Towards Inference Assembly:

Next steps for deep learning, simulation-based inference and astrophysical data

powered by
Swyft

SURF

erc

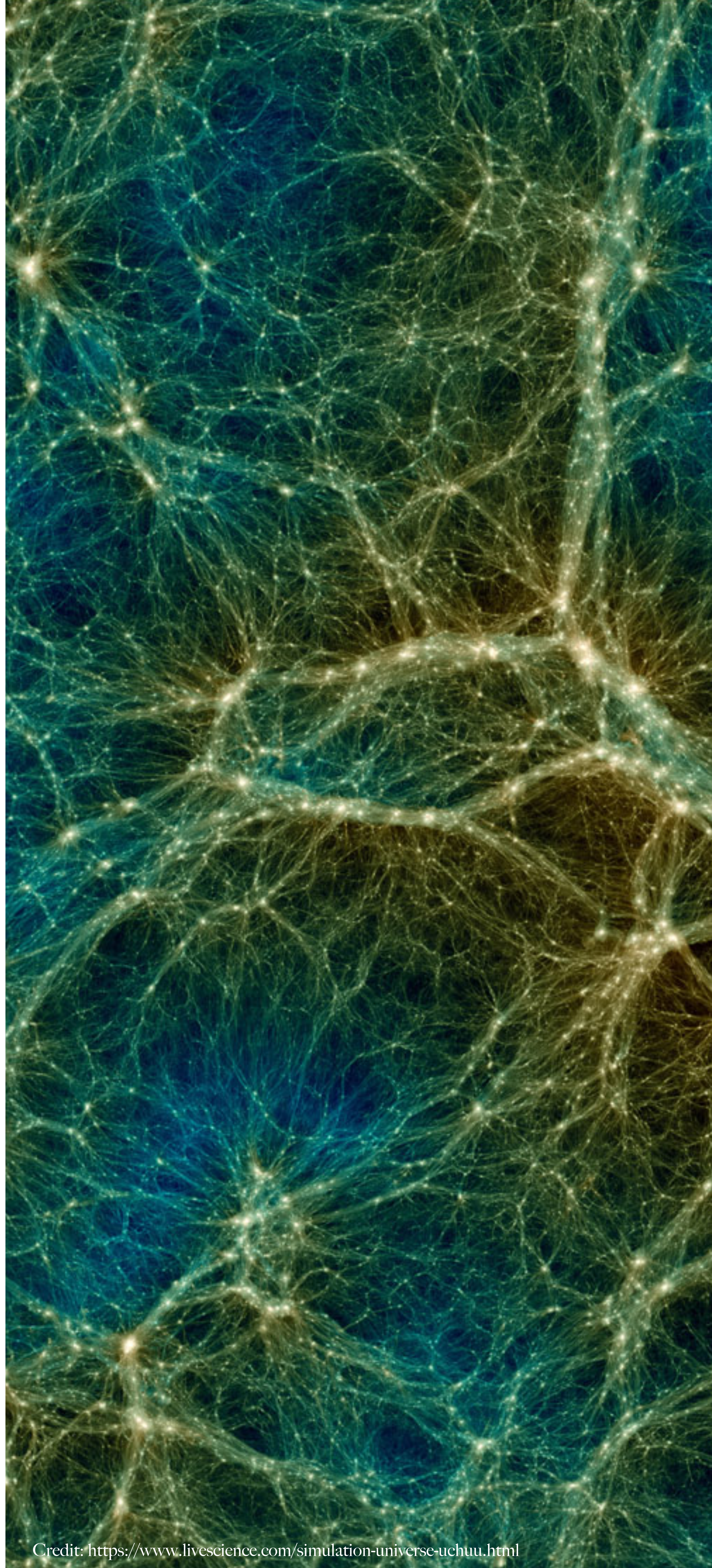
ERC CoG UnDark,
Proj. No. 864035

Christoph Weniger
Dark Side of the Universe
ICTP Kigali, 12 July 2023

GRAPPA
GRavitation AstroParticle Physics Amsterdam

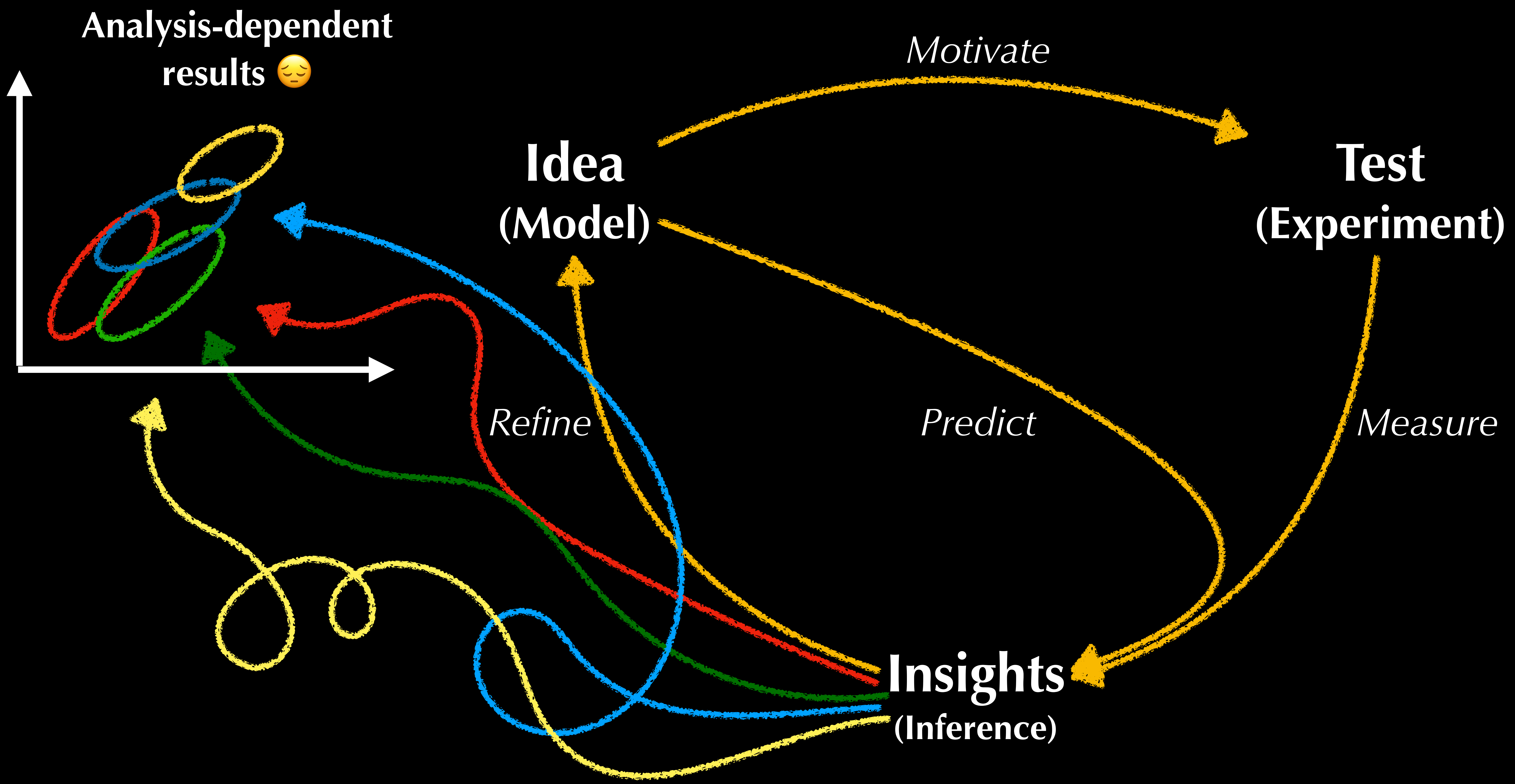
AI4Science Lab

UNIVERSITY
OF AMSTERDAM

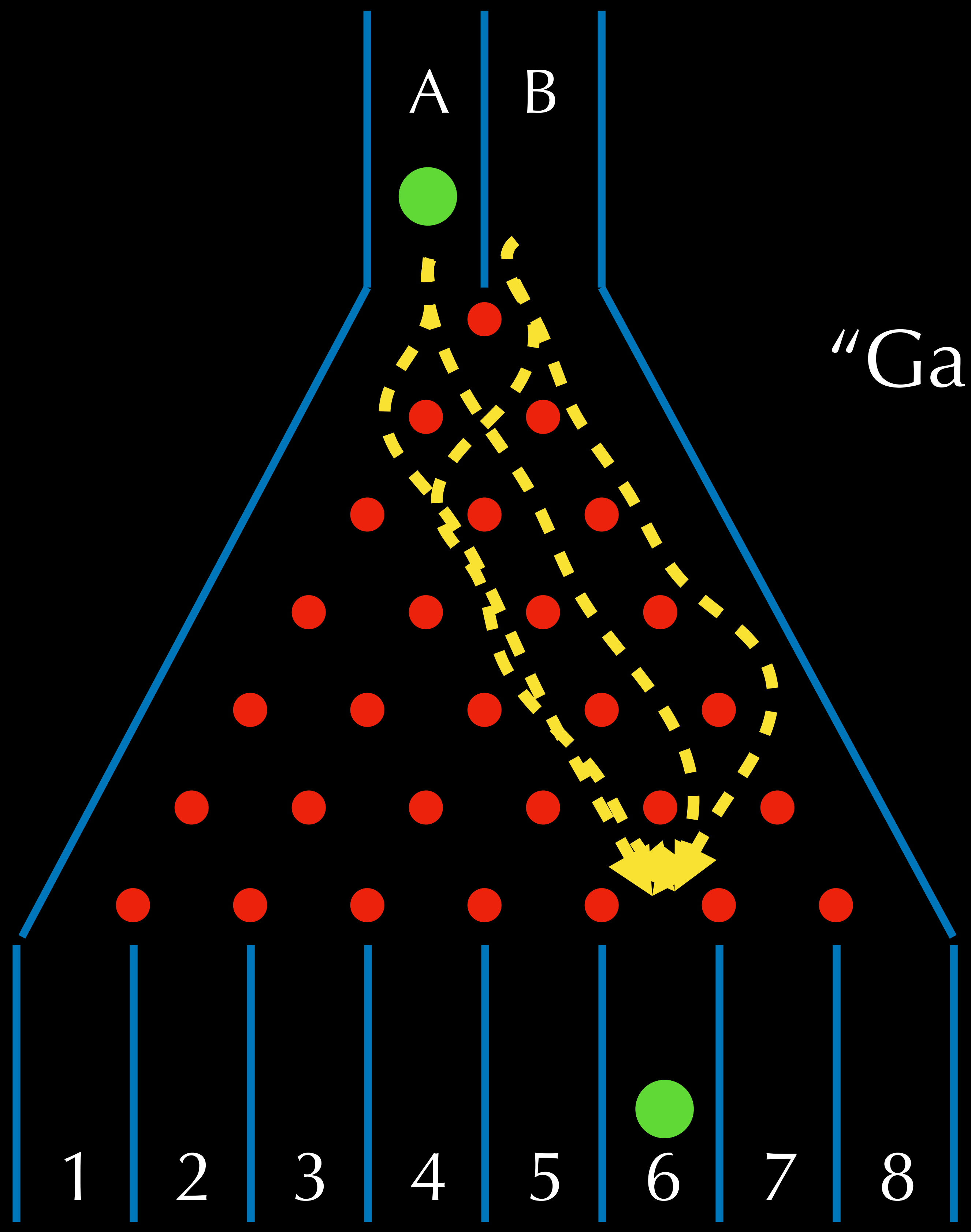

Credit: https://www.livescience.com/simulation-universe-uchuu.html

The limits of likelihood-based inference

