Data-Driven Strong Gravitational Lensing Analysis in the Era of Large Sky Surveys

Laurence Perreault-Levasseur

Ciela Institute

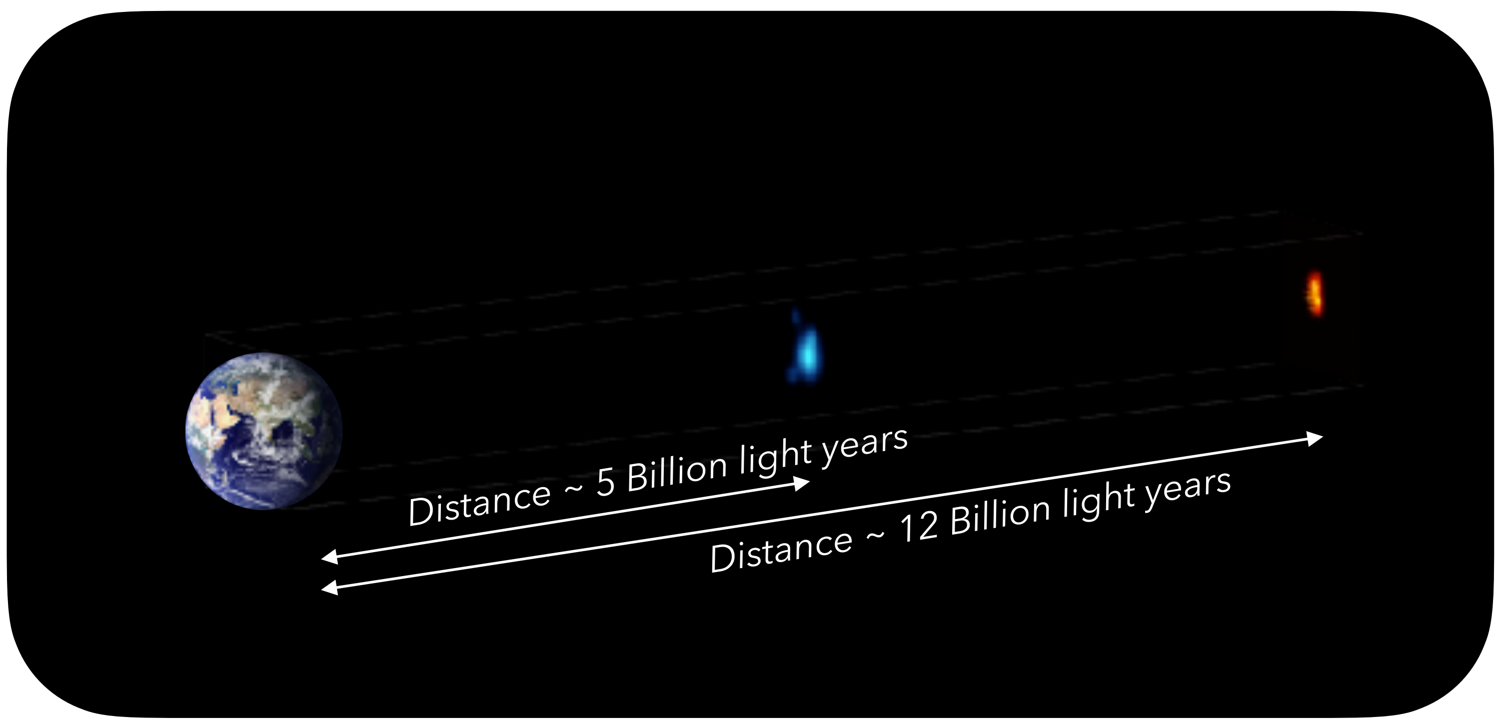
Université de Montréal

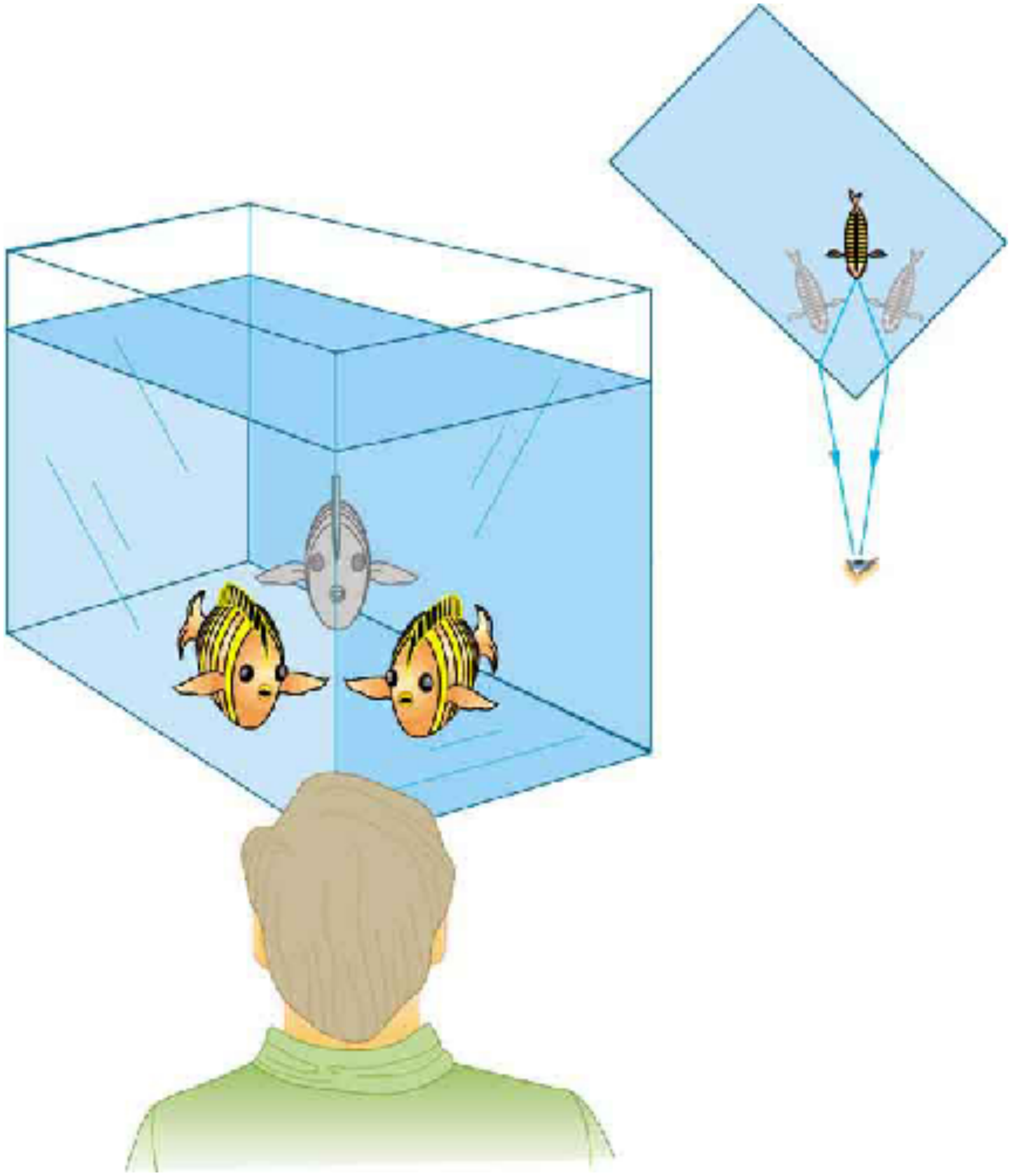
Mila

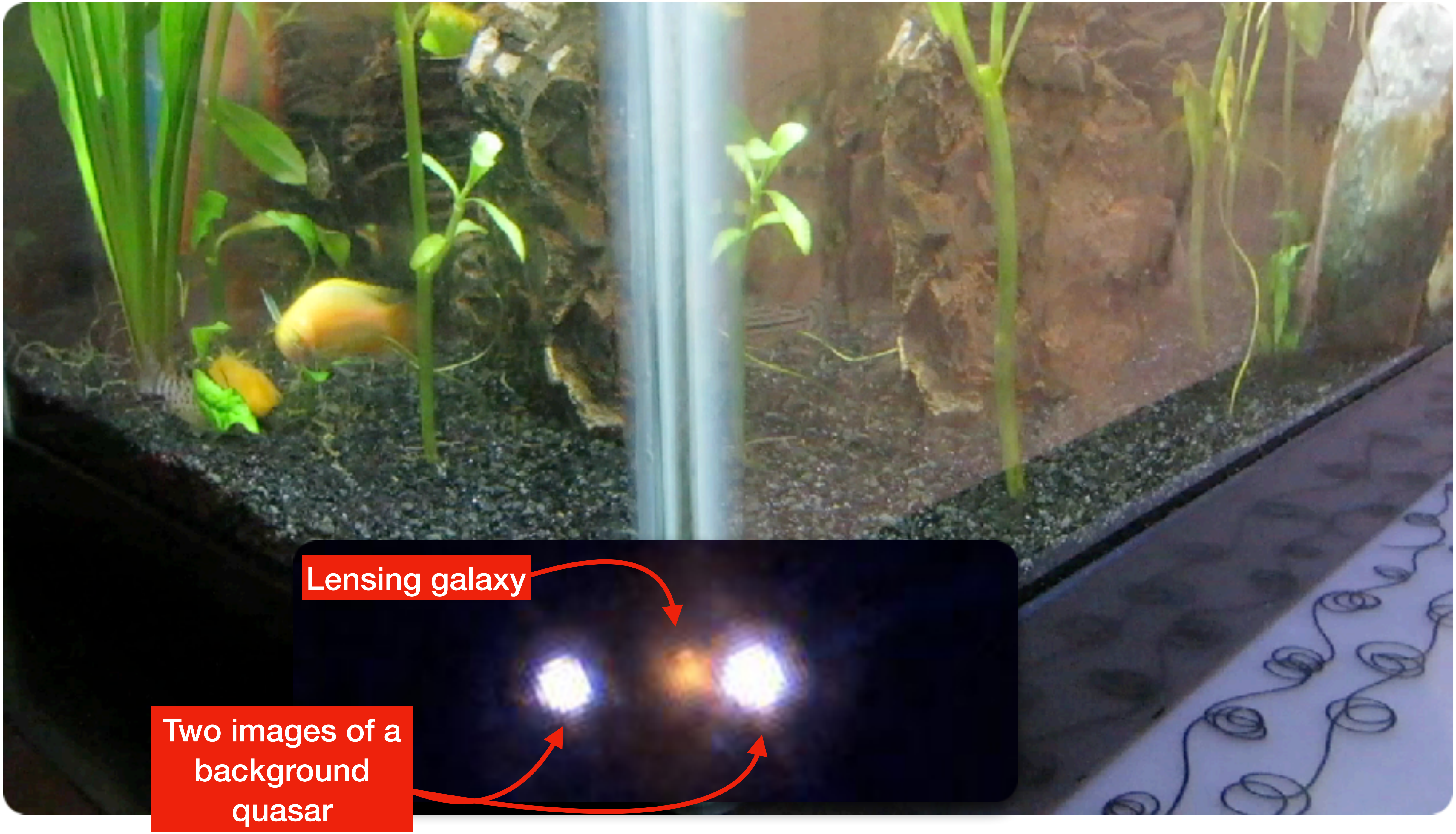
Flatiron Institute
Center for Computational Astrophysics

Strong Gravitational Lensing

Formation of **multiple images** of a single distant object due to the **deflection of its light** by the **gravity** of intervening structures.



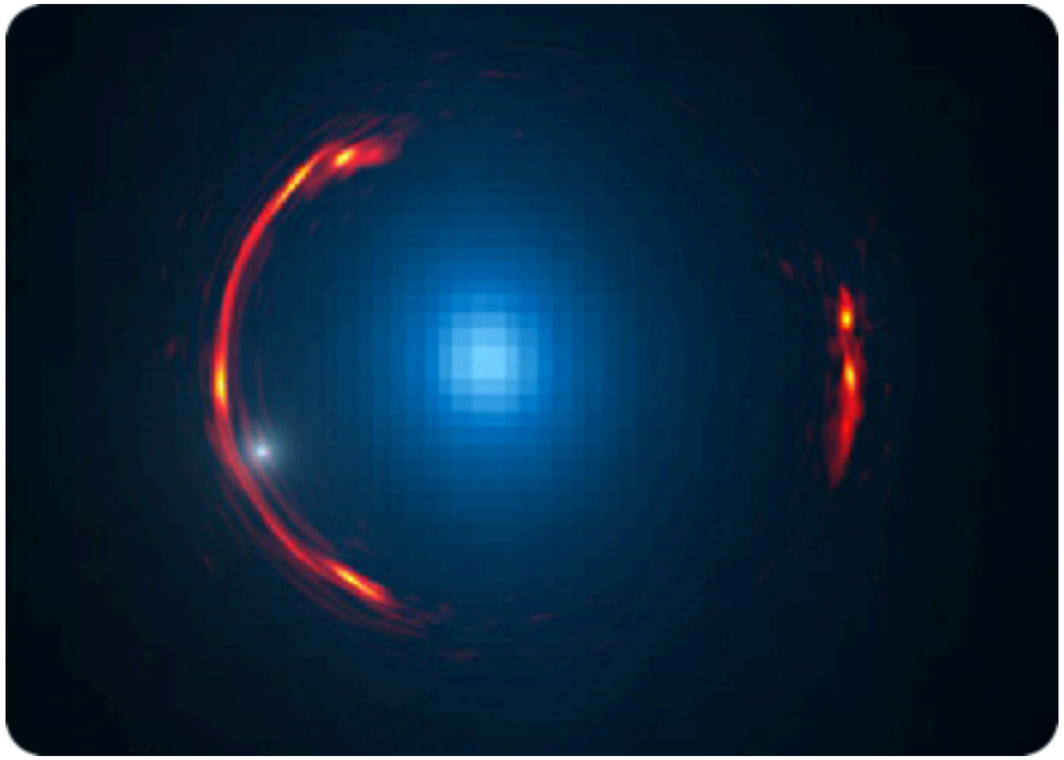



Lensing galaxy
Two images of a background quasar

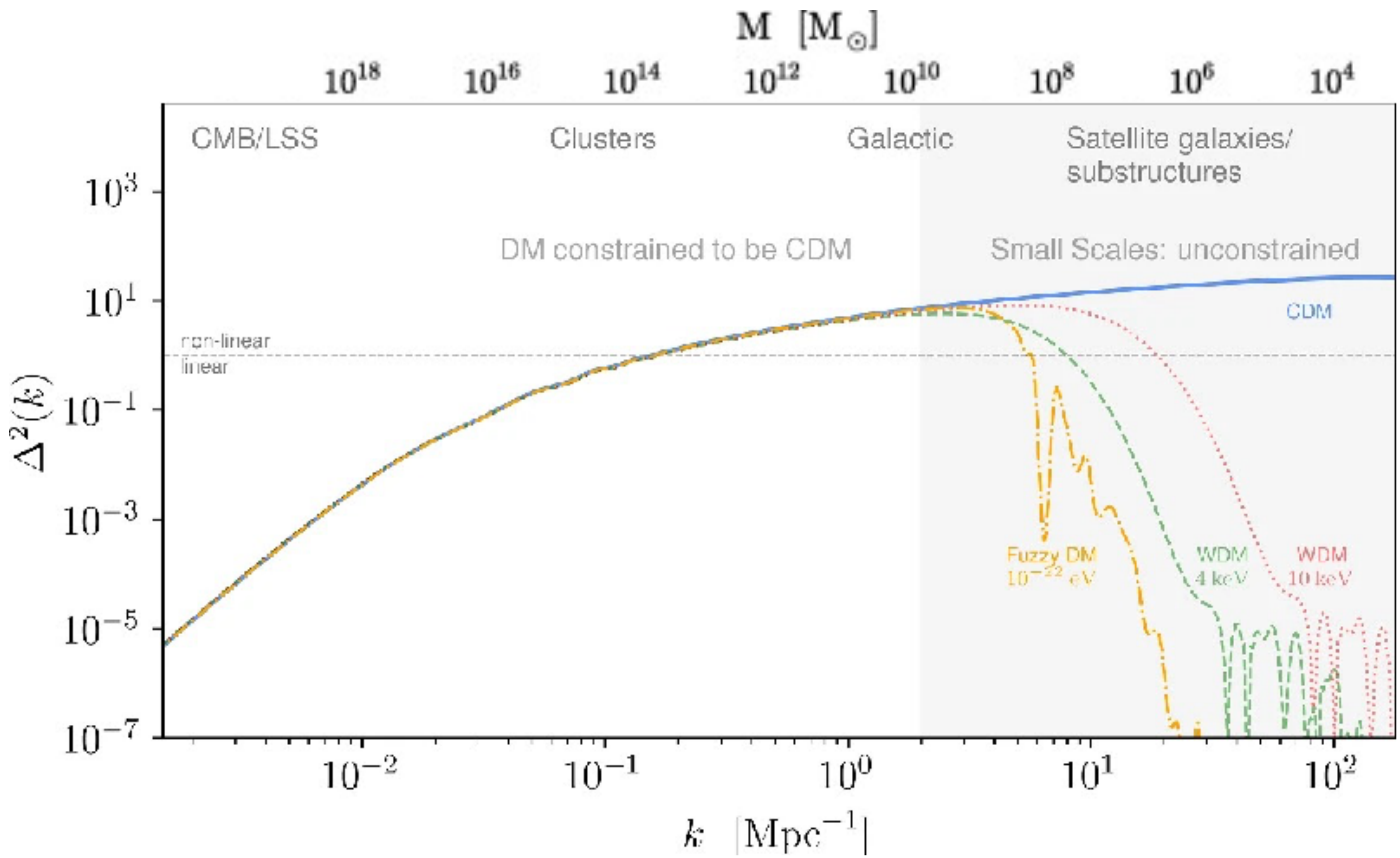
SCIENCE MOTIVATIONS FOR STRONG LENSING

1 - Use lensing to probe the **distribution of matter** in the lensing structures.

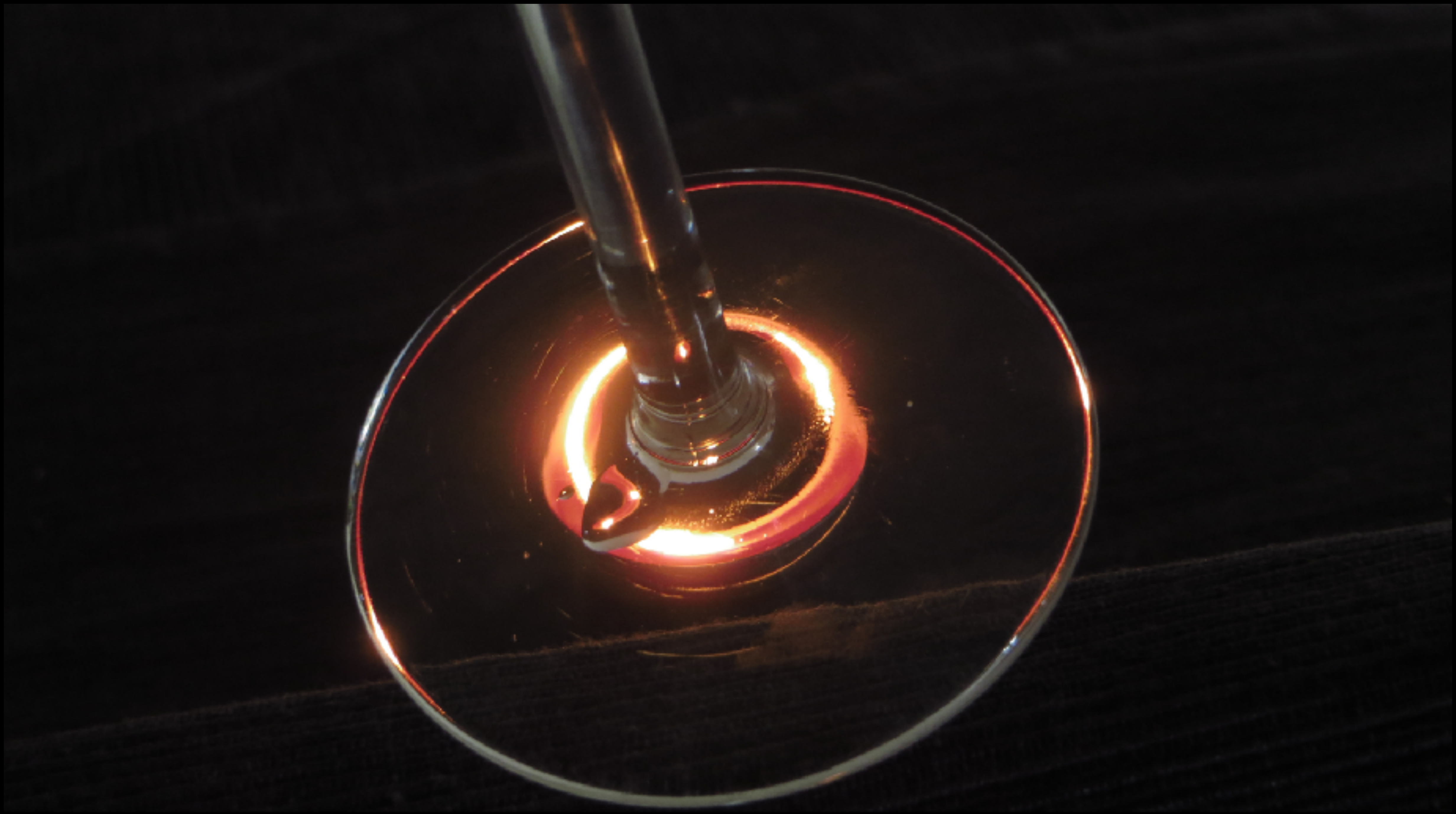
- Distortions in images are caused by **gravity**.
- They can be used to map the **distribution of matter** in the lens.
- Particularly useful for studying **dark matter.**



MATTER POWER SPECTRUM



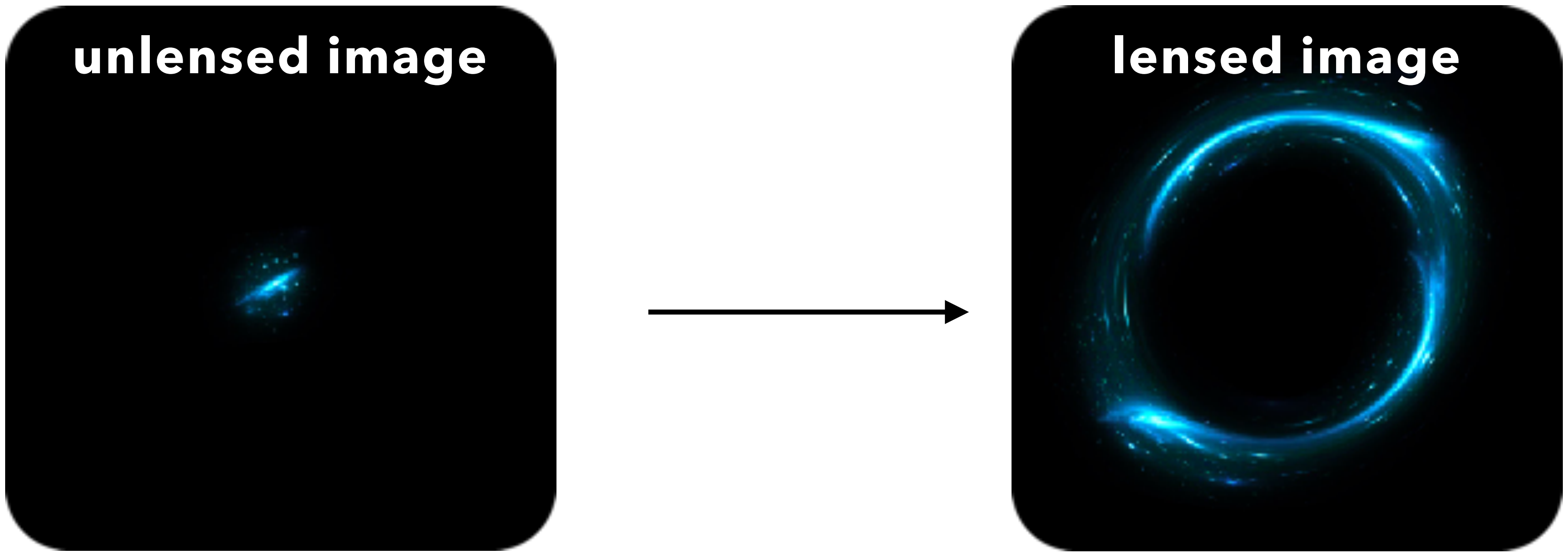
Ferreira et al., A&A Review, 2021



Science motivations for strong lensing

2 - Use strong lensing as a **cosmic telescope**.

- Lensing **magnifies** the images of sources and makes them appear **brighter**.
- This allows us to study some of the most distant galaxies of the universe that would have been otherwise below our sensitivity or resolution limits.

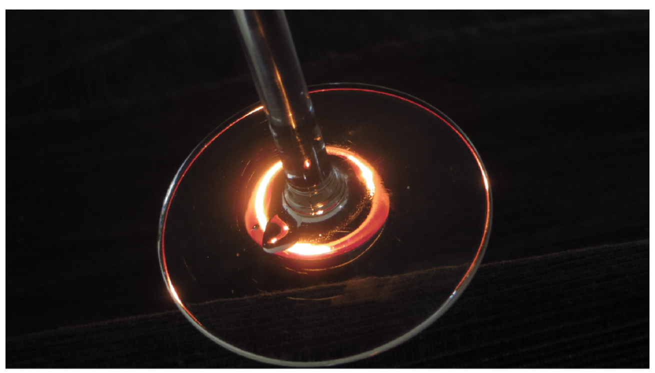


SCIENCE MOTIVATIONS FOR STRONG LENSING

3 - Measure **comological parameters** (H_0).

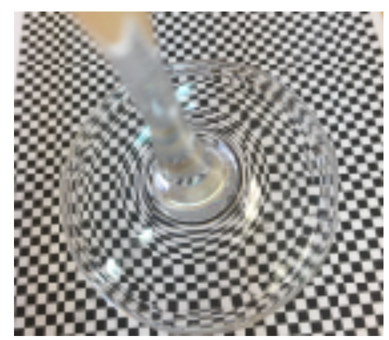
- Different images are produced because light follows **different paths**.
- These paths are of **different lengths**.
- If the source has time variability, this will cause **time delays** between different images.

LENSING ANALYSIS



Data

1: Morphology of the background source (the true, undistorted image of the candle)



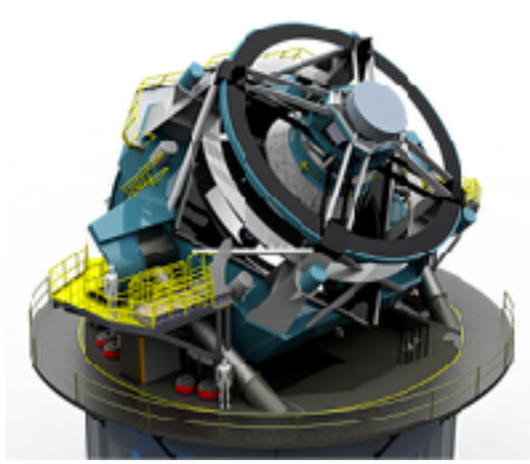
2: Matter distribution in the lens (the shape of the wineglass)

$$y = L(p)x + n$$

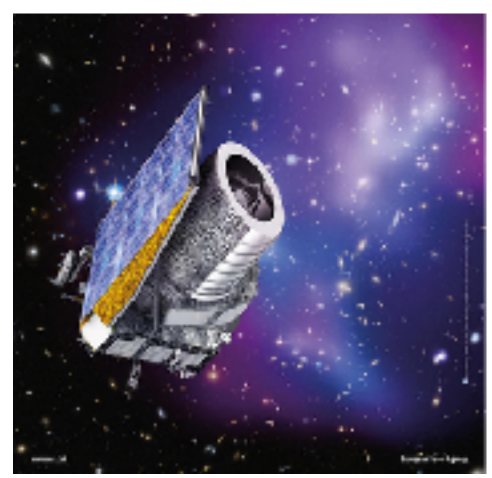
- Data → y
- Source Parameters (linear) → x
- Lens Parameters (non-linear) → p
- Noise → n

LOOKING INTO THE FUTURE

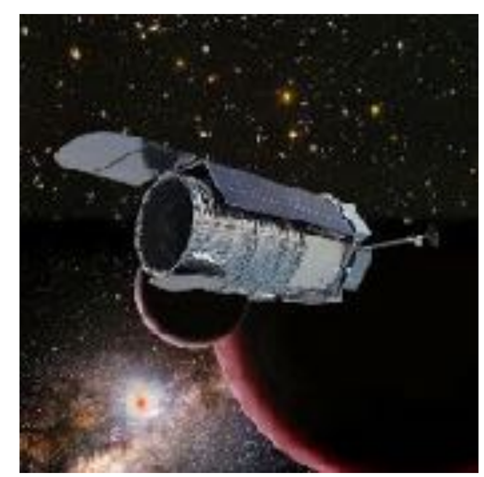
In the next few years, we're expecting to discover more than 170,000 new lenses.



