Data-Driven High-Dimensional Inverse Problems: A Journey Through Strong Gravitational Lensing Data Analysis

Laurence Perreault-Levasseur

Ciela Institute

Université de Montréal

Mila

Flatiron Institute
Center for Computational Astrophysics

Modern Cosmology+Astrophysics

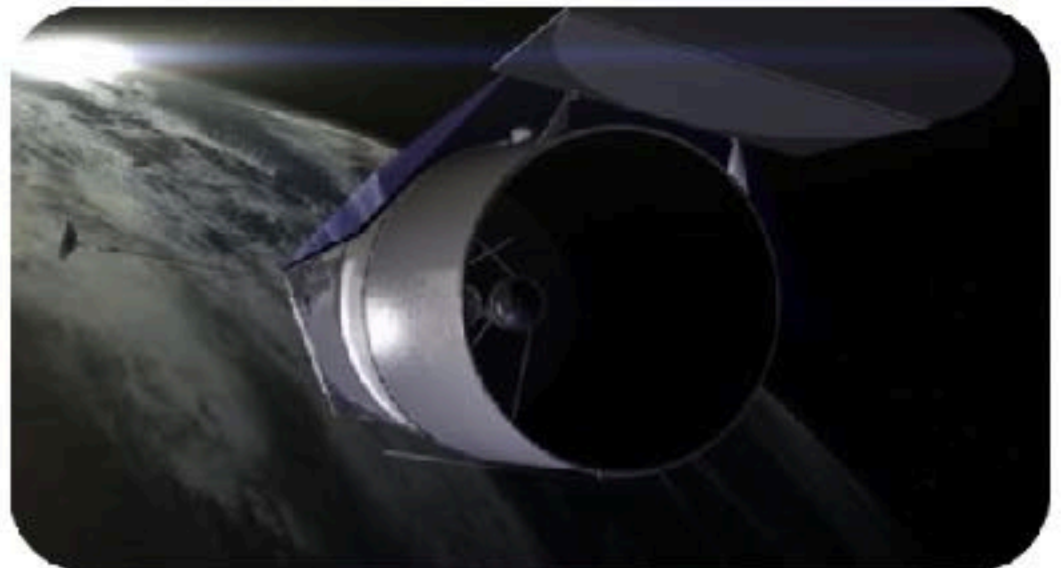

SKA


TMT

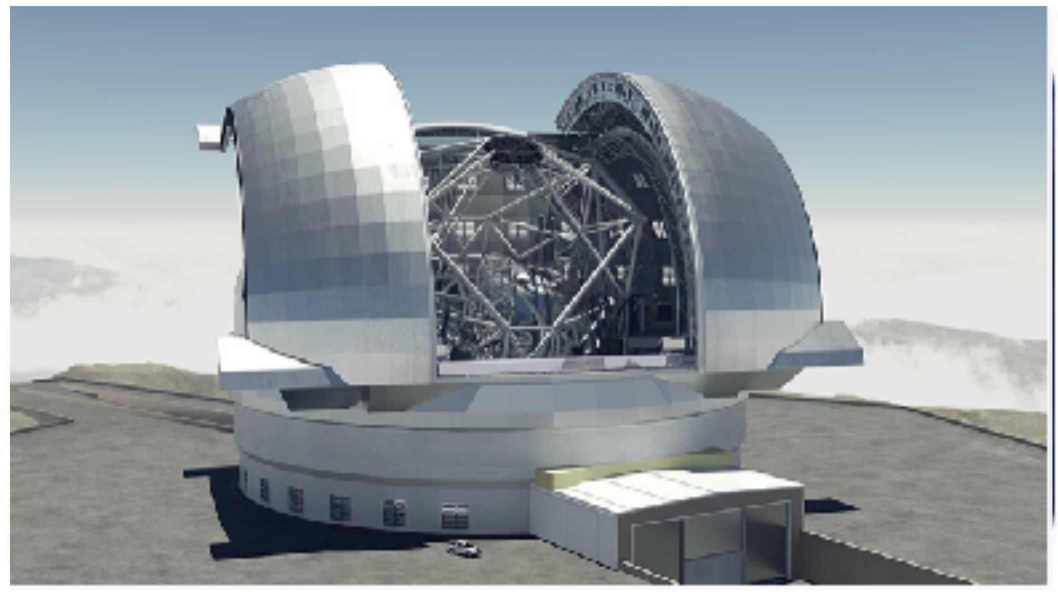

JWST

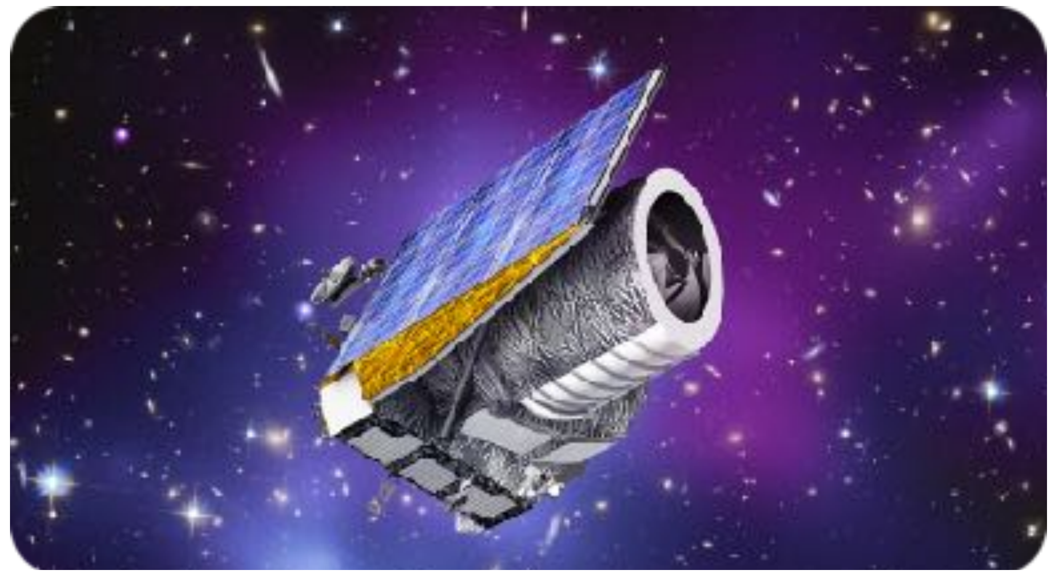
LSST


ALMA


ROMAN

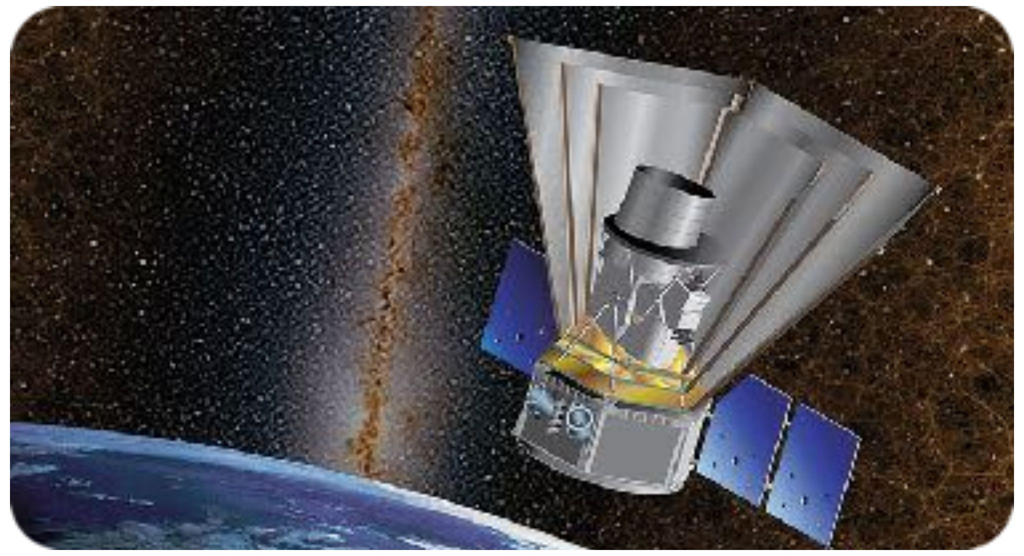

ATHENA


ELT


Euclid


Simons Observatory


CHIME


SPHEREX

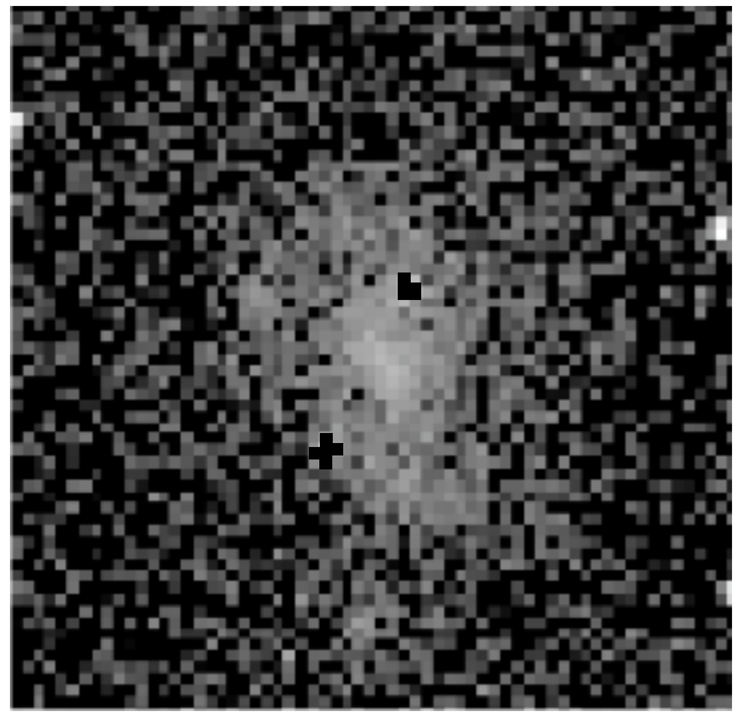
Simple models can be sufficient for simple data

Simple models can be sufficient for simple data

Simple/low resolution/low SNR data ->   simple model might be sufficient

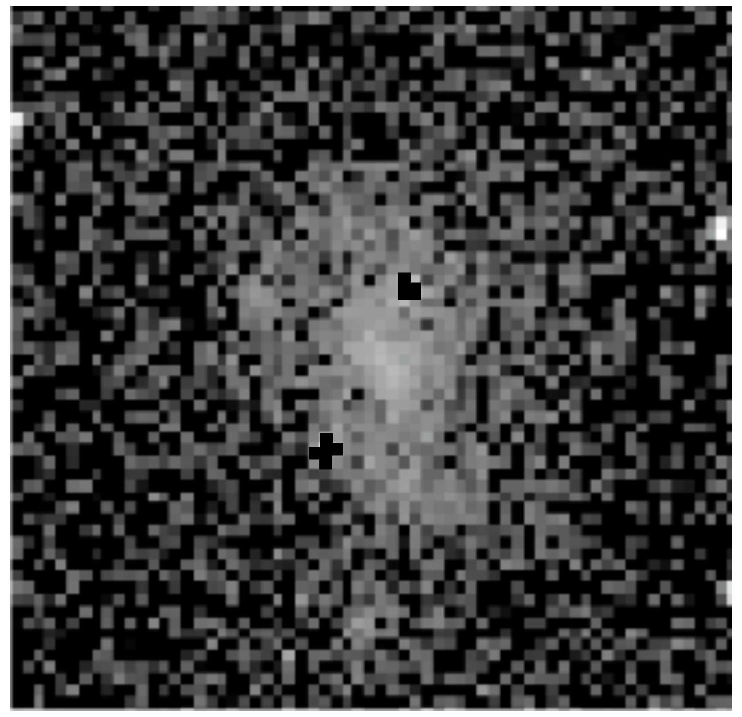
Simple models can be sufficient for simple data

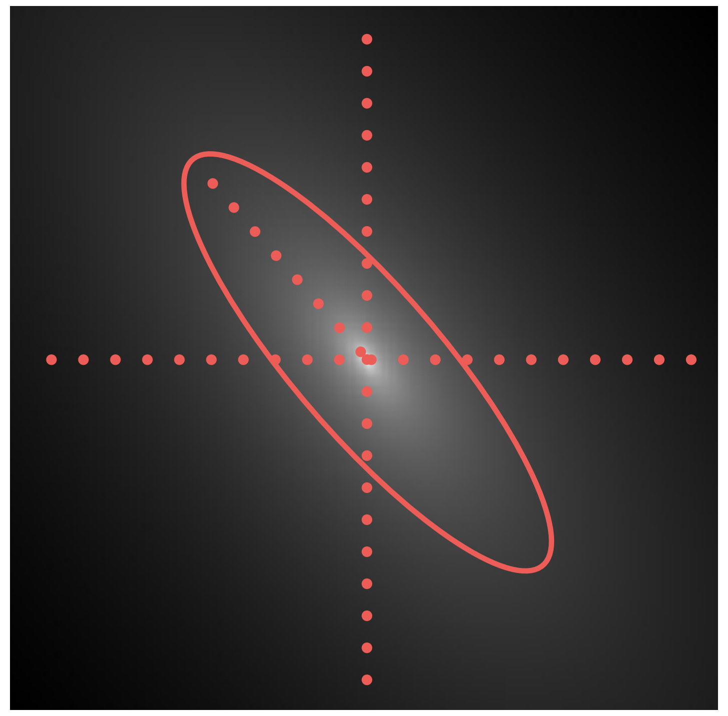
Simple/low resolution/low SNR data -> simple model might be sufficient



Simple models can be sufficient for simple data

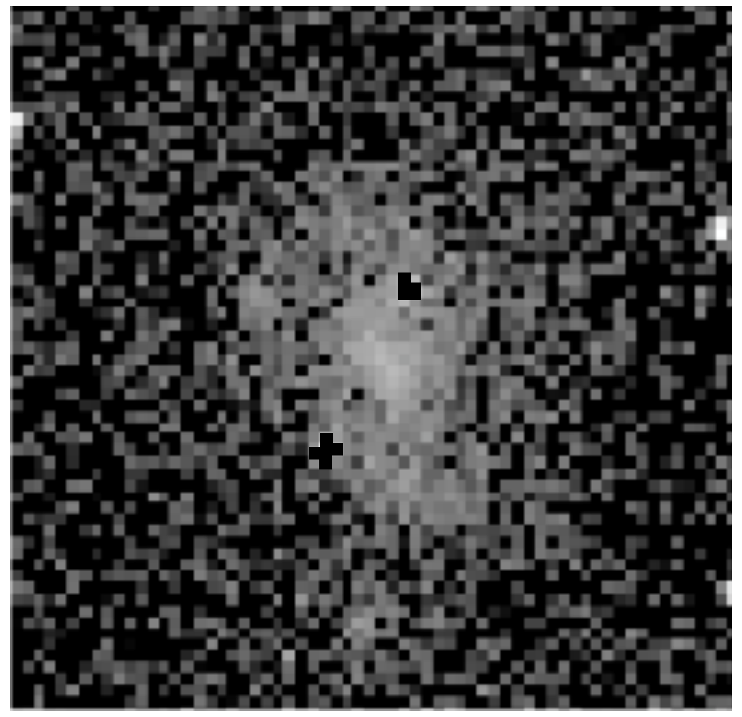
Simple/low resolution/low SNR data -> simple model might be sufficient

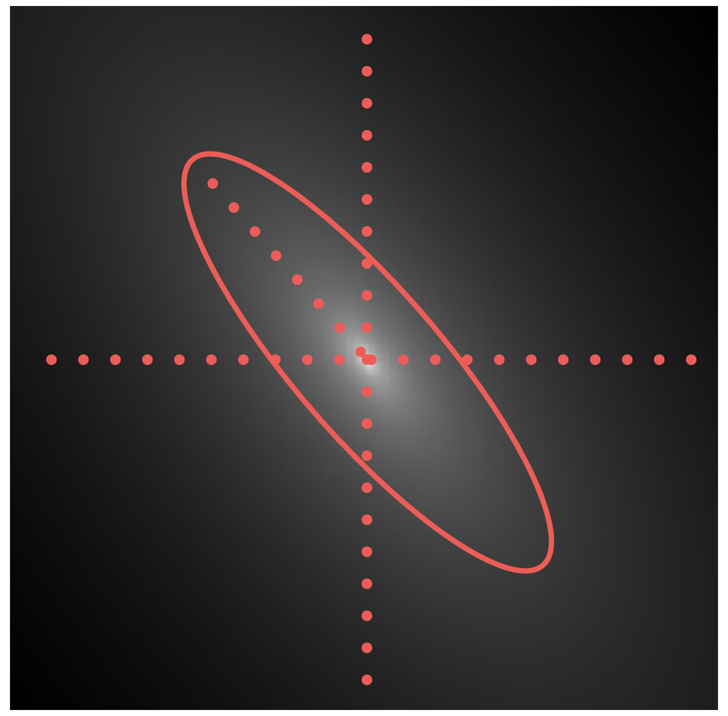




Simple models can be sufficient for simple data

Simple/low resolution/low SNR data -> simple model might be sufficient

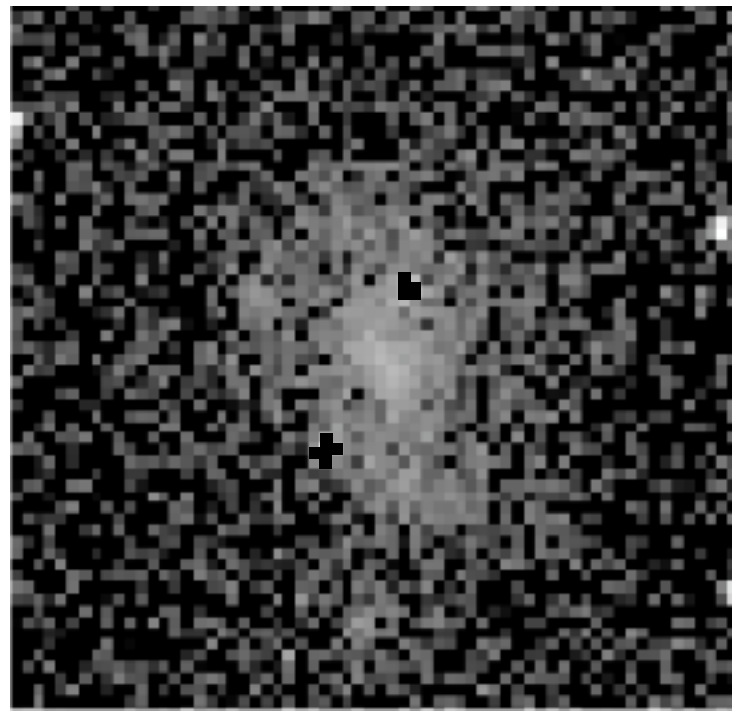


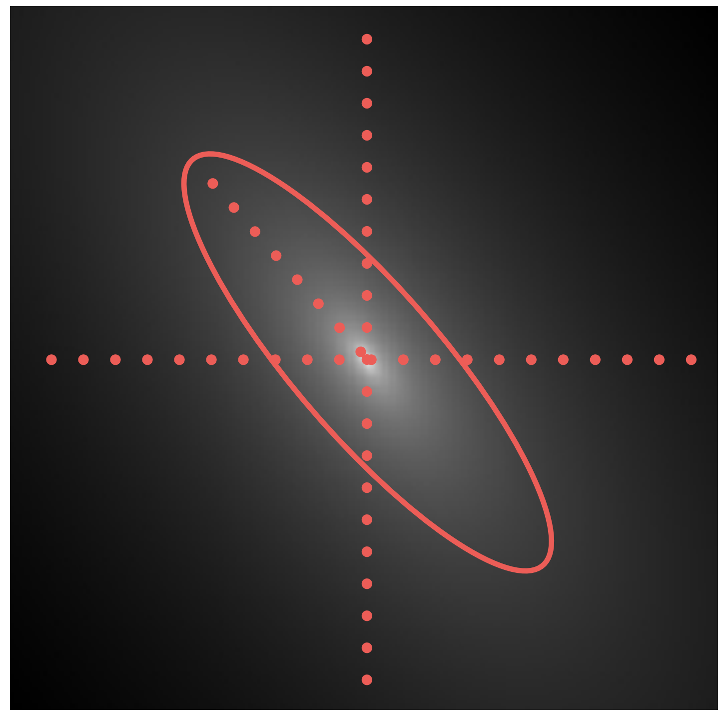


Better data -> simplistic models can bias the inference

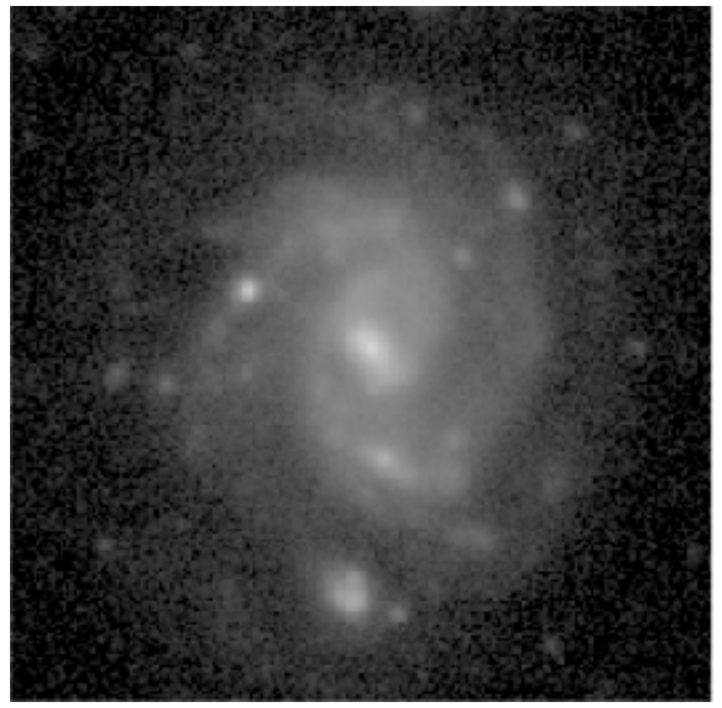
SIMPLE MODELS CAN BE SUFFICIENT FOR SIMPLE DATA

Simple/low resolution/low SNR data -> simple model might be sufficient



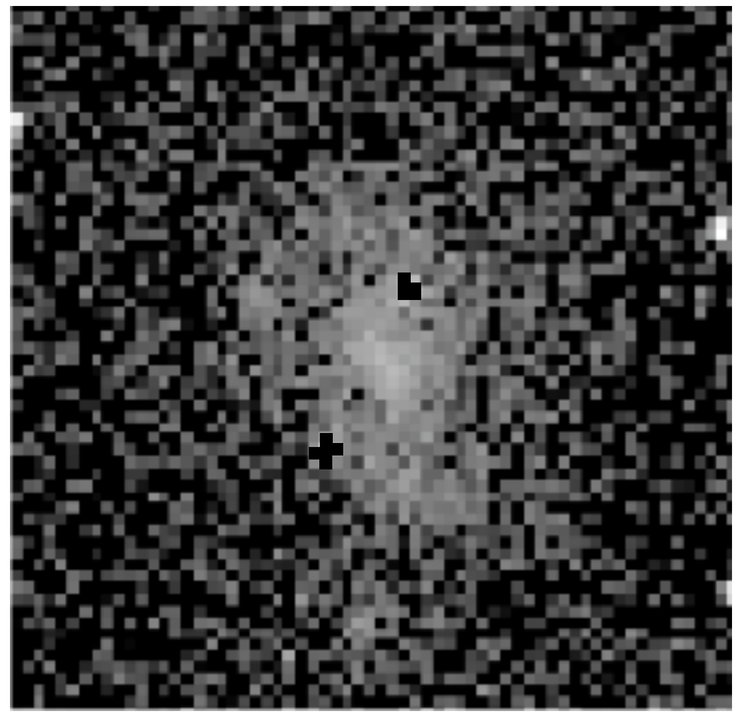


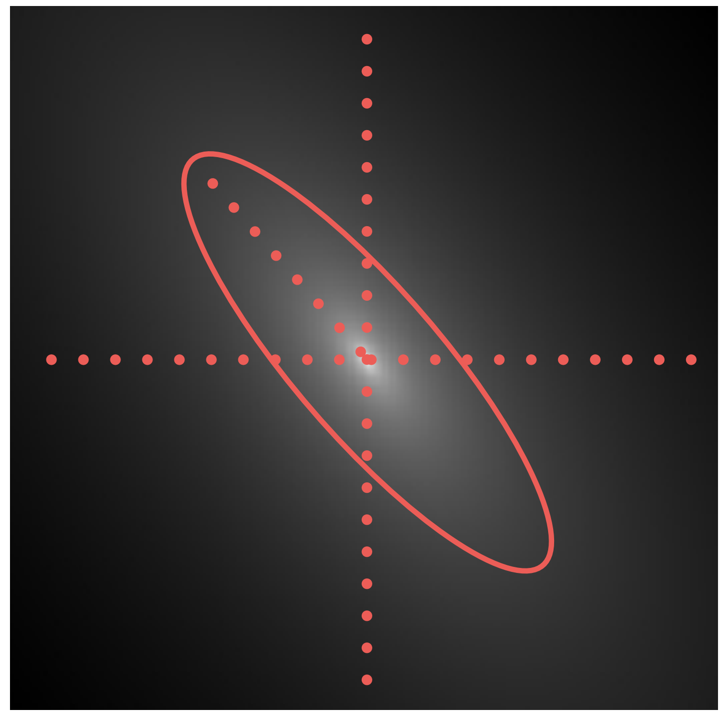
Better data -> simplistic models can bias the inference



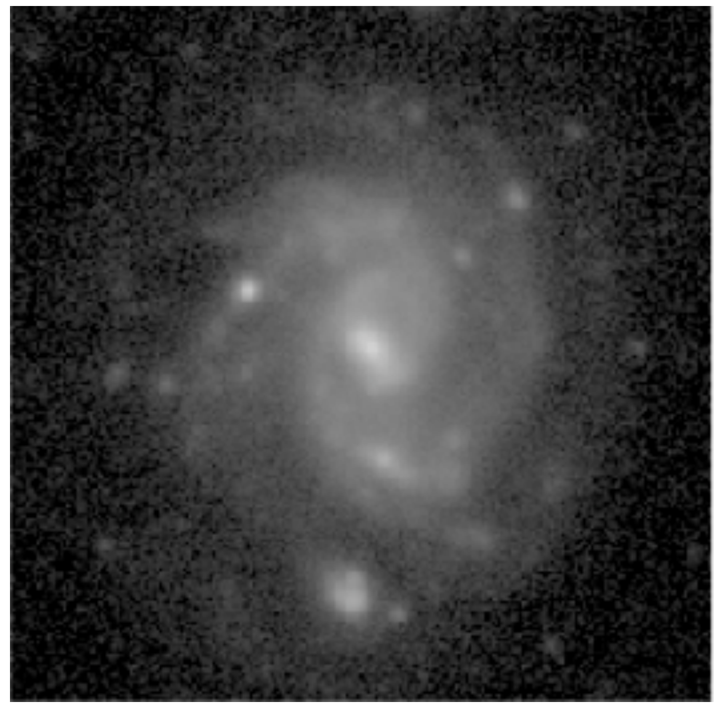
SIMPLE MODELS CAN BE SUFFICIENT FOR SIMPLE DATA

Simple/low resolution/low SNR data -> simple model might be sufficient





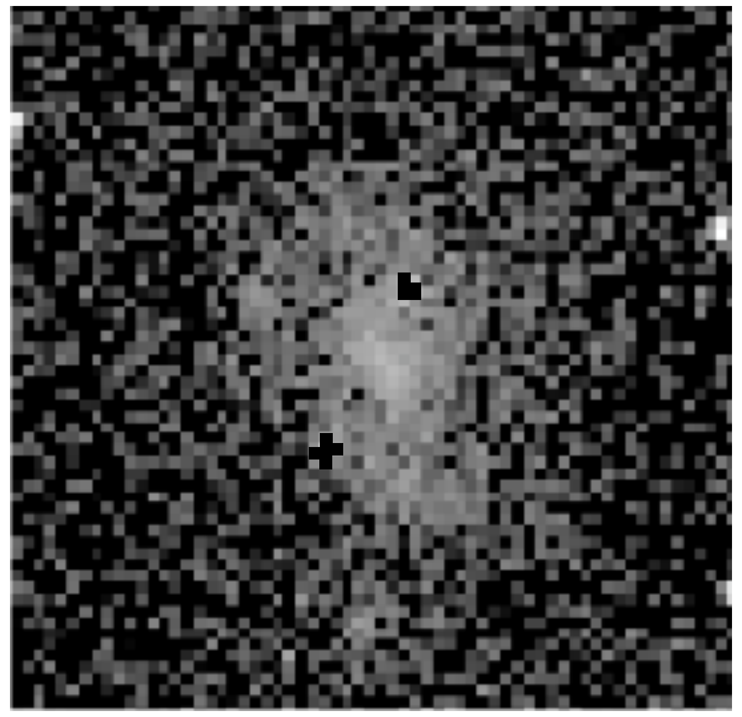
Better data -> simplistic models can bias the inference

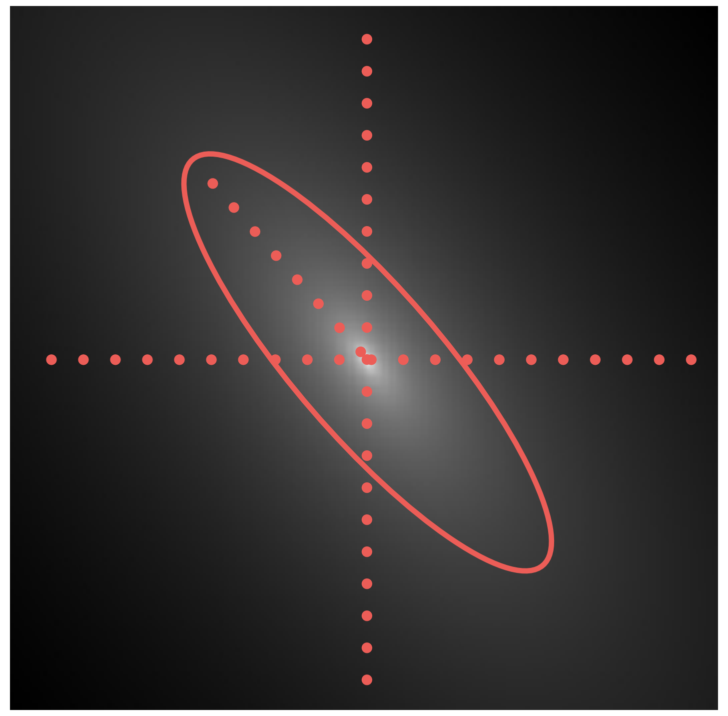


?

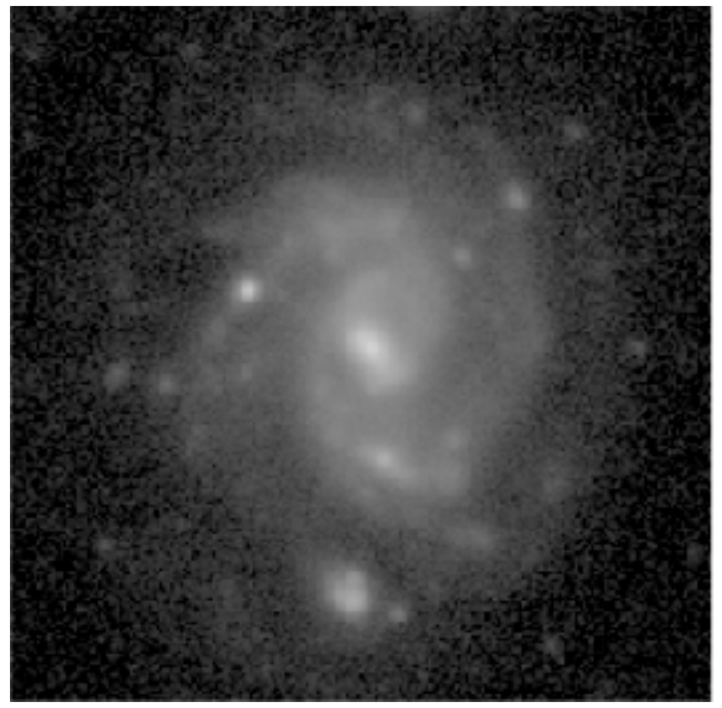
Simple models can be sufficient for simple data

Simple/low resolution/low SNR data -> simple model might be sufficient





Better data -> simplistic models can bias the inference



?

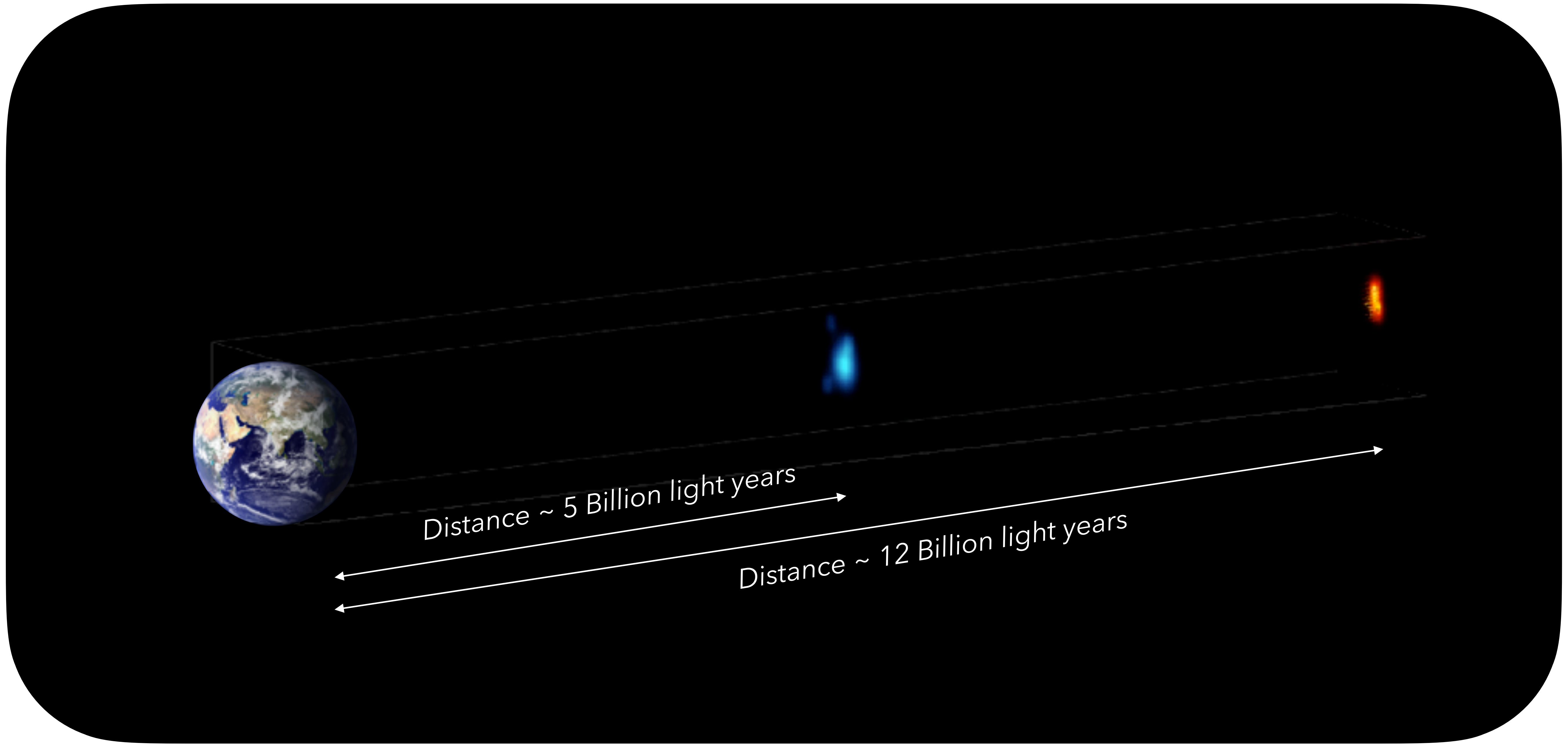
It's becoming more and more challenging to come up with analytical models complex enough to keep up with the increasing complexity of our data.

GAME PLAN

1. Data-driven priors
2. Data-driven likelihoods
3. Accuracy metrics
4. Out-of-Distribution accuracy

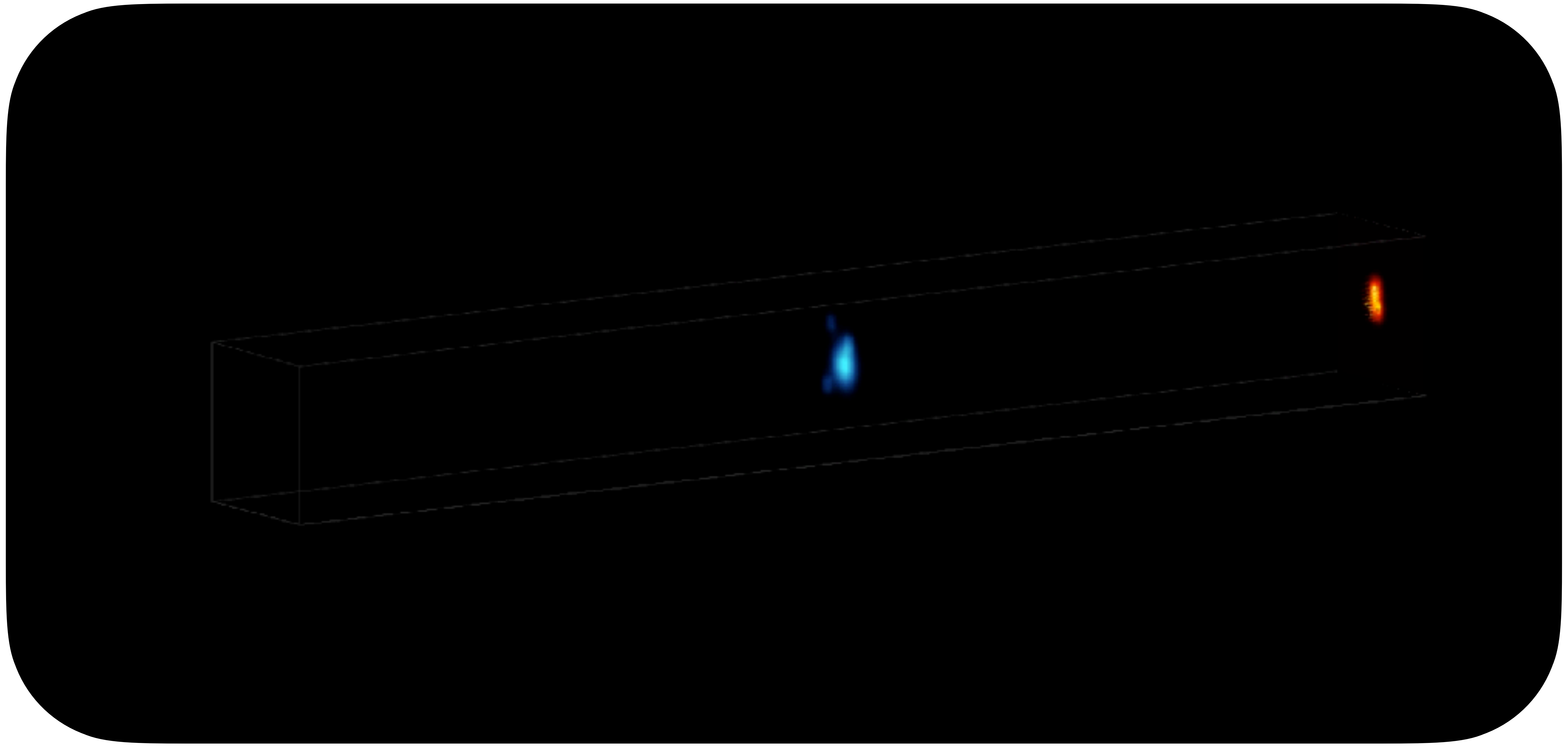
Strong Gravitational Lensing

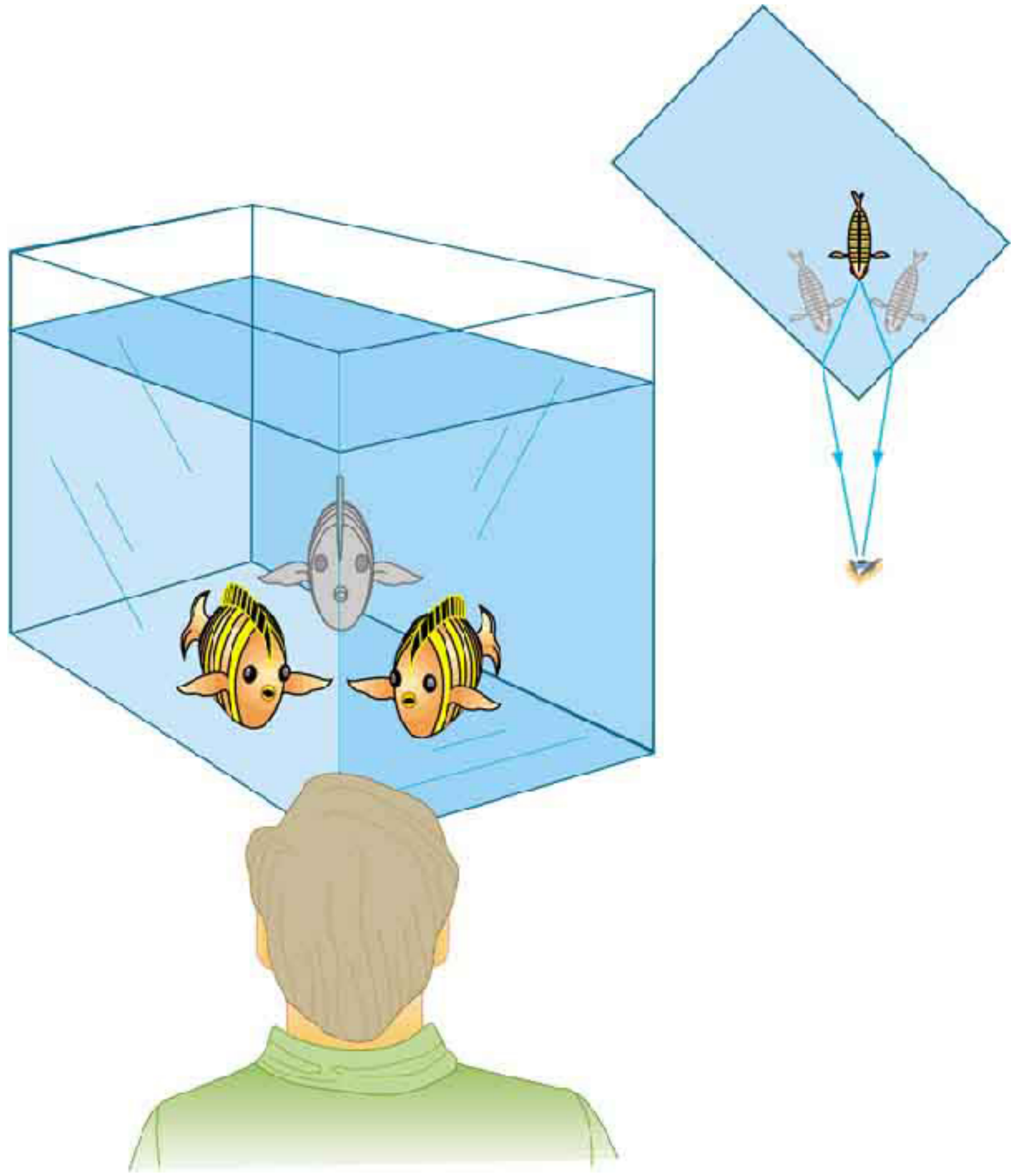
Formation of **multiple images** of a single distant object due to the **deflection of its light** by the **gravity** of intervening structures.

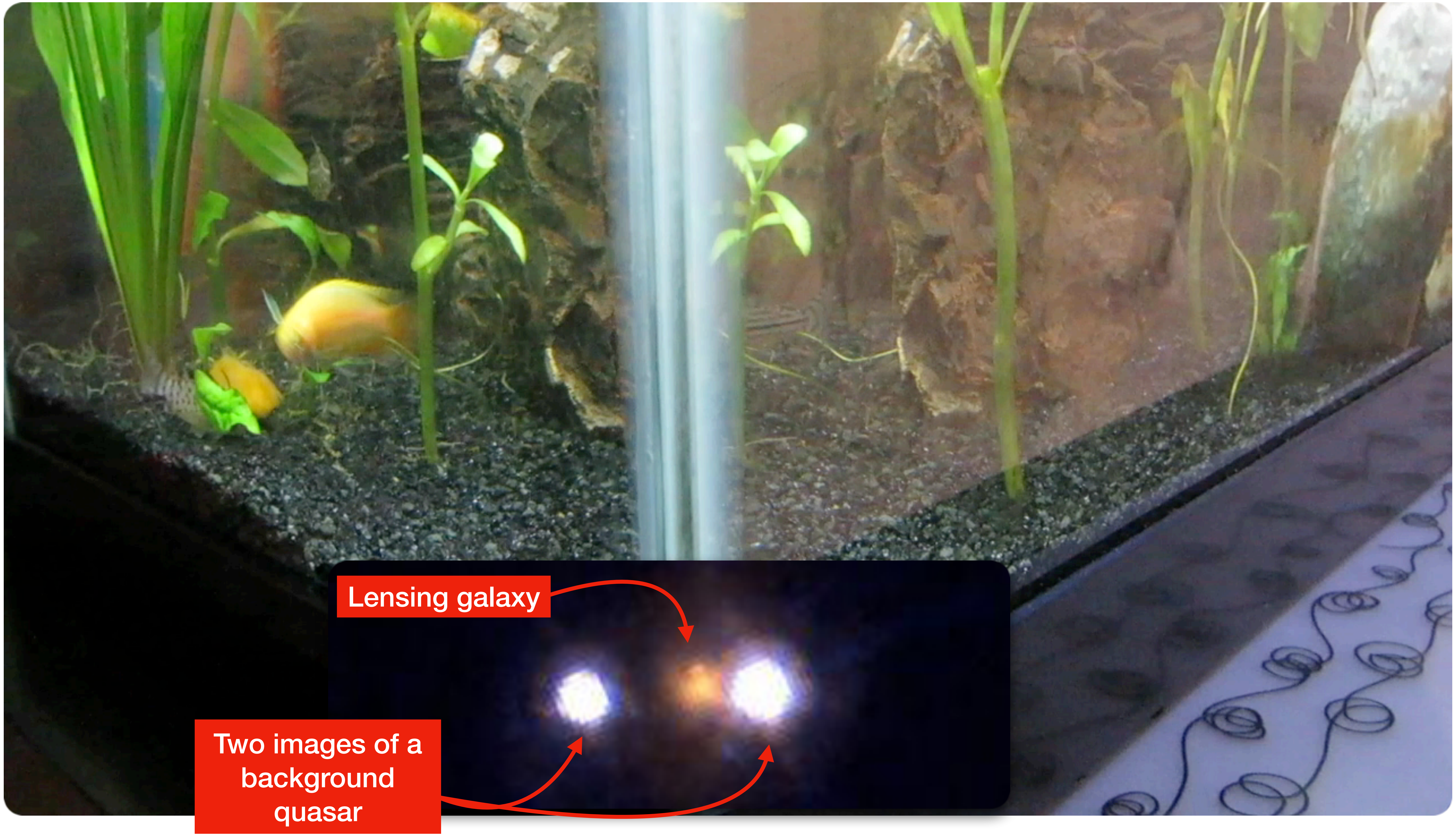


Strong Gravitational Lensing

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

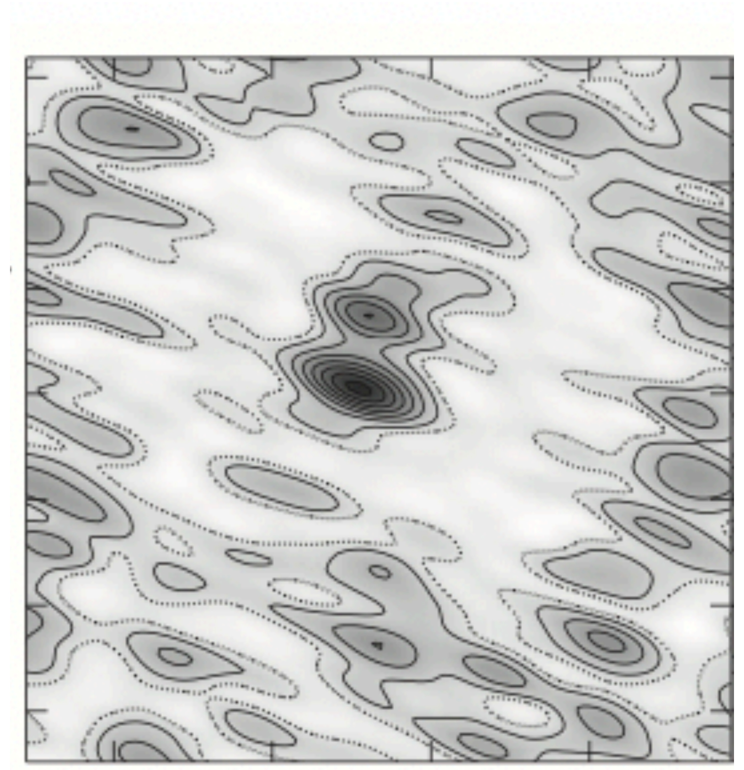


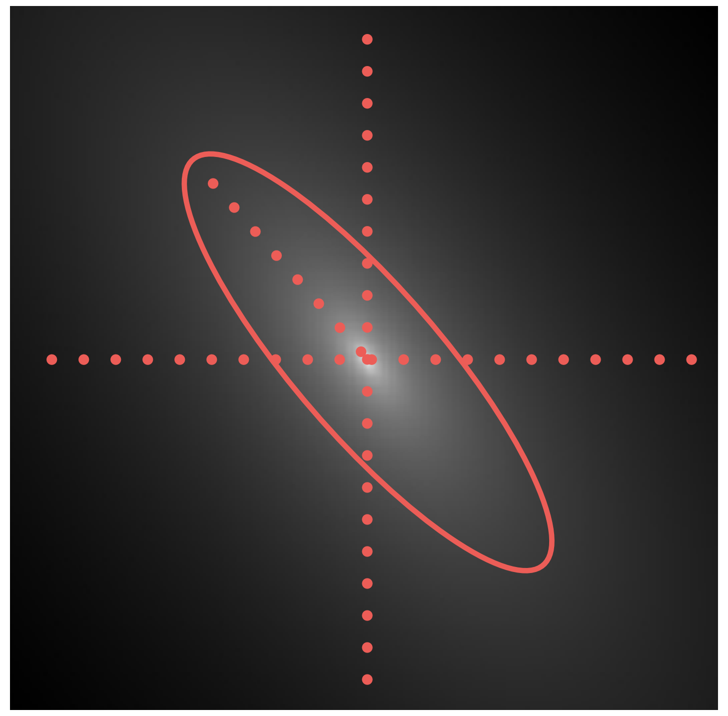



Lensing galaxy
Two images of a background quasar

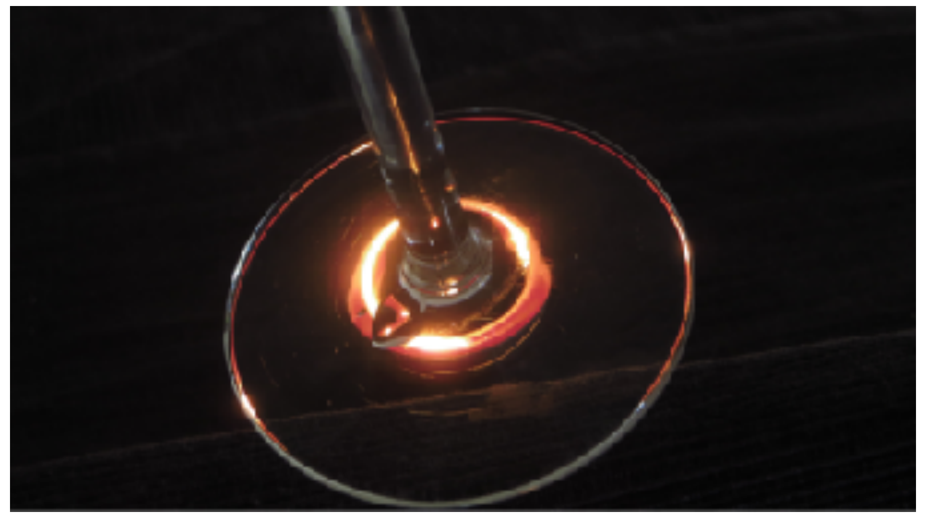
SIMPLE MODELS CAN BE SUFFICIENT FOR SIMPLE DATA

Simple/low resolution/low SNR data -> simple model might be sufficient



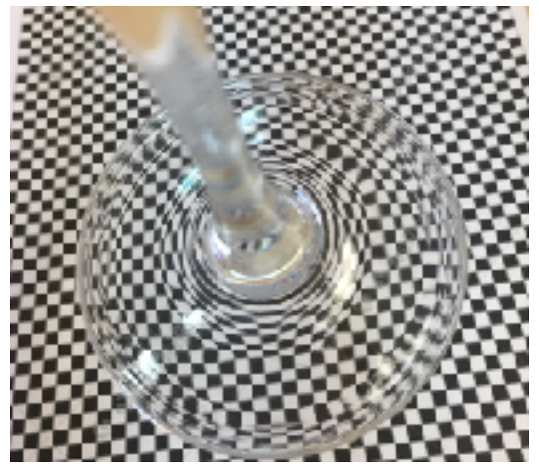


A simple example of Lens modeling



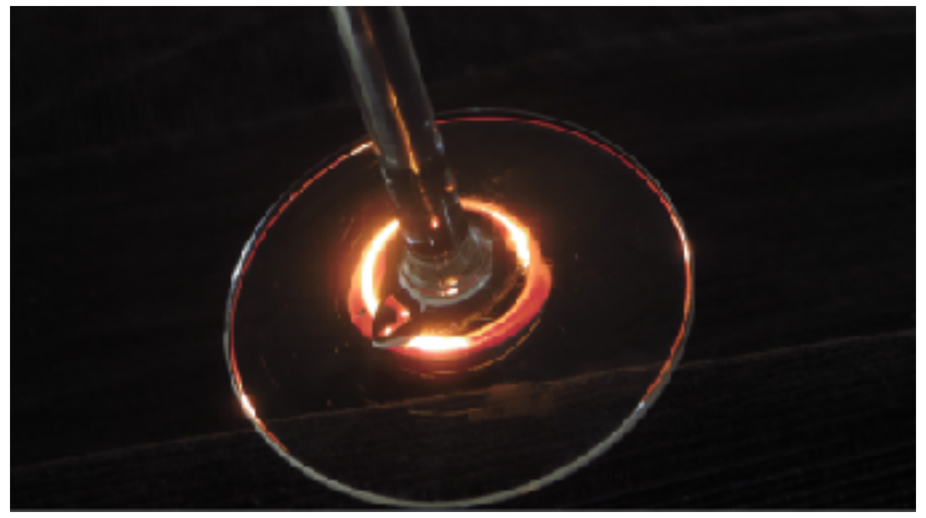
Data

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



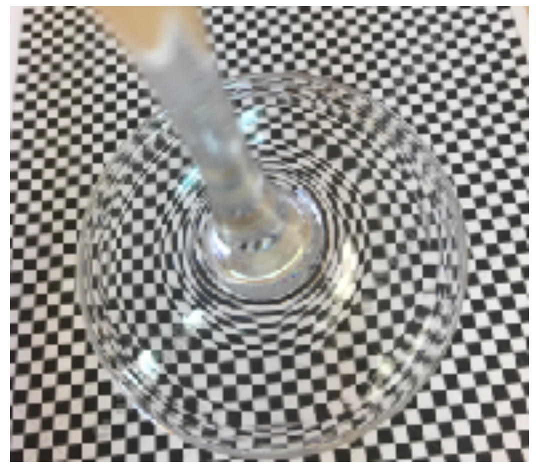
2: Matter distribution in the lens
(the shape of the wineglass)

A simple example of Lens modeling



Data

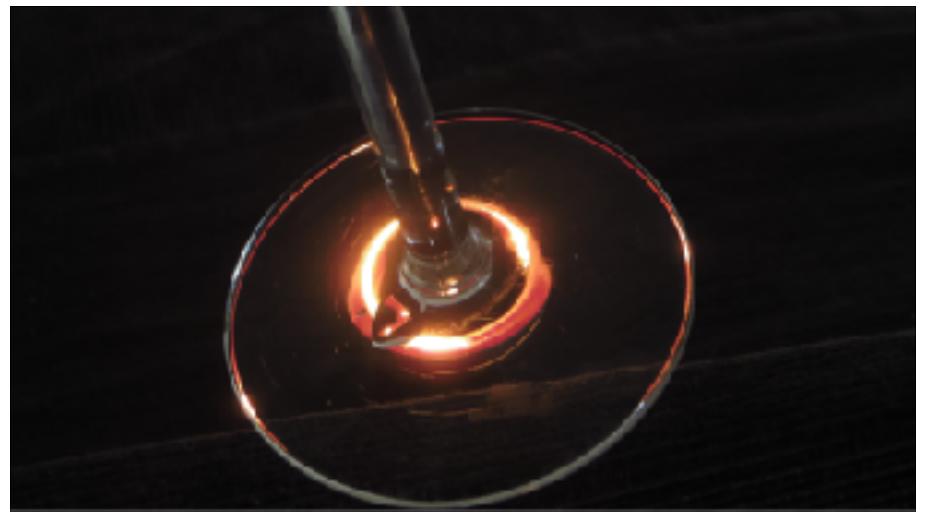
1: Morphology of the background source
(the true, undistorted image of the candle)



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

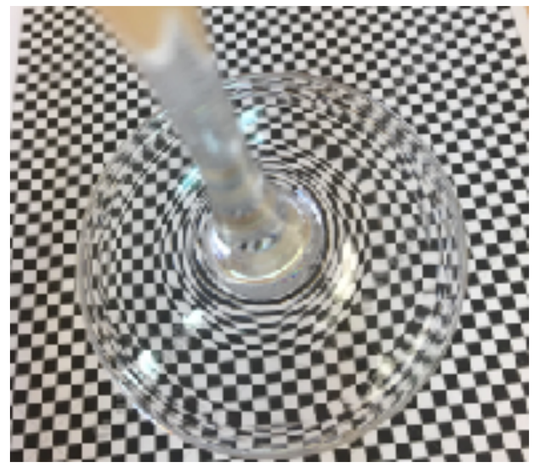
$$y = F(l, s) + n$$

A simple example of Lens modeling



Data

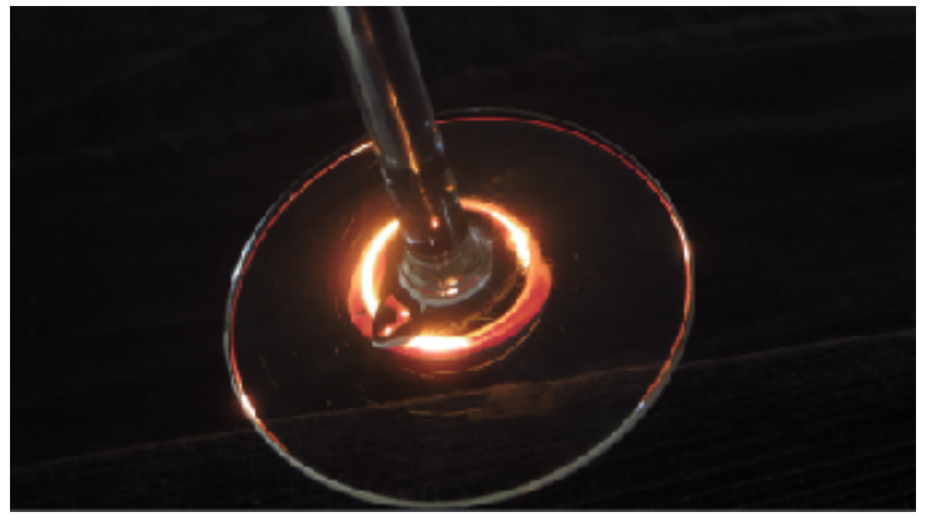
1: Morphology of the background source
(the true, undistorted image of the candle)



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

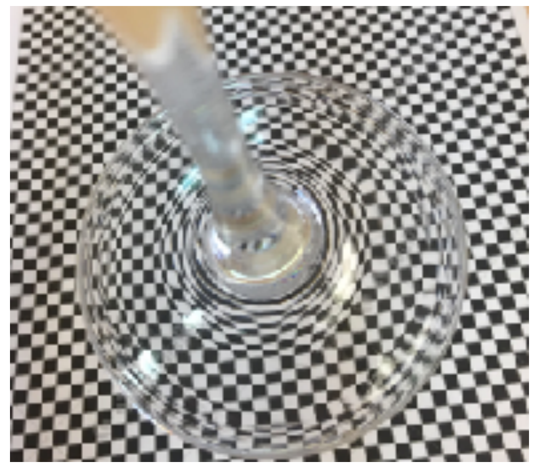
Data $\longrightarrow y = F(l, s) + n$

A simple example of Lens modeling



Data

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

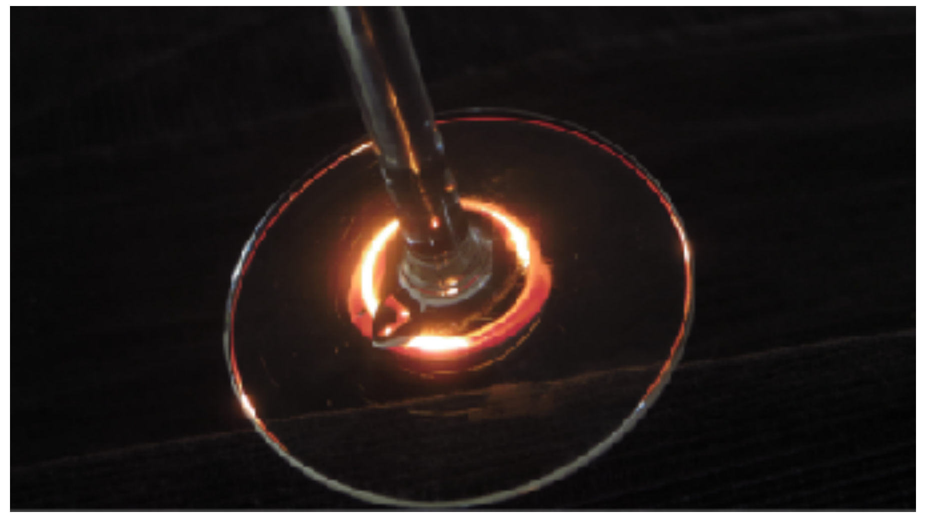


2: Matter distribution in the lens
(the shape of the wineglass)

