Generative models of Galactic foregrounds

In Collaboration with Lloyd Knox, and Karthik Prabhu, at UC Davis

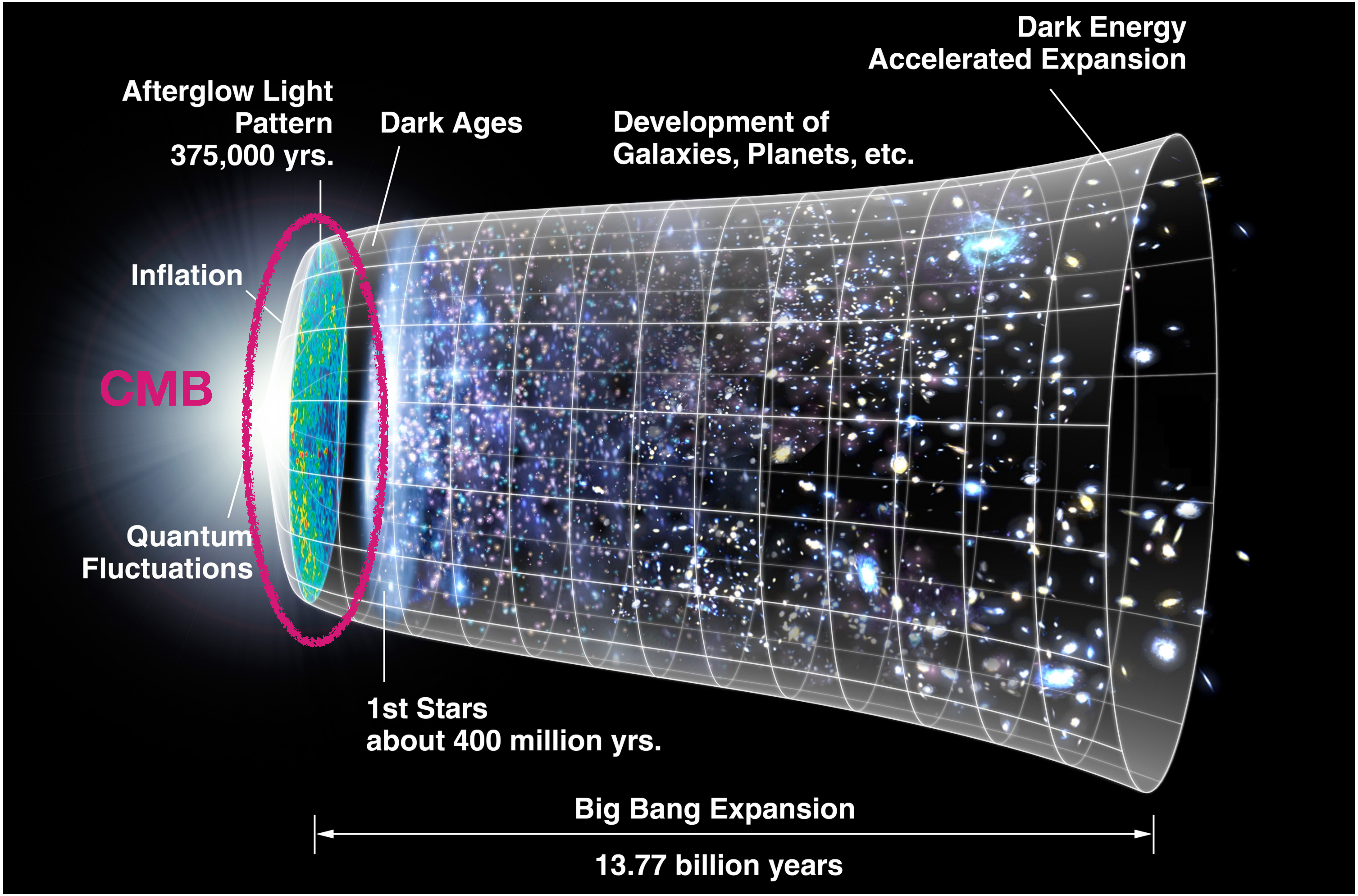
arXiv: 2101.11181

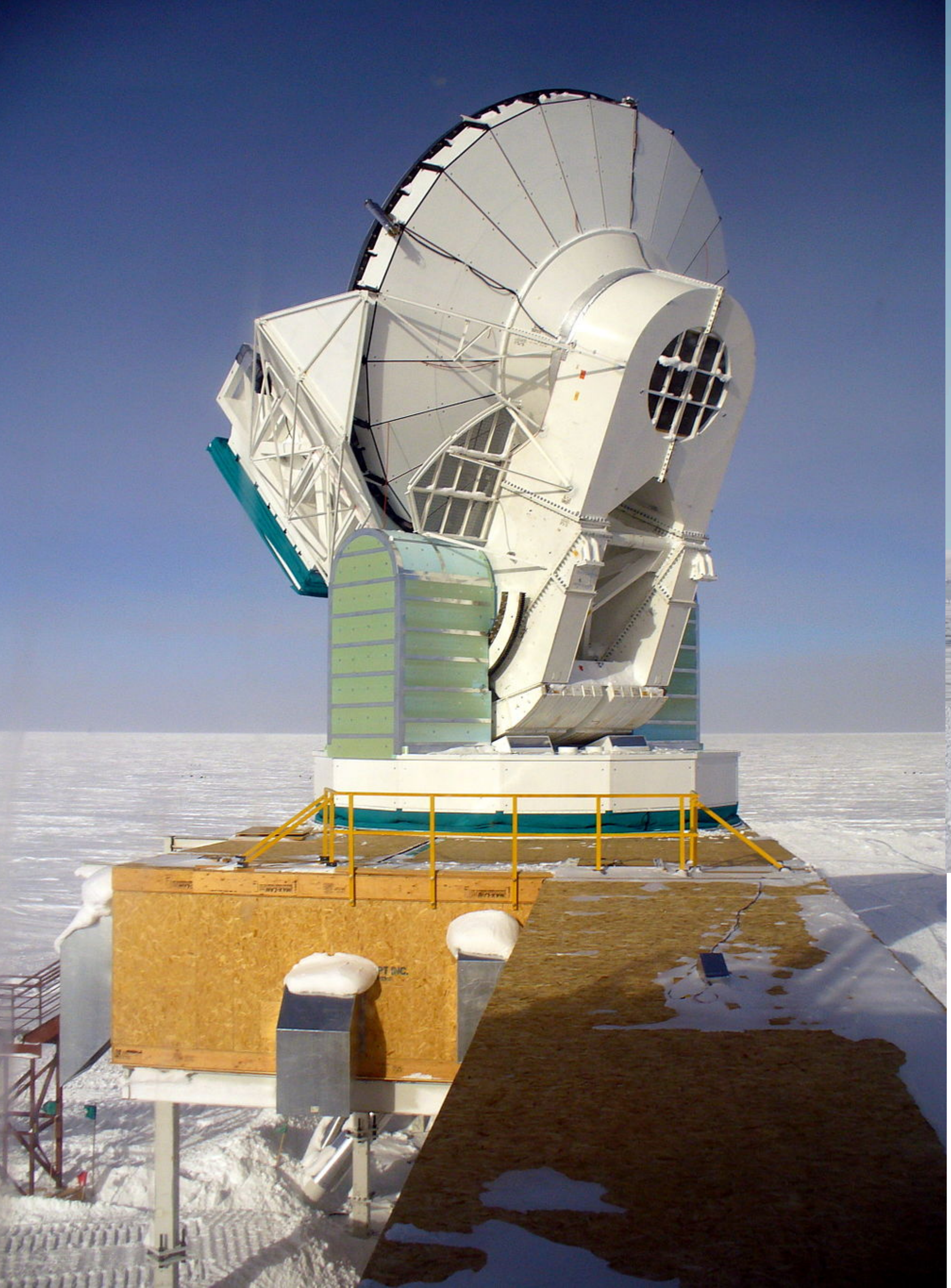
Ben Thorne, Wednesday March 17, 2021 - Deep generative modeling workshop, UC Berkeley

Outline

- Some motivation for studying the CMB.
- Why we have to understand the Galaxy, and why it is hard.
- Recent developments in generative modeling of Galactic components.
- Variational Autoencoders.
- Successes and failures of our VAE application.

Background


Dark Energy
Accelerated Expansion
Afterglow Light
Pattern
375,000 yrs.
Dark Ages
Development of
Galaxies, Planets, etc.
Inflation
CMB
Quantum
Fluctuations
1st Stars
about 400 million yrs.
Big Bang Expansion
13.77 billion years