LSST



Euclid consortium



NANCY GRACE ROMAN SPACE TELESCOPE

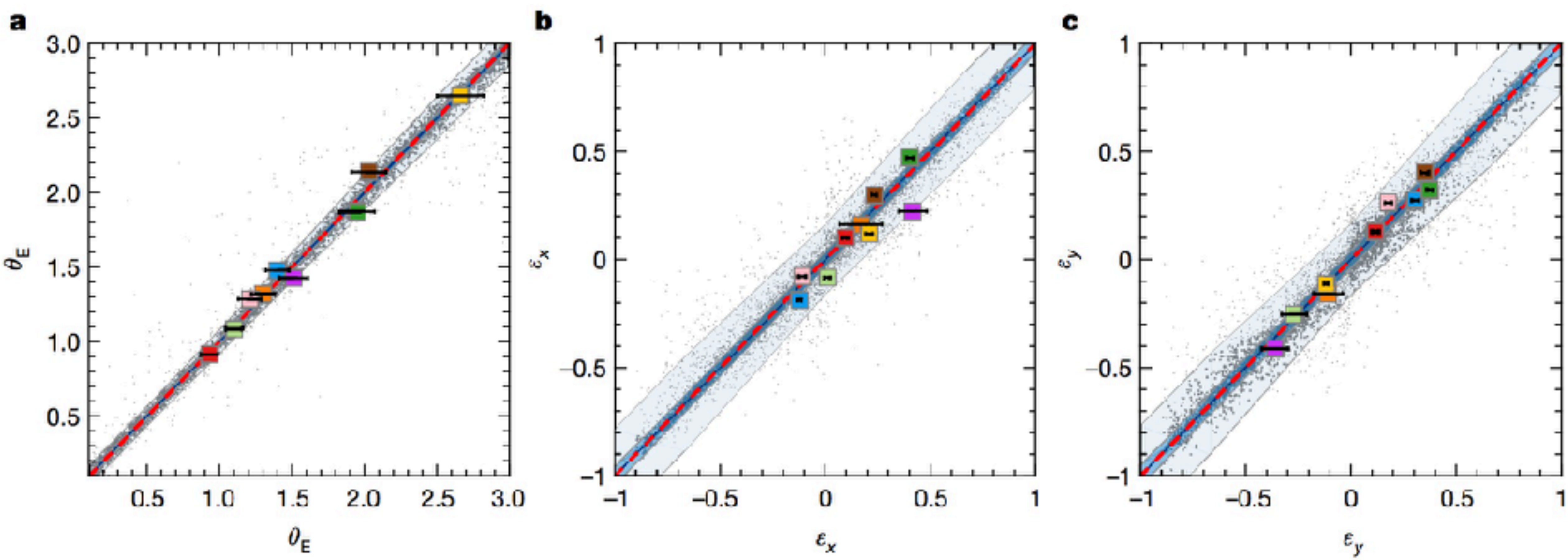
Methods for the future:

How are we going to analyze 170,000 lenses?

- Lens modeling is **very slow**.
- Simple lens model takes ~3 days

=> **1,400 years !**

ESTIMATING THE MATTER DISTRIBUTION PARAMETERS WITH CNNS

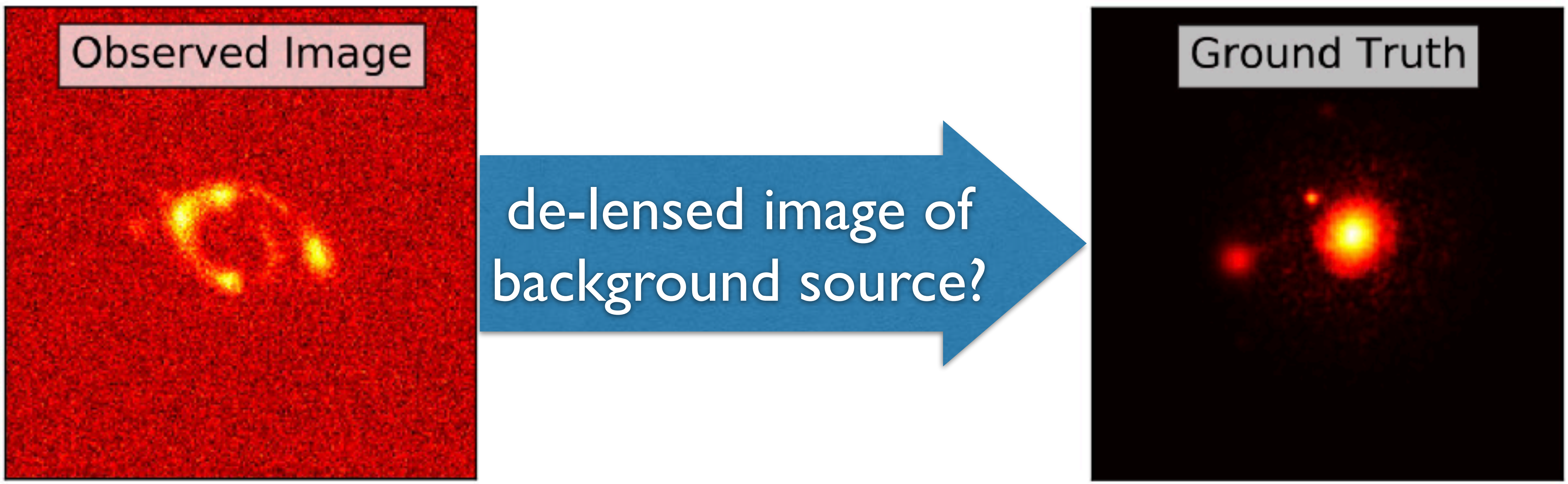


10 million times faster than traditional lens modeling.

0.01 seconds on a **single GPU**

Hezaveh, Perreault Levasseur, Marshall, Nature, 2017

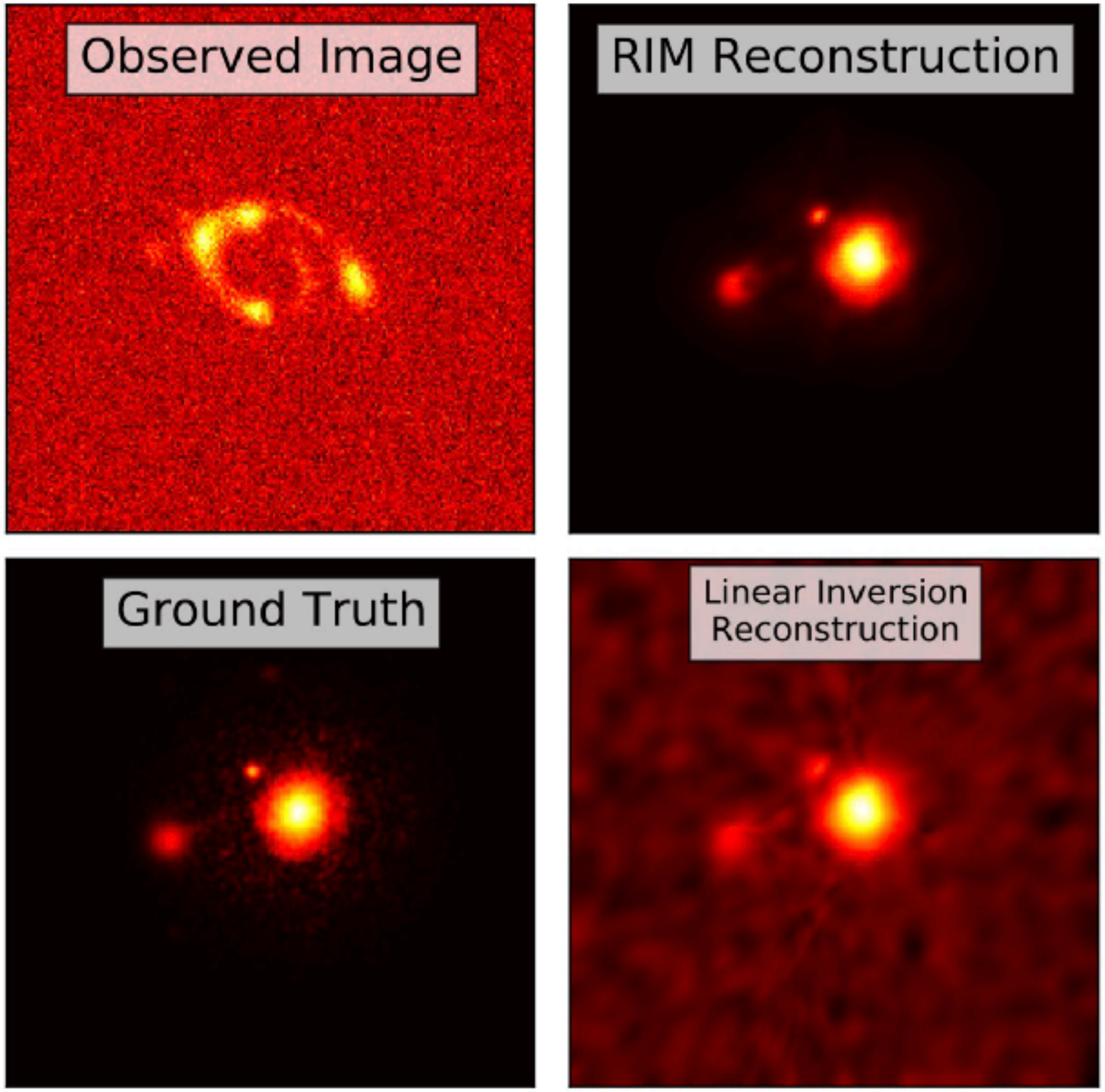
UNDISTORTED IMAGE OF THE BACKGROUND SOURCE



UNDISTORTED IMAGE OF THE BACKGROUND SOURCE

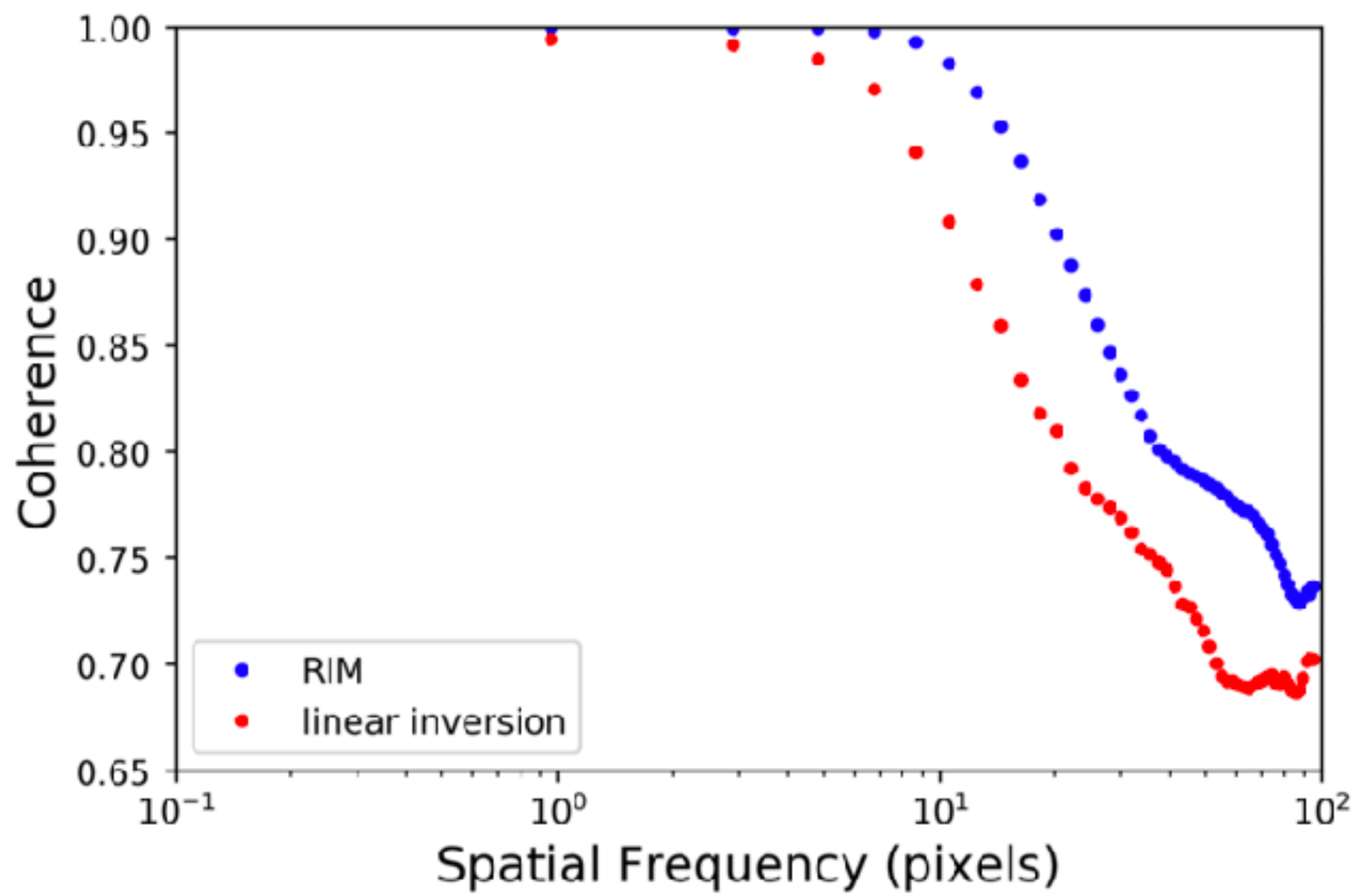
$$S = (L^T C_N^{-1} L + C_p^{-1})^{-1} L^T C_N^{-1} D$$

Undistorted image of the background source With the Recurrent Inference Machine (RIM)



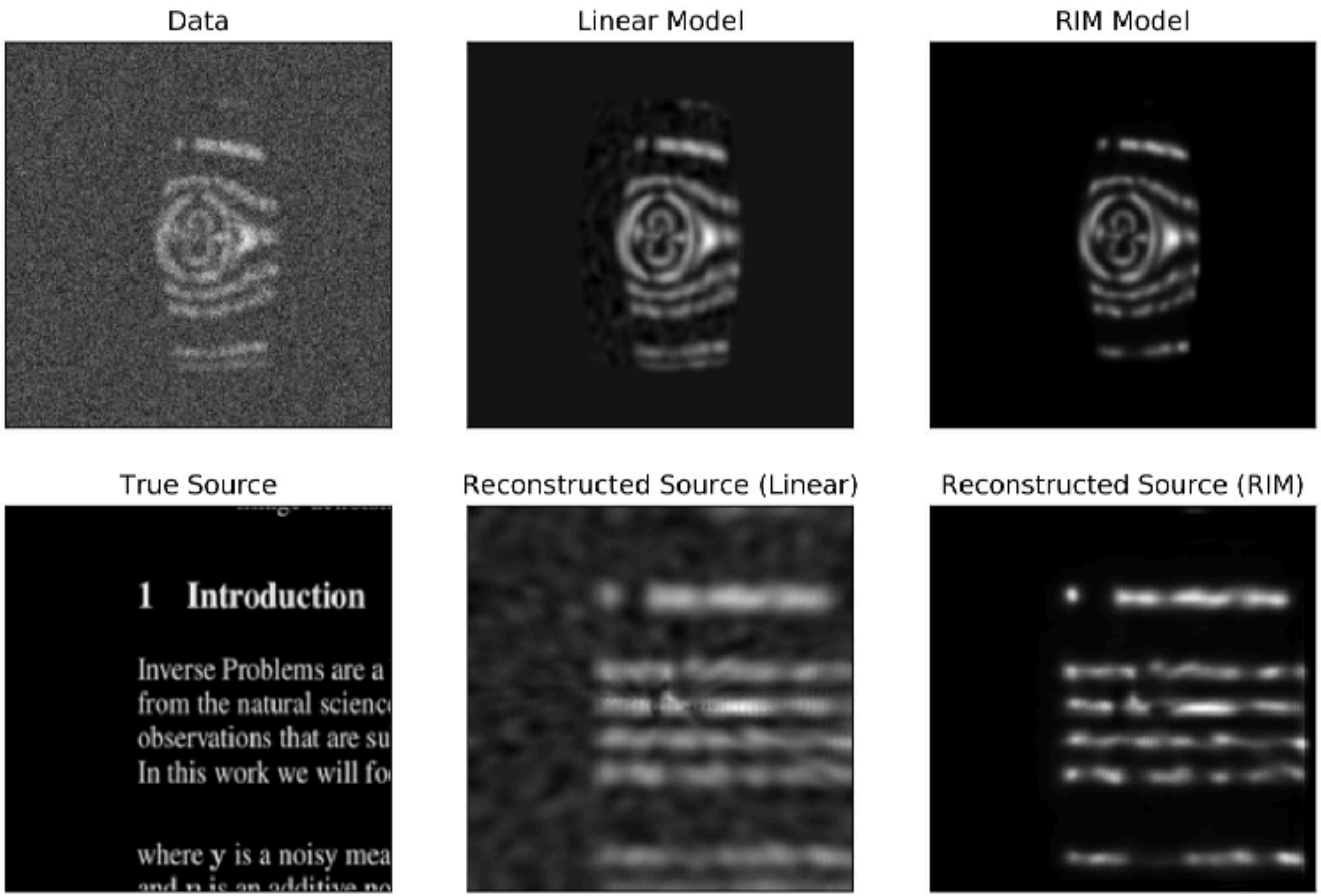
Morningstar, Perreault Levasseur et al., 2019

Background source reconstruction: Comparison to maximum likelihood methods

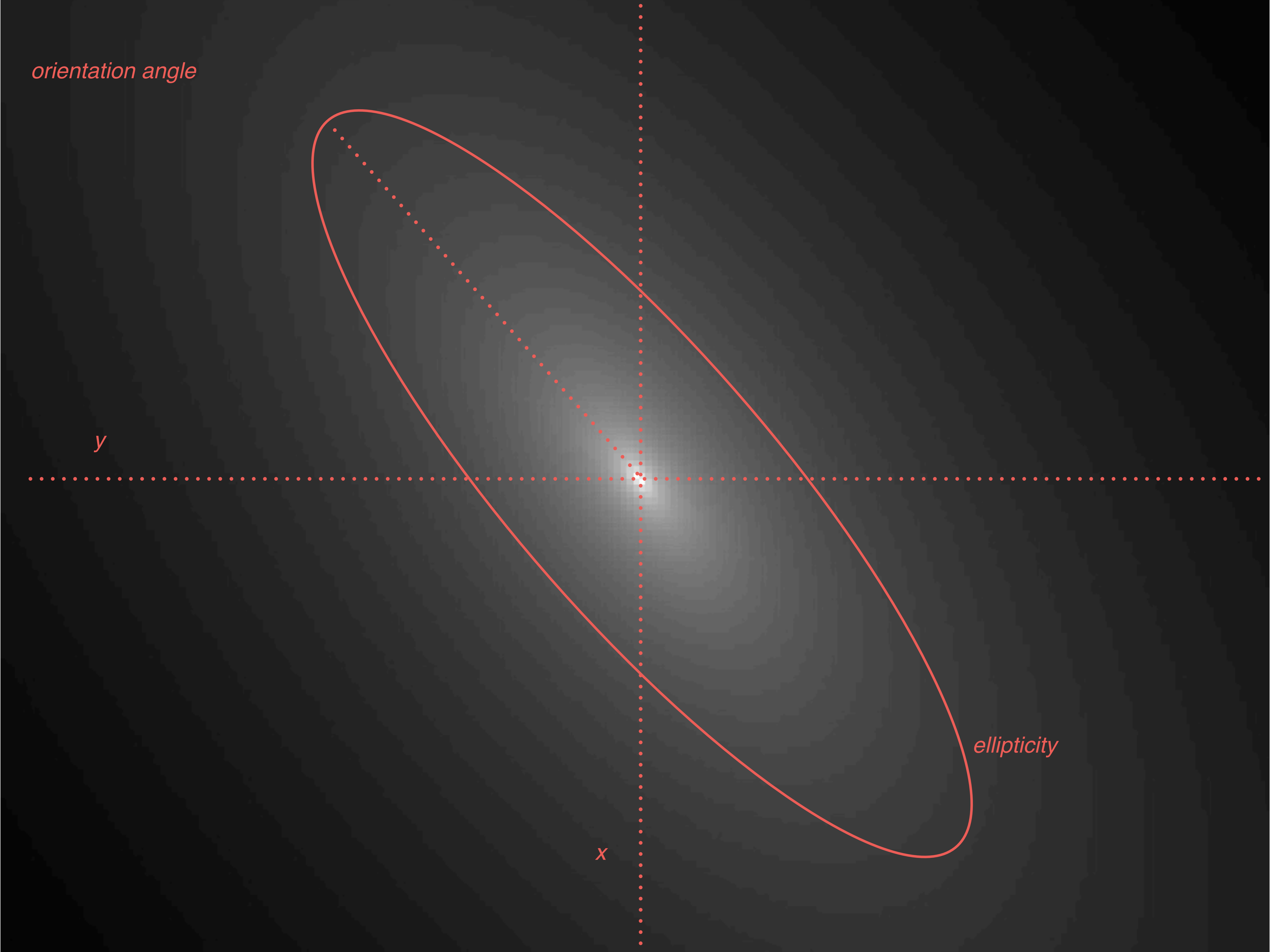


Morningstar, Perreault Levasseur et al., 2019

EXAMPLES OUTSIDE THE TRAINING DATA



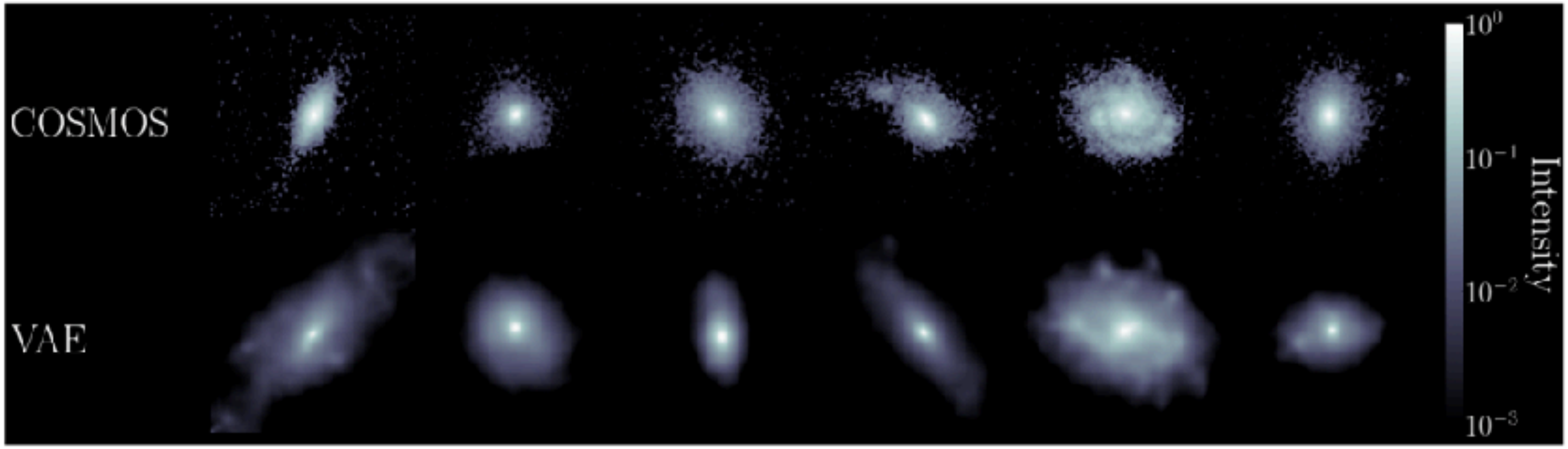
Morningstar, Perreault Levasseur et al., 2019