Data → $$y = F(l, s) + n$$

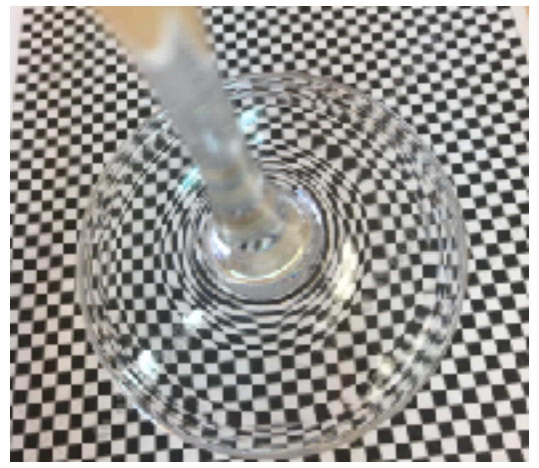
↑ Lens +Source Parameters (non-linear)

A simple example of Lens modeling



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)

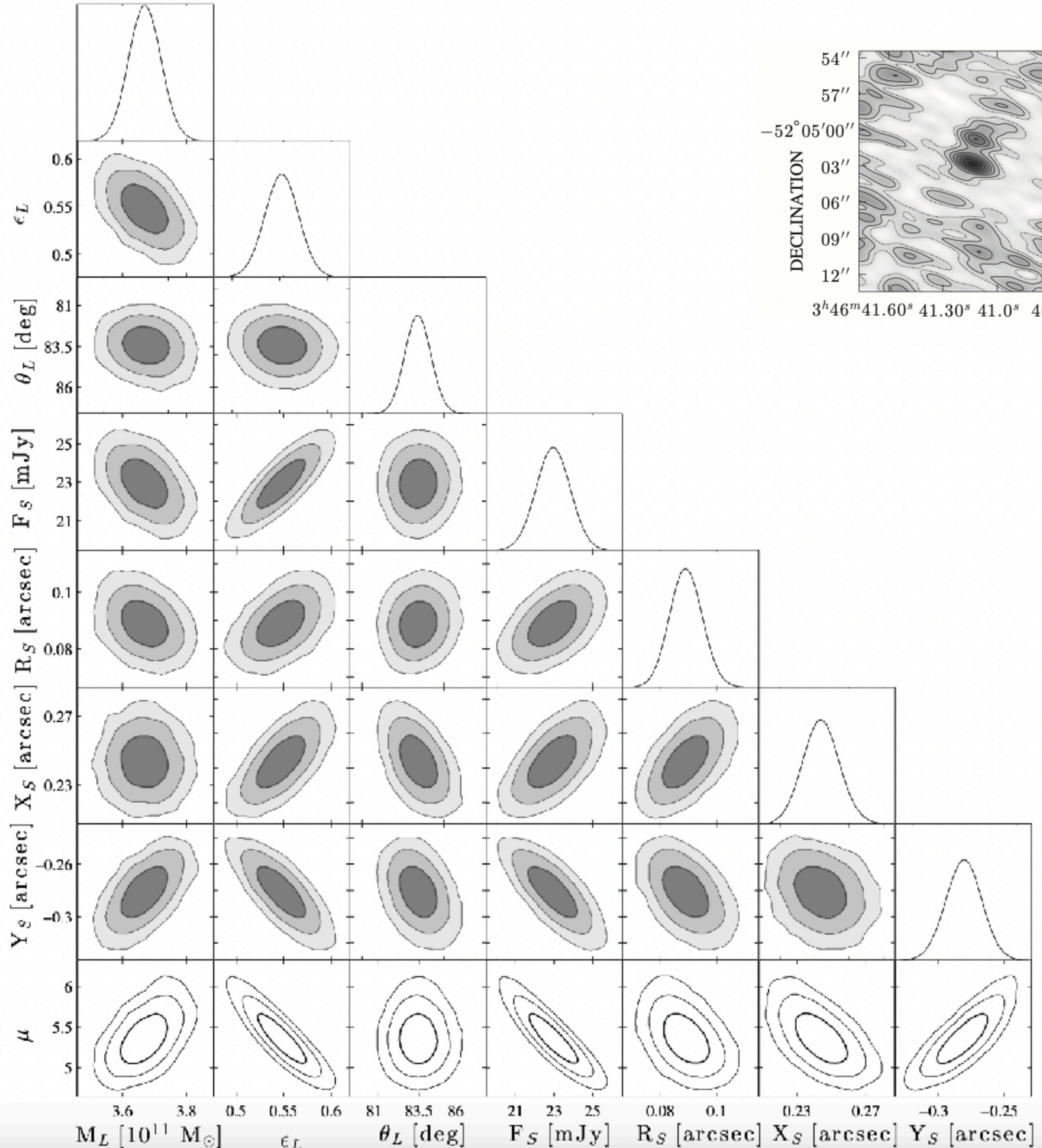
$$y = F(l, s) + n$$

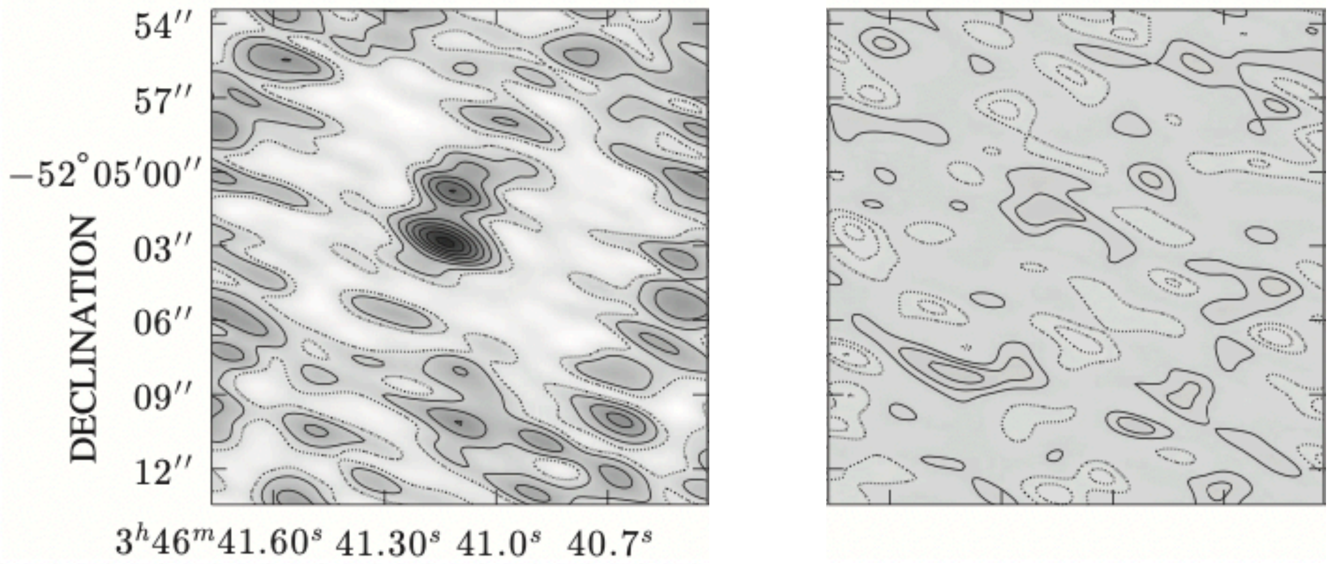
Data → y

Lens +Source Parameters (non-linear) → l, s

Noise → n

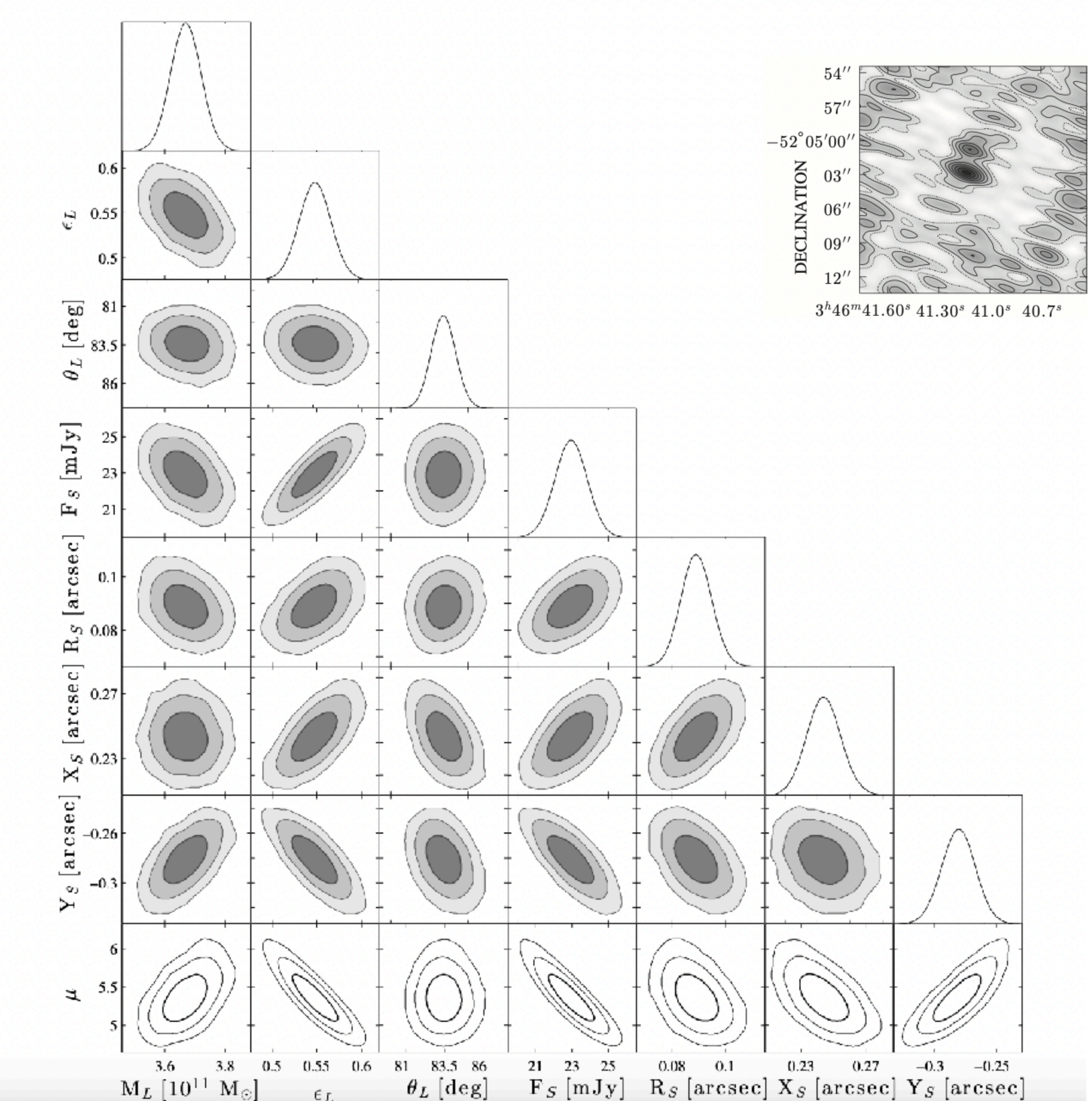
Parametric Background Source Reconstruction

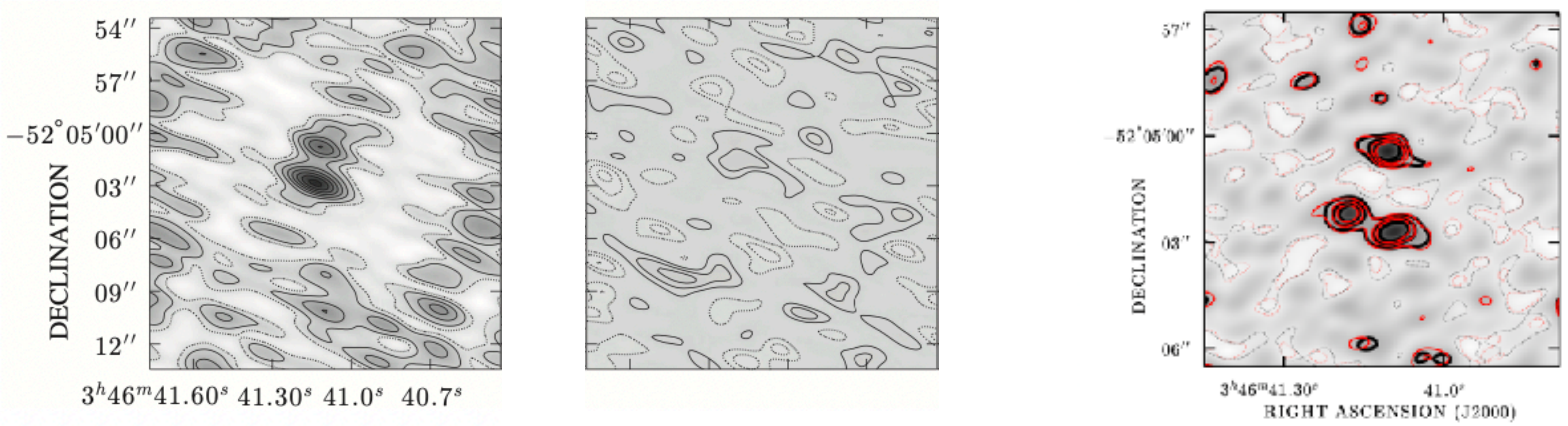




Hezaveh et al., ApJ, 2013

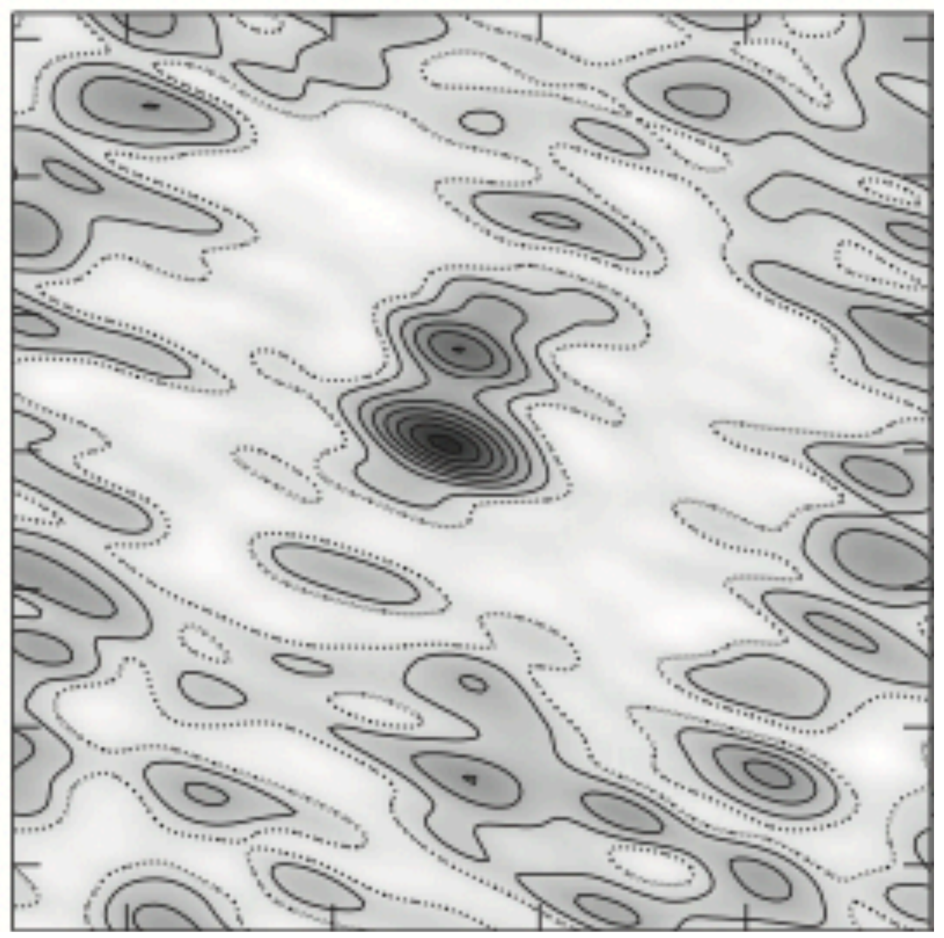
PARAMETRIC BACKGROUND SOURCE RECONSTRUCTION

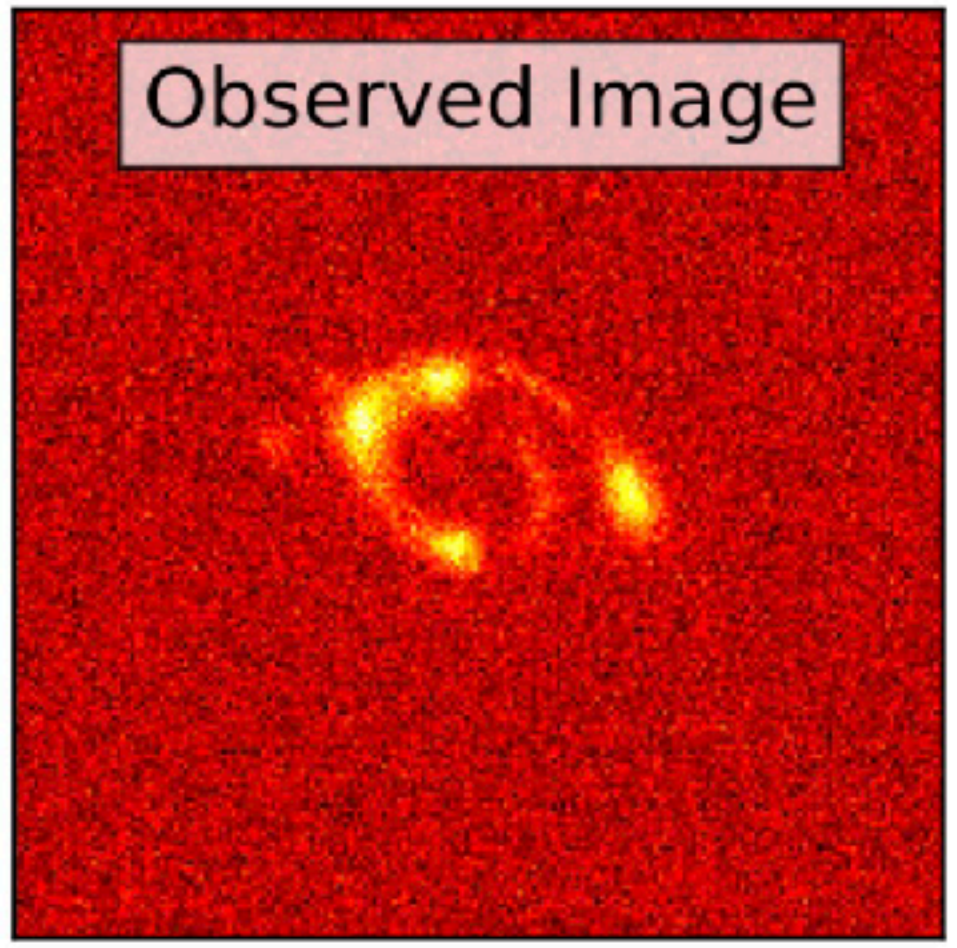




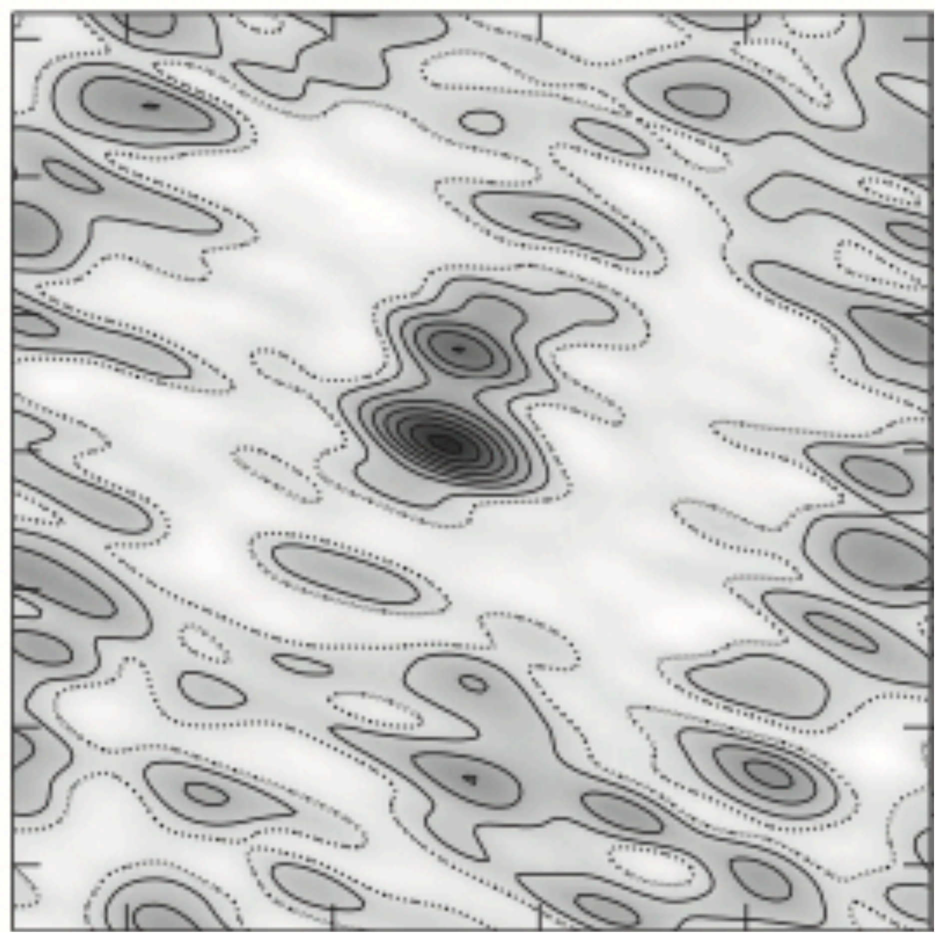
Hezaveh et al., ApJ, 2013

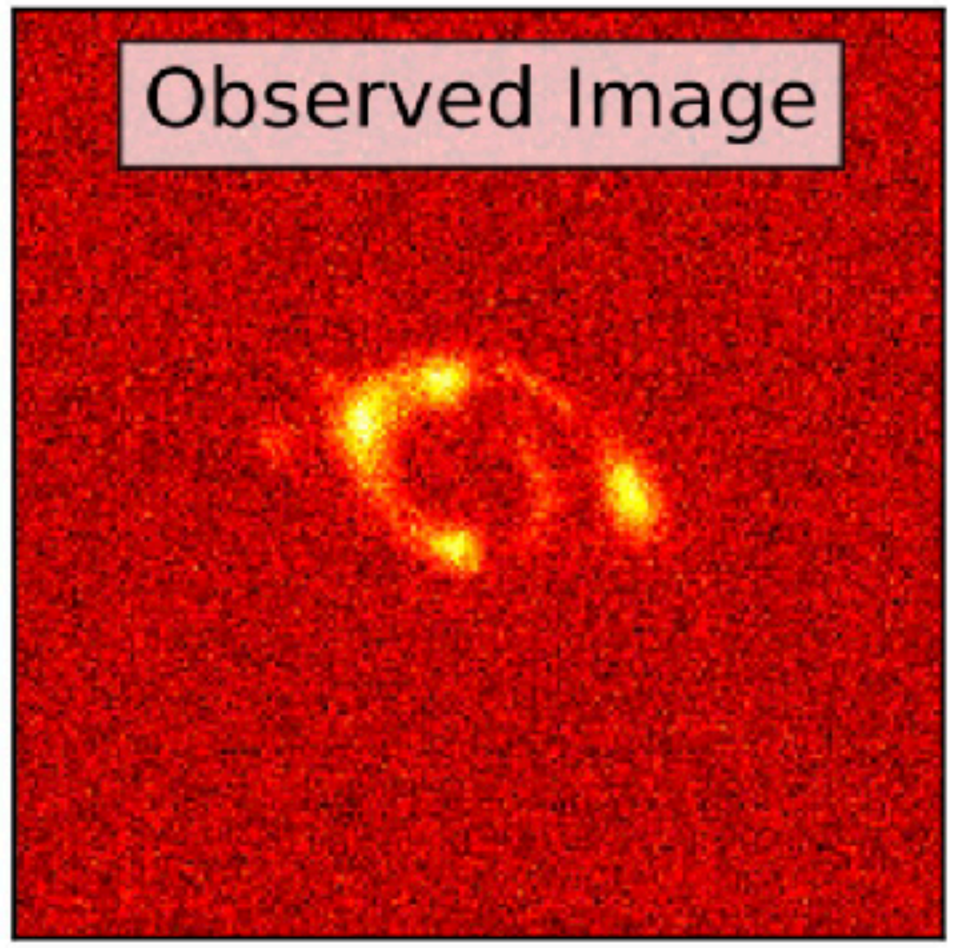
More Complex Data Demands a More Complex Parametrization





More Complex Data Demands a More Complex Parametrization





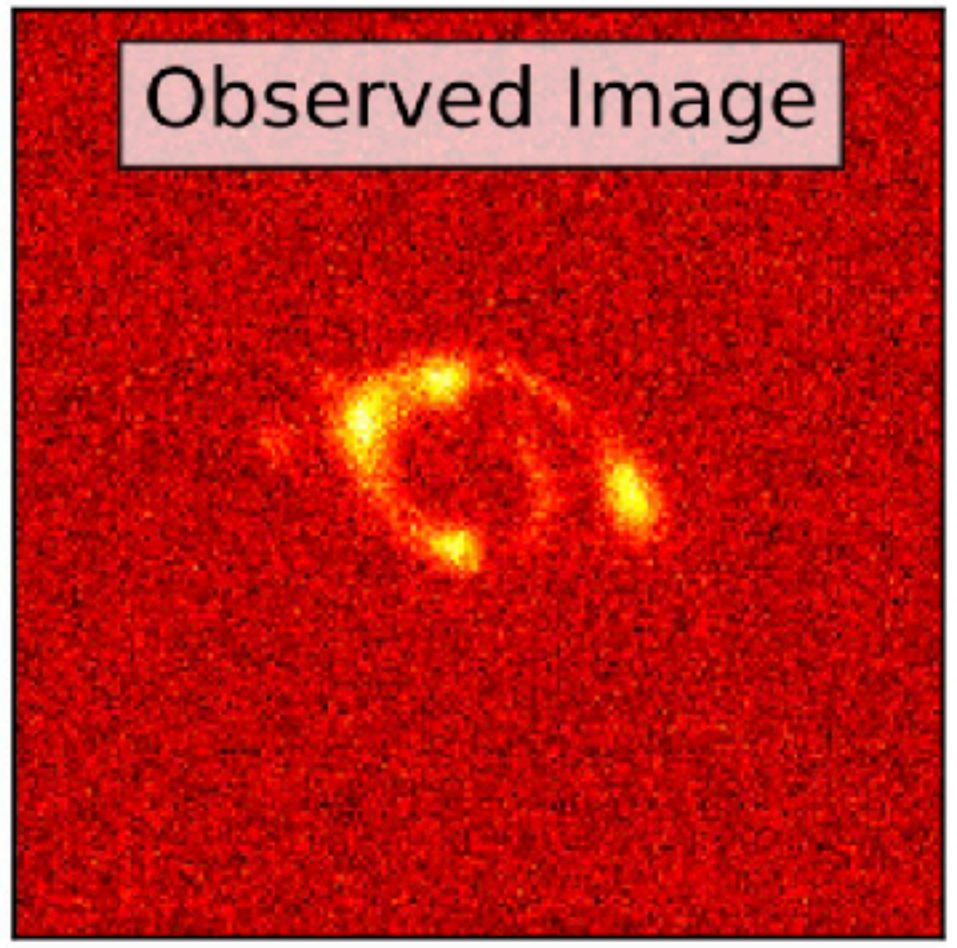
Source Parameters (linear)

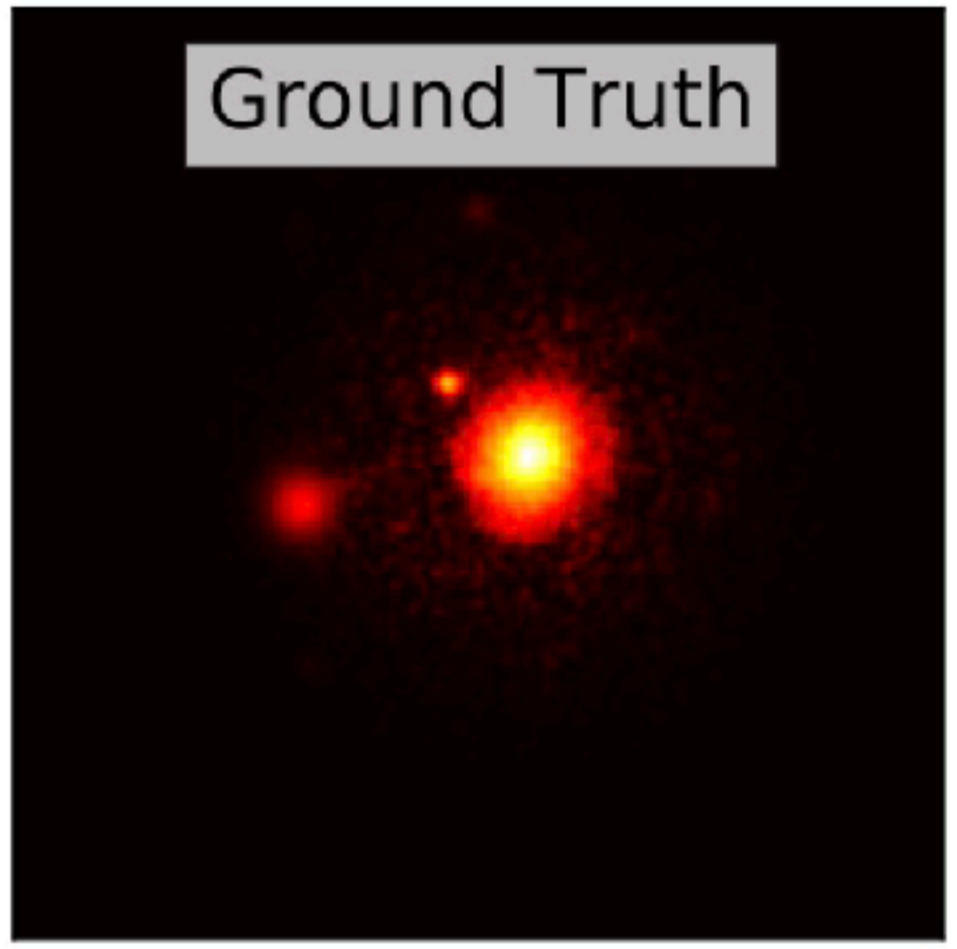
Data $\longrightarrow$ $$y = L(p)x + n$$

Lens Parameters (non-linear)

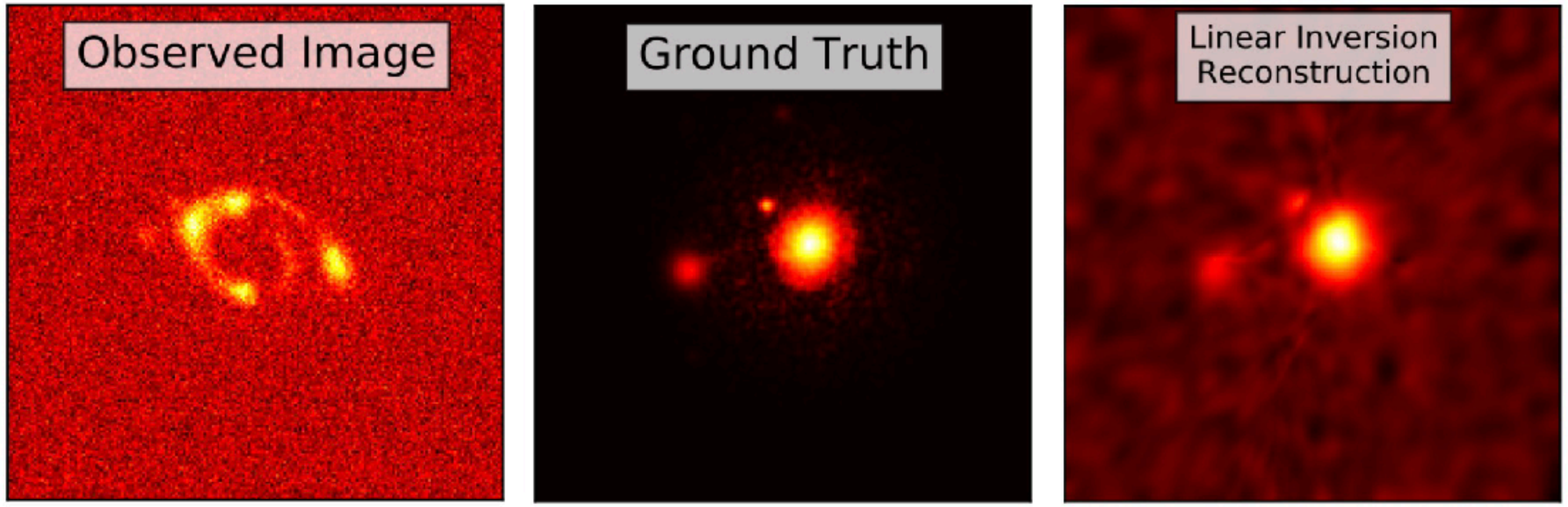
Noise

Pixelated Background Source Reconstruction

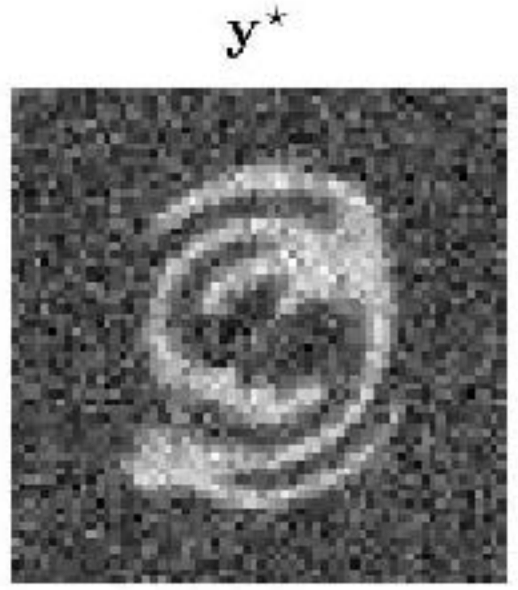




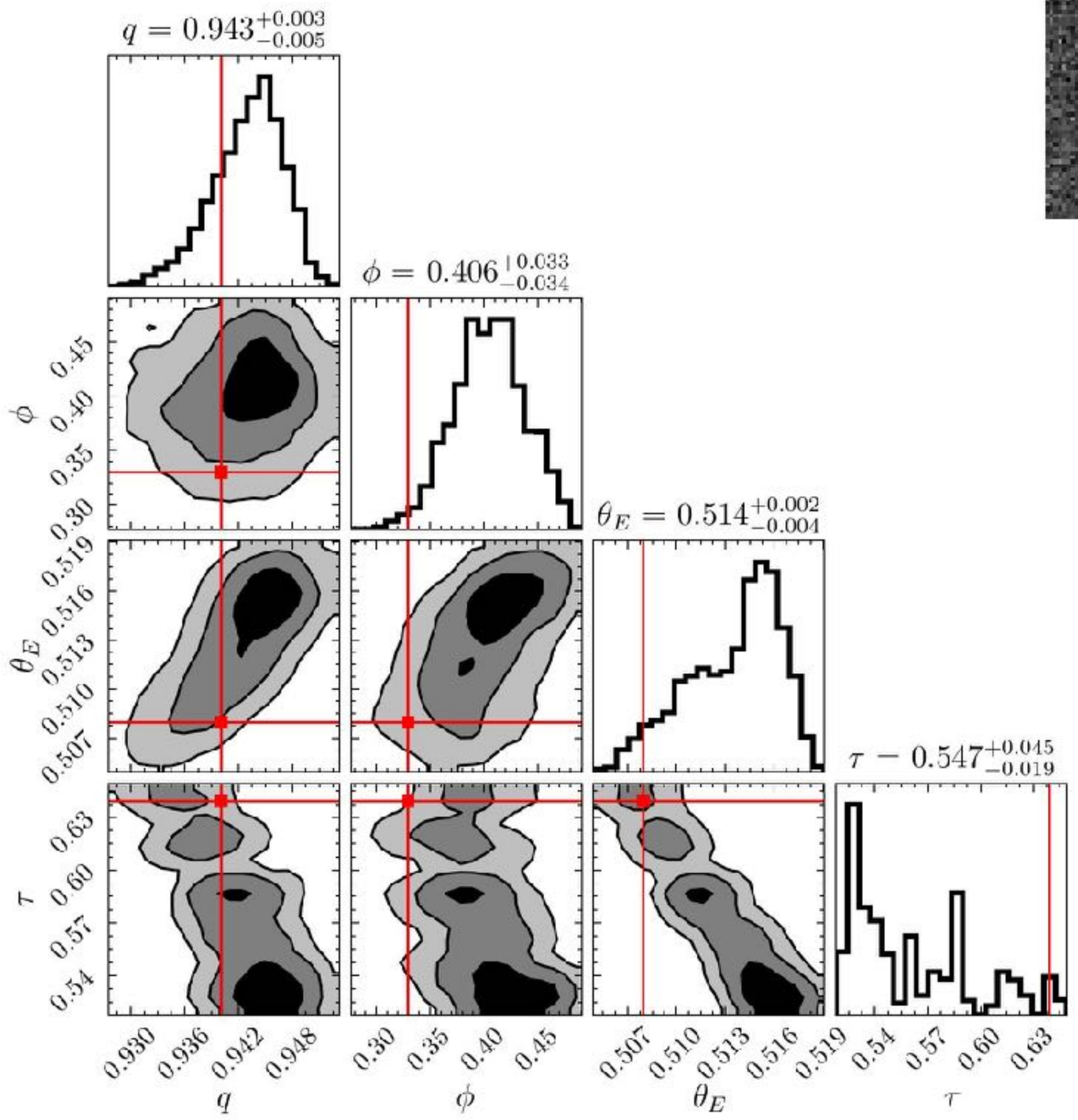
PIXELATED BACKGROUND SOURCE RECONSTRUCTION

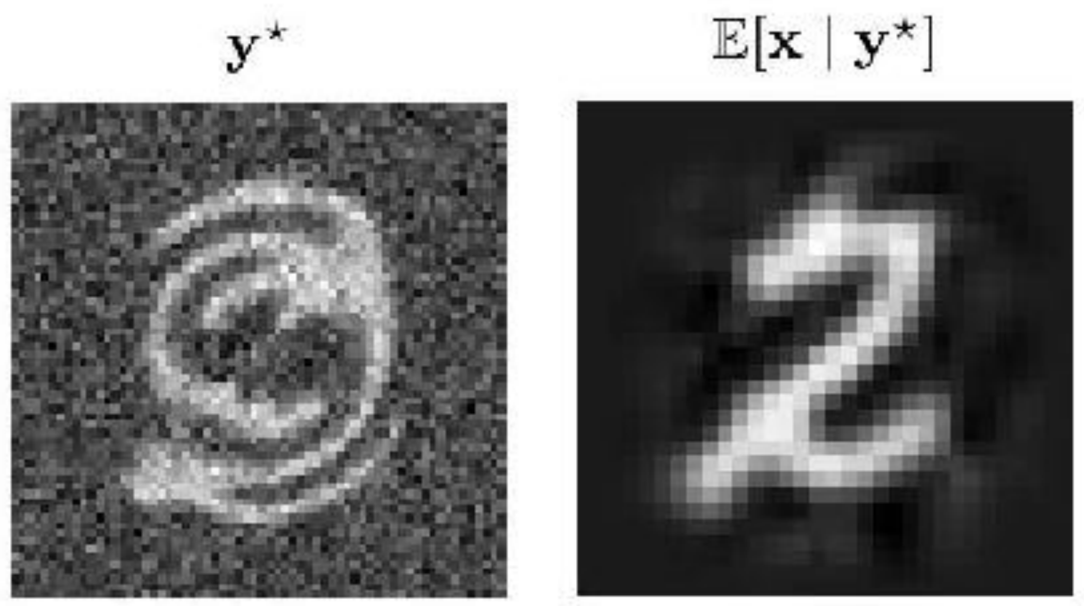


Pixelated Background Source Reconstruction

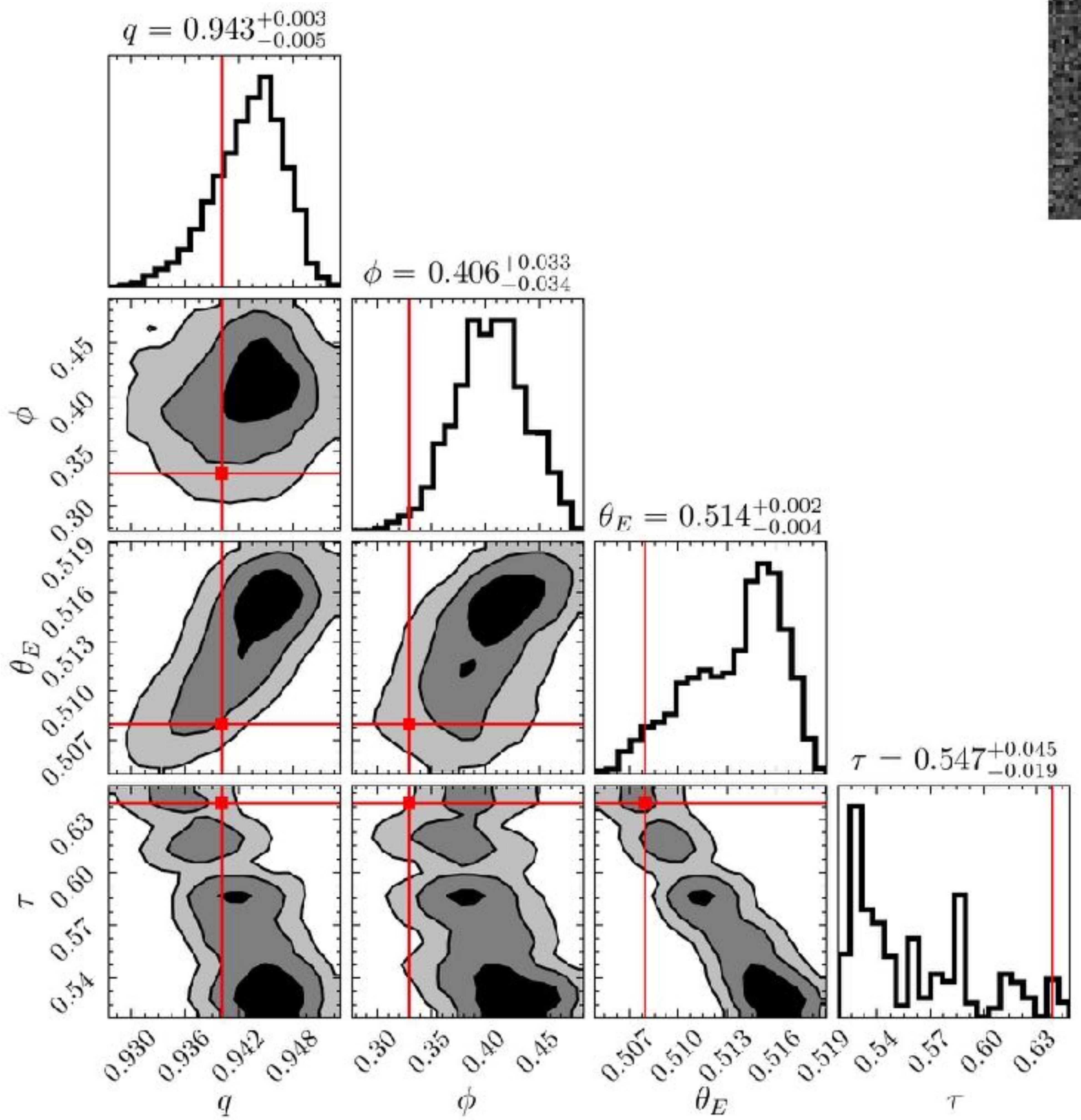


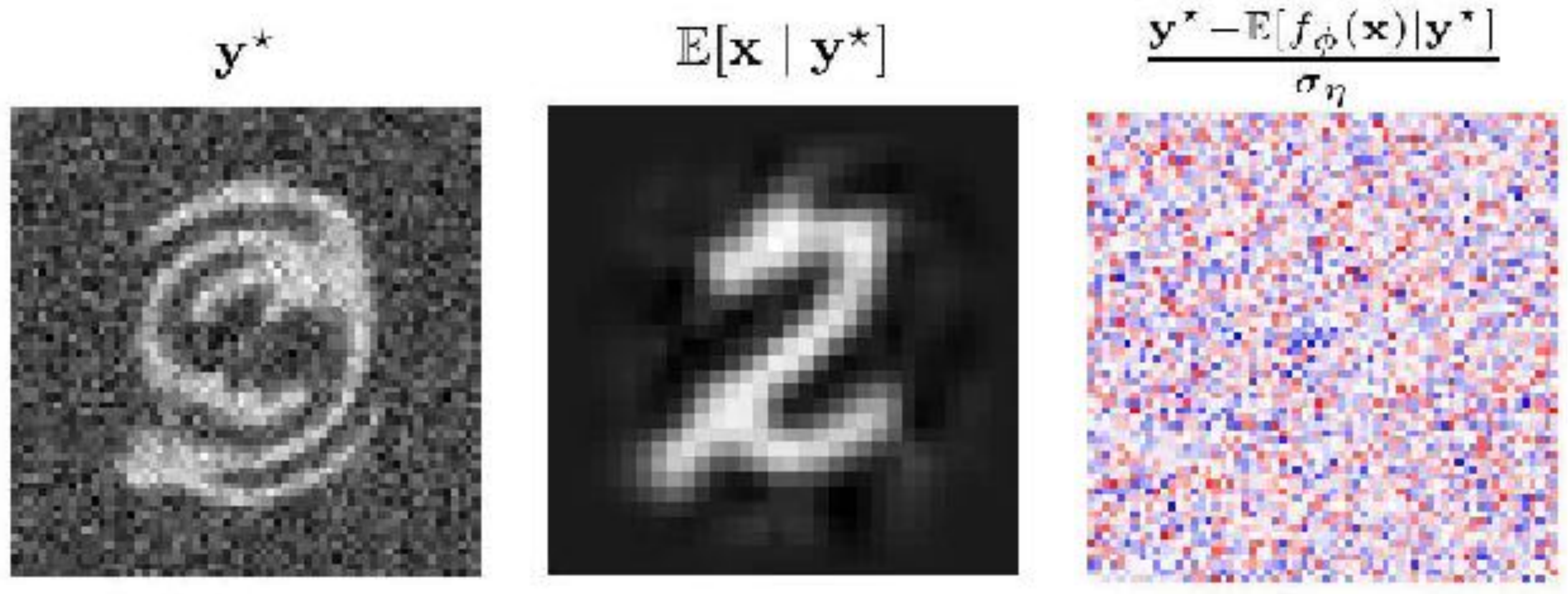
Pixelated Background Source Reconstruction



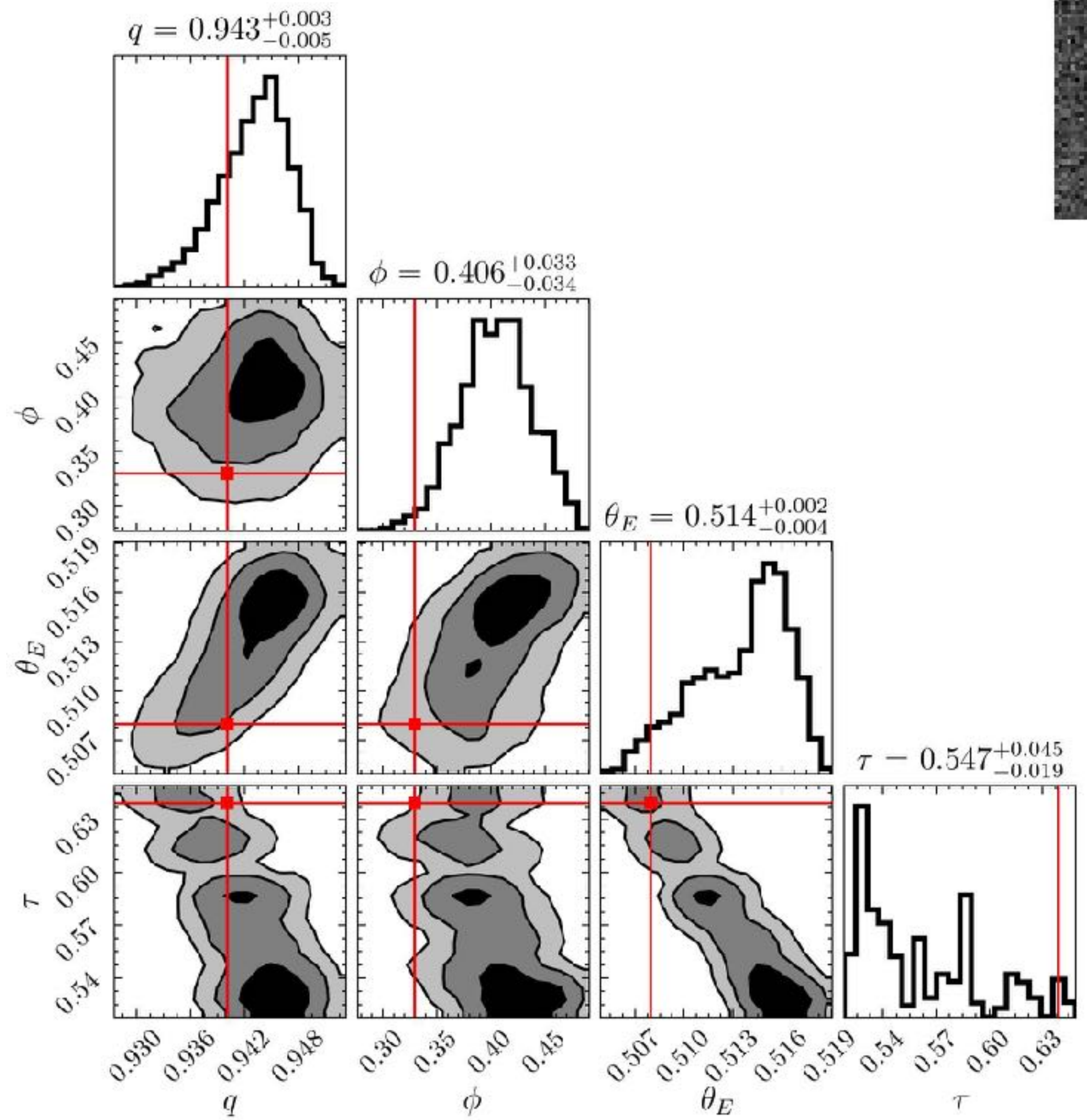


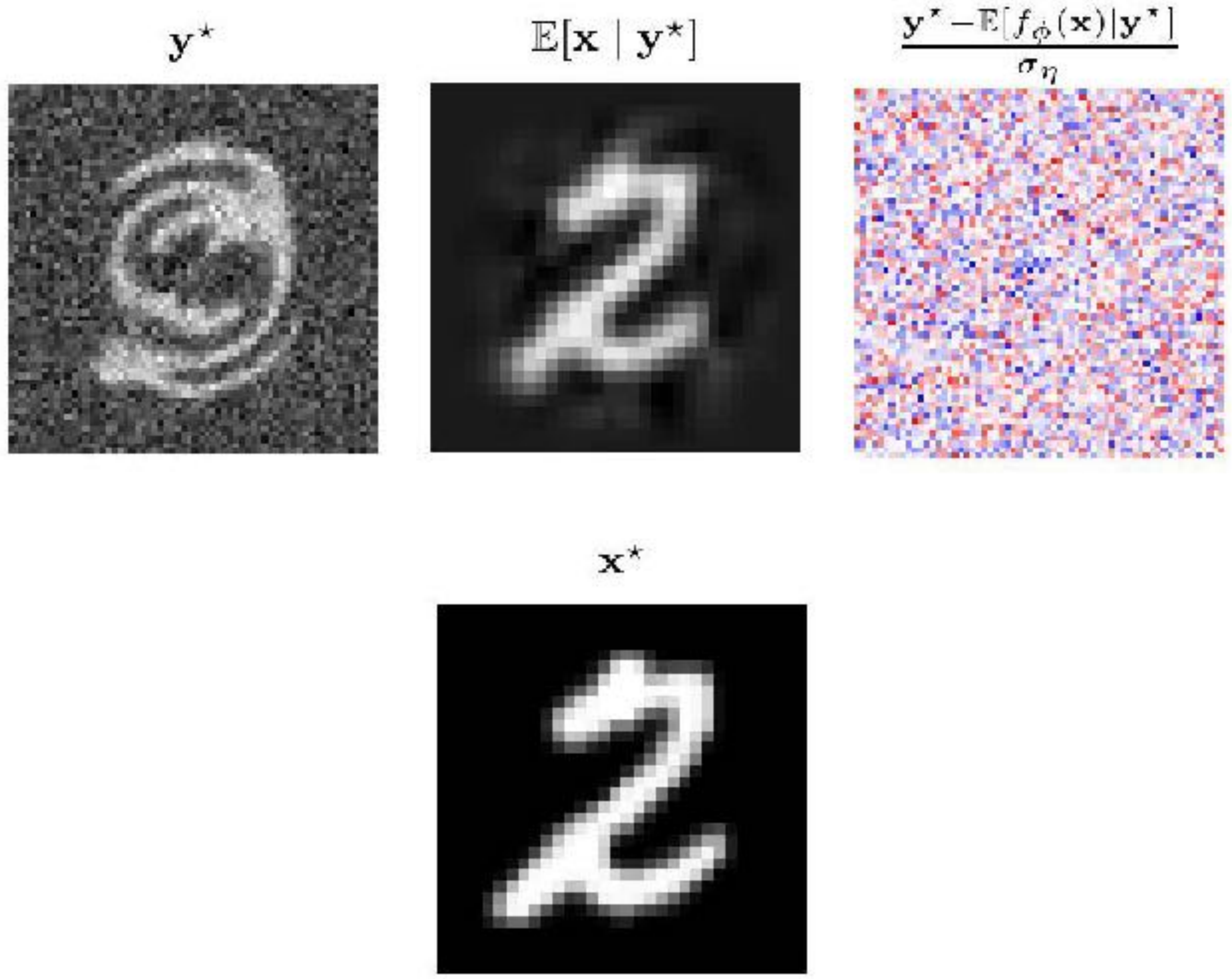
Pixelated Background Source Reconstruction





Pixelated Background Source Reconstruction





GAME PLAN

1. Data-driven priors
2. Data-driven likelihoods
3. Accuracy metrics
4. Out-of-Distribution accuracy

Learning High-Dimensional Priors with Score Modeling

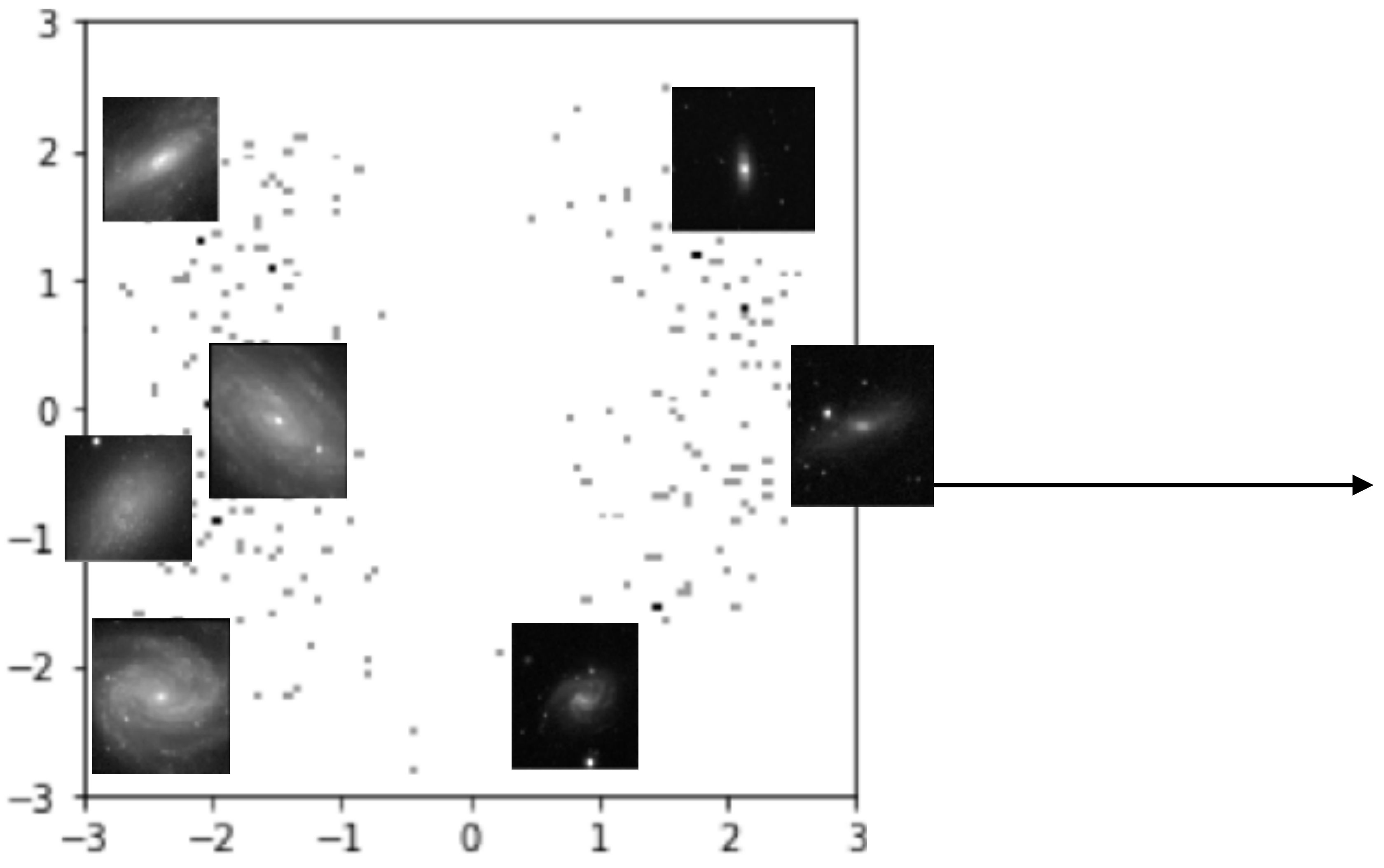
Can we learn our high-dimensional prior from data?

Learning High-Dimensional Priors with Score Modeling

Can we learn our high-dimensional prior from data?

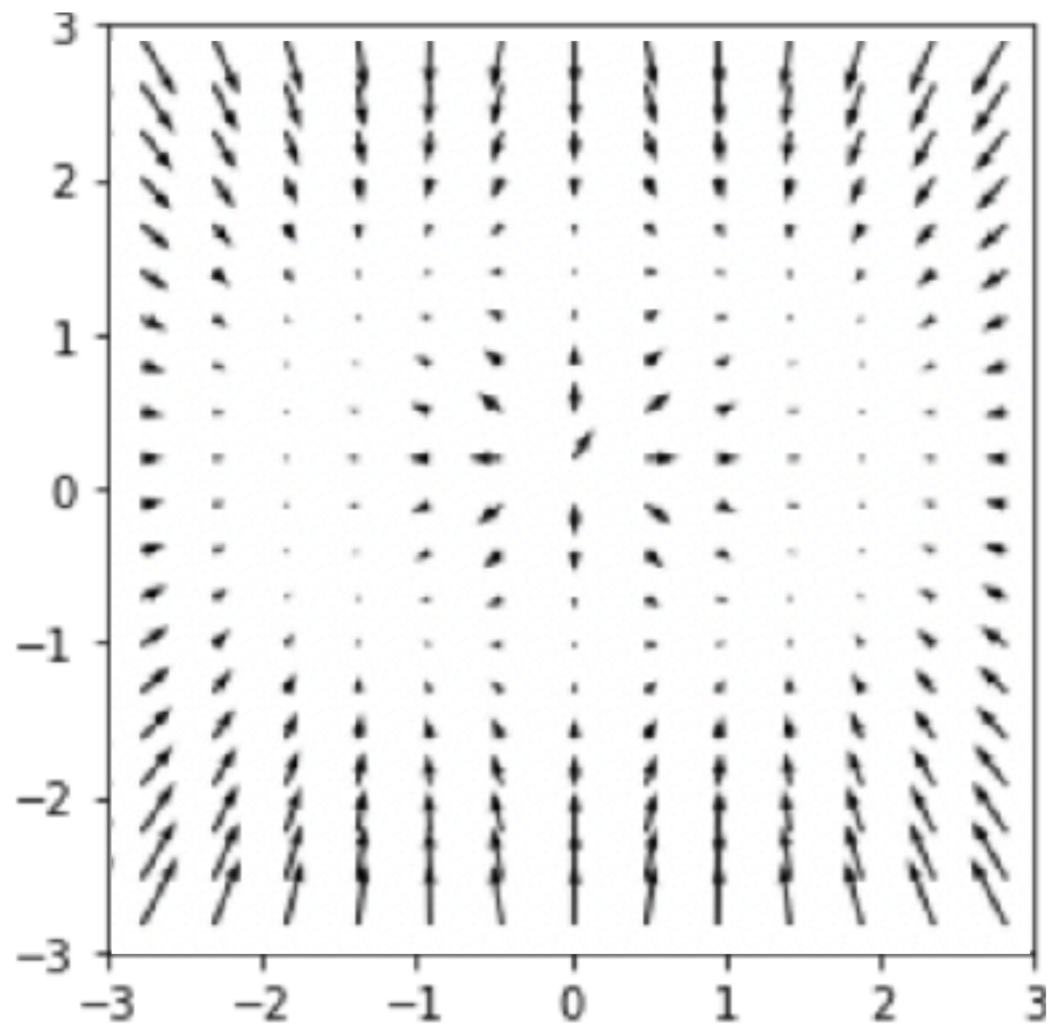
Turns out that we can sample a distribution only knowing its ***score***, which does not include the normalization constant and only uses local information.