Inference-limited science

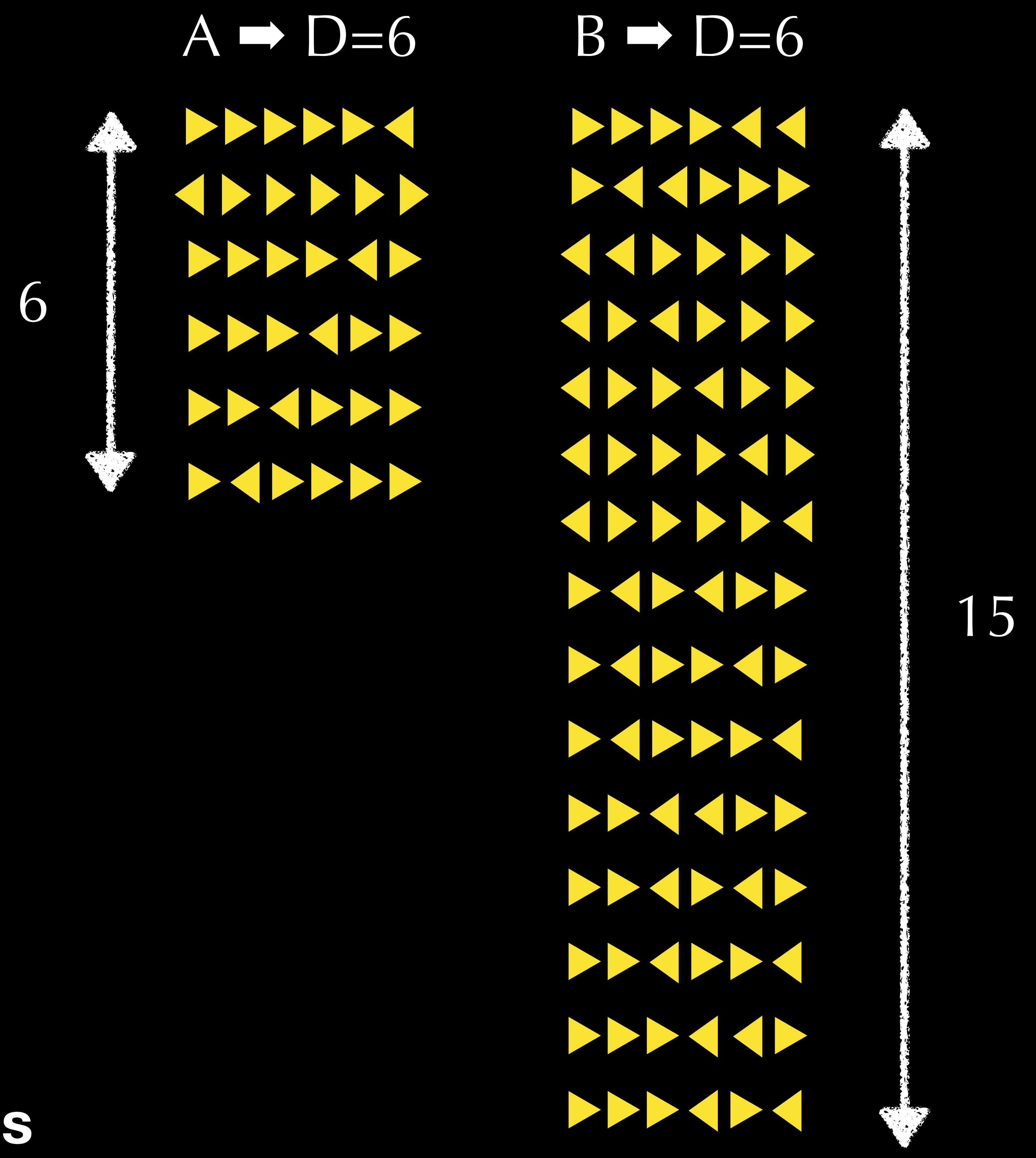


Traditional statistical inference



$D = 6$

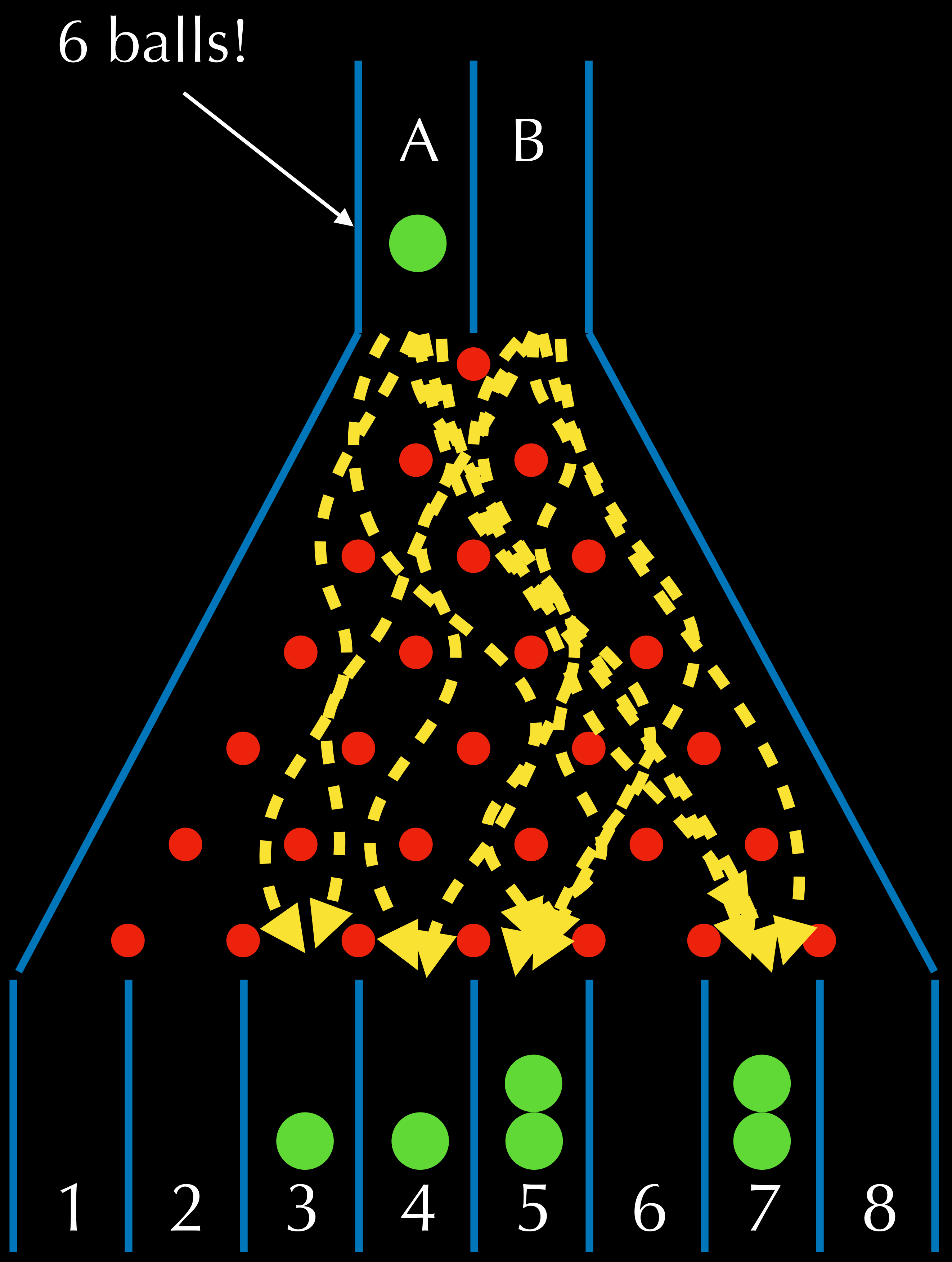
“Galton Board”



Bayesian statistics
(T. Bayes, 1763)

$$P(B \mid D = 6) = \frac{15}{15 + 6} \simeq 71\,\%$$

Traditional statistical inference



$D = (3,4,5,5,7,7)$

A ➡ $D = (3,4,5,5,7,7)$

(▶▶◀▶▶▶, ▶▶◀▶▶▶, ▶▶◀▶▶▶,
▶▶◀▶▶▶, ▶▶◀▶▶▶, ▶▶◀▶▶▶)
…
…
…
…

67500

B ➡ $D = (3,4,5,5,7,7)$

(▶▶◀▶▶▶, ▶▶◀▶▶▶, ▶▶◀▶▶▶,
▶▶◀▶▶▶, ▶▶◀▶▶▶, ▶▶◀▶▶▶)
…
…
…
…

1.296.000

$$P(B \mid D = (3,4,5,5,7,7)) = \frac{1.296.000}{1.296.000 + 67500} \simeq 93\,\%$$