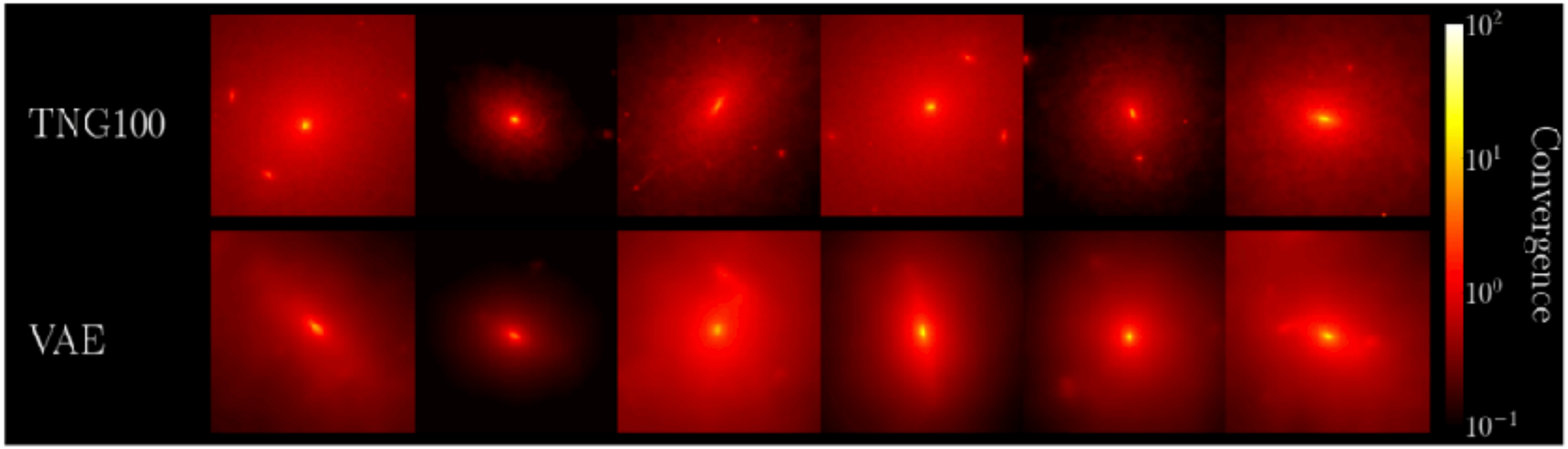


orientation angle
y
x
ellipticity

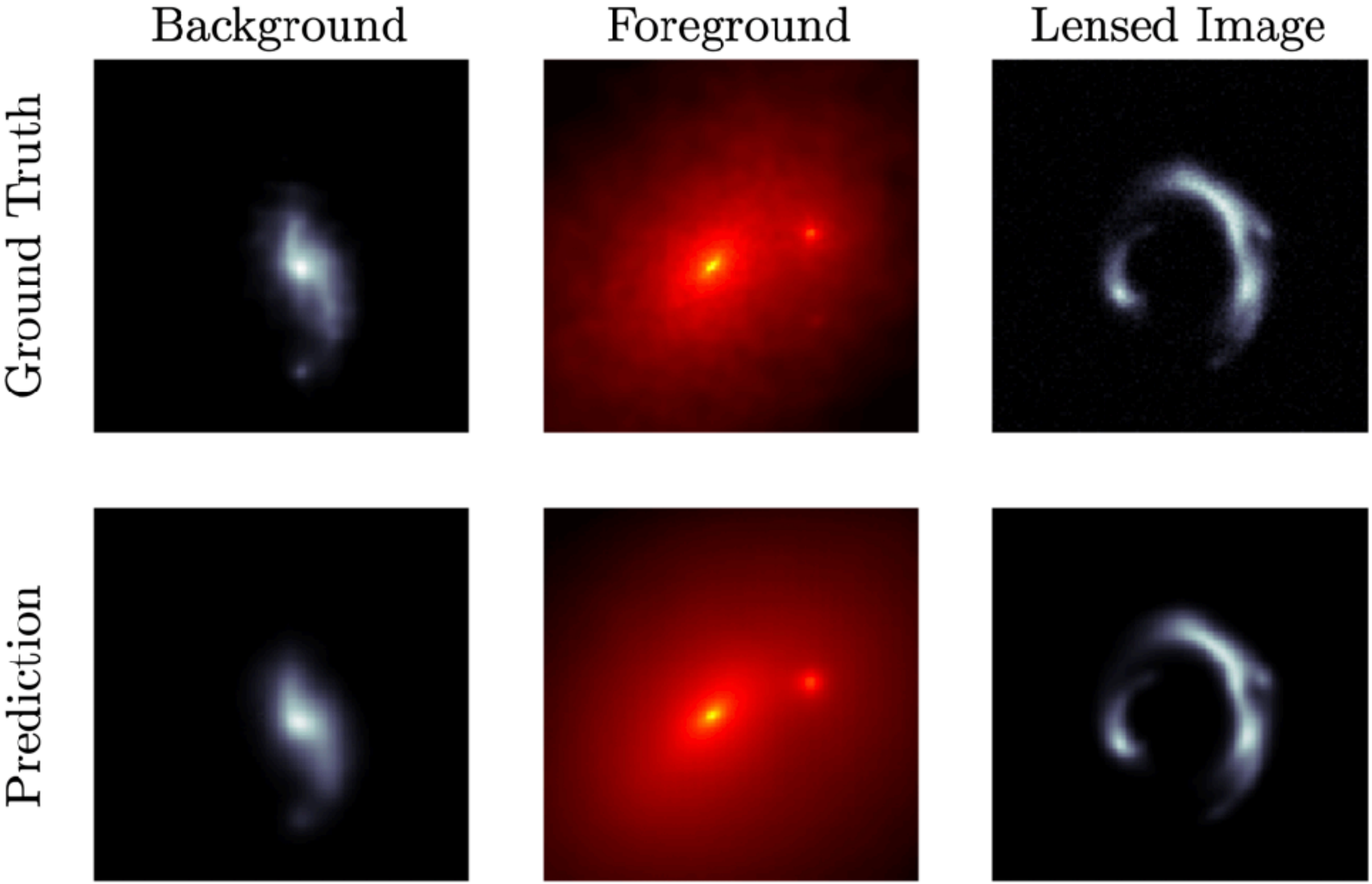
Simulated galaxies generated with a Variational Autoencoder

Alexandre Adam

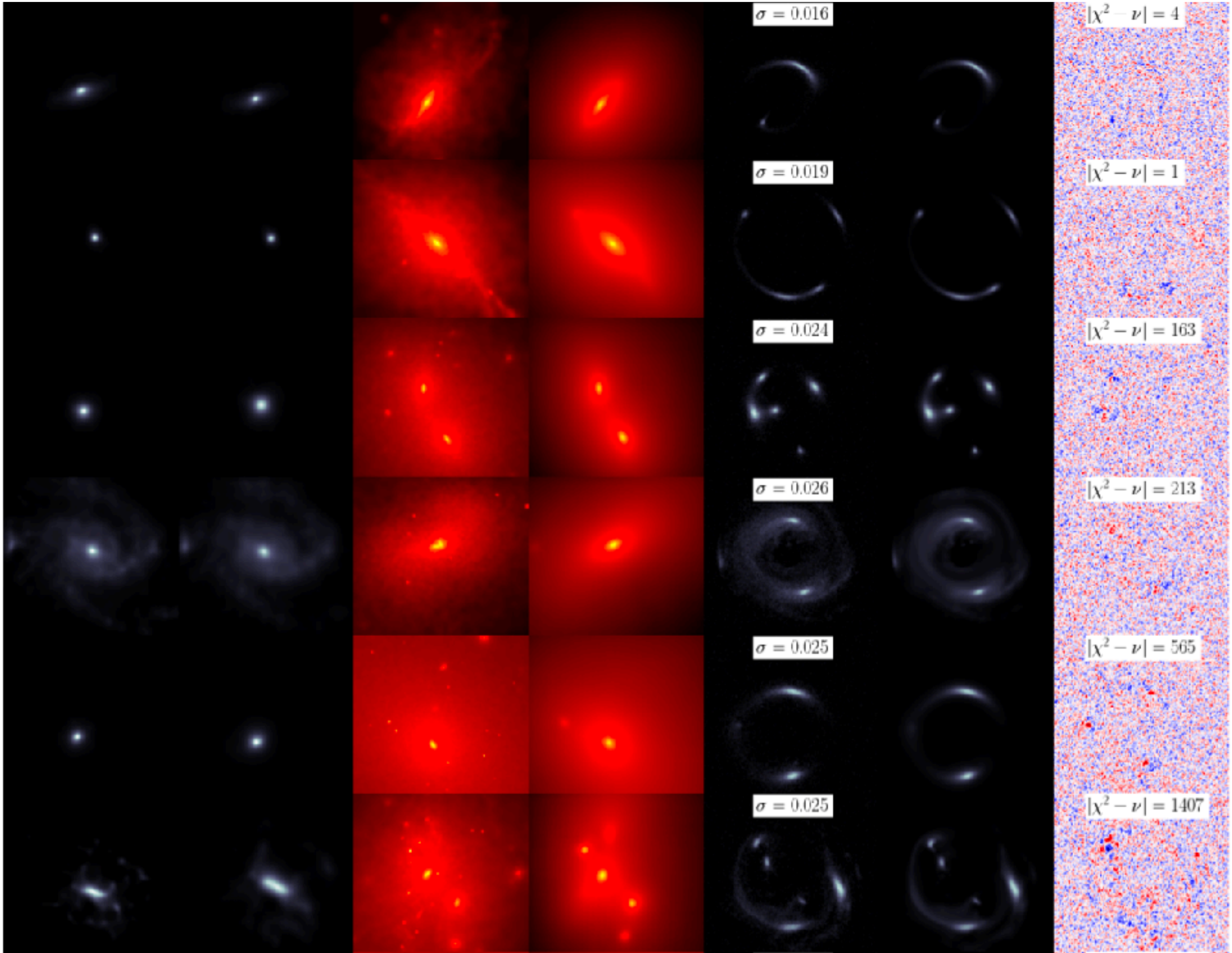




TRAINING ON HYDRODYNAMICAL SIMULATIONS

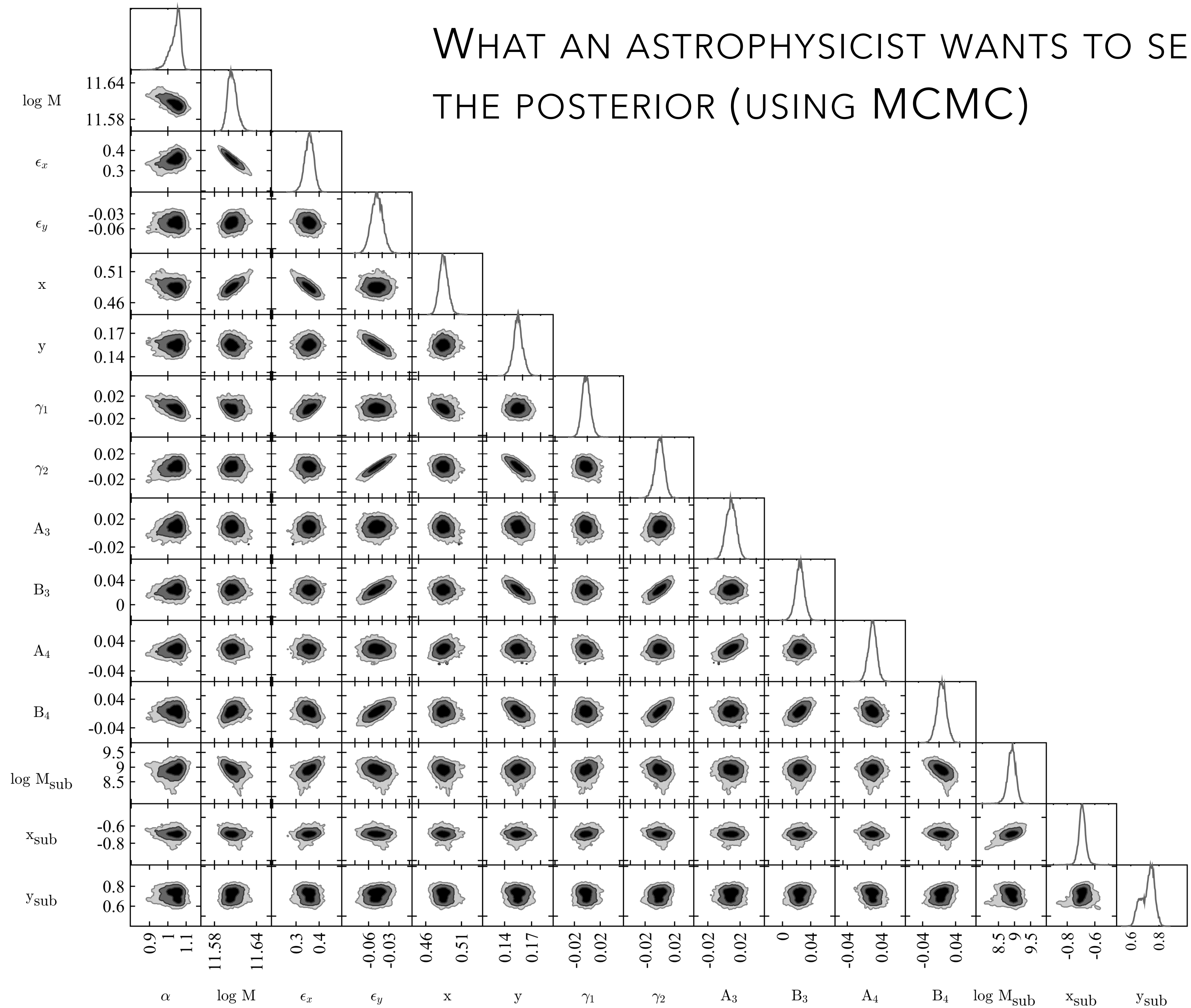


Adam, Perreault-Levasseur, Hezaveh, Welling, ApJ, 2023, arXiv:2301.04168



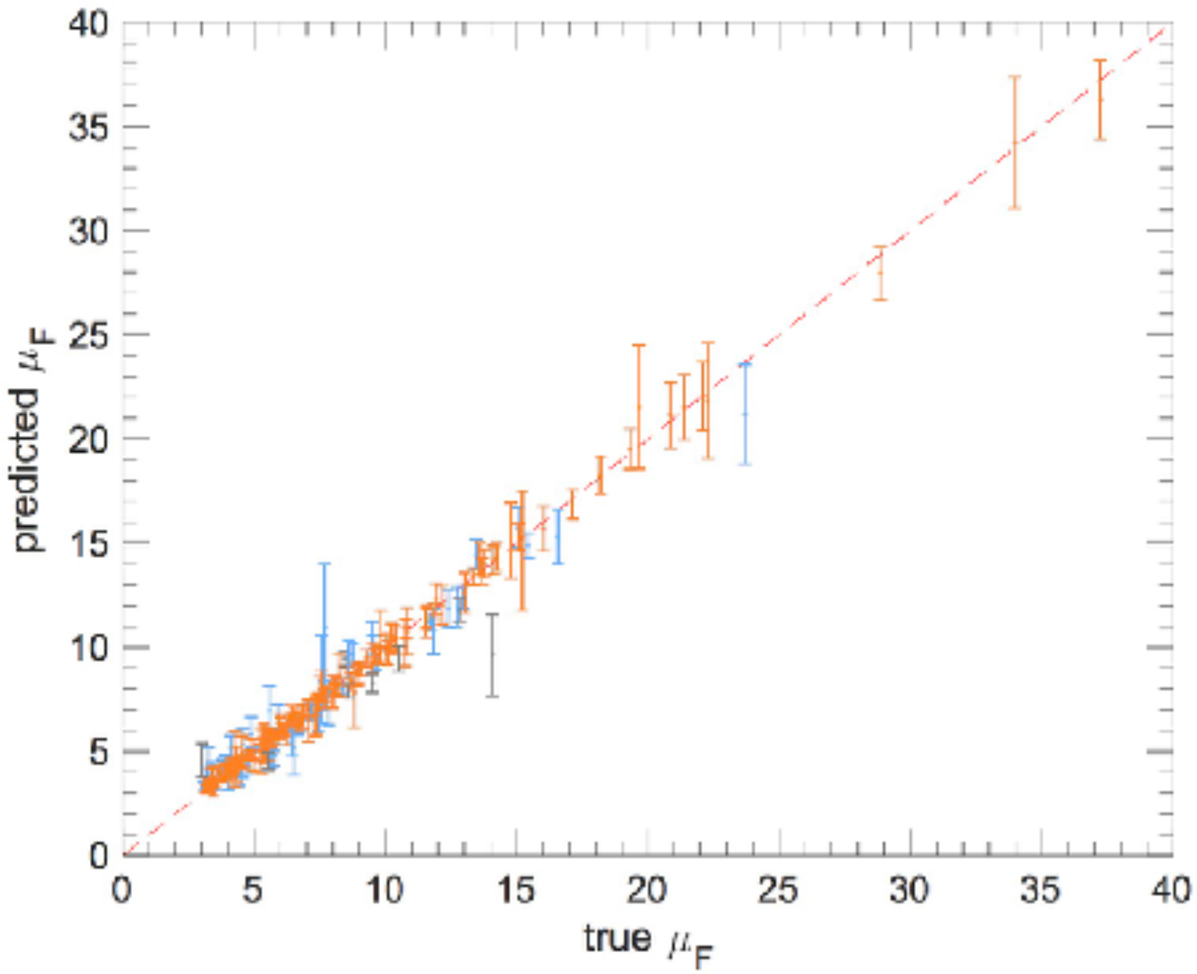
Adam, Perreault-Levasseur, Hezaveh, Welling, ApJ, 2023, arXiv:2301.04168

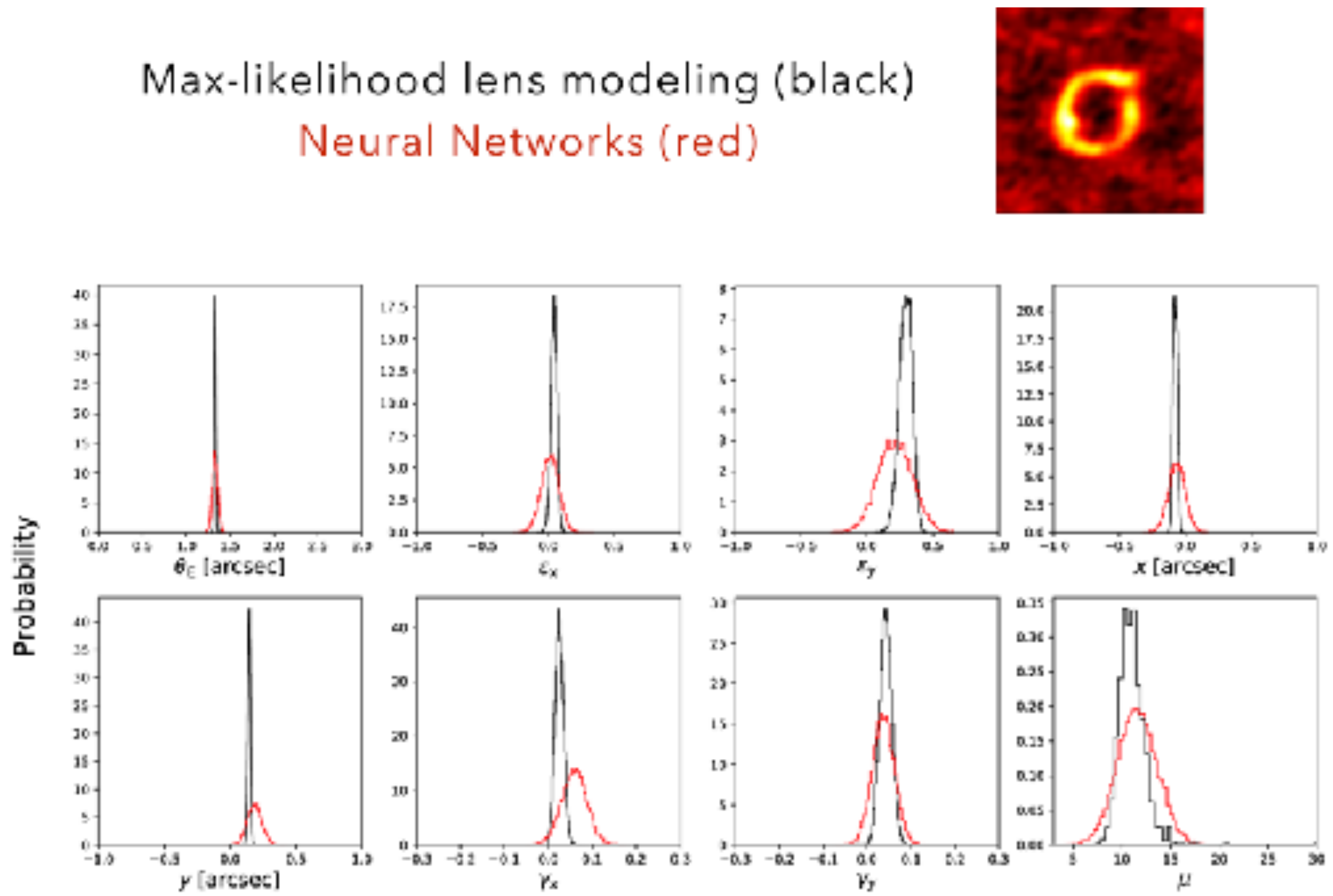
WHAT AN ASTROPHYSICIST WANTS TO SEE: THE POSTERIOR (USING MCMC)



Hezaveh, …,LPL, et al. ApJ 2016

UNCERTAINTY ESTIMATION WITH APPROXIMATE BAYESIAN NEURAL NETWORKS





Perreault-Levasseur et al., 2017

Morningstar et al., 2018

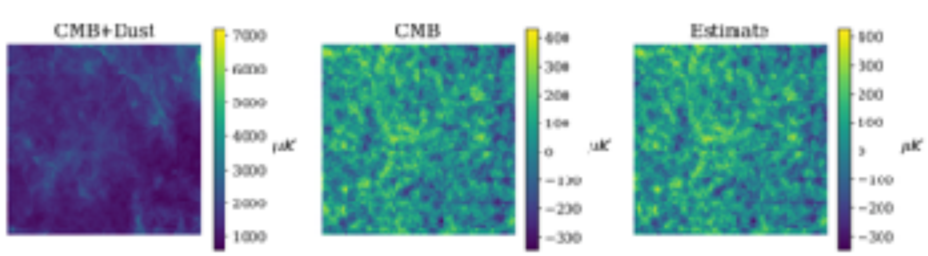
Variational Inference to Approximate Bayesian Neural Networks

Pros:

- Amortized.
- Requires few hundred forward passes at evaluation time (to collect samples). Still very fast.
- Marginalizes implicitly over parameters we do not wish to explicitly model.
- With good coverage probabilities, one can use importance sampling of the output distribution to get an unbiased posterior. (Provided one can actually write this posterior)

Caveats:

- The variational distributions (Bernouilli) are extremely simplistic, therefore even if we attempt to use them to approximate the true weight distributions, that approximation could be bad and yield inaccurate uncertainties.



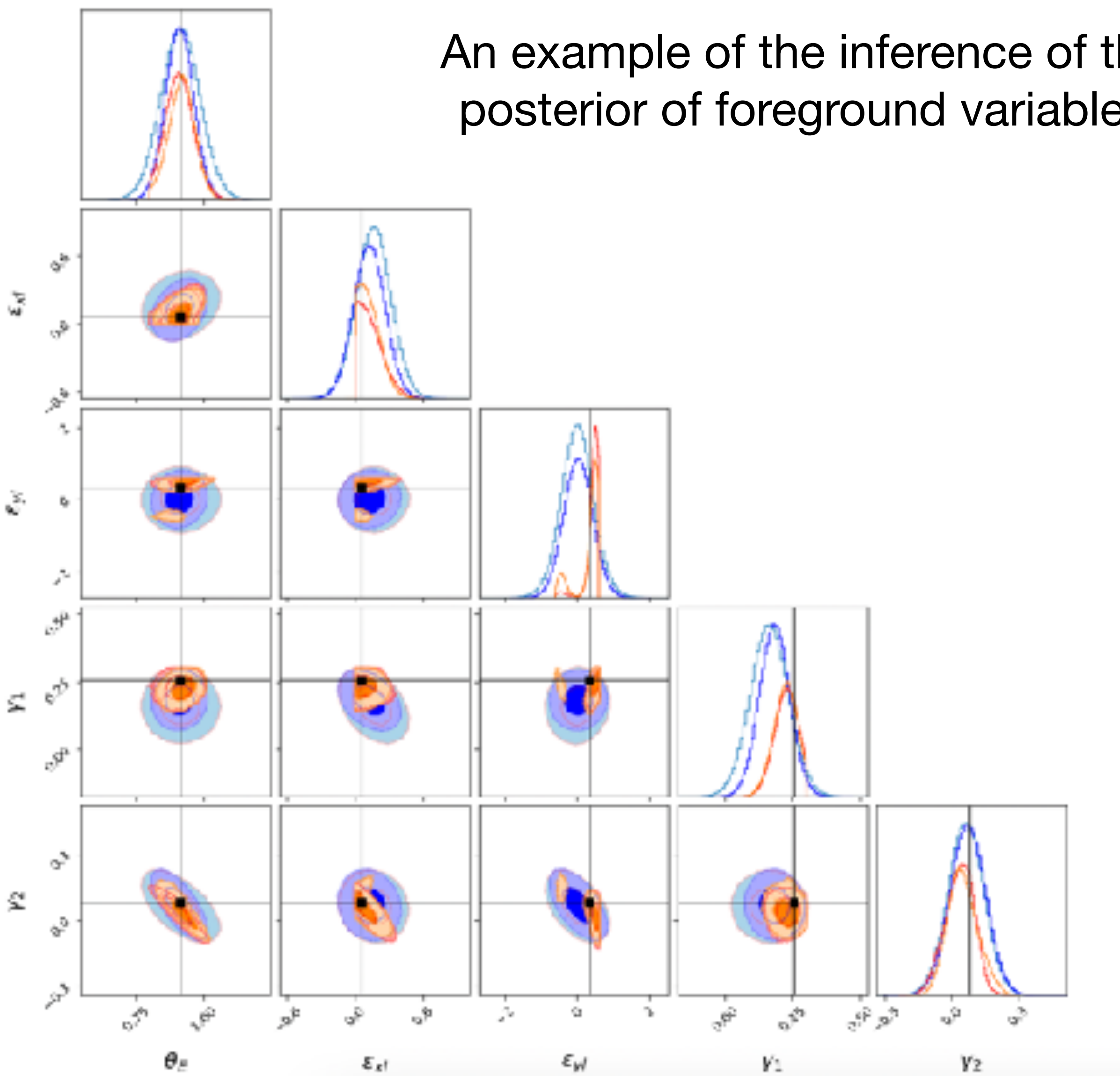
CMB Cleaning

IR Spectrometer De-noising

Same problem remains regardless of the variational distribution used: there is no way of quantifying how well we approximate the true weight distributions

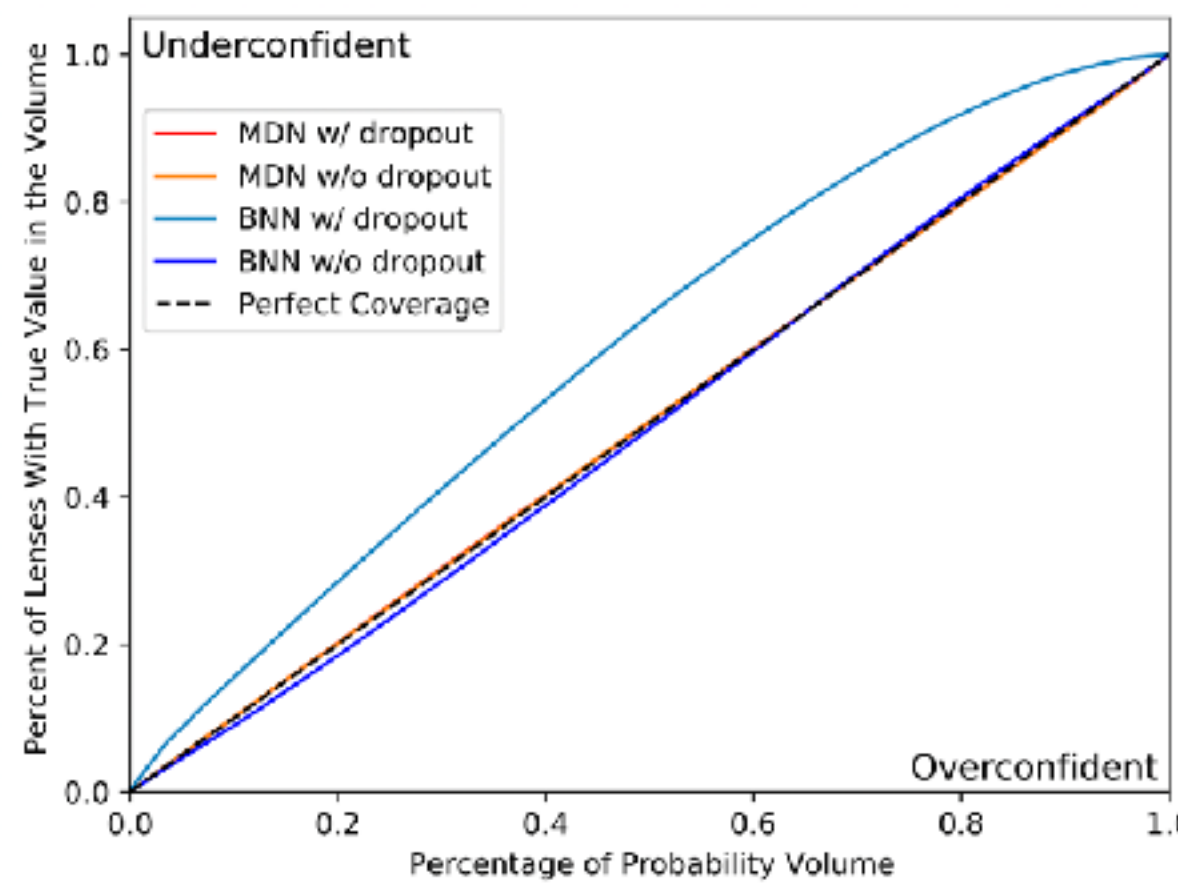
UNCERTAINTY ESTIMATION WITH SIMULATION-BASED INFERENCE METHODS