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



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

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

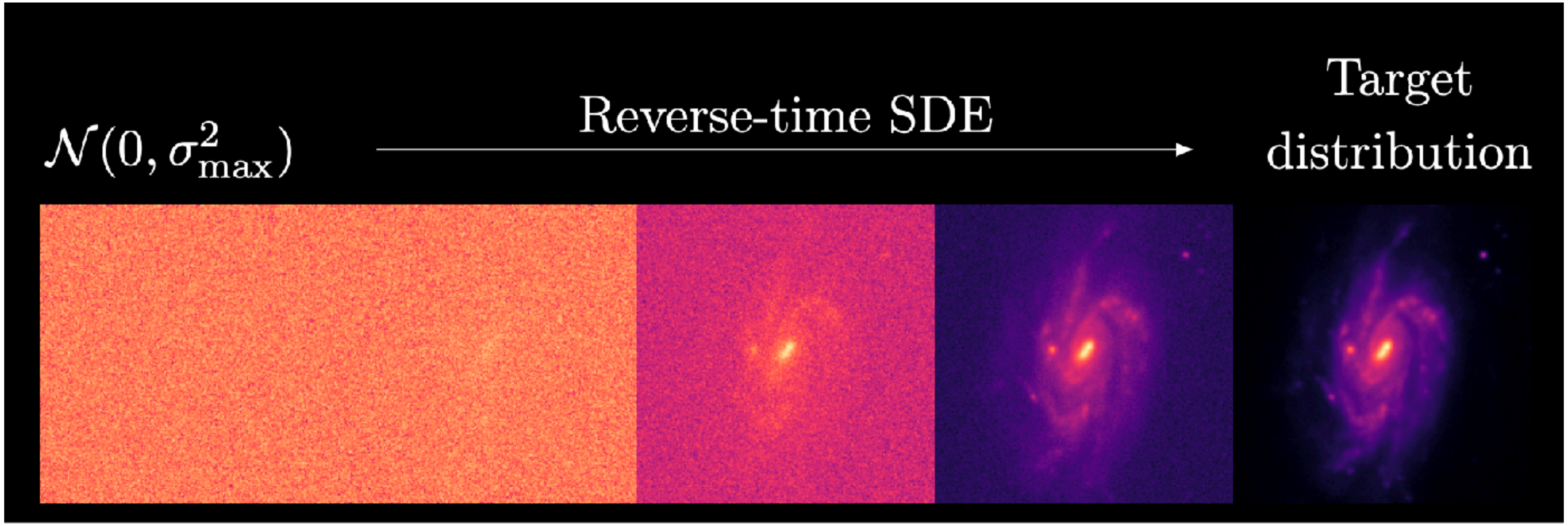
Adam et al. NeurIPS 2022 ML4PS workshop, arXiv:2211.03812

Score-Based Modeling

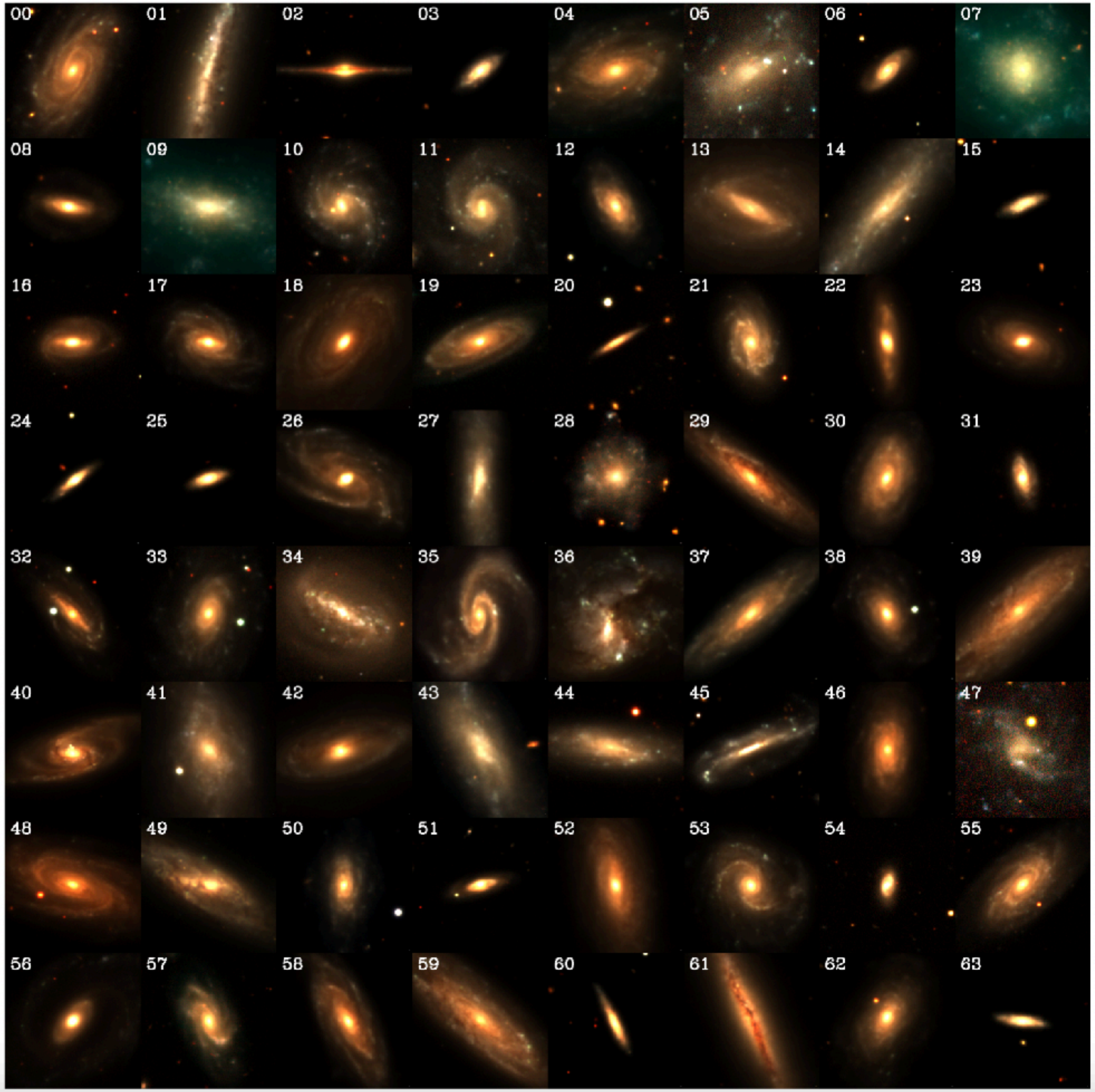
Alexandre Adam

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

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

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

Score-based Modeling

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

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

Score-based Modeling

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

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

This is the prior score we learnt from the training data

Score-based Modeling

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

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

This is intractable!

$$p_t(\mathbf{y} \mid \mathbf{x}_t) = \int_{\mathbf{x}_0} p(\mathbf{y} \mid \mathbf{x}_0) p(\mathbf{x}_0 \mid \mathbf{x}_t) \mathrm{d}\mathbf{x}_0$$

This is the prior score we learnt from the training data

SCORE-BASED MODELING

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

$$\nabla_x \log p(x|y) = \nabla_x \underline{\log p(y|x)} + \nabla_x \underline{\log p_\theta(x)}$$

This is intractable!

$$p_t(\mathbf{y} \mid \mathbf{x}_t) = \int_{\mathbf{x}_0} p(\mathbf{y} \mid \mathbf{x}_0) p(\mathbf{x}_0 \mid \mathbf{x}_t) \mathrm{d}\mathbf{x}_0$$

This is the prior score we learnt from the training data

But!

Many approximations have been proposed:

SCORE-BASED MODELING

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

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

This is intractable!

$$p_t(\mathbf{y} \mid \mathbf{x}_t) = \int_{\mathbf{x}_0} p(\mathbf{y} \mid \mathbf{x}_0) p(\mathbf{x}_0 \mid \mathbf{x}_t) \mathrm{d}\mathbf{x}_0$$

This is the prior score we learnt from the training data

But!

Many approximations have been proposed:

Guidance-Based methods

Using Tweedie formula or other approximations to the convolved likelihood

e.g : Song et al., 2023; Wang et al., 2022; Kawar et al., 2022; Rout et al., 2023; Chung et al., 2023, Adam et al. 2023

Score-based Modeling

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

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

This is intractable!

$$p_t(\mathbf{y} \mid \mathbf{x}_t) = \int_{\mathbf{x}_0} p(\mathbf{y} \mid \mathbf{x}_0) p(\mathbf{x}_0 \mid \mathbf{x}_t) \mathrm{d}\mathbf{x}_0$$

This is the prior score we learnt from the training data

But!
Many approximations have been proposed:

Guidance-Based methods
Using Tweedie formula or other approximations to the convolved likelihood

e.g : Song et al., 2023; Wang et al., 2022; Kawar et al., 2022; Rout et al., 2023; Chung et al., 2023, Adam et al. 2023

Variational Inference
Training a surrogate posterior close to the true one in a KL sense, the DM is a prior in the VI

e.g : Feng et al 2023; Feng & Bouman, 2024

Score-based Modeling

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

$$\nabla_x \log p(x|y) = \nabla_x \underline{\log p(y|x)} + \nabla_x \underline{\log p_\theta(x)}$$

This is intractable!

$$p_t(\mathbf{y} \mid \mathbf{x}_t) = \int_{\mathbf{x}_0} p(\mathbf{y} \mid \mathbf{x}_0) p(\mathbf{x}_0 \mid \mathbf{x}_t) \mathrm{d}\mathbf{x}_0$$

This is the prior score we learnt from the training data

But!

Many approximations have been proposed:

Guidance-Based methods

Using Tweedie formula or other approximations to the convolved likelihood

e.g : Song et al., 2023; Wang et al., 2022; Kawar et al., 2022; Rout et al., 2023; Chung et al., 2023, Adam et al. 2023

Variable splitting

Alternating between a likelihood evaluation step and short prior diffusion step

e.g :Wu et al., 2024 Zhu et al., 2023; Li et al., 2024

Variational Inference

Training a surrogate posterior close to the true one in a KL sense, the DM is a prior in the VI

e.g : Feng et al 2023; Feng & Bouman, 2024

SCORE-BASED MODELING

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

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

This is intractable!

$$p_t(\mathbf{y} \mid \mathbf{x}_t) = \int_{\mathbf{x}_0} p(\mathbf{y} \mid \mathbf{x}_0) p(\mathbf{x}_0 \mid \mathbf{x}_t) \mathrm{d}\mathbf{x}_0$$

This is the prior score we learnt from the training data

But!

Many approximations have been proposed:

Guidance-Based methods
Using Tweedie formula or other approximations to the convolved likelihood

e.g : Song et al., 2023; Wang et al., 2022; Kawar et al., 2022; Rout et al., 2023; Chung et al., 2023, Adam et al. 2023

Variable splitting
Alternating between a likelihood evaluation step and short prior diffusion step

e.g :Wu et al., 2024 Zhu et al., 2023; Li et al., 2024

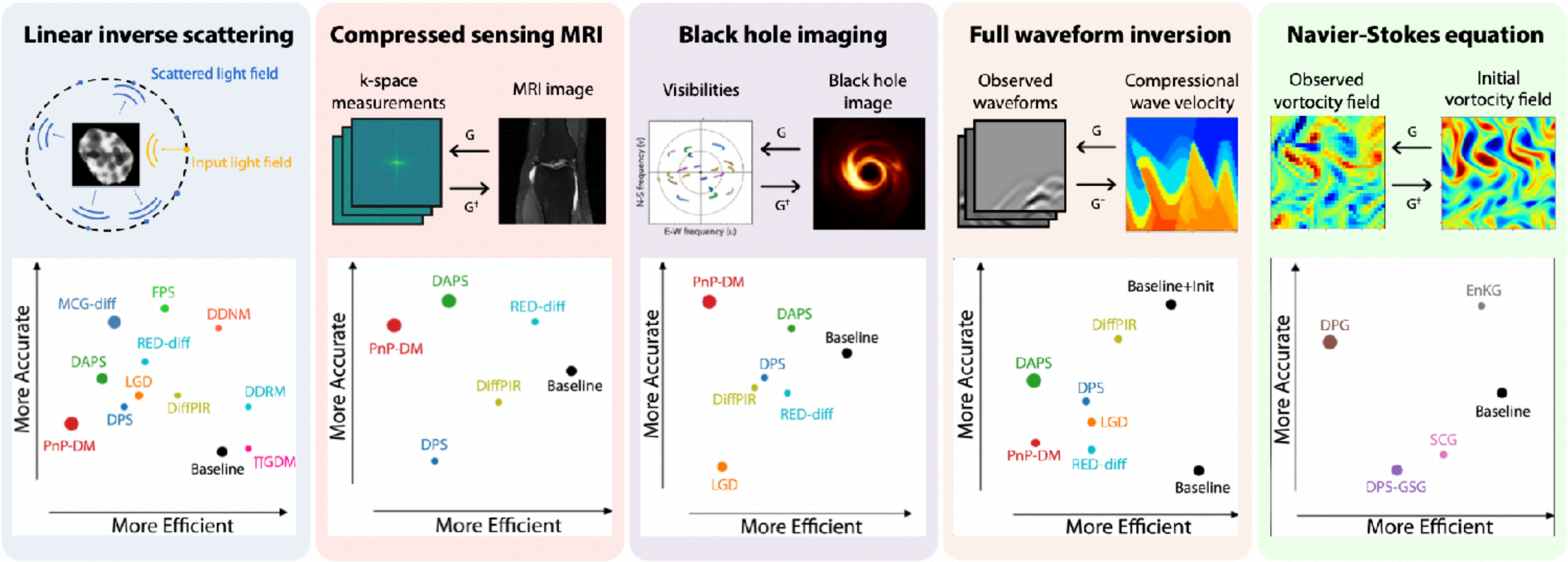
Variational Inference
Training a surrogate posterior close to the true one in a KL sense, the DM is a prior in the VI

e.g : Feng et al 2023; Feng & Bouman, 2024

Sequential MC
Get approx samples, and reweigh them to get a new proposal. Iteratively converges

e.g : Trippe et al., 2023; Dou & Song, 2024, Cardoso et al., 2024

THESE DIFFERENT METHODS PERFORM VERY DIFFERENTLY IN DIFFERENT PHYSICAL SCIENCE PROBLEMS



Zheng et al. ICLR, 2025

Score-based Modeling

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

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

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:

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

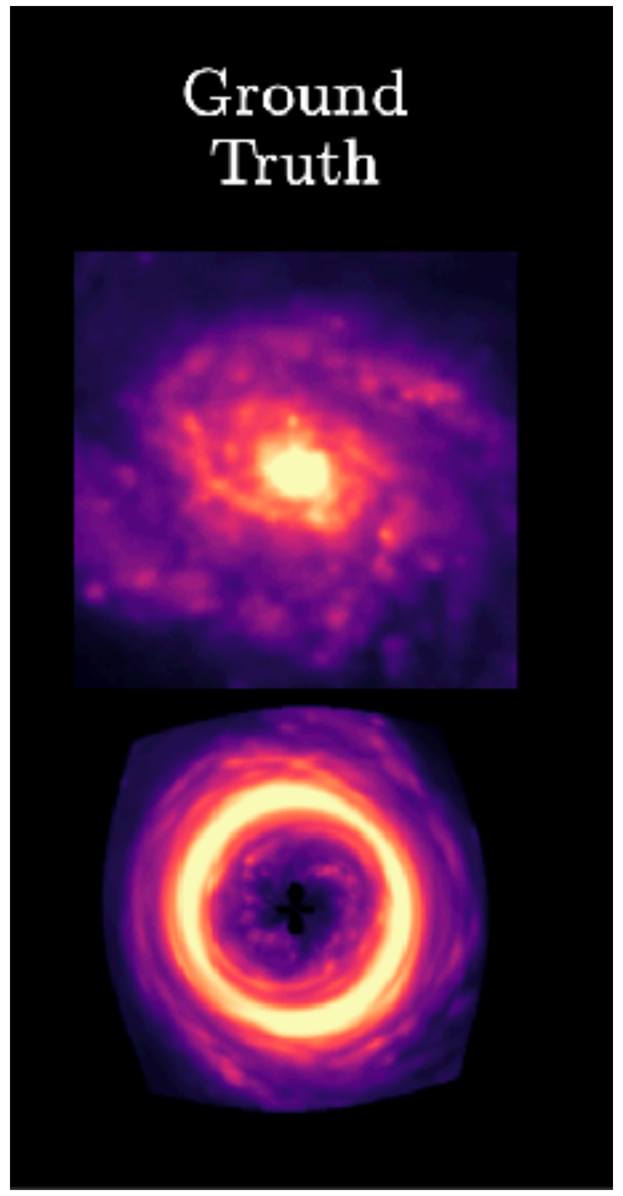
Adam et al. NeurIPS 2022 ML4PS workshop,
arXiv:2211.03812

Score-based Modeling


Alexandre Adam

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

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



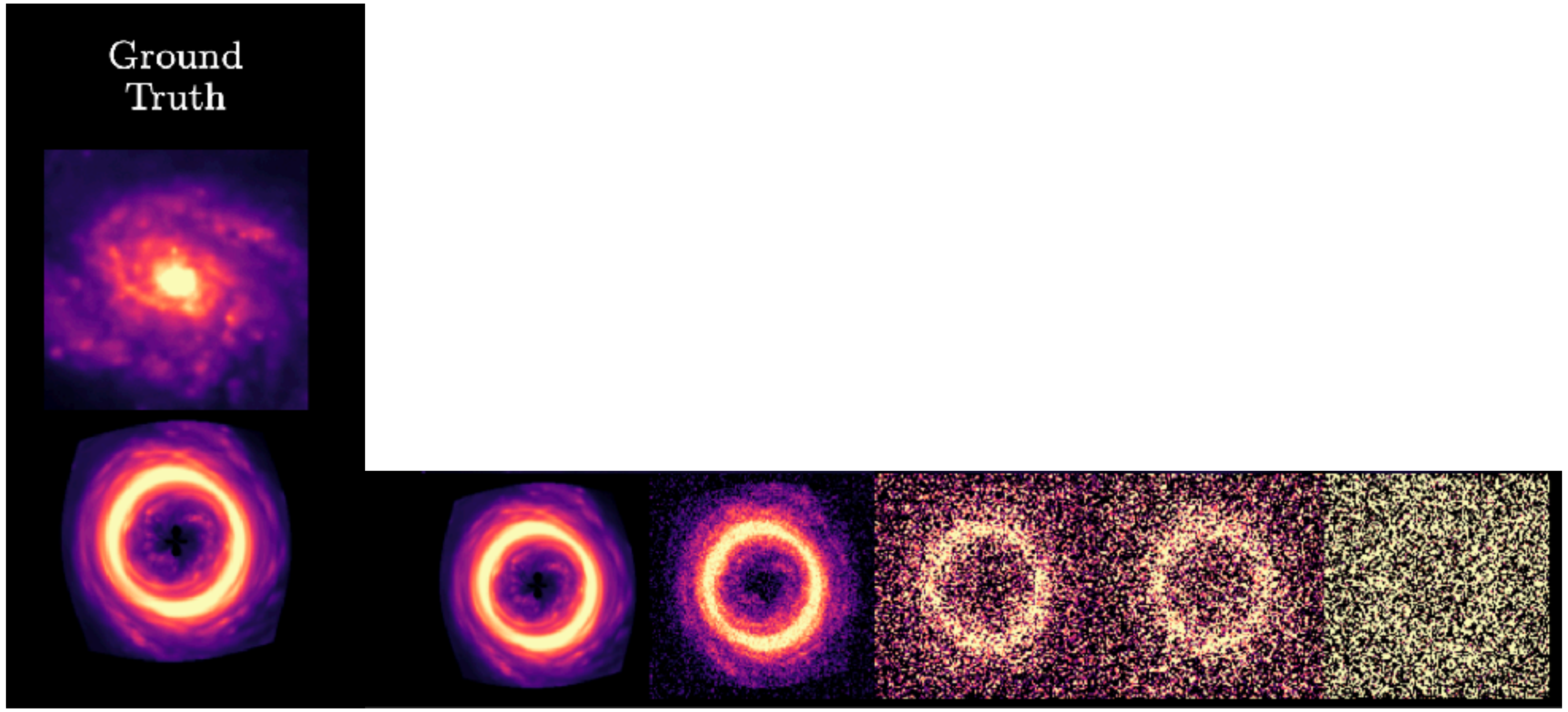
Adam et al. NeurIPS 2022 ML4PS workshop,
arXiv:2211.03812

Score-based Modeling

Alexandre Adam

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

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



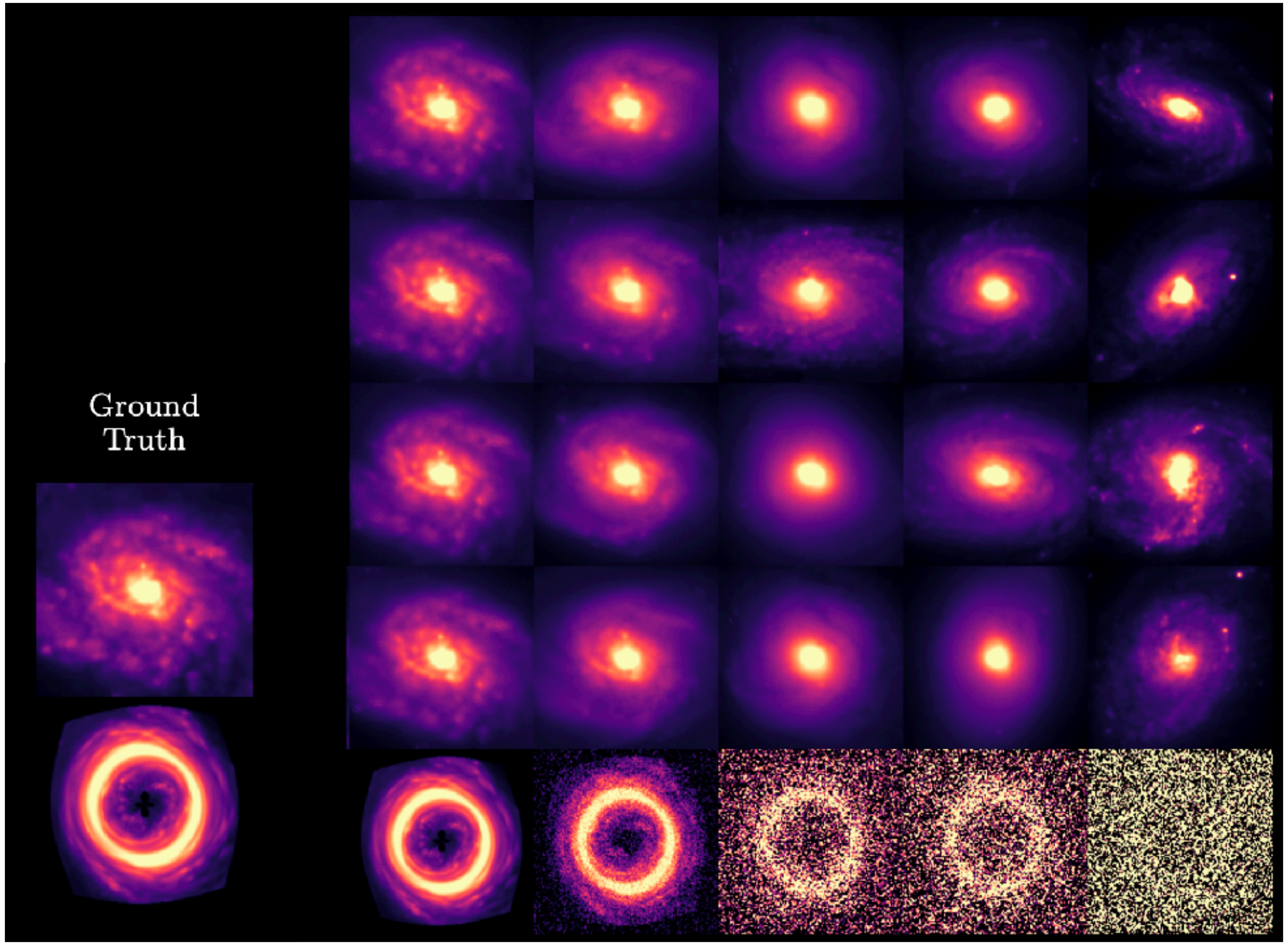
Adam et al. NeurIPS 2022 ML4PS workshop,
arXiv:2211.03812

Score-based Modeling

Alexandre Adam

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

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



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

Out of Distribution Tests

Alexandre Adam

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

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

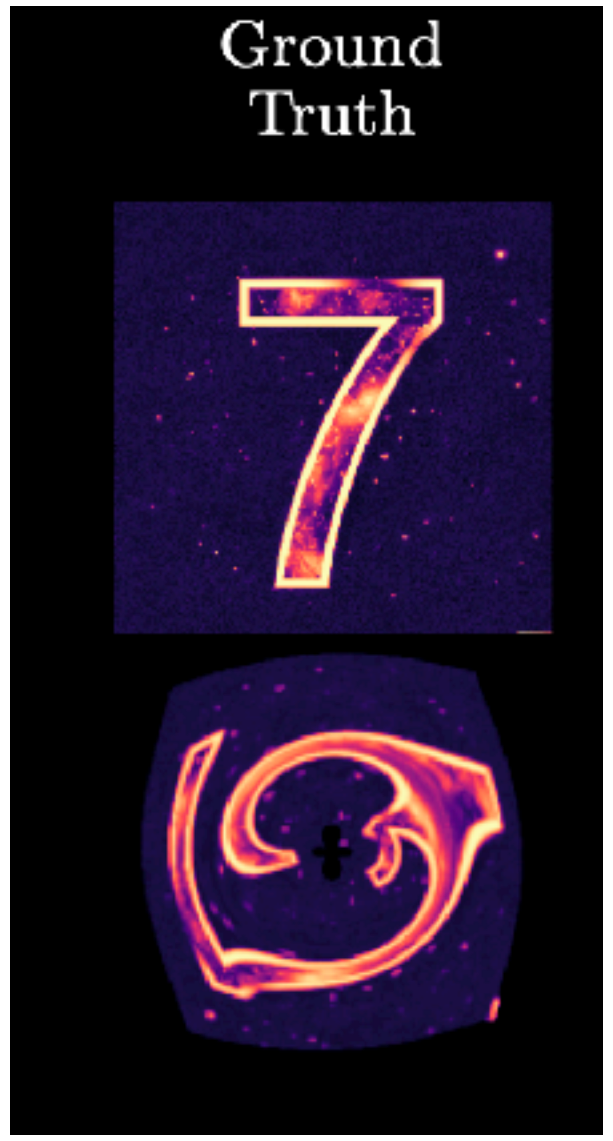
Adam et al. NeurIPS 2022 ML4PS workshop,
arXiv:2211.03812

OUT OF DISTRIBUTION TESTS

Alexandre Adam

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

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



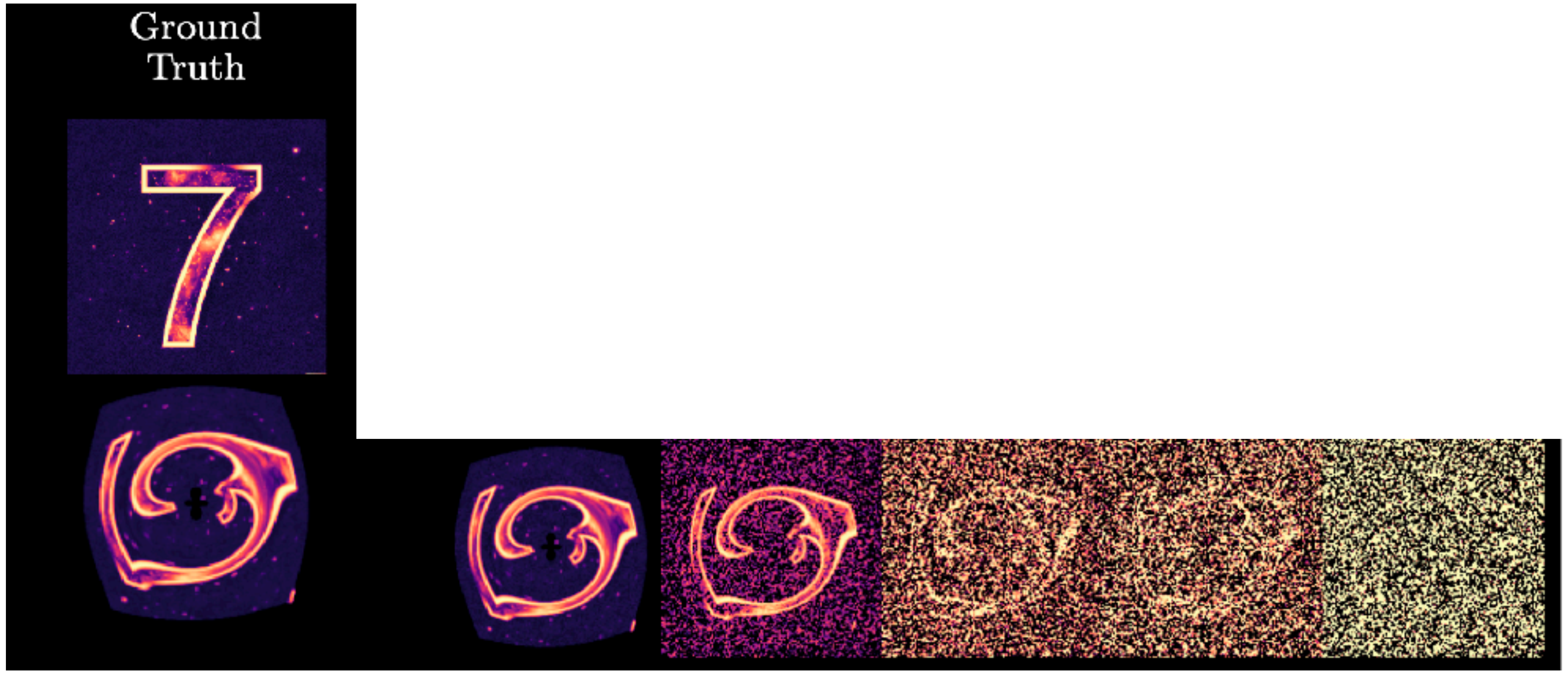
Adam et al. NeurIPS 2022 ML4PS workshop,
arXiv:2211.03812

Out of Distribution Tests

Alexandre Adam

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

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



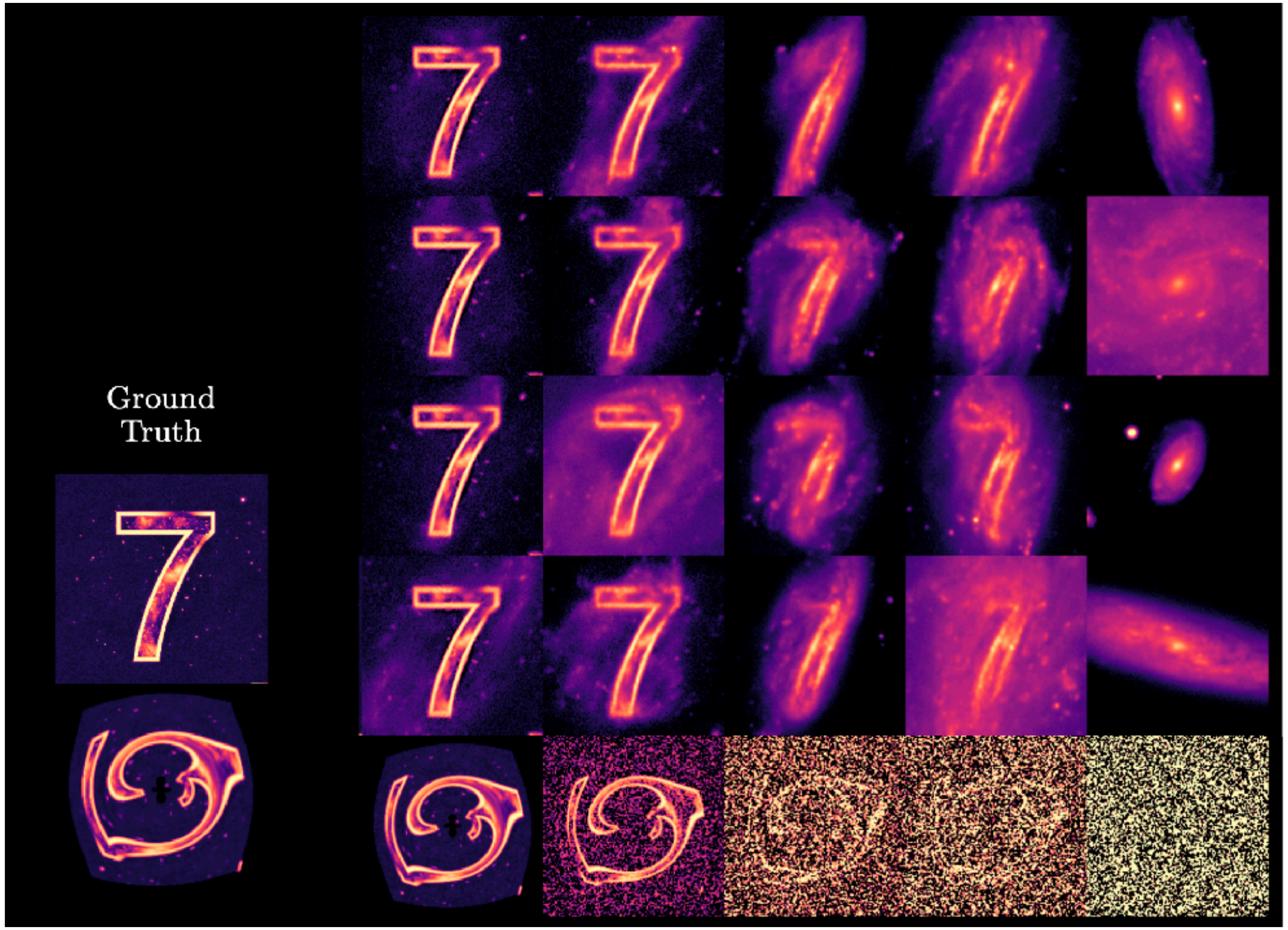
Adam et al. NeurIPS 2022 ML4PS workshop,
arXiv:2211.03812

Out of Distribution Tests

Alexandre Adam

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

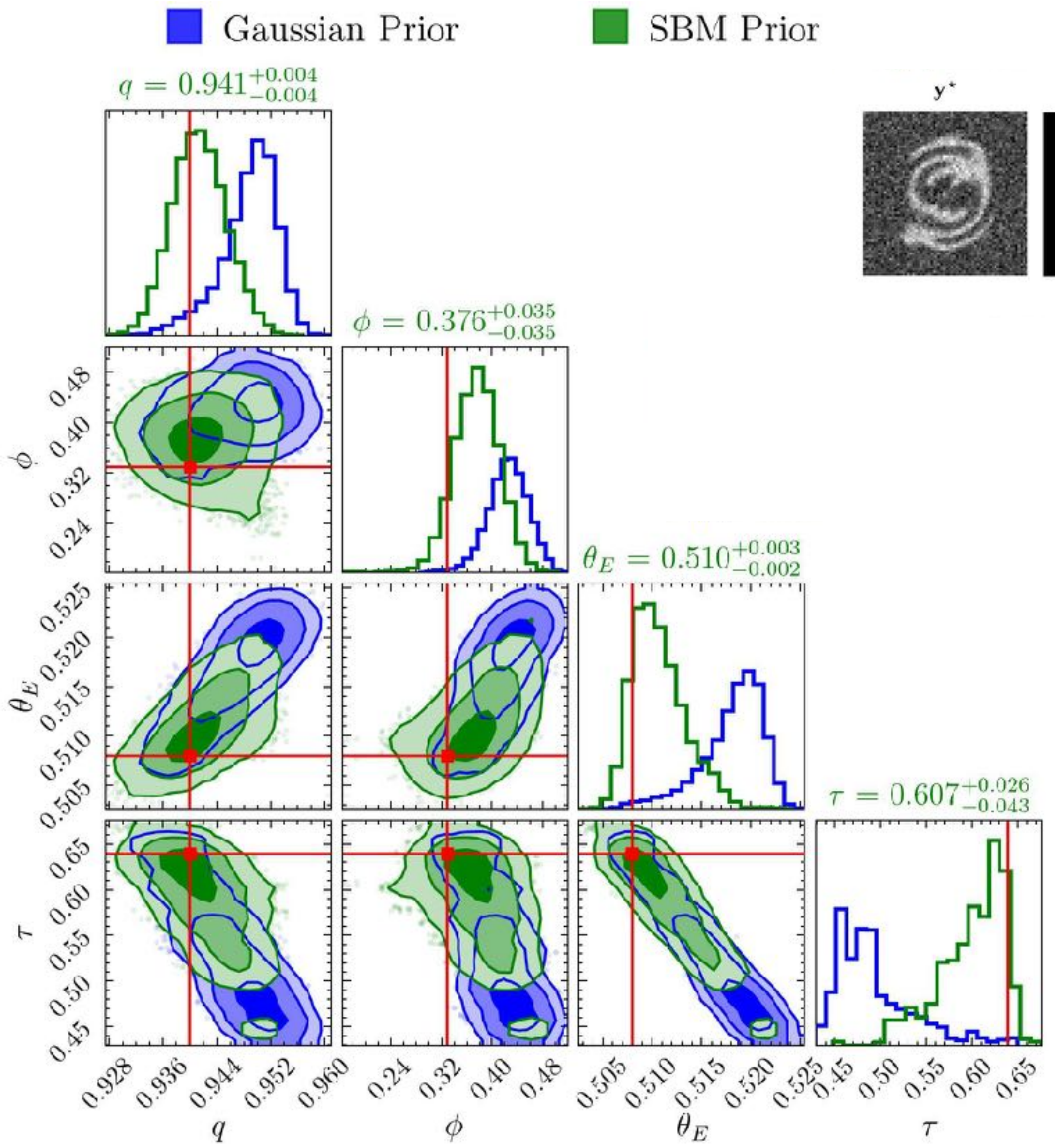
$$\nabla_x \log p(x \mid y) = \nabla_x \log p(y \mid x) + \nabla_x \log p_\theta(x)$$

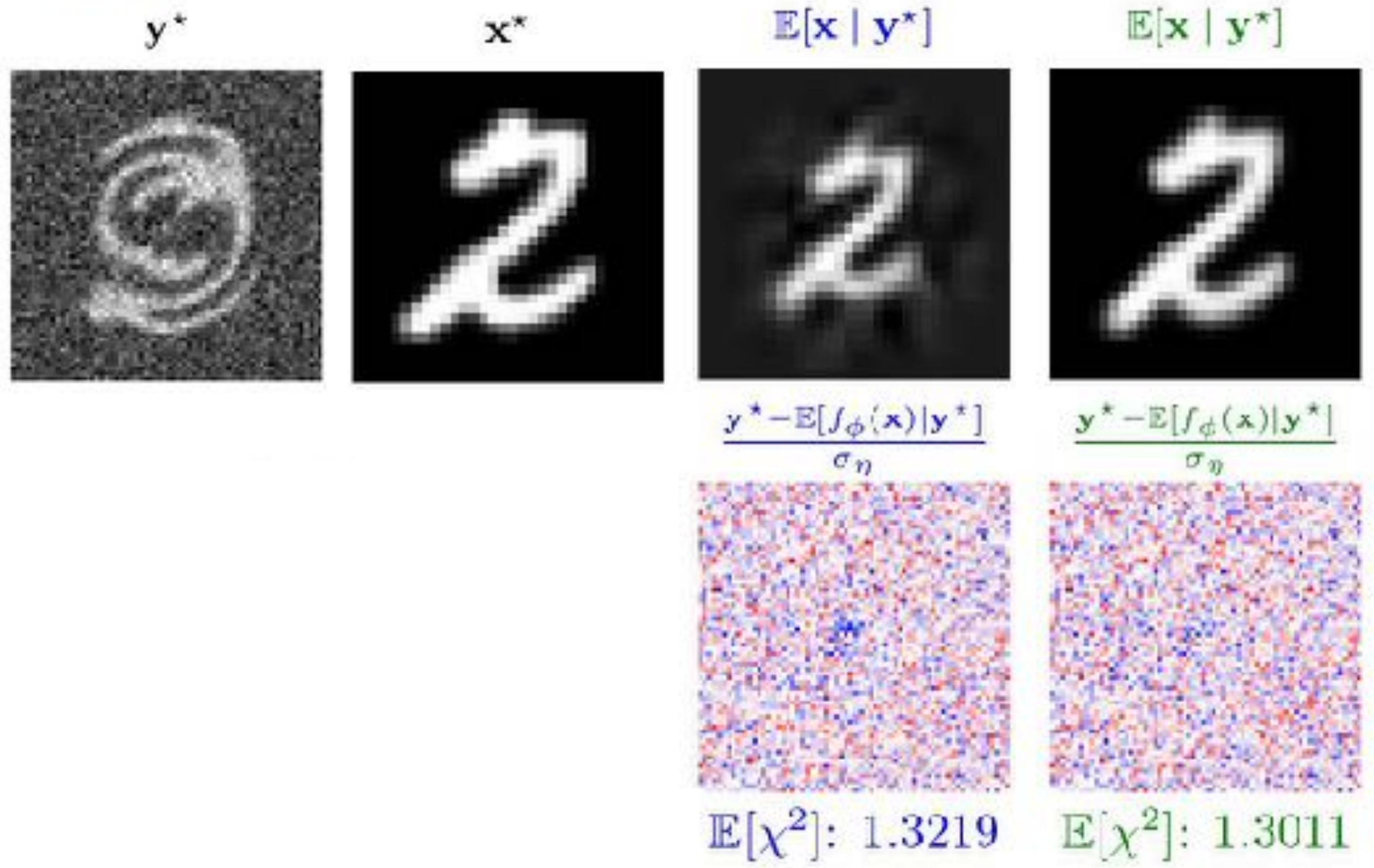


Adam et al. NeurIPS 2022 ML4PS workshop, arXiv:2211.03812

Joint Inference of Lens and Source Parameters

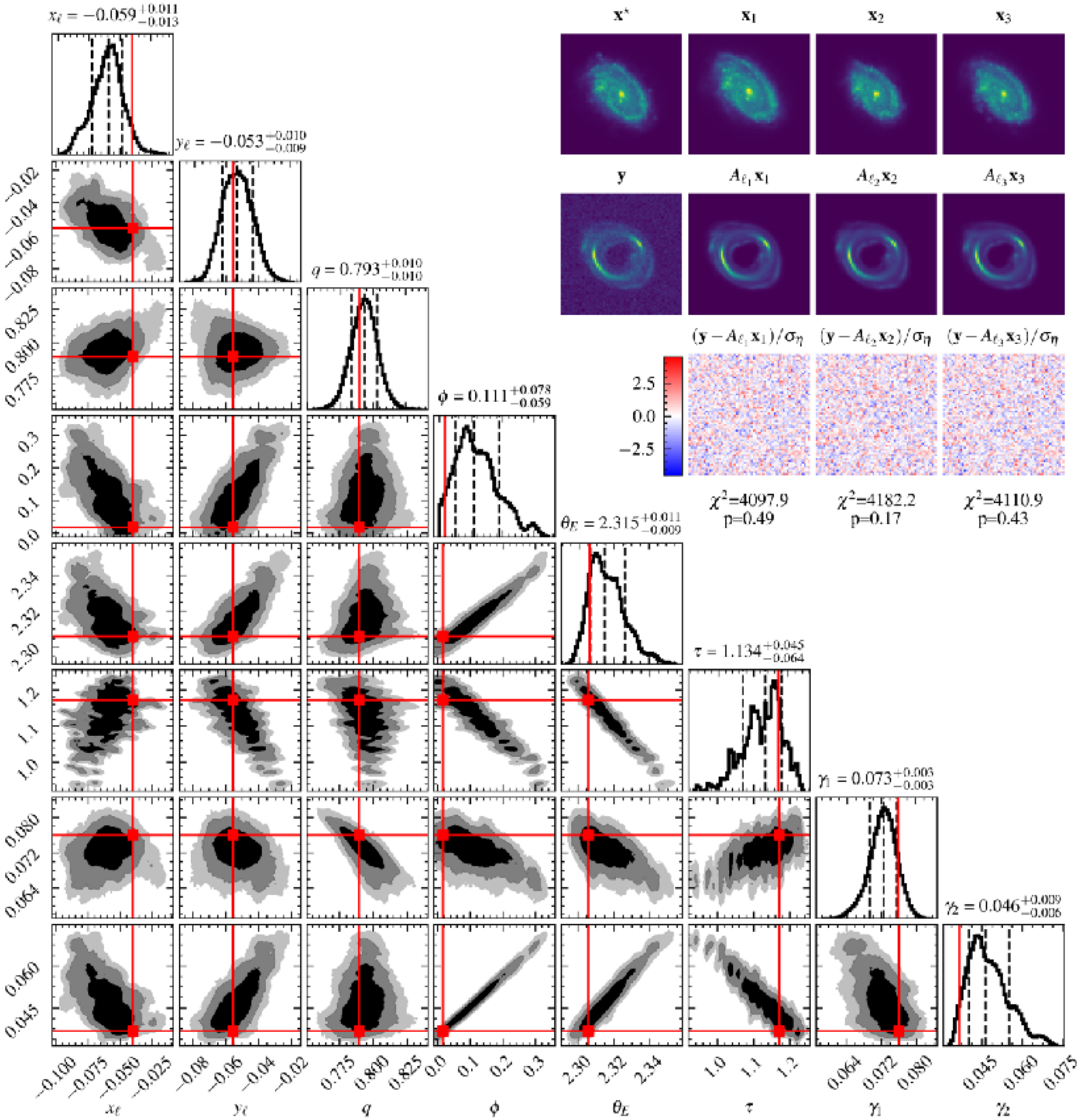
Missa Barco





Barco et al. 2025 ML4Astro workshop at ICML,

Joint Inference of Lens and Source Parameters



Missa Barco

Barco et al. 2025 ML4Astro workshop at ICML, arXiv:2511.04792

GAME PLAN

1. Data-driven priors

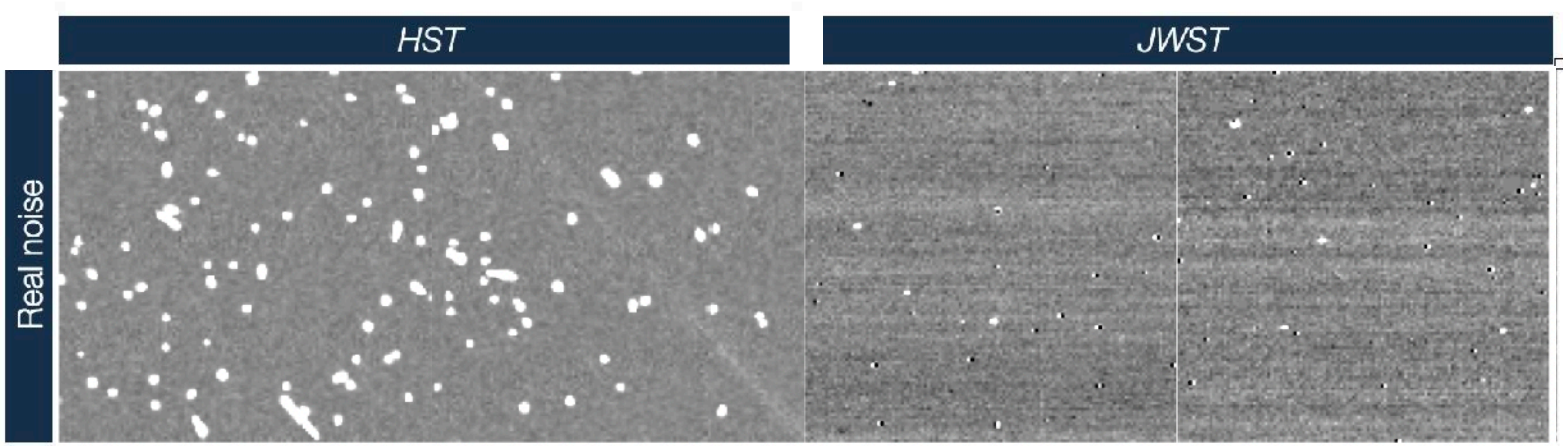
2. Data-driven likelihoods

3. Accuracy metrics

4. Out-of-Distribution accuracy

Dealing With Realistic Noise: Beyond Gaussianity

Alexandre Adam Ronan Legin



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

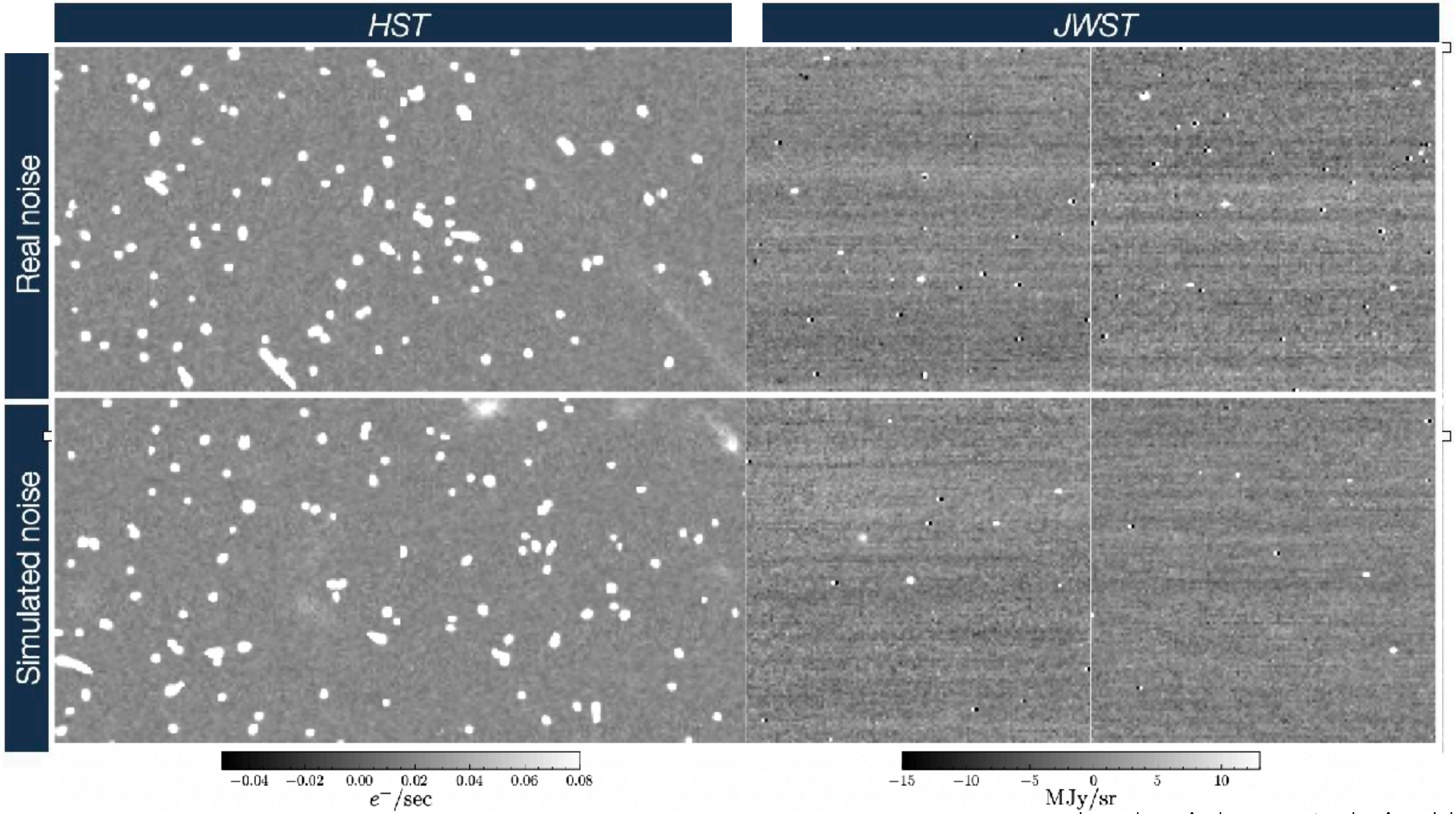
DEALING WITH REALISTIC NOISE: BEYOND GAUSSIANITY
SLIC: SCORE-BASED LIKELIHOOD CHARACTERIZATION


Alexandre Adam


Ronan Legin

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:

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



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

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

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

Alexandre Adam

Ronan Legin

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

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


Alexandre Adam


Ronan Legin

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

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

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

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


Alexandre Adam   Ronan Legin

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

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

$$\eta_{t+1} = \eta_t + \tau \boldsymbol{\nabla}_{\eta} \log Q(\mathbf{x}_{\mathrm{O}} - \mathbf{M}(\eta)) + \sqrt{2\tau}\xi_t$$

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

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


Alexandre Adam


Ronan Legin

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

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

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

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

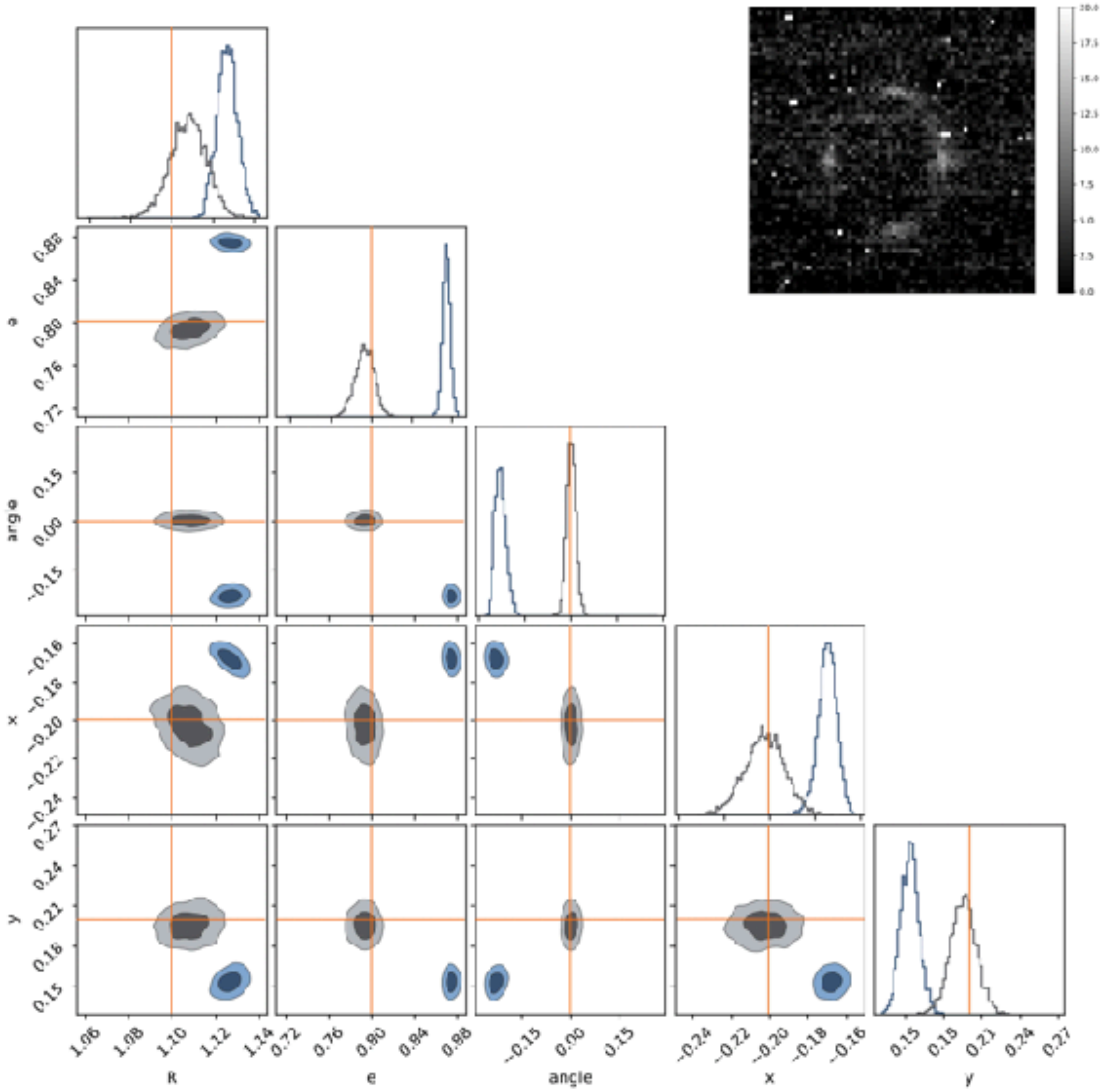
Dealing With Realistic Noise: Beyond Gaussianity
SLIC: Score-based LikelIhood Characterization