Example: Strong gravitational lensing

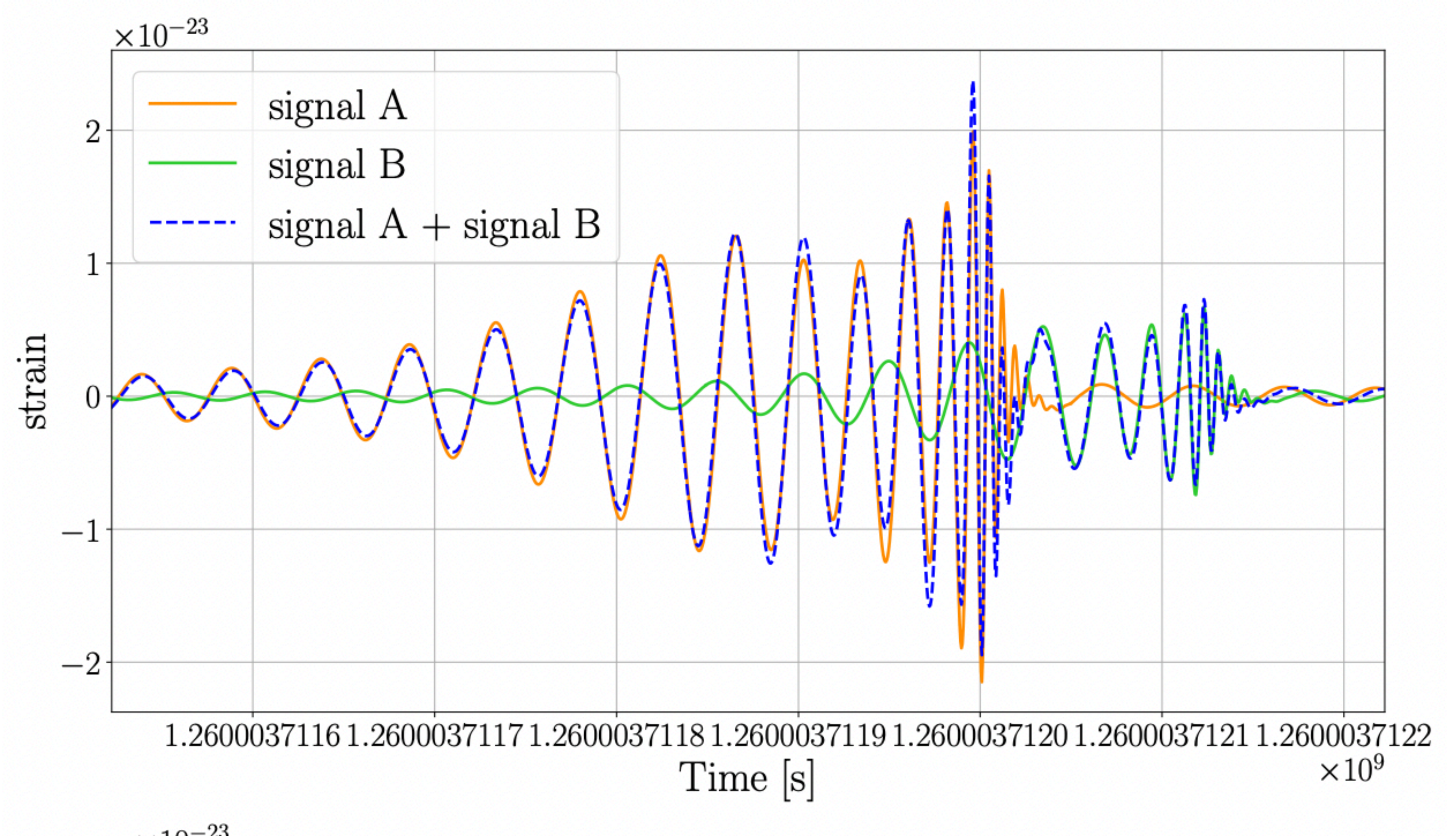
What is the dark matter content of the foreground galaxy?

Credit: ALMA (ESO/NRAO/NAOJ), L. Calçada (ESO), Y. Hezaveh et al.

Example: Overlapping GW signals

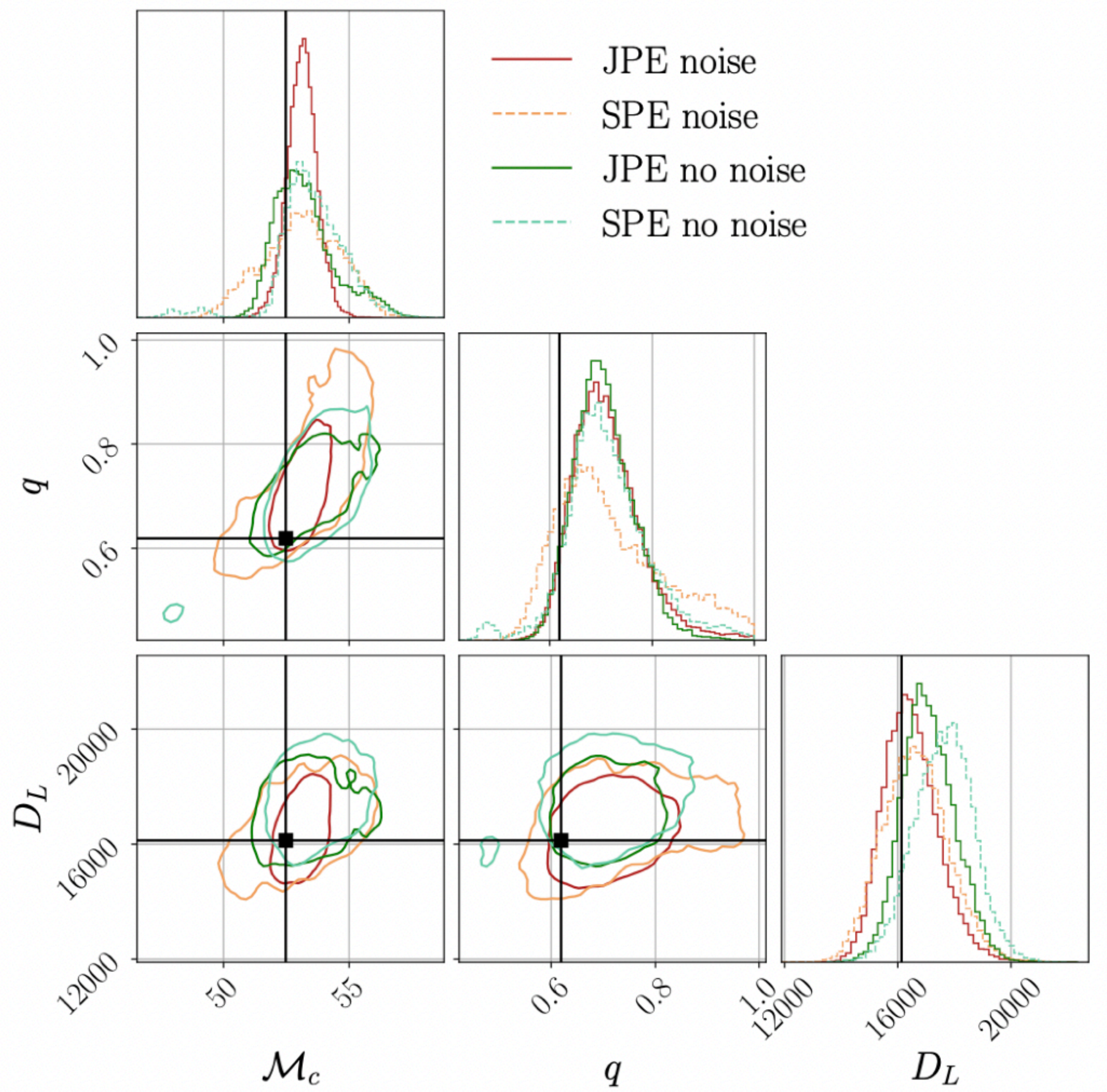
Inference is a challenging task

*"Parameter estimation methods for **analyzing overlapping gravitational wave signals** in the third-generation detector era"*



Nested sampling
(Bilby & Dynesty)

30-dim parameter space



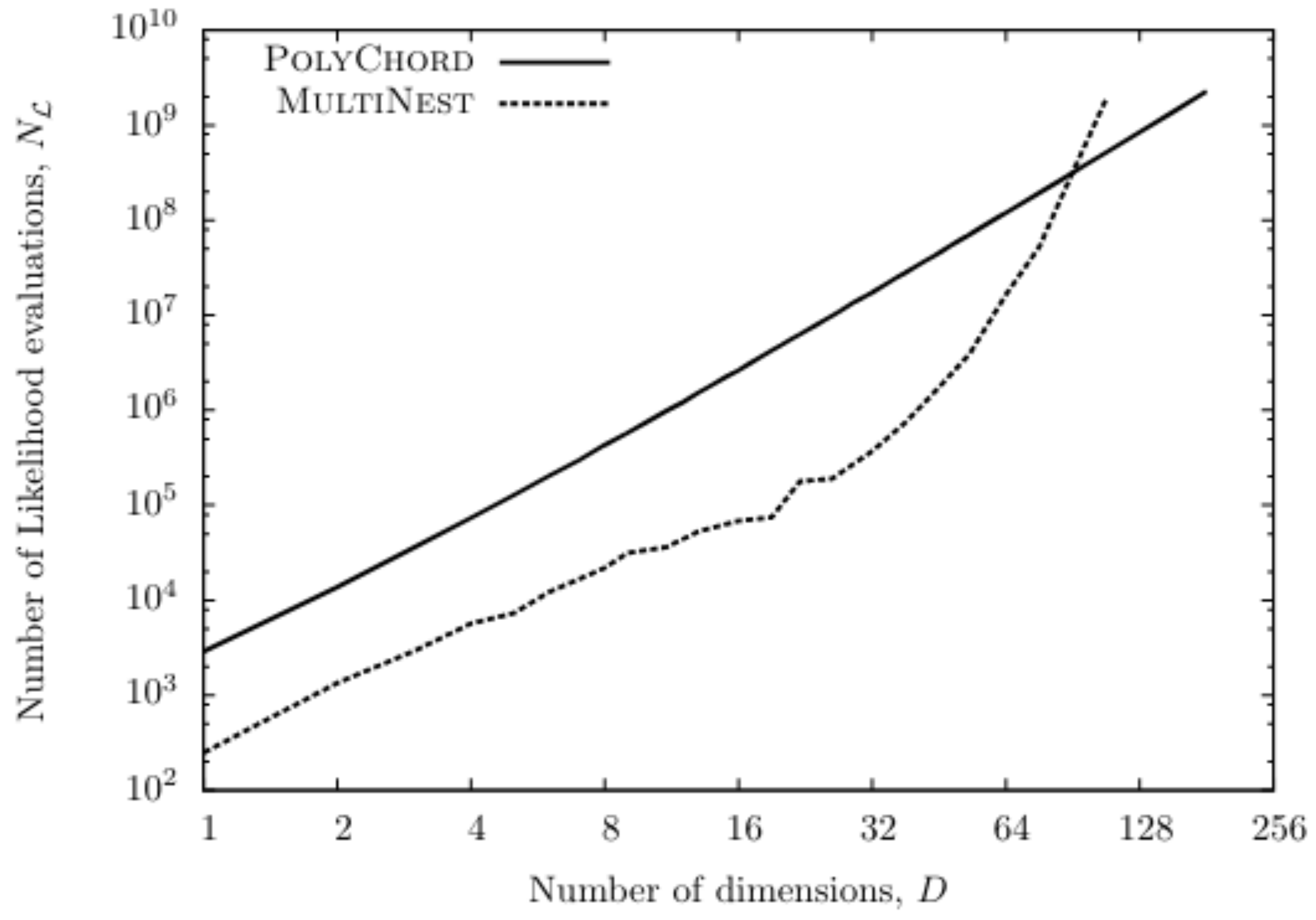
*"…One of the major issues with the methods suggested here is the computational time required, as **the data analysis takes up to a few months** for overlapping binary black hole mergers…"*

Janquart+ 2211.01304

…in practice, that is not easy

(indicative, scaling is problem specific)

- In many cases, the likelihood function $p(\mathbf{x} \mid \mathbf{z})$ cannot be evaluated exactly, we just have the simulator.
- Even if we know the likelihood function, we must "count the paths" (MCMC, HMC, nested sampling, Gibbs sampling, analytic integration, …). This is in general costly.
- Increased model realism typically means more uncertainties and parameters.
- In practice, shortcuts are taken (iterative schemes, simplifying assumptions, surrogate models, perturbative approaches, …). The consequences can be difficult to quantify.