Ronan Legin



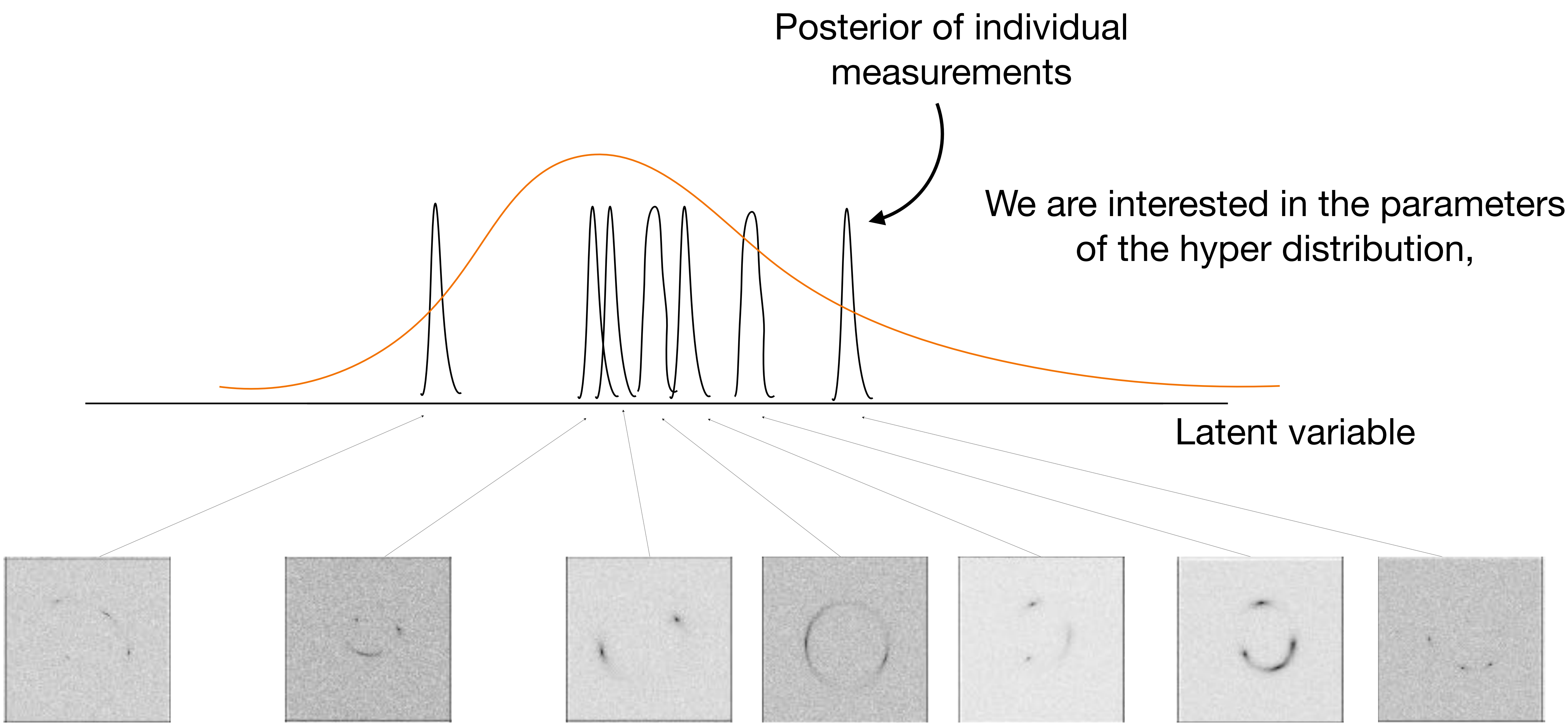
An example of the inference of the posterior of foreground variables

Coverage probabilities



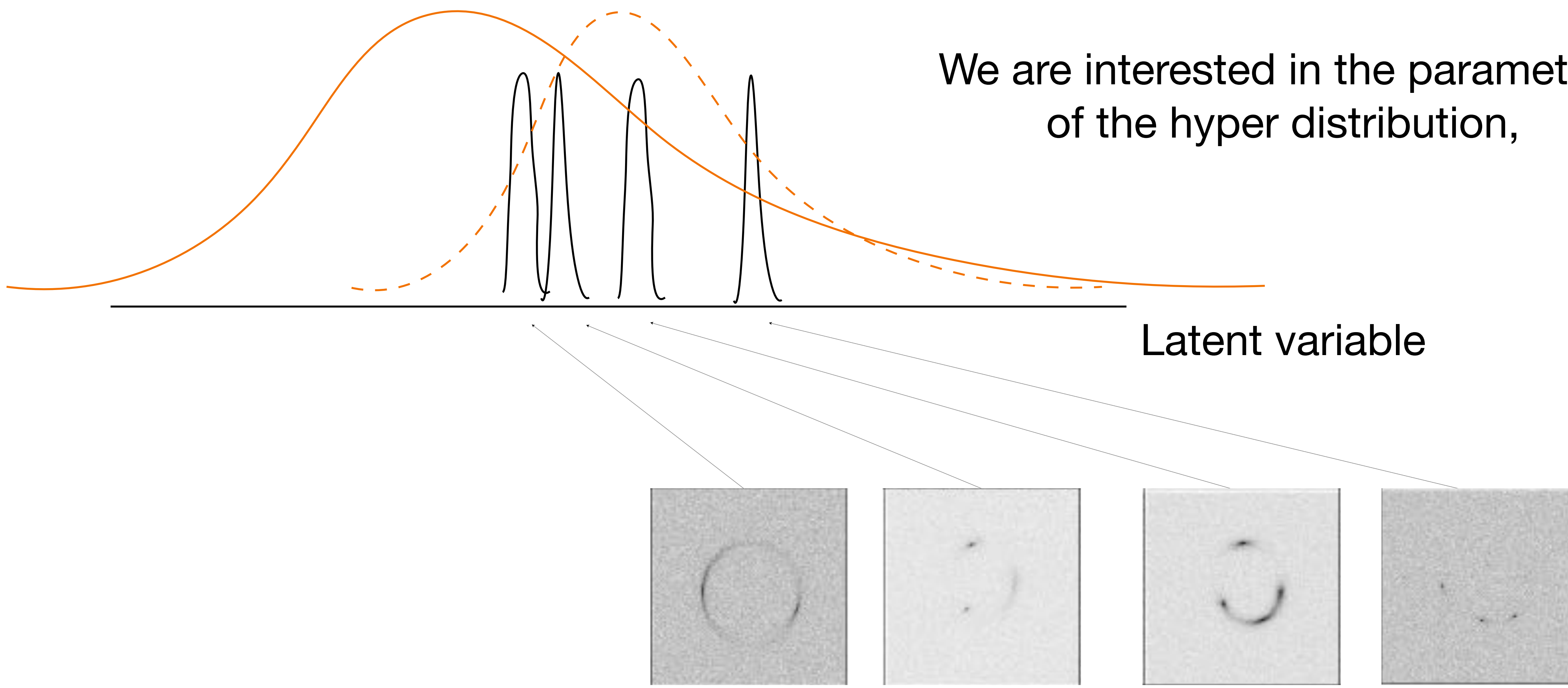
R. Legin, Y. Hezaveh, L. Perreault Levasseur, B. Wandelt, ApJ 2022

Hierarchical Bayesian inference

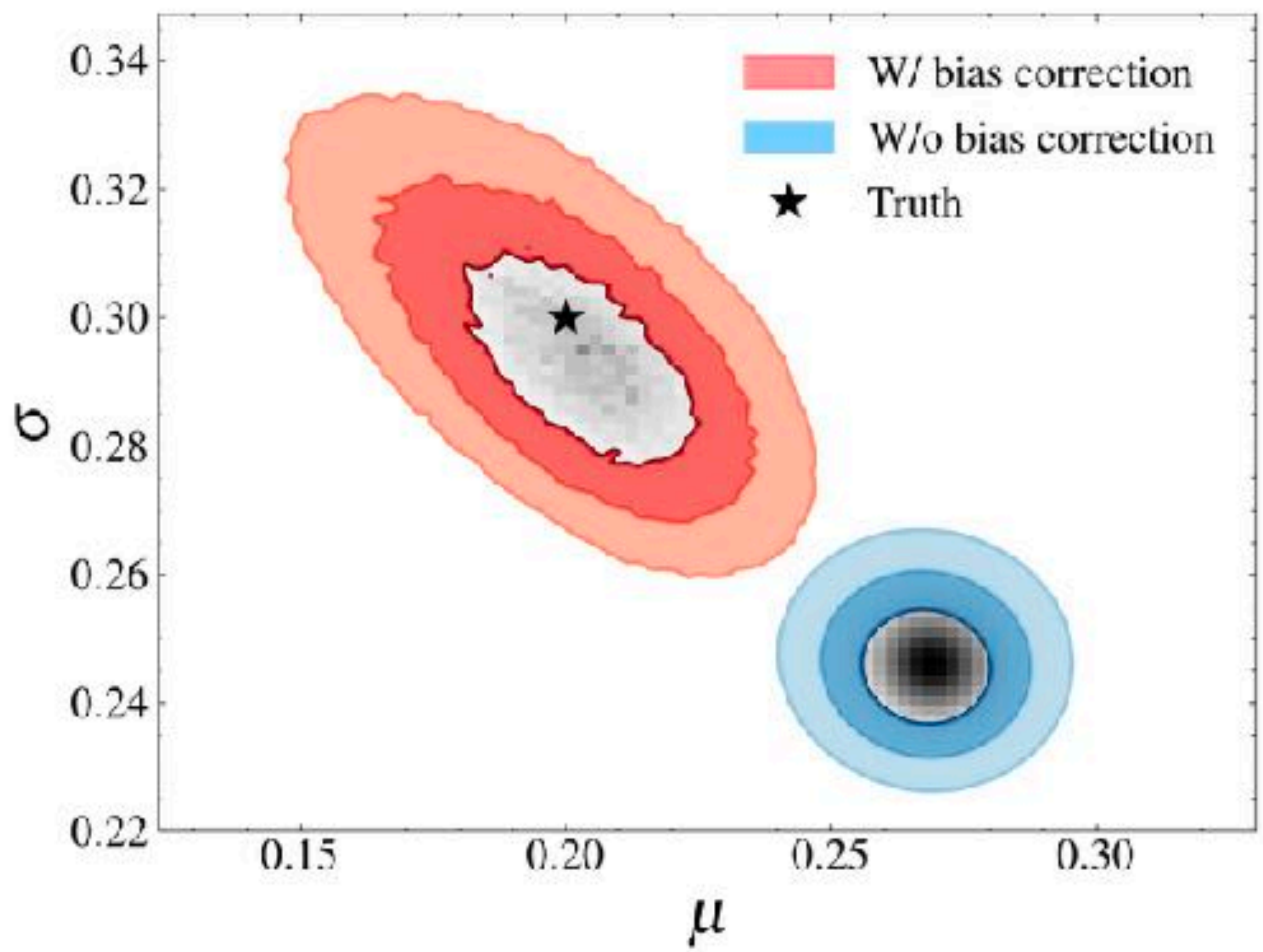


Hierarchical Bayesian inference

We are interested in the parameters of the hyper distribution,



Hierarchical Bayesian inference




Ronan Legin


Connor Stone

Legin, Stone, Hezaveh, Perreault Levasseur,
ICML 2022 - ML4Astro Workshop arXiv:2207.04123

Uncertainty estimation with simulation-based inference methods

<u>Pros:</u>

- Use the **power of ML to find a compress statics**, and even if it is biased, we can get unbiased error estimate, the only drawback would be sub-optimal precision. (Provided the simulation pipeline is accurate!)
- A well-defined statistical framework that can: be relatively **fast**, deal with **complex distributions**, model **joint posteriors**.
- Use a neural density estimator to get the joint distribution p(data, parameter), no need for the epsilon parameter in ABC.
- Can **change the prior** from data point to data point without retraining the ML compressor.
- Once we have the posterior, can **generate samples** that are consistent with data (this is really important for 'interrogating the black box')

<u>Caveats:</u>

- Hard to marginalize implicitly over parameters, we need to **explicitly** model them.
- We don't model the uncertainty of the density estimator itself. (But it's a fairly simple ML model, and except for very pathological problems it's reasonable to expect that we are in interpolation mode).
- Limited to **low-dimensional posteriors** (10s maximum).
- Requires an **accurate simulation pipeline**.

Neural Ratio Estimators

$P(x, \theta)$

Class #1

$\{(x_1, \theta_1), (x_2, \theta_2), \ldots, , (x_N, \theta_N)\}$

$P(x)P(\theta)$

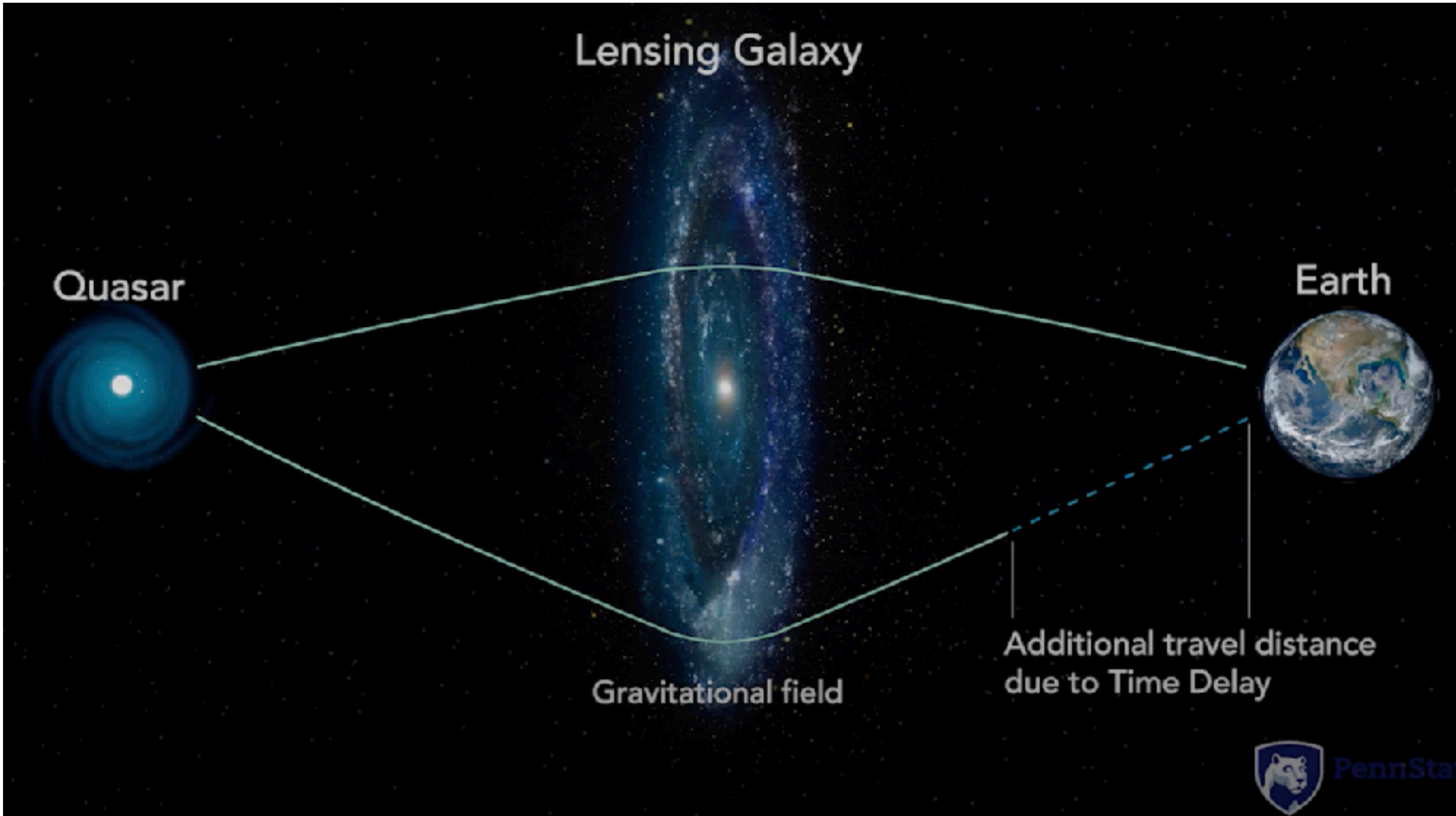
Class #2

$\{(x_1, \theta_1), (x_2, \theta_2), \ldots, , (x_M, \theta_M)\}$

Classify

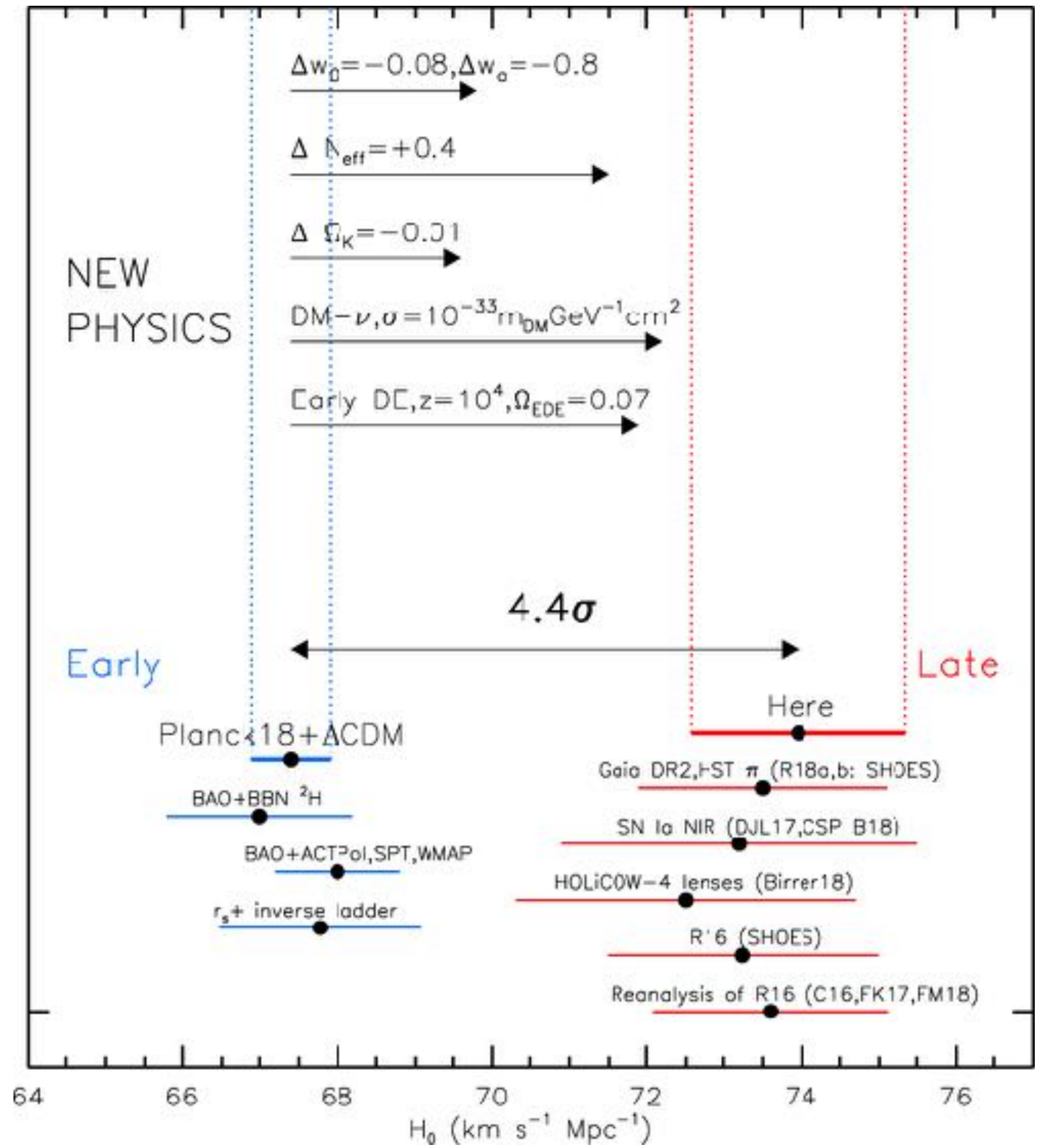
$$r(x, \theta) = \frac{p(x, \theta)}{p(x)p(\theta)} = \frac{p(\theta \mid x)}{p(\theta)}$$

H0 Inference with Time Delay Cosmography



THE HUBBLE CONSTANT

DISCREPANCY BETWEEN MEASUREMENTS



Adam G. Riess *et al* 2019
ApJ **876** 85

H0 Inference with Neural Ratio Estimators

$P(\theta, x)$

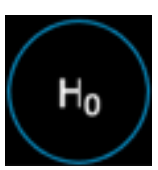
$p($ 

,  $)$

Class #1

$\{(x_1, \theta_1), (x_2, \theta_2), \ldots, , (x_N, \theta_N)\}$

$P(\theta)P(x)$

$p($ 

$)p($  $)$

Class #2

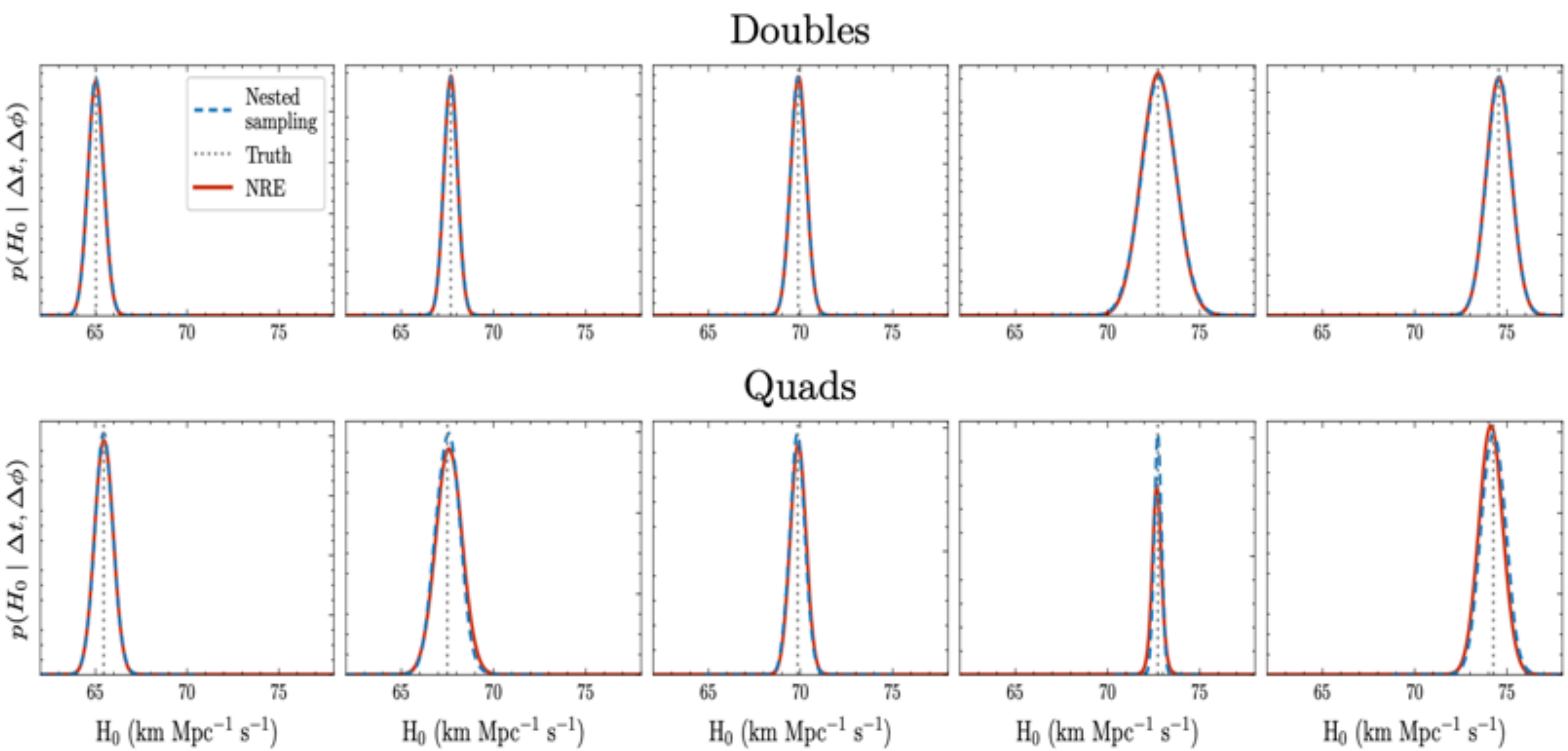
$\{(x_1, \theta_1), (x_2, \theta_2), \ldots, , (x_M, \theta_M)\}$

Classify

$$r(x, \theta) = \frac{p(x, \theta)}{p(x)p(\theta)} = \frac{p(\theta \,|\, x)}{p(\theta)}$$