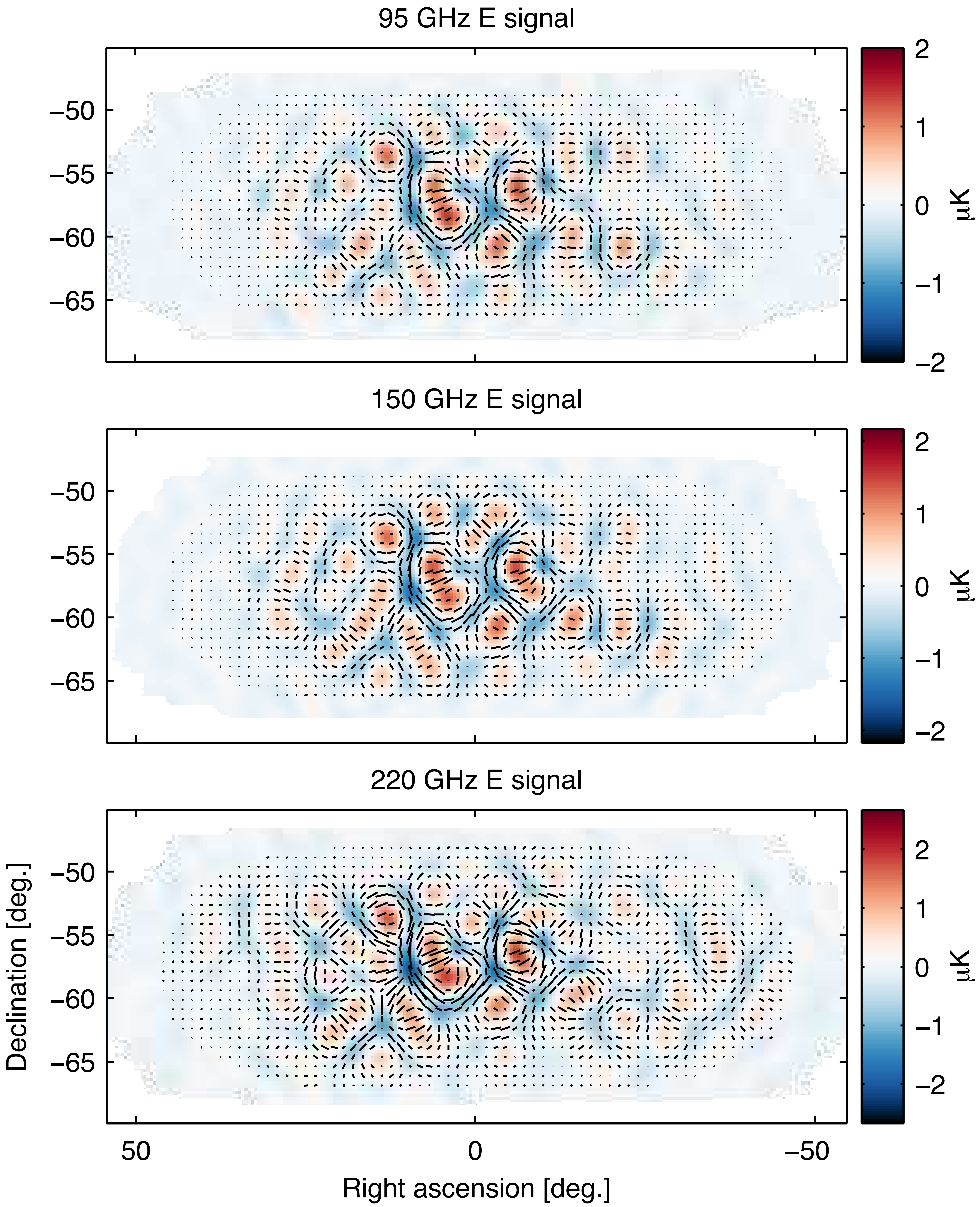
Above: Keck Array

Left: South Pole Telescope

CMB Analysis

Standard approach

- Measure polarization of CMB. Look for signal of primordial gravitational waves present at recombination, smoking gun of **inflation**.
- Curl-like pattern (B-mode), parametrized by a scalar r.
- CMB expected to be **Gaussian** - compress to power spectrum and infer constraint on r from bandpowers.

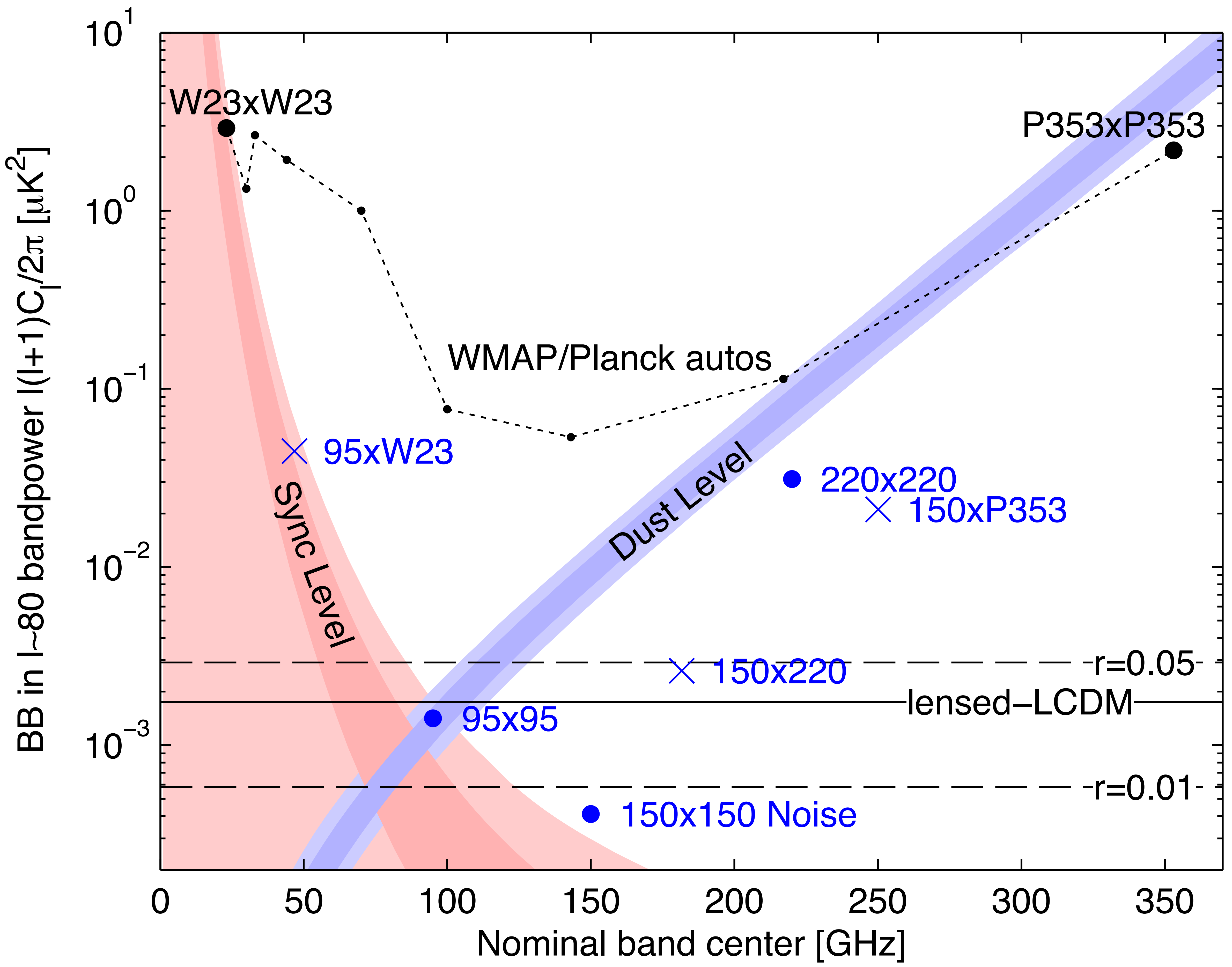


CMB E-mode polarization, as measured by BICEP / Keck Telescope at the South Pole (BKP 2019).

CMB Analysis

Current problems

- Pattern induced by non-zero r is mimicked by:
  - Lensing of the CMB by large scale structure leaks E-mode power into B-mode power.
  - Polarized emission by interstellar dust in our own galaxy.
- Need to model both effects to constrain r in future low-noise observations.



Foreground power on degree scales at cleanest frequency, and in cleanest region of sky, exceeds power due to primordial B-modes (BKP 2019).

Optimal Bayesian Delensing and Foreground Removal

Millea + 2020a,b (Bayesian Delensing Delight, and applied to SPTPol data)

- In order to deal with lensing in an optimal way, forward model the whole problem at the map level. Inherently optimal, but computationally challenging.

CMB field

$$d = \mathbb{A}\mathbb{L}(\phi)f + n$$

Lensing field

Tensor-to-scalar ratio

- Posterior is relatively simple: $\log p(f, \phi, A_\phi, r \,|\, d) \propto \dfrac{(d - \mathbb{A}\mathbb{L}(\phi)f)^2}{\mathbb{C}_n} + \dfrac{f^2}{\mathbb{C}_f(r)} + \dfrac{\phi^2}{\mathbb{C}_\phi(A_\phi)}$

Foreground field

- Adding foregrounds to the data model: $d = \mathbb{A}\mathbb{L}(\phi)f + \mathbb{F}(\beta)g_{\mathrm{dust}} + n$

Foreground models

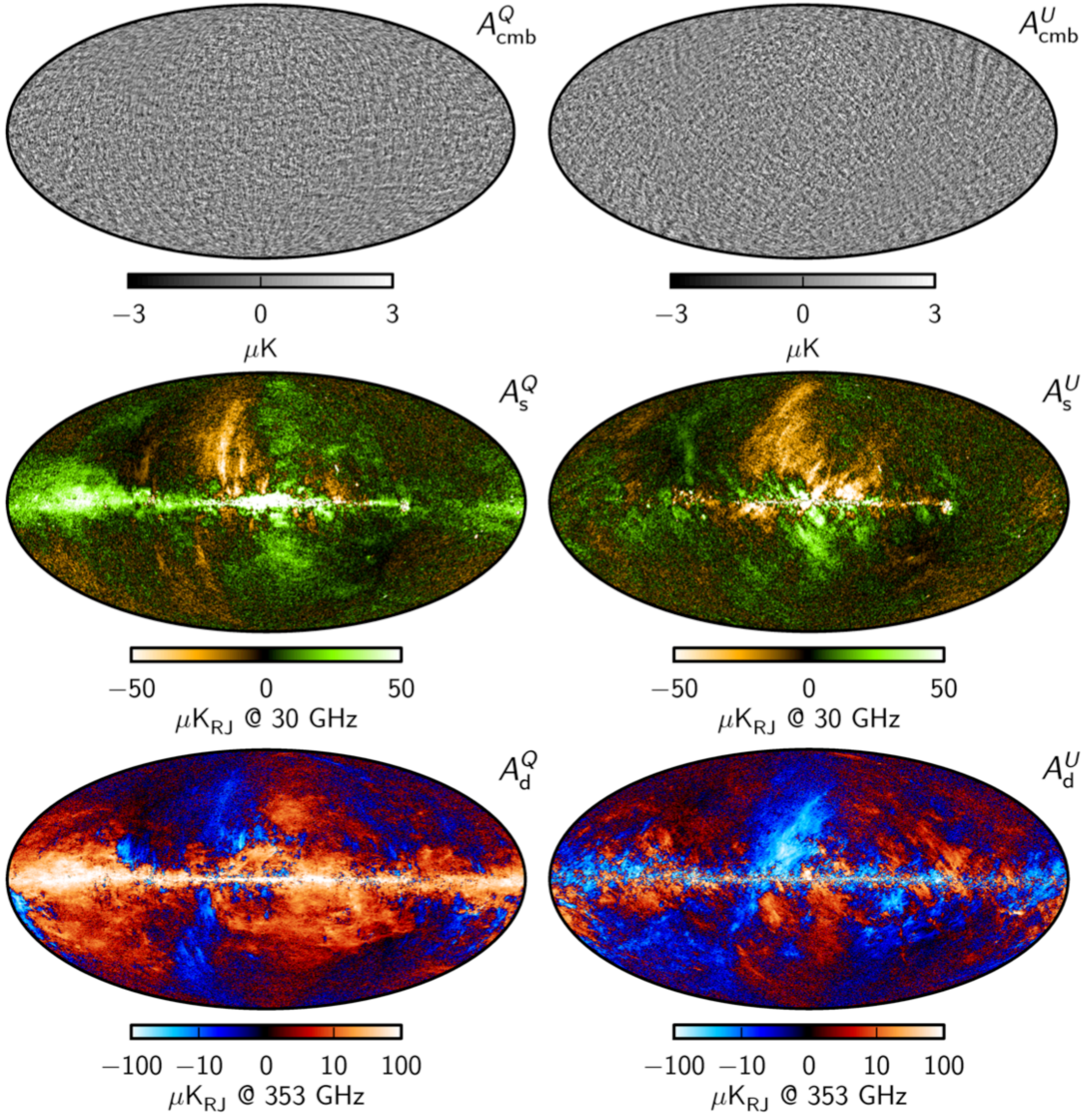
How do we model $p(g_{\mathrm{dust}})$

- Can use a Gaussian prior on this field, equivalent to power spectrum level modeling?