We can afford: $p(\mathbf{z}_A \mid \mathbf{x}, \mathbf{z}_B^{\mathrm{fid}})$

$$\neq$$

We want: $p(\mathbf{z}_A \mid \mathbf{x}) = \int d\mathbf{z}_B \, p(\mathbf{z}_A \mid \mathbf{x}, \mathbf{z}_B) p(\mathbf{z}_B \mid \mathbf{x})$

Feroz+ 0809.3437 (MultiNest)
Handley+ 1502.01856 (PolyChord)

Deep learning and simulation-based inference

Deep neural networks

Training
(evaluating examples & adjusting connections)

$$\sigma(w_1 a_1 + w_2 a_2 + \ldots w_N a_N + b)$$

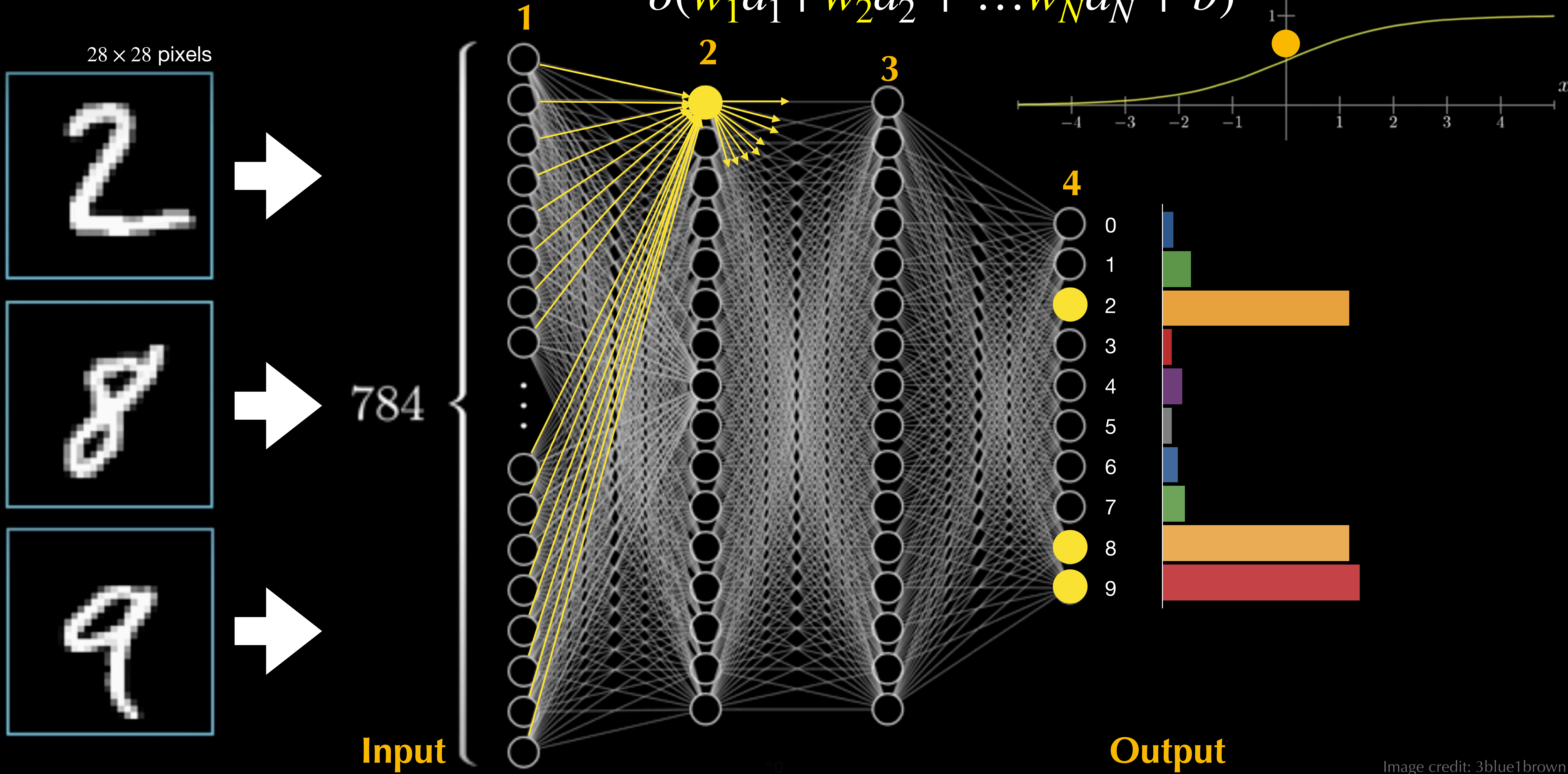


Image credit: 3blue1brown

AI-assisted statistical inference

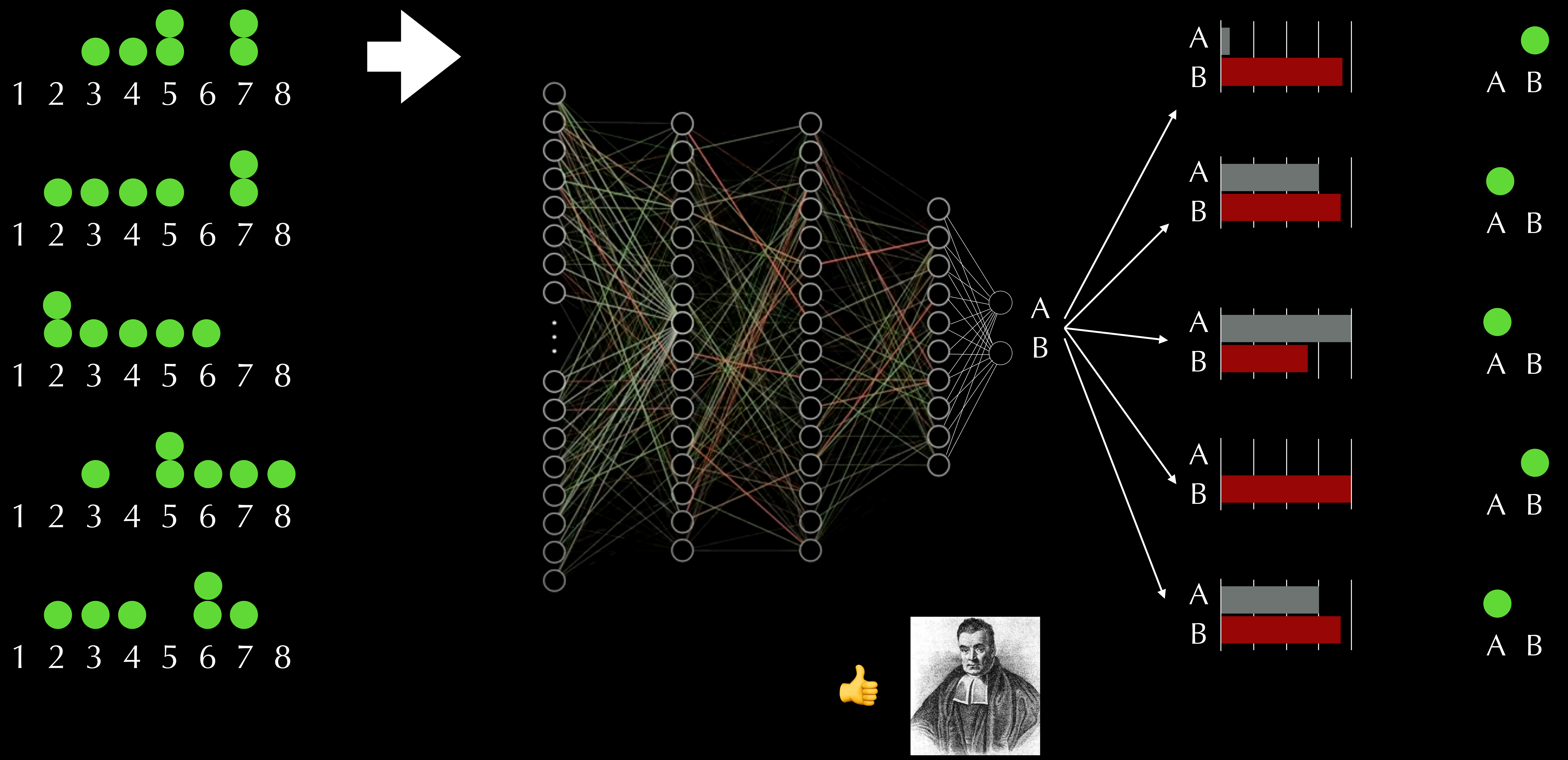
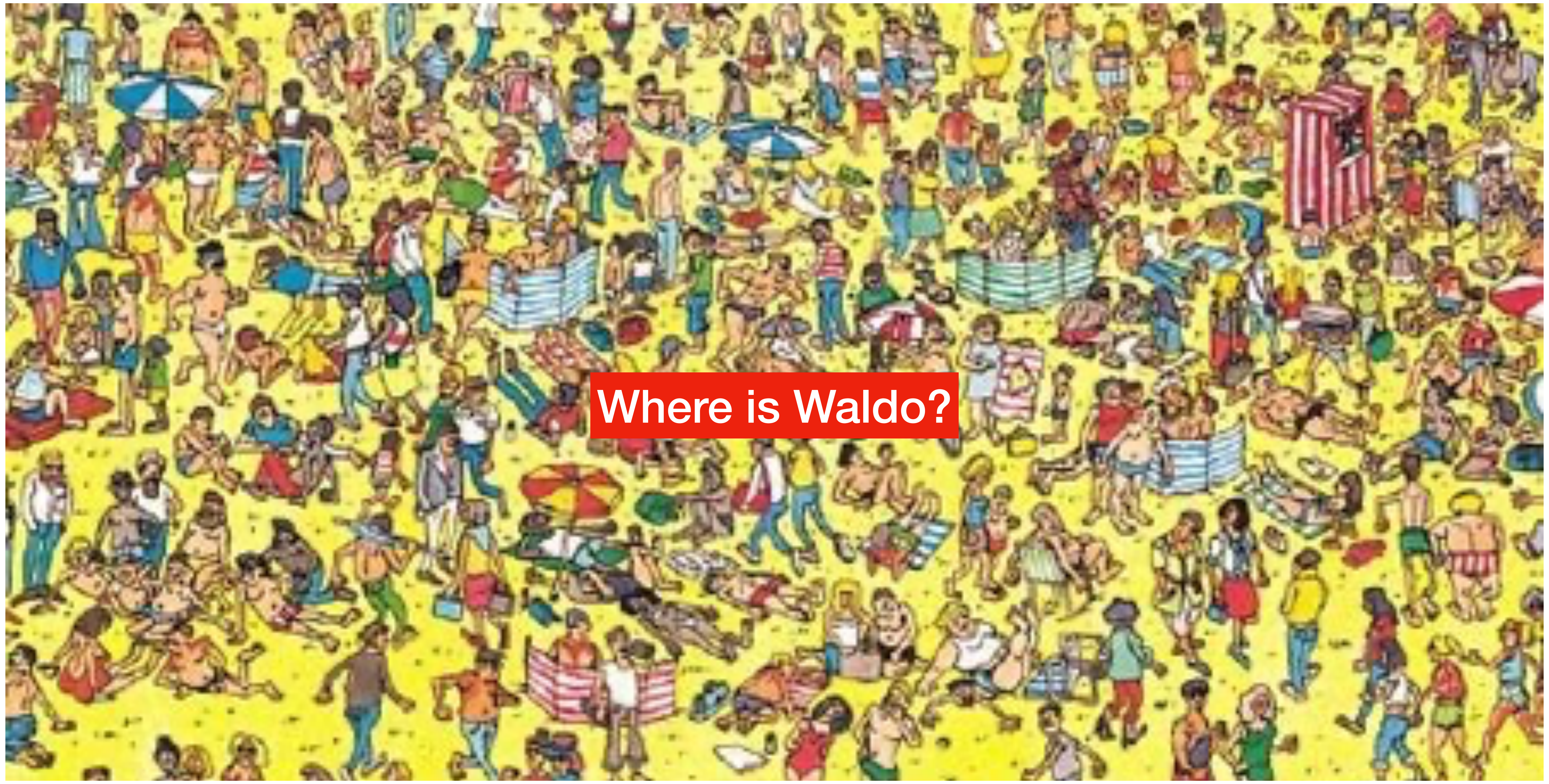