H0 Inference with Neural Ratio Estimators

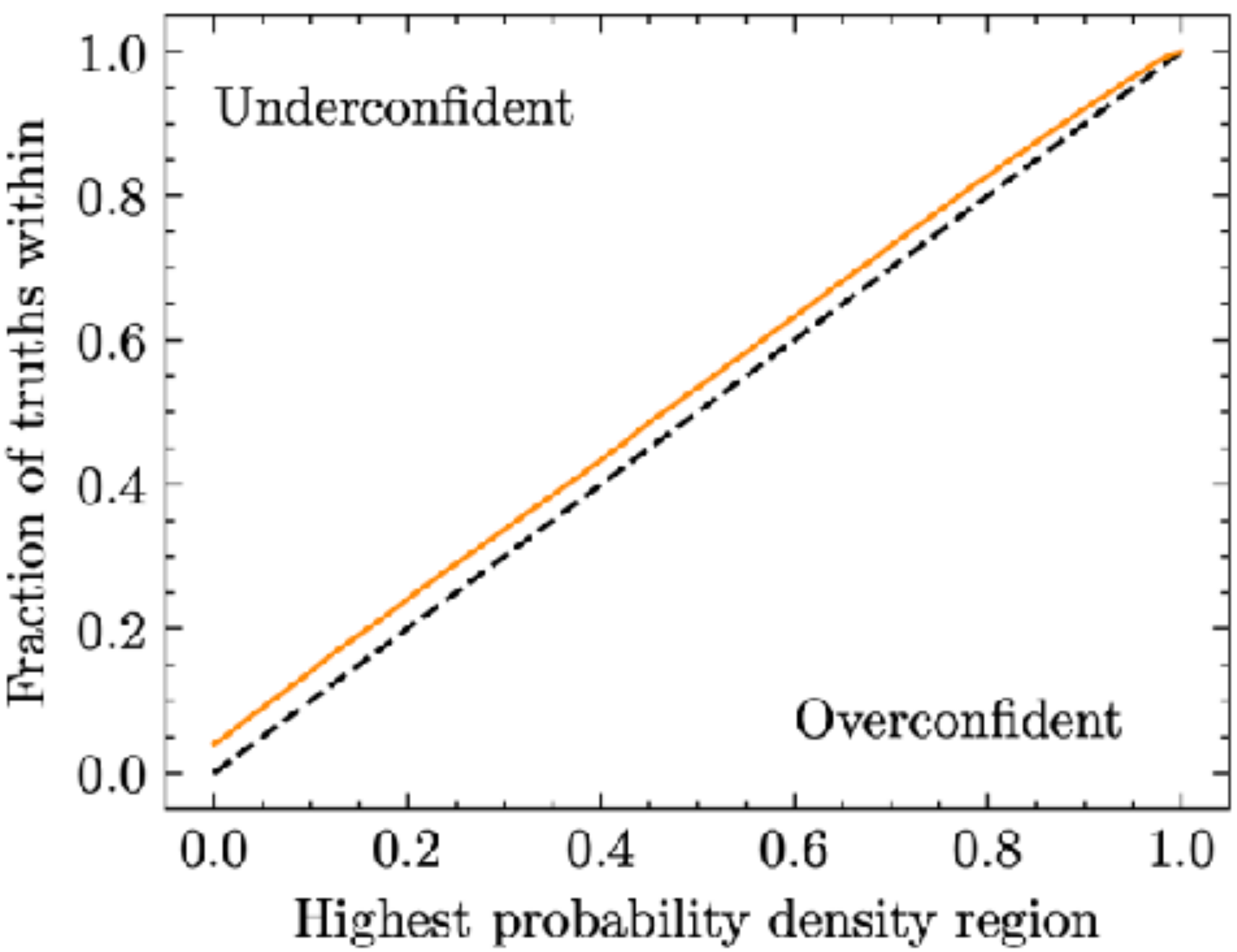
Ève Campeau-Poirier

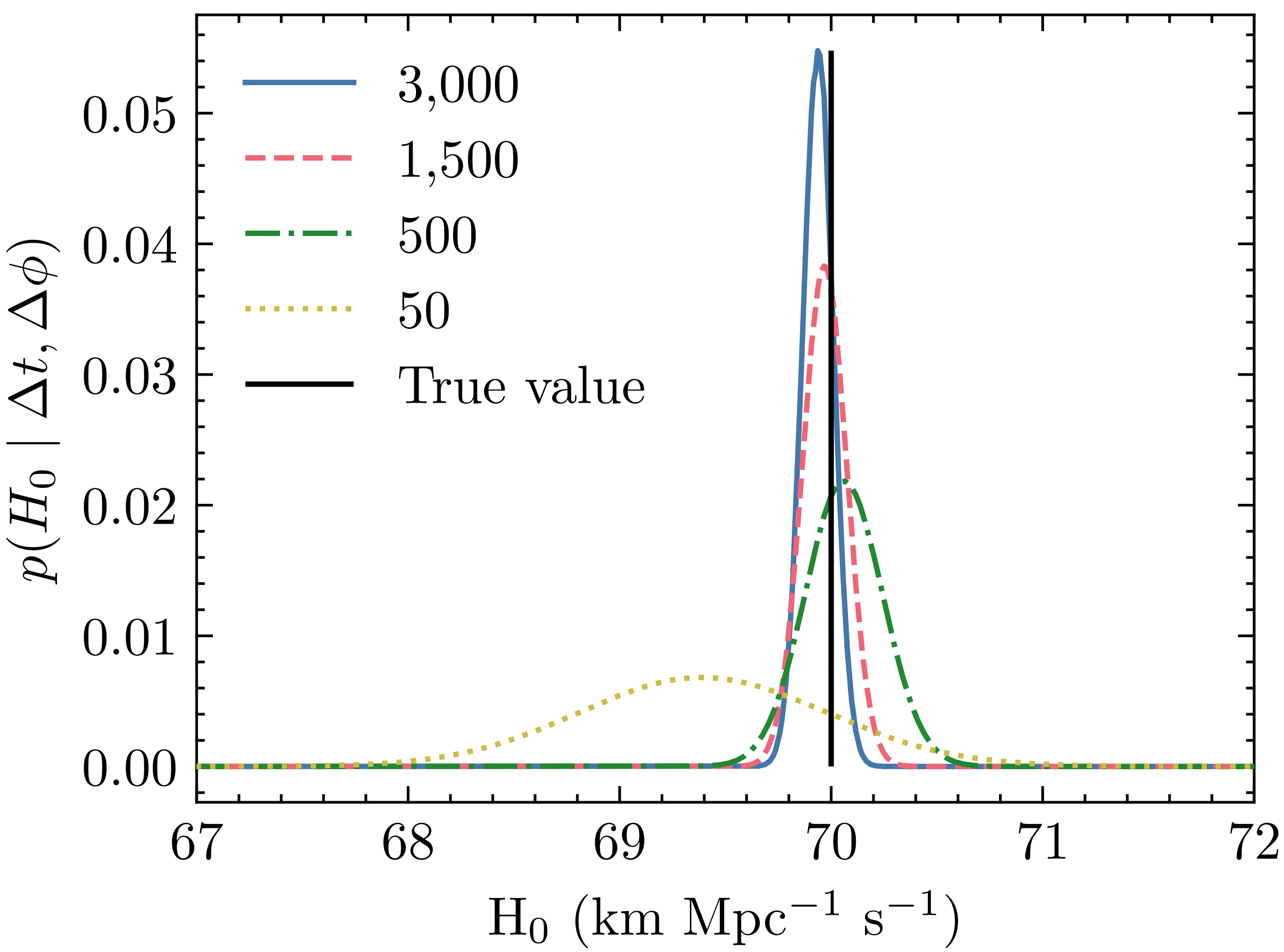


Campeau-Poirier et al. ICML 2023 ML4Astro Workshop, arXiv:2309.16063

H0 Inference with Neural Ratio Estimators

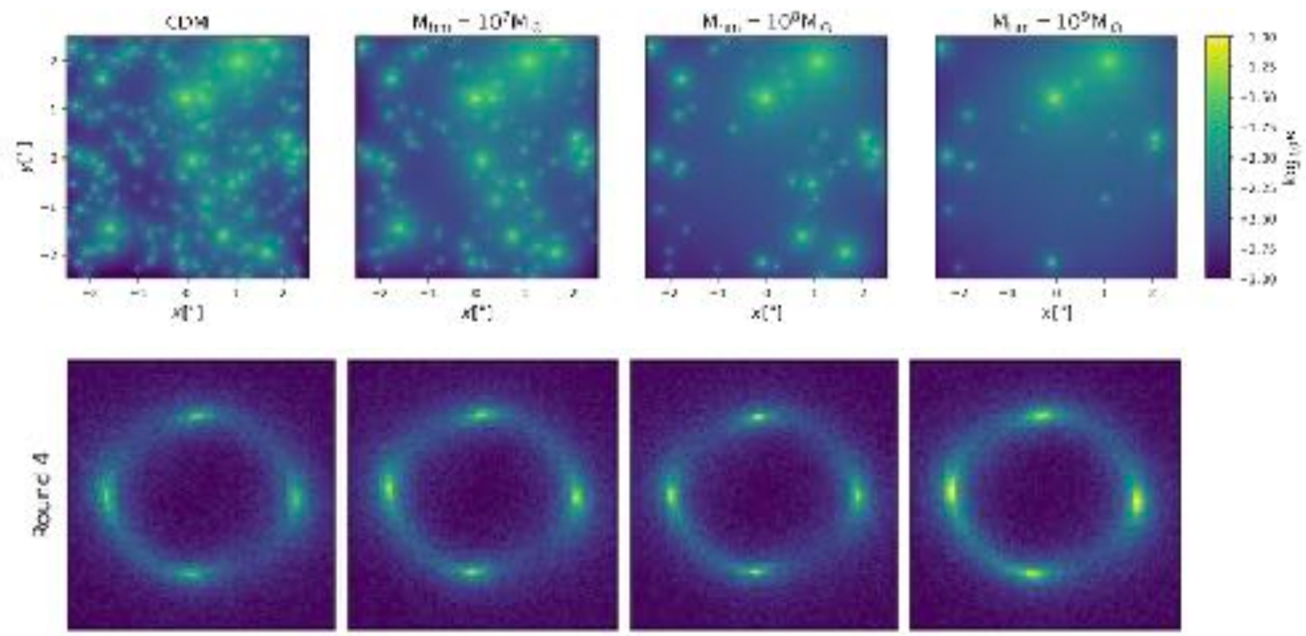
Ève Campeau-Poirier

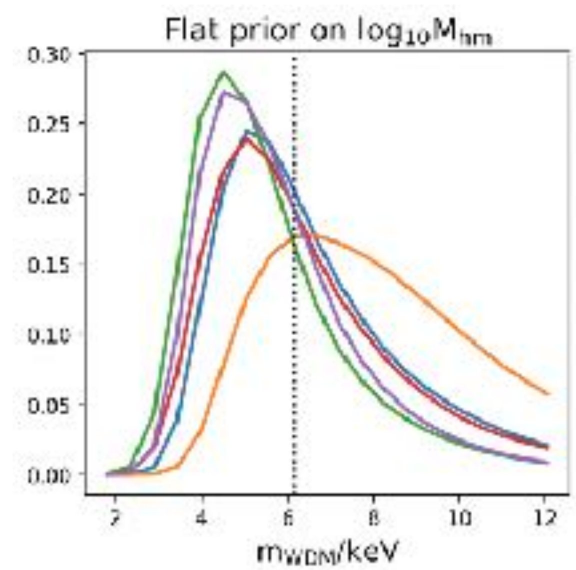




Campeau-Poirier et al. ICML 2023 ML4Astro Workshop, arXiv:2309.16063

Estimating the dark matter particle temperature with Neural Ratio Estimators





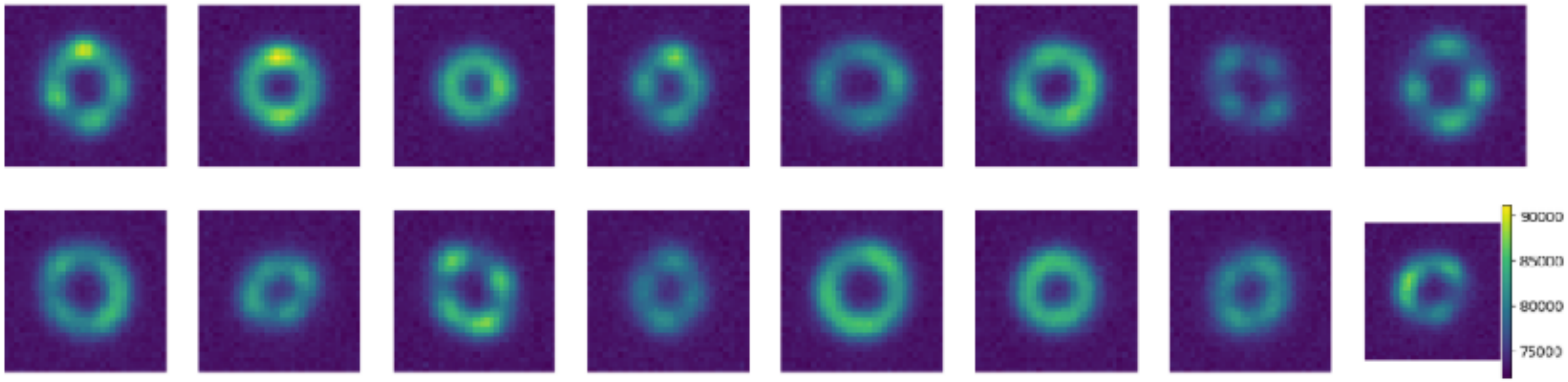
Adam Coogan

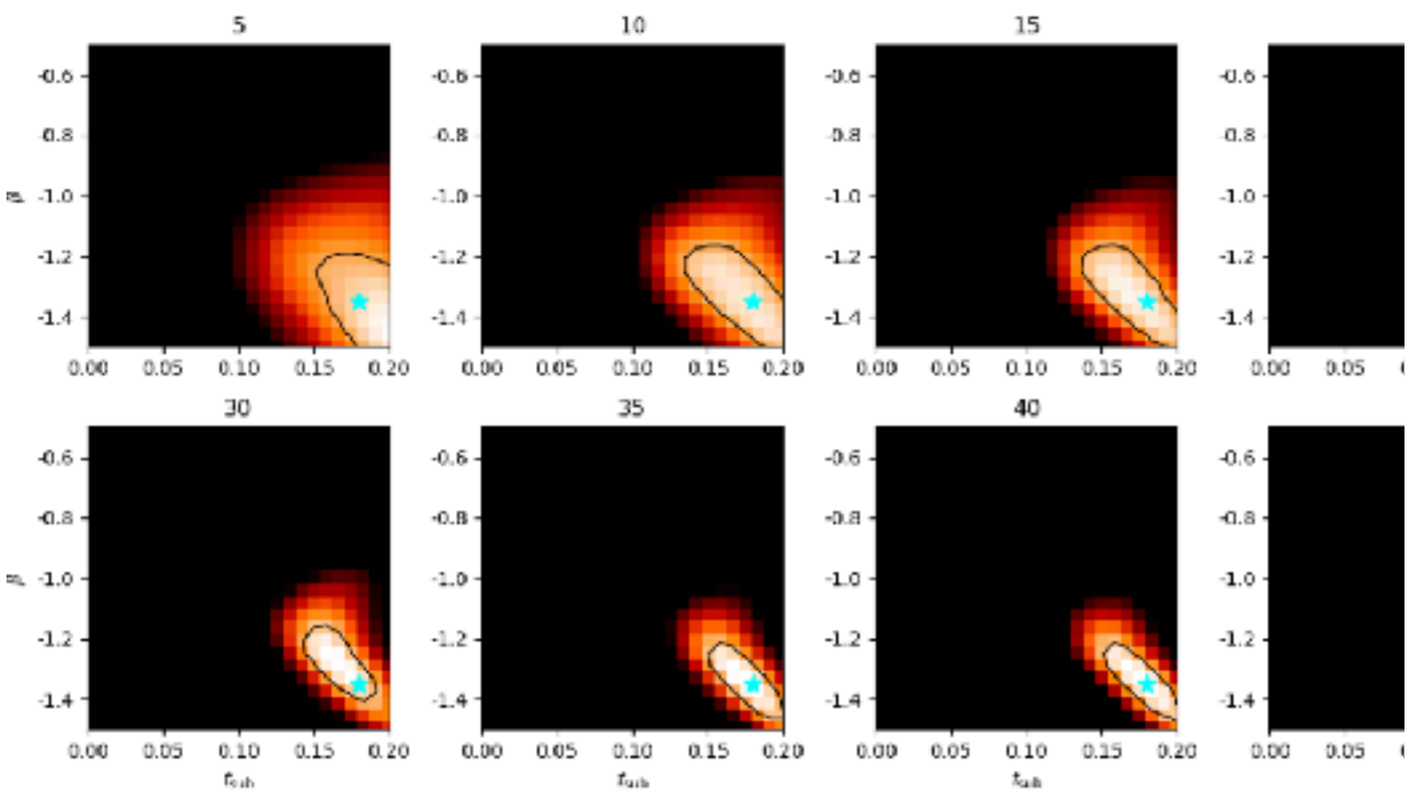
Anau Montel, Coogan et al. 2022, arXiv:2205.09126

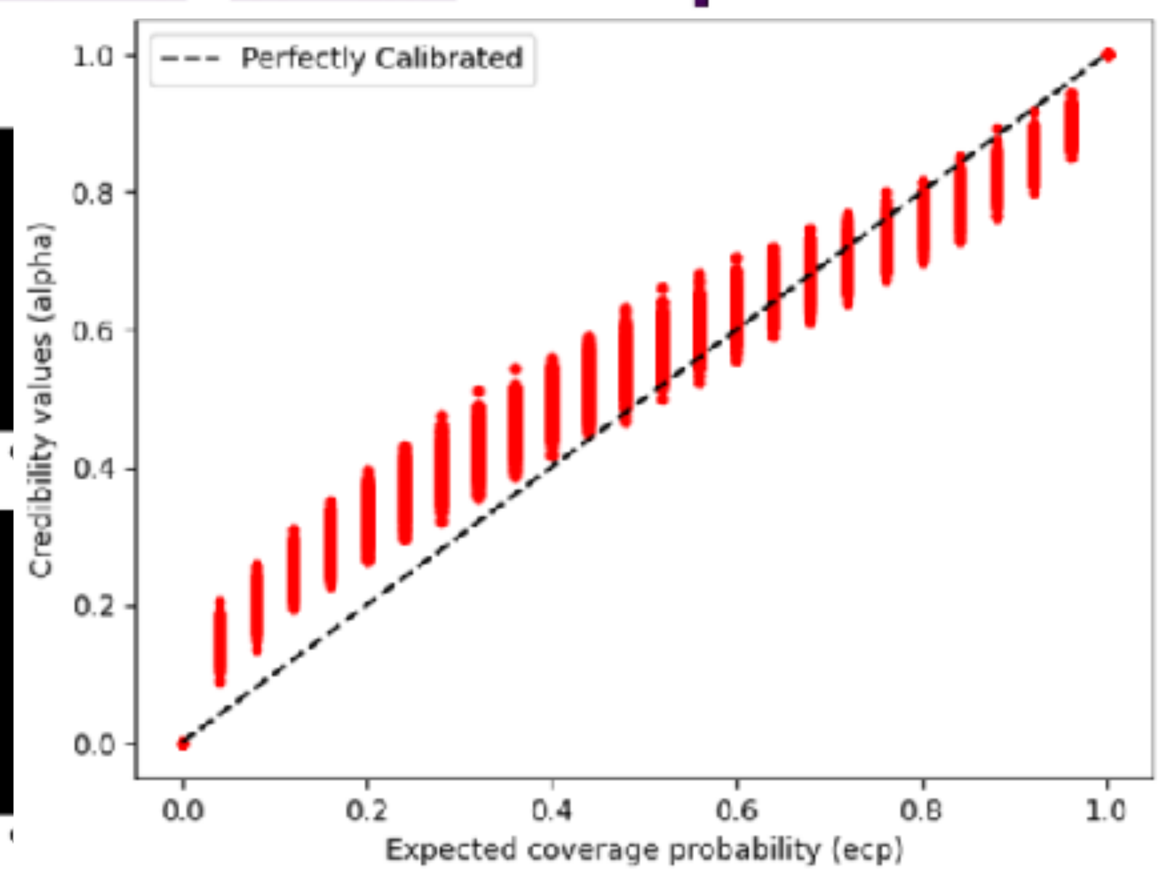
Coogan et al. , NeurIPS 2020 ML4PS Workshop

Estimating the sensitivity of LSST to the break and slope of the dark matter mass function

Andreas Filipp





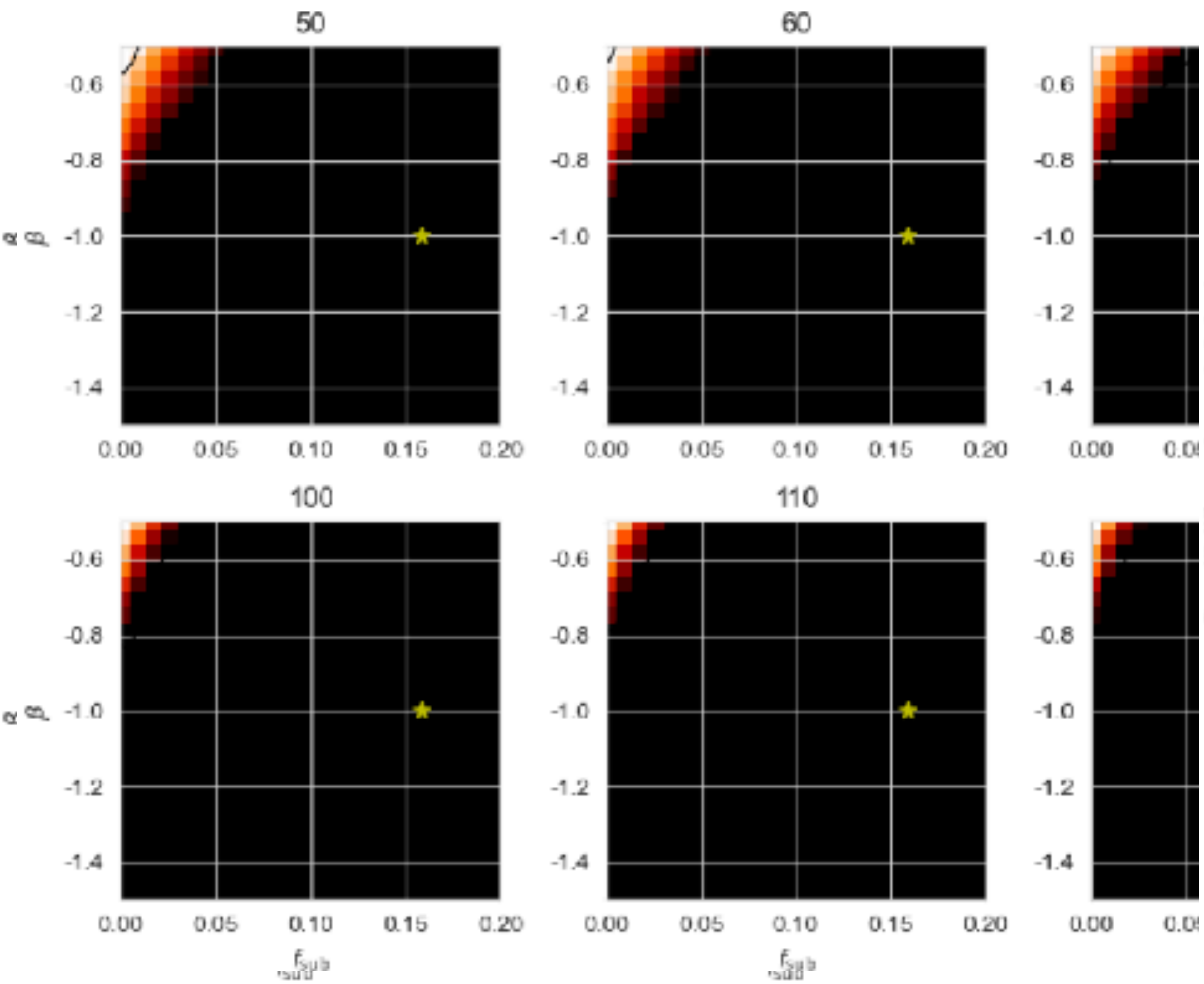


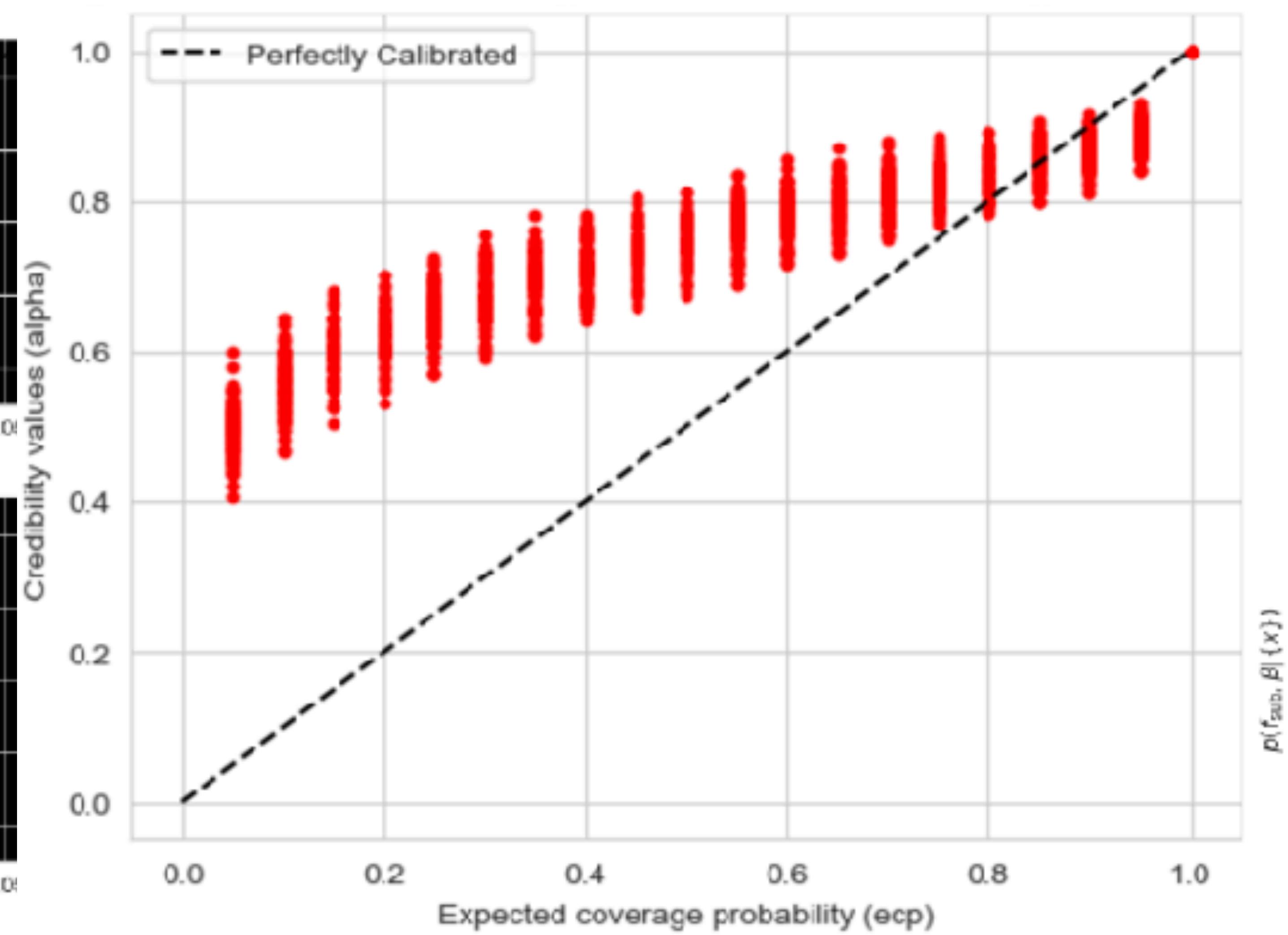
Testing for robustness to model misspecification

Andreas Filipp

Training on SIS (spherically symmetric lenses)…

And testing on galaxies with slight ellipticities, q=0.98





Ratio estimation methods

Pros:

- Can marginalize implicitly over large number of nuisance parameters

Caveats:

- Because we have marginalized, we've lost the capability to generate samples consistent with the observations.
- From experiments, it seems easy to find examples where the NN is very brittle and sensitive to model misspecification.
- So far: no real way of quantifying the uncertainty of the ratio estimator itself. All the guarantees are in terms of convergence to a specific ratio in the limit of perfect training. Is this always realistic?

TACKLING AN UNSOLVED PROBLEM: HIGH DIMENSIONAL INFERENCE

How do we infer the posteriors of high-dimensional parameters (e.g., an image or spectra)?

Obstacles:

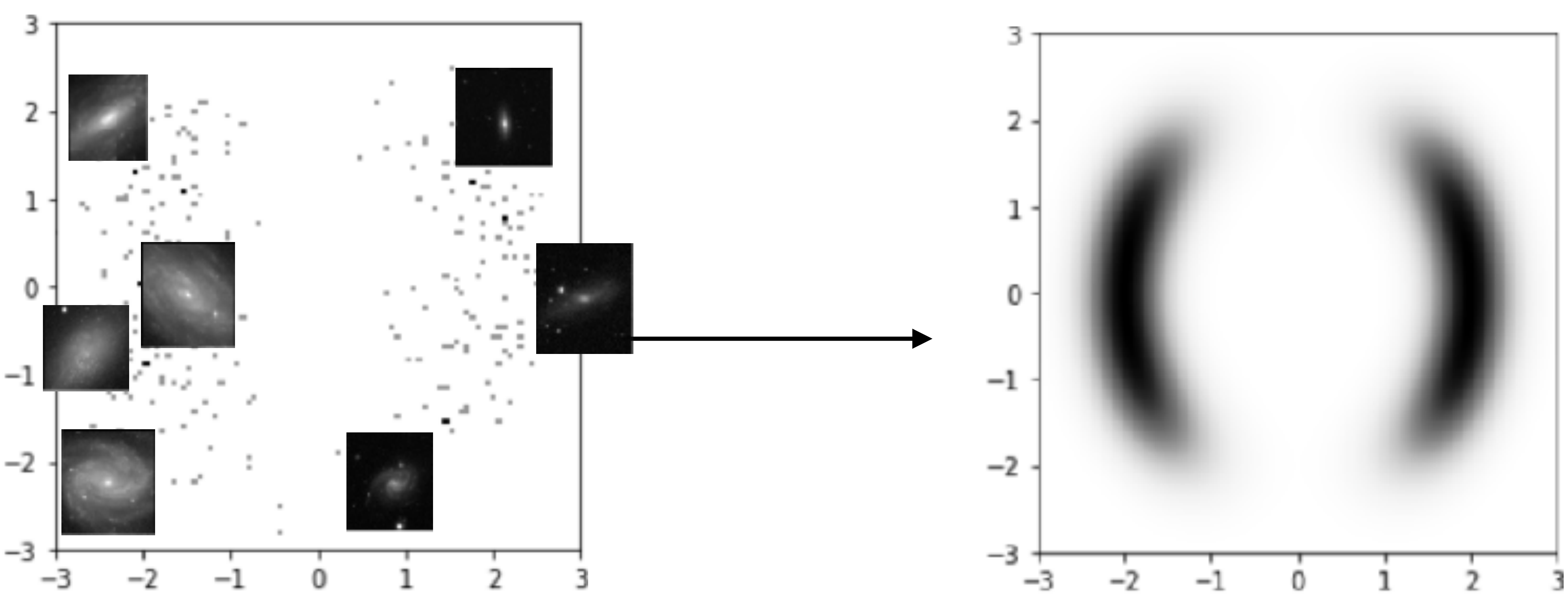
1) How do we encode complex priors

2) How we sample such high-dimensional posteriors (even if we could compute them)

LEARNING THE PRIOR EXPLICITLY

Can we learn our high-dimensional prior explicitly from data?
i.e. can we learn a generative model that will produce samples from that distribution?