- Can we learn more about its statistics, and perhaps even a Bayesian inference model, using machine learning?

- At the least we would like to develop simulations with non-Gaussian foregrounds to test our Bayesian inference pipeline.

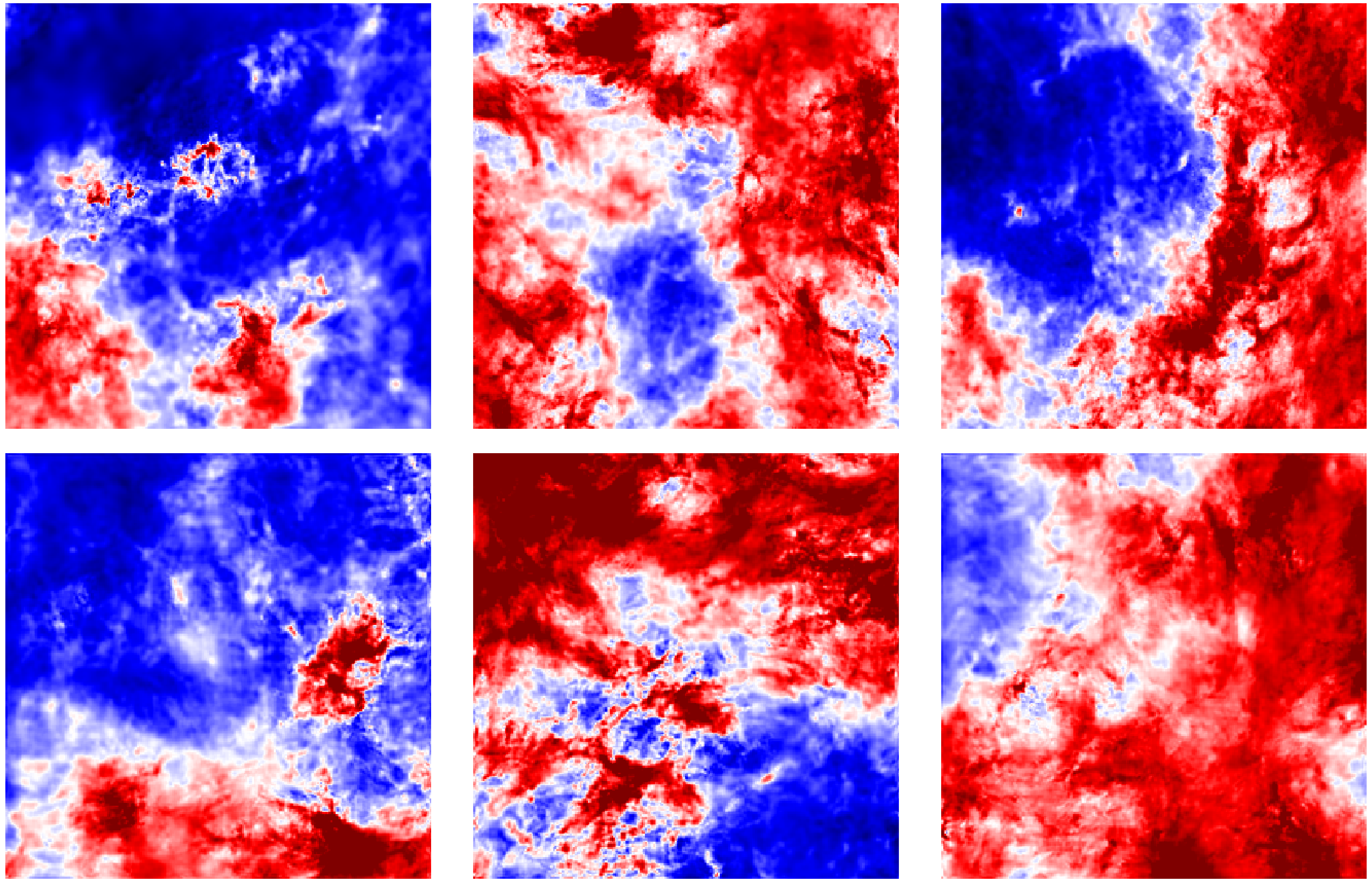
- Difficulty is that we have only one sky to measure, and simulations do not yet cover the full range of angular scales we are interested in for CMB experiments.



CMB (top), synchrotron Q and U, and dust Q and U (Planck Collaboration 2015)

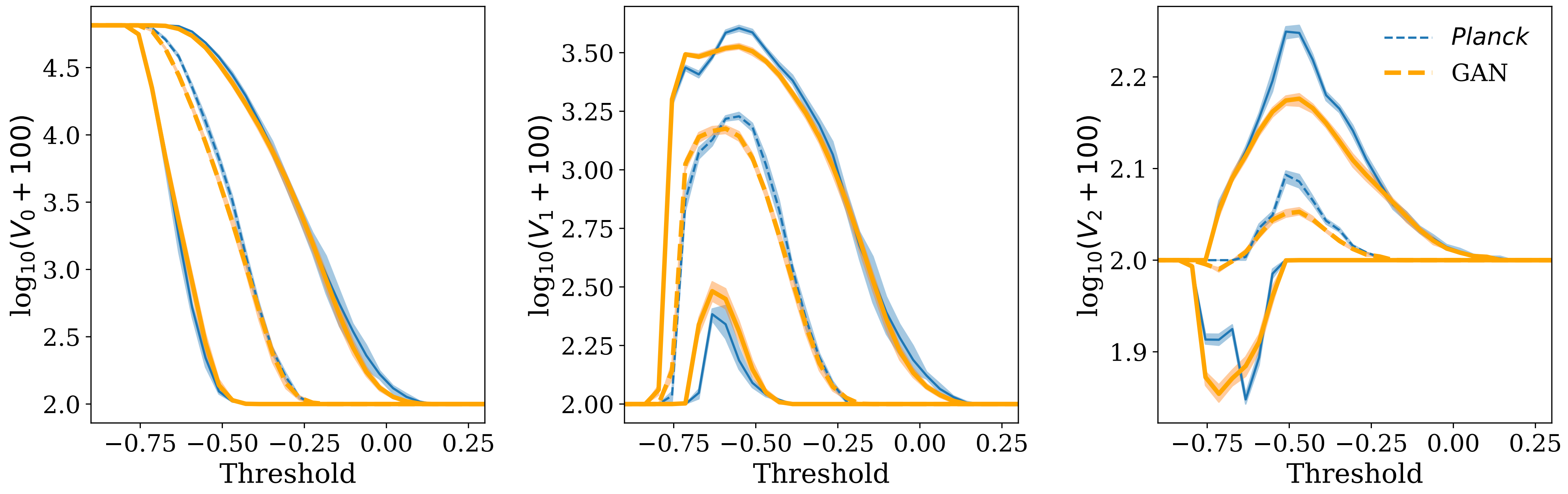
Aylor + 2021

Trained GAN to produce intensity samples



Test set data (top row) and samples from trained GAN (bottom row) presented in Aylor+2021.

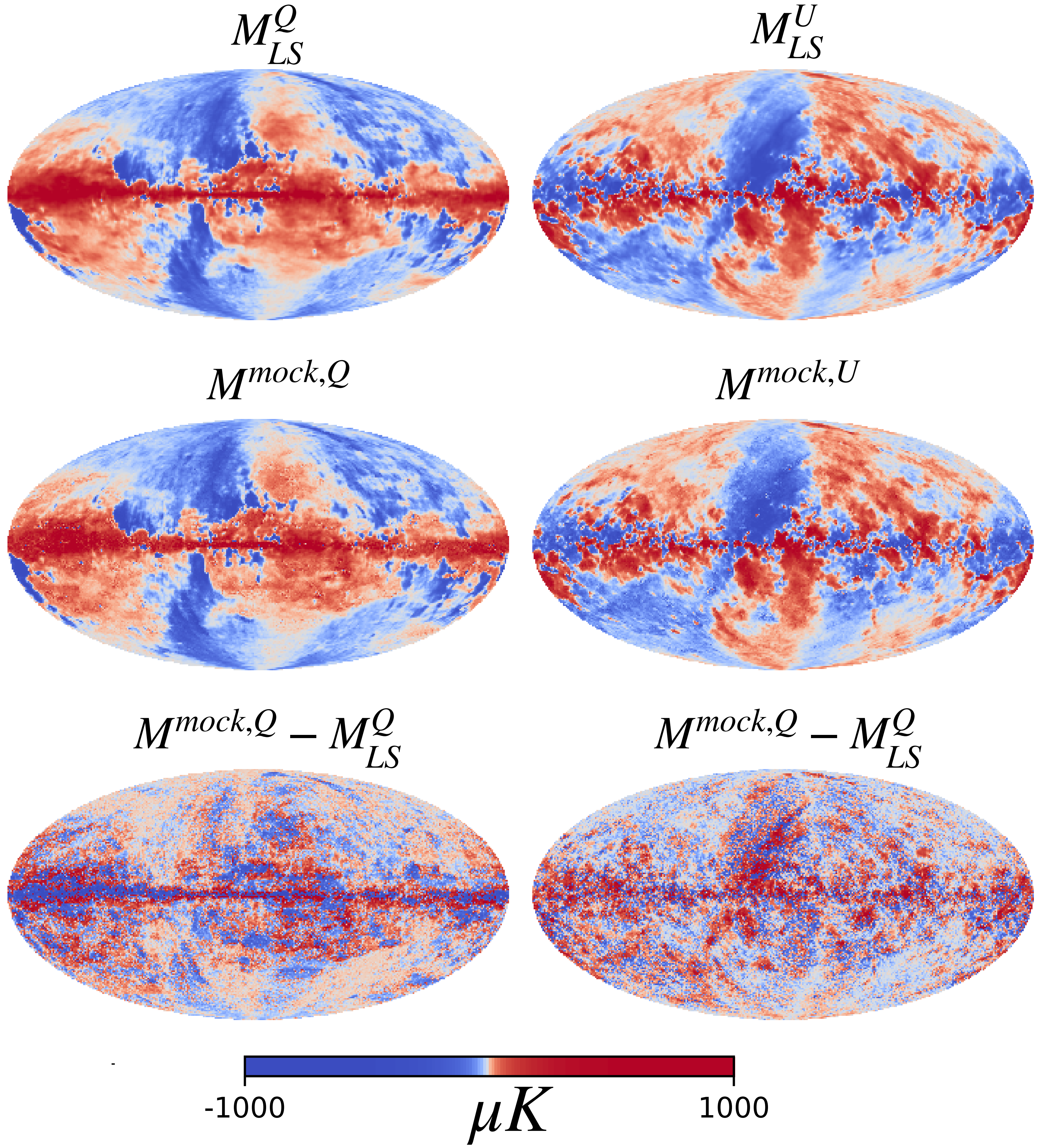
- GAN trained on dataset consisting of cutouts from Planck 545 GHz GNILC map.
- Samples reproduced correct summary statistics: pixel value histograms, power spectra, Minkowski functionals.
- Unstable training process, difficult to reproduce.
- Stochastic, but not particularly diverse images are generated.