Image credit: 3blue1brown

Finding Waldo: MCMC vs deep learning



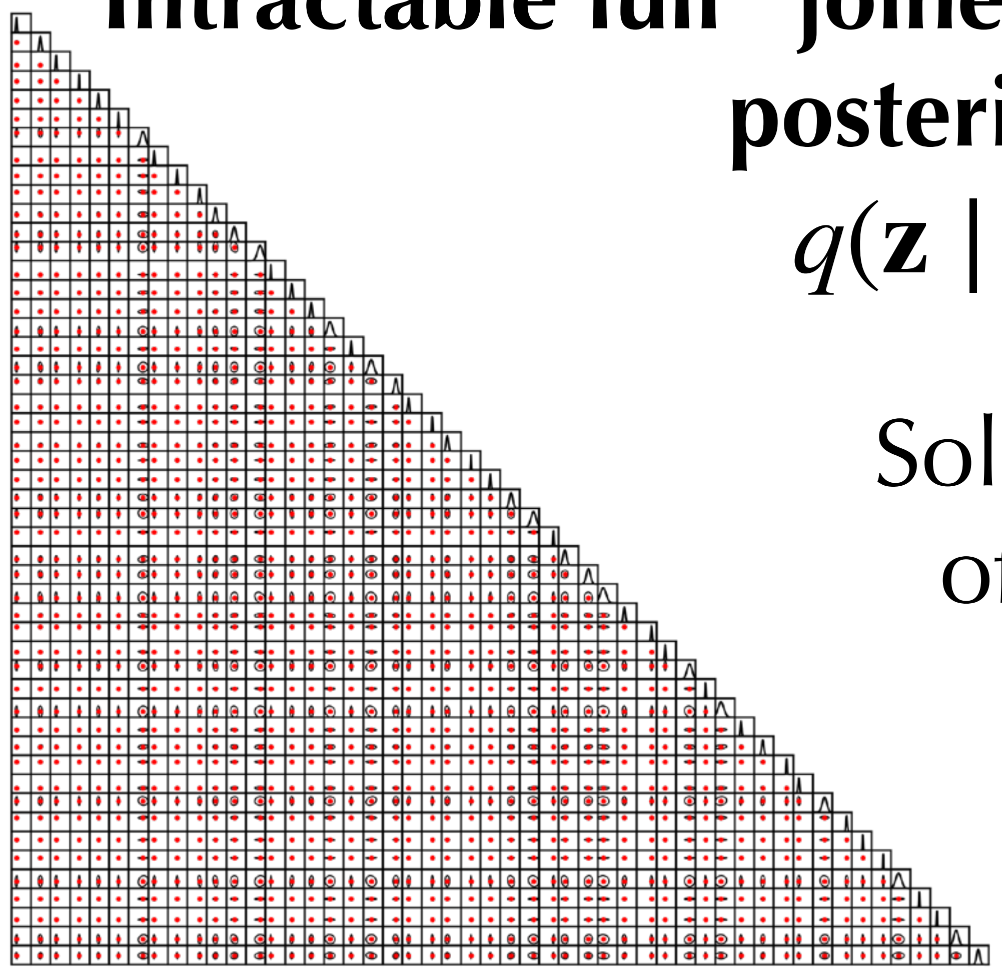
$$\underbrace{q_\phi(\mathbf{z}_{\text{Waldo}} \mid \mathbf{x}_o)} = \int d\mathbf{z}_{\text{Lucia}} d\mathbf{z}_{\text{Oleg}} d\mathbf{z}_{\text{Sibilla}} d\mathbf{z}_{\text{Dion}} \cdots d\mathbf{z}_{\text{Noemi}} \; \underbrace{q_\phi(\mathbf{z}_{\text{Waldo}}, \mathbf{z}_{\text{Lucia}}, \mathbf{z}_{\text{Oleg}}, \mathbf{z}_{\text{Sibilla}}, \mathbf{z}_{\text{Dion}}, \cdots, \mathbf{z}_{\text{Noemi}} \mid \mathbf{x}_o)}$$

Simulation-based inference
Train a neural network to find Waldo's marginal posterior.

Joined inference
Run MCMC to explore ultra-high-dimensional model for every single aspect in the image. Then marginalise.

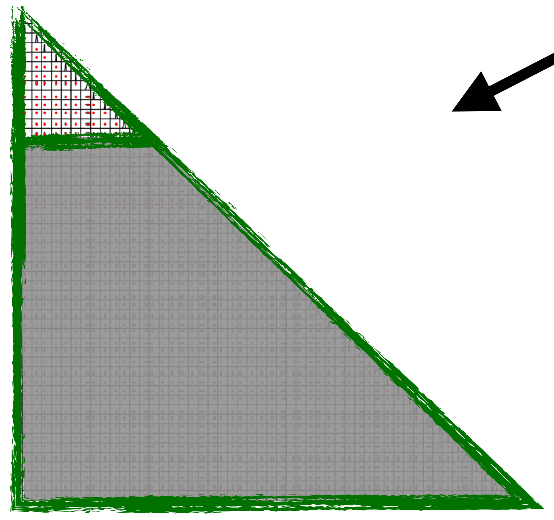
“Inference Assembly”

Intractable full “joined” posterior $q(\mathbf{z} \mid \mathbf{x})$



Traditional approach

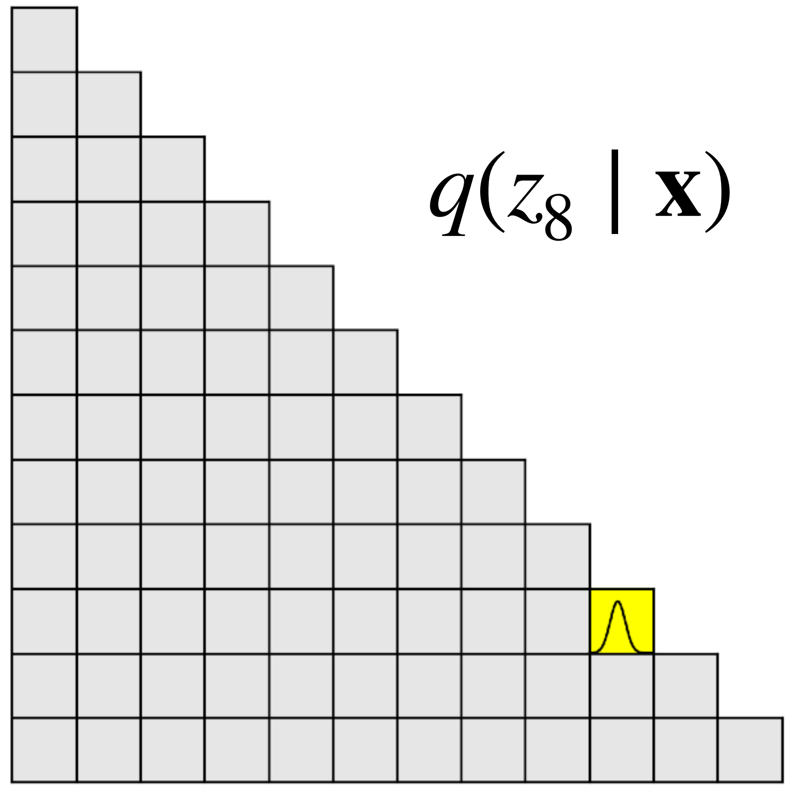
Solve simpler but approximate problem, and hope for the best



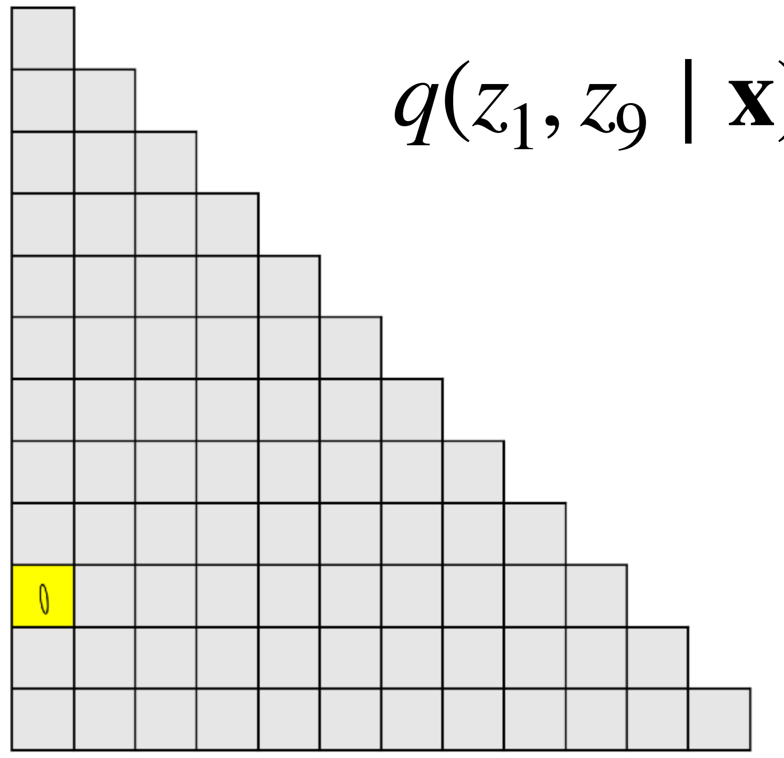
$q_{\text{approx}}(\mathbf{z}_1 \mid \mathbf{x})$

Simulation-based approach

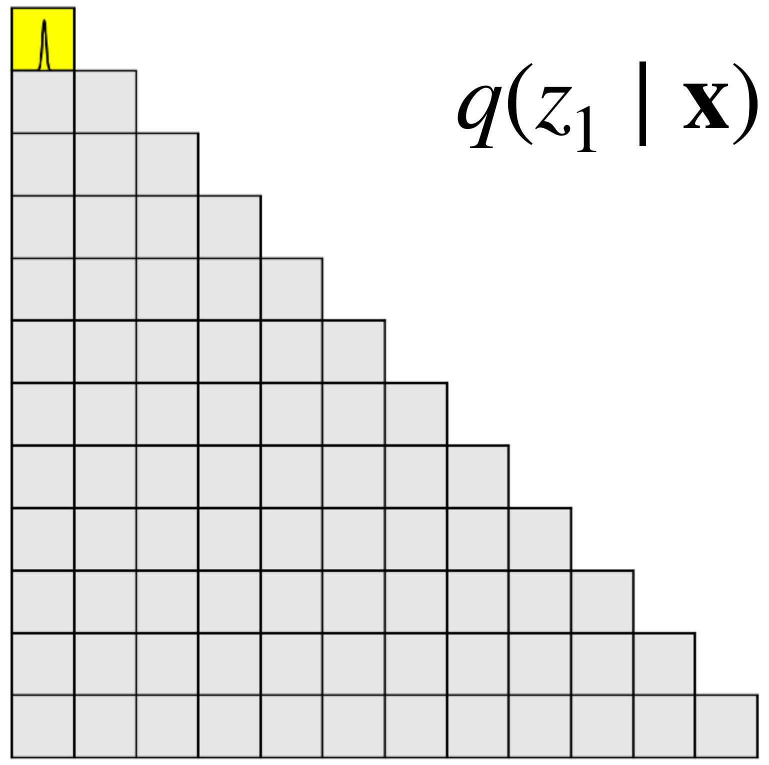
Solve arbitrary aspects of the *full* problem.