How can we do this from samples (e.g. data)? Modeling the density?



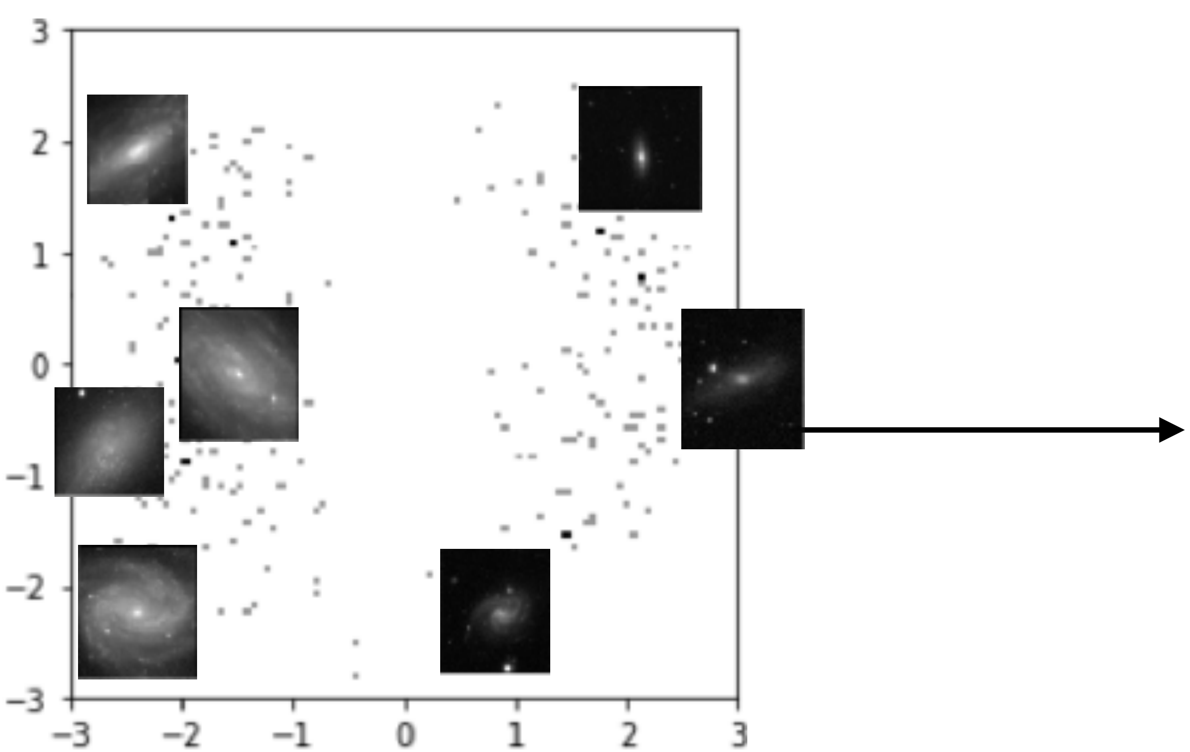
Training data
$\{\mathbf{x}_1, \ldots, \mathbf{x}_n,\} \overset{iid}{\sim} \pi_{data}(\mathbf{x})$

Density function
$p_\theta(\mathbf{x}) \approx \pi_{data}(\mathbf{x})$

Score Modeling

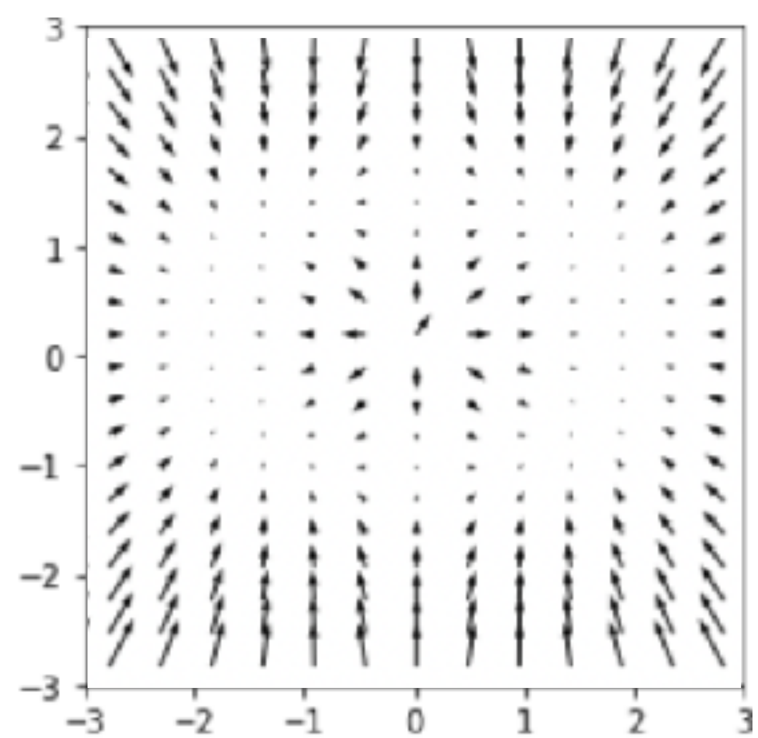
Turns out that if I want to sample a distribution, the only thing I need to learn is its ***score***, which does not include the normalization constant and only uses local information

$$\mathbf{s}(\mathbf{x}) = \nabla_{\mathbf{x}} \log(\pi(\mathbf{x}))$$



Training data

$\{\mathbf{x}_1, \ldots, \mathbf{x}_n,\} \overset{iid}{\sim} \pi_{data}(\mathbf{x})$



Score function

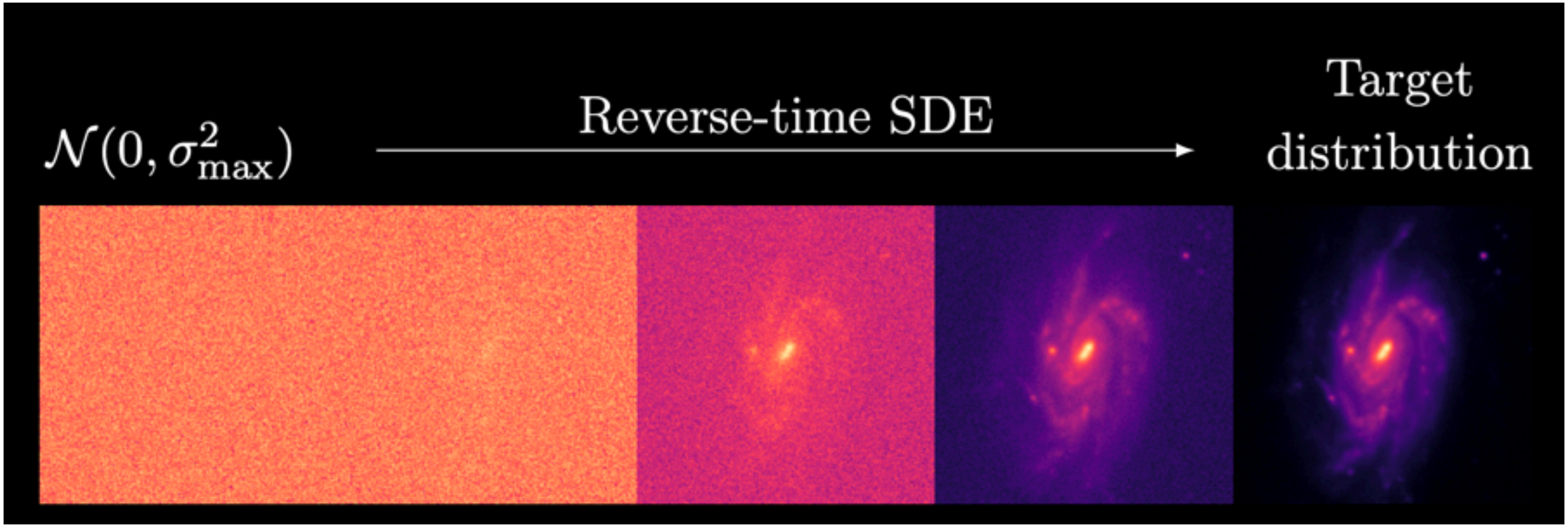
$\mathbf{s}_\theta(\mathbf{x}) \approx \nabla \log(\pi(\mathbf{x}))$

Score-Based Modeling

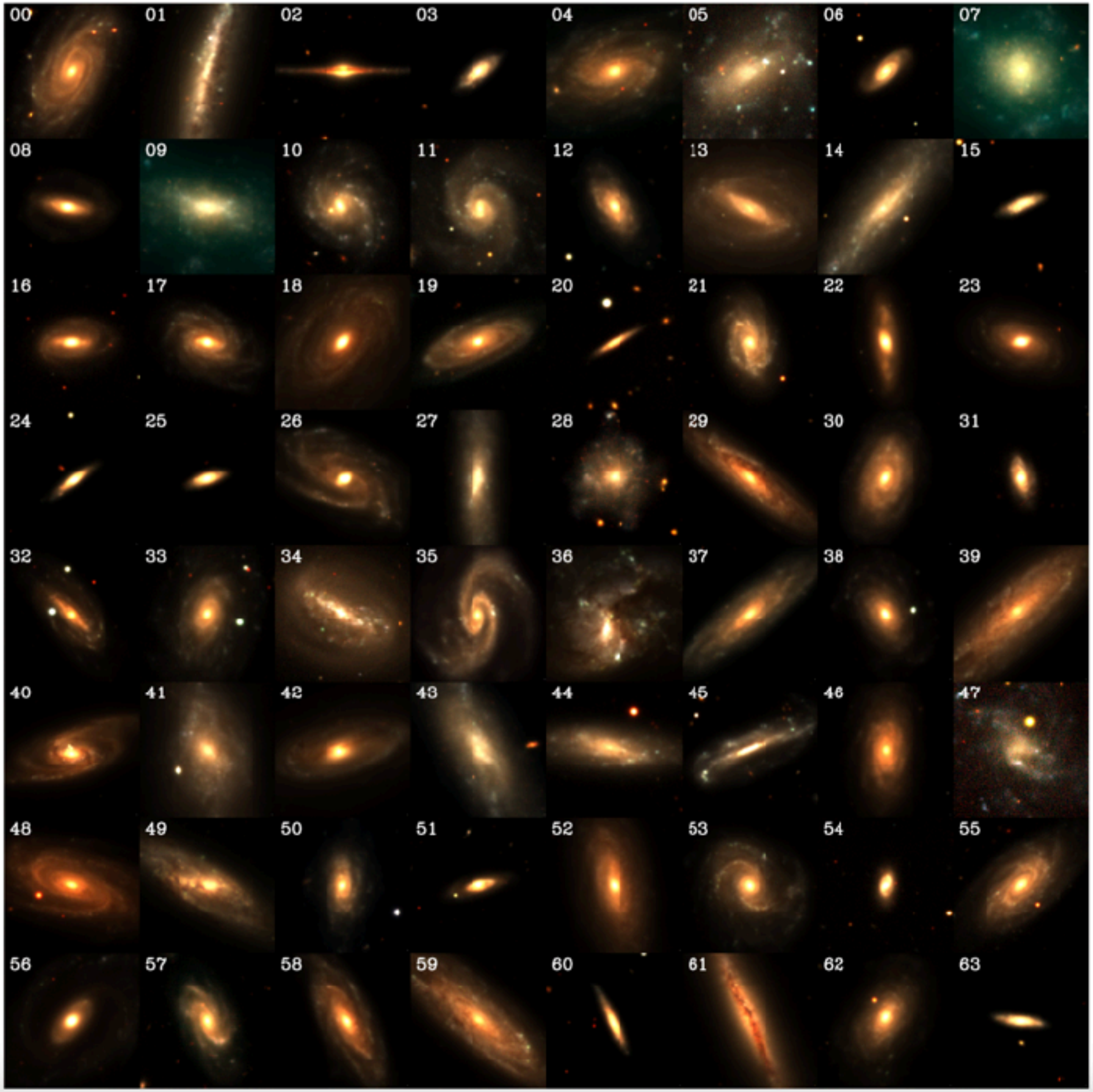

Alexandre Adam

We model the score of the prior

$$s_\theta(x) \equiv \nabla_x \log p_\theta(x)$$



Adam et al. NeurIPS 2022 ML4PS workshop, arXiv:2211.03812

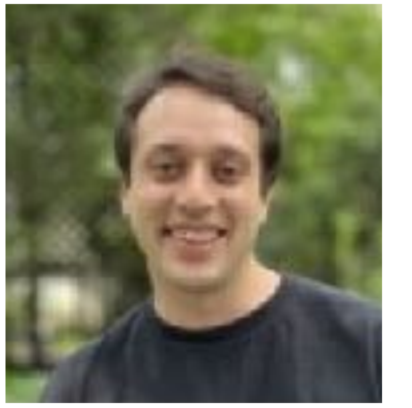


Connor Stone

http://www.mjjsmith.com/thisisnotagalaxy/

Smith et al. arXiv:2111.01713

SCORE-BASED MODELING



Alexandre Adam

Now if we want to sample from the posterior, its score is all we need:

$$\nabla_x \log p(x \mid y)$$

To a good approximation, we can calculate the likelihood score analytically if we assume it's Gaussian and we know the lensing matrix.

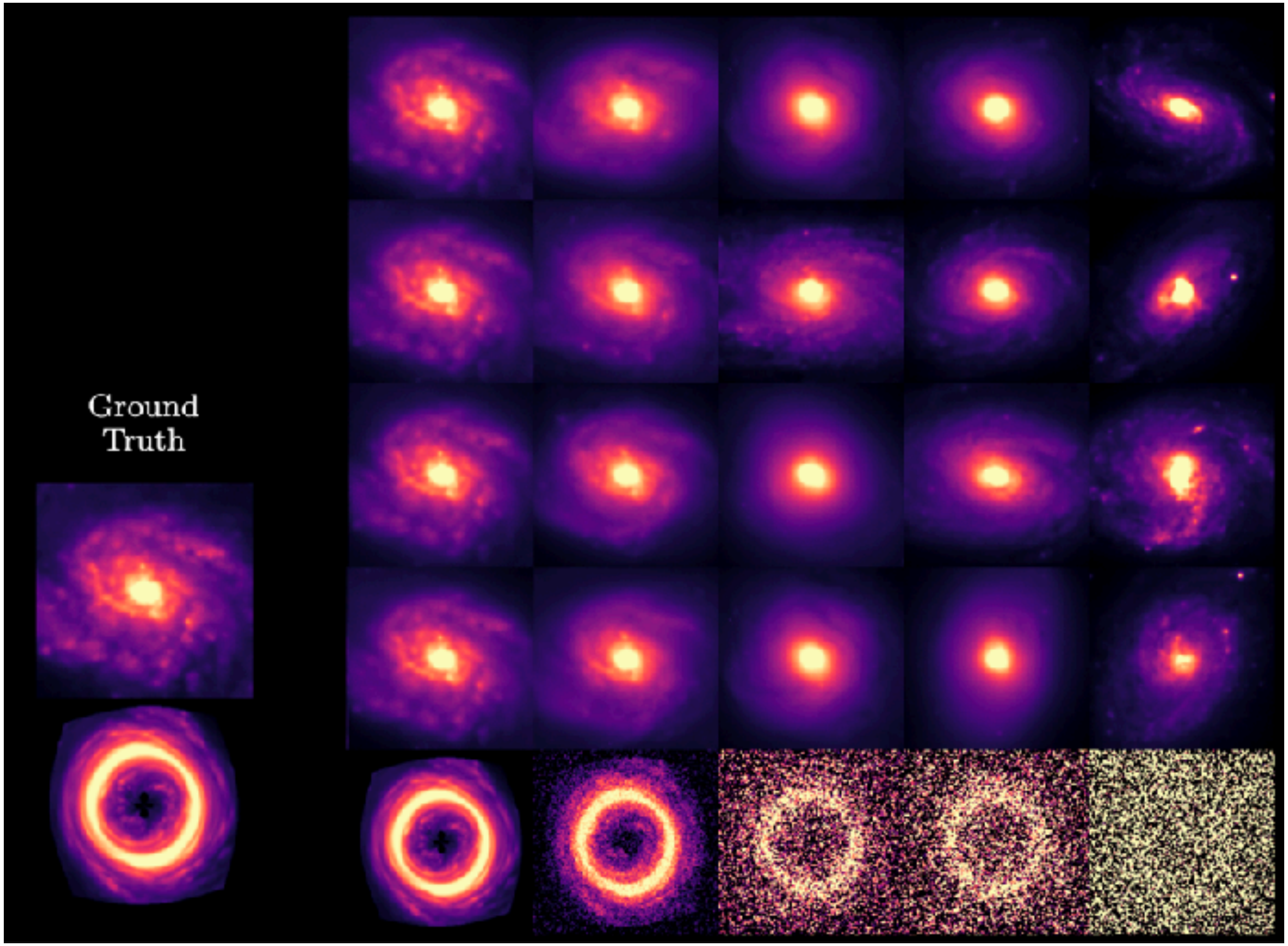
This is the prior score we learnt from the training data

Score-based Modeling

Alexandre Adam

Now if we want to sample from the posterior, its score is all we need:

$$\nabla_x \log p(x \mid y) = \nabla_x \log p(y \mid x) + \nabla_x \log p_\theta(x)$$

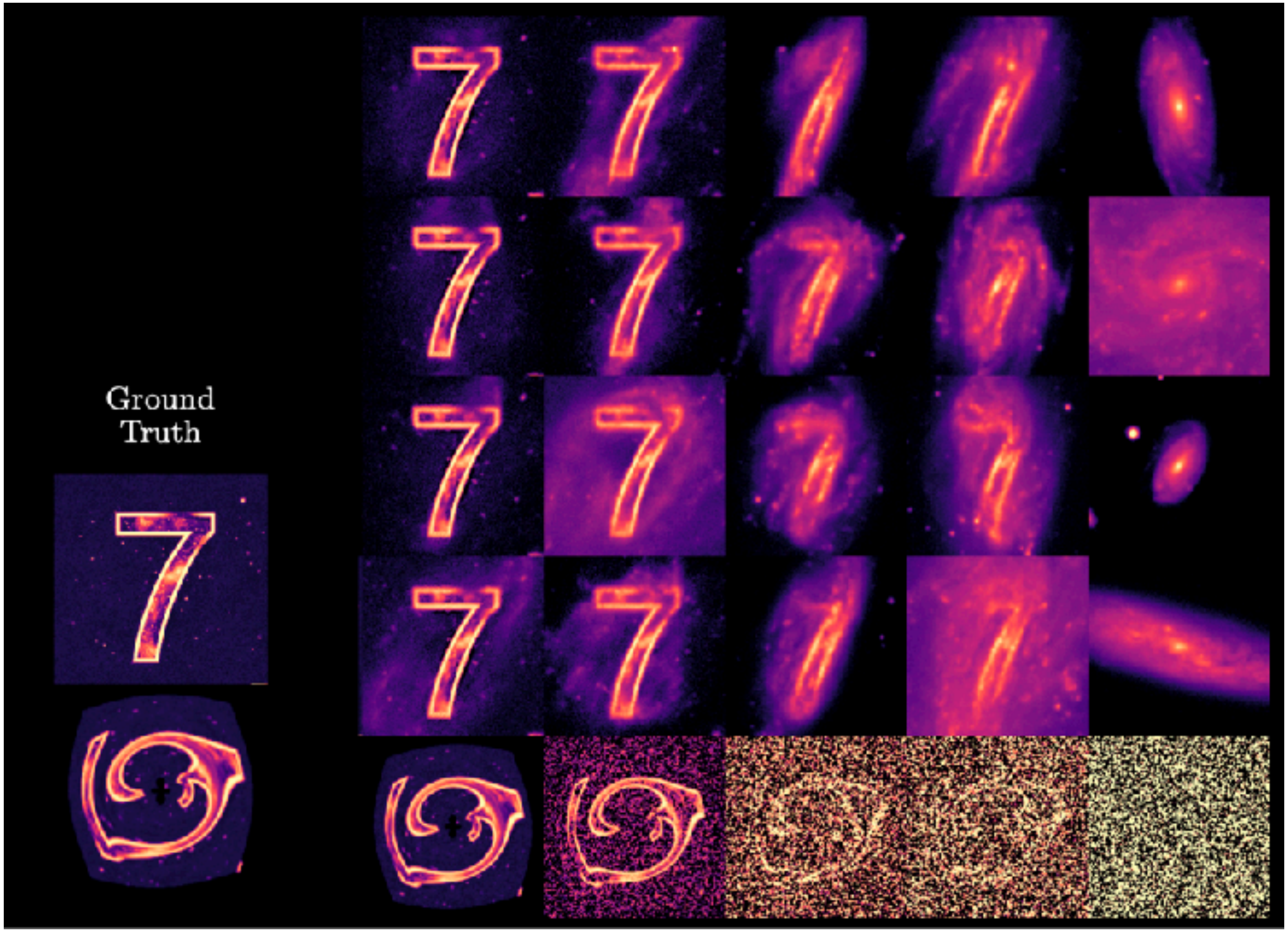


Out of Distribution Tests


Alexandre Adam

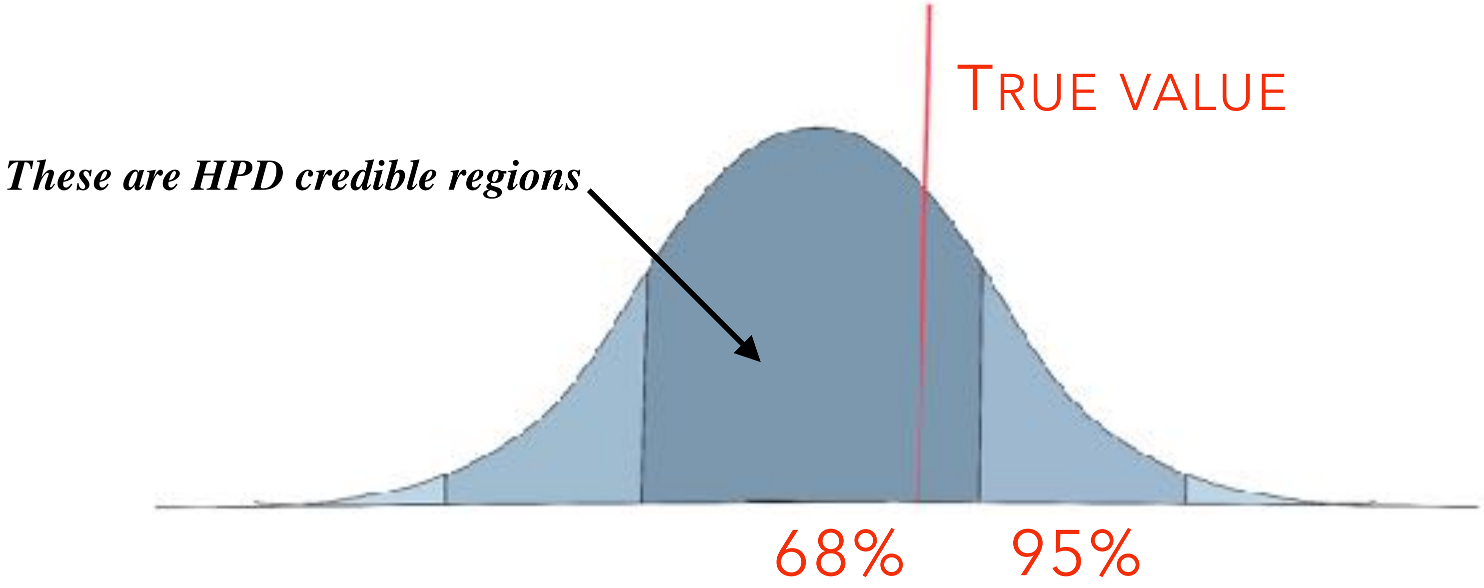
Now if we want to sample from the posterior, its score is all we need:

$$\nabla_x \log p(x \mid y) = \nabla_x \log p(y \mid x) + \nabla_x \log p_\theta(x)$$



ARE THESE UNCERTAINTIES ACCURATE?

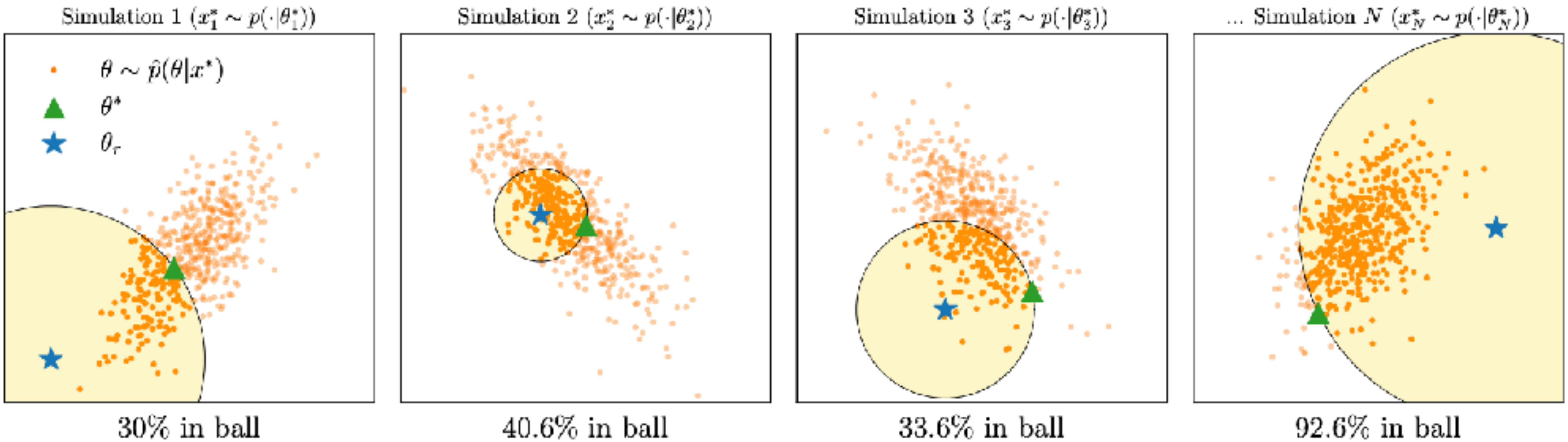
The expected coverage probability of a credible region is the proportion of the time that the region contains the true value of interest.



For an accurate posterior estimator, the expected coverage probability is equal to the probability mass of the credible region.