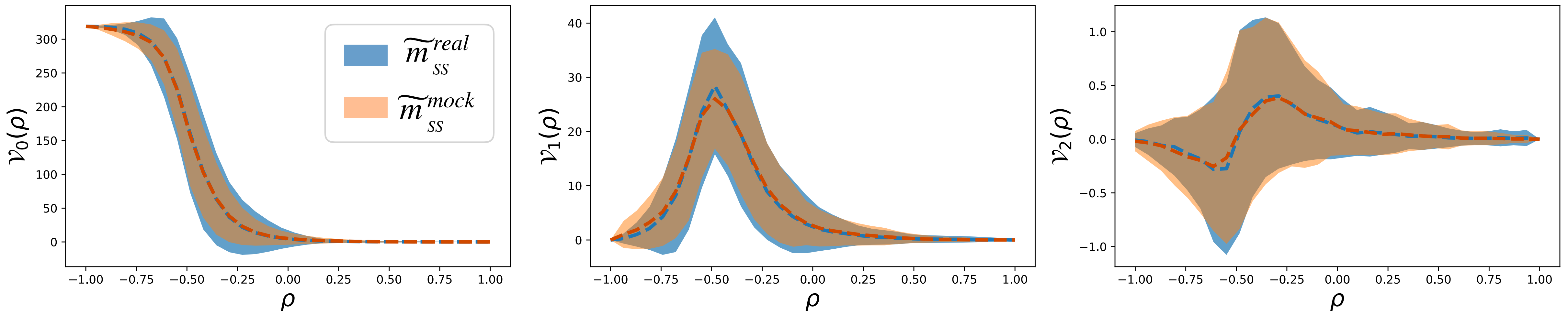
Minkowski functionals of GAN-generated images compared to data.

Krachmalnicoff & Puglisi 2020

Small-scale realizations using GANs



- Developed to generate high-resolution polarization simulations, conditioned on low-resolution observations.
- Full sky through a tiling method.
- Deterministic - due to method of training there is no stochasticity in the generated dust images.
- Some summary statistics are not reproduced (e.g. EE / BB ratio of small scales)



Minkowski functionals of GAN-generated small-scales compared to data.

Model Setup

Variational Autoencoders

Background

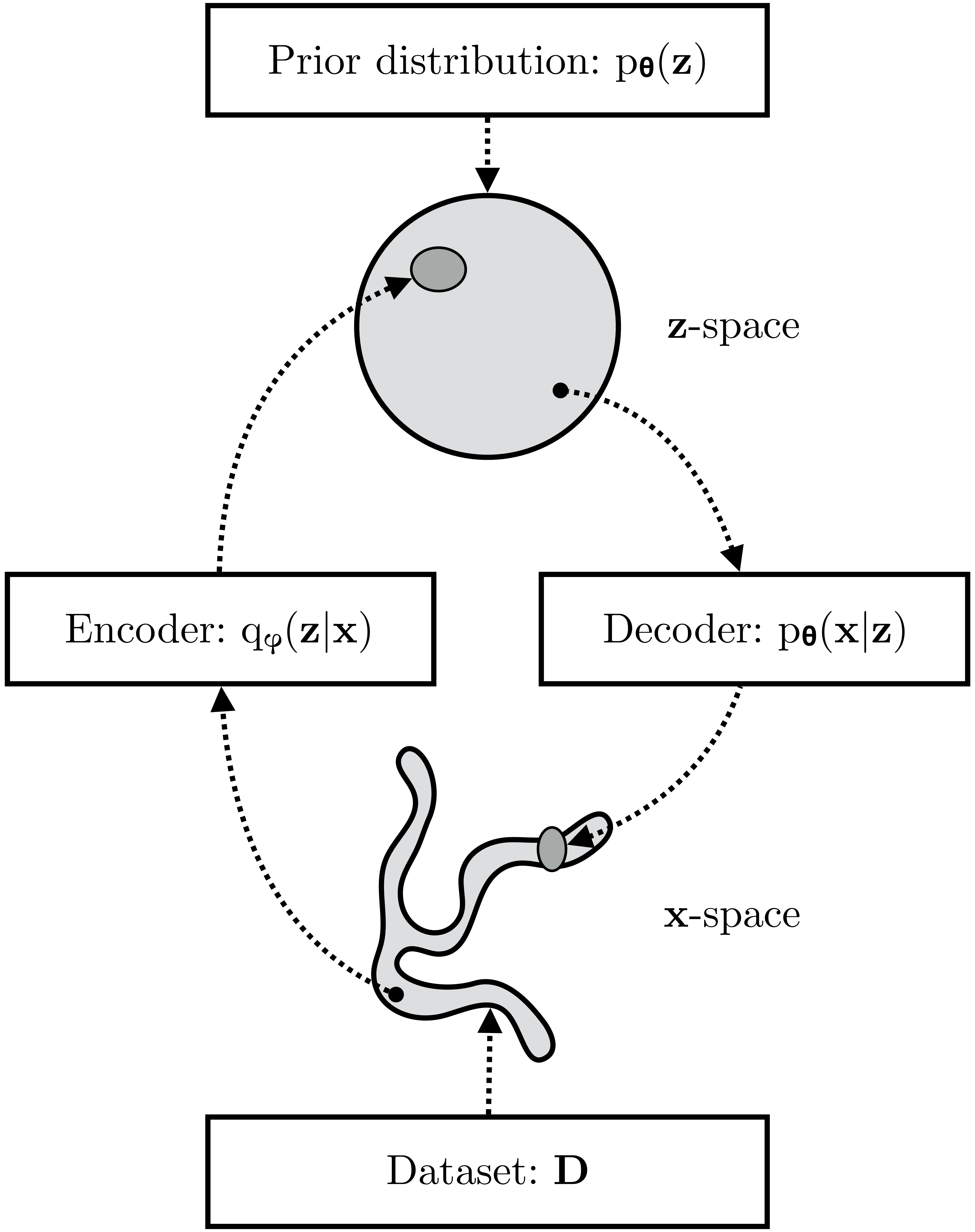
- Model density as $p(\mathbf{x}, \mathbf{z}) = p(\mathbf{x} \mid \mathbf{z})p(\mathbf{z})$.

- Approximate conditional density with a neural network: $p(\mathbf{x} \mid \mathbf{z}) \approx p_\theta(\mathbf{x} \mid \mathbf{z})$.

- Introduce second network to approximate posterior: $p(\mathbf{z} \mid \mathbf{x}) \approx q_\phi(\mathbf{z} \mid \mathbf{x})$.

- Infer $\theta,\ \phi$ by maximizing the evidence lower bound on the log likelihood:

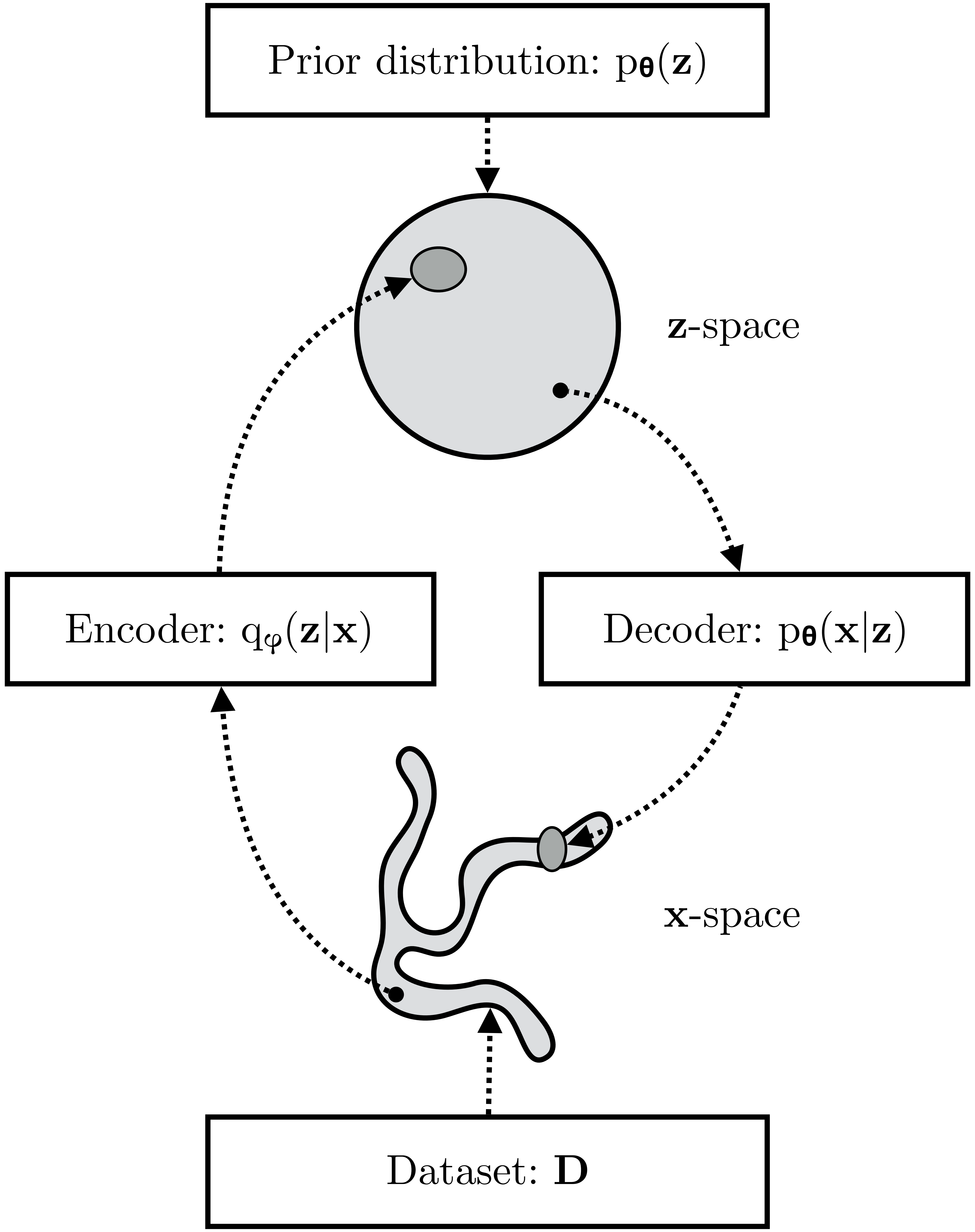
$$\mathscr{L}(\theta, \phi; \mathbf{x}) = \mathbb{E}_{q_\phi(\mathbf{z}_\mathbf{x})}\left[\ln p_\theta(\mathbf{x} \mid \mathbf{z})\right] - \mathbb{D}_{\mathrm{KL}}(q_\phi(\mathbf{z} \mid \mathbf{x}) \,||\, p_\theta(\mathbf{z}))$$



Variational Autoencoders

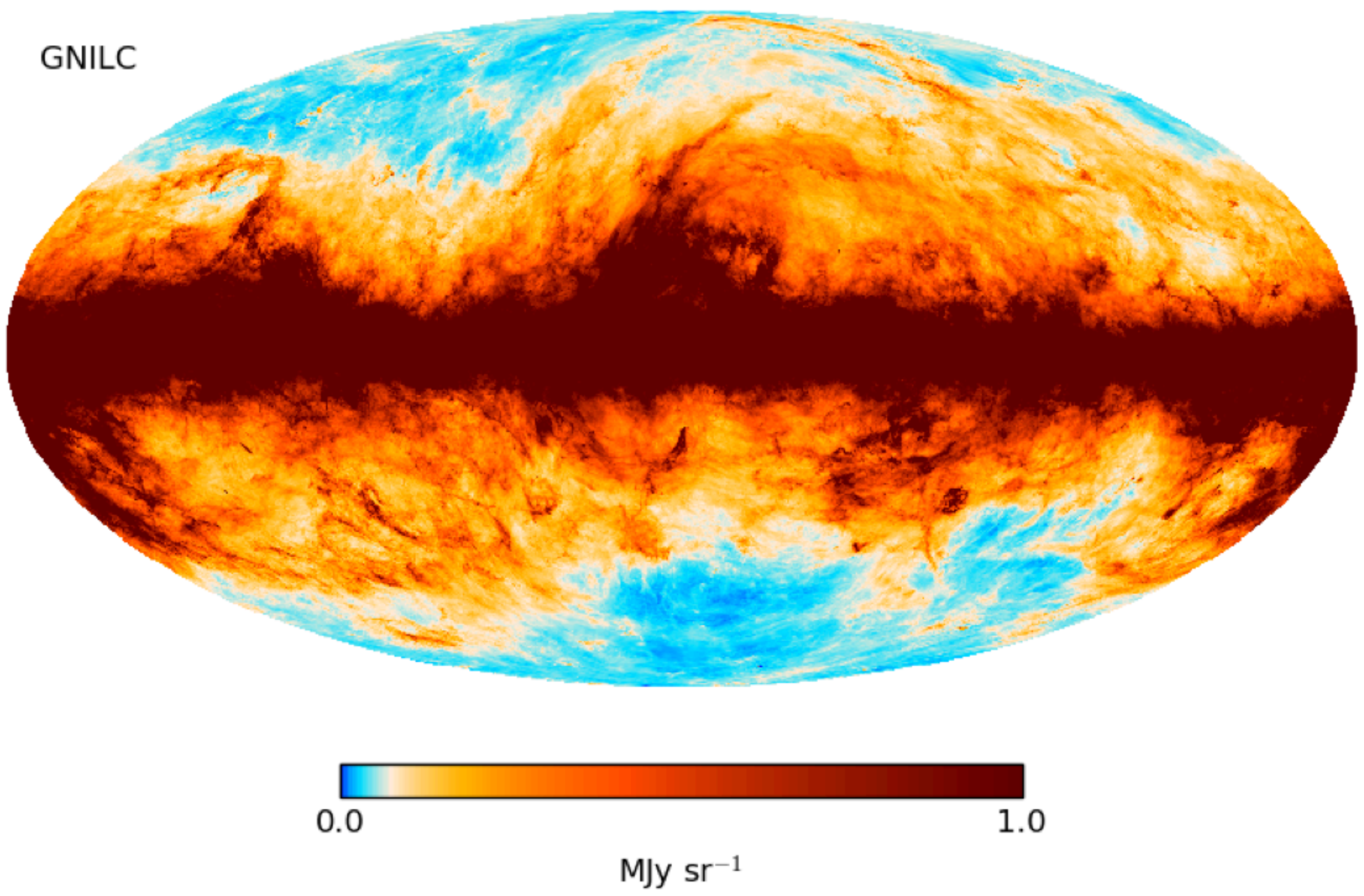
Architecture

- $p_\theta(\mathbf{z})$ is a 256-dimensional normal distribution.
- $p_\theta(\mathbf{x} \mid \mathbf{z})$ is assumed to be a normal distribution, parametrized by a set of simple transpose convolutions.
- $q_\phi(\mathbf{z} \mid \mathbf{x})$ is assumed to be Gaussian, and parametrized by a set of simple convolutions.



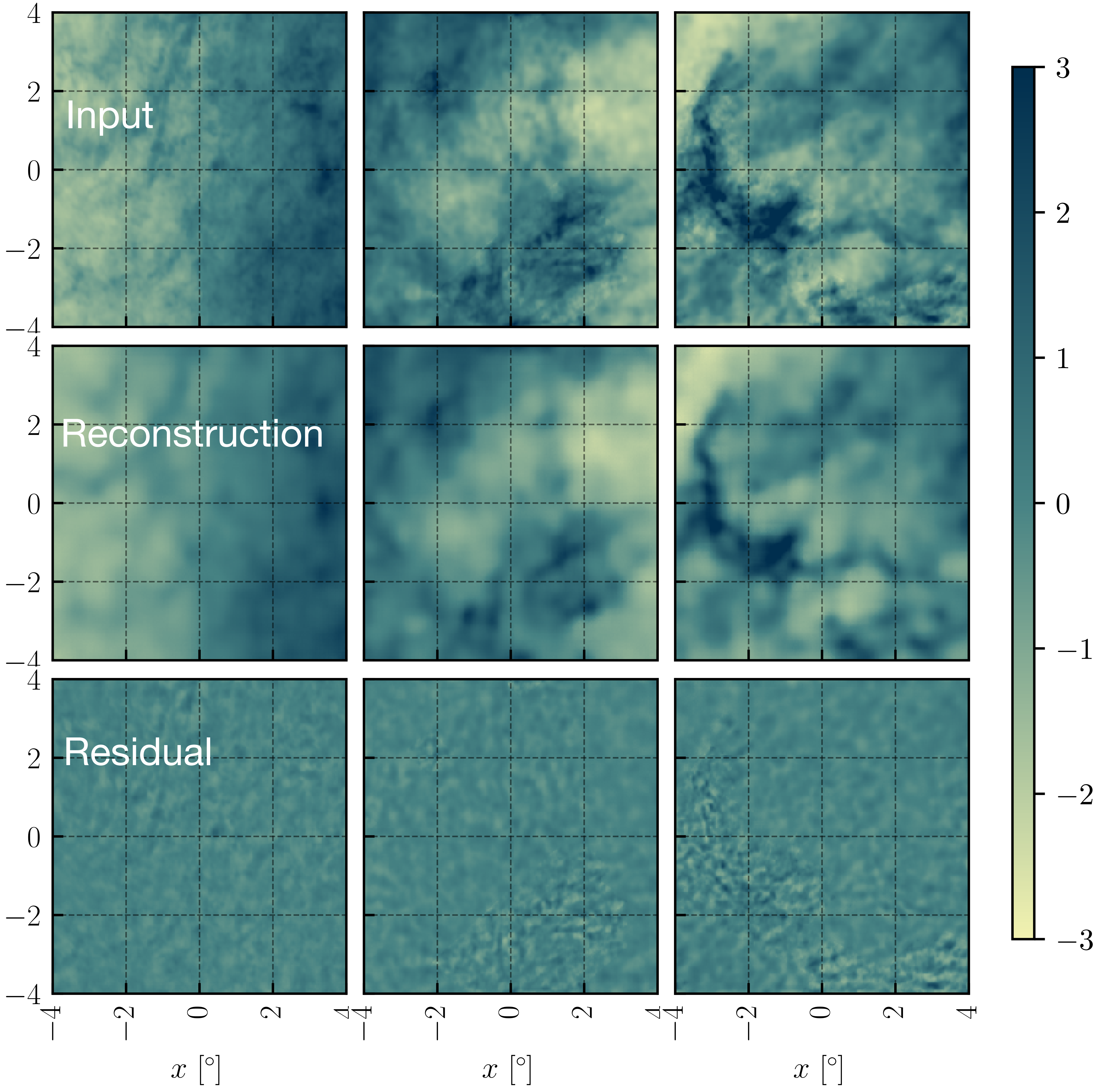
Training Set
Planck GNILC maps

- Planck GNILC intensity map at 545 GHz.
- Perform Cartesian projection on 64 sq. degree patches at $l > 15^{\circ}$, roughly 75% of sky, and 2254 cutouts.
- Randomly separate maps into 70% training set, 15% validation set, 15% test set.
- Augment training set with random flips and rotation applied on a per-image basis.