$q(z_8 \mid \mathbf{x})$



$q(z_1, z_9 \mid \mathbf{x})$



$q(z_1 \mid \mathbf{x})$

Going back to Bayes theorem
What can we approximate?

Data likelihood
↳ **Neural likelihood estimation** (NLE)

$$p(\mathbf{x}\,|\,\mathbf{z})$$

Posterior density function
↳ **Neural posteriors estimation** (NPE)

$$p(\mathbf{z}\,|\,\mathbf{x})$$

$$p(\mathbf{z}\,|\,\mathbf{x}) = \frac{p(\mathbf{x}\,|\,\mathbf{z})}{p(\mathbf{x})}p(\mathbf{z})$$

Likelihood-to-evidence ratio
↳ **Neural ratio estimation** (NRE)

$$r(\mathbf{x};\mathbf{z}) \equiv \frac{p(\mathbf{x}\,|\,\mathbf{z})}{p(\mathbf{x})} = \frac{p(\mathbf{z}\,|\,\mathbf{x})}{p(\mathbf{z})} = \frac{p(\mathbf{x},\mathbf{z})}{p(\mathbf{x})p(\mathbf{z})}$$

$\mathbf{x}$: Data

$\mathbf{z}$: Parameters

NRE = binary classification

Strategy: We estimate posteriors-to-prior ratio by training a binary classifier to discriminate between matching and scrambled (data, parameter) pairs.

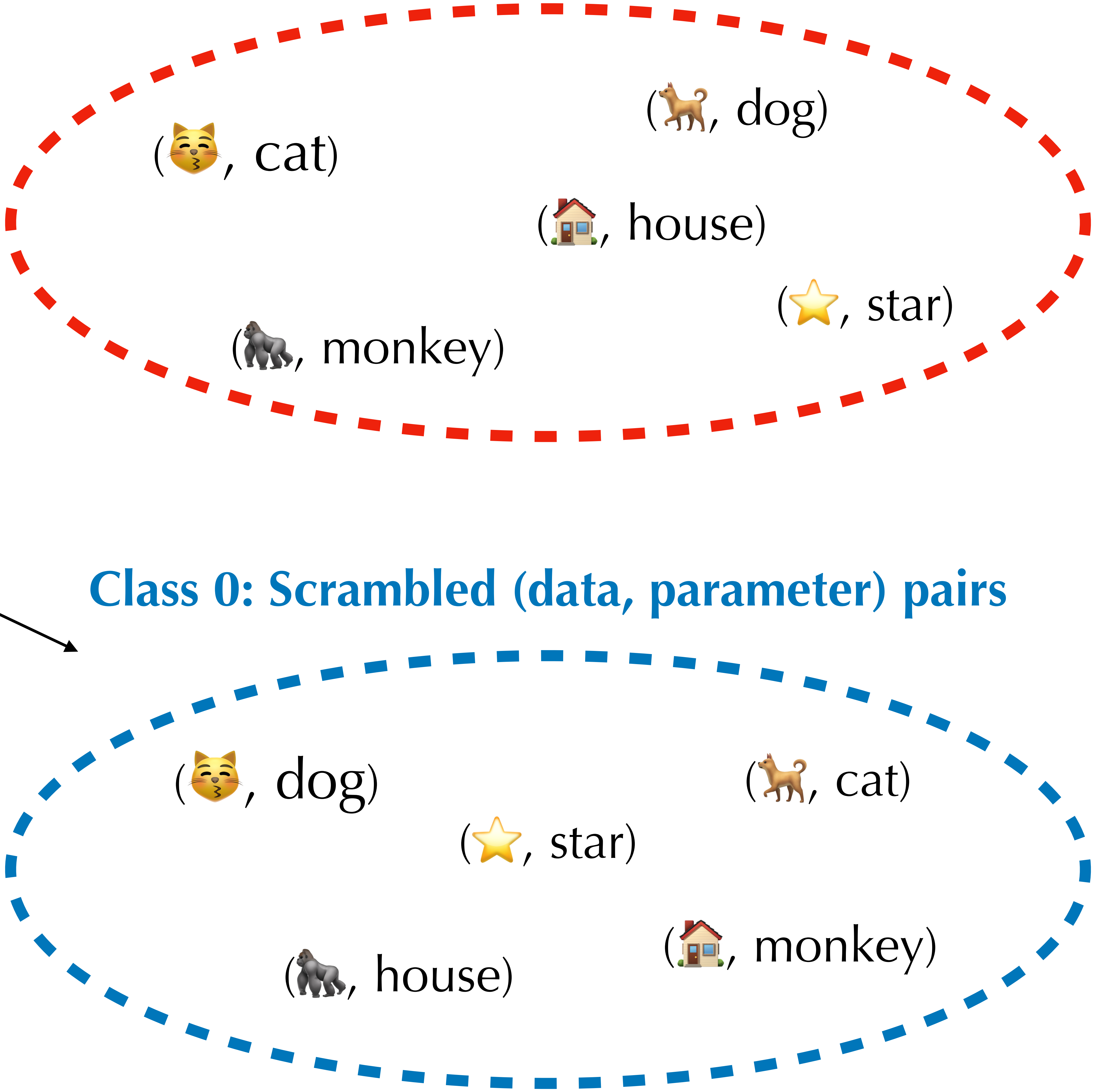
$$r(\mathbf{x};\mathbf{z}) \equiv \frac{p(\mathbf{x}\,|\,\mathbf{z})}{p(\mathbf{x})} = \frac{p(\mathbf{z}\,|\,\mathbf{x})}{p(\mathbf{z})} = \frac{p(\mathbf{x},\mathbf{z})}{p(\mathbf{x})p(\mathbf{z})}$$

$\mathbf{x},\mathbf{z} \sim p(\mathbf{x},\mathbf{z})$

Class 1: Matching (data, parameter) pairs

$\mathbf{x},\mathbf{z} \sim p(\mathbf{x})p(\mathbf{z})$

Class 0: Scrambled (data, parameter) pairs



Data: $\mathbf{x}$
Parameter: $\mathbf{z}$

Hermans+ 1903.04057

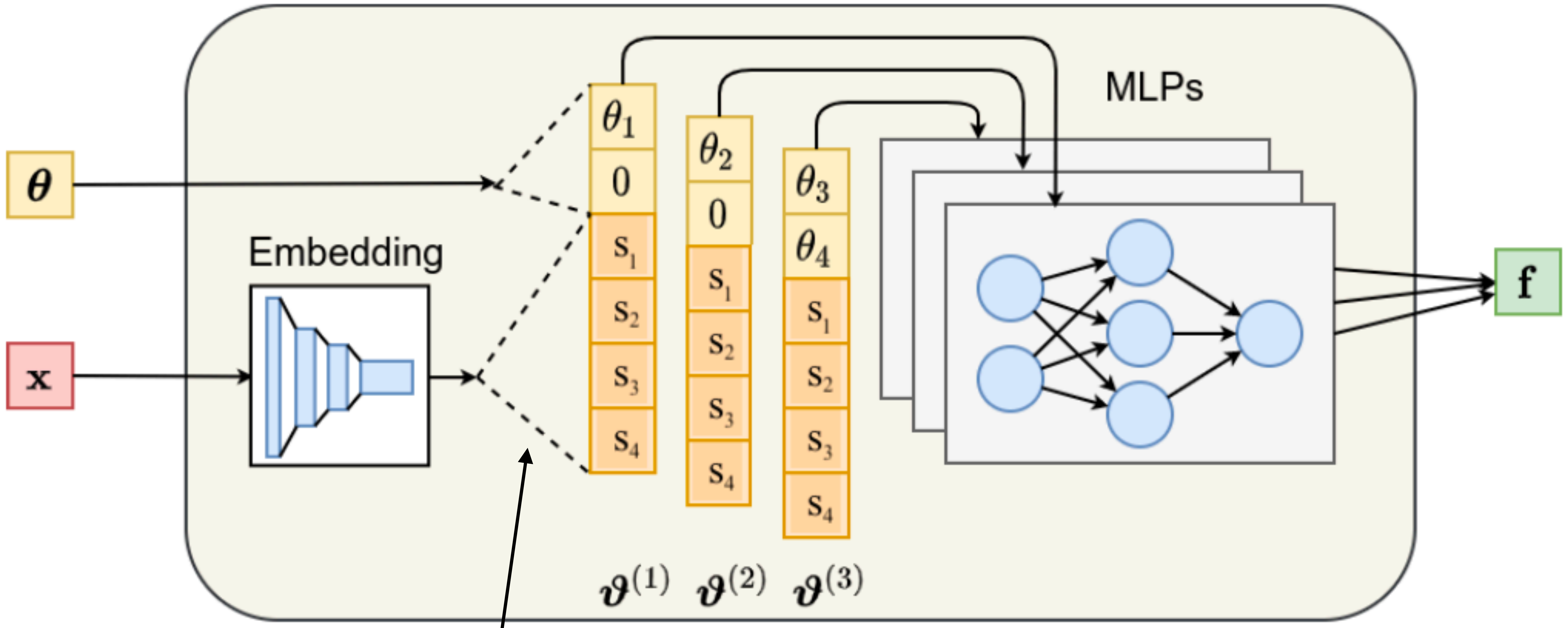
Neural ratio estimation

TMNRE

Architecture

Typical network architecture

- Embedding network for data $\mathbf{x}$ (e.g. a CNN), yielding data summaries $\mathbf{s} = \mathrm{S}_\phi(\mathbf{x})$.
- Correlated (usually MLP) combining data summaries $\mathbf{s}$ and parameters $\mathbf{z}$, $f = \ln r = \mathrm{M}_\phi(\mathbf{x}, \mathbf{s})$.