Alexandre Adam   Ronan Legin

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

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

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



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

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

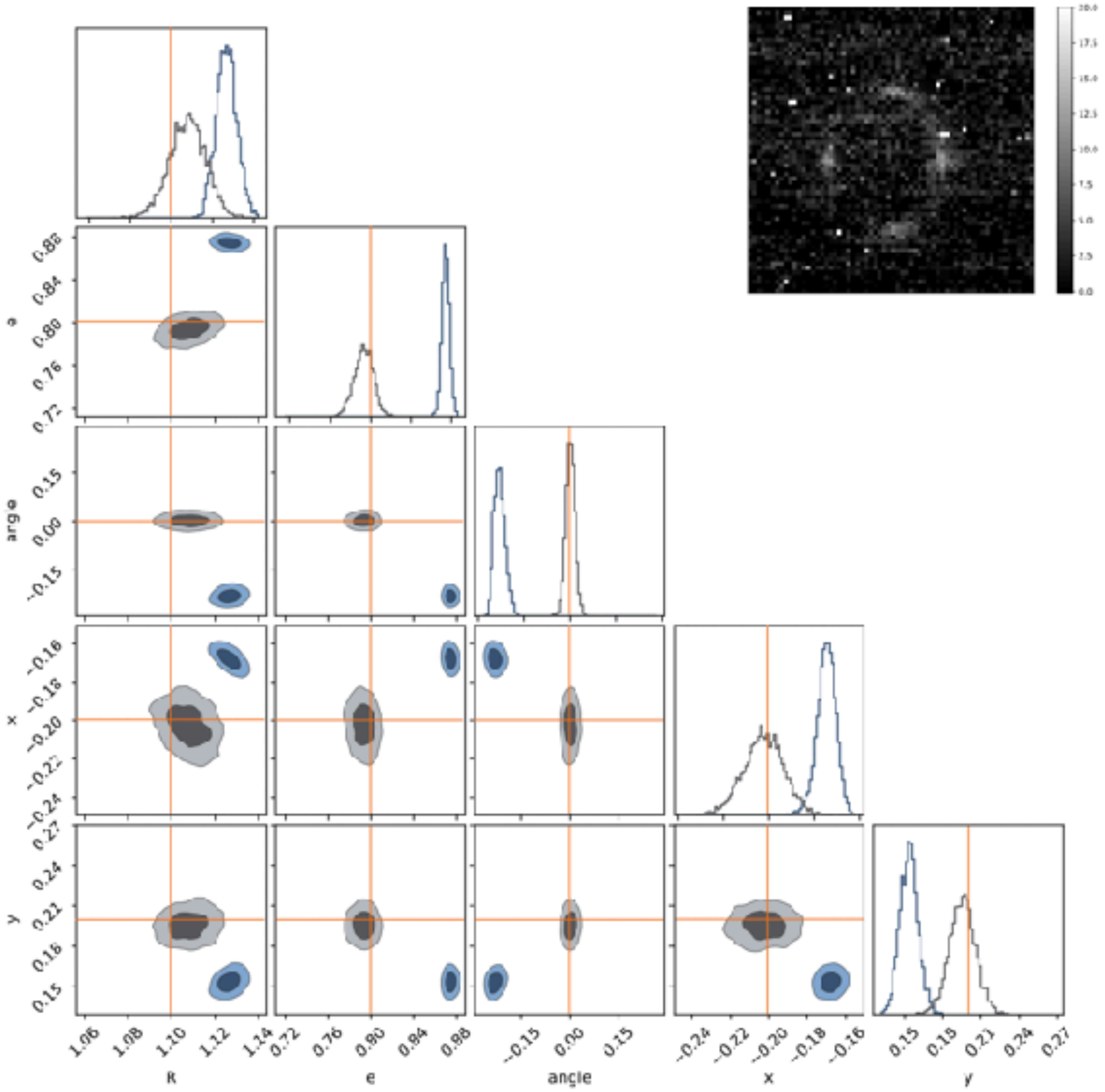

Alexandre Adam

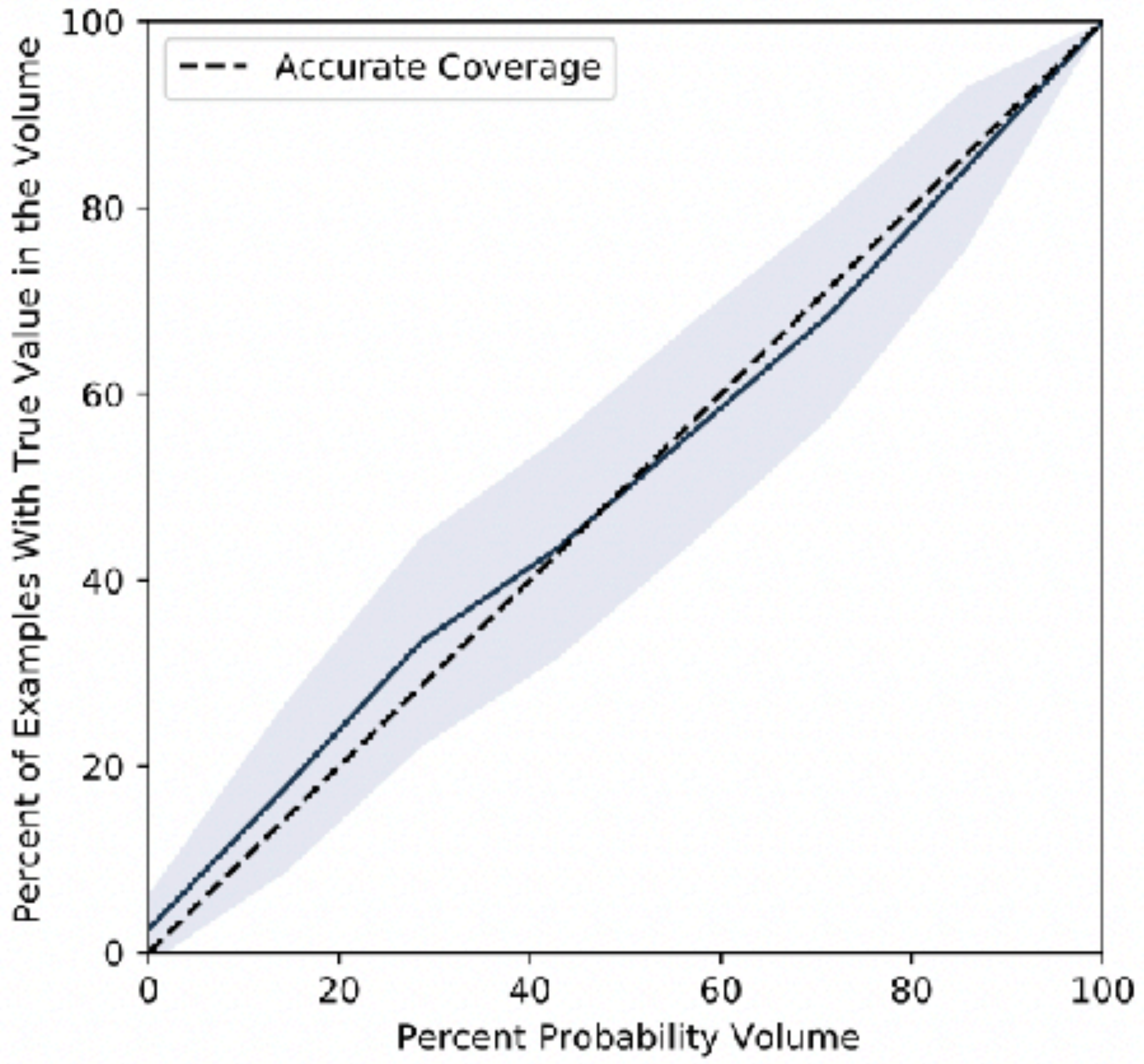

Ronan Legin

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

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

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





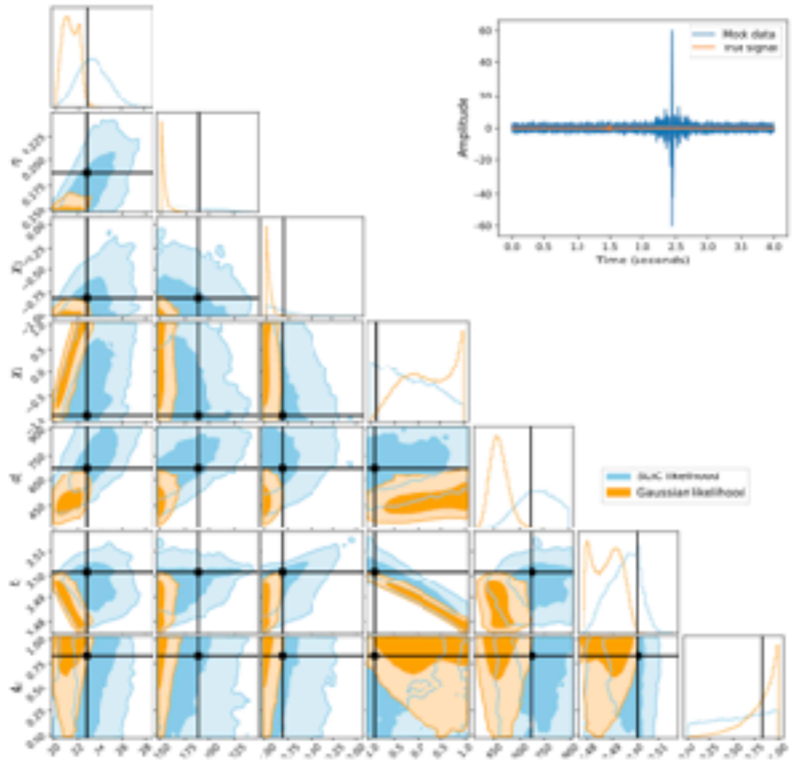
Legin, Adam, et al. ApJ Letters 2023, 2302.03046

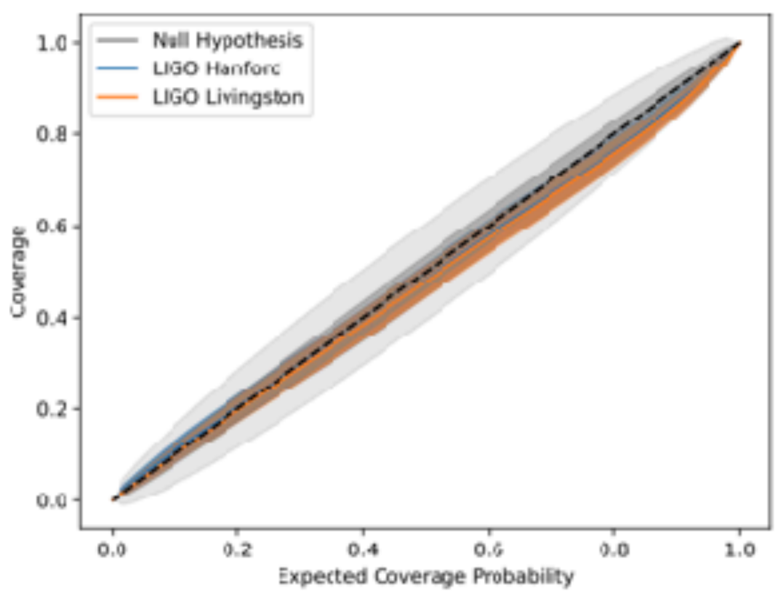
SLIC can be applied to multiple datasets where non-Gaussian noise is an issue:

SLIC CAN BE APPLIED TO MULTIPLE DATASETS WHERE NON-GAUSSIAN NOISE IS AN ISSUE:

GRAVITATIONAL WAVE PARAMETER INFERENCE IN THE PRESENCE OF GLITCHES

Ronan Legin



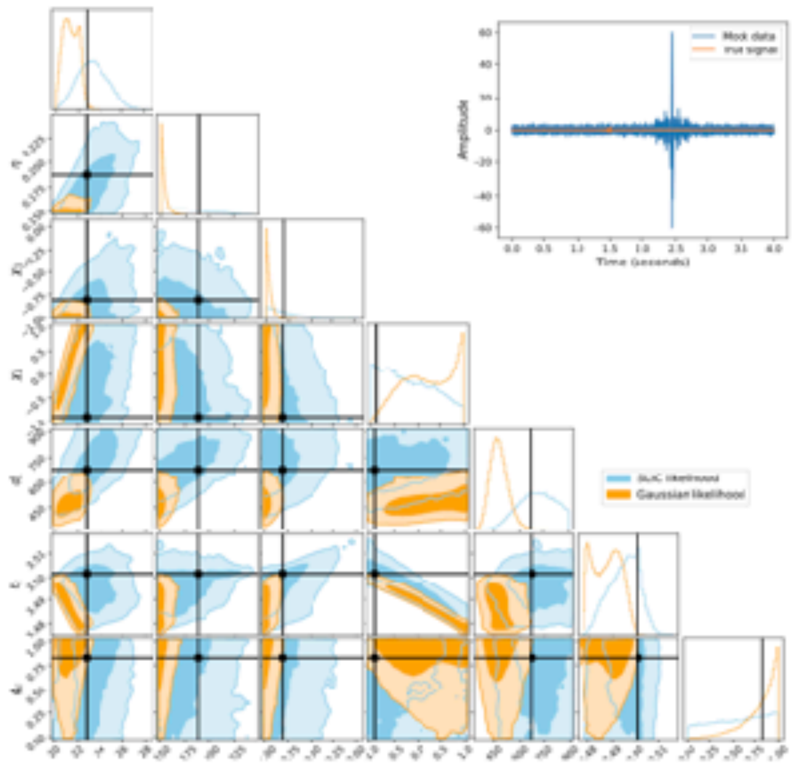


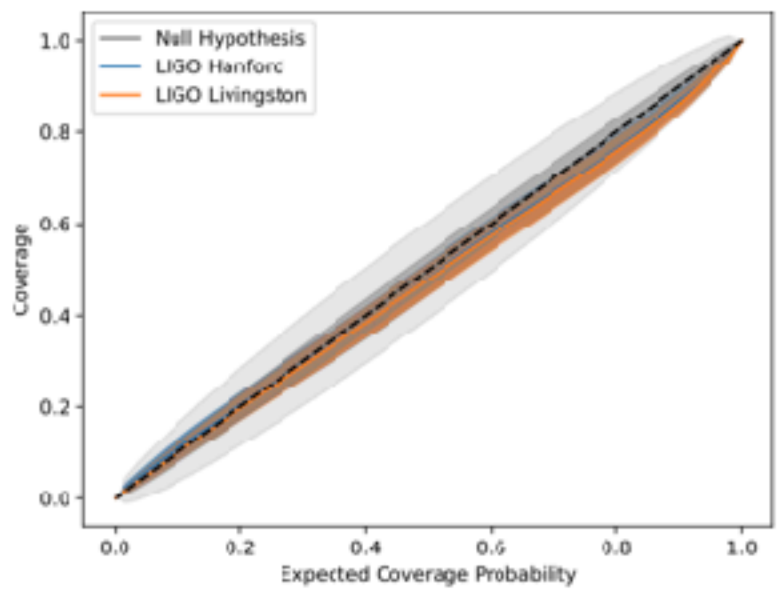
Legin et at, ApJ Letters 2026, 2410.19956

SLIC can be applied to multiple datasets where non-Gaussian noise is an issue:

Gravitational wave parameter inference in the presence of glitches

Ronan Legin

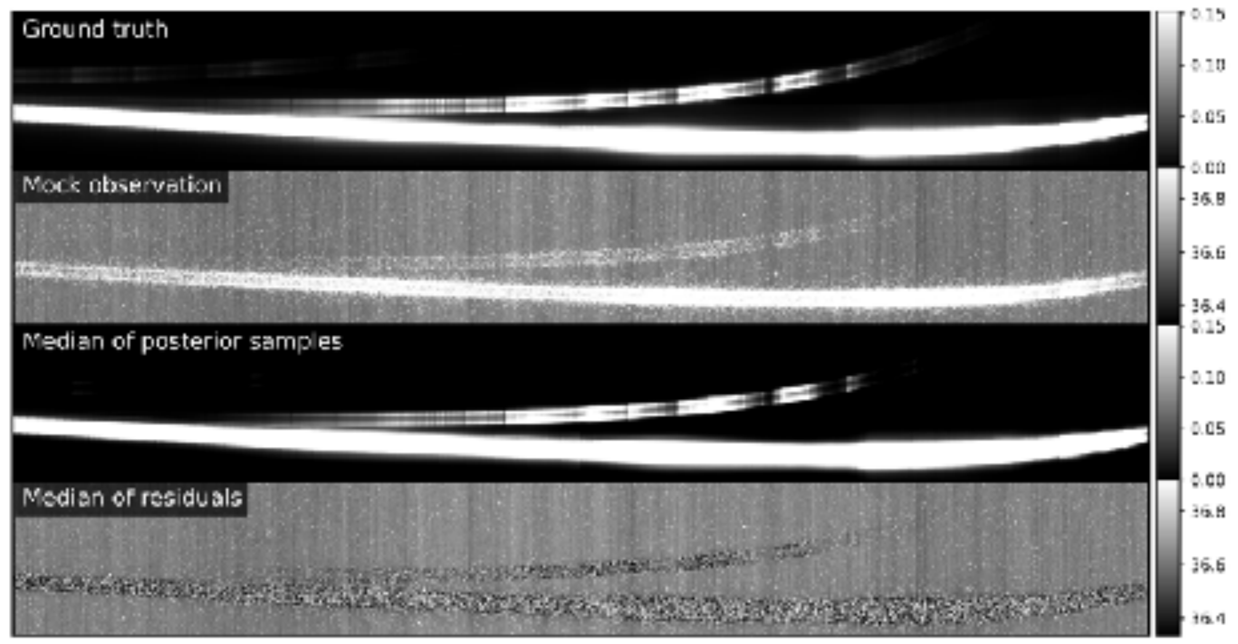


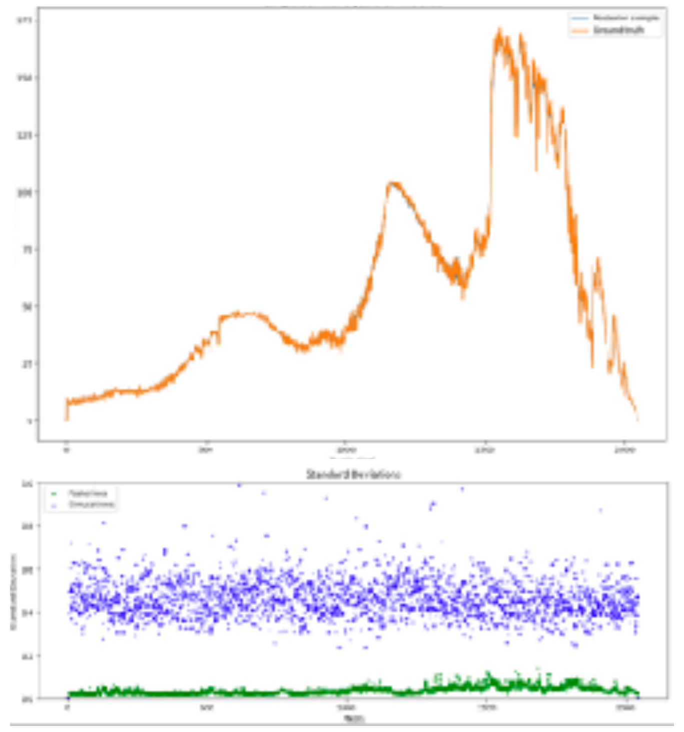


Legin et at, ApJ Letters 2026, 2410.19956

Noise modeling for JWST spectroscopy

Salma Salhi



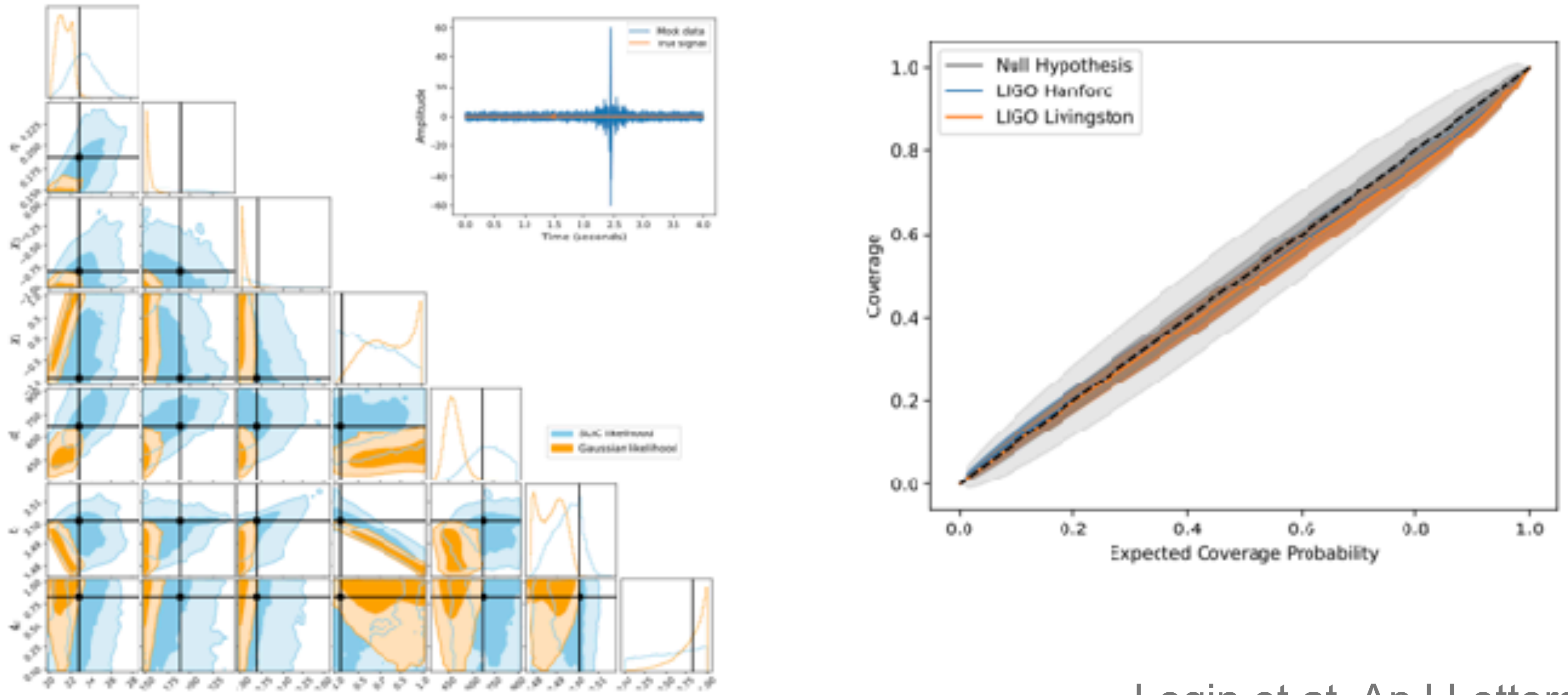


Salhi et a;, NeurIPS ML4PS workshop 2024

SLIC can be applied to multiple datasets where non-Gaussian noise is an issue:

Gravitational wave parameter inference in the presence of glitches

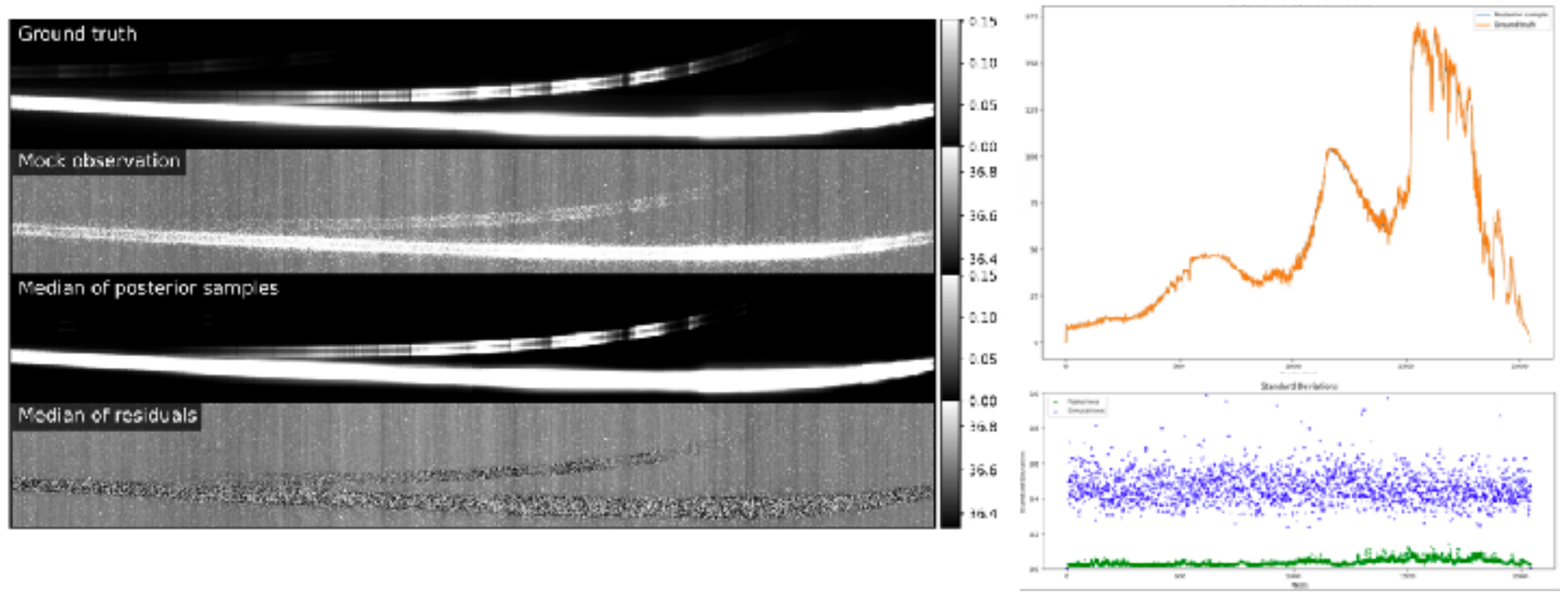
Ronan Legin



Legin et at, ApJ Letters 2026, 2410.19956

Noise modeling for JWST spectroscopy

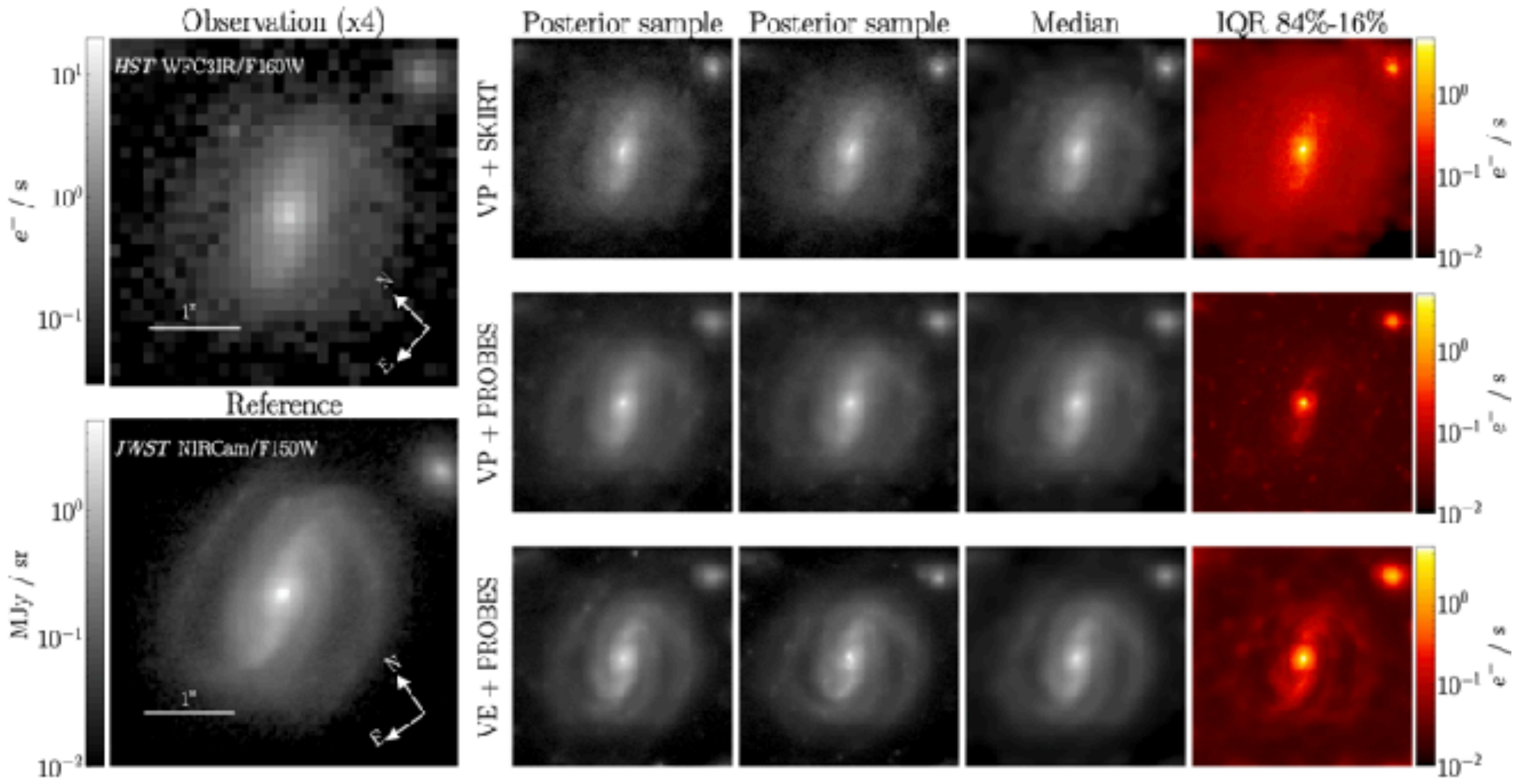
Salma Salhi



Salhi et a;, NeurIPS ML4PS workshop 2024

PSF-deconvolution (for HST)

Alexandre Adam

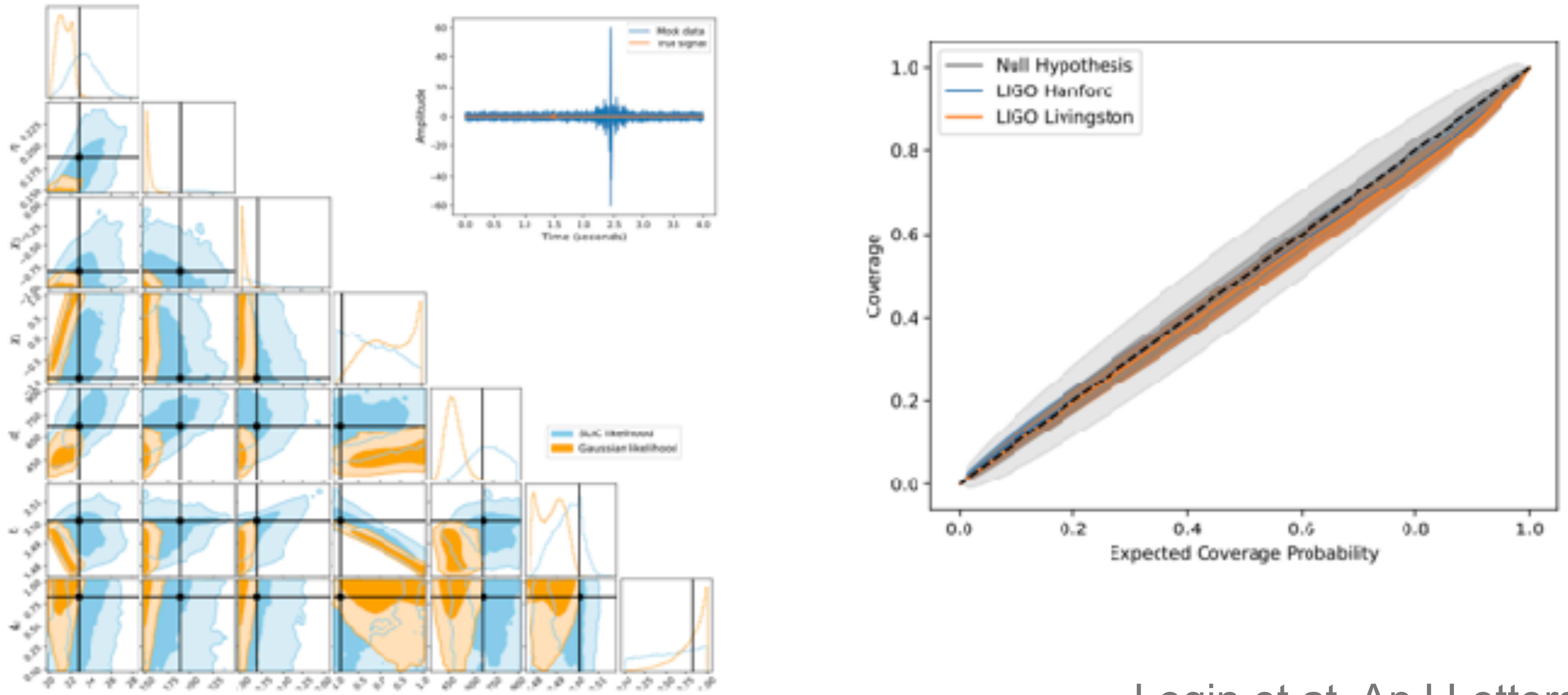


Adam et al. NeurIPS 2023 ML4PS workshop

SLIC can be applied to multiple datasets where non-Gaussian noise is an issue:

Gravitational wave parameter inference in the presence of glitches

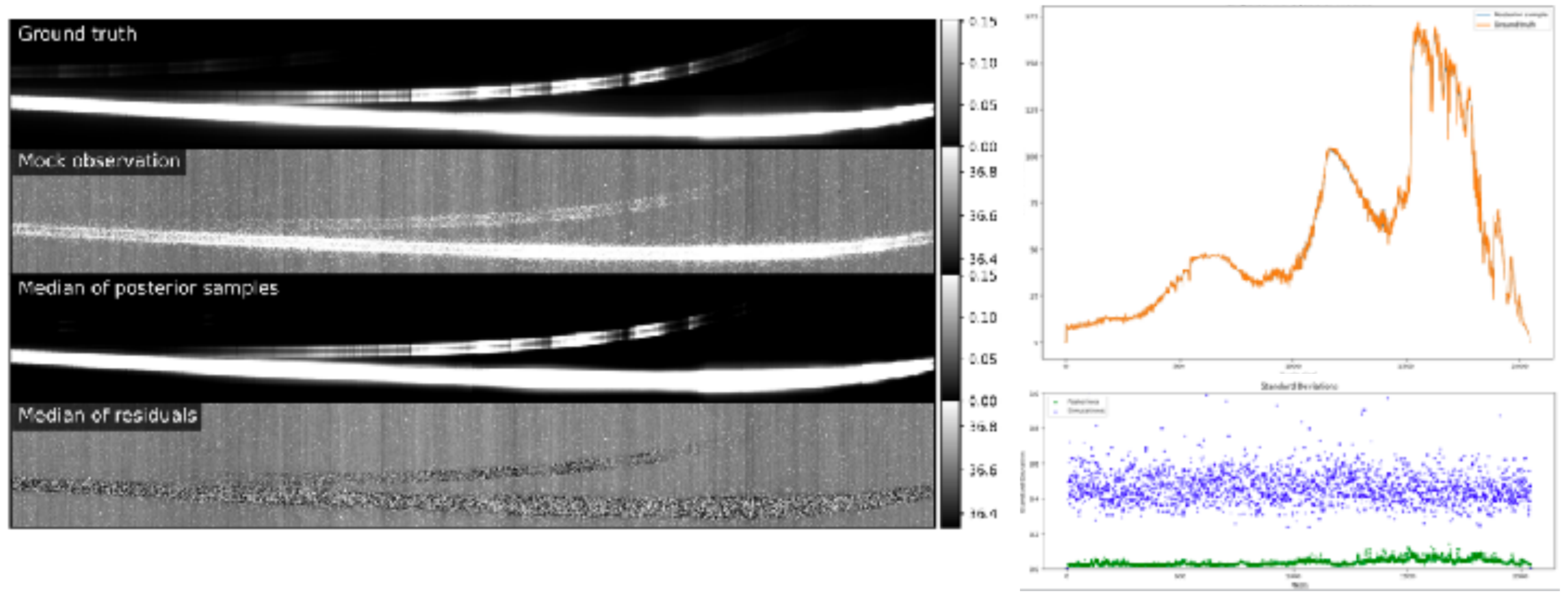

Ronan Legin



Legin et at, ApJ Letters 2026, 2410.19956

Noise modeling for JWST spectroscopy

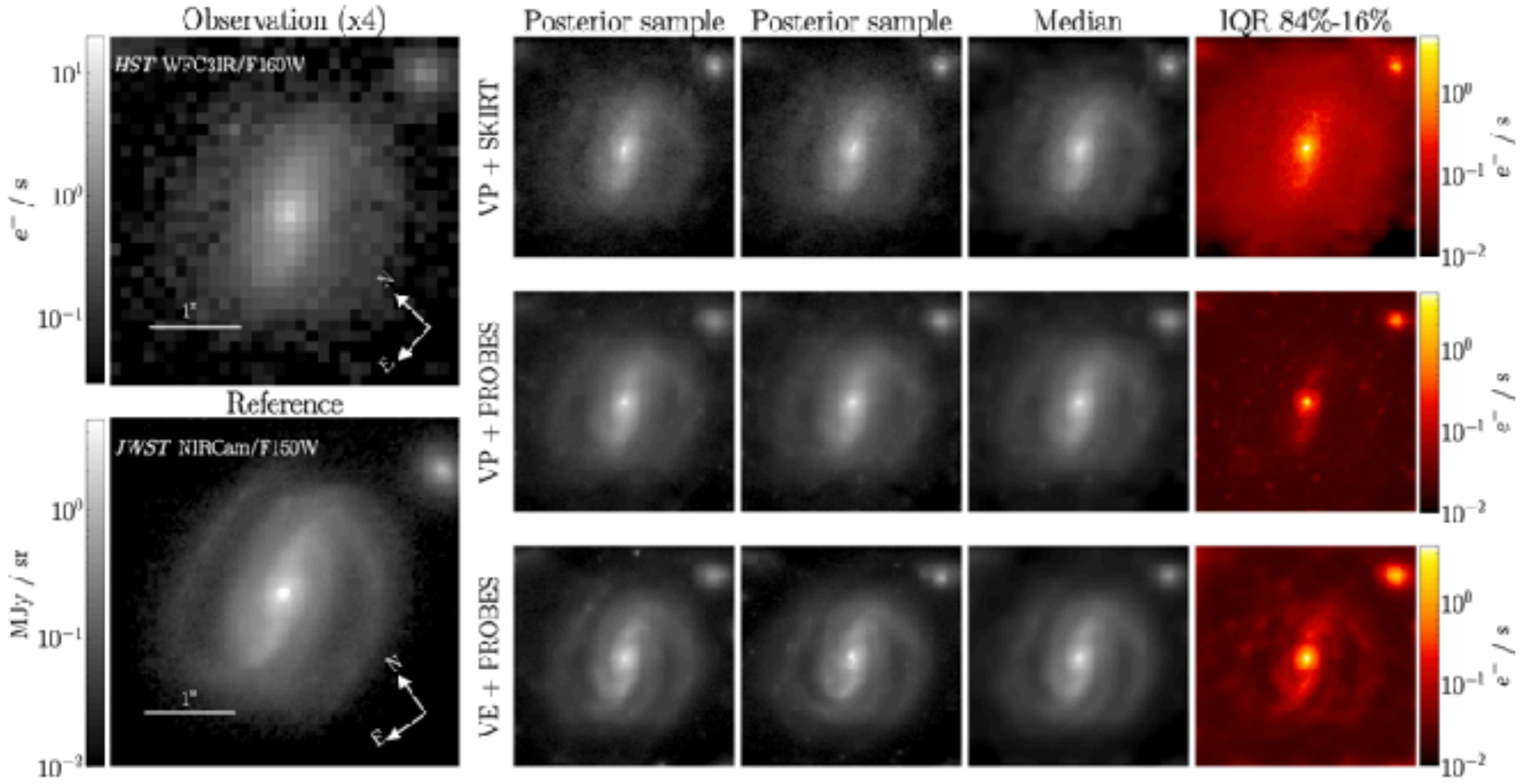

Salma Salhi



Salhi et a;, NeurIPS ML4PS workshop 2024

PSF-deconvolution (for HST)


Alexandre Adam



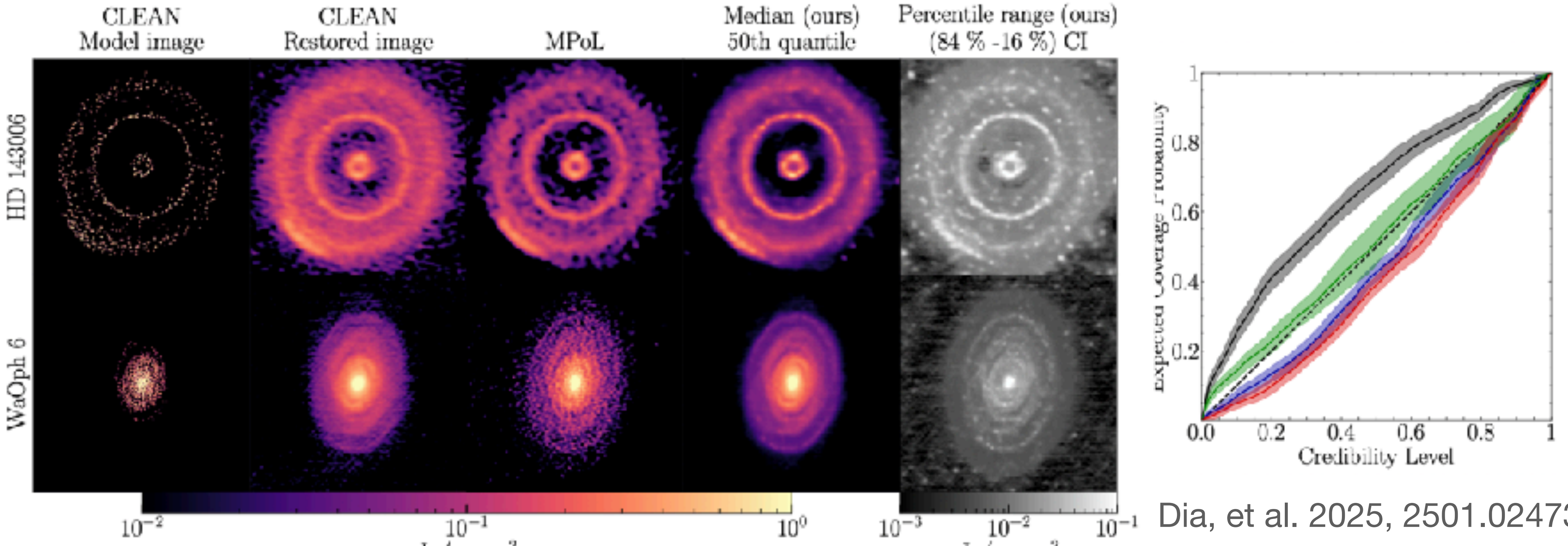
Adam et al. NeurIPS 2023 ML4PS workshop

PSF-deconvolution for interferometric data

With Score-Based Diffusion Models

$$\mathcal{V} = \mathcal{SFP}_{\text{beam}}\mathbf{x} + \eta$$


Noé Dia



Dia, et al. 2025, 2501.02473

GAME PLAN

1. Data-driven priors

2. Data-driven likelihoods

3. Accuracy metrics

4. Out-of-Distribution accuracy

Machine learning is enabling us to do inference in high-dimensional spaces, and tackle previously intractable questions

Machine learning is enabling us to do inference in high-dimensional spaces, and tackle previously intractable questions

In these high-dimensional spaces,
how do we actually define and measure accuracy?

Machine learning is enabling us to do inference in high-dimensional spaces, and tackle previously intractable questions

In these high-dimensional spaces,
how do we actually define and measure accuracy?

We need to answer 2 questions:

Machine learning is enabling us to do inference in high-dimensional spaces, and tackle previously intractable questions

In these high-dimensional spaces,
how do we actually define and measure accuracy?

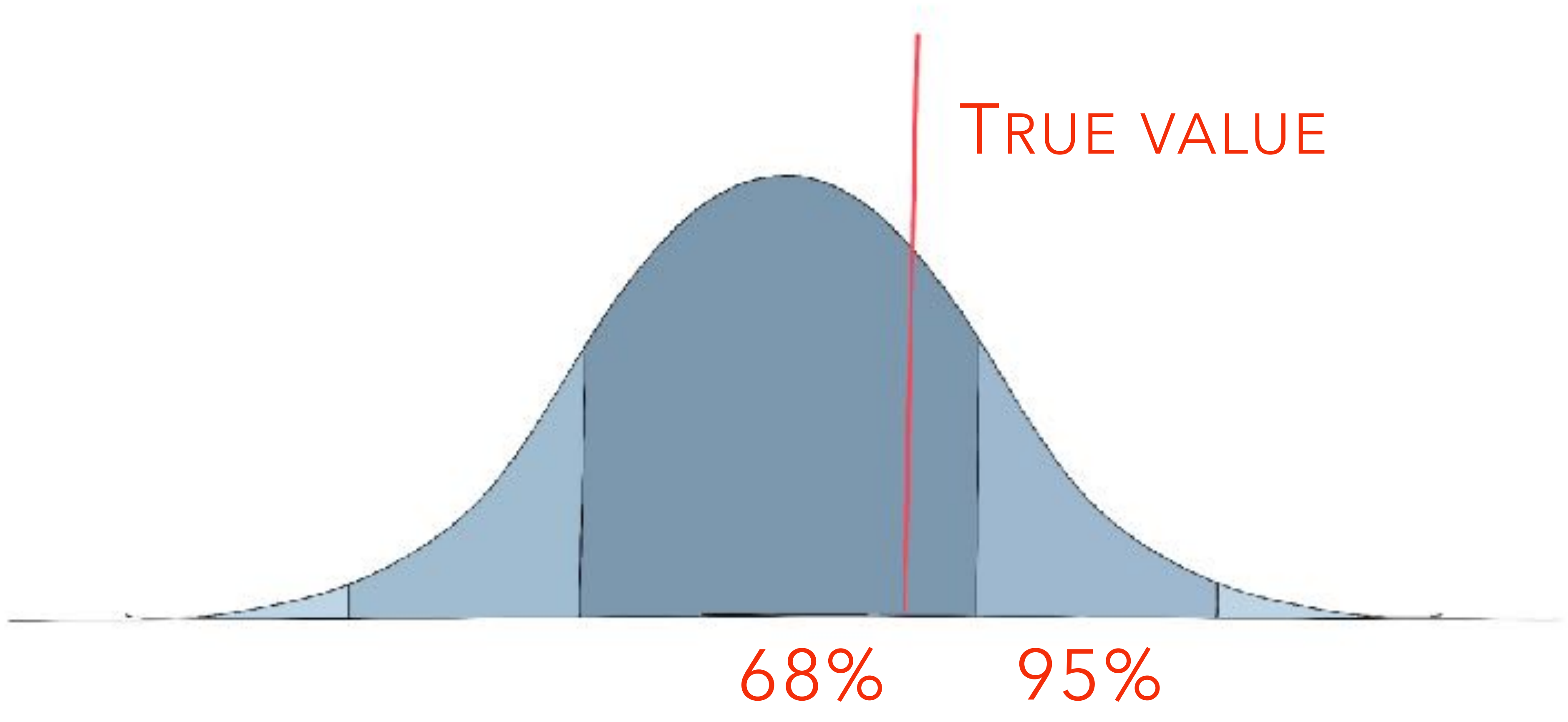
We need to answer 2 questions:

- If we have a pipeline that infers high-dimensional variables, like images of galaxies, how do we assess the accuracy of this pipeline?

- If we are using generative models as components or inputs to those inference pipelines, how do we quantify the accuracy with which they represent the underlying distribution?

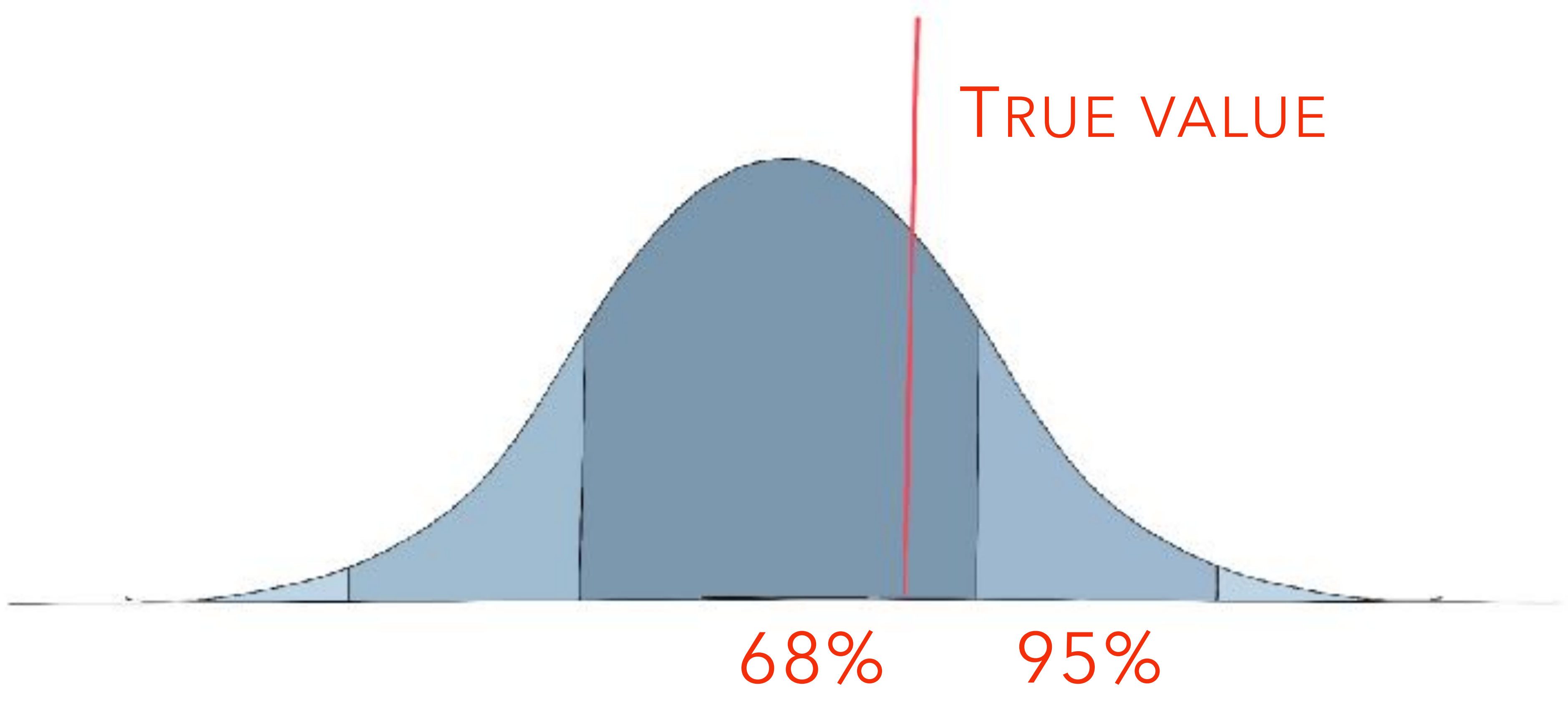
ARE THESE UNCERTAINTIES ACCURATE?
COVERAGE TEST ASSESSMENT

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



Are these Uncertainties Accurate?
Coverage test assessment

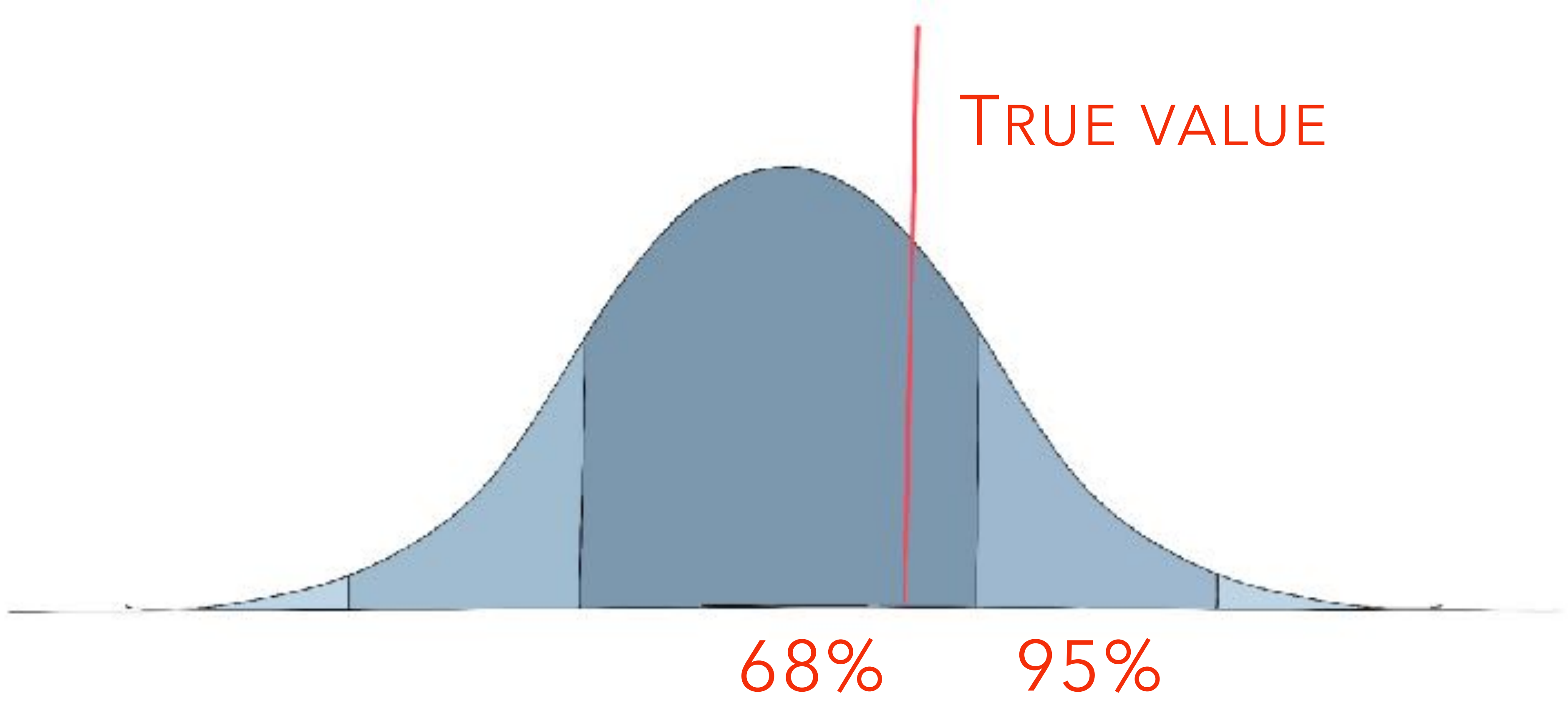
The expected coverage probability of a credible region is the proportion of the time that the region contains the true value of interest.



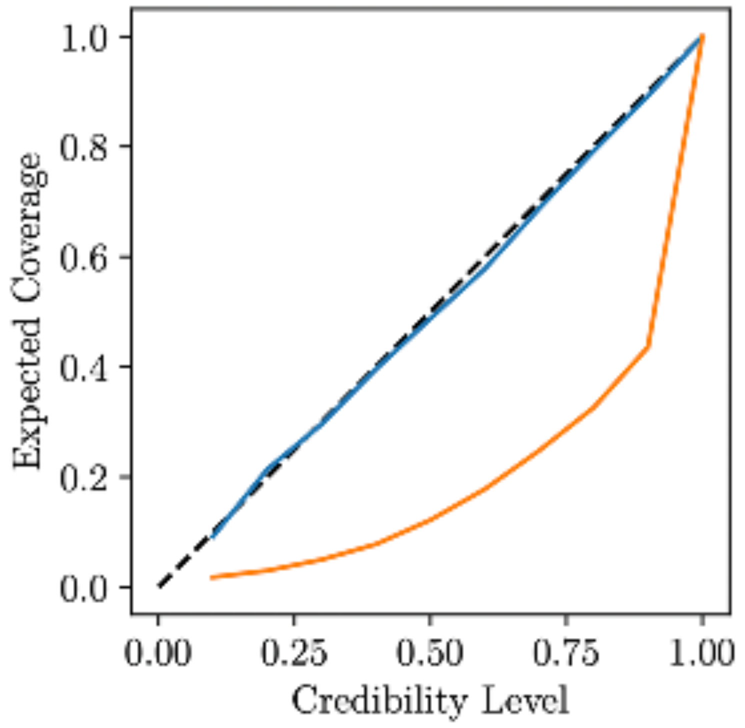
For an accurate posterior estimator, the expected coverage probability is equal to the probability mass of the credible region.

ARE THESE UNCERTAINTIES ACCURATE?
COVERAGE TEST ASSESSMENT

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

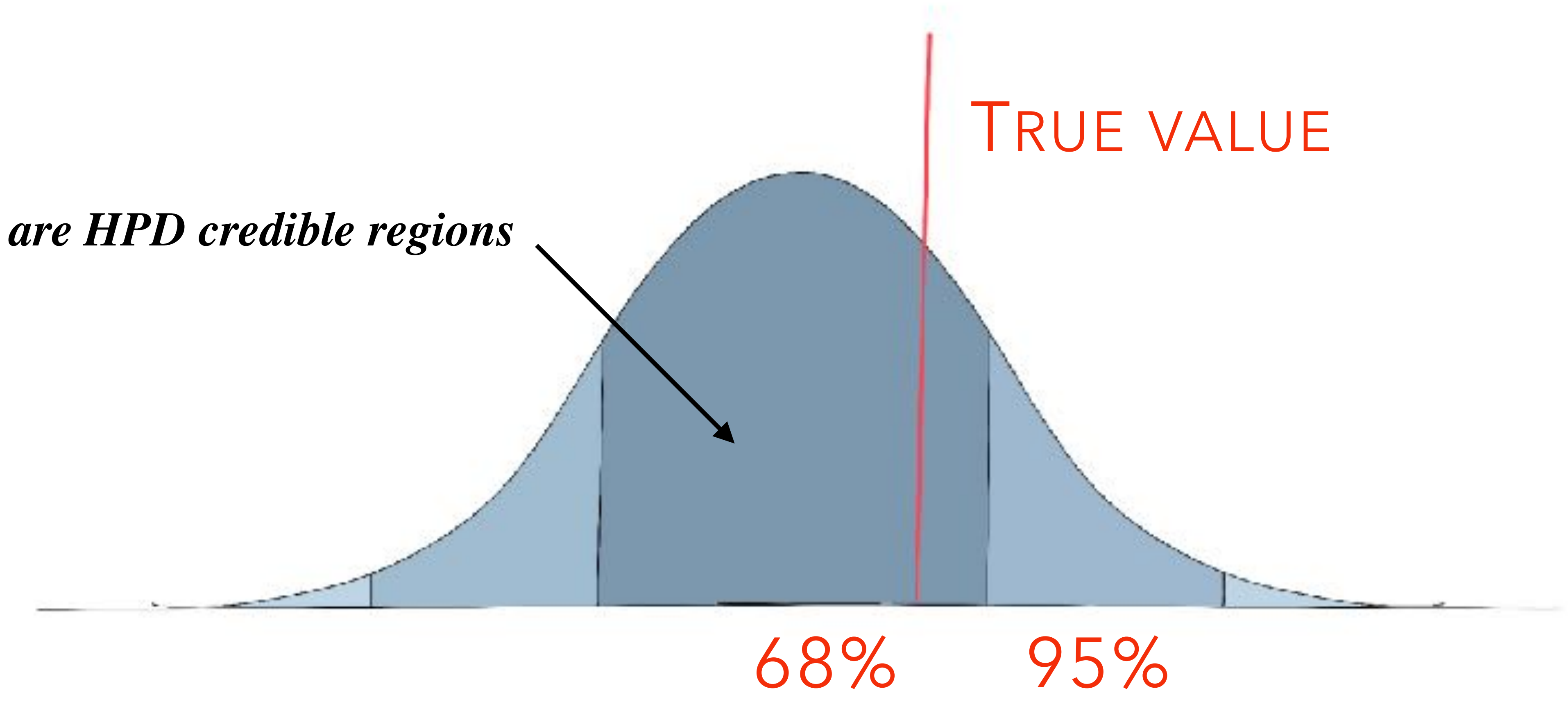


For an accurate posterior estimator, the expected coverage probability is equal to the probability mass of the credible region.

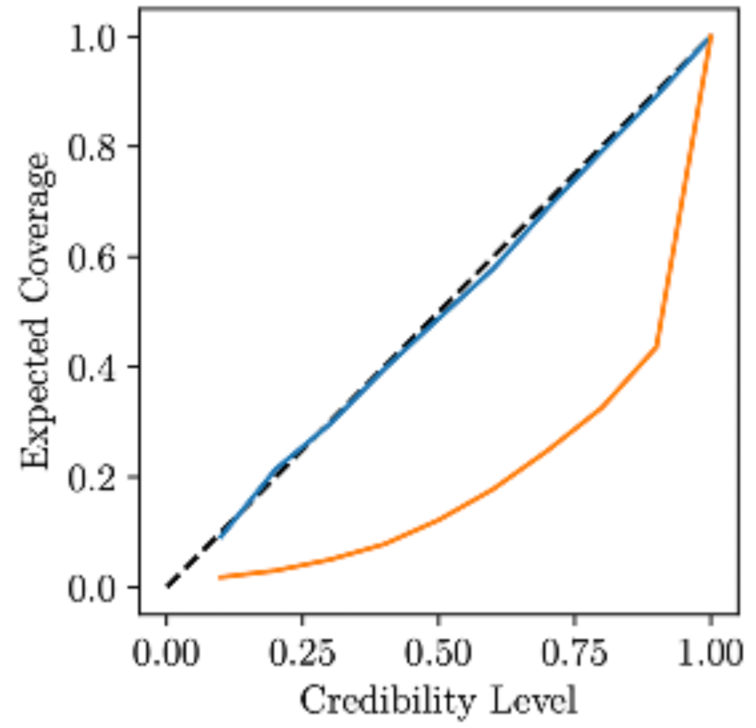


Are these Uncertainties Accurate?
Coverage test assessment

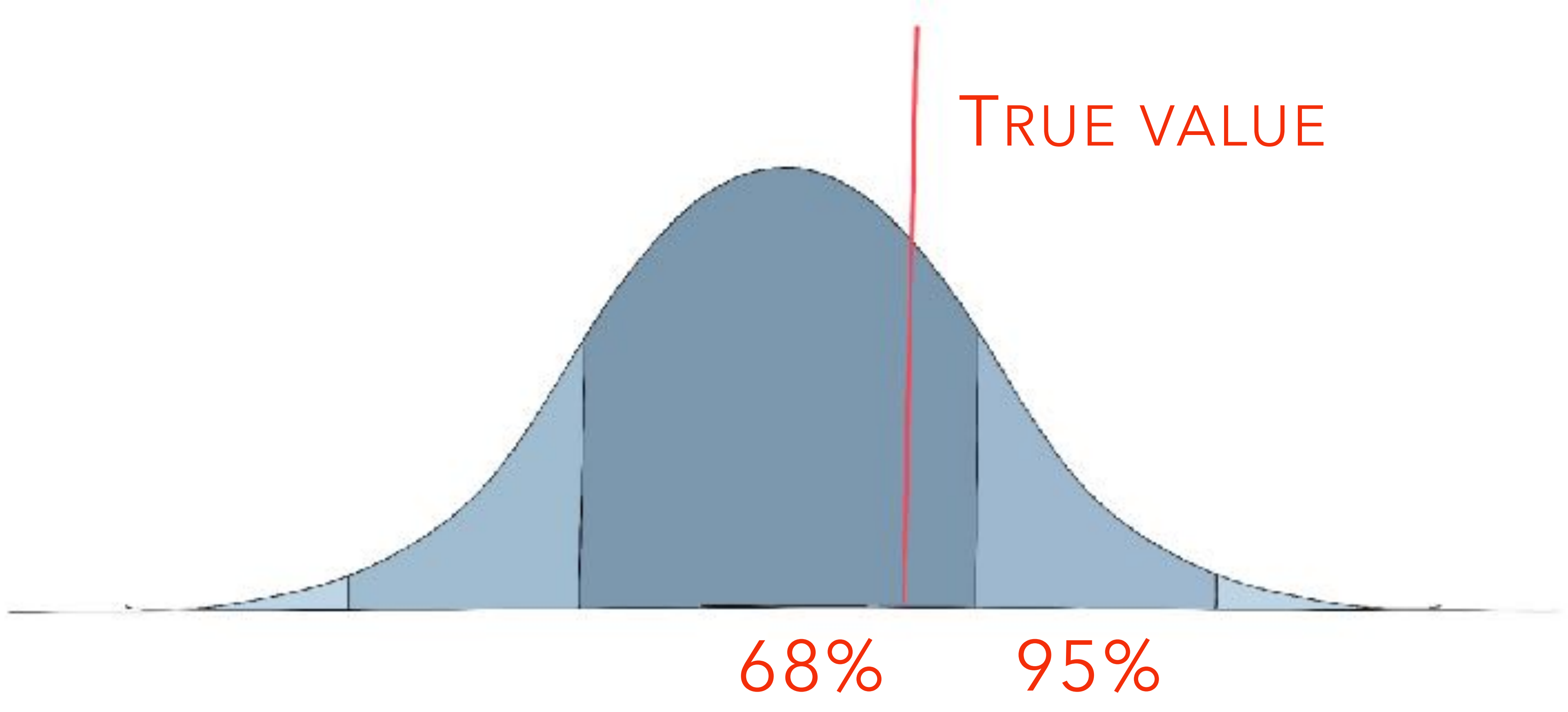
The expected coverage probability of a credible region is the proportion of the time that the region contains the true value of interest.