Coverage Test for Accuracy

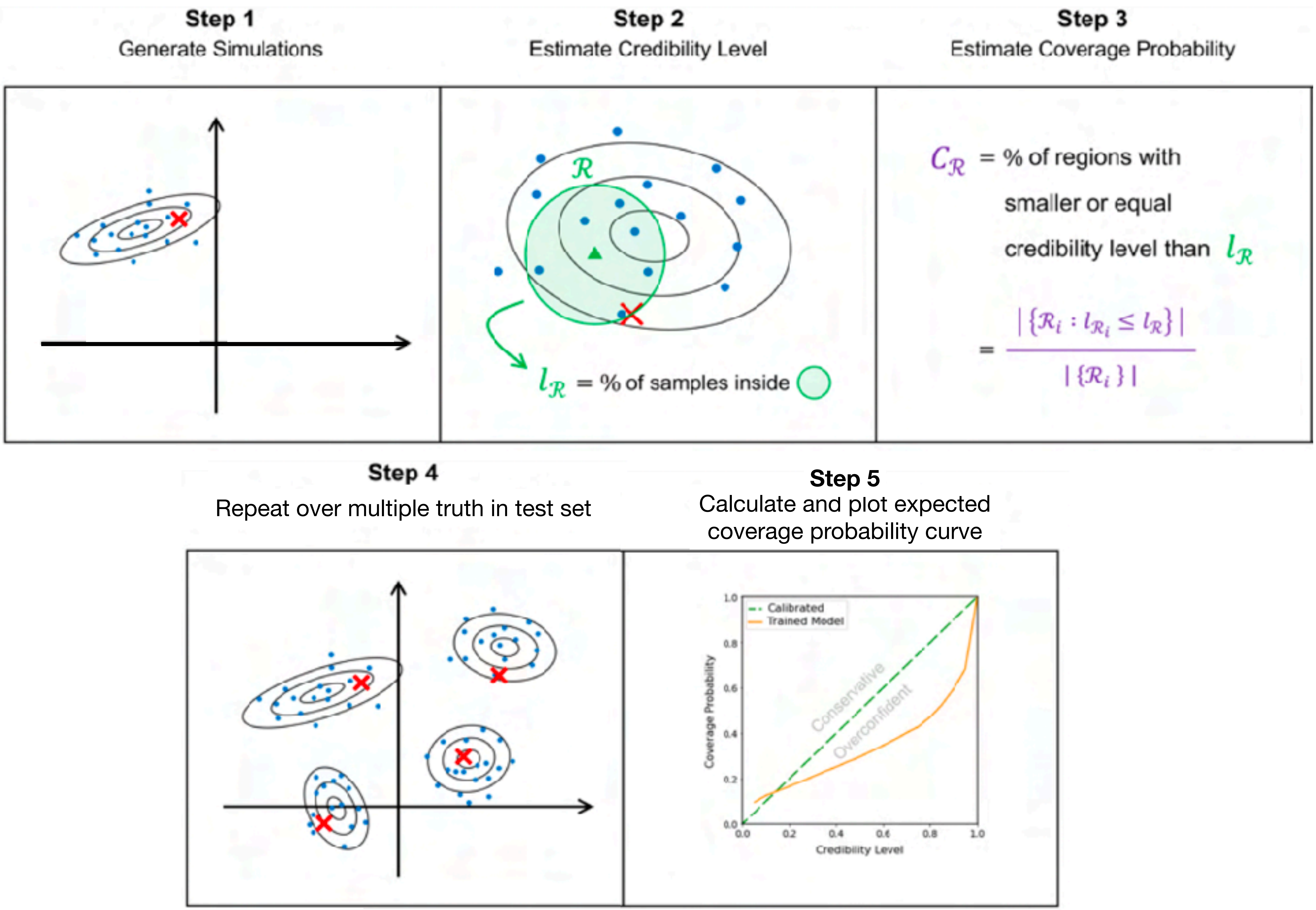
Pablo Lemos



pip install tarp

Lemos, et al. ICML 2023, 2302.03026

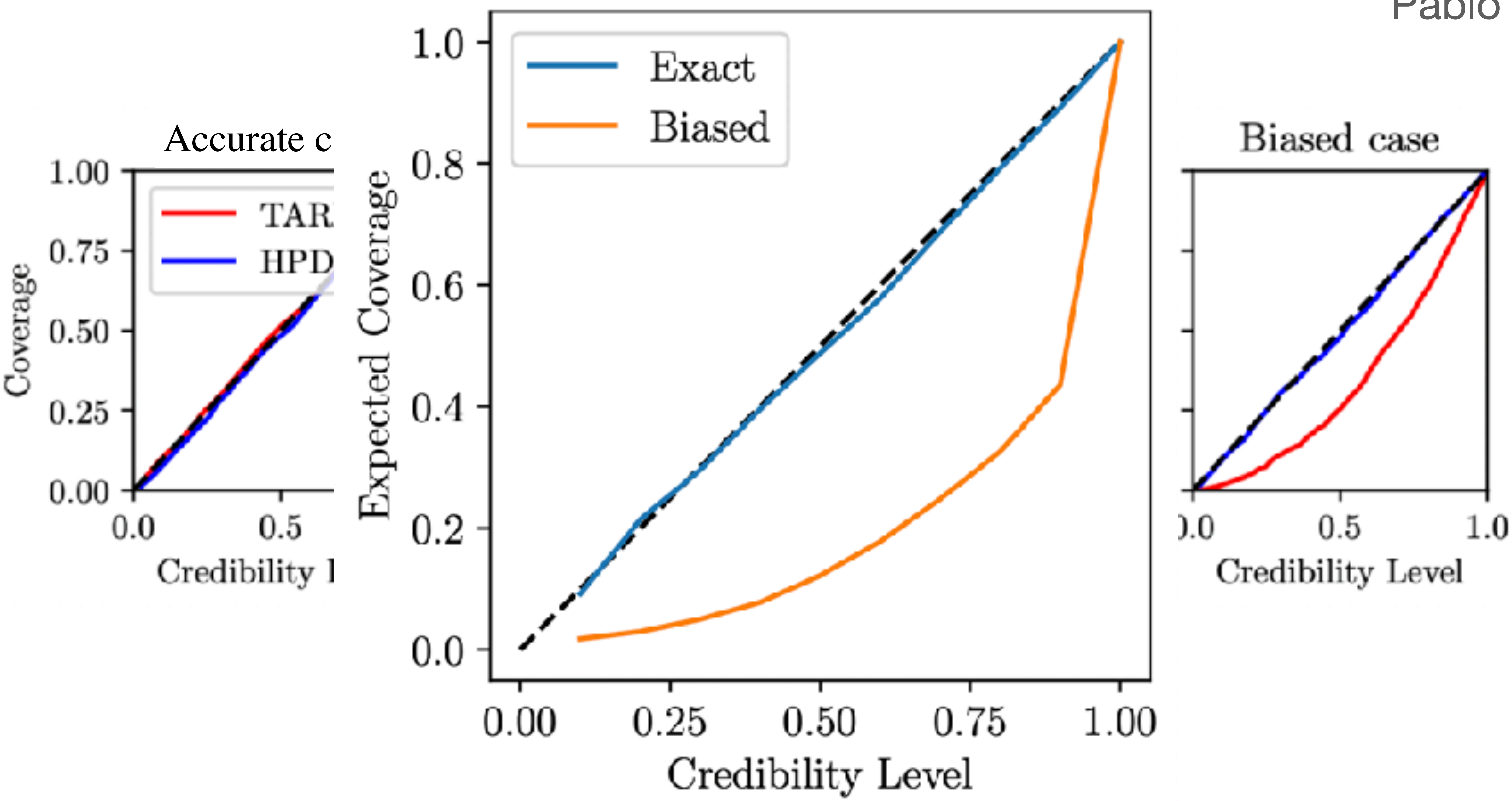
Coverage Test for Accuracy with Random Points (TARP)



Lemos, et al. ICML 2023, 2302.03026

Coverage Test for Accuracy

Pablo Lemos

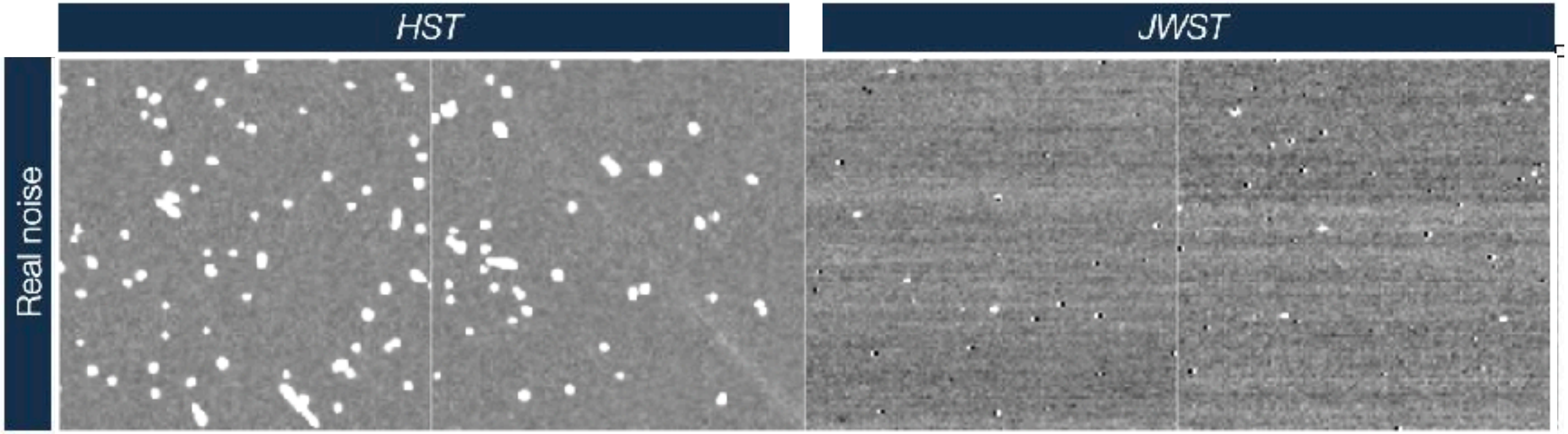


Lemos, et al. ICML 2023, 2302.03026

Dealing With Realistic Noise: Beyond Gaussianity

Alexandre Adam

Ronan Legin

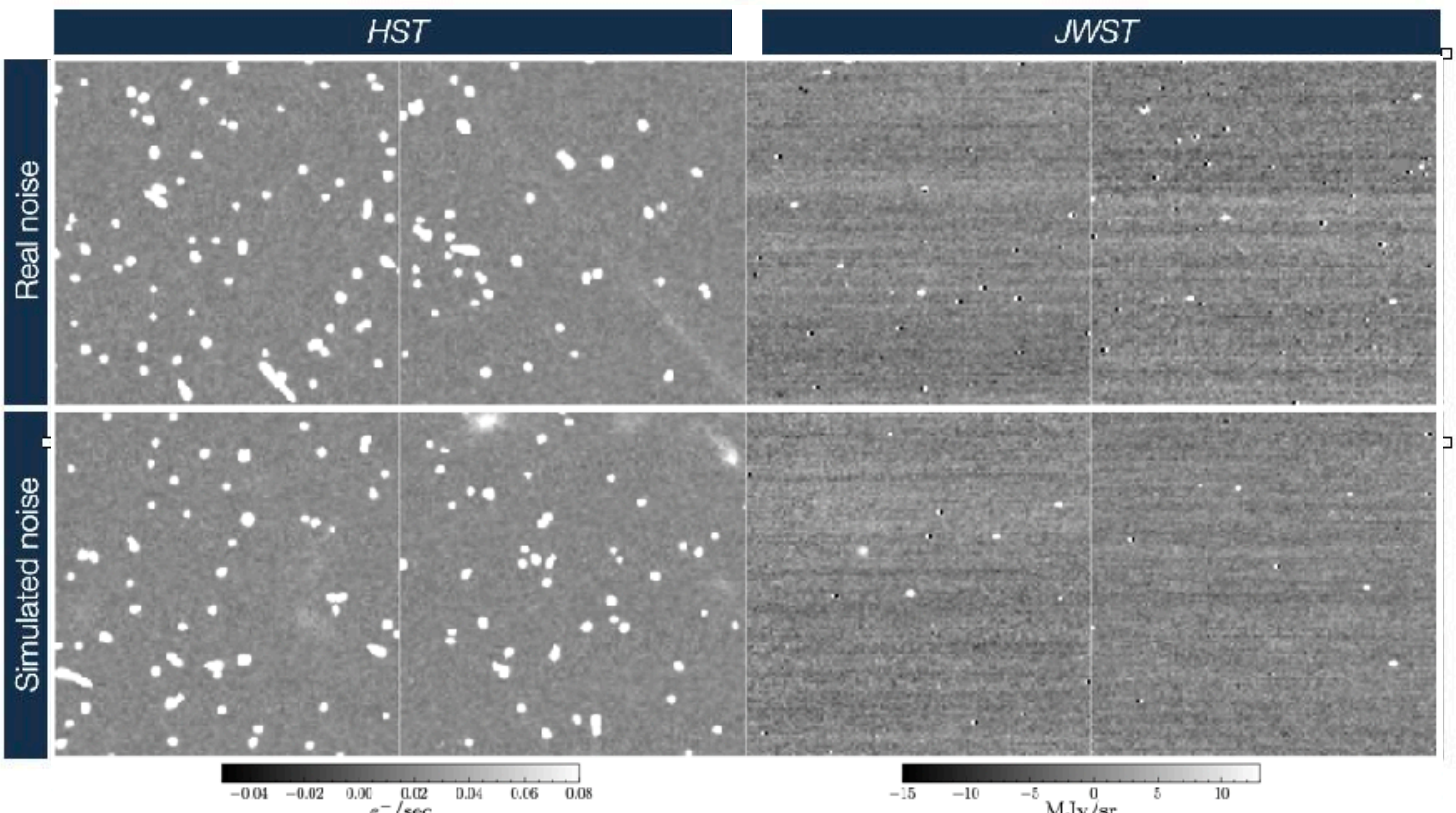


Legin, Adam, et al. ApJ Letters 2023, 2302.03046

Dealing With Realistic Noise: Beyond Gaussianity
SLIC: Score-based Likelihood Characterization

Since we have learnt a generative model of the additive noise, it can now be used in a simulation pipeline to get new, independent realizations of noise:

$$P(\mathbf{x}_{\mathrm{O}}|\eta) = Q(\mathbf{x}_{\mathrm{O}} - \mathbf{M}(\eta))$$



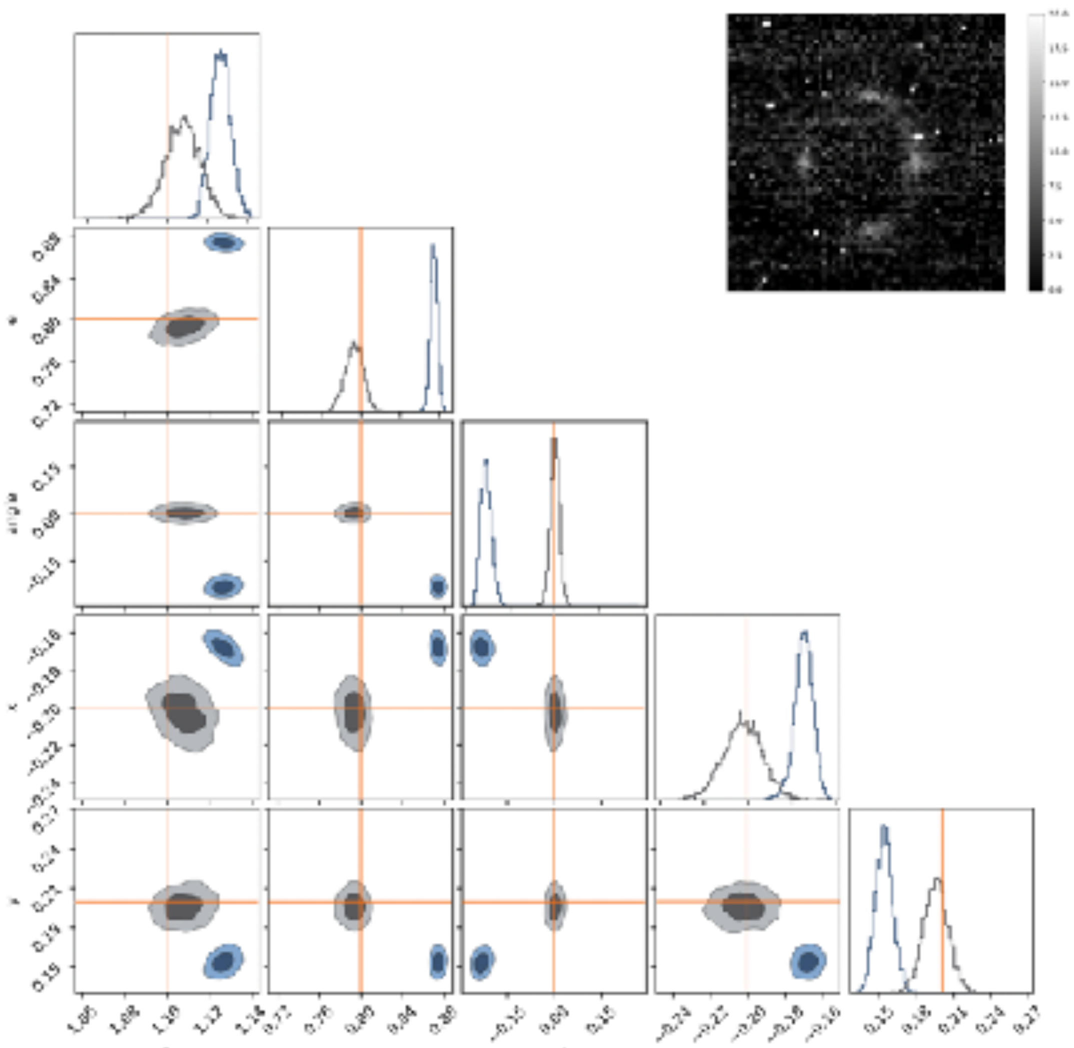
Legin, Adam, et al. ApJ Letters 2023, 2302.03046

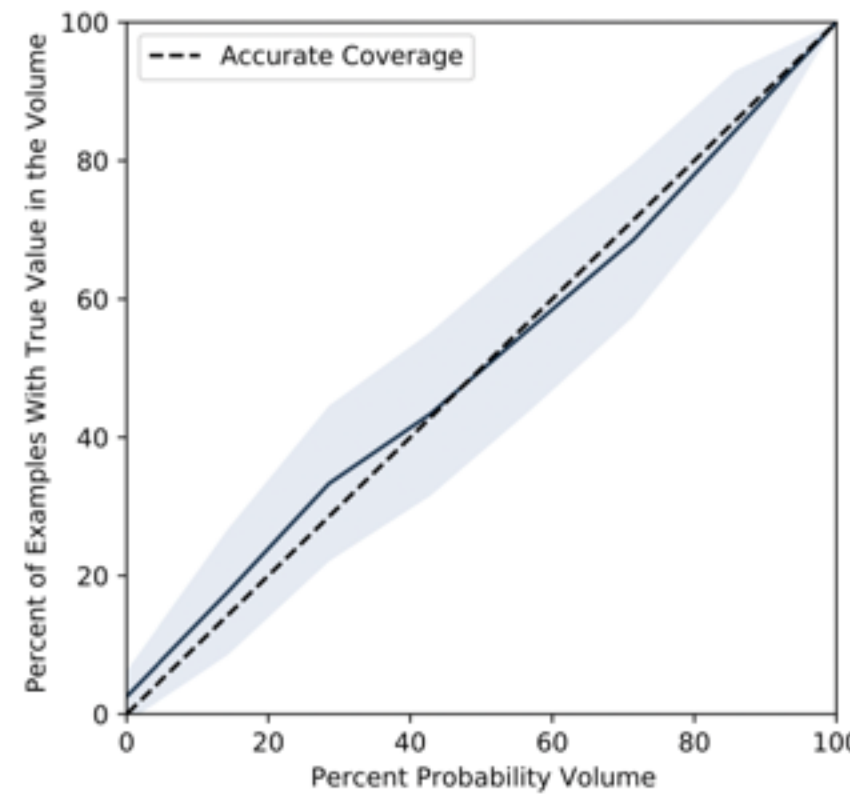
DEALING WITH REALISTIC NOISE: BEYOND GAUSSIANITY
SLIC: SCORE-BASED LIKELIHOOD CHARACTERIZATION

$$\mathbf{s}(\mathbf{x}_0) = \partial \log Q(\mathbf{x}_0)/\partial \mathbf{x}$$

$$P(\mathbf{x}_O|\eta) = Q(\mathbf{x}_O - \mathbf{M}(\eta))$$

$$\eta_{i+1} = \eta_i + \tau\, \boldsymbol{\nabla}_\mathbf{x} \log Q(\mathbf{x}_o - \mathbf{M}(\eta)) \boldsymbol{\nabla}_\eta M(\eta_i) + \sqrt{2\tau}\xi$$

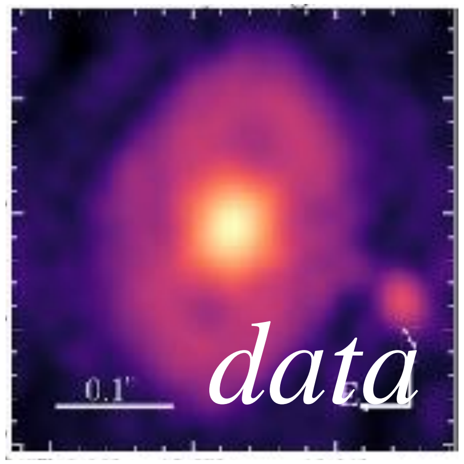




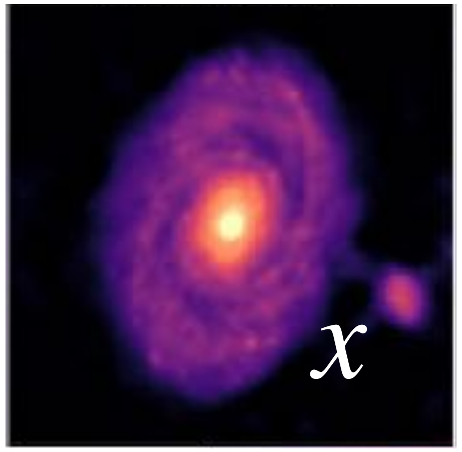
Legin, Adam, et al. ApJ Letters 2023, 2302.03046

PSF-DECONVOLUTION

Alexandre Adam



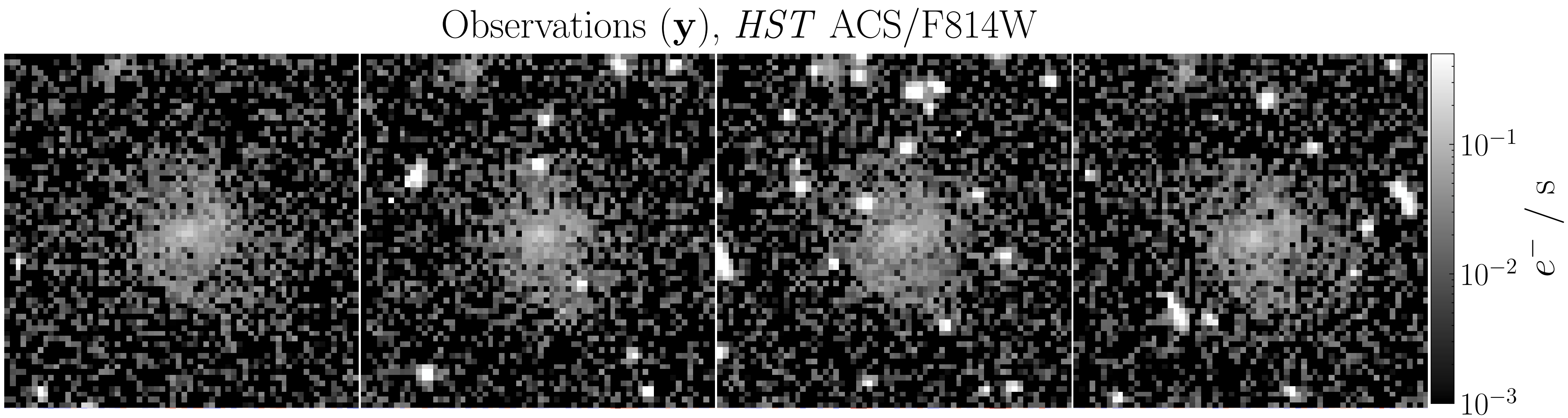
$= PSF\Big($



$\Big) + n$

PSF-DECONVOLUTION (FOR HST)

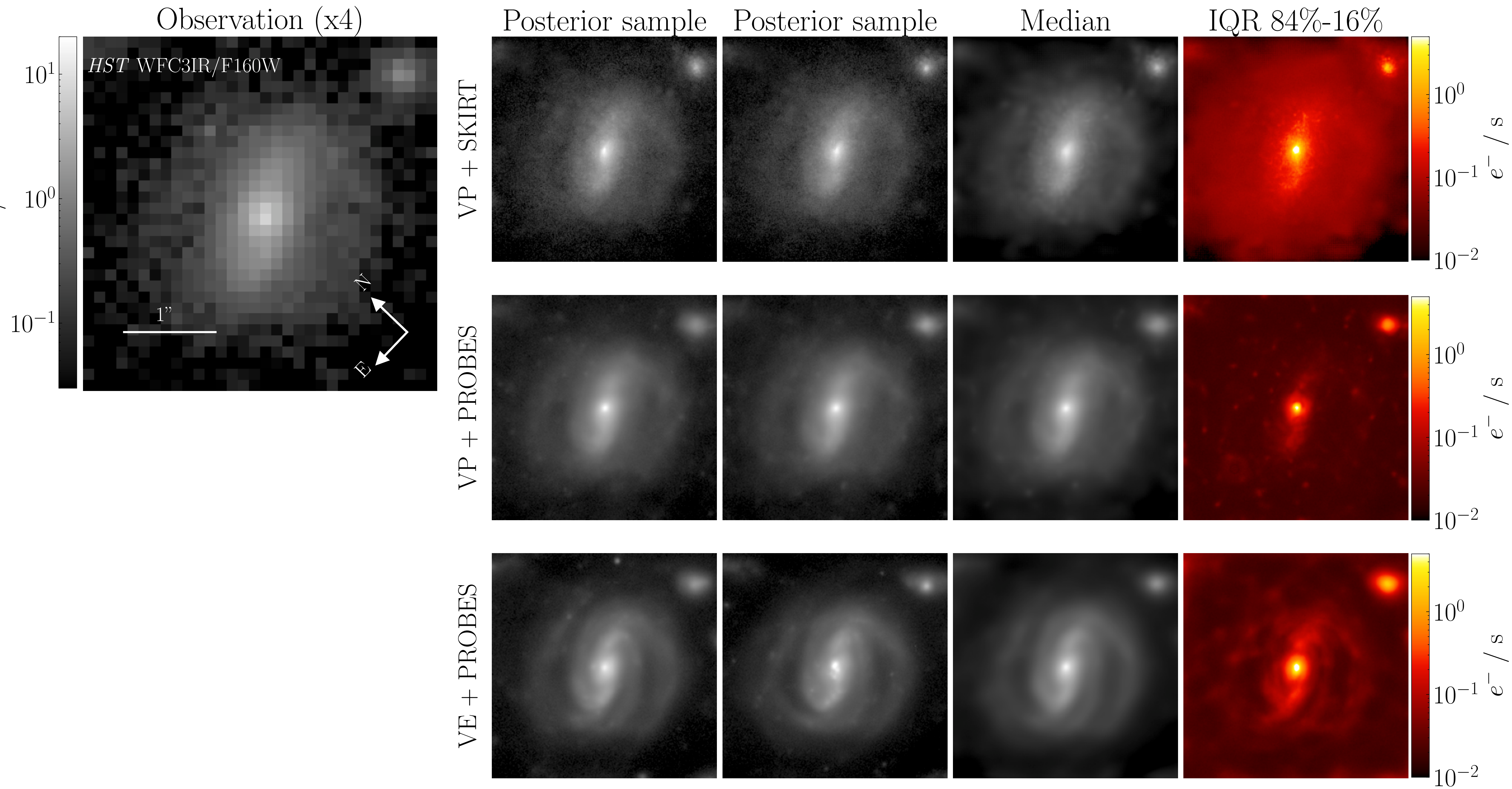
Alexandre Adam



Adam et al. NeurIPS 2023 ML4PS workshop

PSF-DECONVOLUTION (FOR HST)