Automatically learned **data summaries**

$$f(\mathbf{x}, \mathbf{z}) = \ln r(\mathbf{x}; \mathbf{z}) = \ln \frac{p(\mathbf{x}, \mathbf{z})}{p(\mathbf{x})p(\mathbf{z})}$$

Data summary maximises distance between $p(\mathbf{z} \,|\, \mathbf{s}(\mathbf{x}))$ and $p(\mathbf{z})$ in terms of JS divergence.

Miller+ 2011.13951, 2107.01214 - swyft & TMNRE

Cole+ 2111.08030

Neural ratio estimation

TMNRE

Visualised

Neural ratio estimation (NRE)
Train a neural network to discriminate

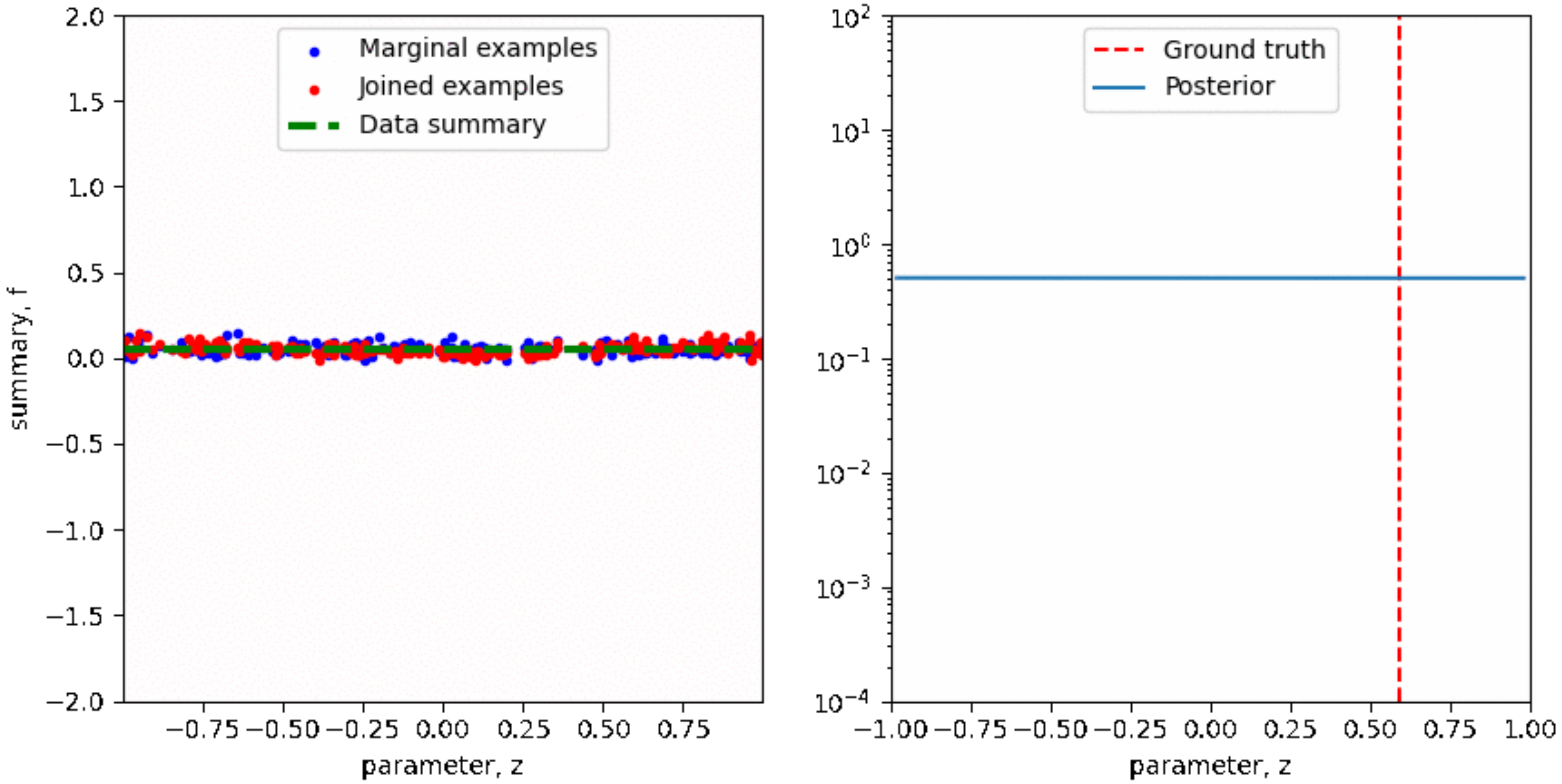
- Real sims: $z, \mathbf{x} \sim p(\mathbf{x}\,|\,z)p(z)$
- Scrambled sims: $z, \mathbf{x} \sim p(\mathbf{x})p(z)$

Embedding network $\mathbf{s} = \mathrm{S}_\phi(\mathbf{x})$ is trained such that $\mathbf{s}$ is a (hopefully) sufficient statistic,

$$p(z\,|\,\mathbf{s}) \simeq p(z\,|\,\mathbf{x})$$

MLP is trained to estimate ratio of interest,

$$\ln r(\mathbf{x}, z) \equiv \mathrm{M}_\phi(\mathbf{s}, \mathbf{z}) \simeq \ln \frac{p(\mathbf{x}, \mathbf{z})}{p(\mathbf{x})p(\mathbf{z})}$$



Toy model: $\mathbf{x} = \mathbf{v} \cdot z^2 + \epsilon$ $\mathbf{S}(\mathbf{x}) = \mathrm{Linear}(D,1)(\mathbf{x})$ $\mathrm{M}_\phi(s, z)$: ResNet

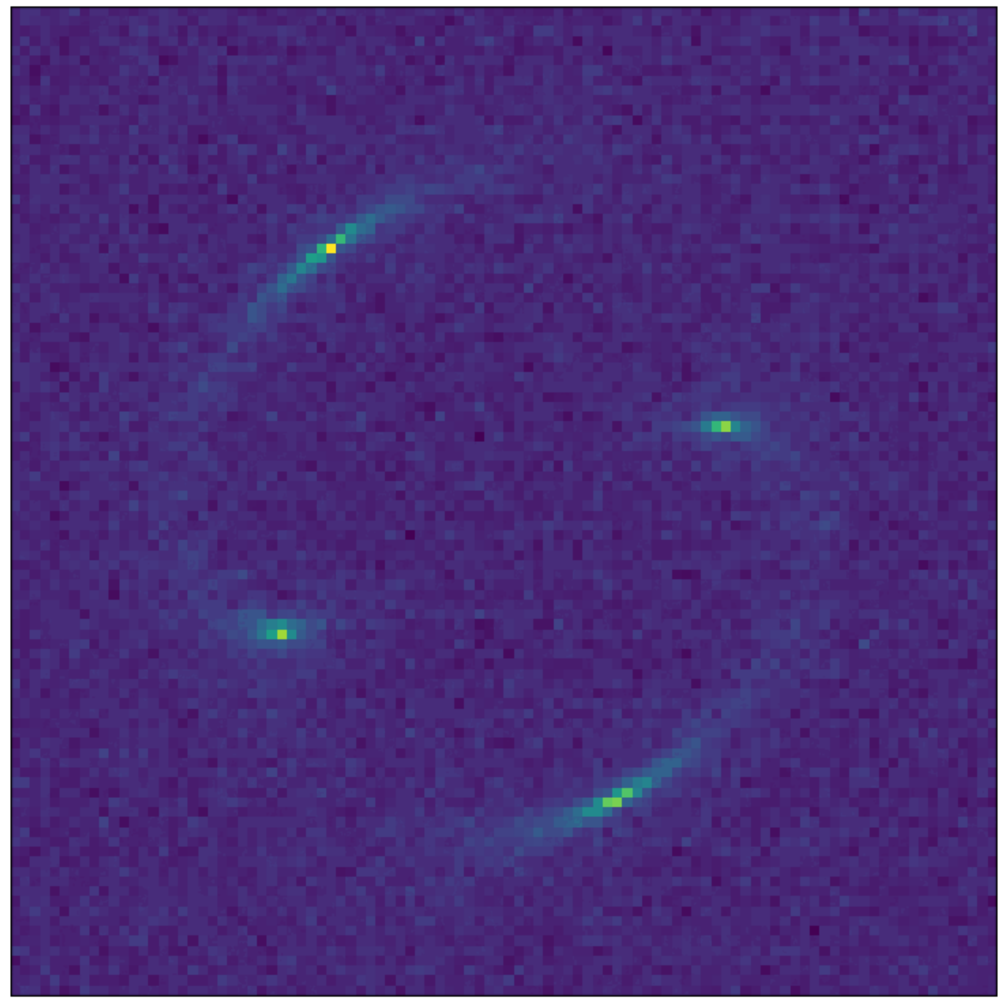
Miller+ 2011.13951, 2107.01214 - swyft & TMNRE

Sequential inference

TMNRE

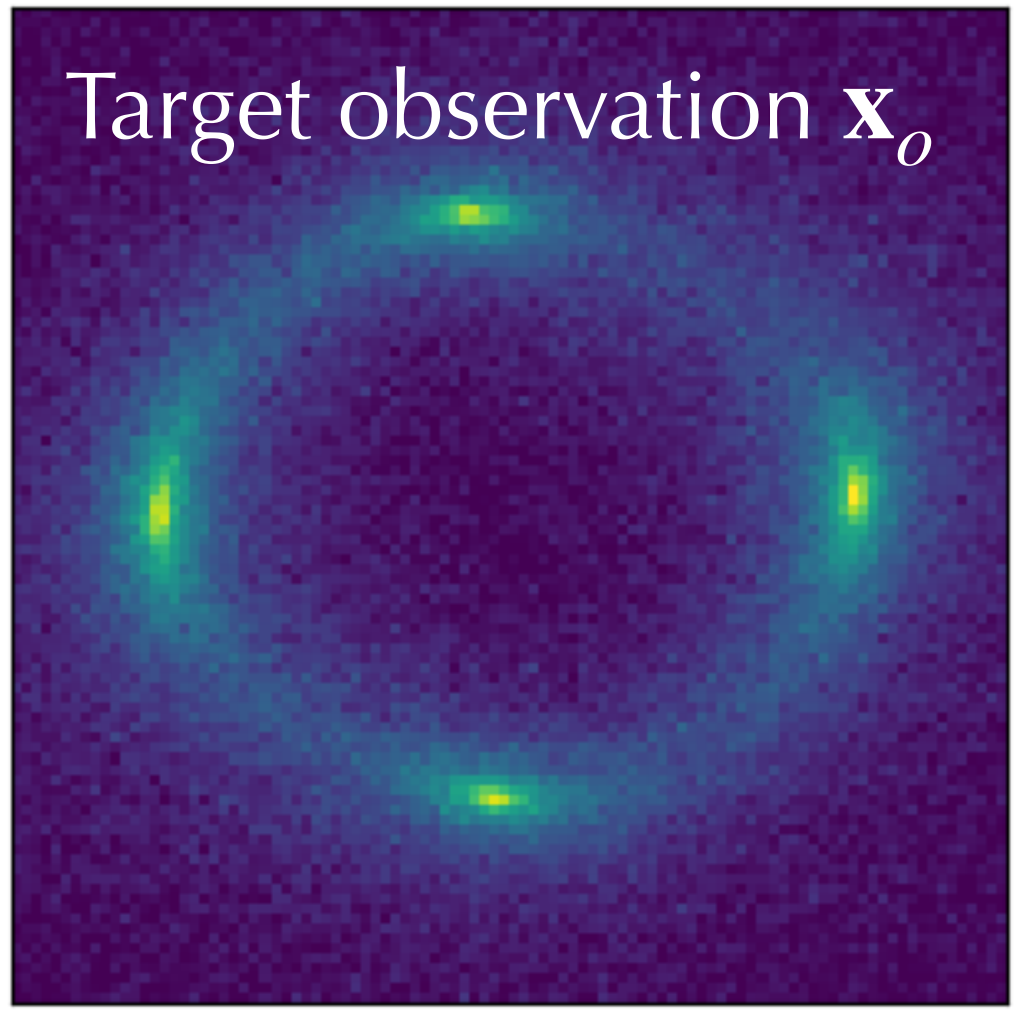
Gaining precision through targeted simulations