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



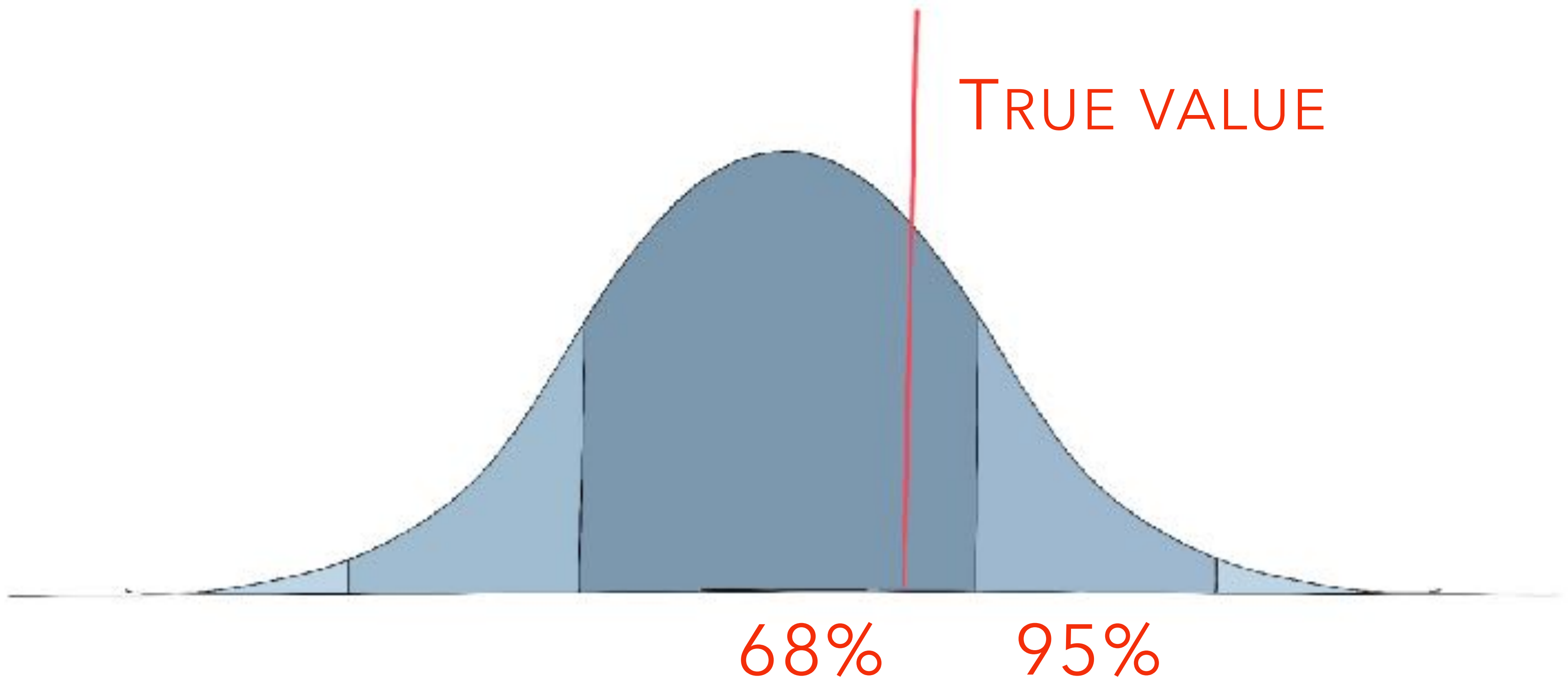
PROBLEMS WITH TRADITIONAL HPD COVERAGE TESTS:



1. In high dimension, like in pixel space for an image, how do we find a the HDP region for a given volume?

We could estimate the density (e.g. with KDE) but very crude approximation with limited samples, especially over complex densities like images.

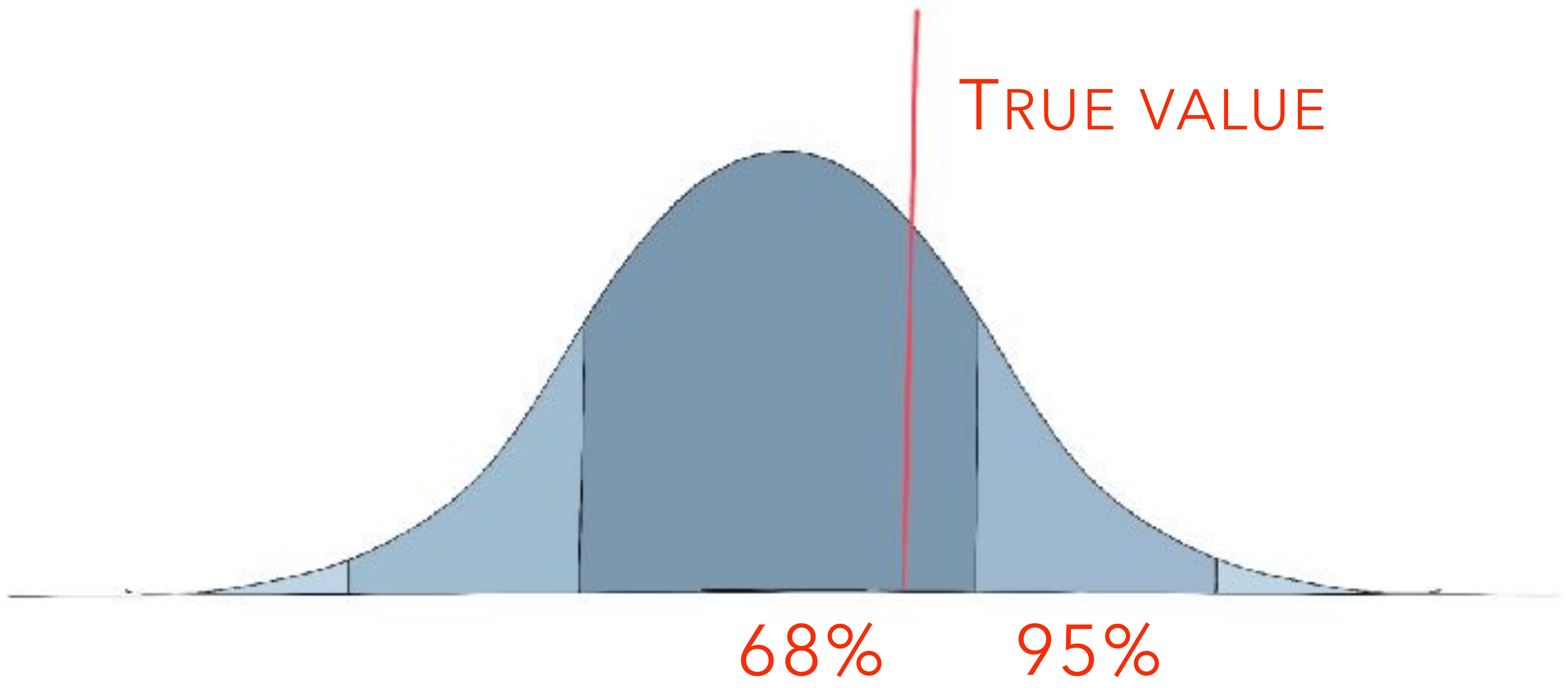
PROBLEMS WITH TRADITIONAL HPD COVERAGE TESTS:



2. This is a necessary but **not sufficient** condition for calibration

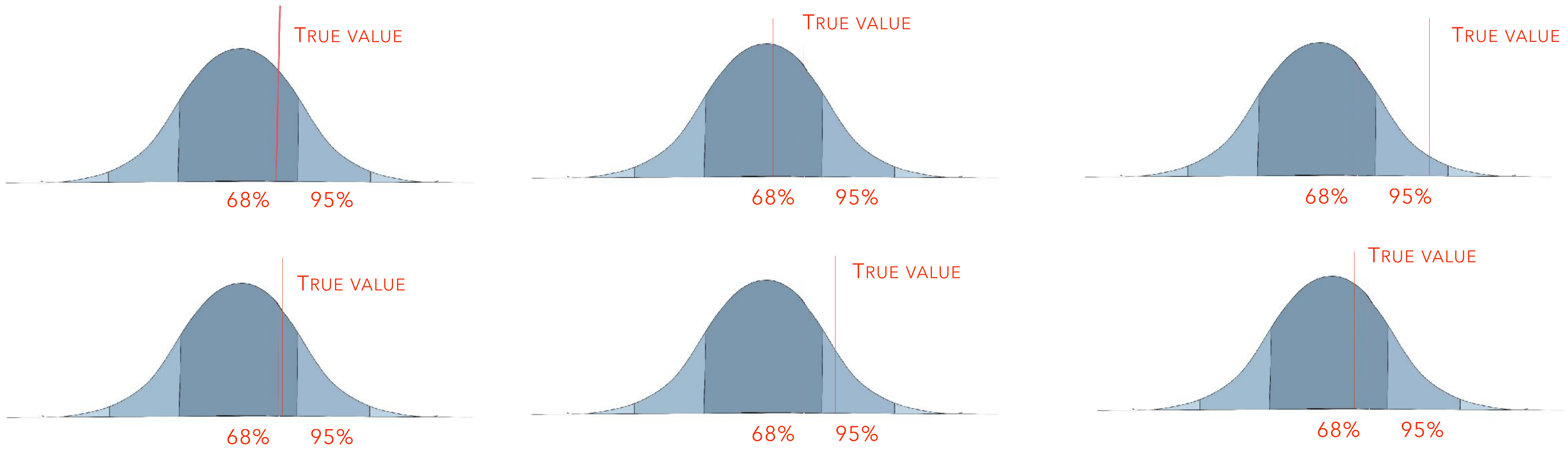
Lots of inaccuracies could be hiding in a diagonal plot!

PROBLEMS WITH TRADITIONAL HPD COVERAGE TESTS:



2. This is a necessary but **not sufficient** condition for calibration

Lots of inaccuracies could be hiding in a diagonal plot!

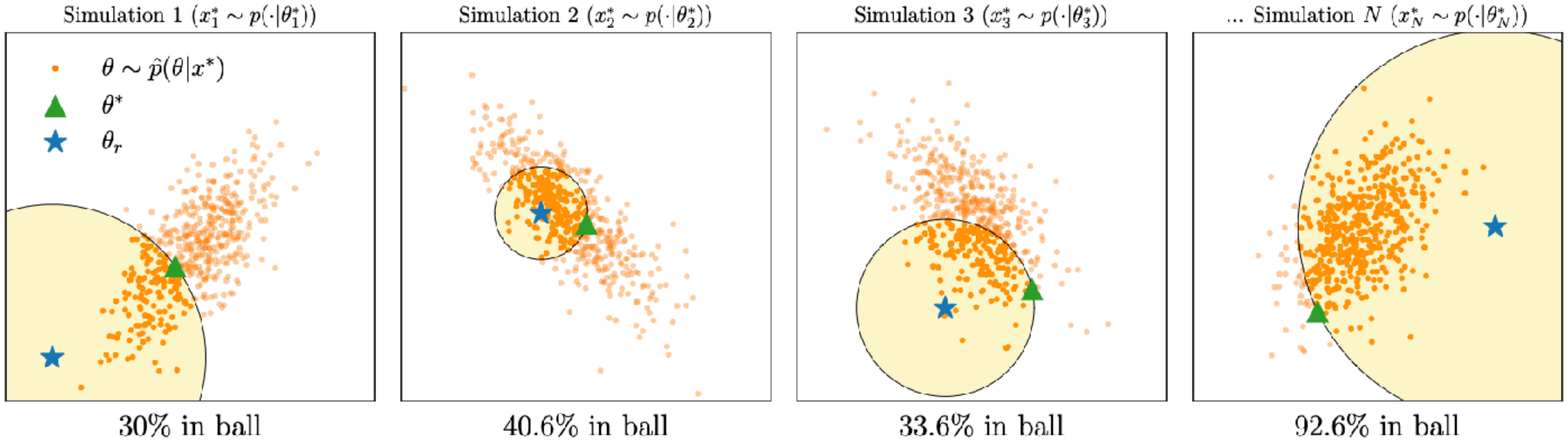


Coverage Test for Accuracy with Random Points

-TARP-


Adam Coogan

Pablo Lemos

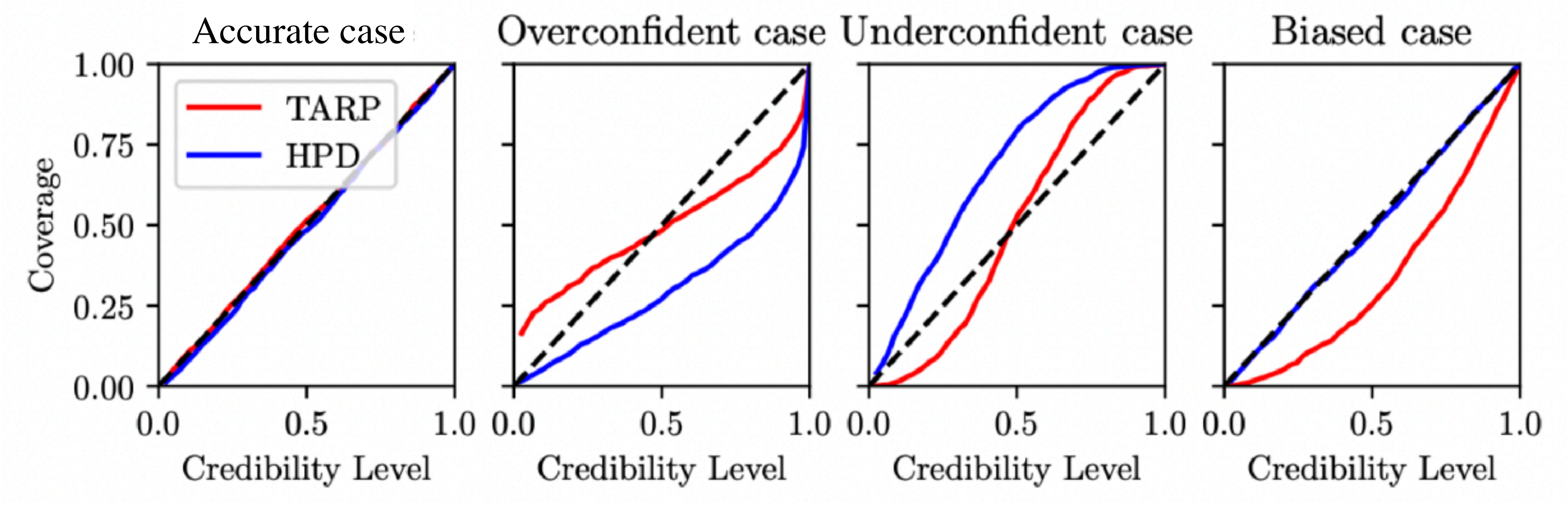


pip install tarp

Lemos, et al. ICML 2023, 2302.03026

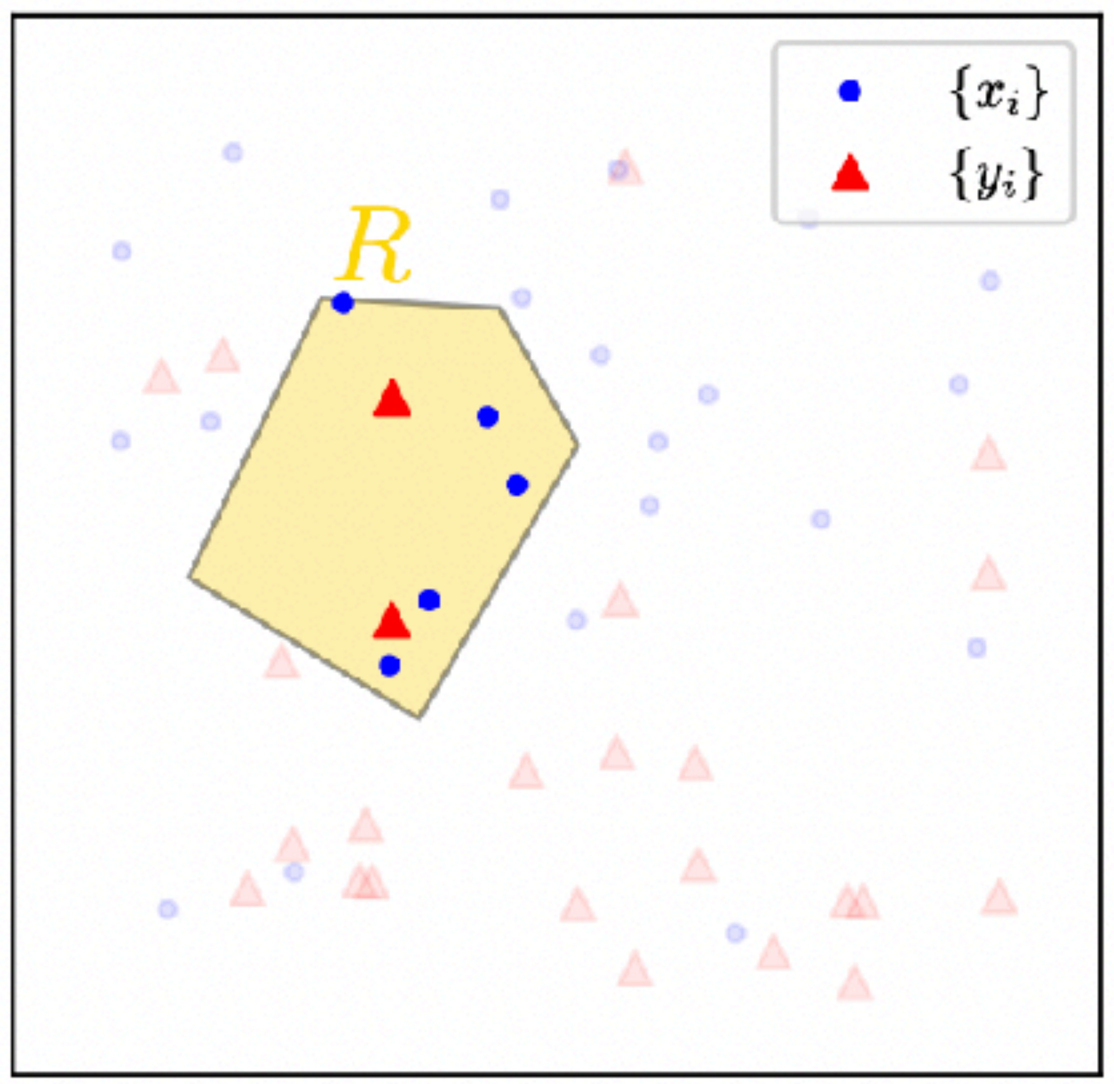
COVERAGE TEST FOR ACCURACY


Pablo Lemos



Lemos, et al. ICML 2023, 2302.03026

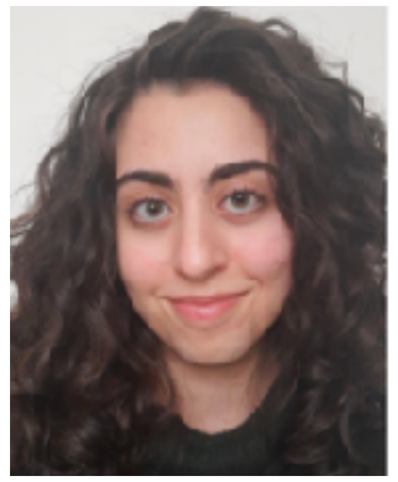
PQMass: Probabilistic assessment of Generative Models using Probability Mass Estimation



Sammy Sharief

Pablo Lemos

Kolya Malkin



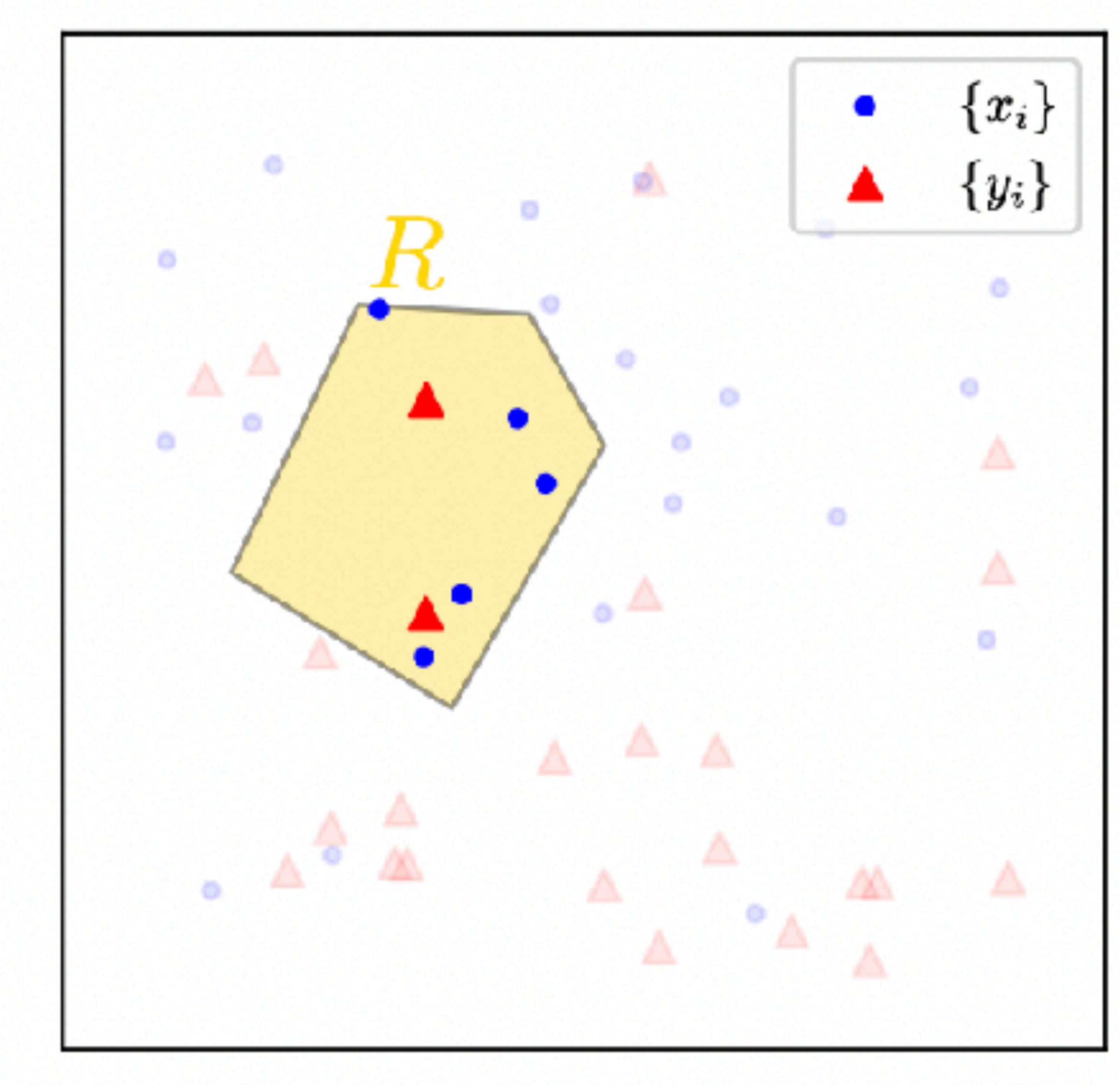
Connor Stone

Salma Salhi

pip install pqm

Lemos et al. ICLR 2025, 2402.04355

PQMass: Probabilistic assessment of Generative Models using Probability Mass Estimation



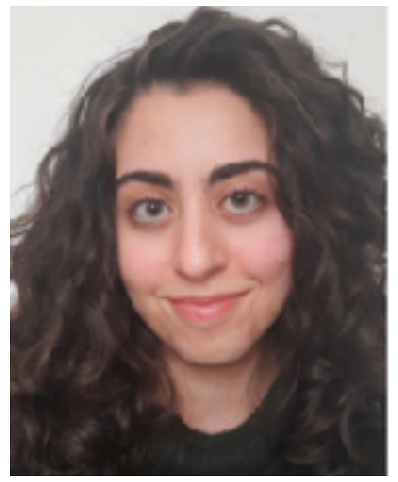
$$k(\mathbf{x}, R) \sim \mathcal{B}\left(n, \mathbb{P}_p(R)\right)$$


Sammy Sharief


Pablo Lemos


Kolya Malkin


Connor Stone


Salma Salhi

pip install pqm

Lemos et al. ICLR 2025, 2402.04355

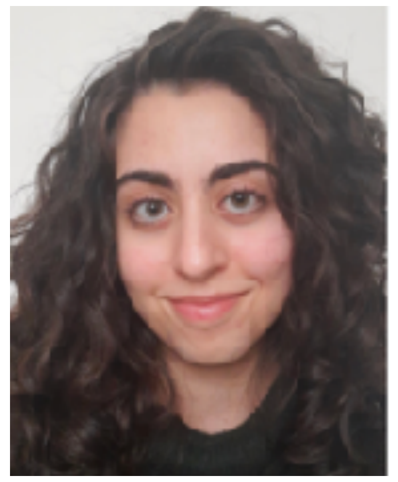
PQMass: Probabilistic assessment of Generative Models using Probability Mass Estimation

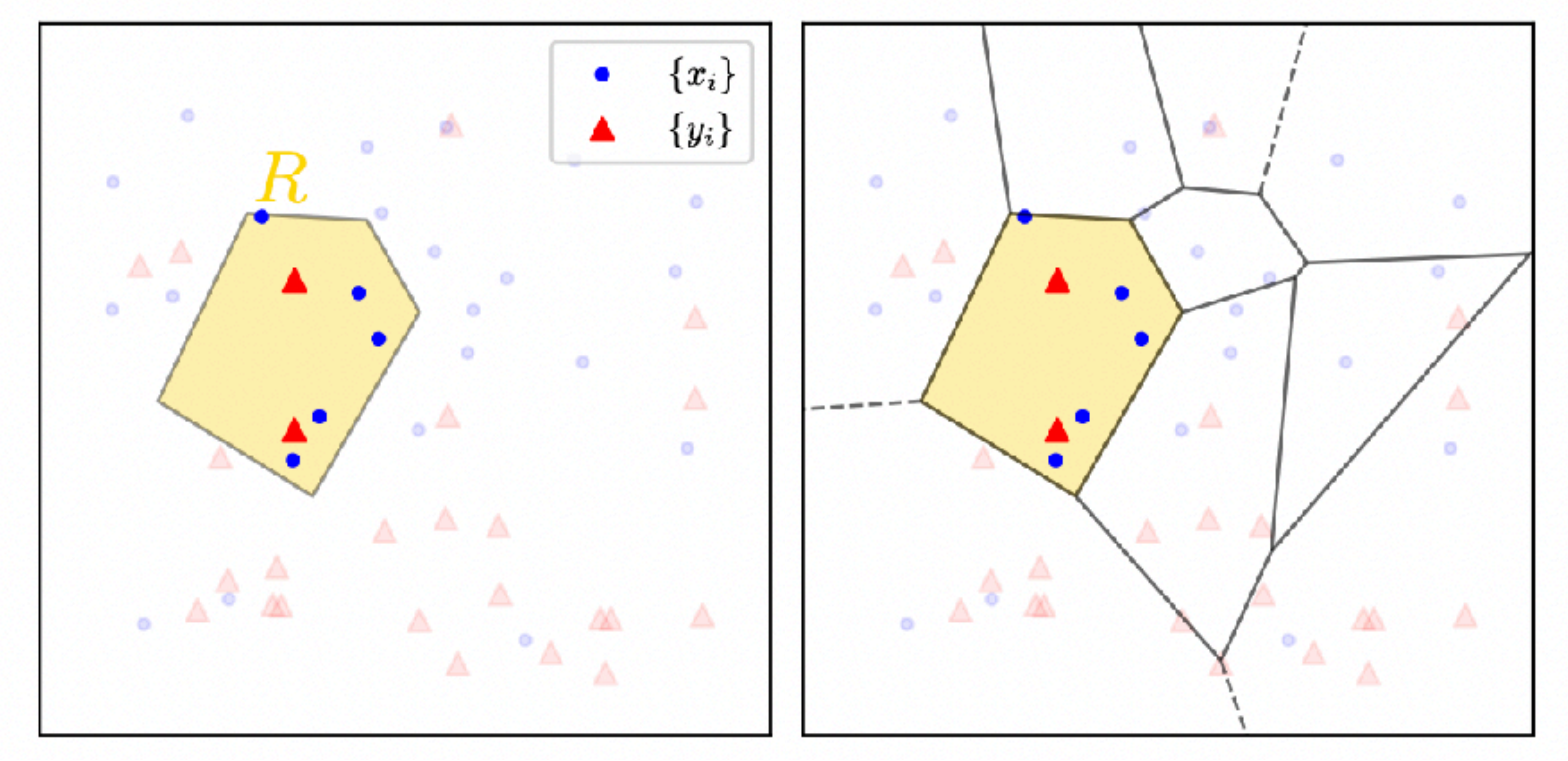

Sammy Sharief


Pablo Lemos


Kolya Malkin


Connor Stone


Salma Salhi



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

pip install pqm

Lemos et al. ICLR 2025, 2402.04355

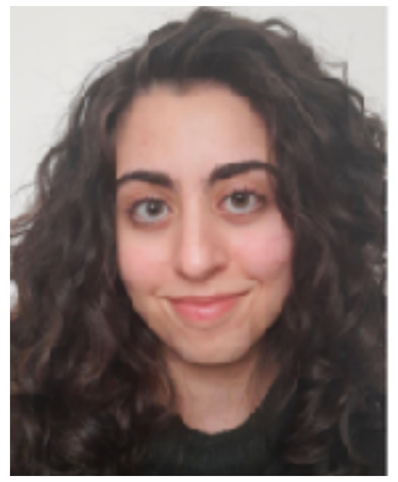
PQMass: Probabilistic assessment of Generative Models using Probability Mass Estimation

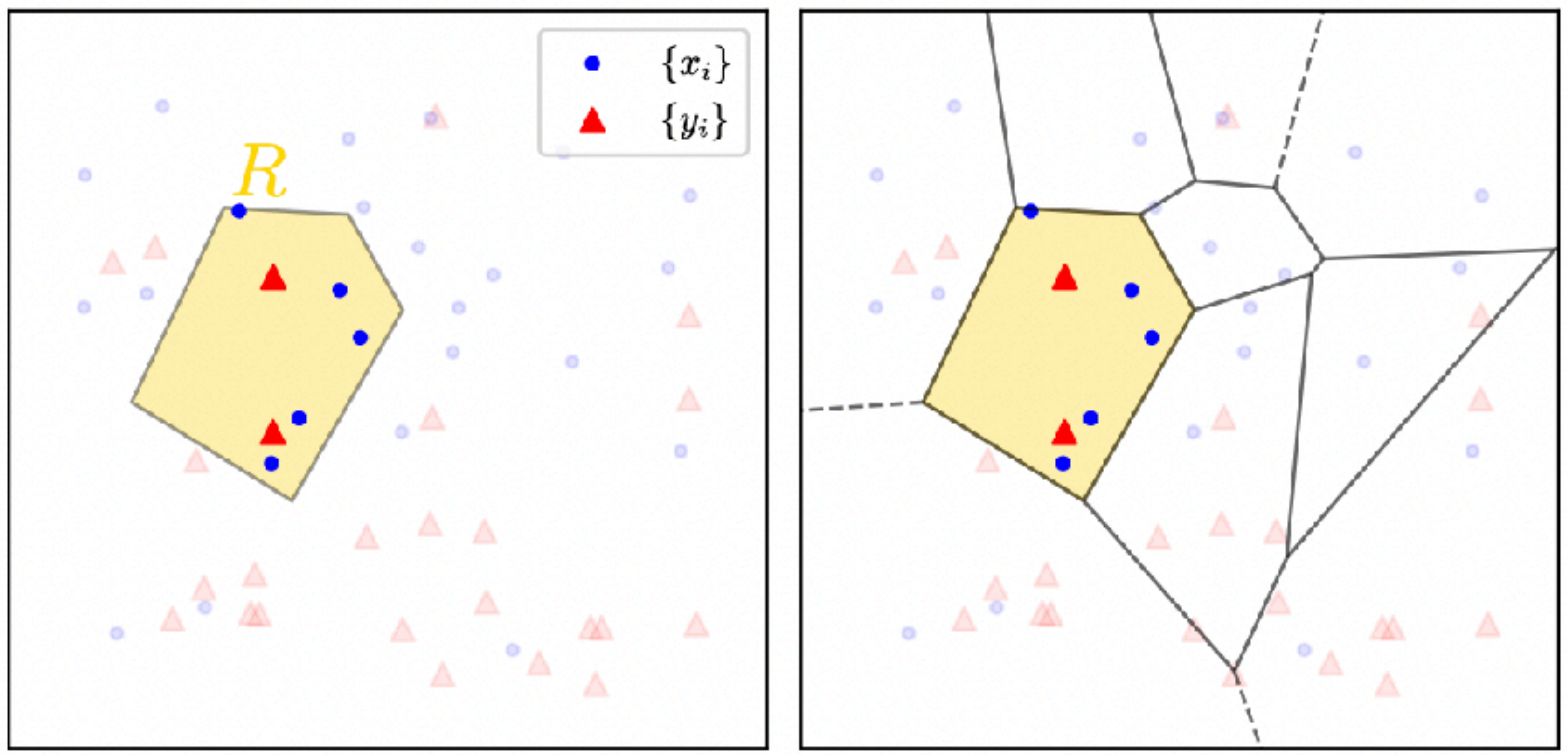

Sammy Sharief


Pablo Lemos


Kolya Malkin


Connor Stone


Salma Salhi



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

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

pip install pqm

Lemos et al. ICLR 2025, 2402.04355

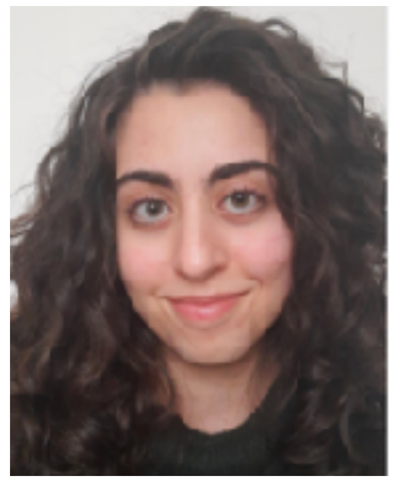
PQMass: Probabilistic assessment of Generative Models using Probability Mass Estimation

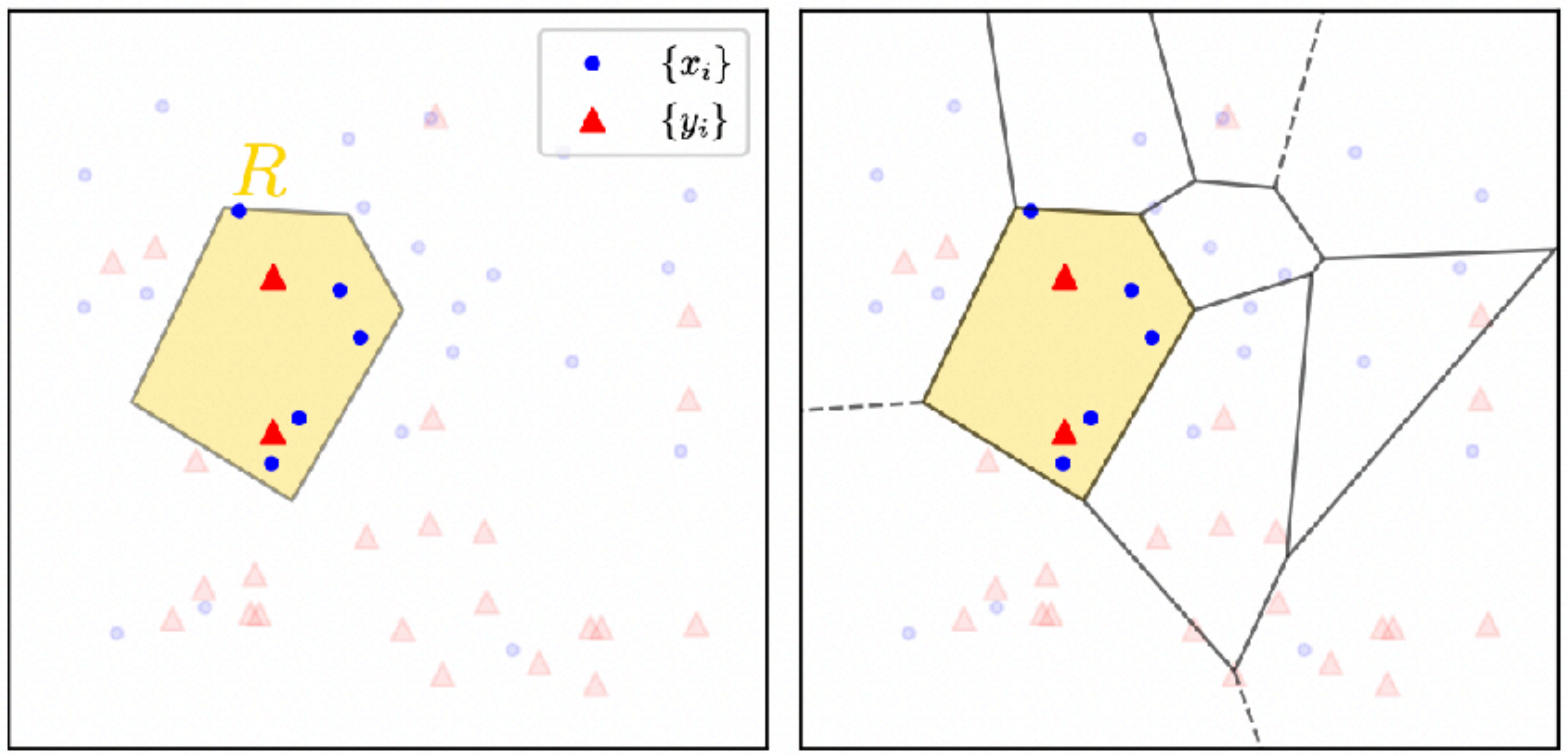

Sammy Sharief


Pablo Lemos


Kolya Malkin


Connor Stone


Salma Salhi



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

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

pip install pqm

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

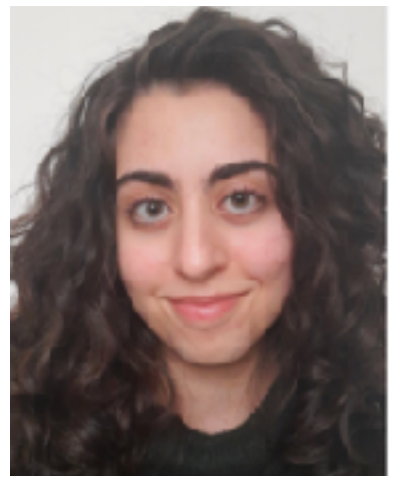
Lemos et al. ICLR 2025, 2402.04355

PQMass: Probabilistic assessment of Generative Models using Probability Mass Estimation

Sammy Sharief

Pablo Lemos

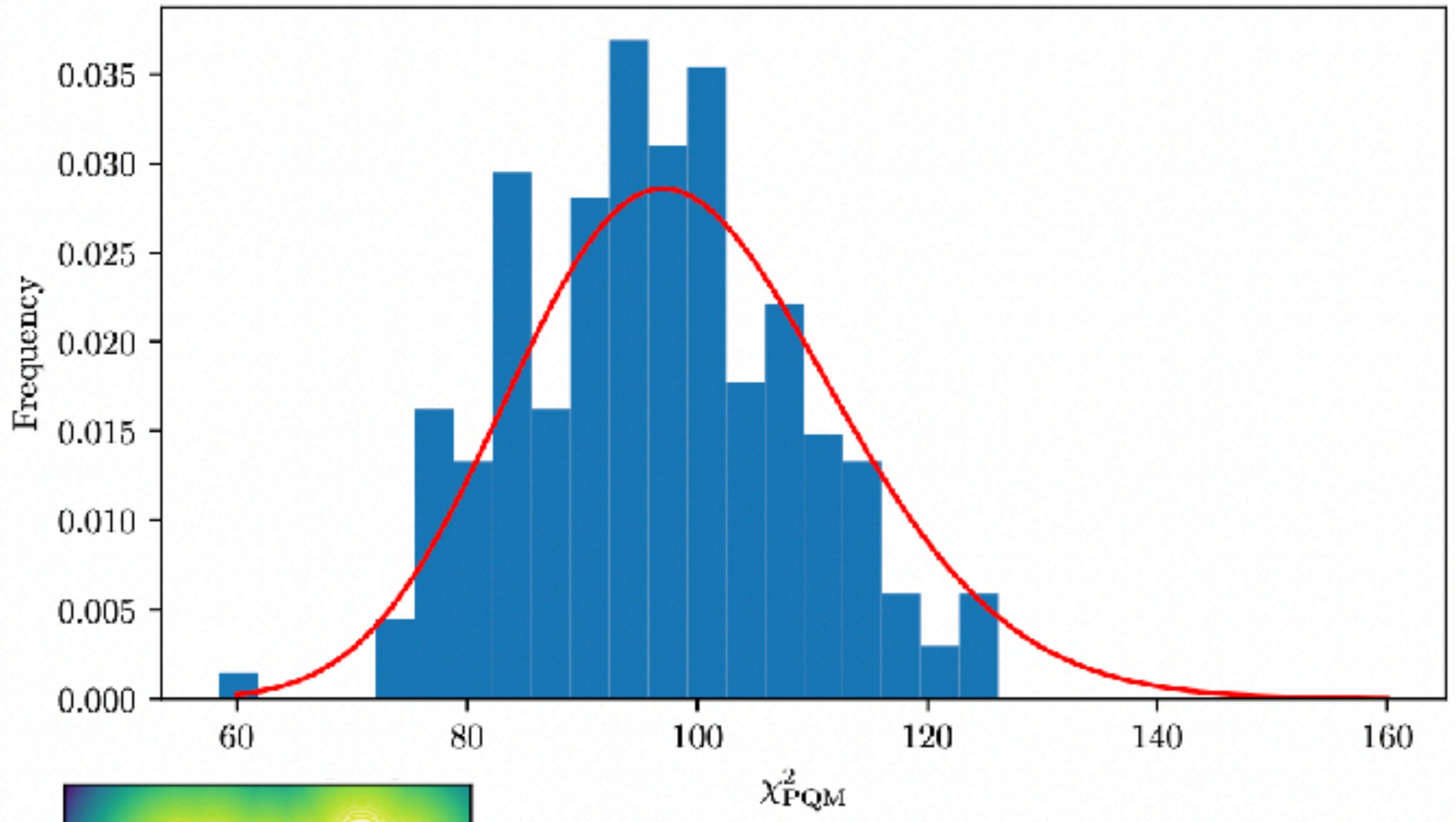
Kolya Malkin

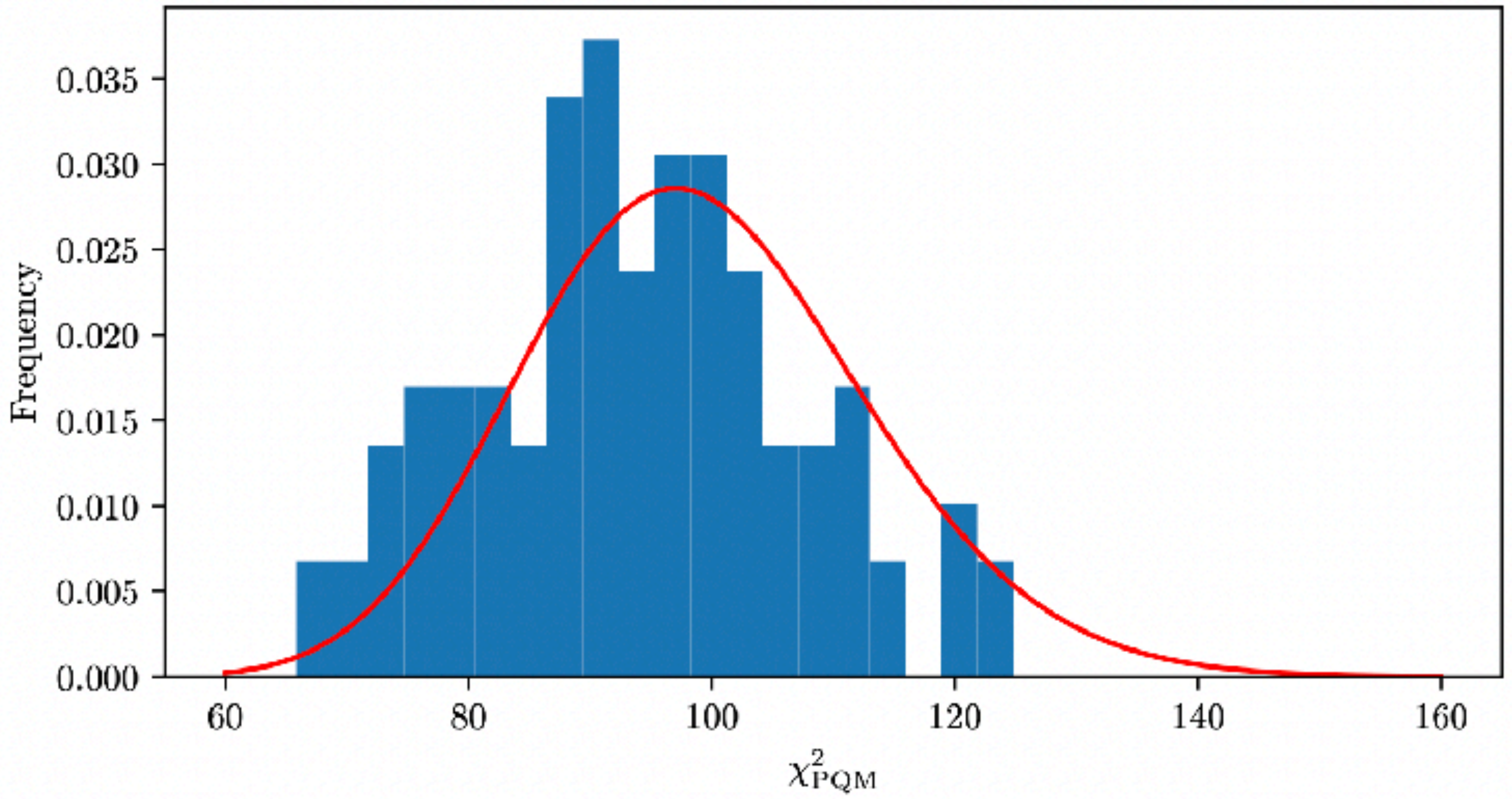


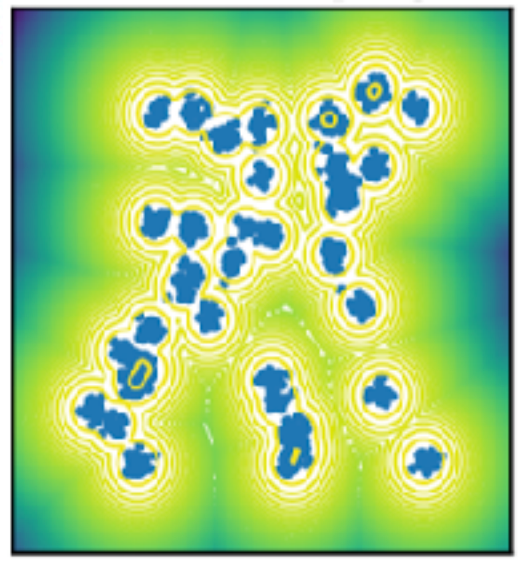
Connor Stone

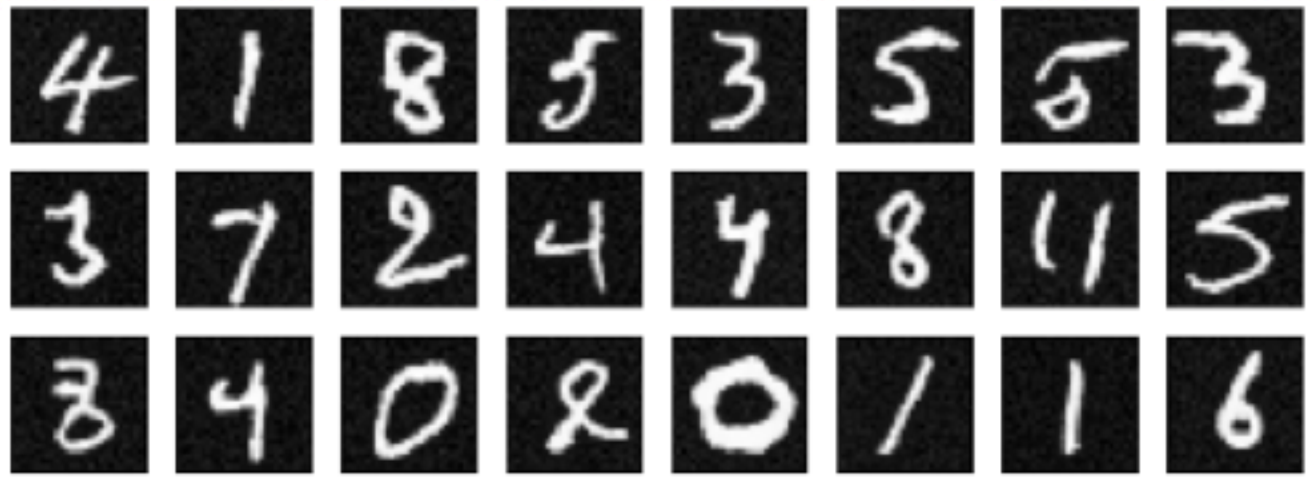
Salma Salhi

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.*









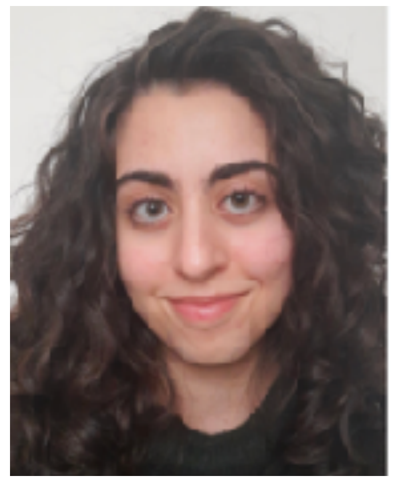
PQMass: Probabilistic assessment of Generative Models using Probability Mass Estimation

Sammy Sharief

Pablo Lemos

Kolya Malkin

Connor Stone



Salma Salhi

Power of PQMass:

- Works directly in pixel space (or other data space) without relying on dimensionality reduction

- Only relies on defining a distance between points, so scales extremely well to high-dimensional spaces (>500,000 dimensions)

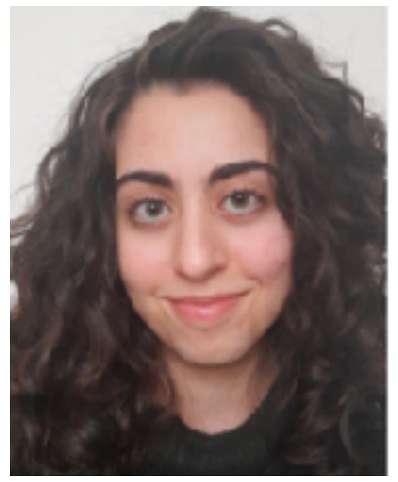
PQMass: Probabilistic assessment of Generative Models using Probability Mass Estimation

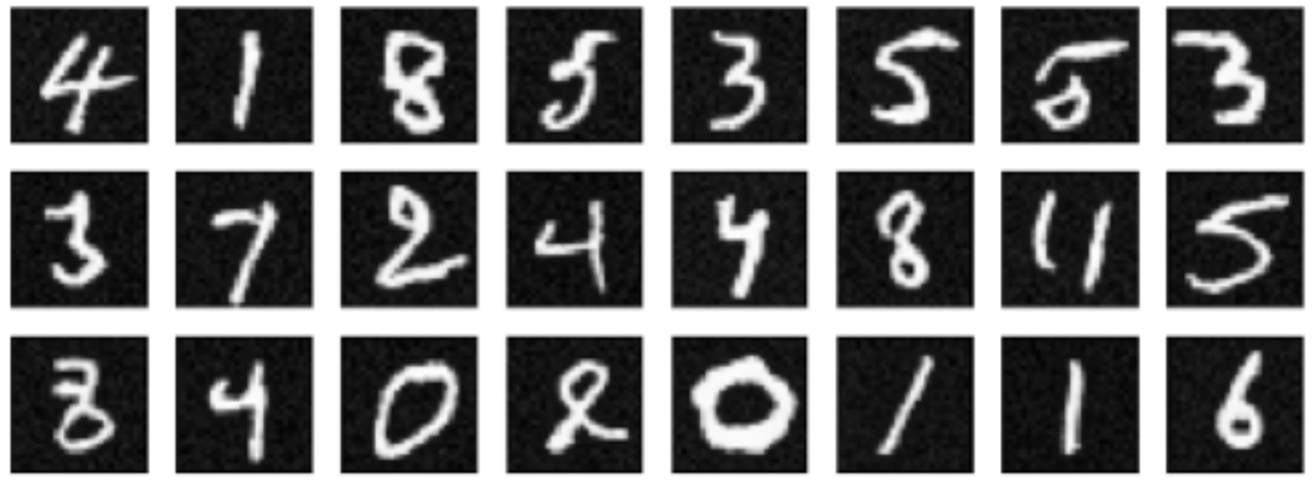

Sammy Sharief

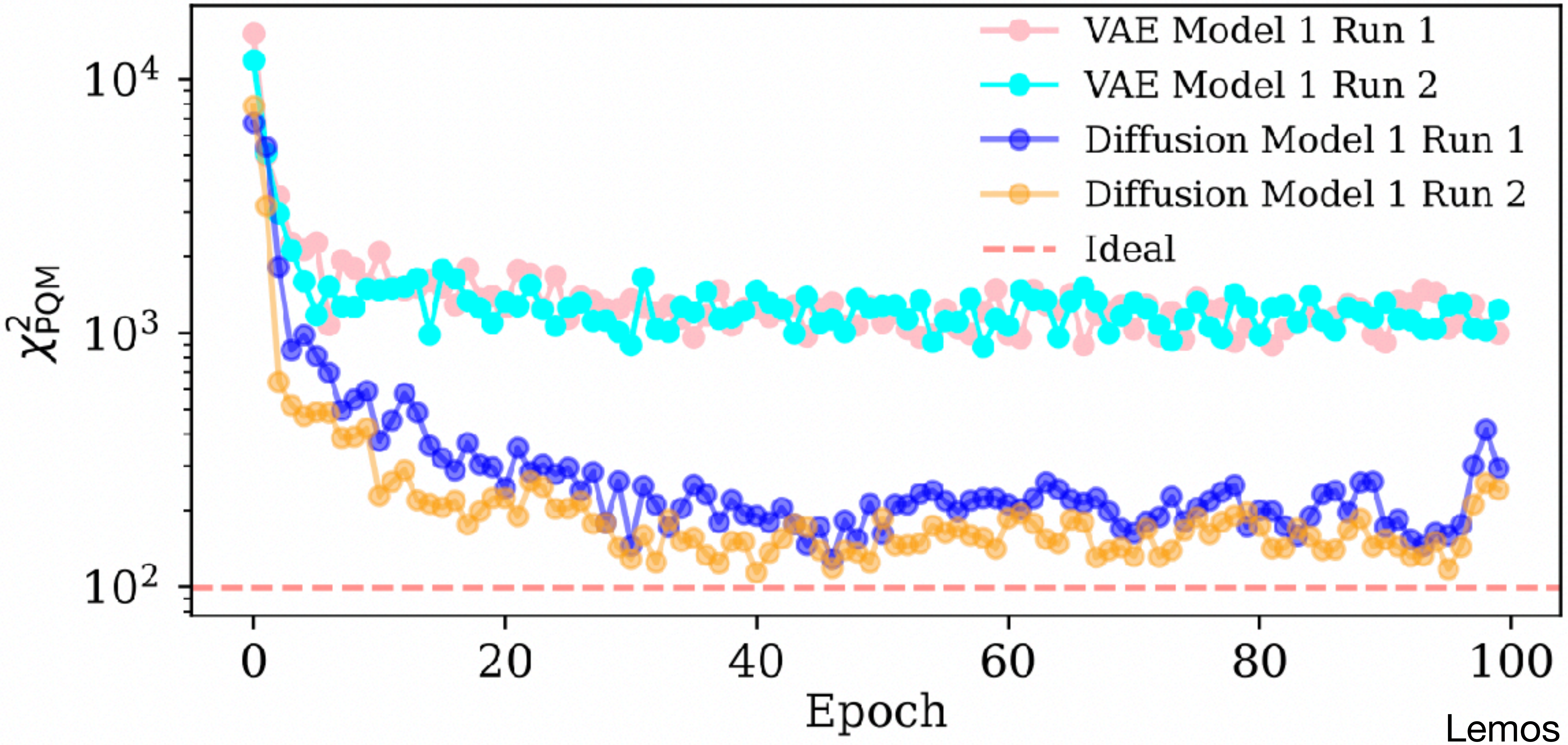

Pablo Lemos


Kolya Malkin


Connor Stone


Salma Salhi

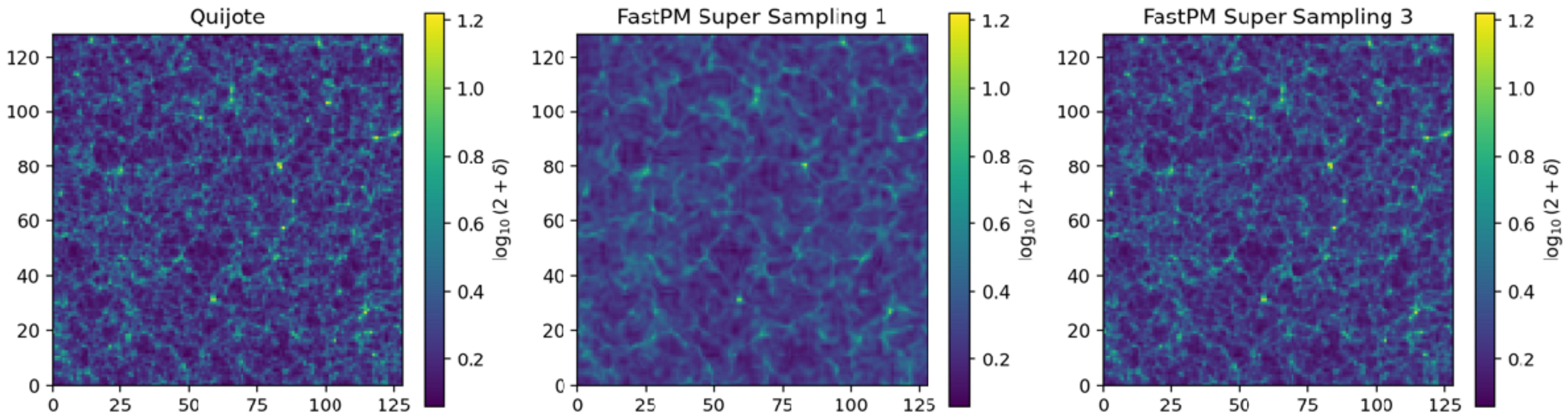


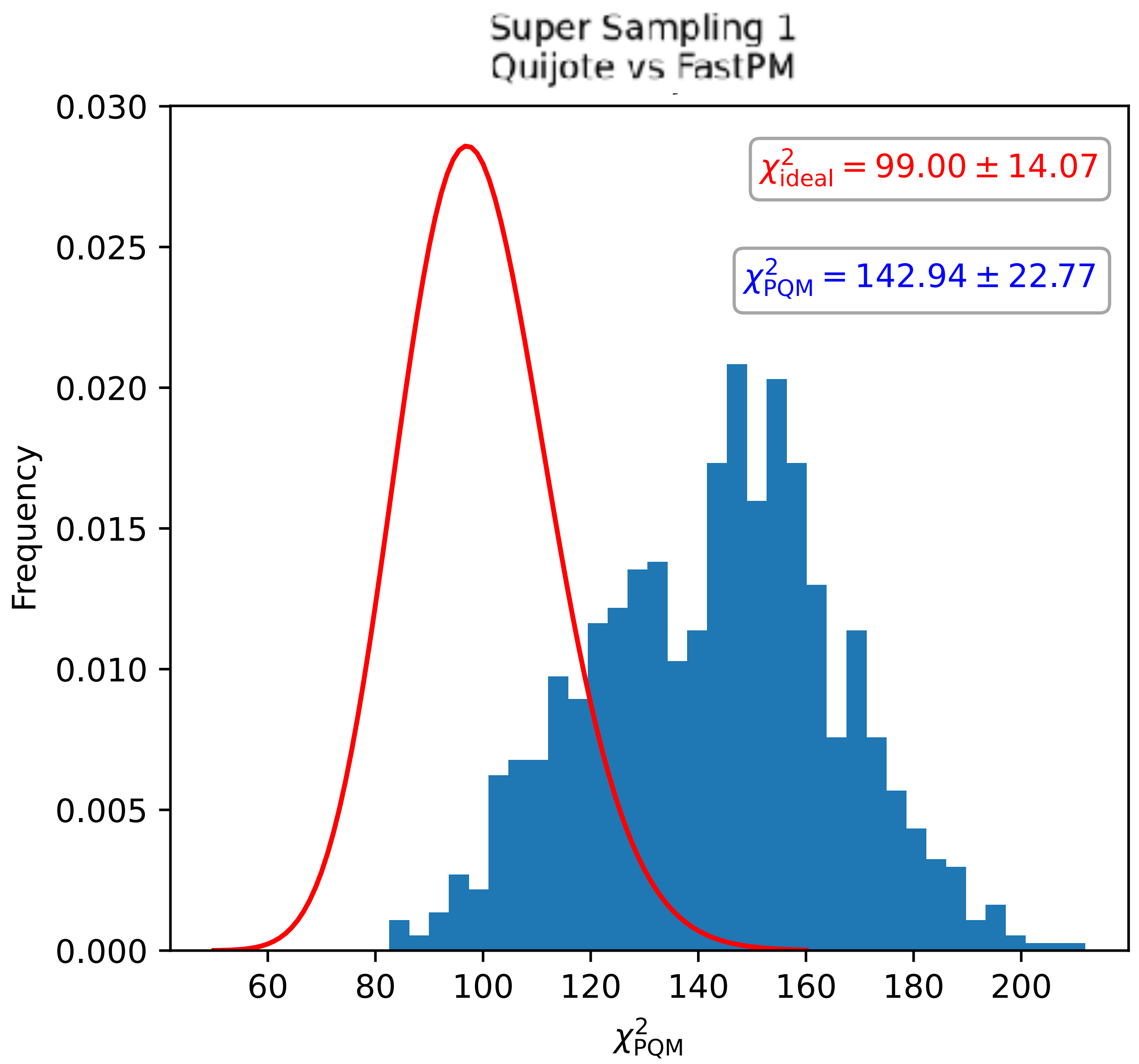


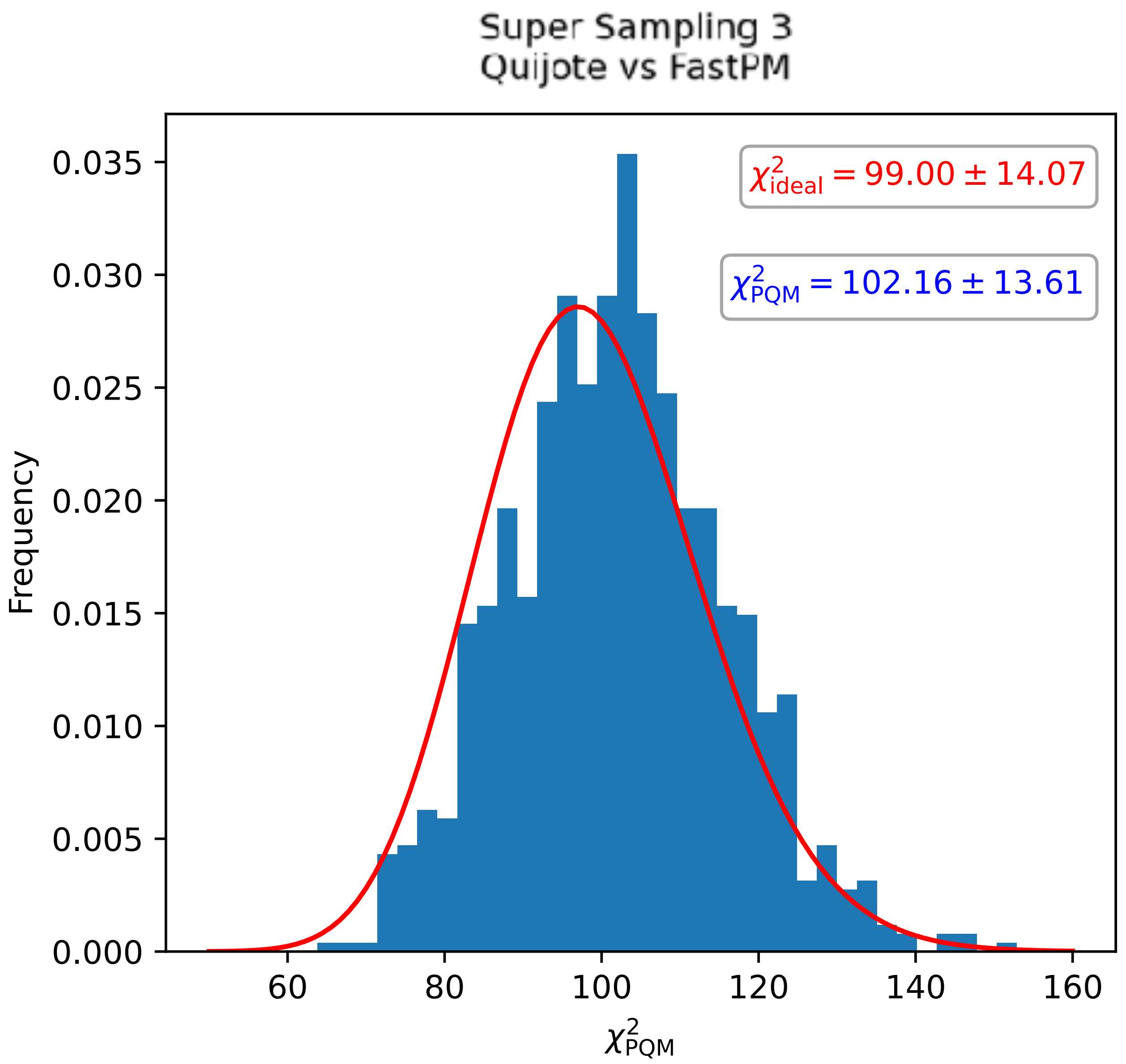
Lemos et al. ICLR 2025, 2402.04355

PQMass: Example of application to test accuracy of Simulators and emulators at the field level