Alexandre Adam



Adam et al. NeurIPS 2023 ML4PS workshop

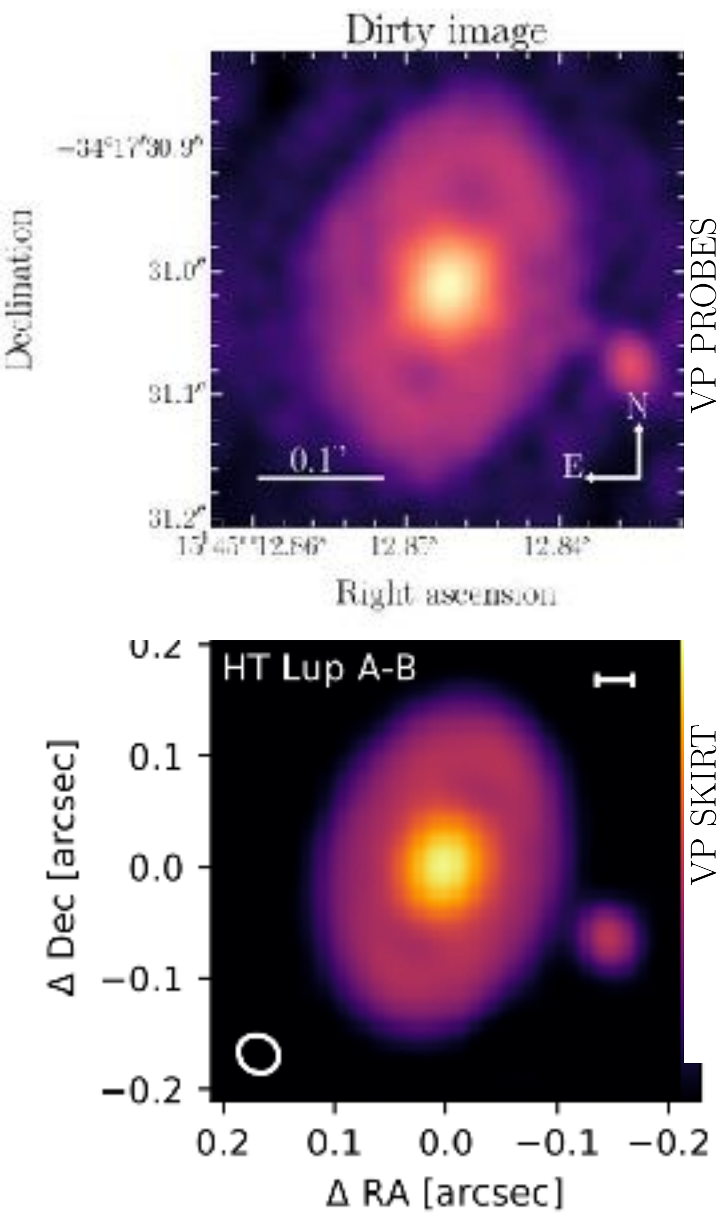
PSF-DECONVOLUTION FOR INTERFEROMETRIC DATA

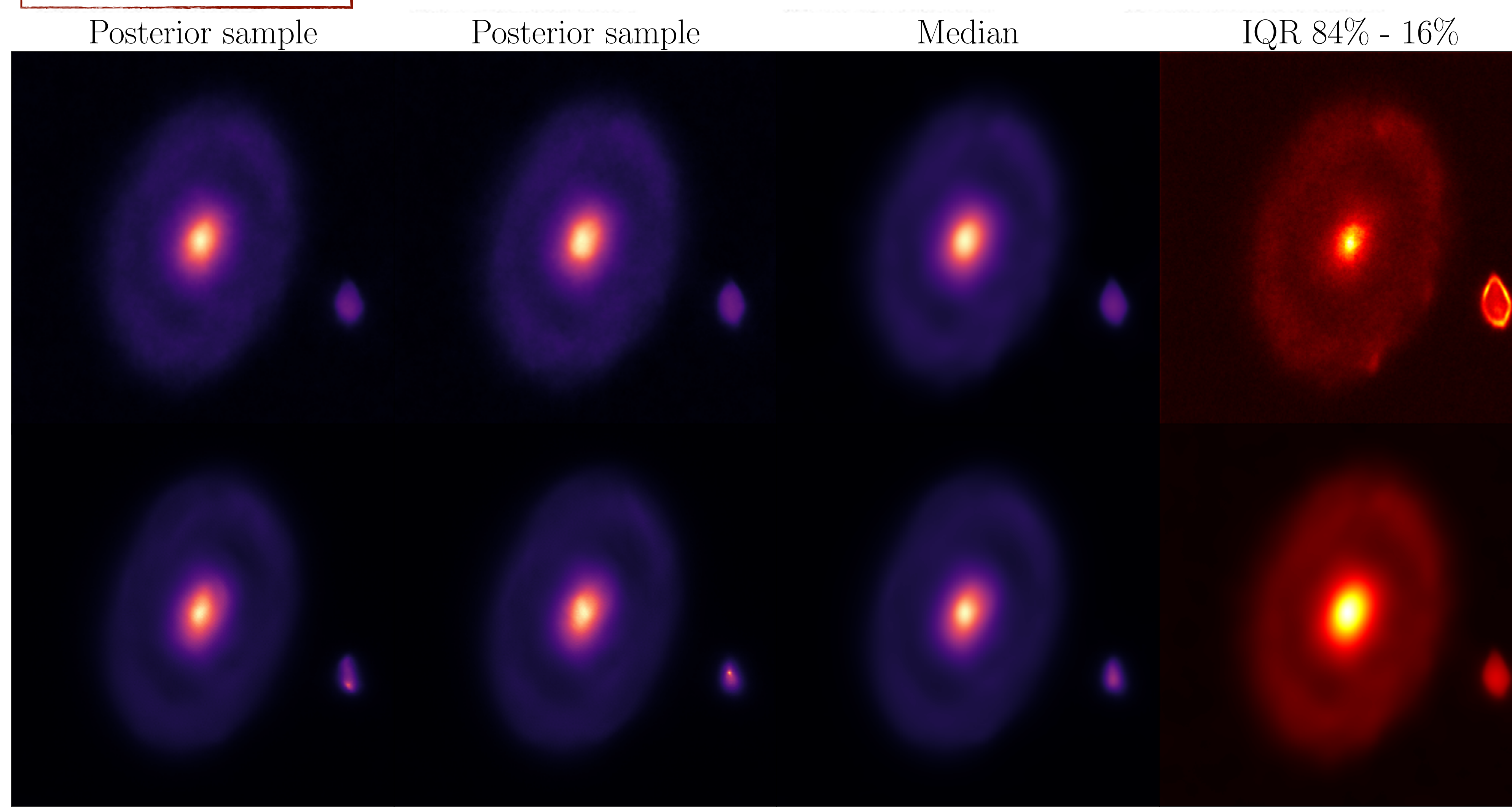
With Score-Based Diffusion Models

Noé Dia

$$\mathcal{V} = S\mathcal{F}P_{\text{beam}}\mathbf{x} + \boldsymbol{\eta}$$

- Observed visibilities → $\mathcal{V}$
- Mask, subsampling function → S
- Dense, unitary, Fourier operator → $\mathcal{F}$
- Primary beam → P_{beam}
- Sky emission → $\mathbf{x}$
- Gaussian noise → $\boldsymbol{\eta}$





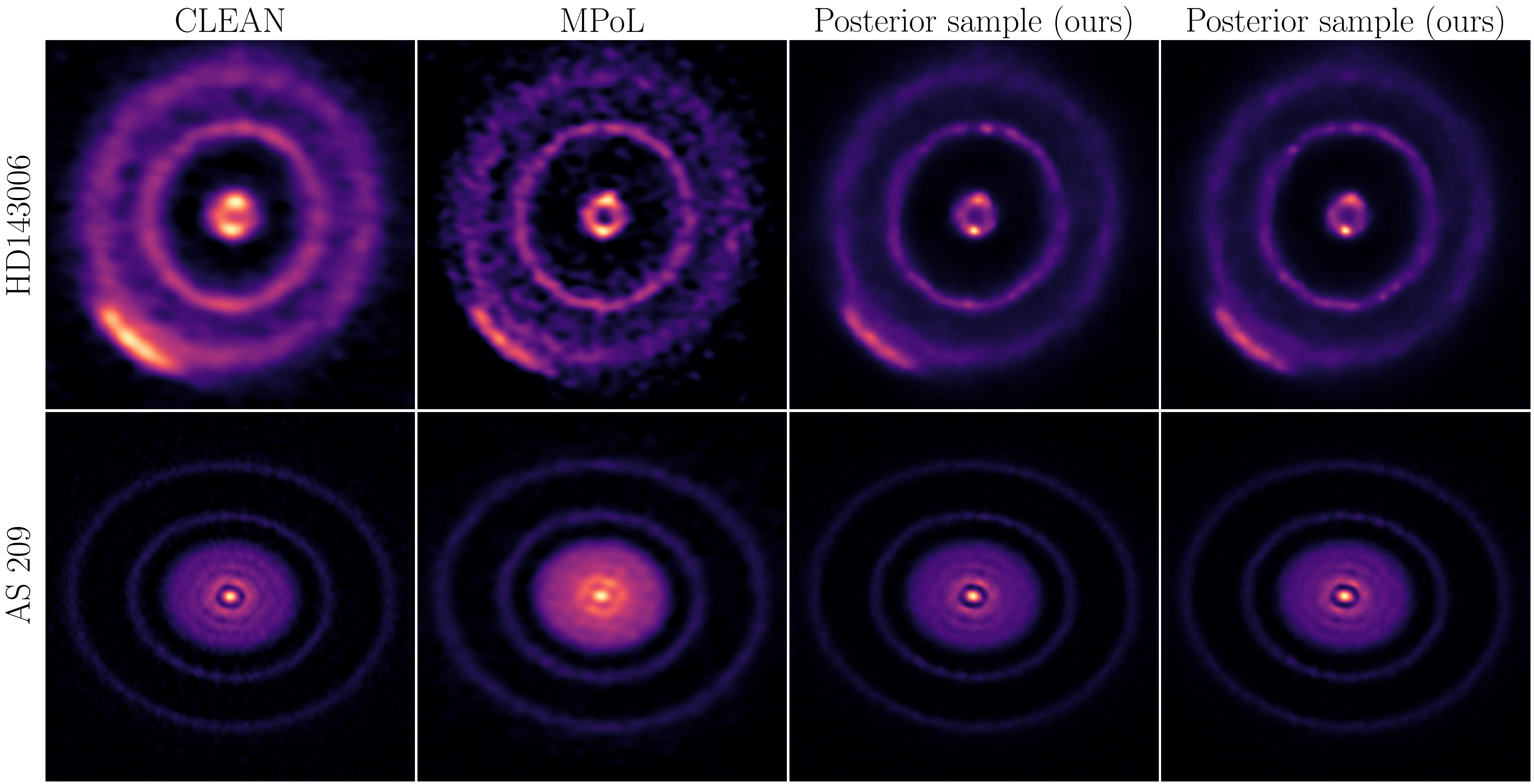
Dia, Adam, Barth, et al. NeurIPS 2023 ML4PS workshop

PSF-DECONVOLUTION FOR INTERFEROMETRIC DATA

With Score-Based Diffusion Models

$$\mathcal{V} = S\mathcal{F}P_{\text{beam}}\mathbf{x} + \boldsymbol{\eta}$$

Noé Dia

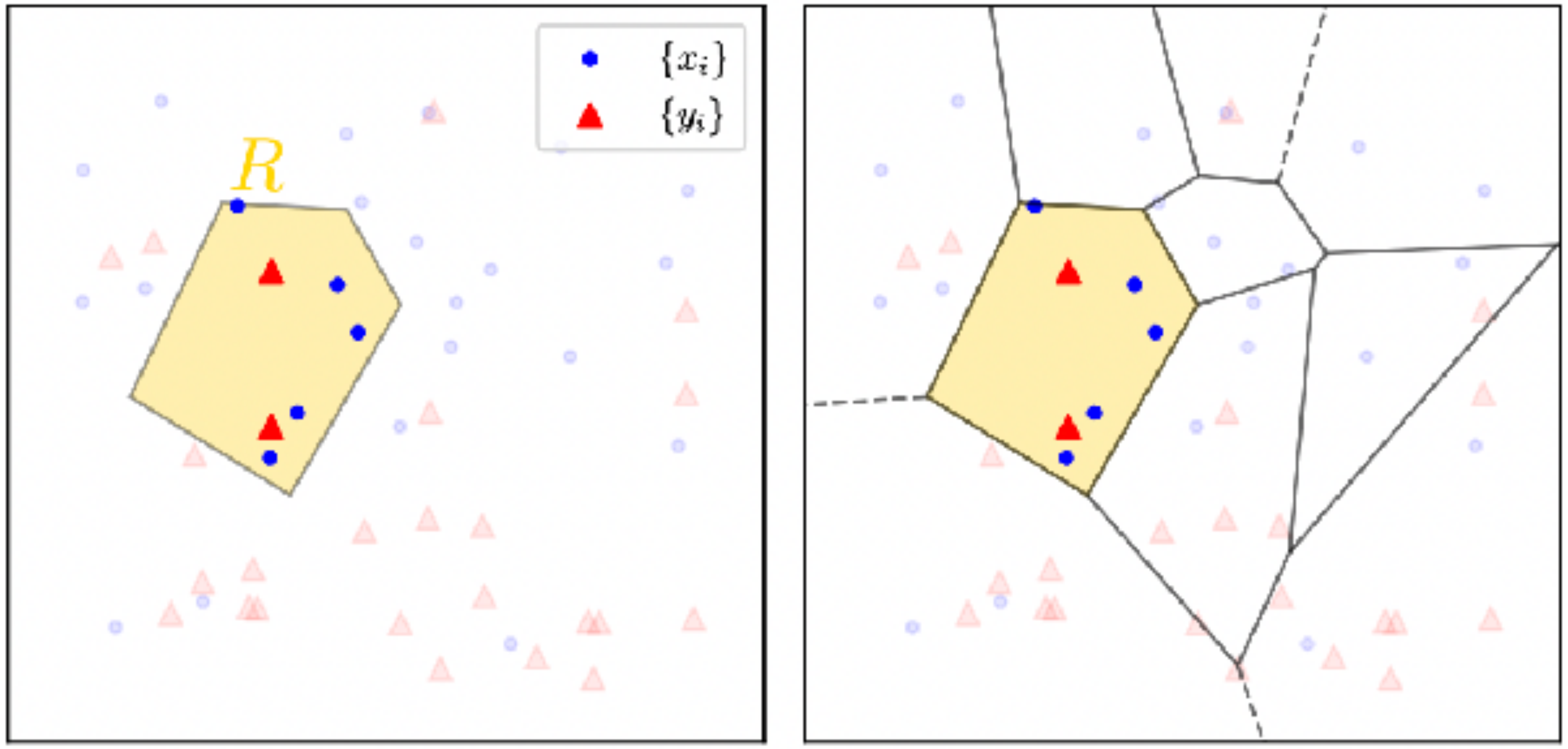


Dia, Adam, Barth, et al. NeurIPS 2023 ML4PS workshop

PQMass: Probabilistic assessment of Generative Models using Probability Mass Estimation

Sammy Sharief

Pablo Lemos



$$k(\mathbf{x}, R) \sim \mathcal{B}(n, \mathbb{P}_p(R))$$

$$\{k(\mathbf{x}, R_i)\}_{i=1\ldots n_R} \sim \mathcal{M}\left(n, \{\mathbb{P}_p(R_i)\}_{i=1\ldots n_R}\right)$$

$$\hat{N}_{x,i} \equiv n\hat{p}_{R_i}, \qquad \hat{N}_{y,i} \equiv m\hat{p}_{R_i},$$

$$\hat{p}_{R_i} \equiv \frac{k(\mathbf{x}, R_i) + k(\mathbf{y}, R_i)}{n+m}.$$

$$\chi^2_{\mathrm{PQM}} \equiv \sum_{i=1}^{n_R} \left[\frac{(k(\mathbf{x}, R_i) - \hat{N}_{x,i})^2}{\hat{N}_{x,i}} + \frac{(k(\mathbf{y}, R_i) - \hat{N}_{y,i})^2}{\hat{N}_{y,i}} \right]$$

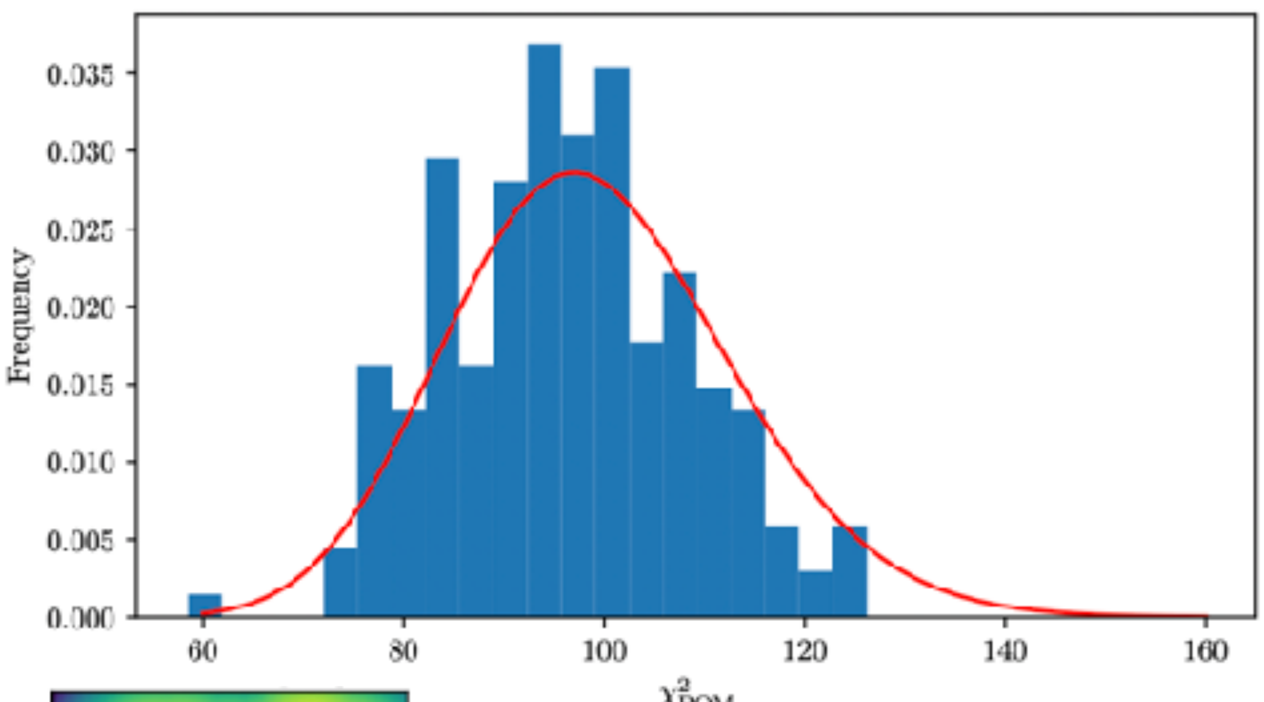
Lemos et al. 2402.04355

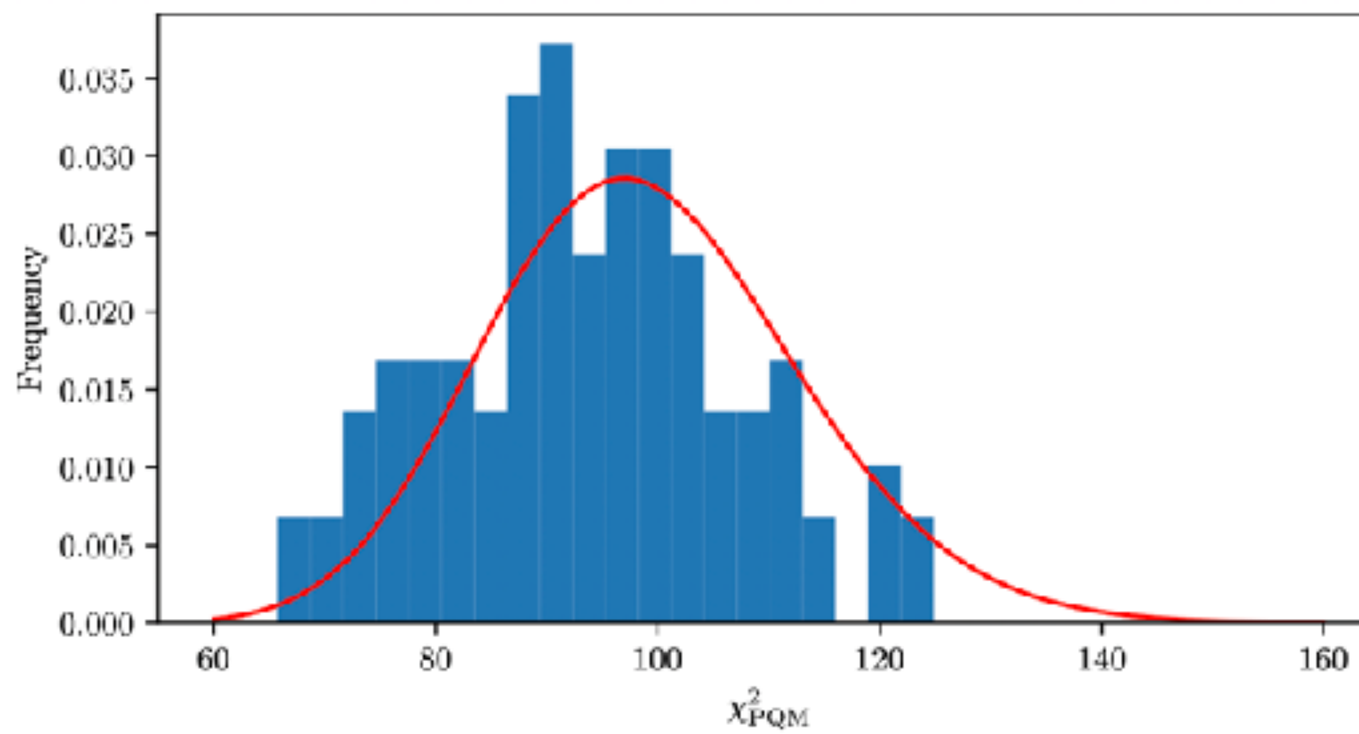
PQMass: Probabilistic assessment of Generative Models using Probability Mass Estimation

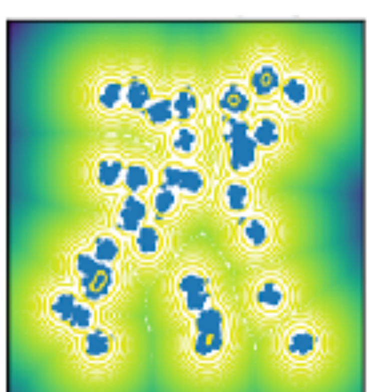
Sammy Sharief

Pablo Lemos

Given **any** *sampling distribution, or generative model, if two sets of samples are generated from the same distribution, then the statistic χ^2_{PQM} follows a chi-square distribution with $n_R - 1$ degrees of freedom.*



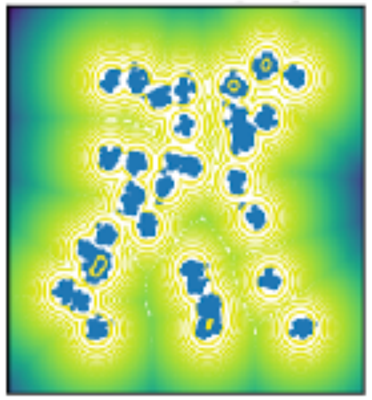


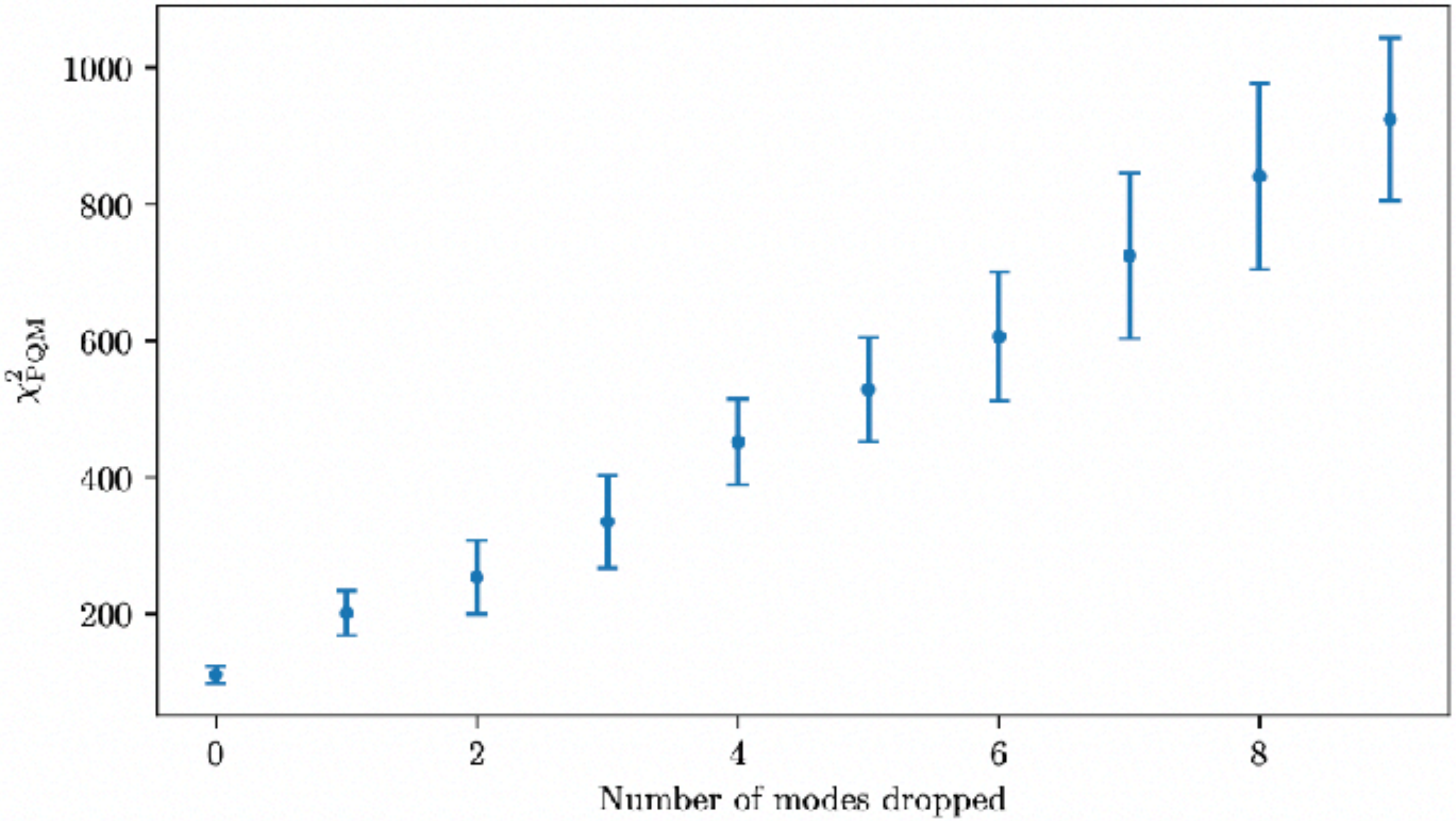


PQMASS: PROBABILISTIC ASSESSMENT OF GENERATIVE MODELS USING PROBABILITY MASS ESTIMATION

Sammy Sharief

Pablo Lemos





PQMass: Probabilistic assessment of Generative Models using Probability Mass Estimation

Sammy Sharief

Pablo Lemos