
MAP
Probability Density Function (PDF)
Pixel Intensity
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
−8
−6
−4
−2
0
2
4
6
8

Results



Minkowski Functionals with Confidence Intervals
M_0 (Area)
M_1 (Perimeter)
M_2 (Euler Number)
Threshold
Threshold
Threshold
posterior
s
d
MAP
posterior
s
d
MAP
posterior
s
d
MAP

Conclusion

- Challenging setting : one data and no external prior model for s
- Applicable to a lot of other data than LSS
- Recover a posterior of ST statistics of s and other usual astrophysical statistics
- Computationally expensive $\simeq$ 10 hours H100 GPU (work in progress)

Proposal distribution

- Use the data and the algorithm from Régaldo+21 to get multiple estimation $\tilde{s}$ of s_{truth}

  Gradient descent in pixel space to minimise the objective $||\phi(f(\tilde{s})) - \phi(d_{obs})||^2$

- Fit a the proposal distribution $q(\mu_s)$ as a gaussian distribution over these estimations $q(\mu_s) = \mathcal{N}(\rho, \Gamma)$ and scale the covariance to cover as much as possible the high likelihood region