Sammy Sharief





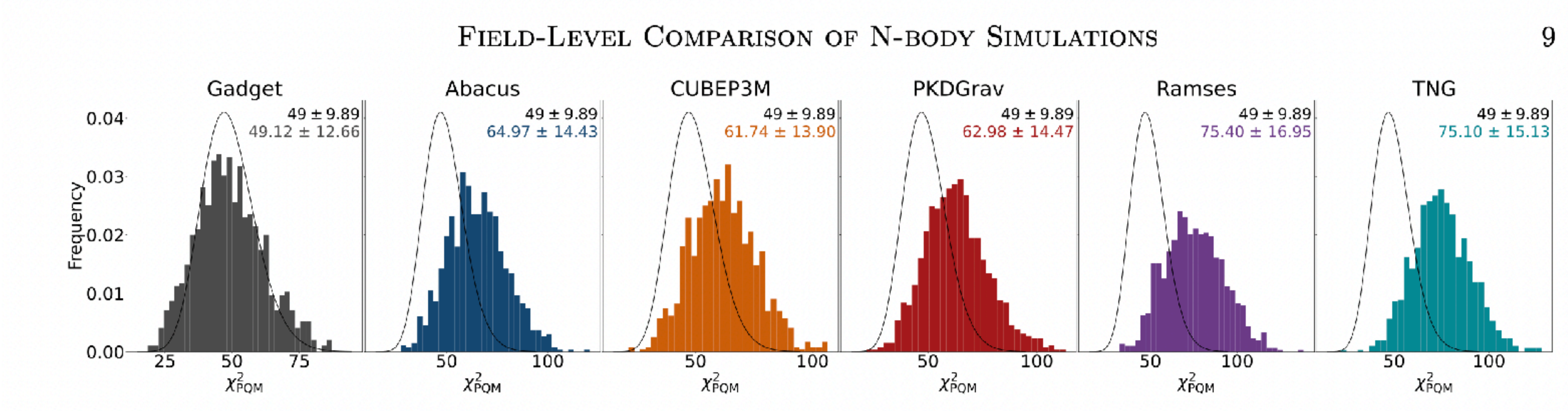


Lemos et al. ICLR 2025, 2402.04355

PQMass: Example of application to test accuracy of simulators and emulators at the field level

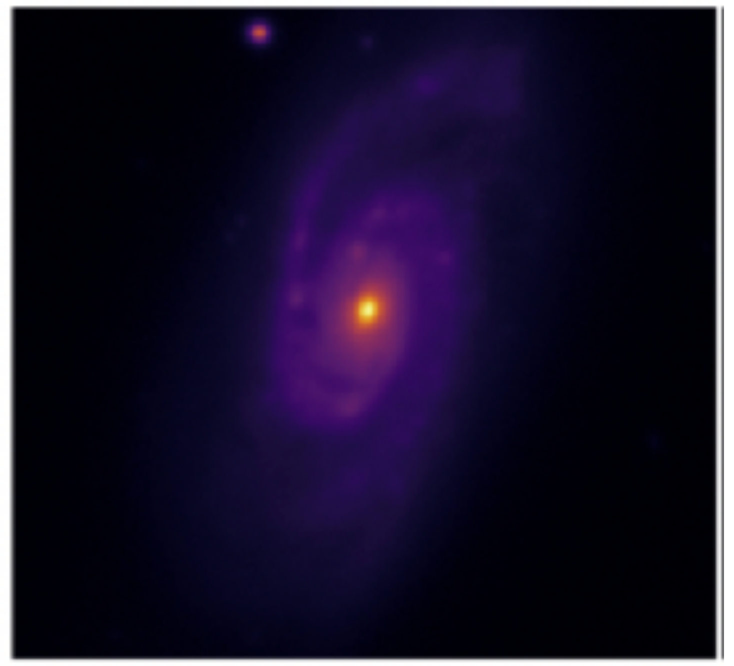
Adrian Bayer

Sammy Sharief



A. Bayer et al. ApJ2025, 2505.13620

PQMass: Can be applied to data of varied modality

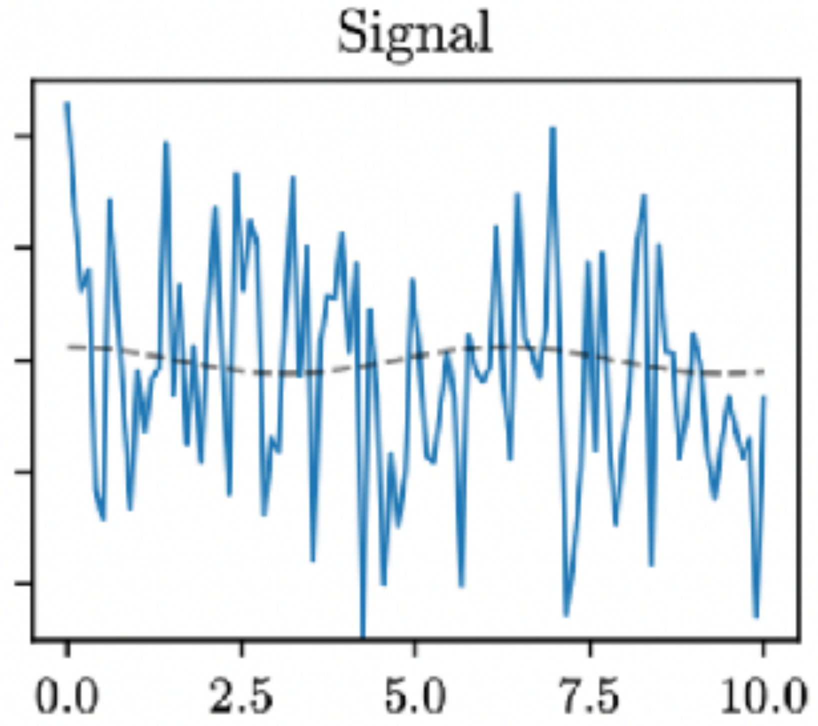


Images

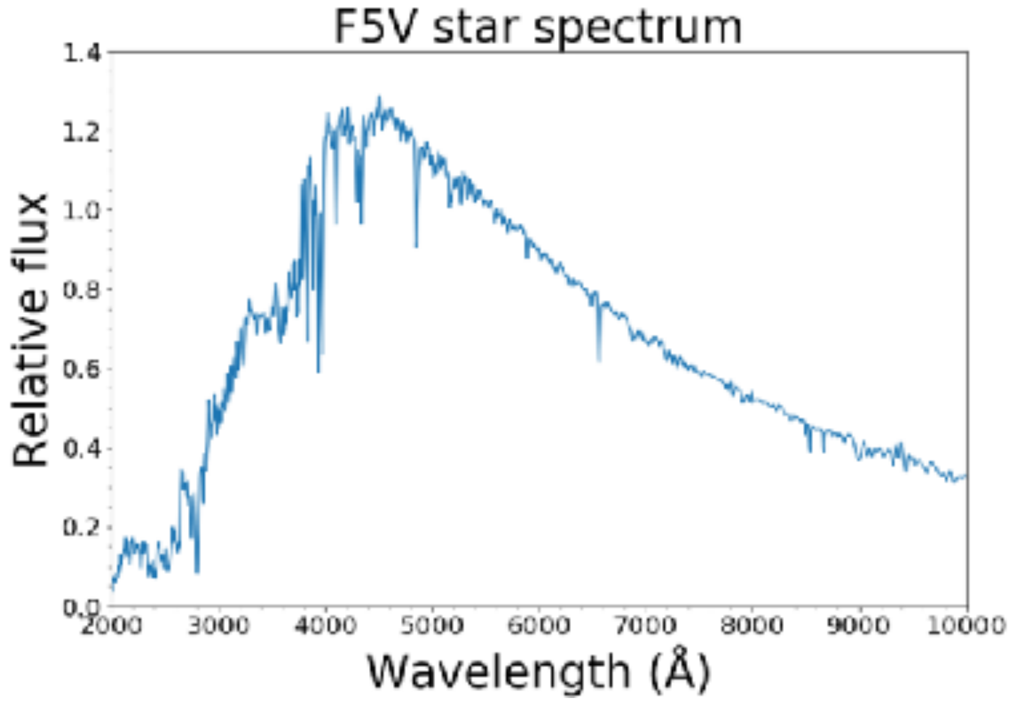
Variables Table

Variable Name	Role	Type	Demographic	Description
age	Feature	Integer	Age	N/A
workcass	Feature	Categorical	Income	Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, Never-worked.
fnlwgt	Feature	Integer		
education	Feature	Categorical	Education Level	Bachelors, Some-college, 11th, HS-grad, Prof-school, Assoc-acdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate, 5th-6th, Preschool.
education-num	Feature	Integer	Education Level	

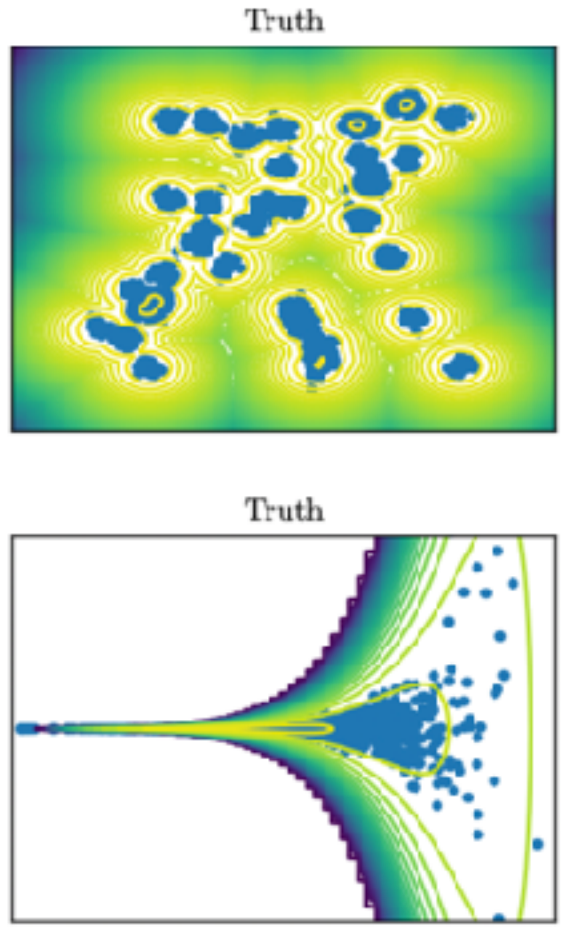
Tabular data



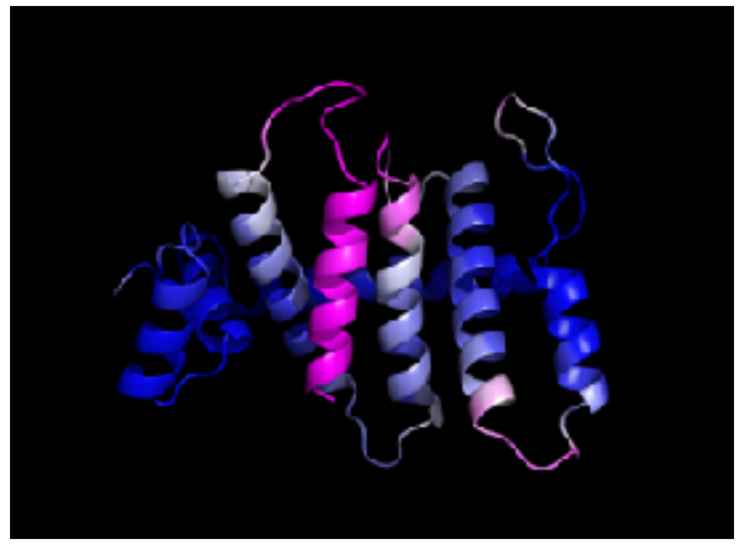
Time series data



Spectra

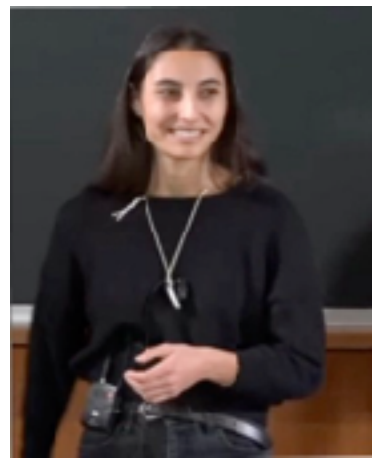


Low-dimensional parametric distributions

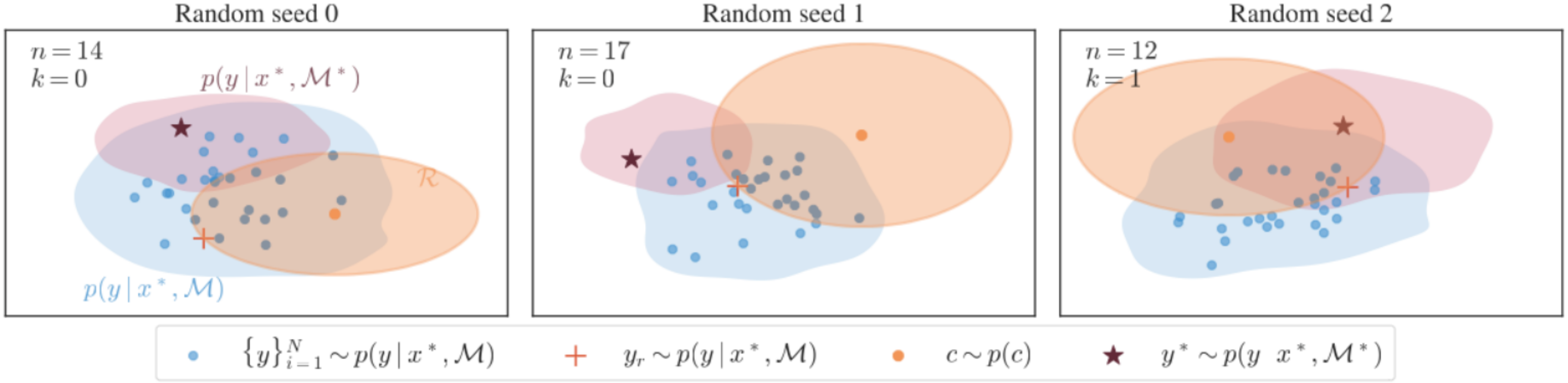


Protein sequences

Mira: PQMass for Posterior Inference and conditional models transforming TARP into an accuracy metric

Sammy Sharief
Justine Zeghal



$$p(k \mid n) = \frac{n+1}{N+2}\mathbb{1}(k=1) + \frac{N-n+1}{N+2}\mathbb{1}(k=0)$$

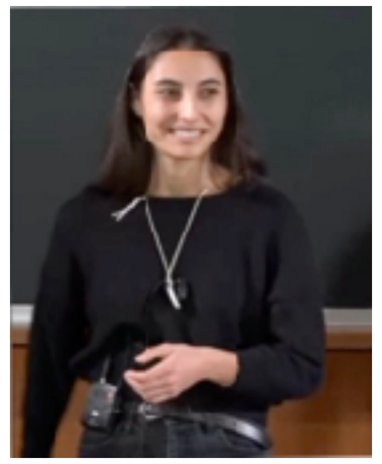
$$\mu_{\text{Mira}}(\mathcal{M}) = \mathbb{E}_{p(Z)}\left[\mathbb{E}_{p(k,n|Z)}\left[p(k \mid n)\right]\right]$$

$$\mathbb{E}_{p(k,n)}\left[P_N\right] \to \frac{2}{3} \quad as\ N \to \infty$$

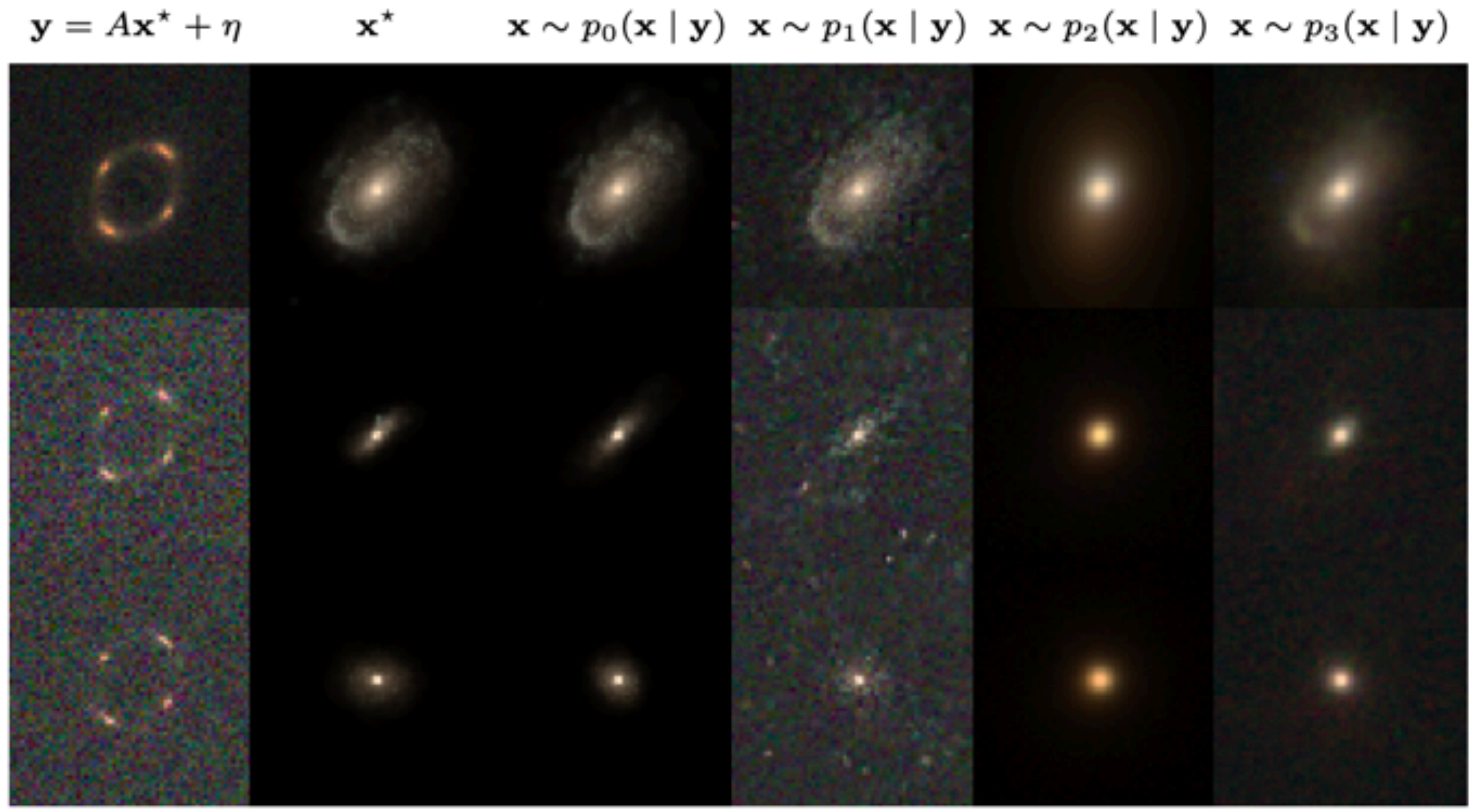
$$Var_{p(k,n)}\left[P_N\right] \to \frac{1}{18} \quad as\ N \to \infty$$

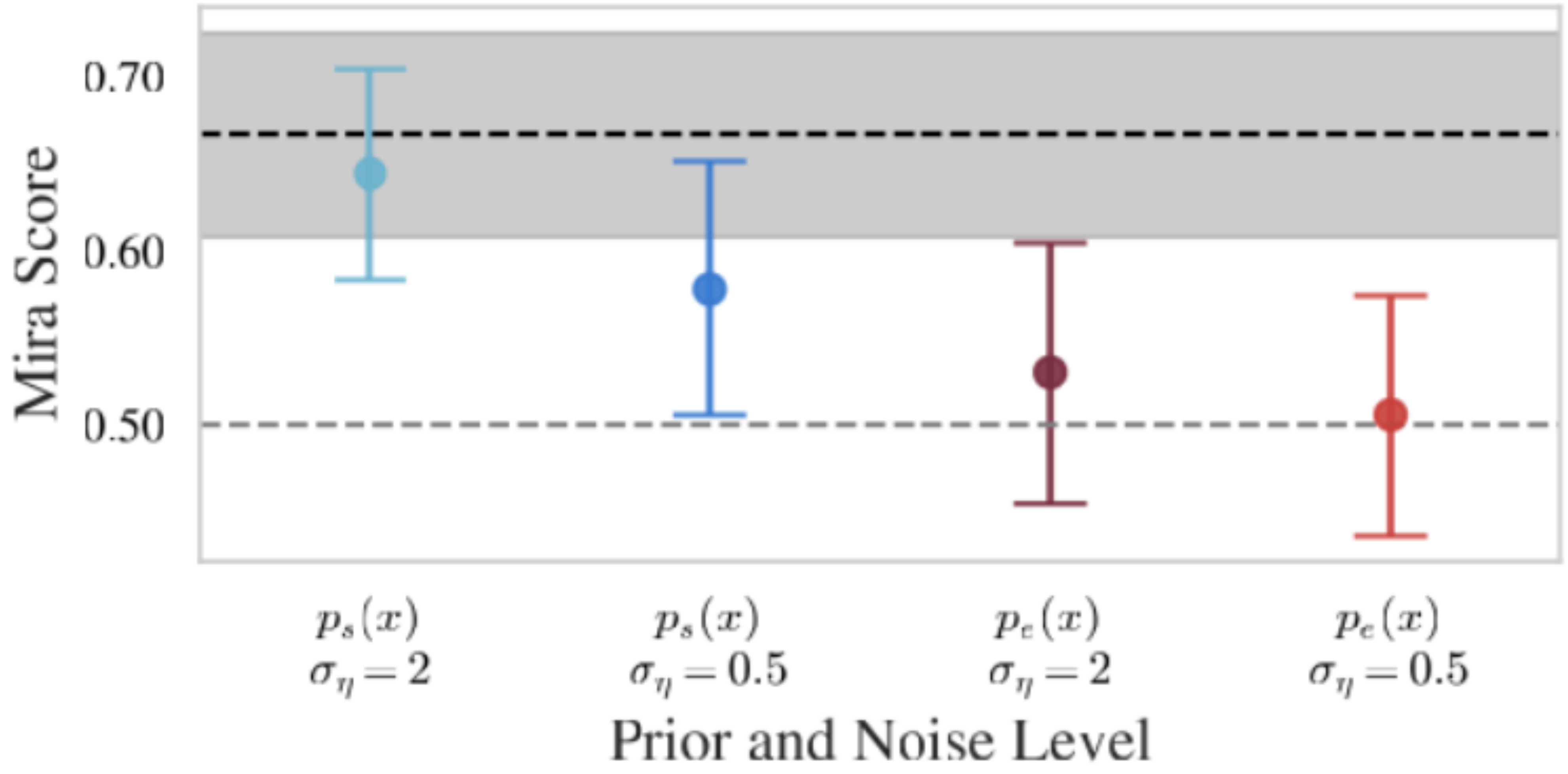
2605.02014, ICML 2026

MIRA (Mass In Random Areas): A Score for Conditional Distribution Accuracy and Model Comparison



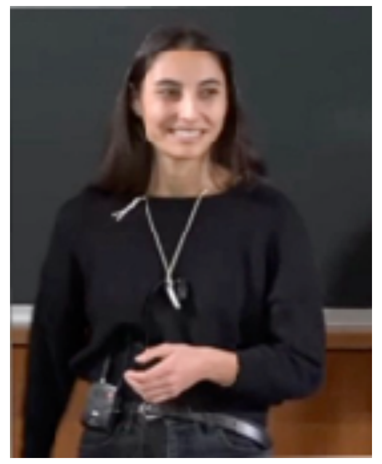
Sammy Shariel   Justine Zeghal



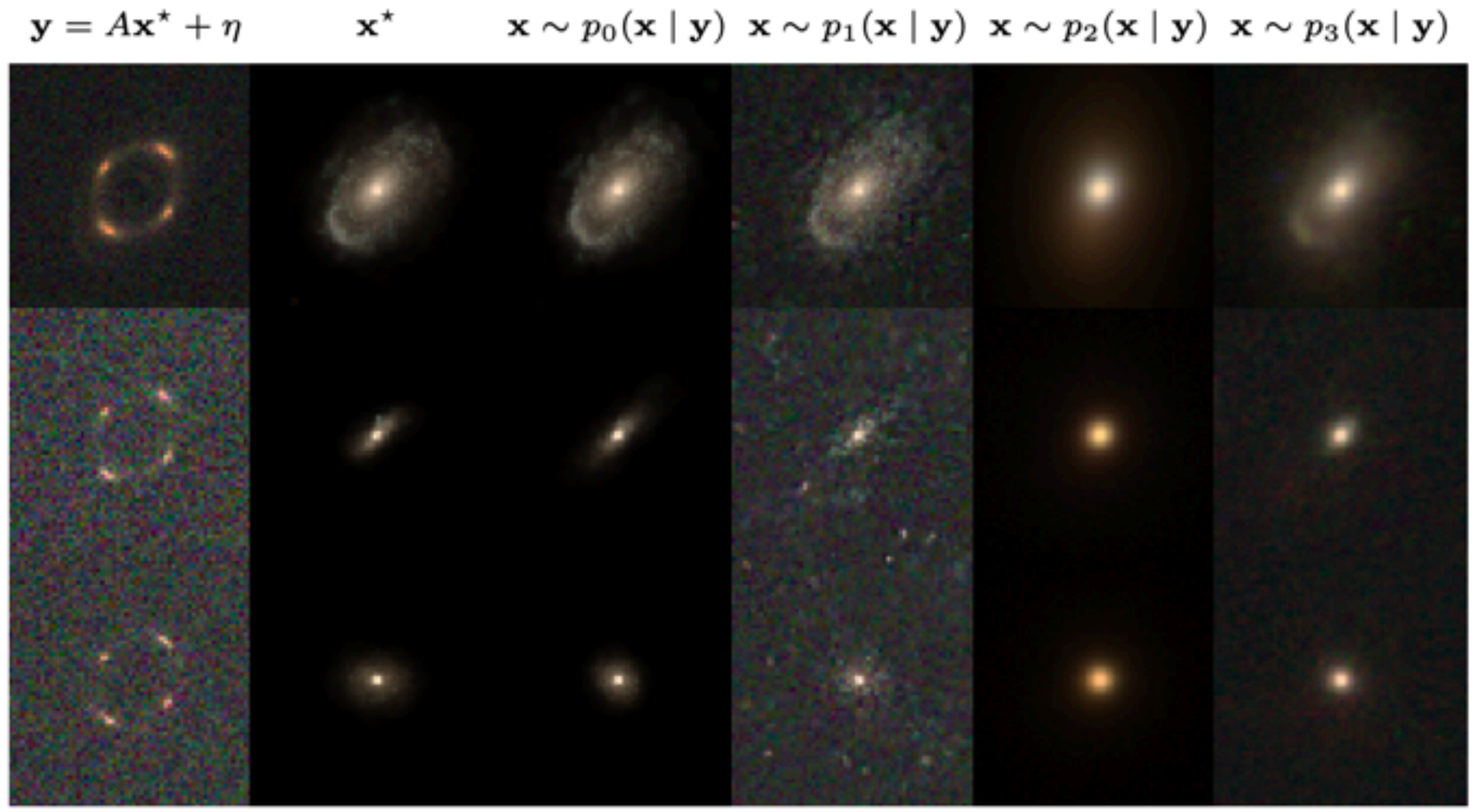


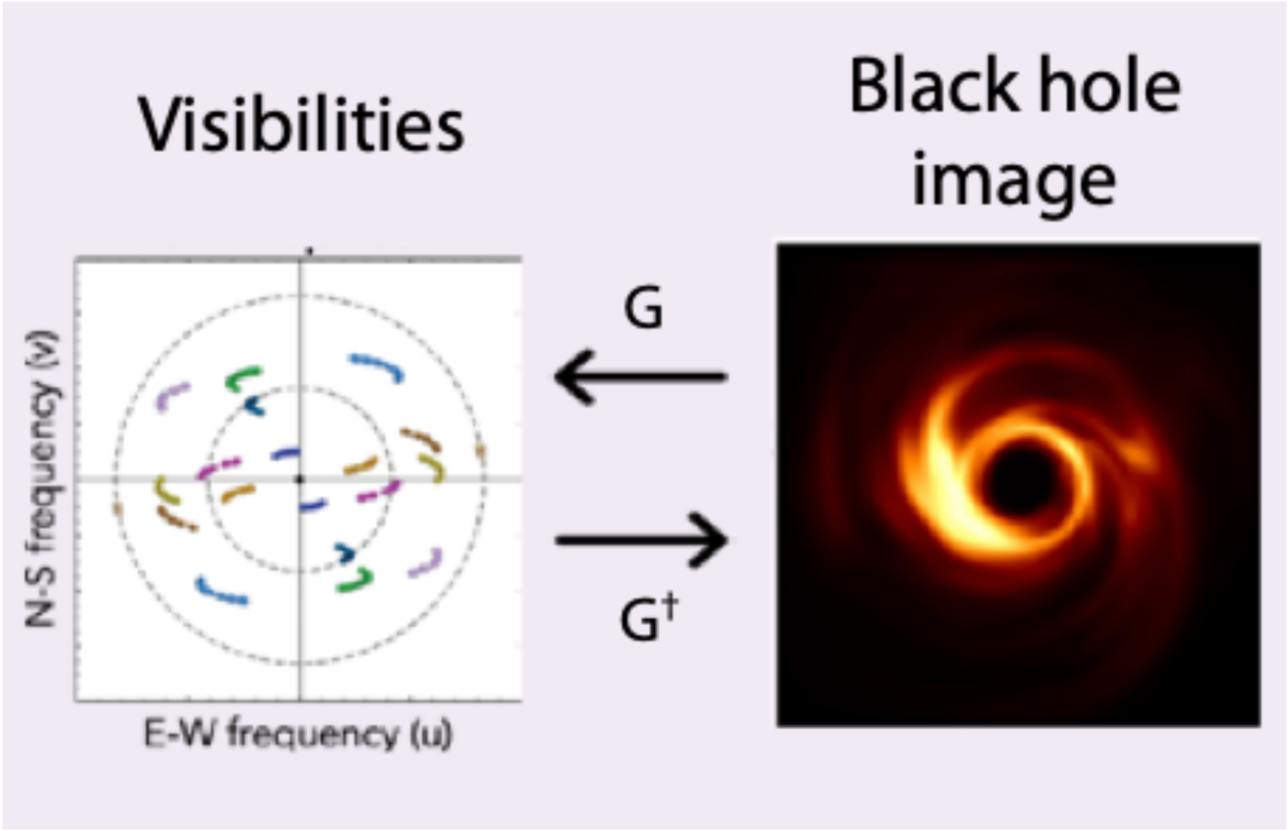
2605.02014, ICML 2026

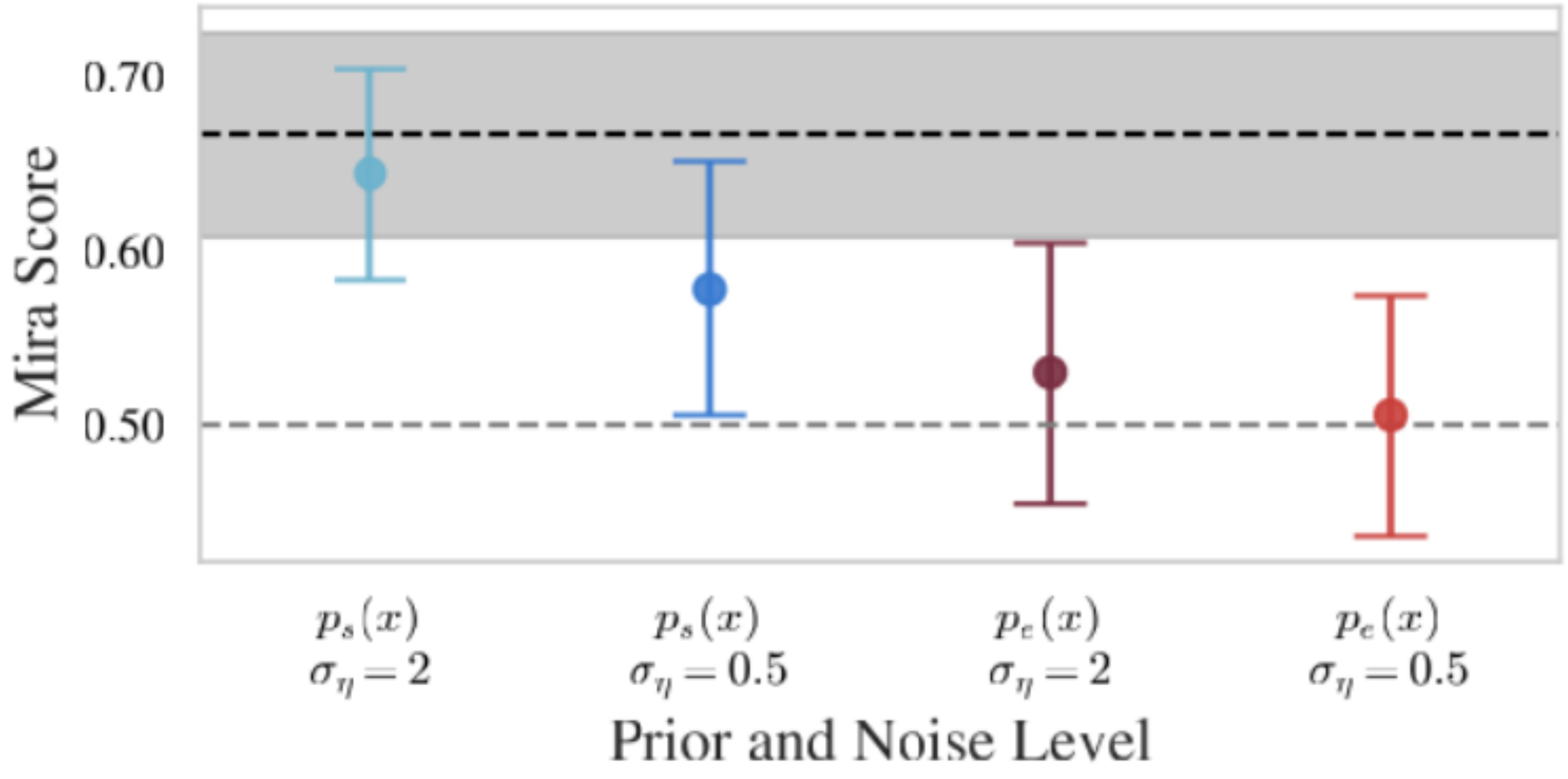
MIRA (Mass In Random Areas): A Score for Conditional Distribution Accuracy and Model Comparison

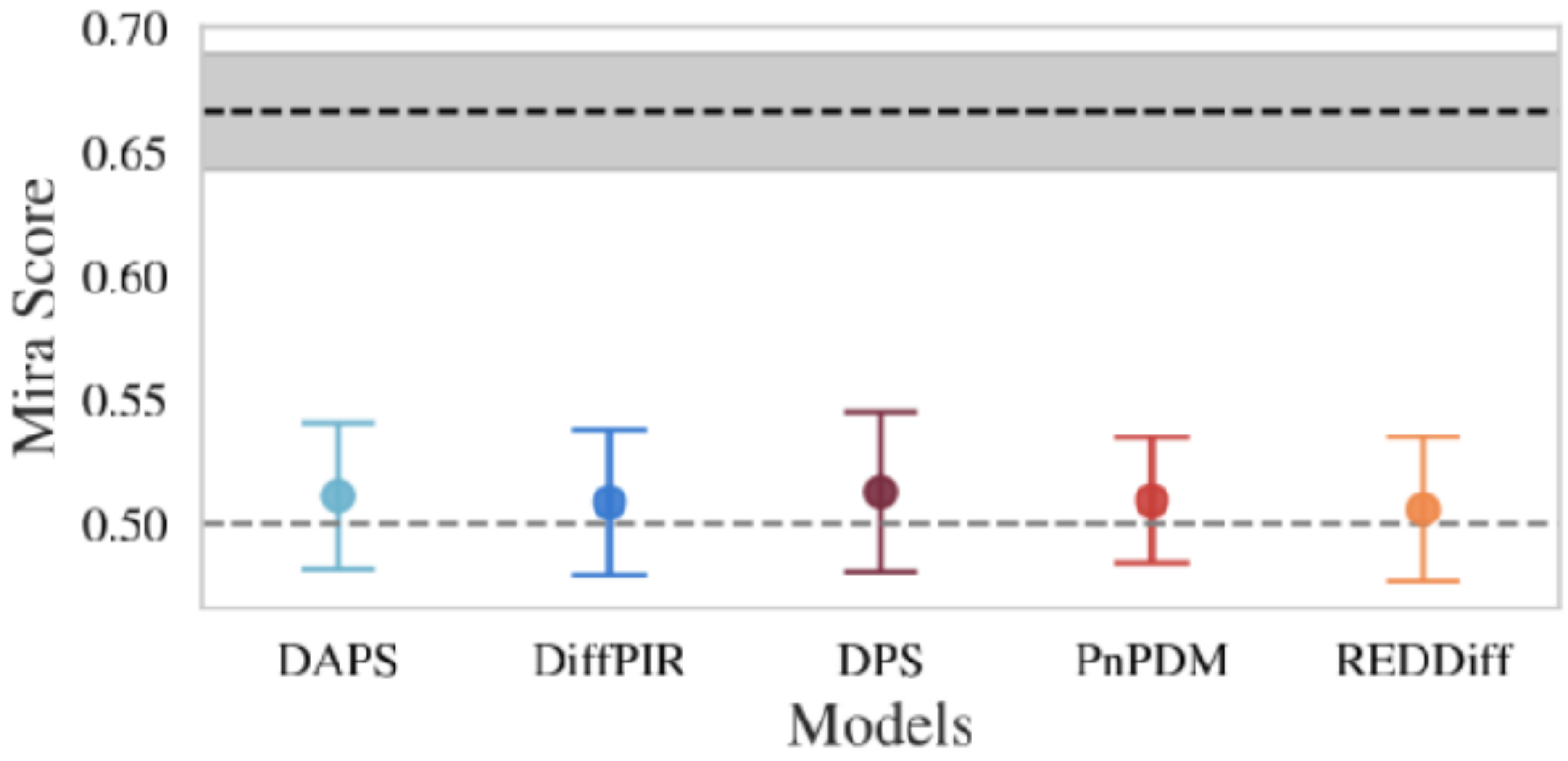



Sammy Sharief
Justine Zeghal









2605.02014, ICML 2026

GAME PLAN

1. Data-driven priors

2. Data-driven likelihoods

3. Accuracy metrics

4. Out-of-Distribution accuracy

Addressing Distributional Shifts: Prior Misspecification

So far, we've used **data** (corrupted by noise, psf, etc.) or **simulations** (potentially out of distribution) to learn our prior distributions.

Should we worry?

Addressing Distributional Shifts: Prior Misspecification

So far, we've used **data** (corrupted by noise, psf, etc.) or **simulations** (potentially out of distribution) to learn our prior distributions.

Should we worry?

We will soon be in the regime where we have $\mathcal{O}(10^5)$ strong lenses.
PQMass can allow us to detect if our posterior reconstructions are significantly shifted with respect to our prior
=> we can detect if the prior is out of distribution even in very high dimensional spaces.

Addressing Distributional Shifts: Prior Misspecification

So far, we've used **data** (corrupted by noise, psf, etc.) or **simulations** (potentially out of distribution) to learn our prior distributions.

Should we worry?

We will soon be in the regime where we have $\mathcal{O}(10^5)$ strong lenses.
PQMass can allow us to detect if our posterior reconstructions are significantly shifted with respect to our prior => we can detect if the prior is out of distribution even in very high dimensional spaces.

Can we use some of our observations to learn the prior hierarchically?

ADDRESSING DISTRIBUTIONAL SHIFTS: PRIOR MISSPECIFICATION

Missa
Barco

What we would like to do:

$$P(\theta | d_{1,\ldots,i}) \propto \Pi_i \int P(d_i | x_i) p(x_i | \theta) p(\theta) dx$$

Barco et al. 2407.17667

Addressing Distributional Shifts: Prior Misspecification

Missa Barco

What we would like to do:

$$P(\theta \mid d_{1,\ldots,i}) \propto \Pi_i \int P(d_i \mid x_i) p(x_i \mid \theta) p(\theta) dx$$

Where θ parametrize our prior (in our case the score-based model parameters).

However, in our case, evaluating $p(x_i \mid \theta)$ is not straightforward.

Barco et al. 2407.17667

ADDRESSING DISTRIBUTIONAL SHIFTS: PRIOR MISSPECIFICATION

Missa Barco

What we would like to do:

$$P(\theta | d_{1,\ldots,i}) \propto \Pi_i \int P(d_i | x_i) p(x_i | \theta) p(\theta) dx$$

Where θ parametrize our prior (in our case the score-based model parameters). However, in our case, evaluating $p(x_i | \theta)$ is not straightforward.

What we can do instead:

Barco et al. 2407.17667

Addressing Distributional Shifts: Prior Misspecification

Missa Barco

What we would like to do:

$$P(\theta | d_{1,...,i}) \propto \Pi_i \int P(d_i | x_i) p(x_i | \theta) p(\theta) dx$$

Where θ parametrize our prior (in our case the score-based model parameters). However, in our case, evaluating $p(x_i | \theta)$ is not straightforward.

What we can do instead:

$$\hat{\theta} = \arg\max_{\theta \in \Theta} p(\{\mathbf{y}_i\}_{i=1}^{N} | \theta)$$

Barco et al. 2407.17667

ADDRESSING DISTRIBUTIONAL SHIFTS: PRIOR MISSPECIFICATION

Missa Barco

What we would like to do:

$$P(\theta | d_{1,\ldots,i}) \propto \Pi_i \int P(d_i | x_i) p(x_i | \theta) p(\theta) dx$$

Where θ parametrize our prior (in our case the score-based model parameters).
However, in our case, evaluating $p(x_i | \theta)$ is not straightforward.

What we can do instead:

$$\hat{\theta} = \arg\max_{\theta \in \Theta} p(\{\mathbf{y}_i\}_{i=1}^N | \theta)$$

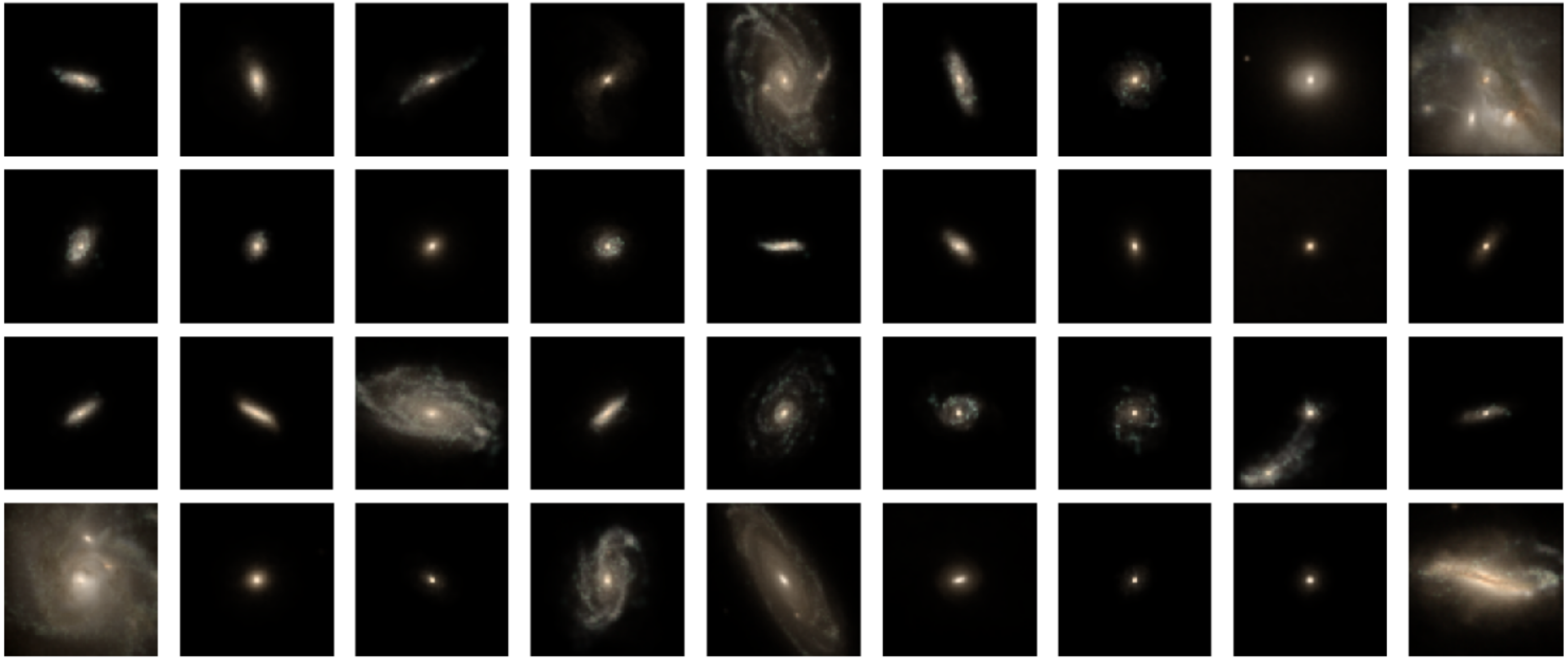
And we do this iteratively using posterior samples to retrain a new prior.

Barco et al. 2407.17667

Addressing Distributional Shifts: Prior Misspecification

Missa Barco

Applications to galaxies: can we learn that spiral galaxies exist starting from a prior with just (real, noisy) ellipticals?



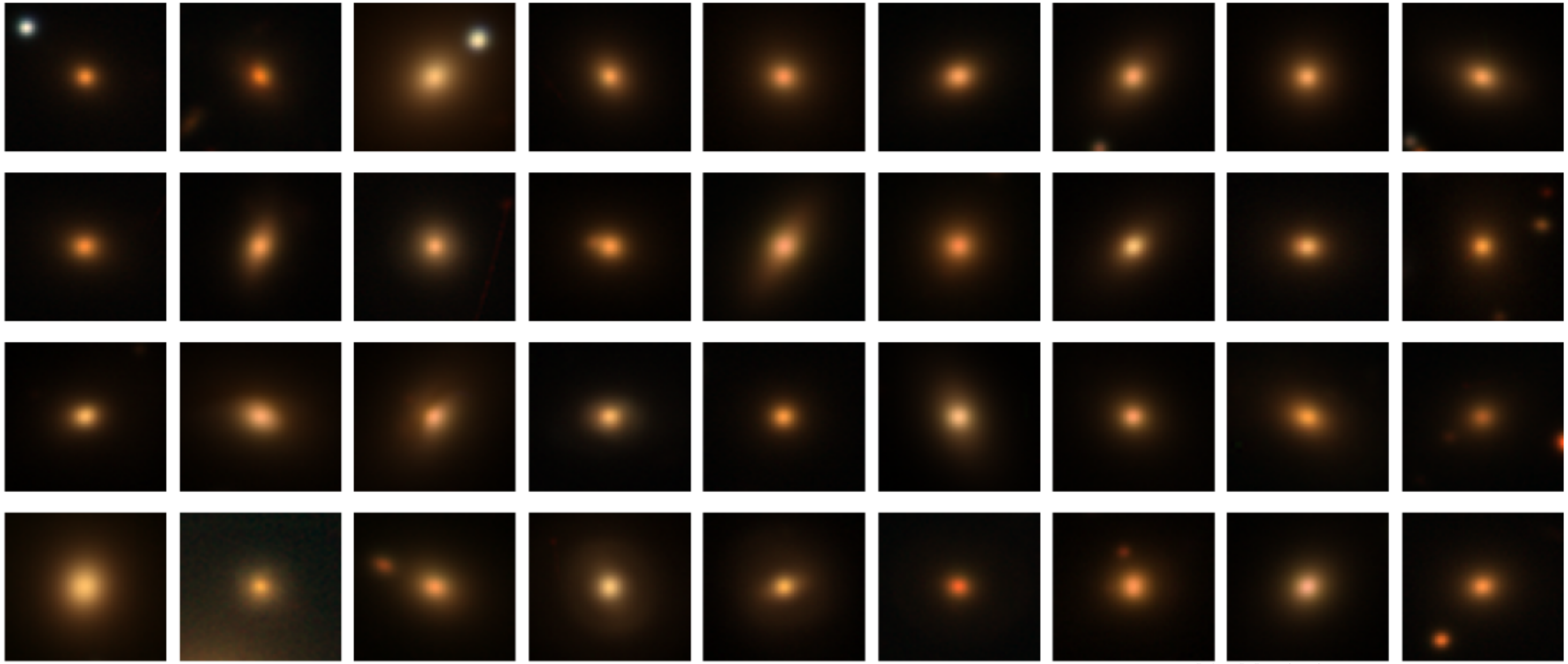
Data credit: IllustrisTNG, Bottrell et al. 2024

Barco et al., ApJ 2025, 2407.17667

Addressing Distributional Shifts: Prior Misspecification

Missa Barco

Applications to galaxies: can we learn that spiral galaxies exist starting from a prior with just (real, noisy) ellipticals?