GNILC map in intensity at 545 GHz (Planck Collaboration XLVIII 2016).

Validation



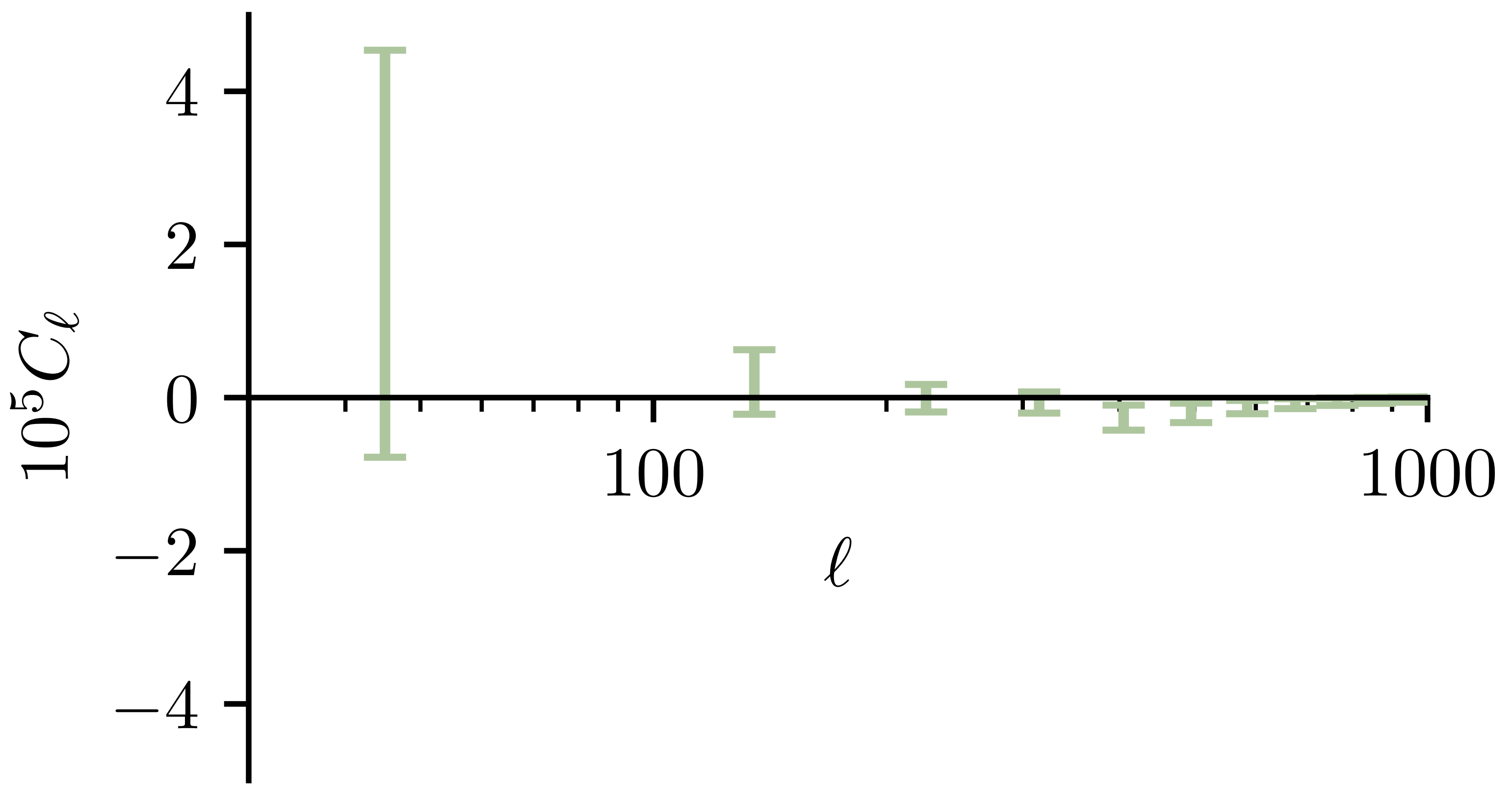
Validation

Reconstructions

- For each test set image compute the mean predicted map:

$$\tilde{\mathbf{x}} \approx \frac{1}{L}\sum_{l=1}^{L} \mathbf{g}_\theta(\mathbf{z}_{\text{test}}^{(l)})$$

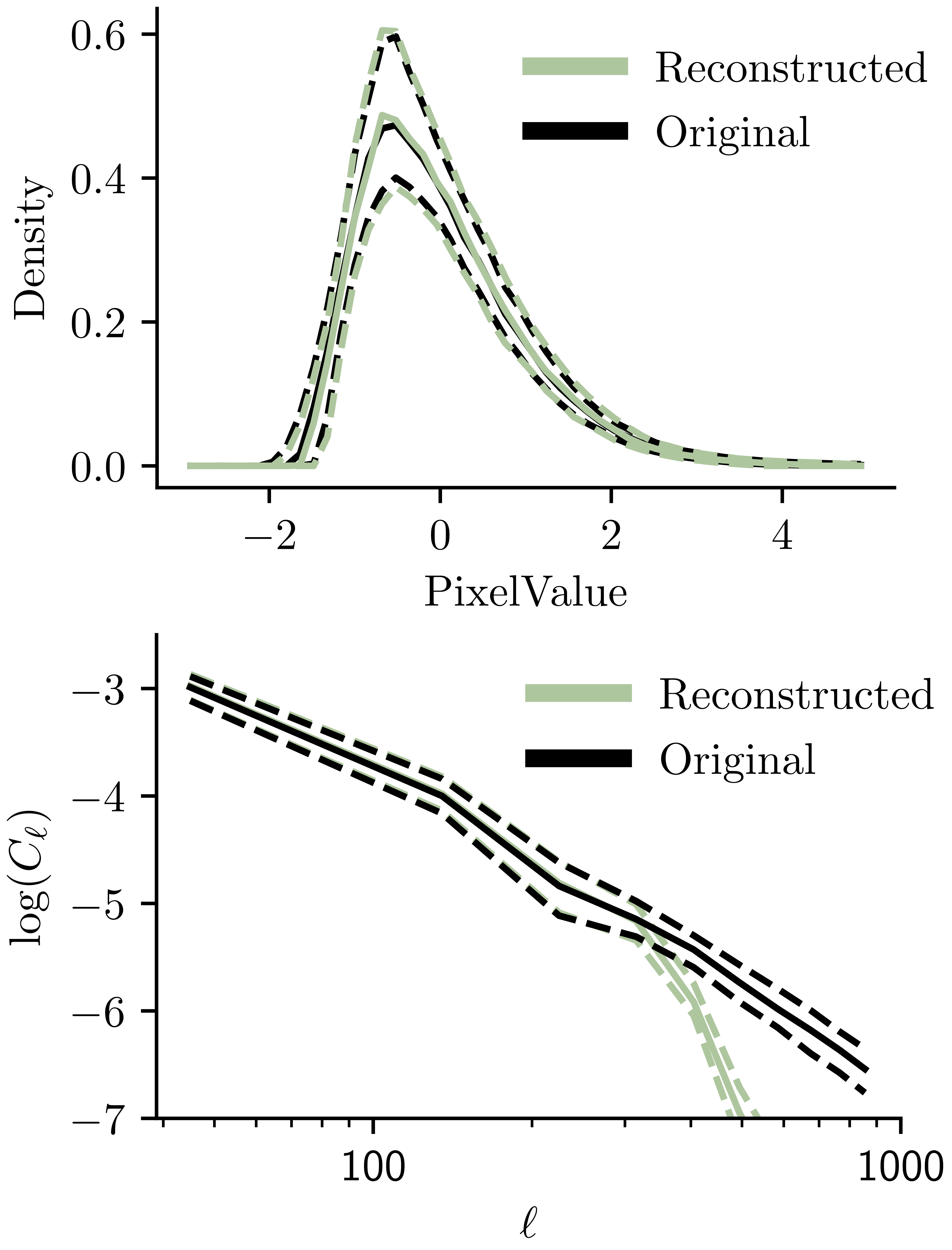
- Compute the residuals, $\mathbf{x}^{\text{res}} = \tilde{\mathbf{x}} - \mathbf{x}$, and cross correlate them with the input map: $C_\ell^{\text{res}} = \langle \mathbf{x}\mathbf{x}^{\text{res}} \rangle$



Validation

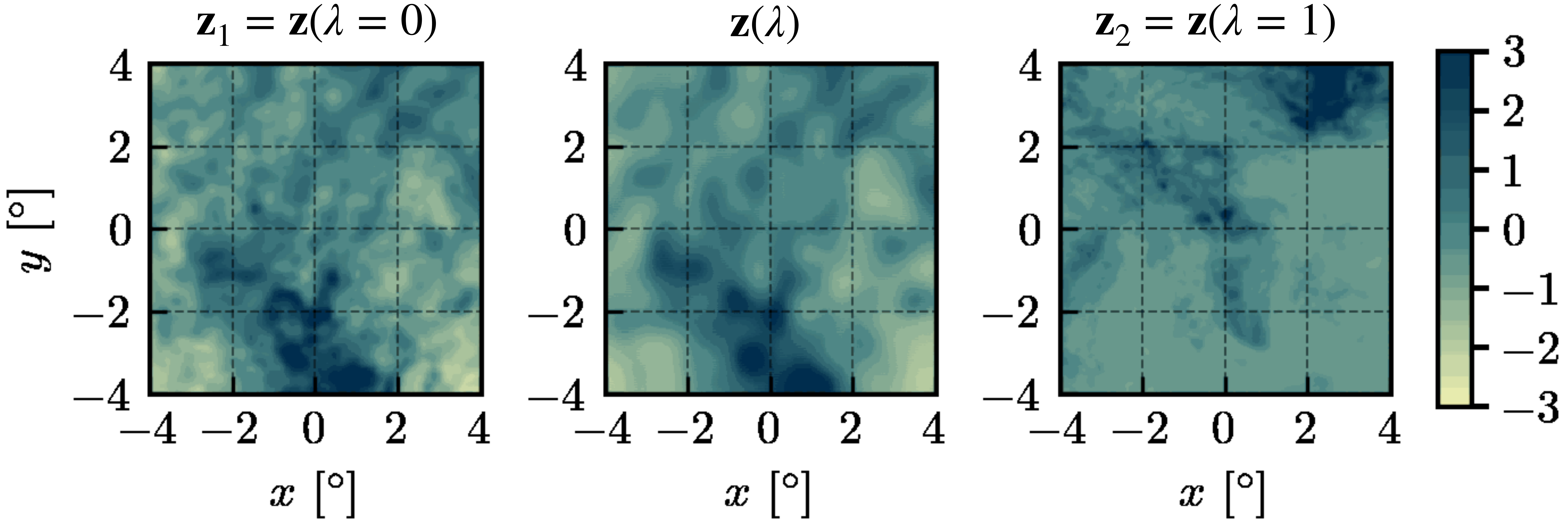
Summary Statistics

- Compute pixel histograms and power spectra on each test set image.
- Importantly, we recover summary statistics consistent with the test set, up to a particular *scale.*
- Past scales corresponding to $\ell \approx 400$, reconstruction quality declines sharply.
- This is uniform across all reconstructions in the test set.