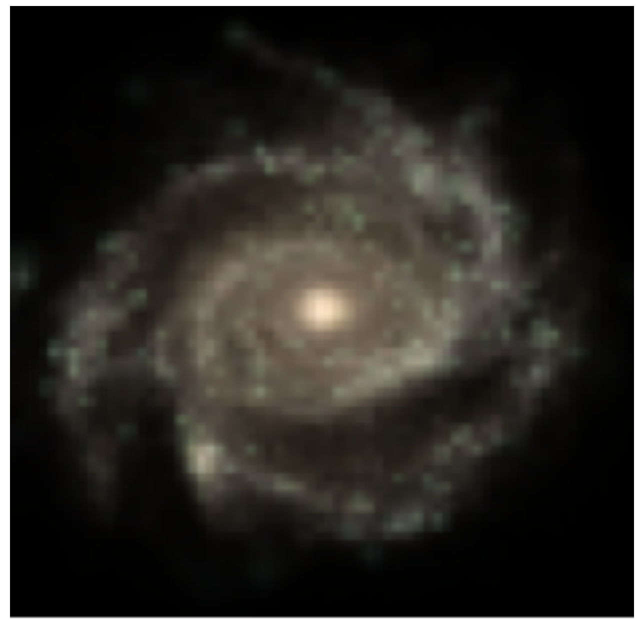


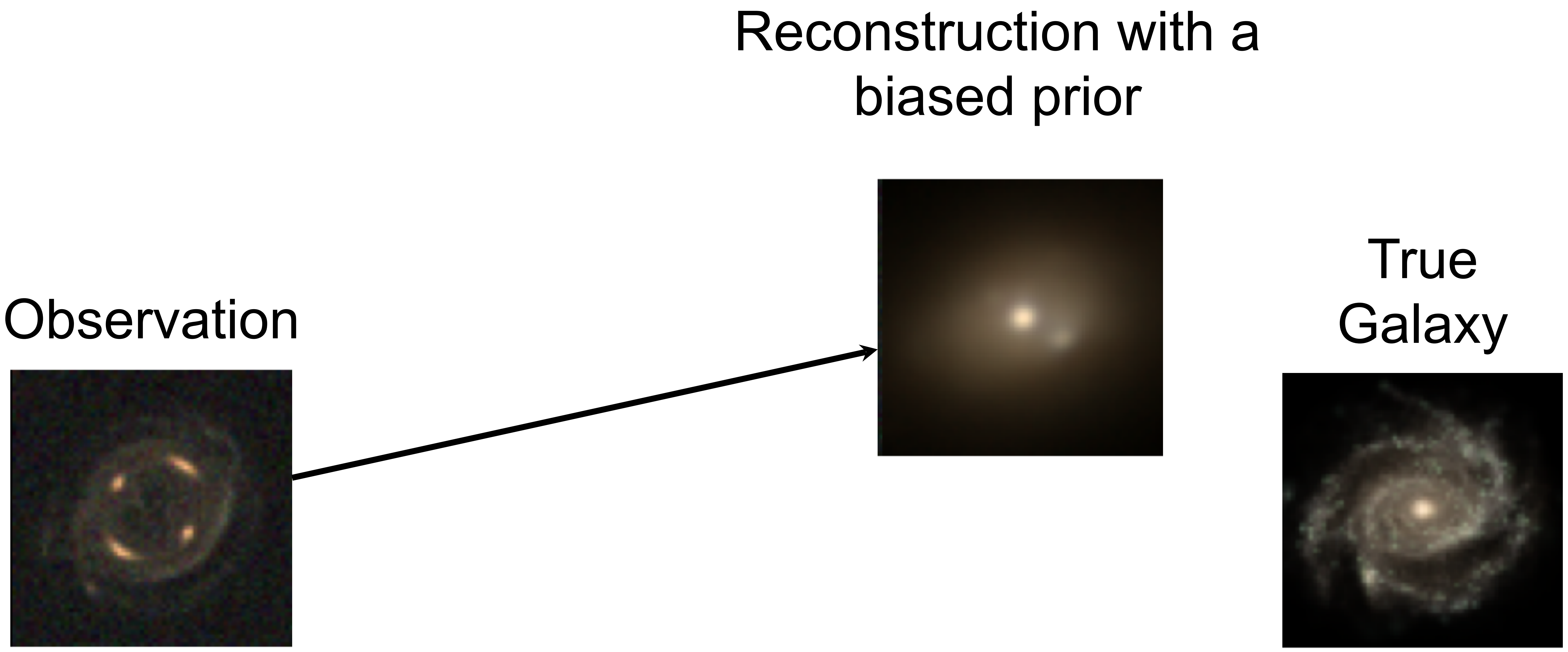
Barco et al., ApJ 2025, 2407.17667

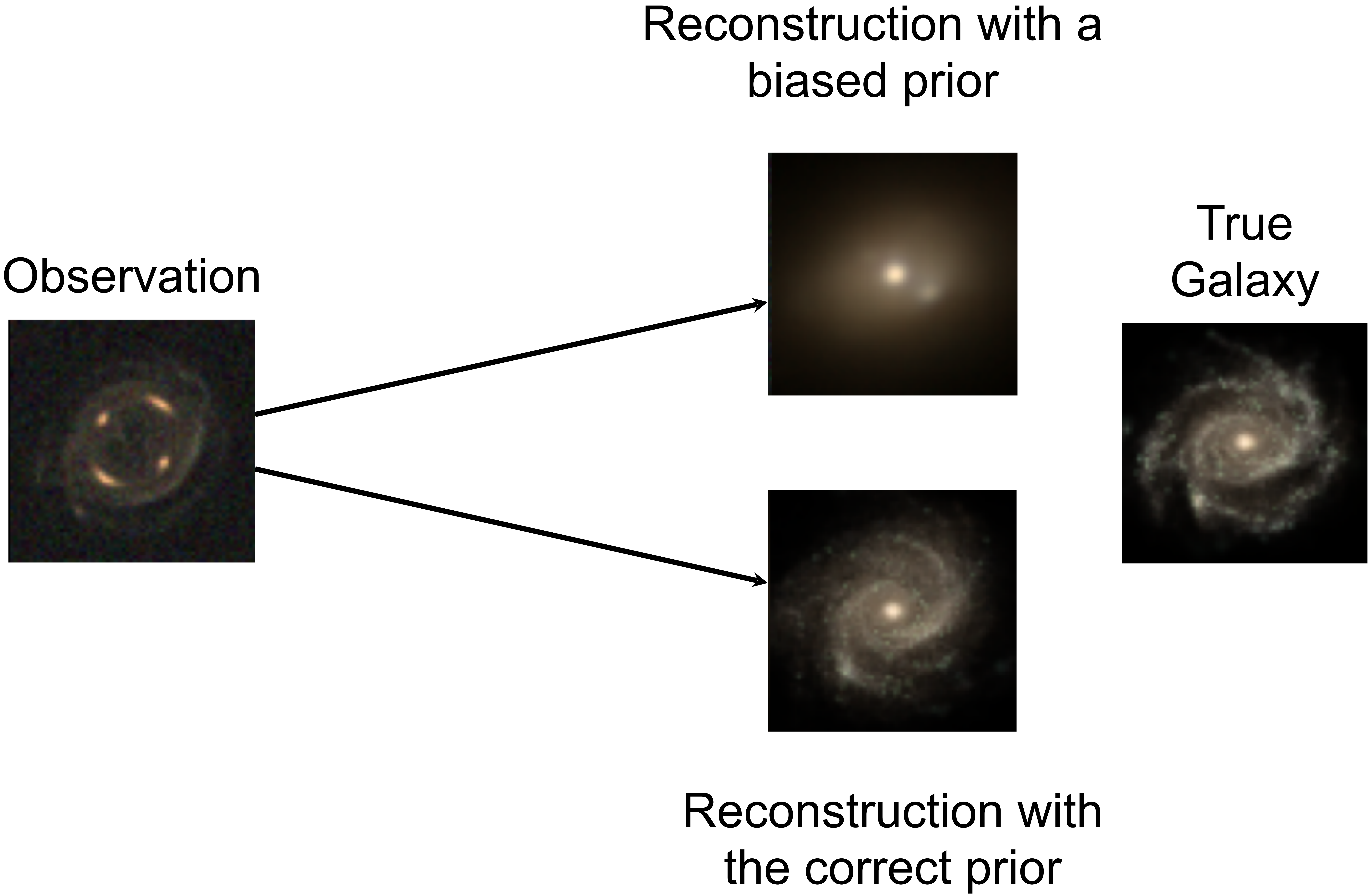
Observation

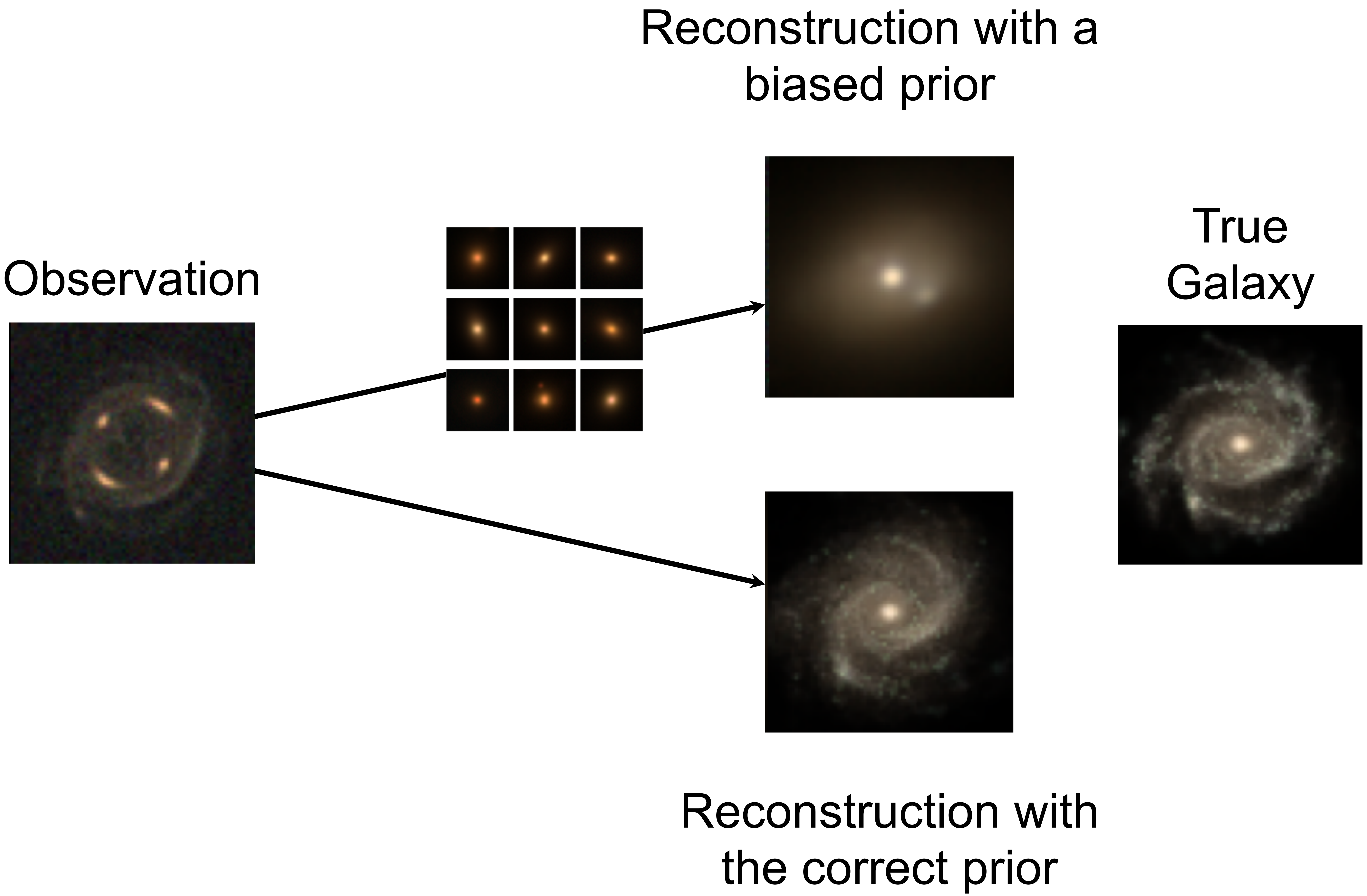
Observation

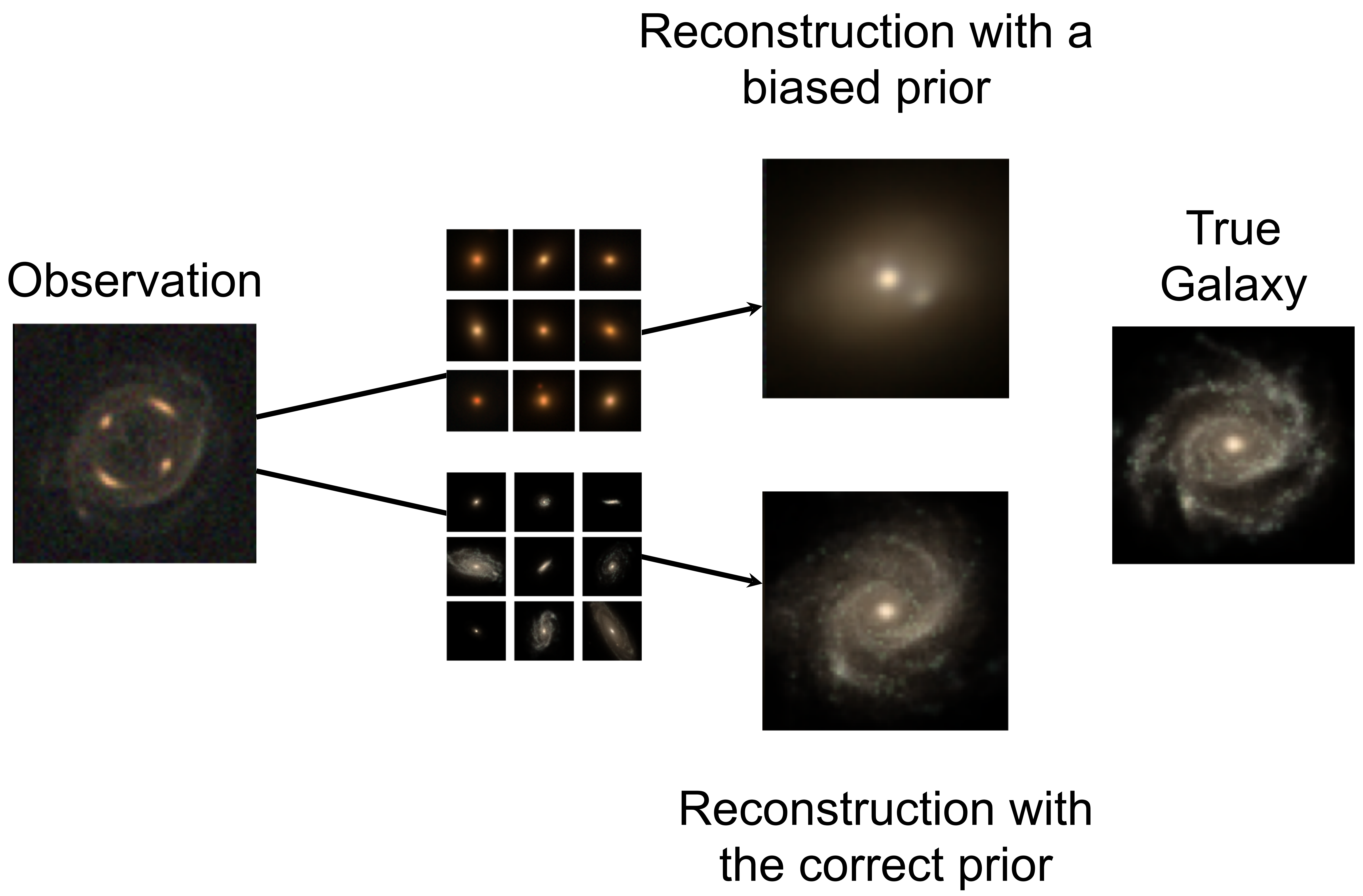
True
Galaxy

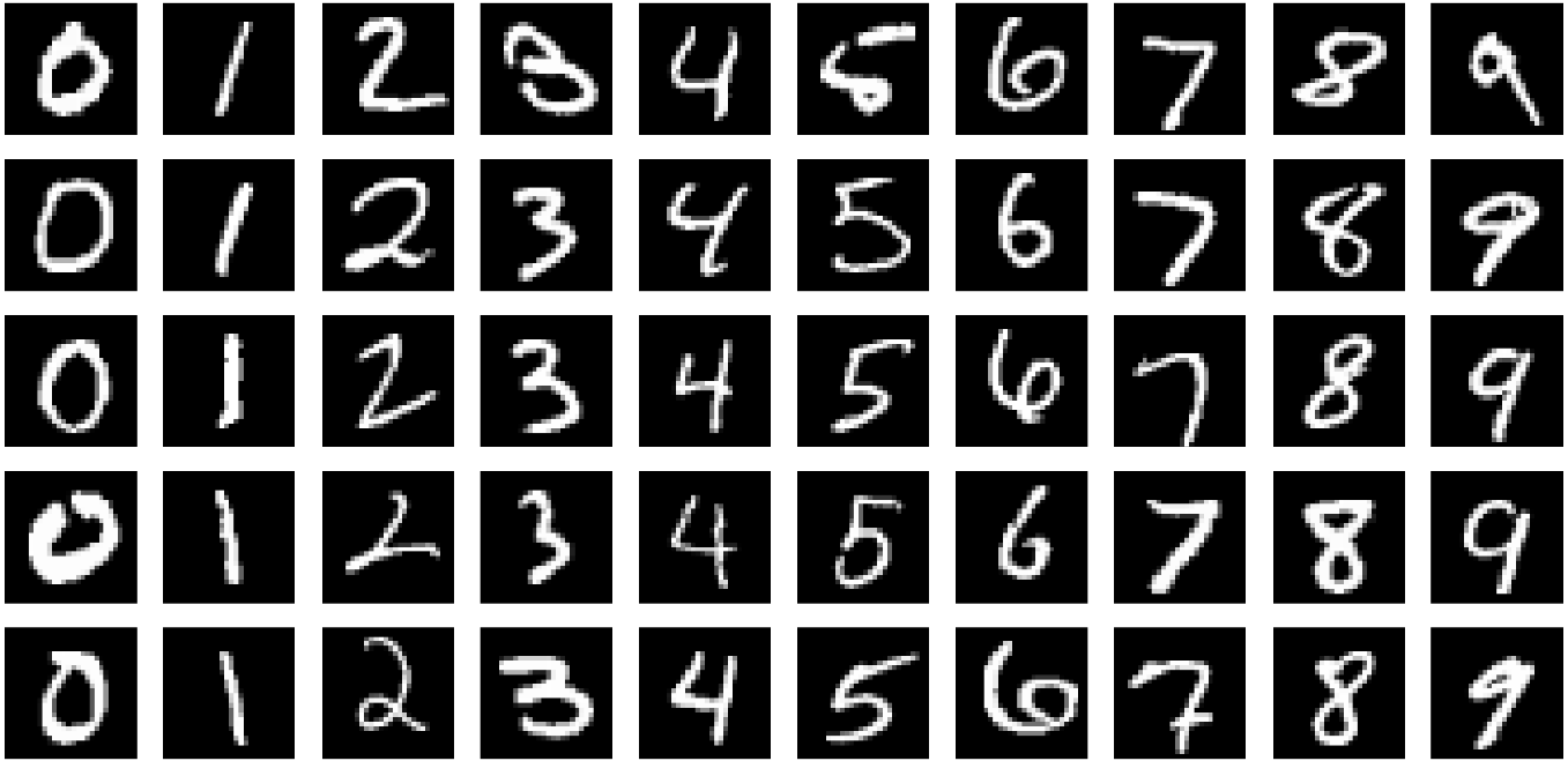


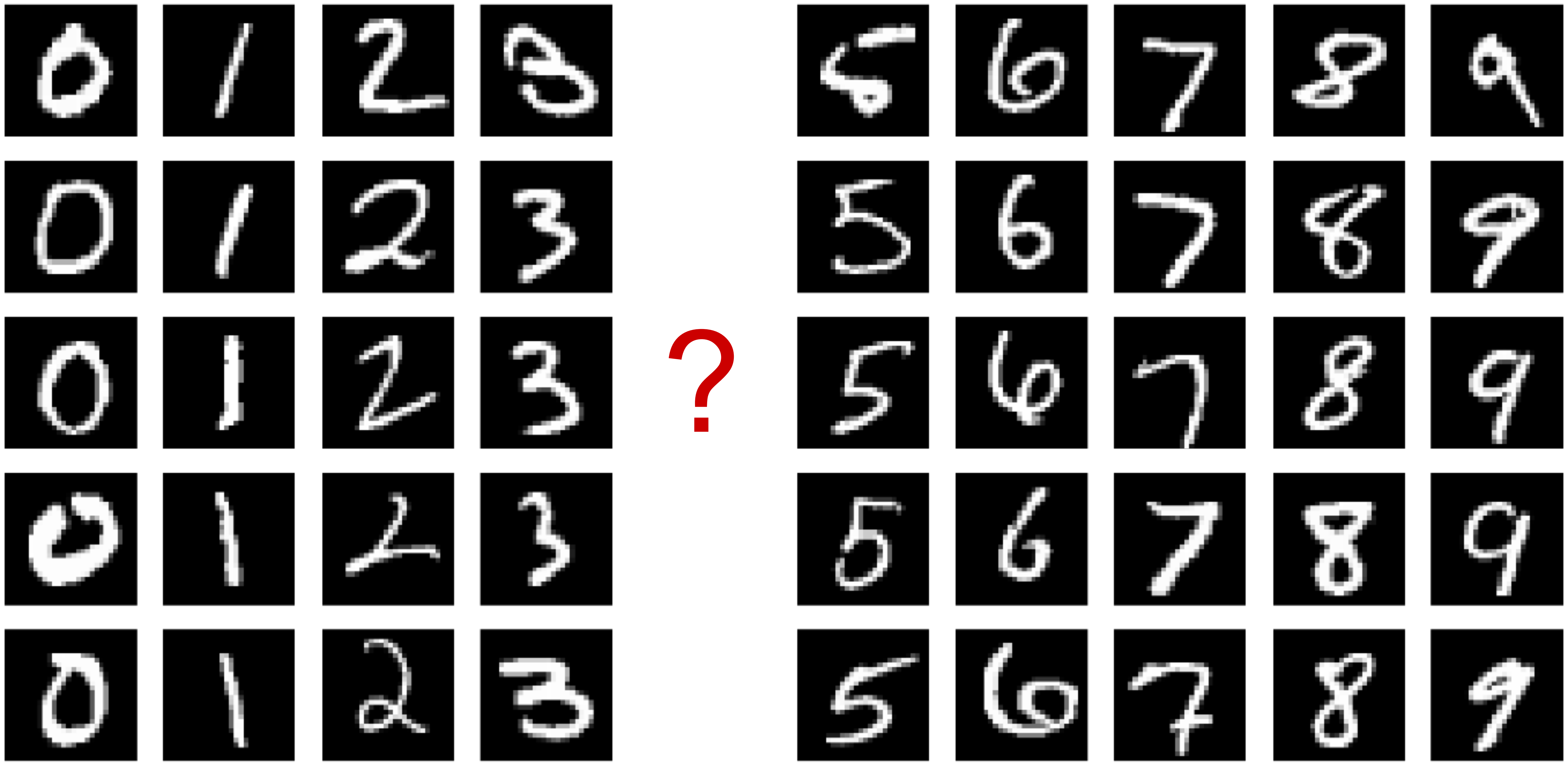

Observation
Reconstruction with a biased prior
True Galaxy


Reconstruction with a biased prior
Observation
True Galaxy
Reconstruction with the correct prior

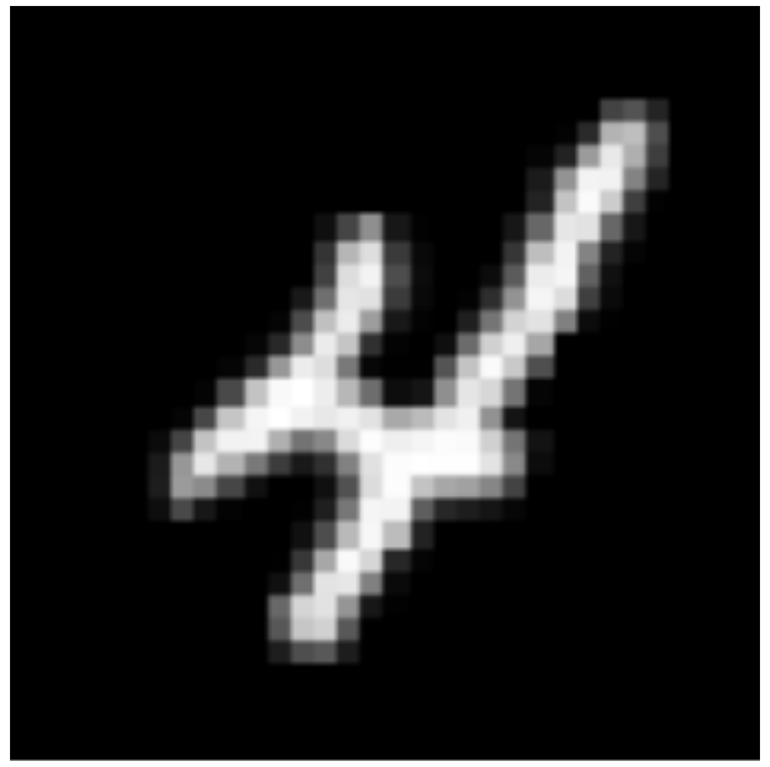

Reconstruction with a
biased prior
Observation
True
Galaxy
Reconstruction with
the correct prior

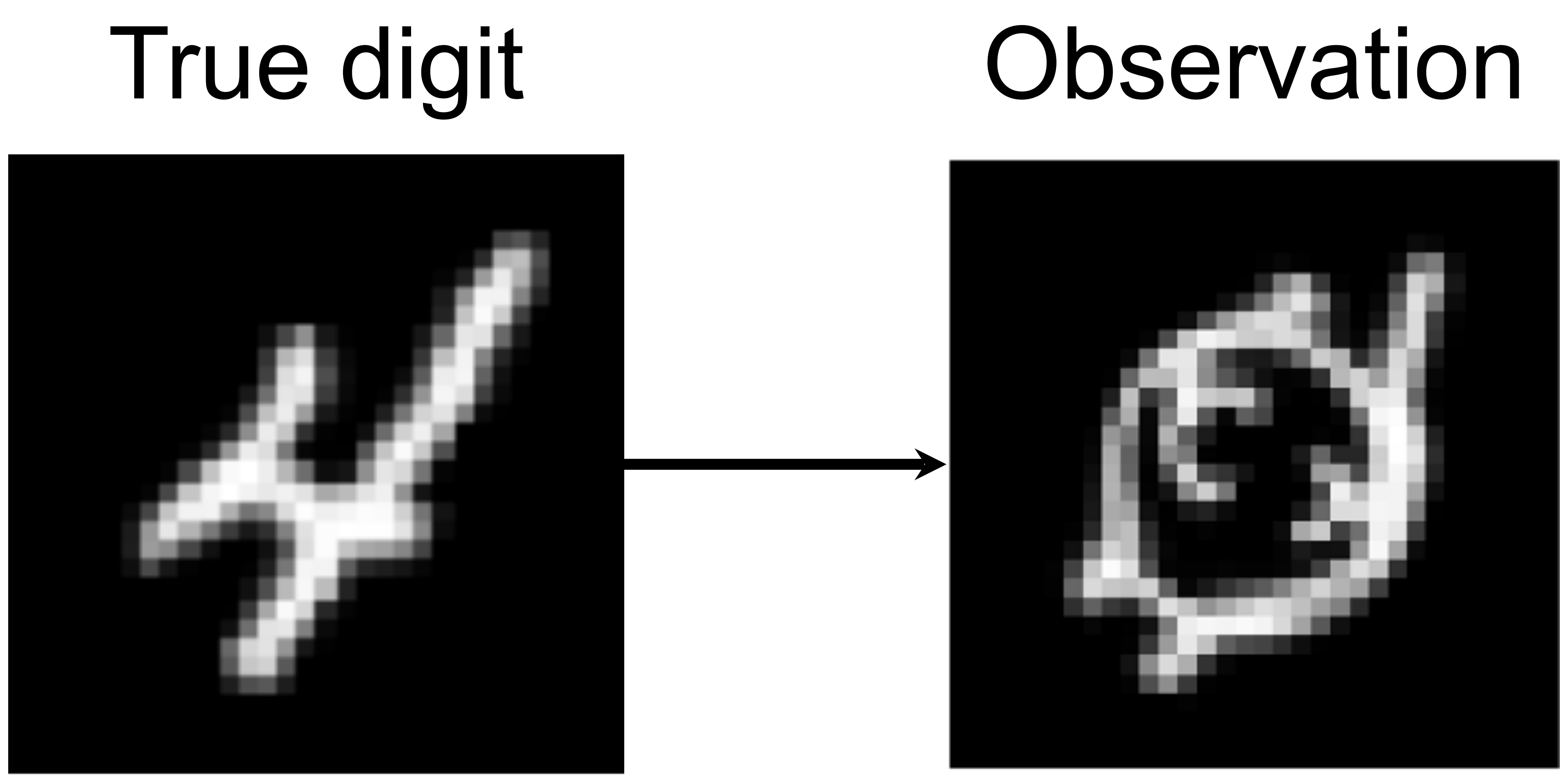

Reconstruction with a
biased prior
True
Galaxy
Observation
Reconstruction with
the correct prior

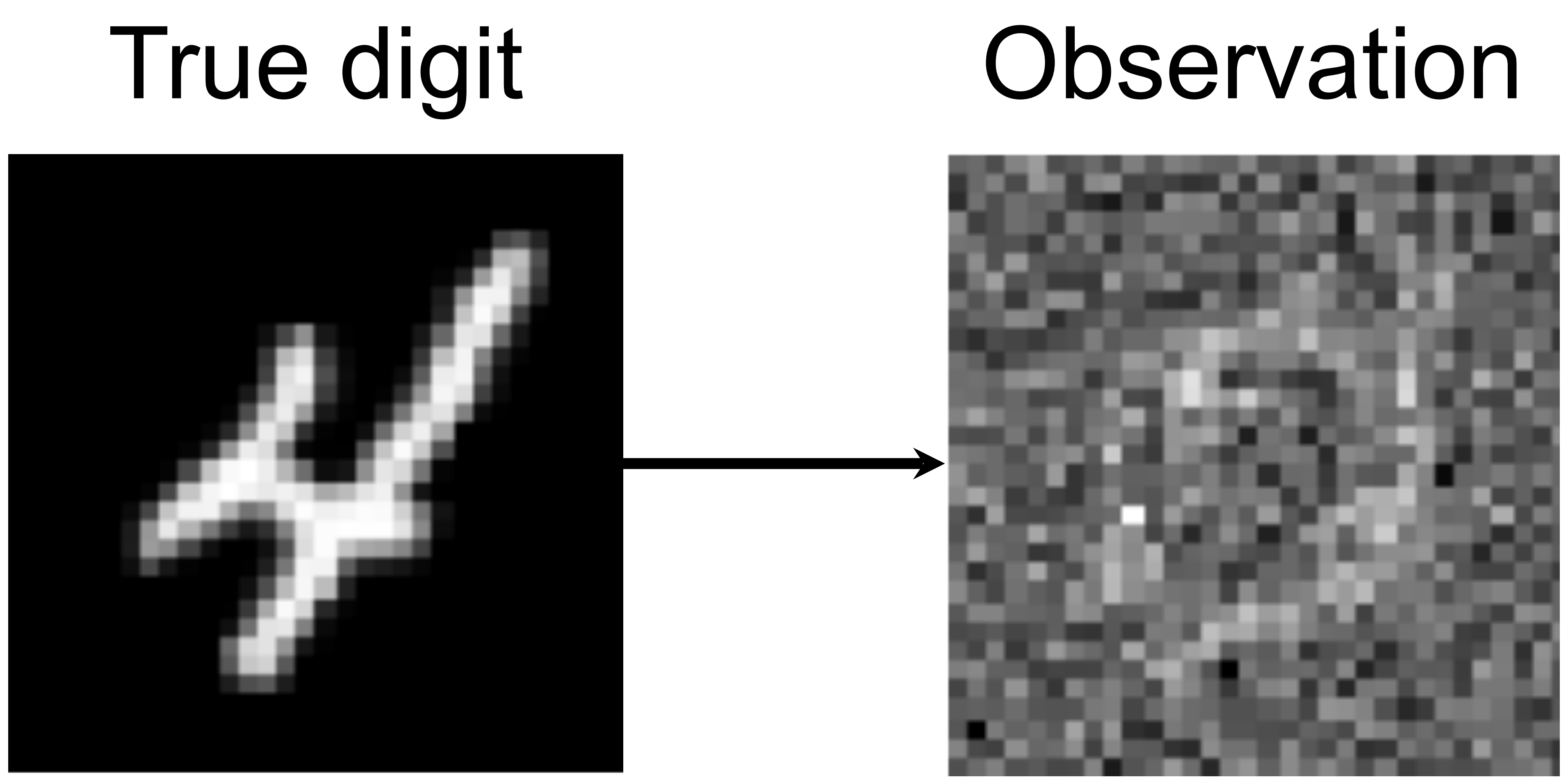


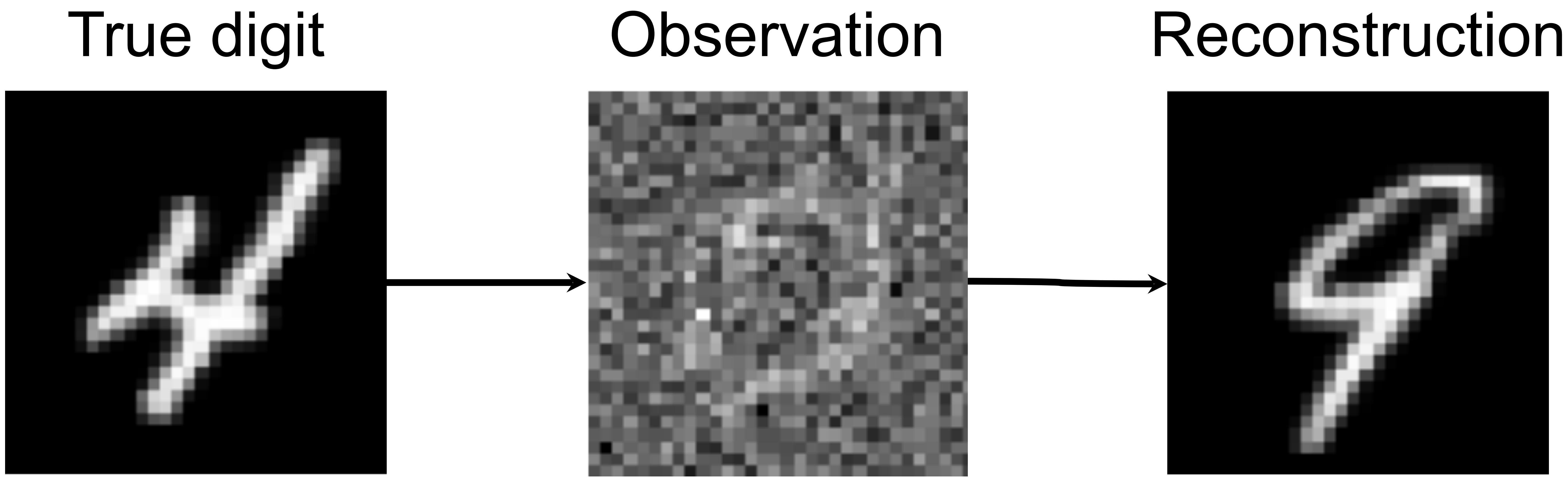


True digit

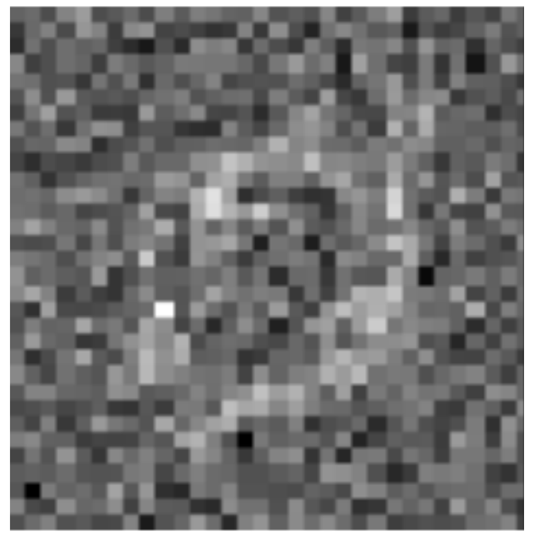


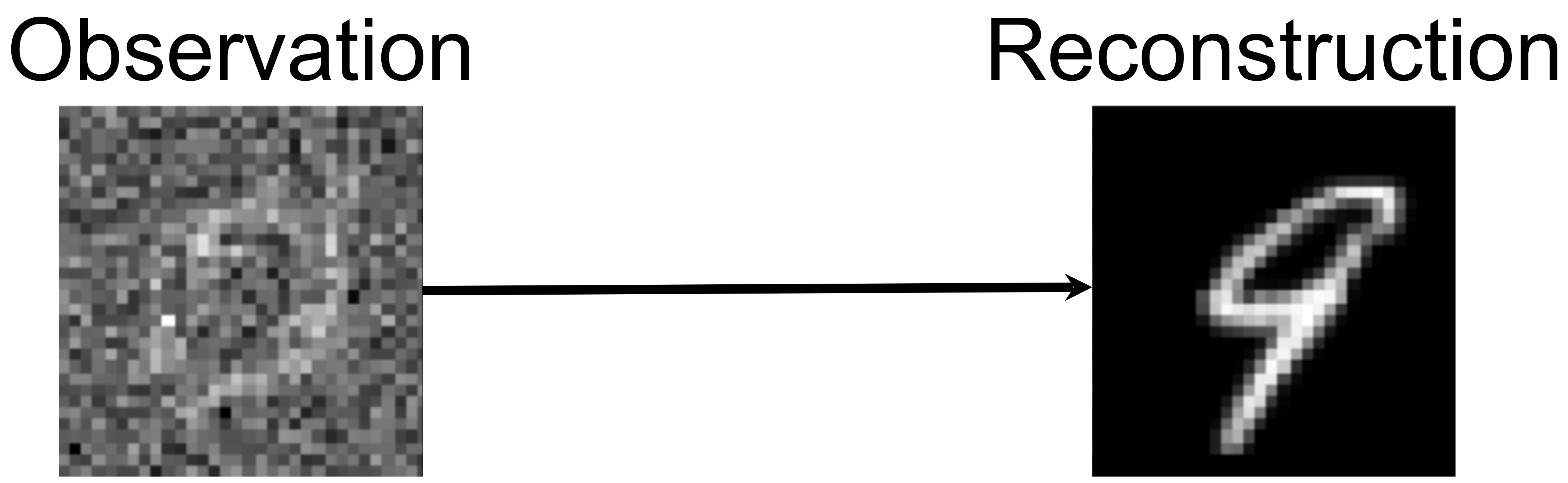

True digit
Observation

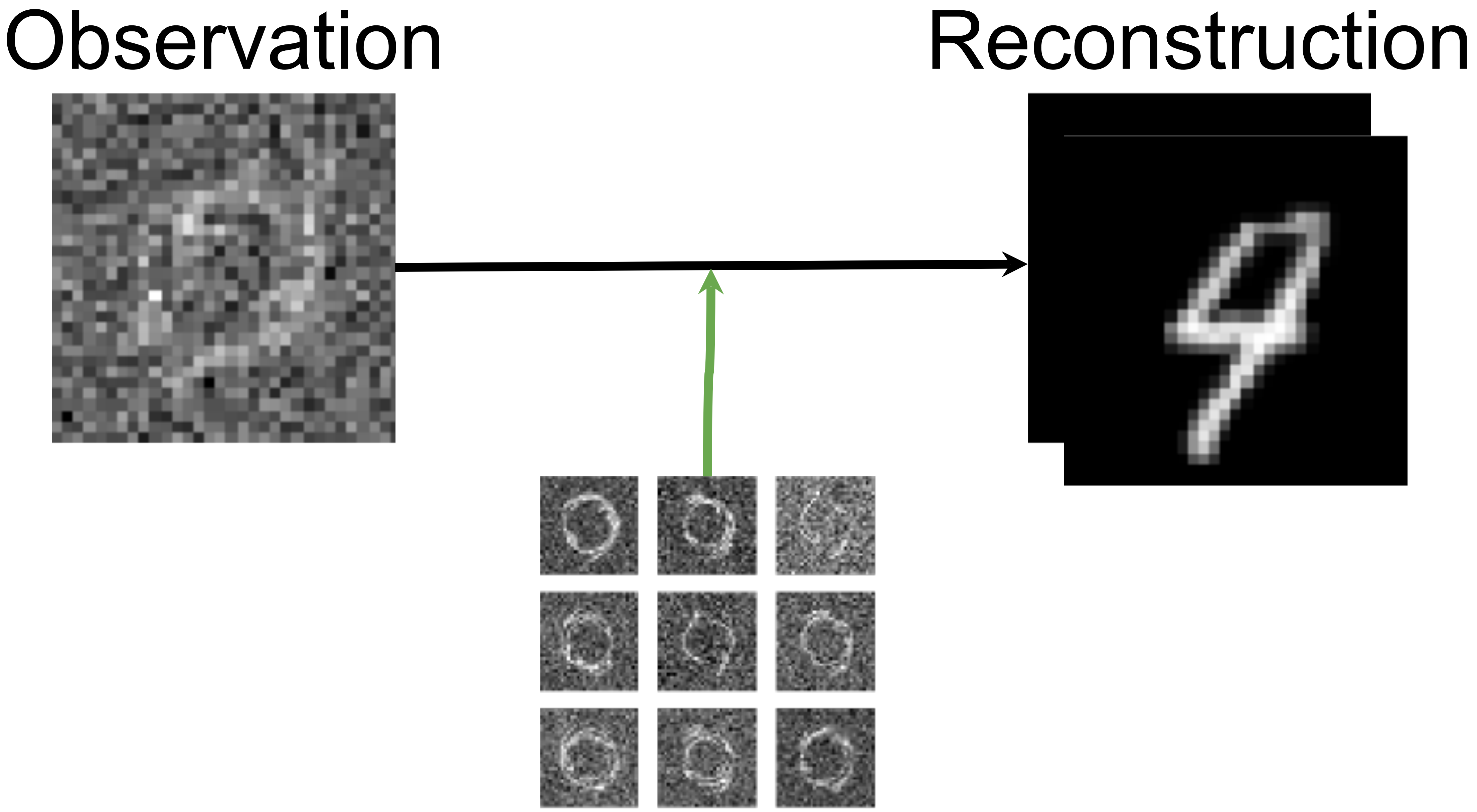

True digit
Observation

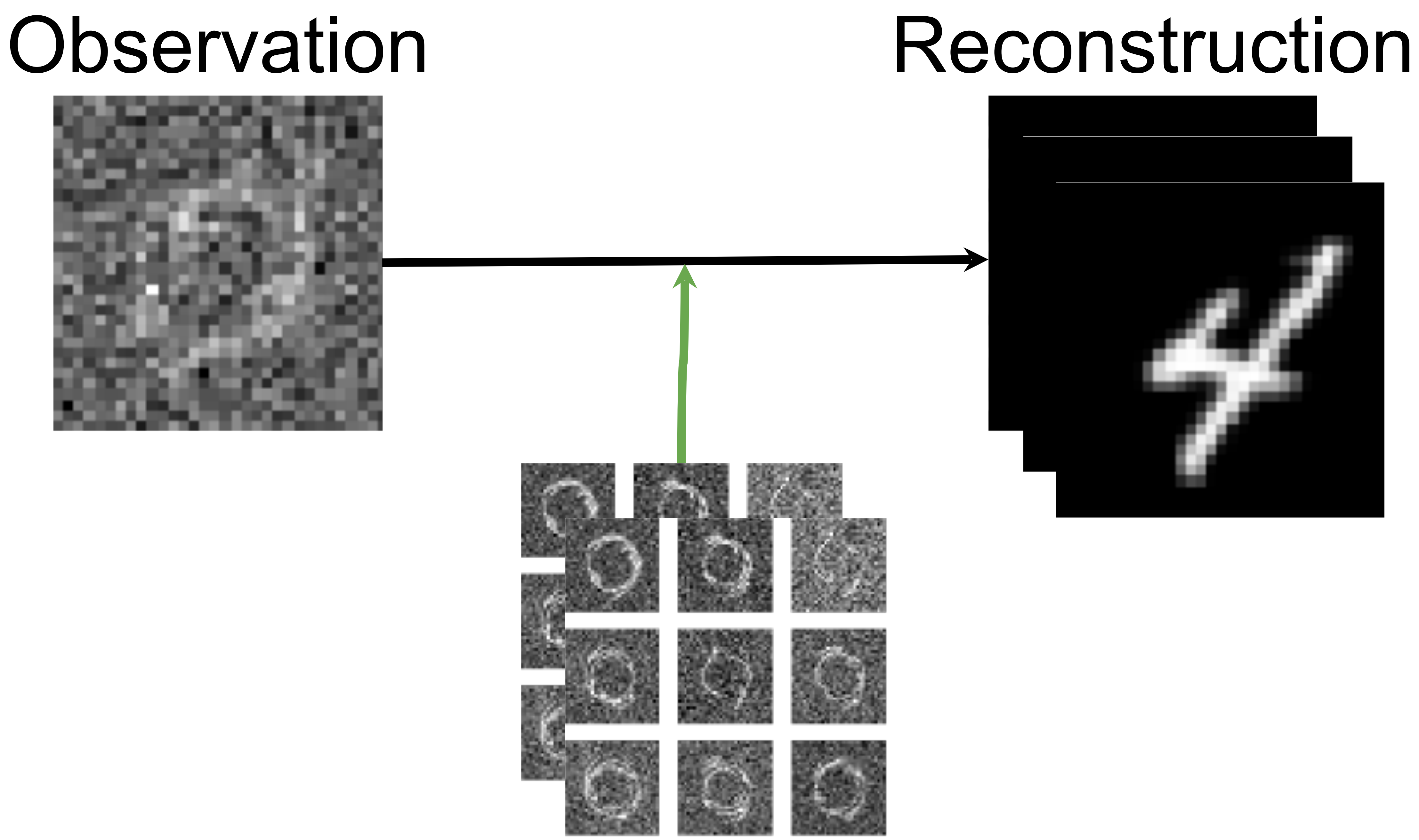

True digit
Observation
Reconstruction

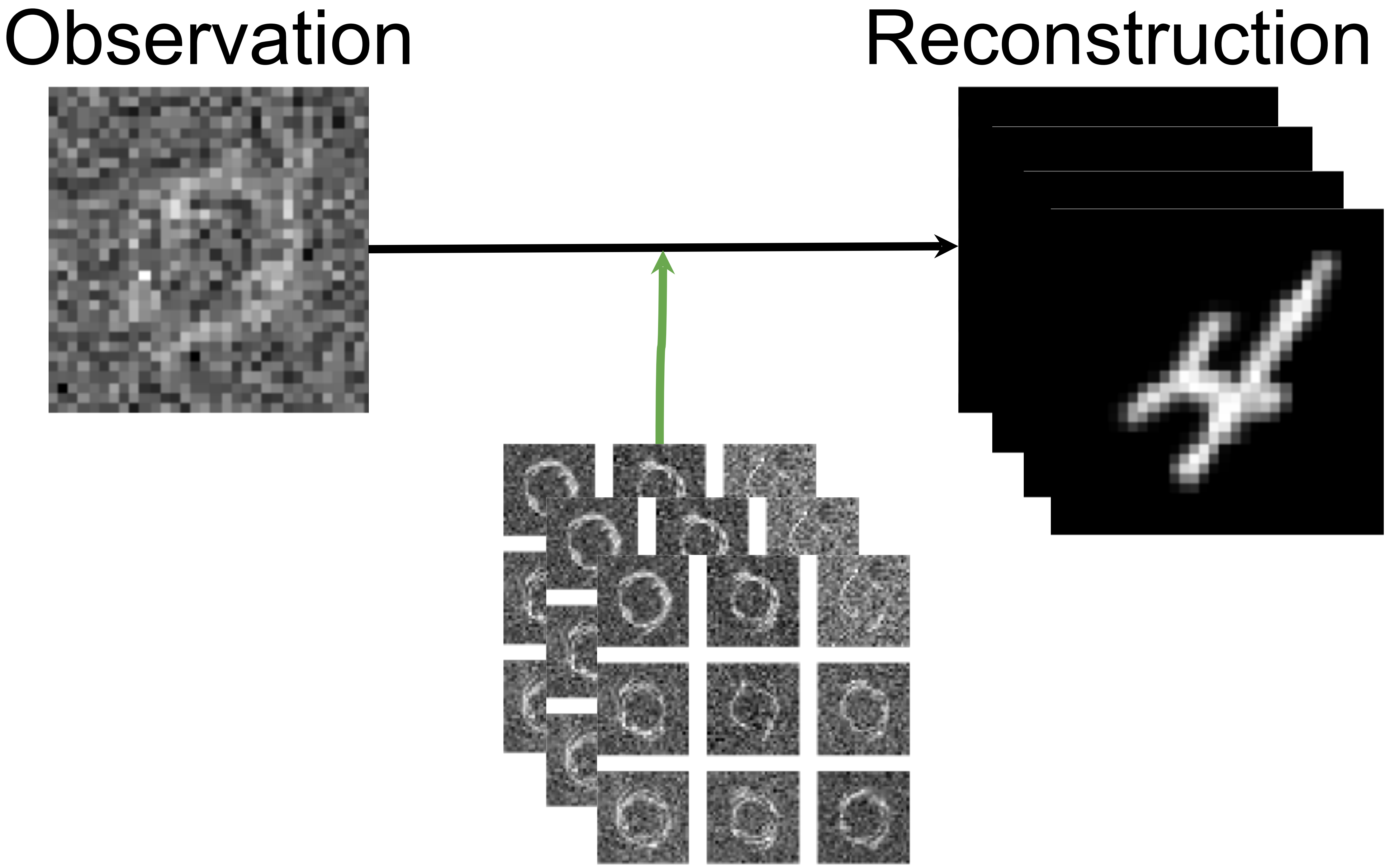
Observation

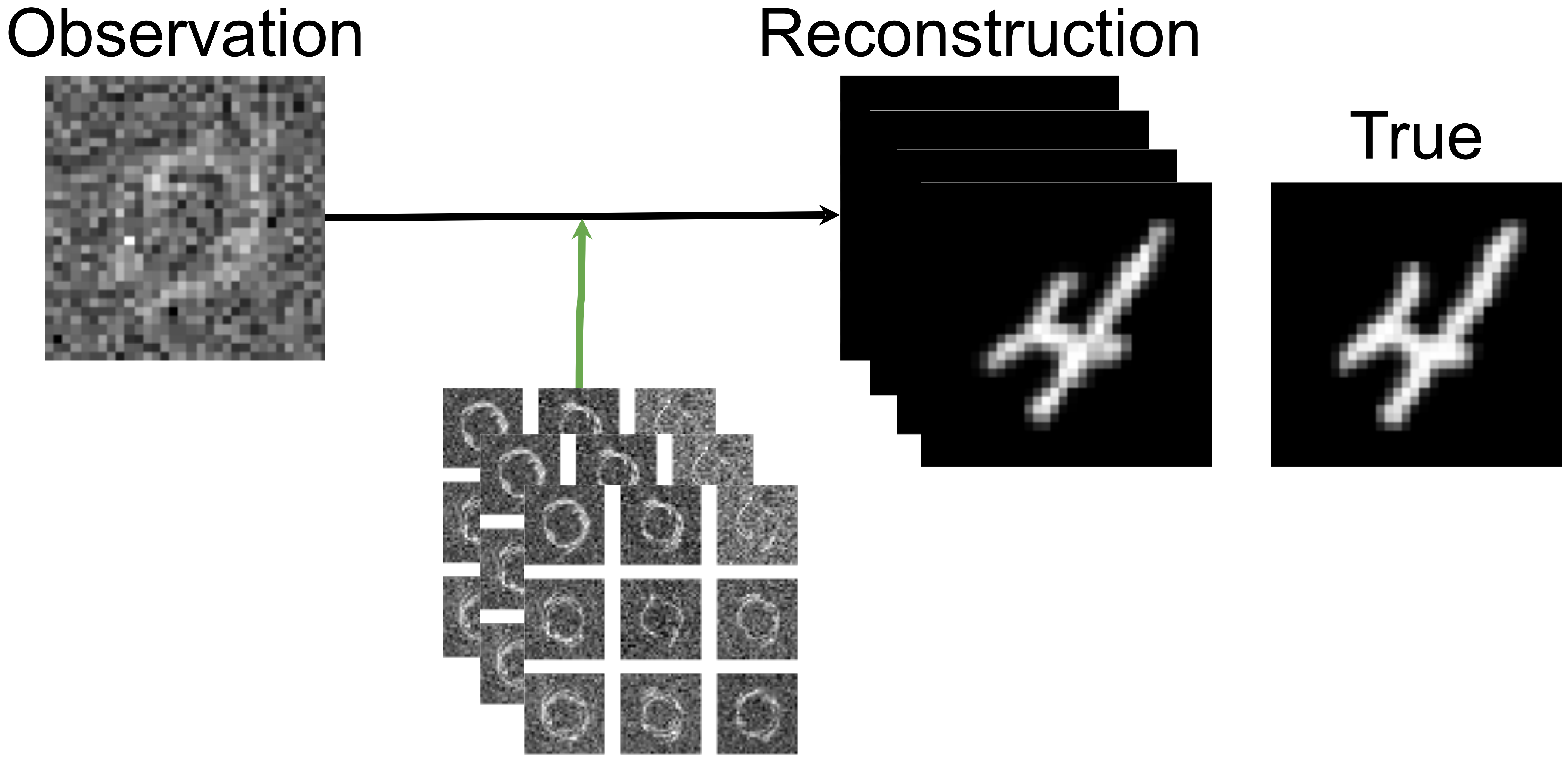



Observation
Reconstruction


Observation
Reconstruction


Observation
Reconstruction

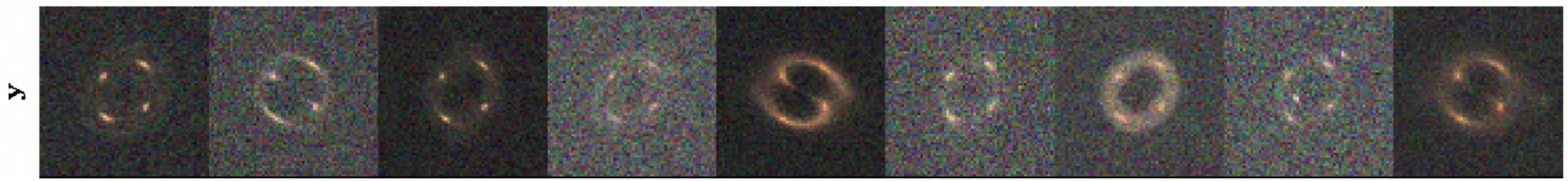

Observation
Reconstruction

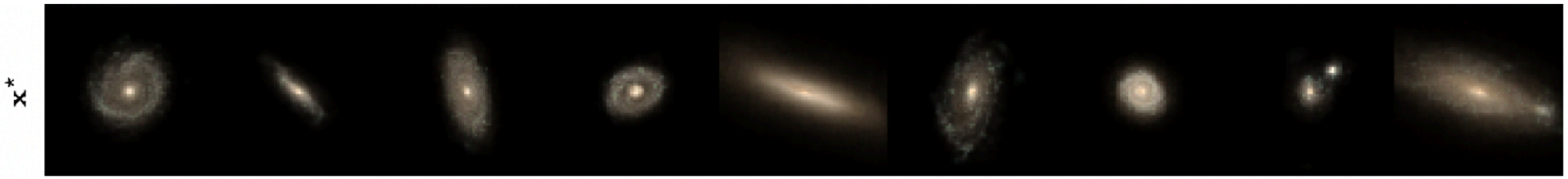

Observation
Reconstruction
True

Addressing Distributional Shifts: Prior Misspecification

Missa Barco

Applications to galaxies: can we learn that spiral galaxies exist starting from a prior with just (real, noisy) ellipticals?



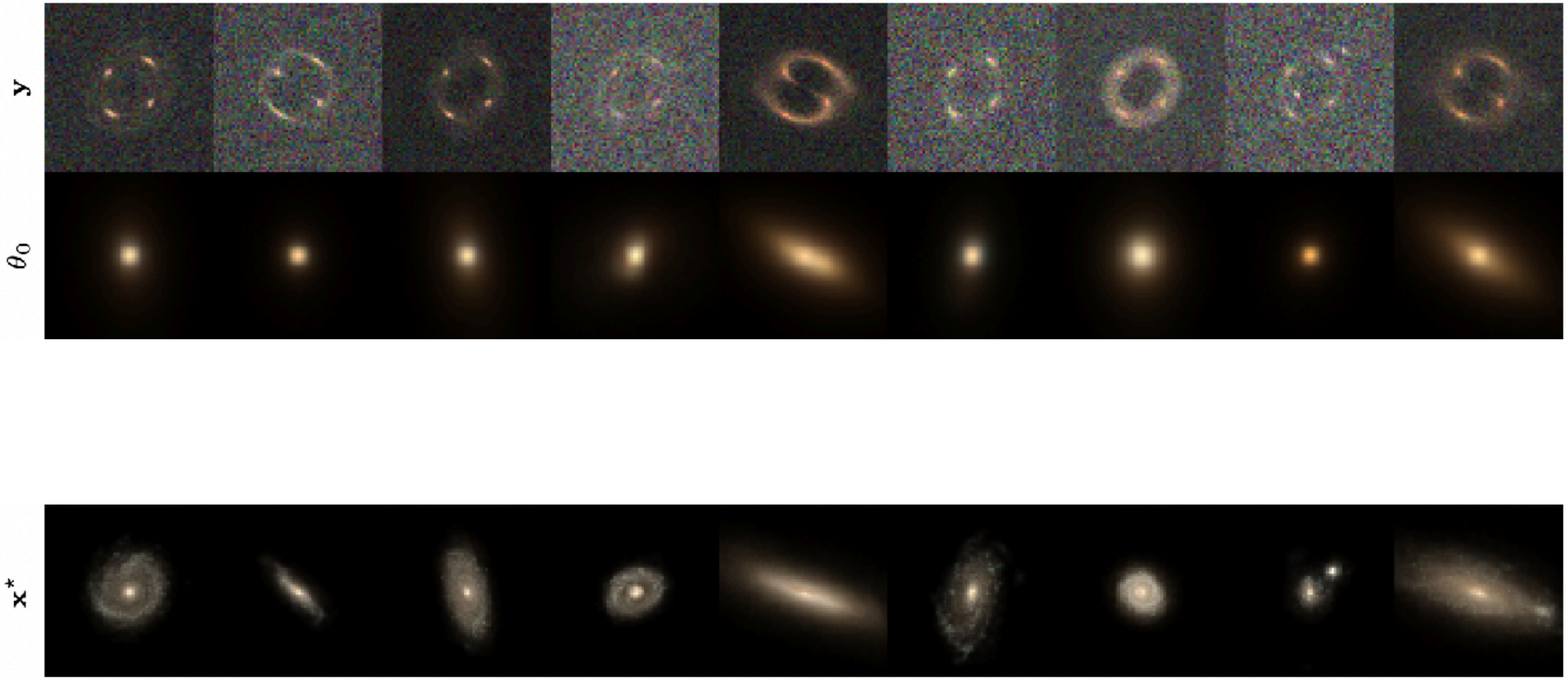


Barco et al., ApJ 2025, 2407.17667

Addressing Distributional Shifts: Prior Misspecification

Missa Barco

Applications to galaxies: can we learn that spiral galaxies exist starting from a prior with just (real, noisy) ellipticals?

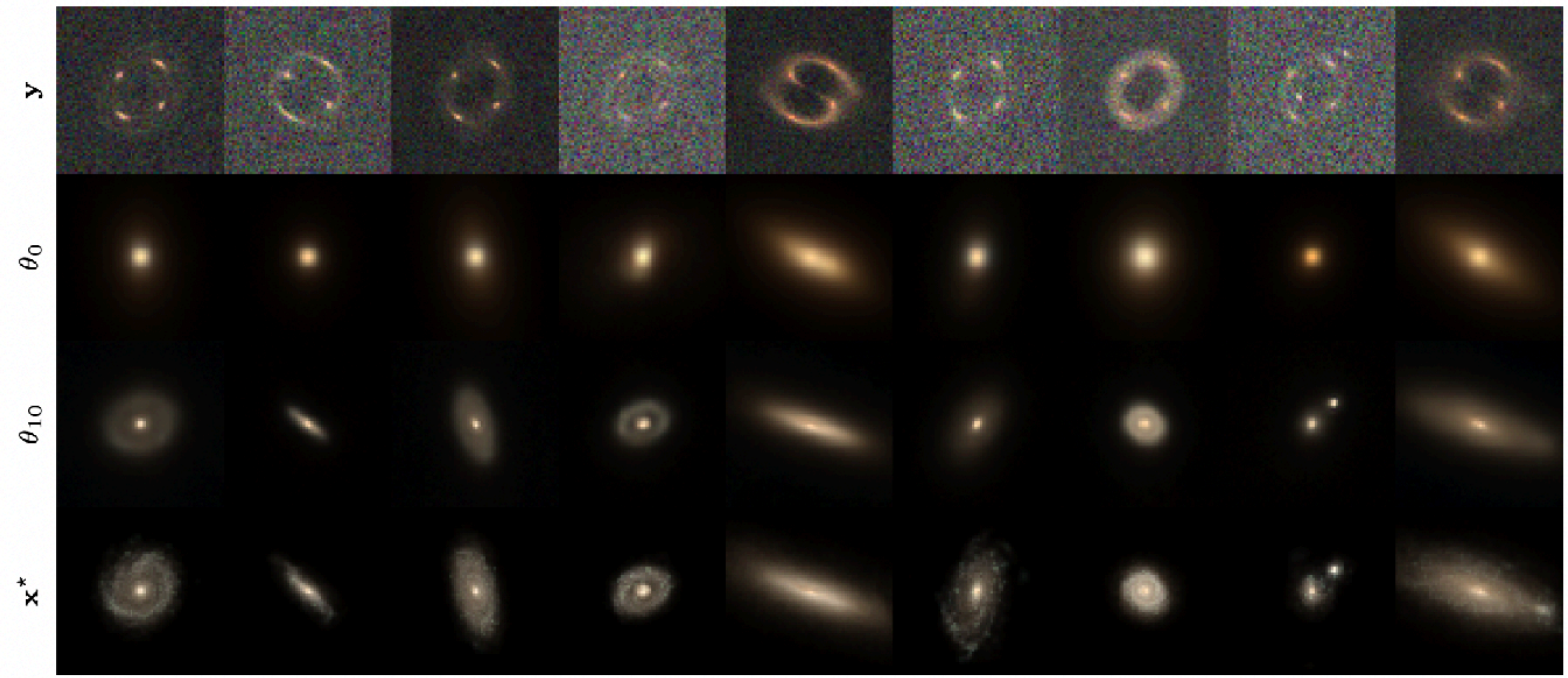


Barco et al., ApJ 2025, 2407.17667

ADDRESSING DISTRIBUTIONAL SHIFTS: PRIOR MISSPECIFICATION

Missa
Barco

Applications to galaxies: can we learn that spiral galaxies exist starting from a prior with just (real, noisy) ellipticals?



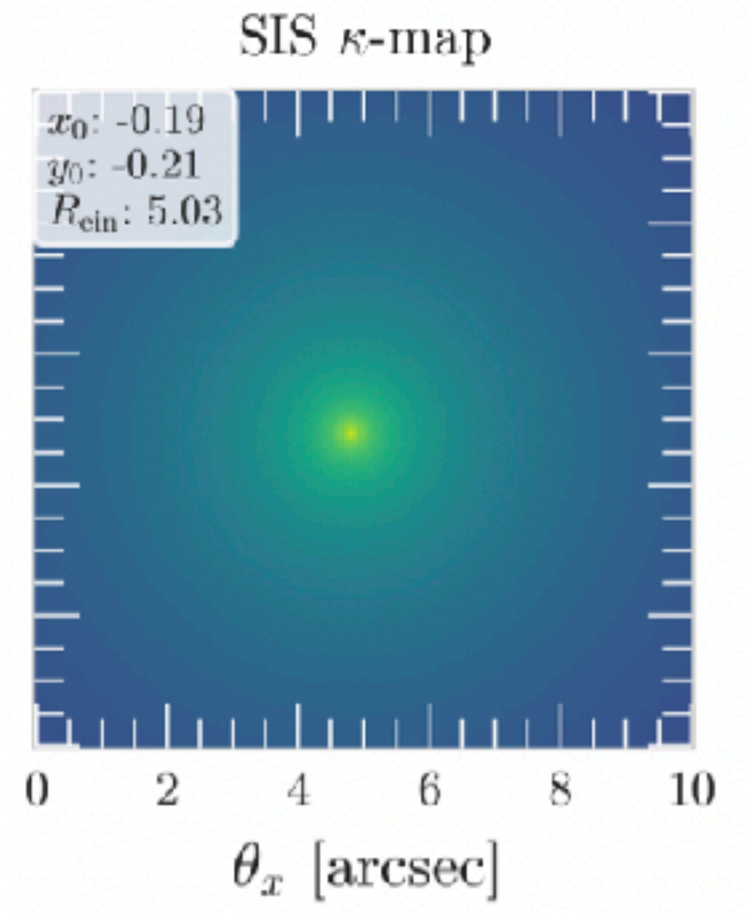
Barco et al., ApJ 2025, 2407.17667

Addressing Distributional Shifts: Model Misspecification?

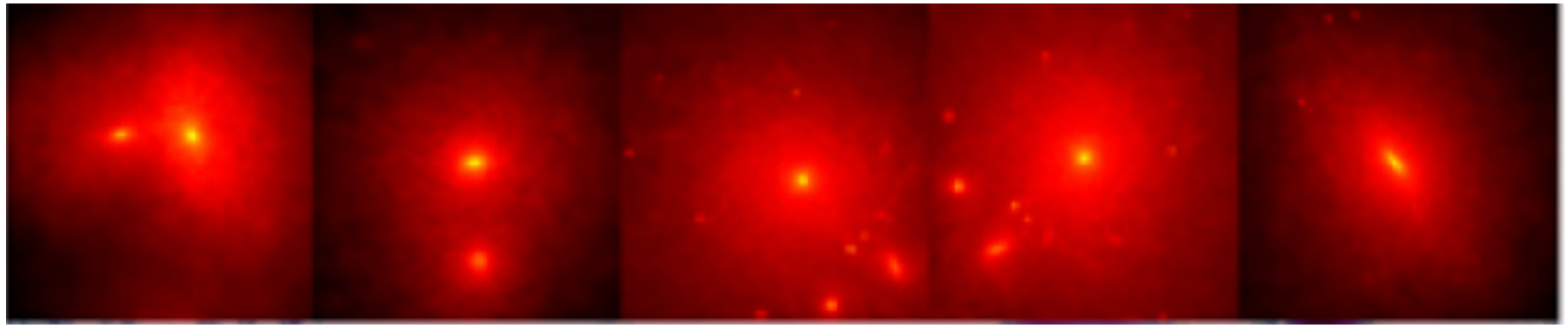
Nicolas Payot

Could we use a similar framework to learn misspecified ***physical models***?