Validation

Semantic Interpolation: $$\mathbf{z}_{1,2}(\lambda) = \frac{\sin((1-\lambda)\theta)}{\sin\theta}\mathbf{z}_1 + \frac{\sin(\lambda\theta)}{\sin\theta}\mathbf{z}_2$$



Application

Application

Data Imputation

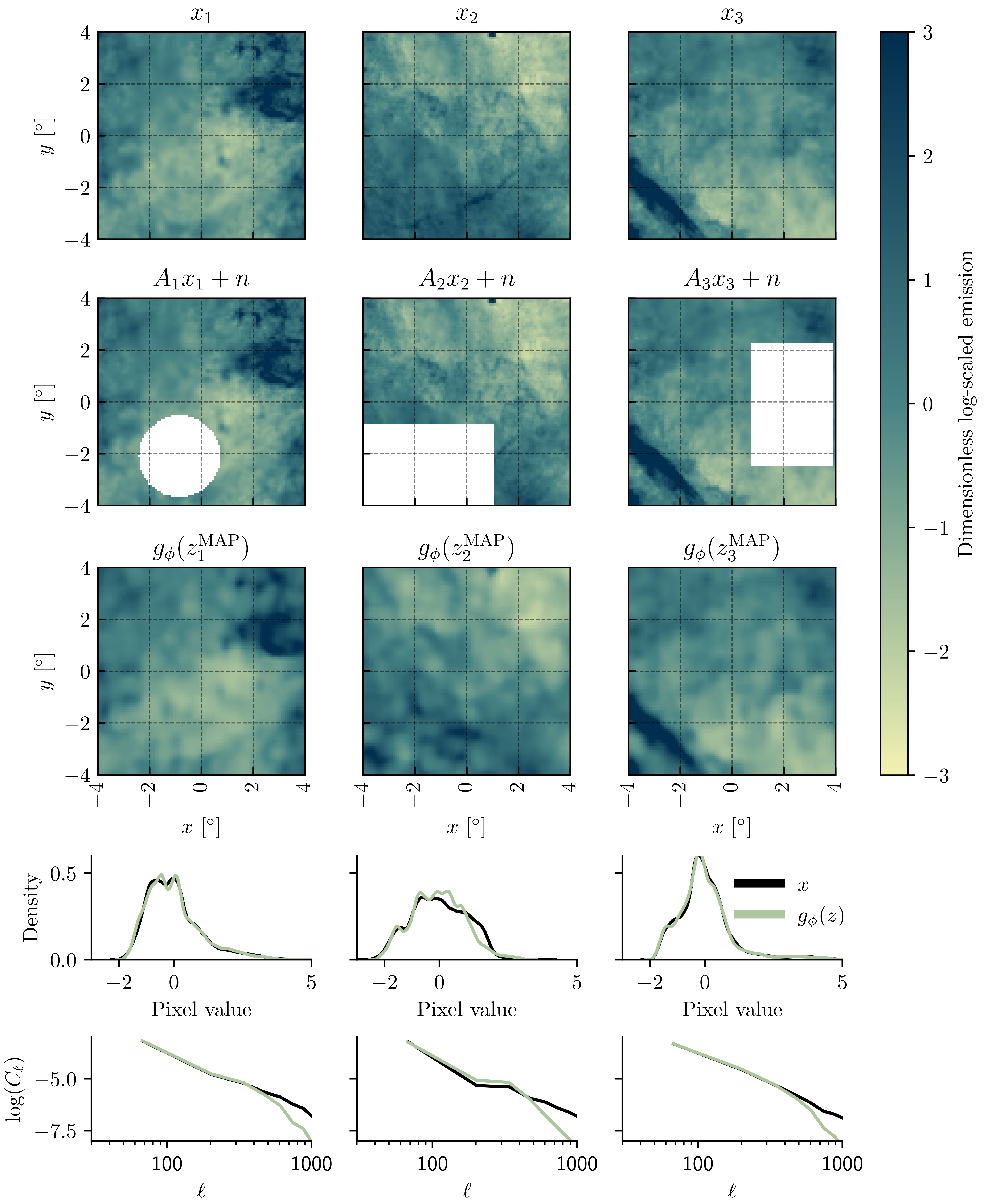
- Applying this to a denoising and inpainting problem. For data containing white noise, σ, and pixel-wise masking A, we can write down an expression for log posterior:

$$\log p(\mathbf{z}\,|\,\mathbf{d}) = \log p(\mathbf{z}) + \log p_\theta(\mathbf{d}\,|\,\mathbf{z}) - \log p(\mathbf{d})$$

- Substituting expressions for the prior, and in the case that the reconstruction error is smaller than the noise:

$$-2\log p(\mathbf{z}\,|\,\mathbf{d}_{\text{test}}) \propto \mathbf{z}^T\mathbf{z} + \frac{\mu_\theta(\mathbf{z})^T\mu_\theta(\mathbf{z})}{\sigma^2}$$

- In principle, this posterior can be sampled from to characterize the posterior uncertainty in the reconstructed image. Here we simply obtain a single MAP estimate by maximizing this quantity.



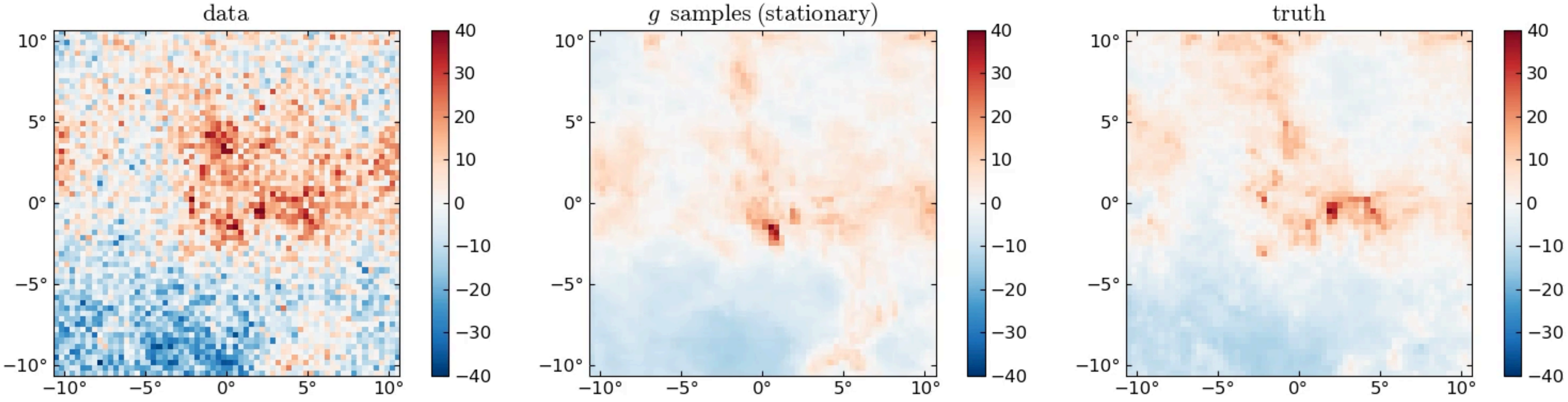
Sampling

(Very preliminary work with Marius Millea)

Remembering the Bayesian analysis pipeline we are interested in, we can write down our model:

$$d = \mathbb{A}\mathbb{L}(\phi)f + \mathbb{F}(\beta)g_{\text{dust}} + n$$

We simulate some mock data by taking a test set image, adding simulated CMB, beam smoothing, and adding noise. We then jointly sample the CMB fields, lensing potential, and dust latent space parameters:



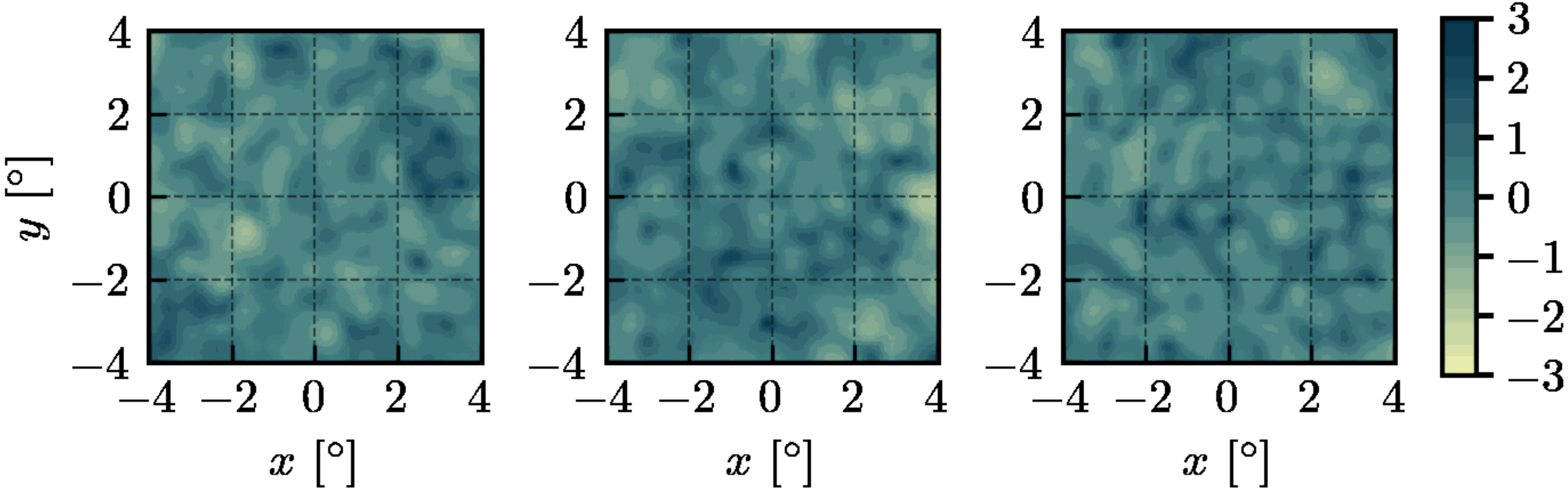
Conclusions

- We have built a variational auto encoder to perform density estimation for dust images.
- We successfully reconstruct test set images up to a scale of $\ell \approx 400$, beyond which we suffer rapid suppression of power.
- Novel samples are of a poor visual quality, and do not recover the correct statistics.
- Conditioned on corrupted real data, we perform realistic inpainting of Galactic foregrounds.
- The failure points seem consistent with a prior that is ill-fitting the true data posterior distribution. Currently we are exploring ameliorating these, e.g. 2-stage variational auto encoder, or considering more flexible flow-based encoders.

Backup Slides

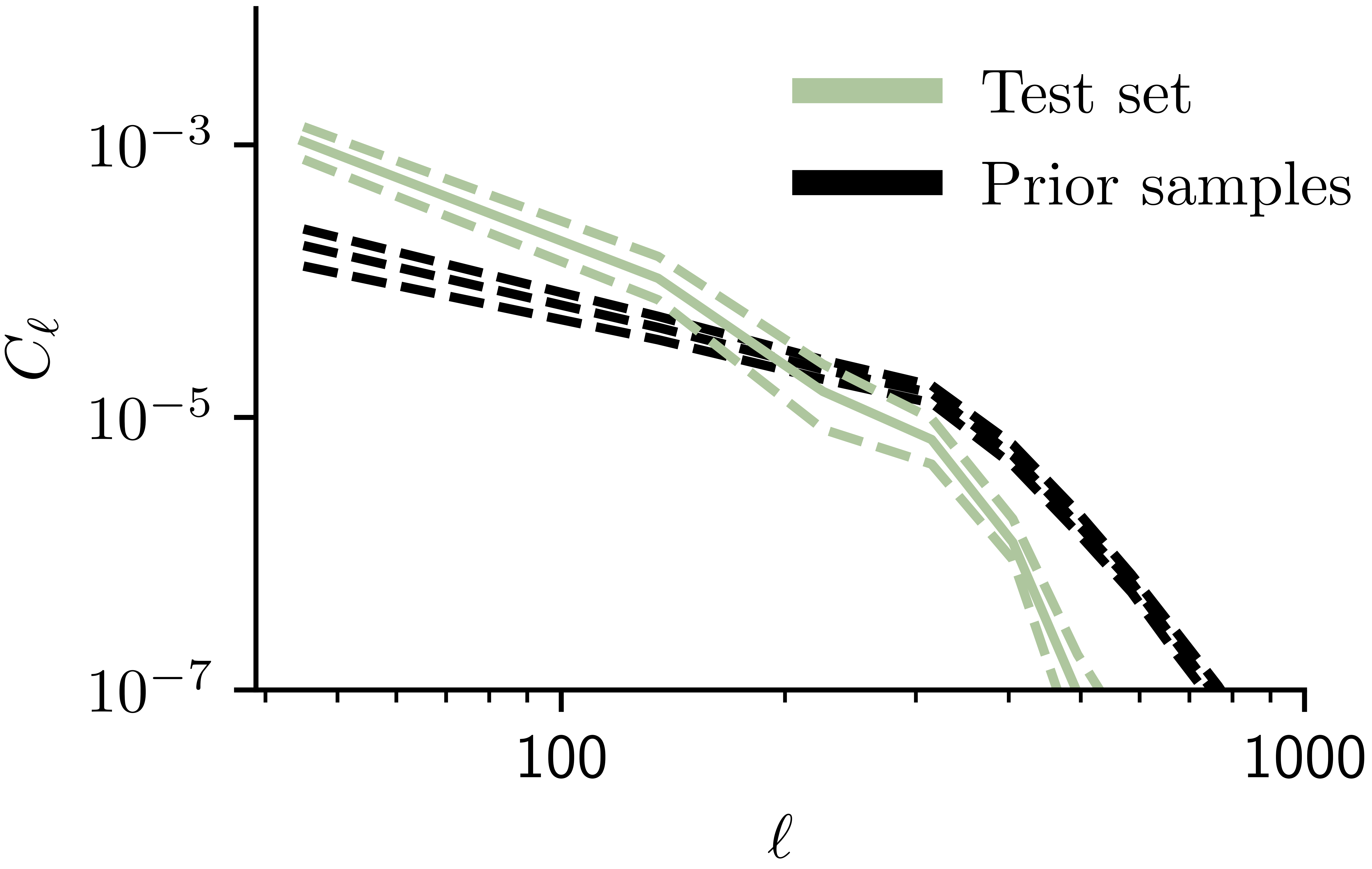
Validation

Prior Samples, $\mathbf{x} \sim p_\theta(\mathbf{x} \mid \mathbf{z})p(\mathbf{z})$



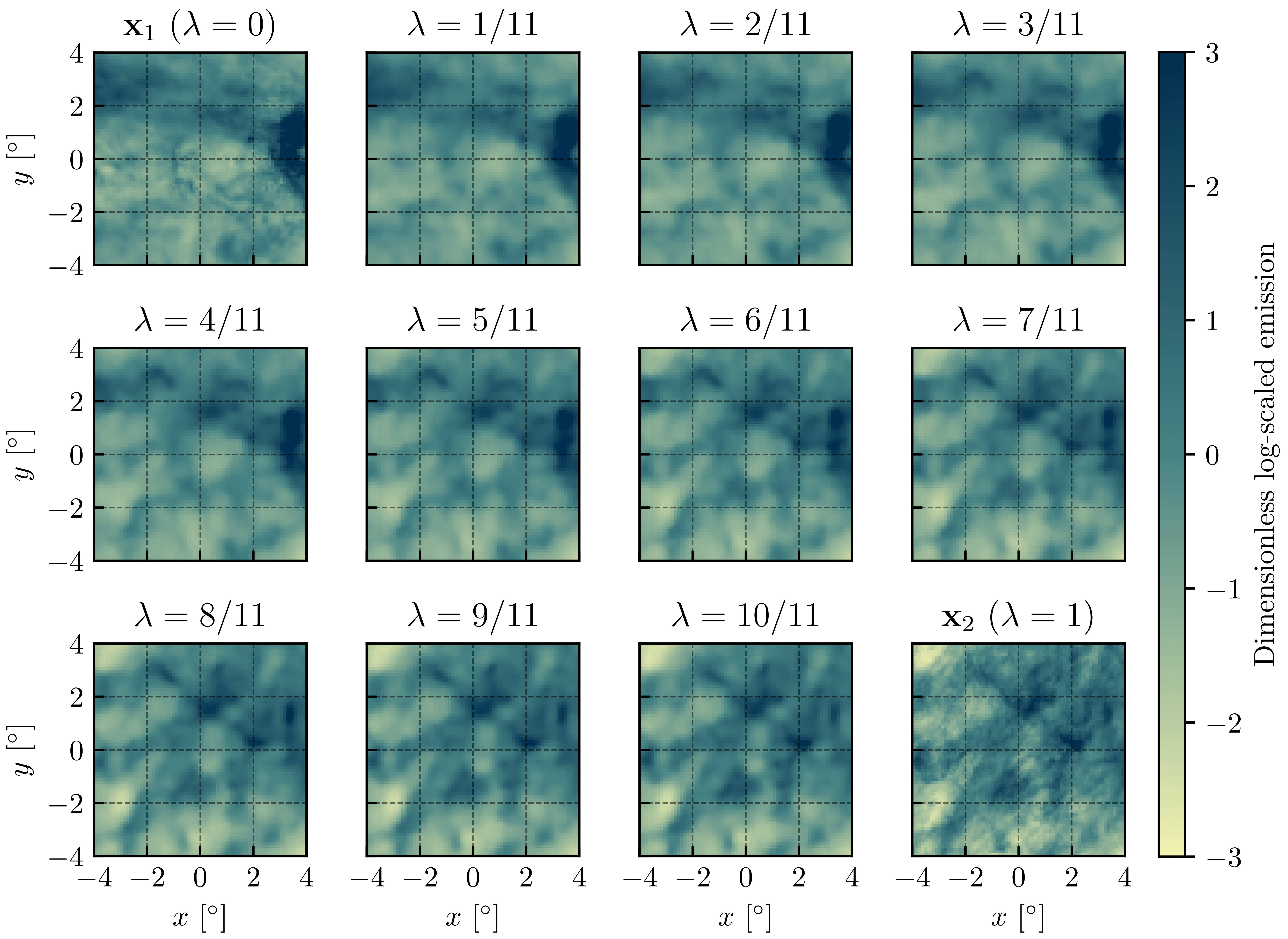
Validation

Power spectra of prior samples



Comparison of generative models

	Variational Autoencoders	GANs	Normalizing Flows
Bayesian Inference	✓	✗	✓
Stable Training	✓	✗	✓
Competitive Resolution	✗	✓	✓
Tractable in High Dimensions	✓	✓	✗



$\mathbf{x}_1$ ($\lambda = 0$)
$\lambda = 1/11$
$\lambda = 2/11$
$\lambda = 3/11$
$\lambda = 4/11$
$\lambda = 5/11$
$\lambda = 6/11$
$\lambda = 7/11$
$\lambda = 8/11$
$\lambda = 9/11$
$\lambda = 10/11$
$\mathbf{x}_2$ ($\lambda = 1$)
y [°]
x [°]
Dimensionless log-scaled emission