Parametric model for the lens:



Realistic mass map from simulations:



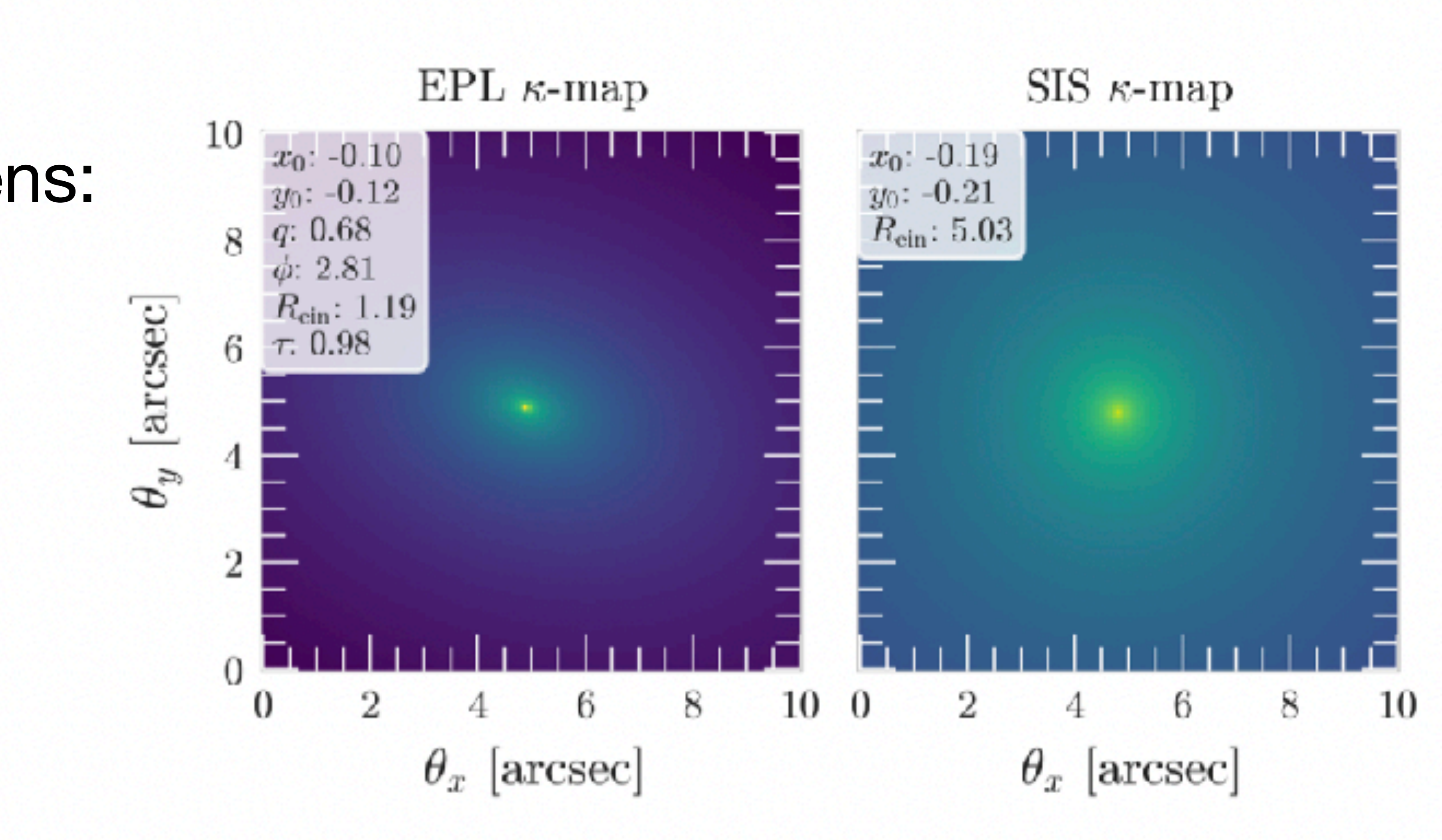
Payot et al., ML4Astro Workshop at ICML 2025,

ADDRESSING DISTRIBUTIONAL SHIFTS: MODEL MISSPECIFICATION?

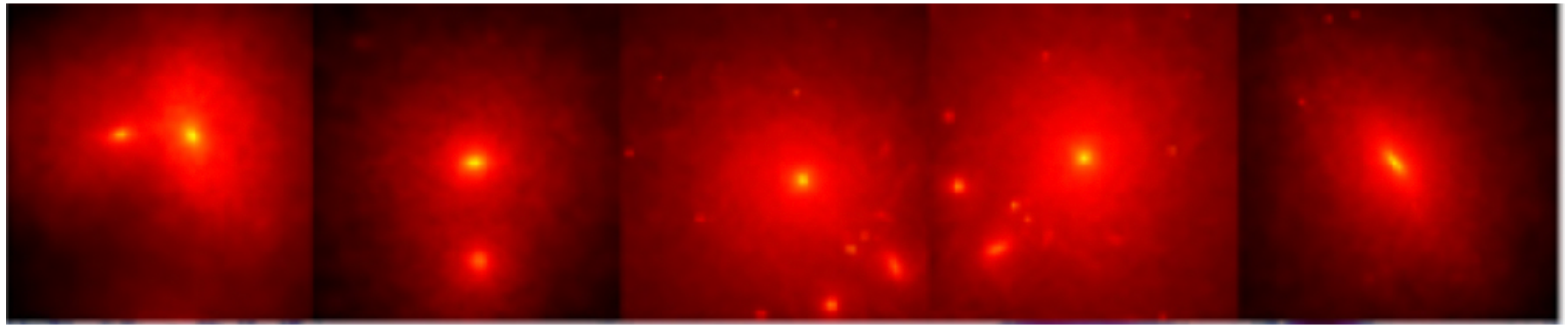
Nicolas Payot

Could we use a similar framework to learn misspecified ***physical models***?

Parametric model for the lens:



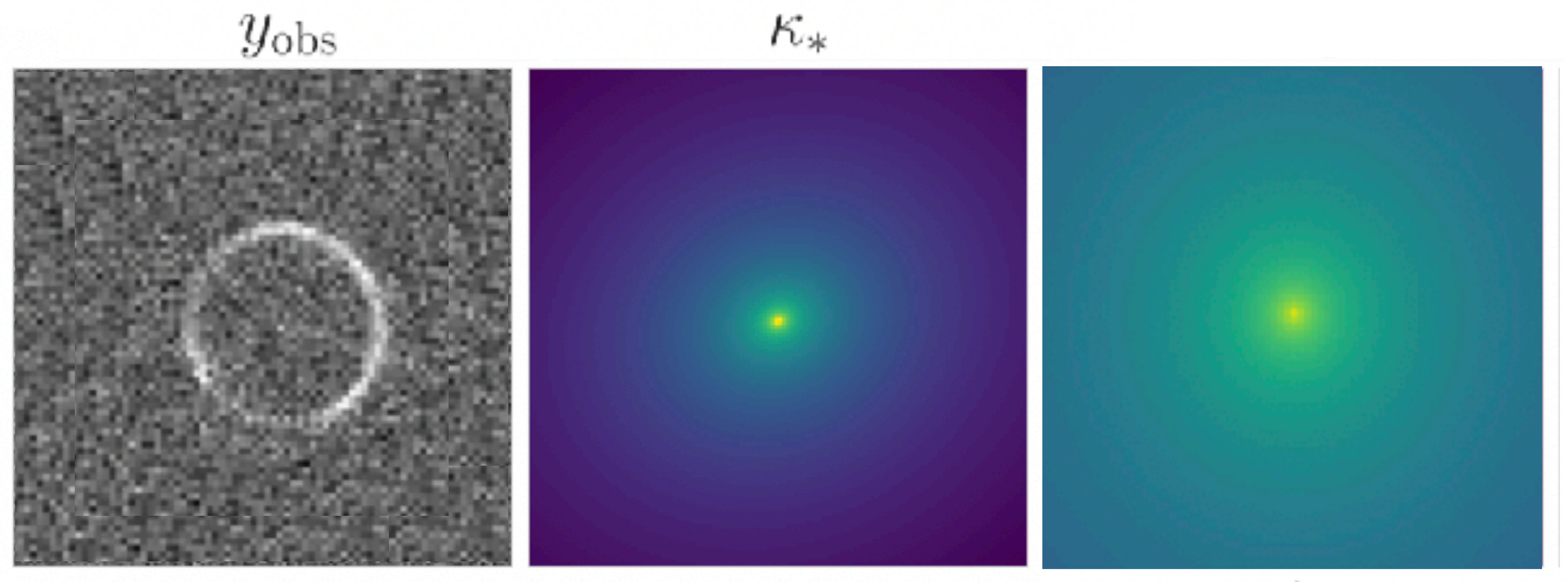
Realistic mass map from simulations:



Payot et al., ML4Astro Workshop at ICML 2025,

Addressing Distributional Shifts: Model Misspecification?

We can map our over-constrained physical model to an over-parametrized latent space, and apply the same techique!

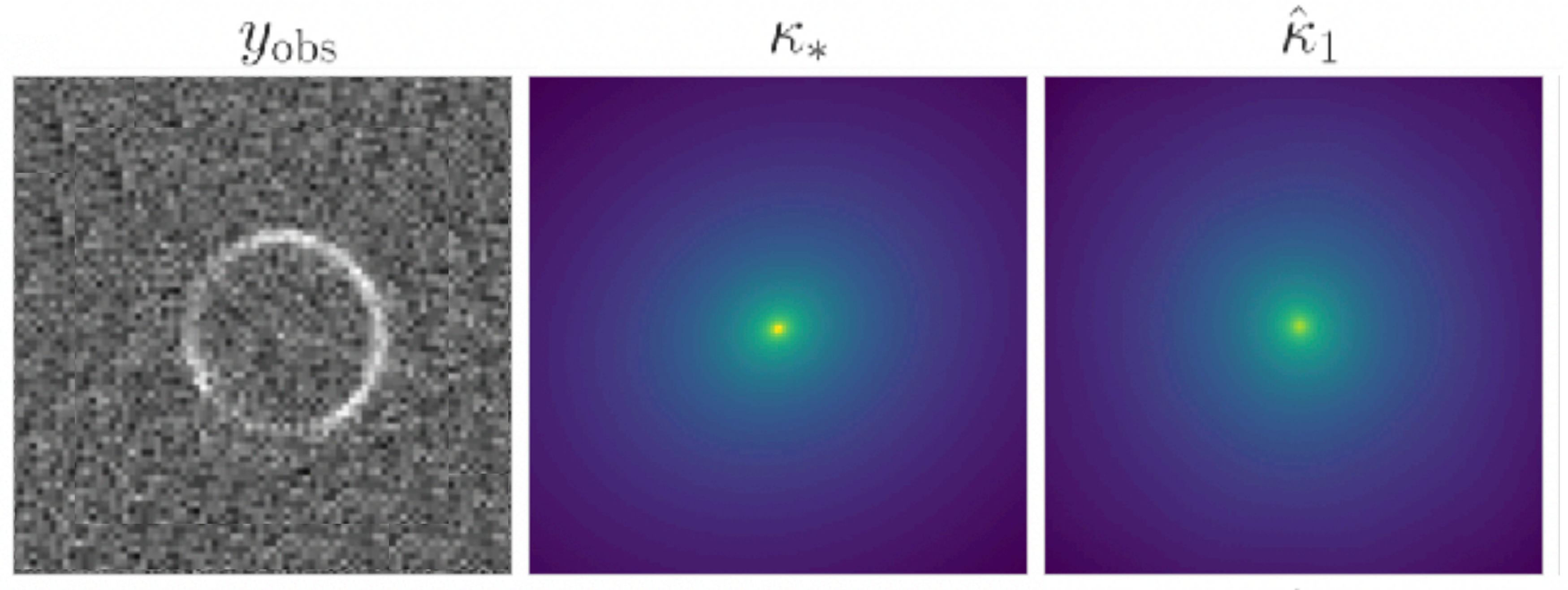


Payot et al., ML4Astro Workshop at ICML 2025,

Addressing Distributional Shifts: Model Misspecification?

Nicolas Payot

We can map our over-constrained physical model to an over-parametrized latent space, and apply the same techique!

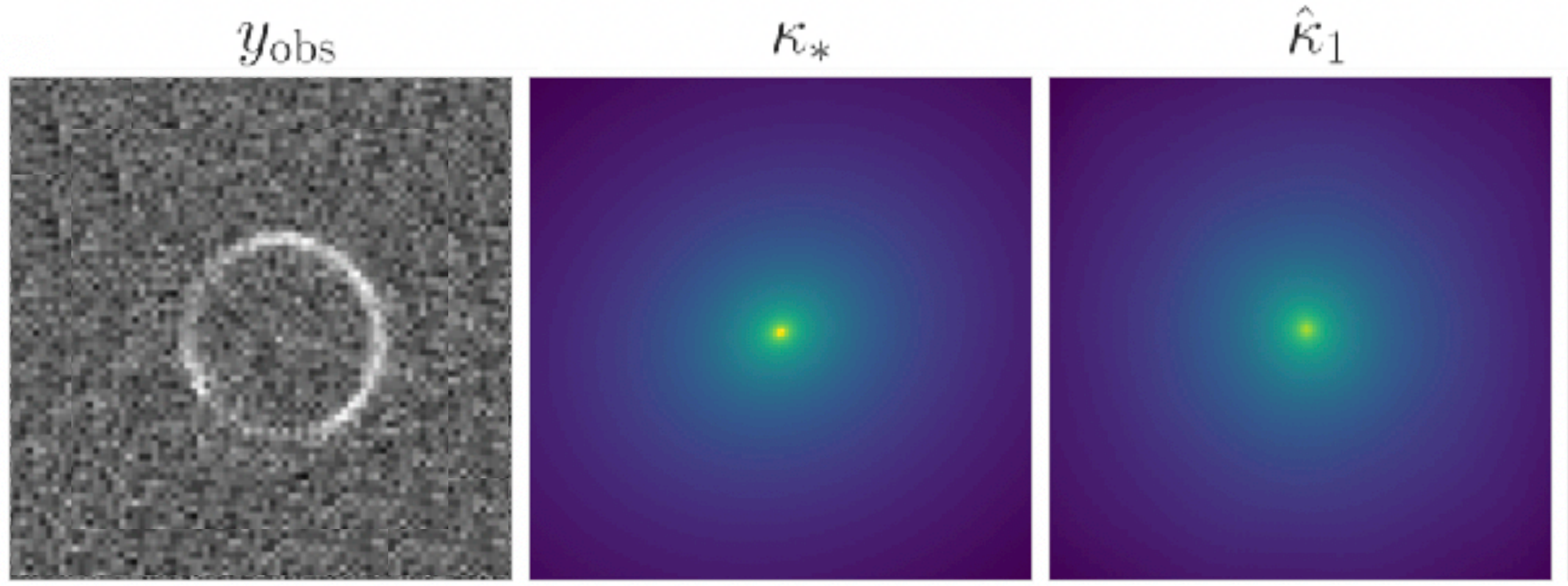


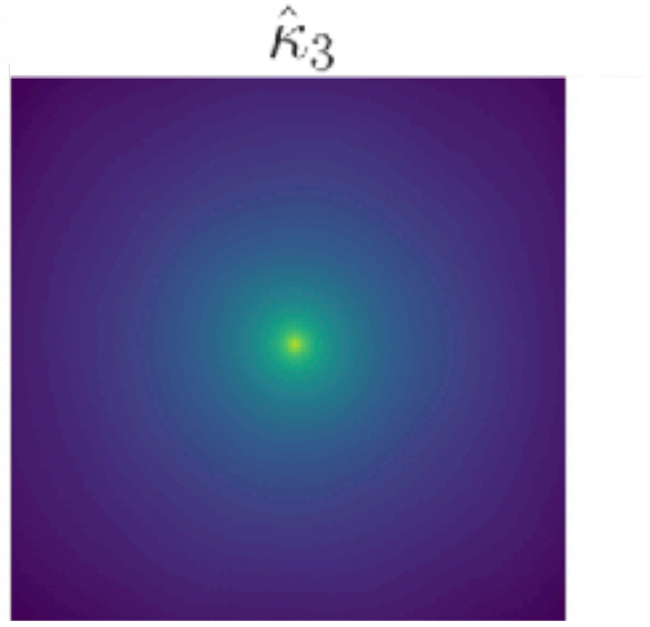
Payot et al., ML4Astro Workshop at ICML 2025,

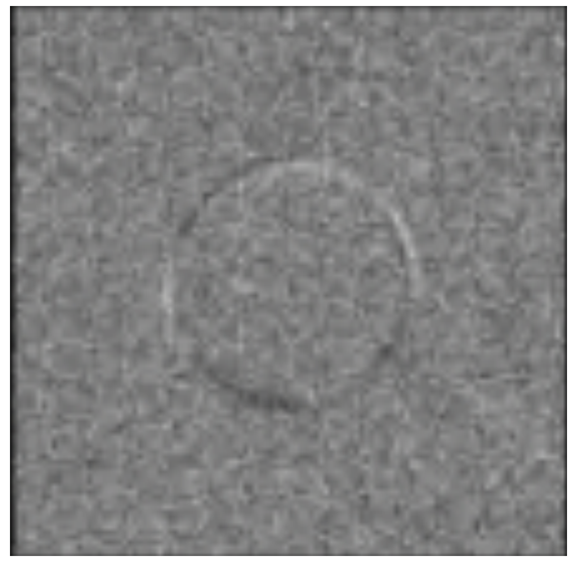
Addressing Distributional Shifts: Model Misspecification?

Nicolas
Payot

We can map our over-constrained physical model to an over-parametrized latent space, and apply the same techique!





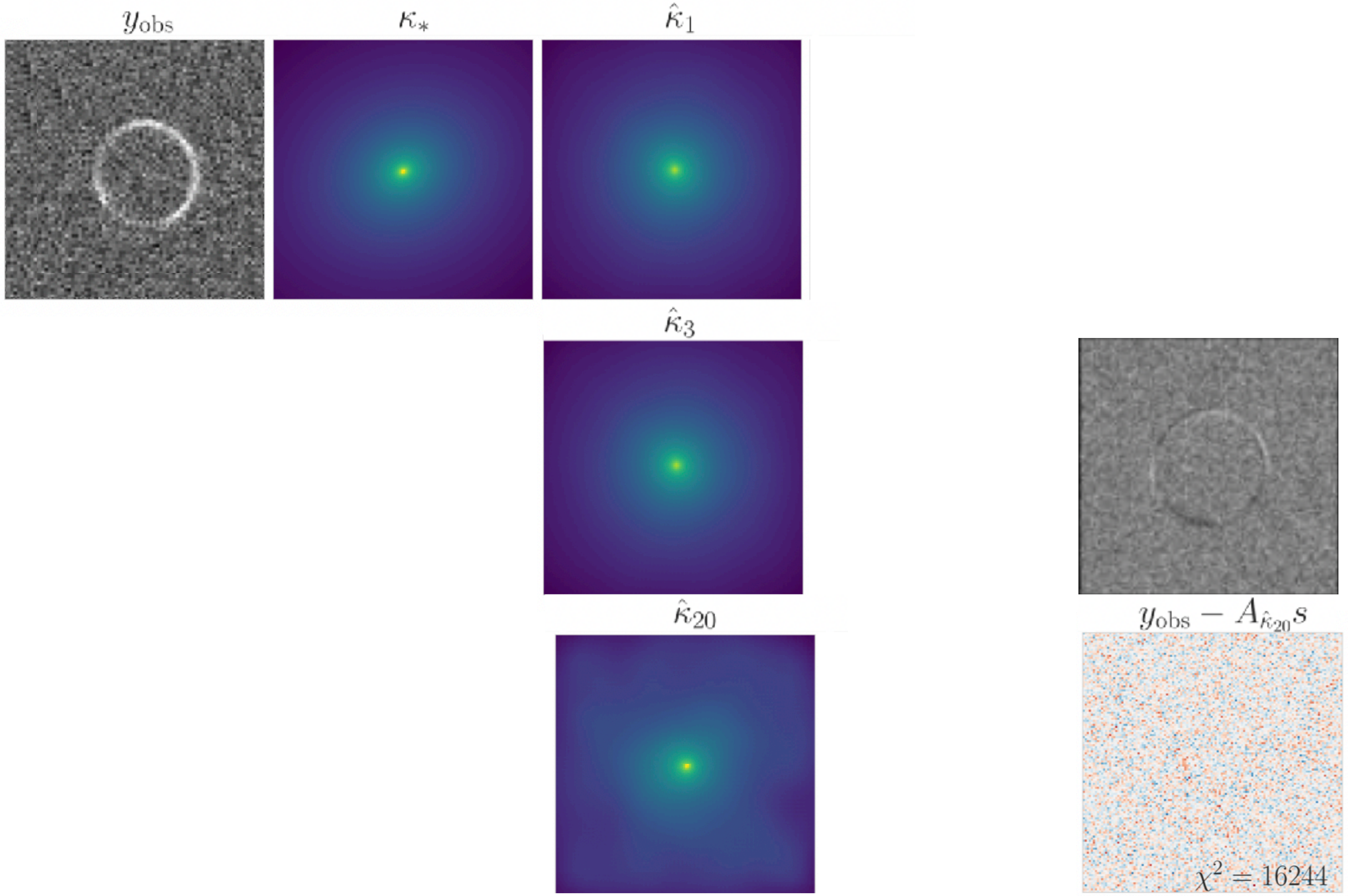


Payot et al., ML4Astro Workshop at ICML 2025,

Addressing Distributional Shifts: Model Misspecification?

We can map our over-constrained physical model to an over-parametrized latent space, and apply the same techique!

Nicolas Payot



Payot et al., ML4Astro Workshop at ICML 2025,

STRONG LENSING SIMULATION PIPELINE:
CAUSTIC

A fast, GPU-compatible, auto-differentiable, extremely modular simulation pipeline for all your strong lensing needs.

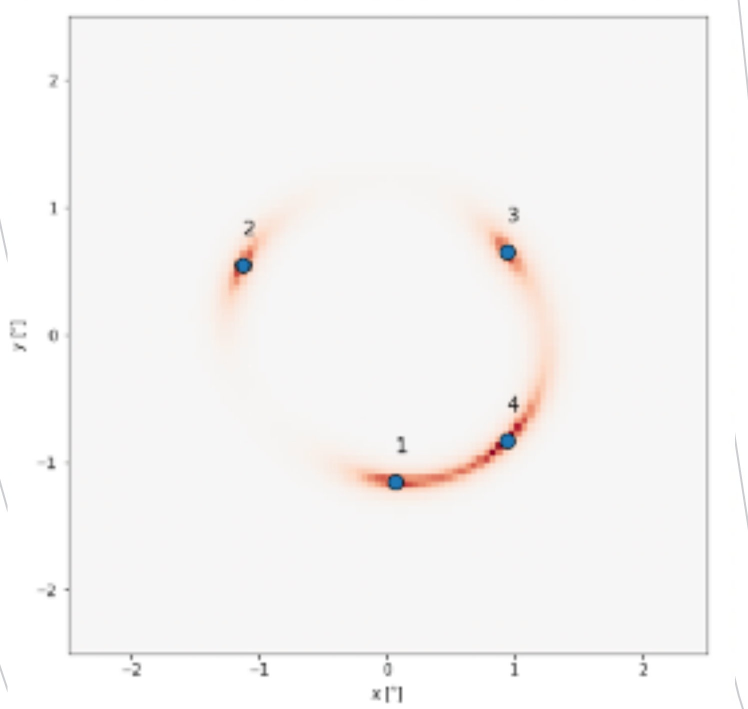
1) Lens and source from analytic profiles or pixelated images/densities

2) Multiplane lensing

3) Line of sight mass distributions

4) Fast microlensing simulations

5) Time-delays



Connor Stone

Adam Coogan

Andi Filipp

Alex Adam

Misha Barth

Charles Wilson

https://github.com/Ciela-Institute/caustics

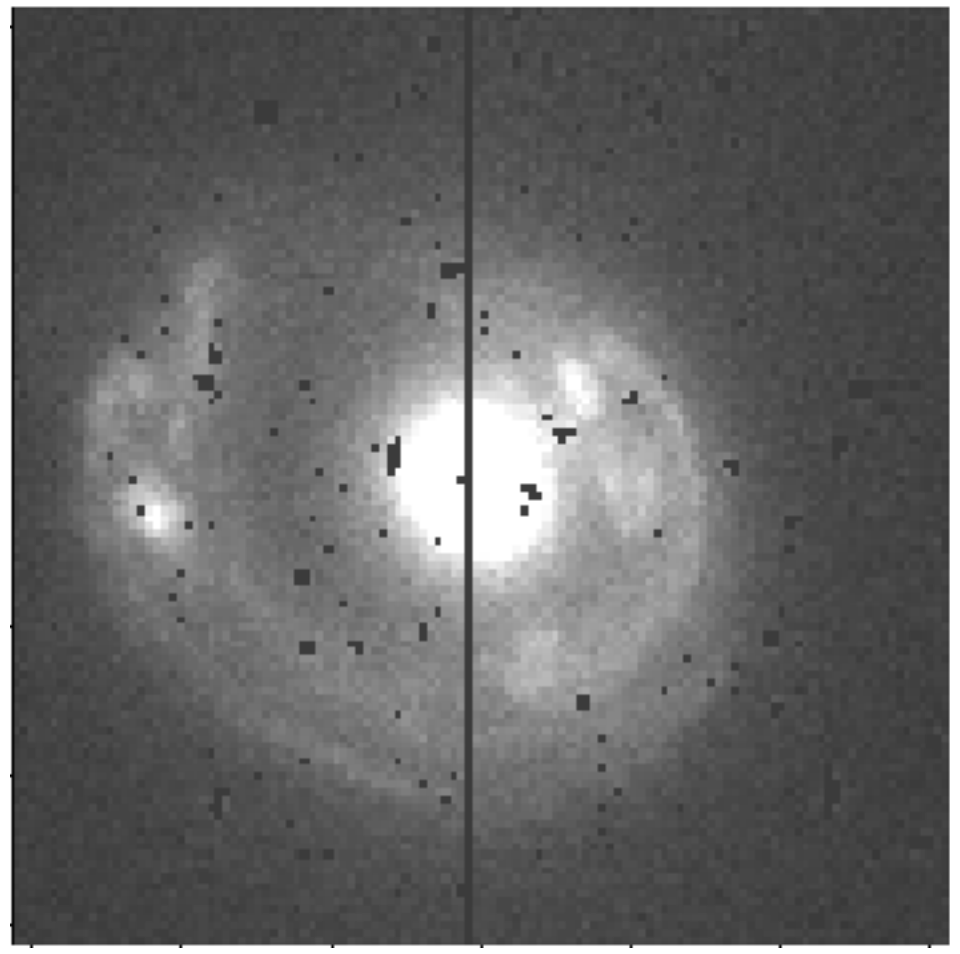
https://caustics.readthedocs.io/en/latest/intro.html

BONUS!

1. Data-driven priors
2. Data-driven likelihoods
3. Accuracy metrics
4. Out-of-Distribution accuracy
5. Data!

TEST CASE: SDSSJ1430+4105

Lens Observation



Discovered in 2006 from SDSS;

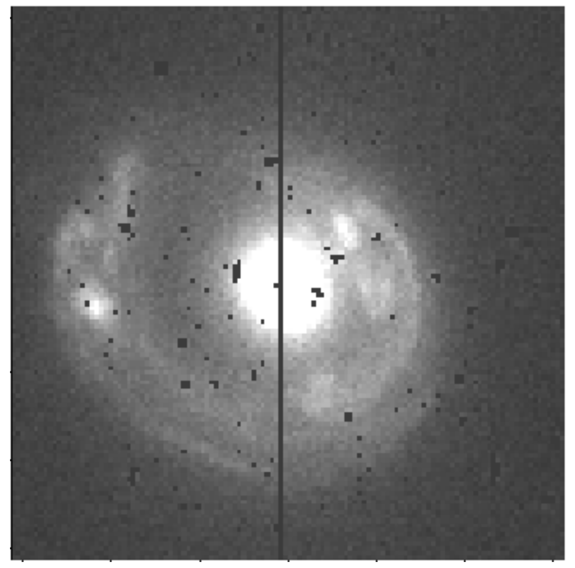
Hubble Space Telescope Imaging;

Distance: z_l ~ 0.285, z_s ~ 0.575.

Reconstructing the source galaxy is challenging, even with non-MNIST sources.

Joint inference of lens and source

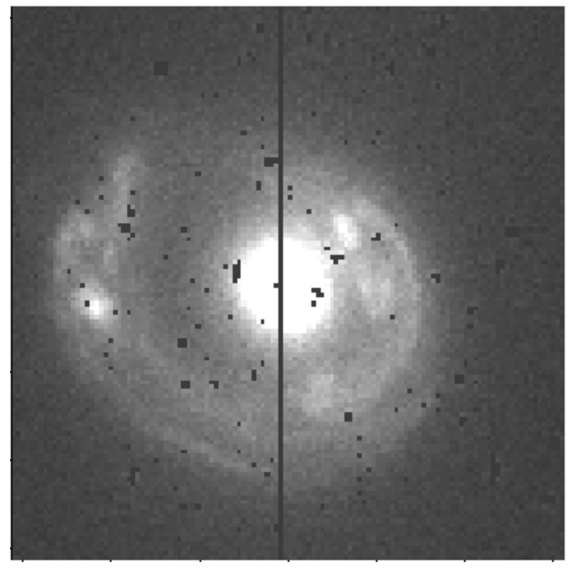
Lens
Observation



RECONSTRUCTING THE SOURCE GALAXY IS CHALLENGING, EVEN WITH NON-MNIST SOURCES.

Joint inference of lens and source

Lens Observation

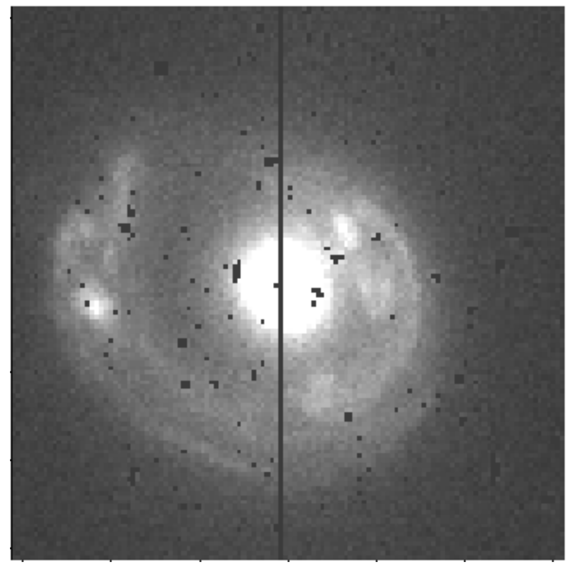


Bolton et al. 2008

RECONSTRUCTING THE SOURCE GALAXY IS CHALLENGING, EVEN WITH NON-MNIST SOURCES.

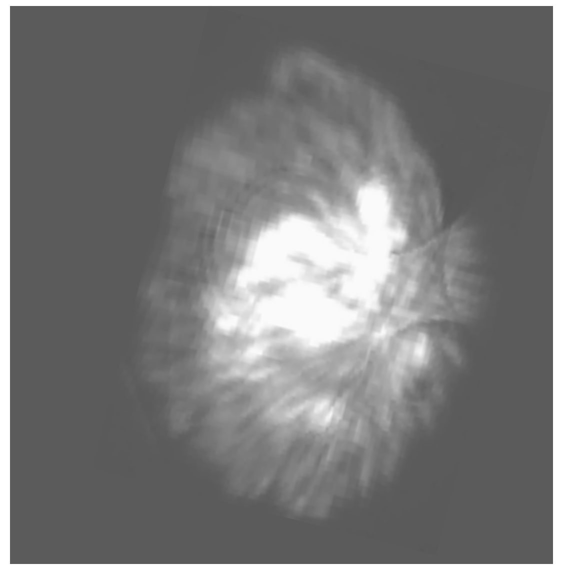
Joint inference of lens and source

Lens
Observation



Bolton et al.
2008

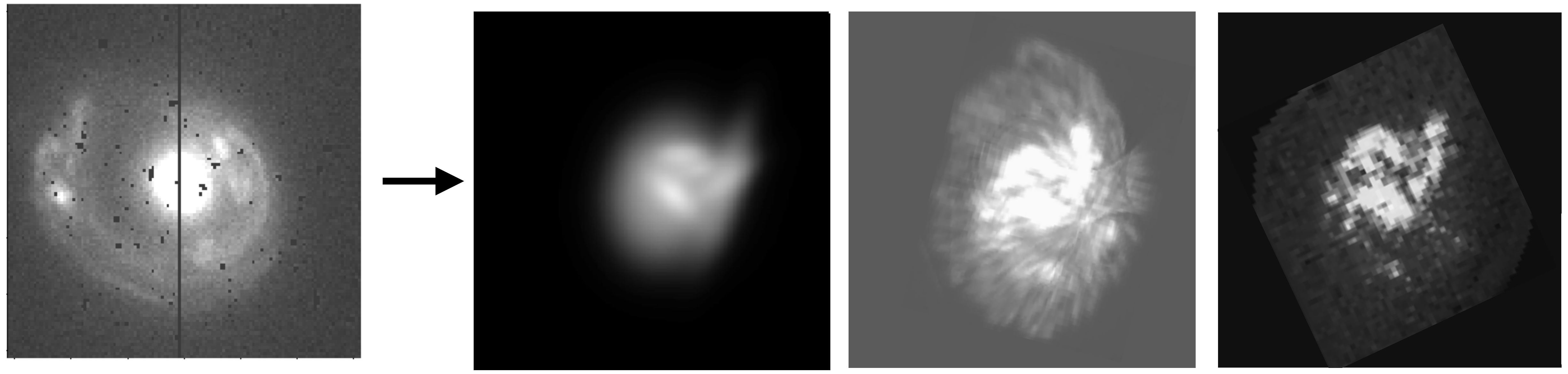
Eichner et al.
2012



Reconstructing the source galaxy is challenging, even with non-MNIST sources.

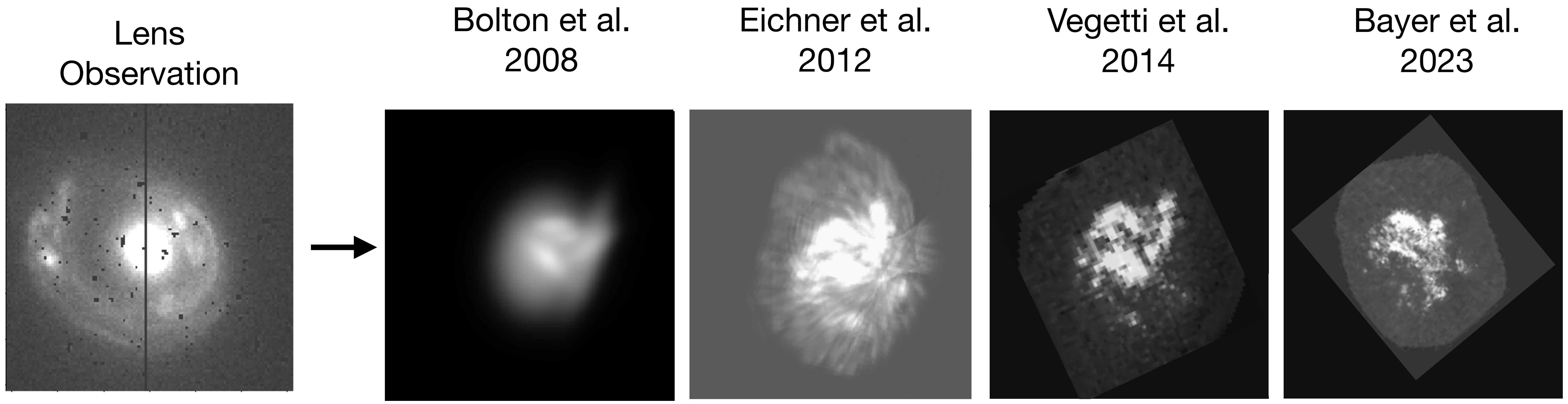
Joint inference of lens and source

Lens Observation | Bolton et al. 2008 | Eichner et al. 2012 | Vegetti et al. 2014



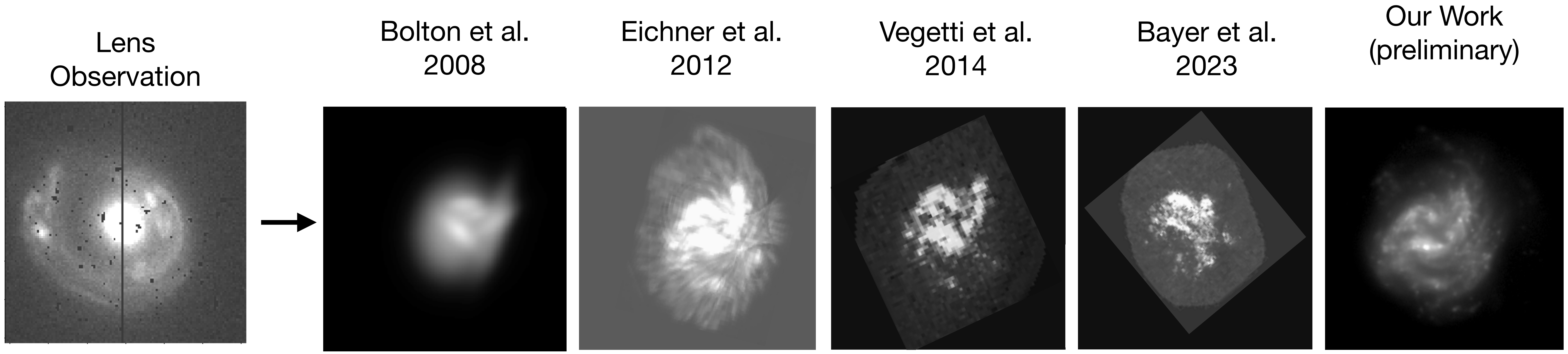
RECONSTRUCTING THE SOURCE GALAXY IS CHALLENGING, EVEN WITH NON-MNIST SOURCES.

Joint inference of lens and source



Reconstructing the source galaxy is challenging, even with non-MNIST sources.

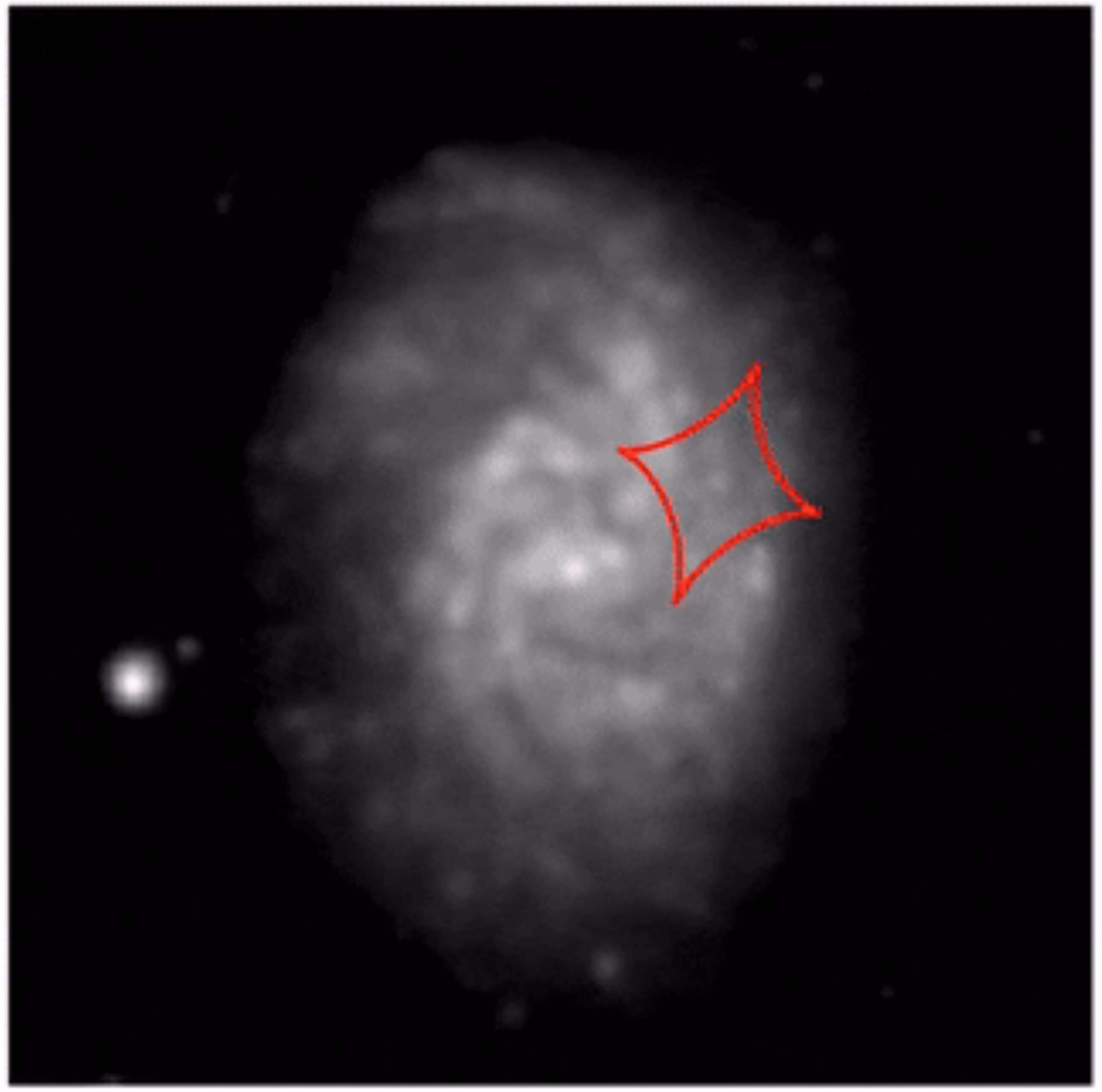
Joint inference of lens and source



We can sample the posterior over possible sources, marginalizing over possible lens models

Uncertainties show that the complex morphology is reconstructed with high confidence.

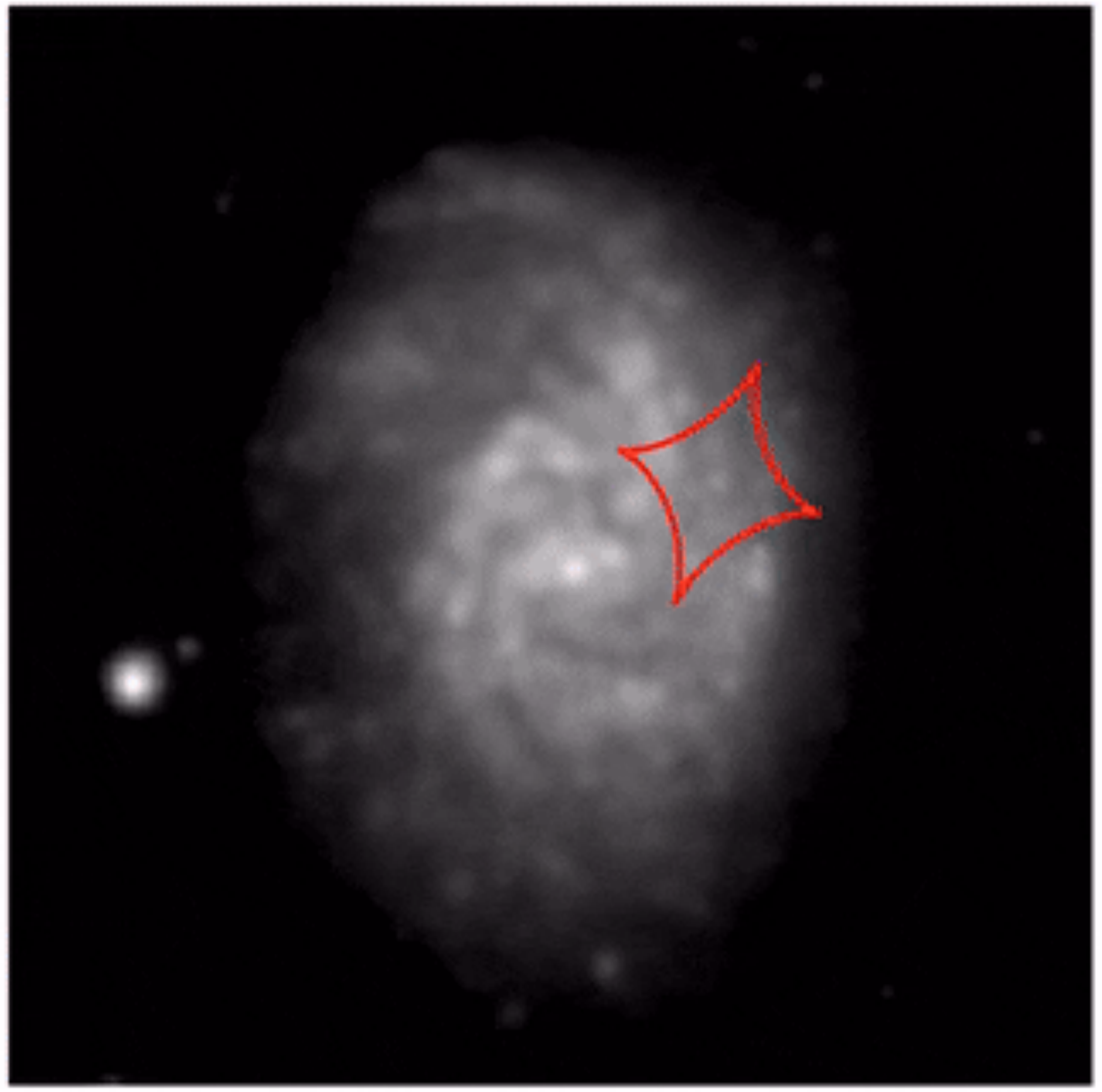
Galaxy Model



WE CAN SAMPLE THE POSTERIOR OVER POSSIBLE SOURCES, MARGINALIZING OVER POSSIBLE LENS MODELS

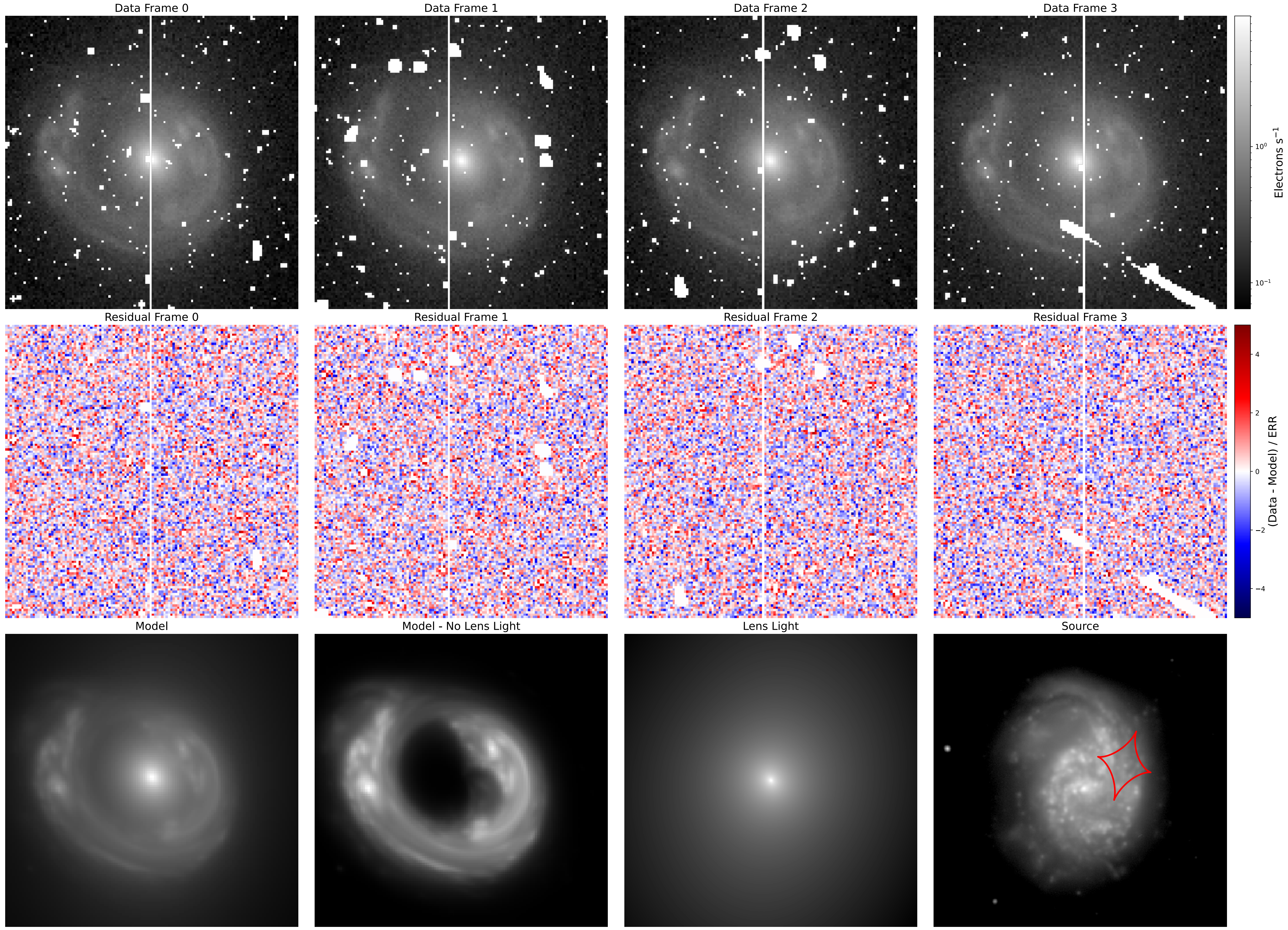
Uncertainties show that the complex morphology is reconstructed with high confidence.

Galaxy Model



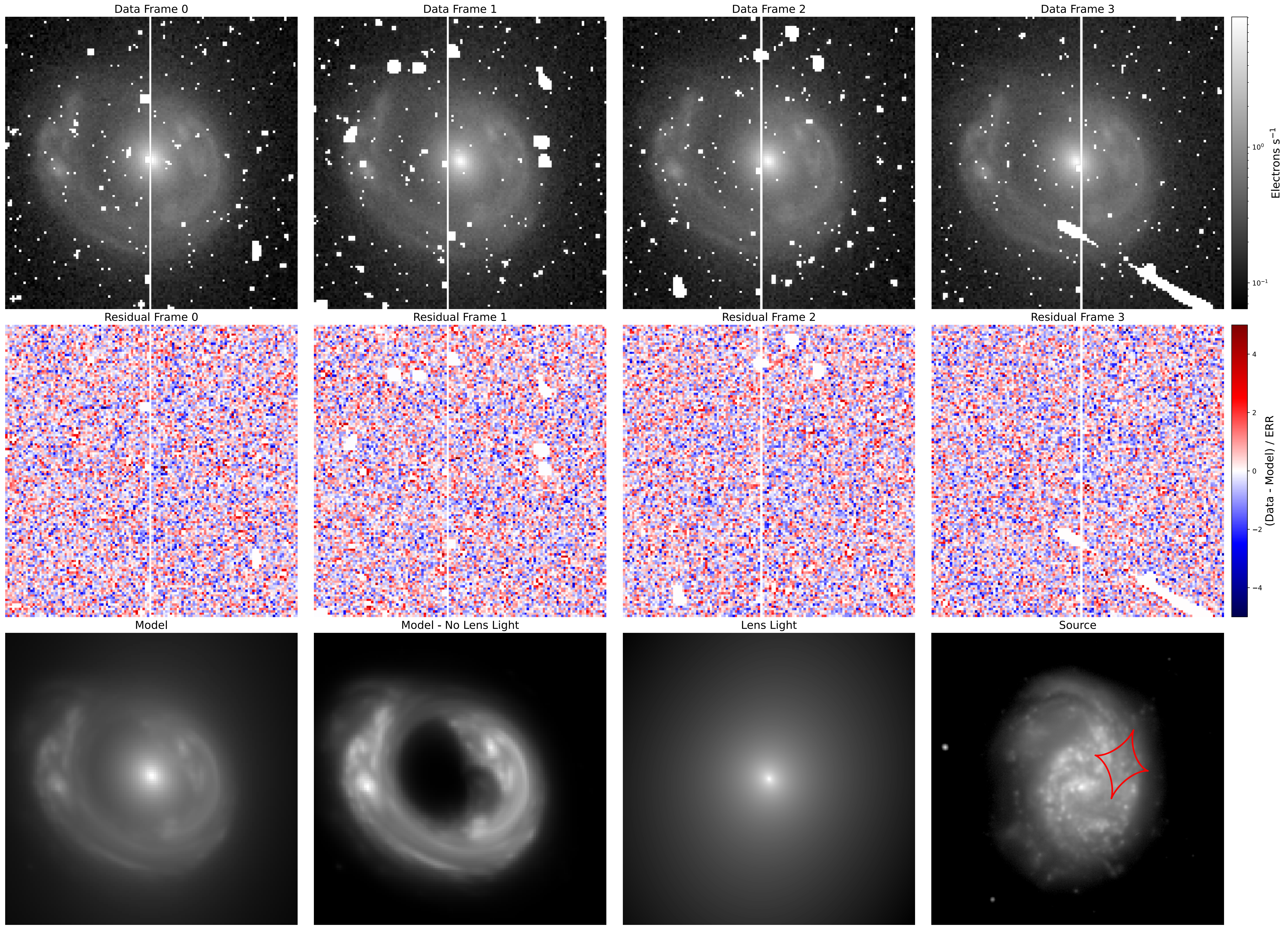
How well does our galaxy model reconstruct the lensed observation?

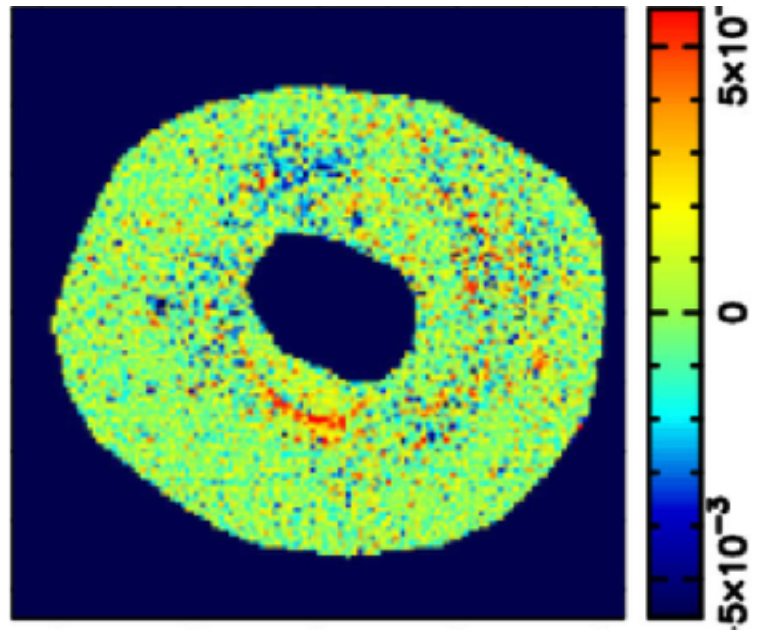
We can extracted near-maximal information from the data.



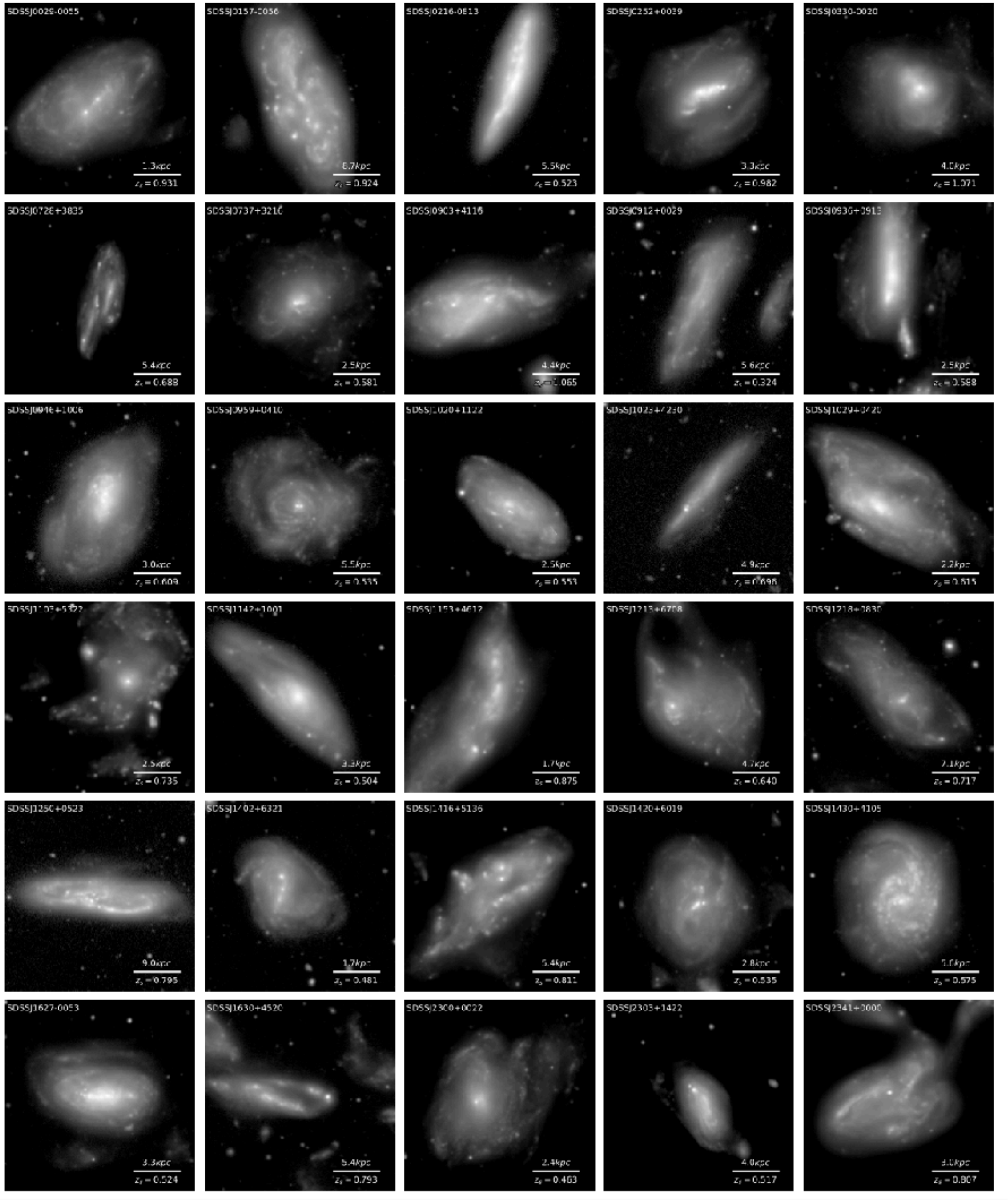
How well does our galaxy model reconstruct the lensed observation?

We can extracted near-maximal information from the data.



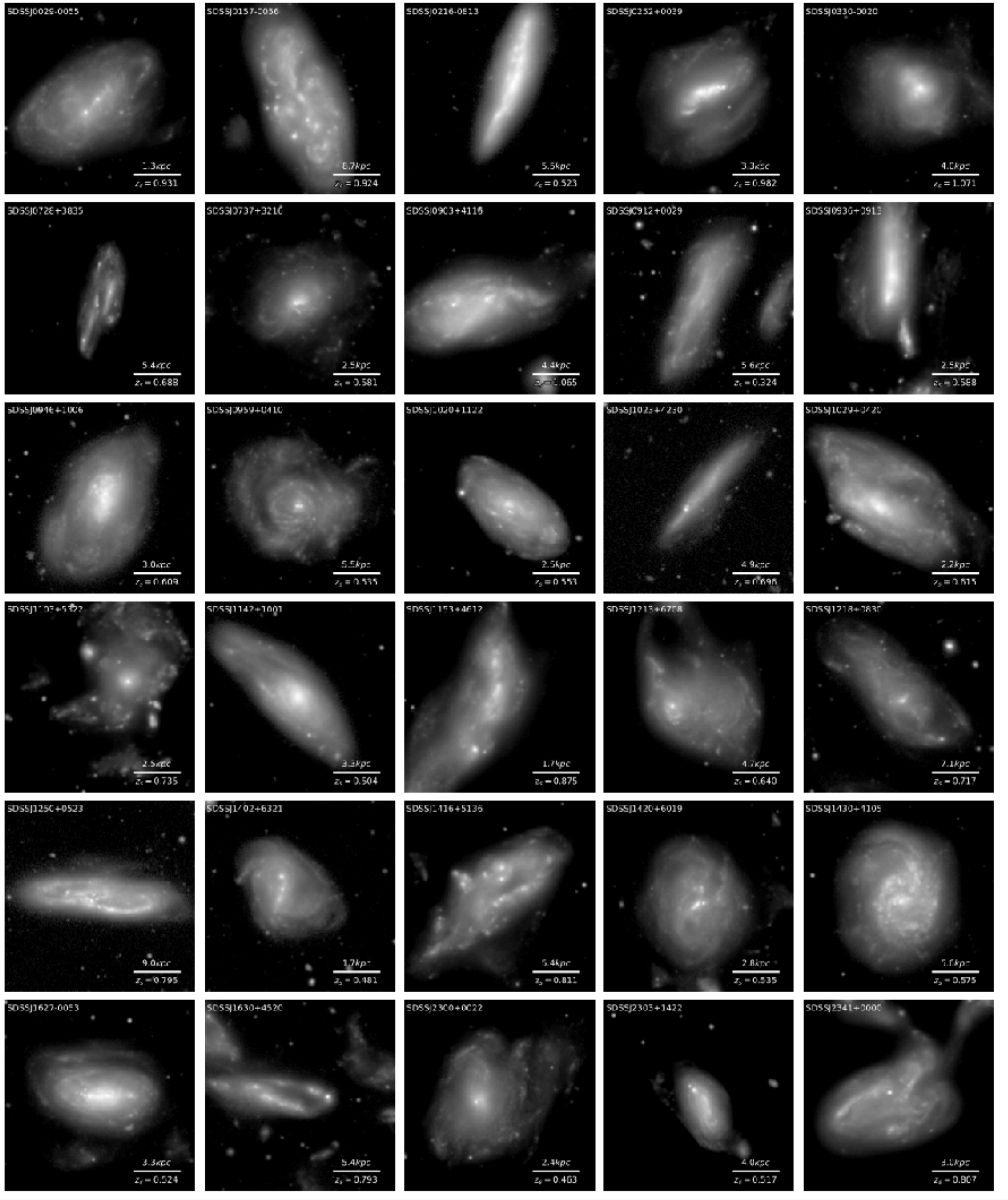


Bayer et al. 2023



'Family photo' of reconstructed galaxies

z~0.6 to 1 (5 to 8 billion lightyears away)



'Family photo' of reconstructed galaxies
z~0.6 to 1 (5 to 8 billion lightyears away)

Source: SDSS DR10

z=0.7 About 6.3 billion light-years away

Final thoughts…

I presented a lot of exemples related to strong lensing data analysis, but the structure of many inverse problems in astrophysics and cosmology is similar, and they face similar challenges.

By tackling to the problem of explicit, high-dimensional inference, we have access to our posterior samples, and that allows us to open the black box, and understand where the information in our inference is coming from. This is (one of) the way we are going to build trust in inference methods that incorporate ML.