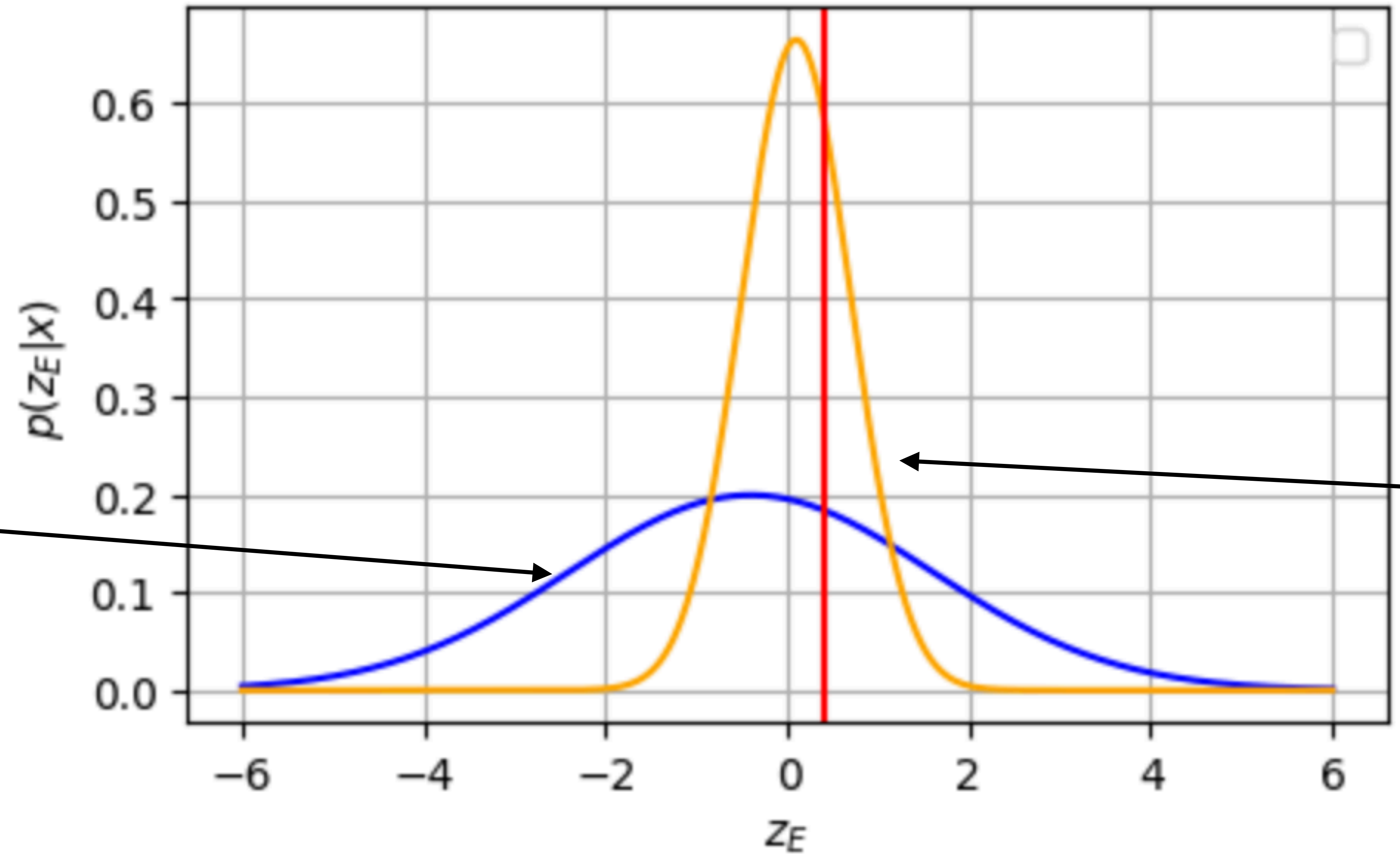
Samples from full prior:
$\mathbf{x}, \mathbf{z} \sim p(\mathbf{x}\,|\,\mathbf{z})p(\mathbf{z})$



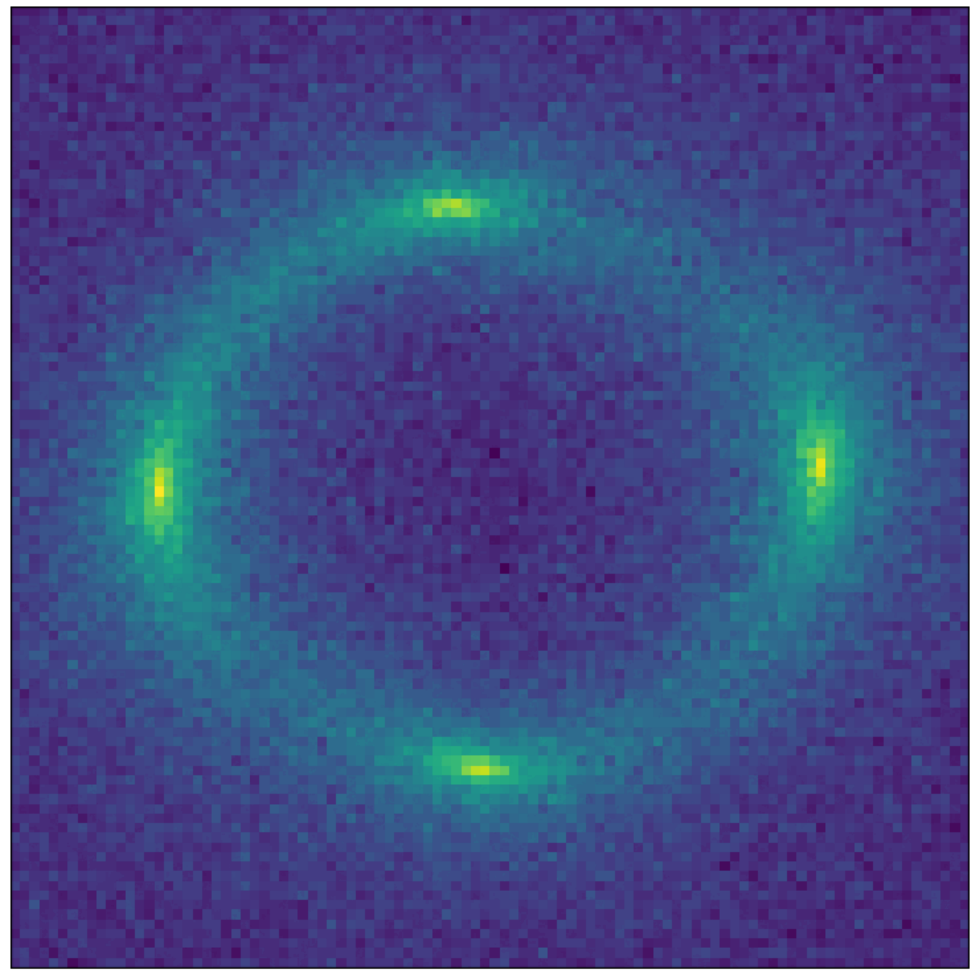
NRE training

$r(\mathbf{x}; z_E)$





Samples from some constrained prior:
$\mathbf{x}, \mathbf{z} \sim p(\mathbf{x}\,|\,\mathbf{z})\tilde{p}_{\mathbf{x}_o}(\mathbf{z})$



NRE training

$r(\mathbf{x}; z_E)$

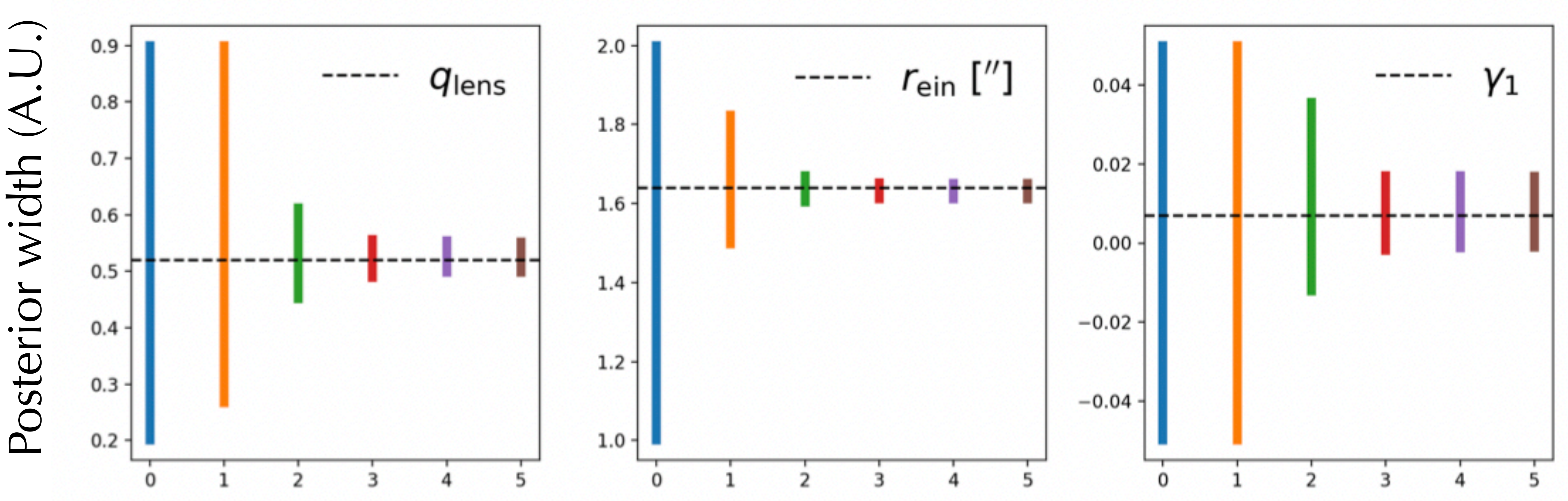
Durkan+ 2002.03712 for a discussion

Image credit: Noemi Anau Montel

Sequential inference

TMNRE

Learning precise posteriors in multiple rounds



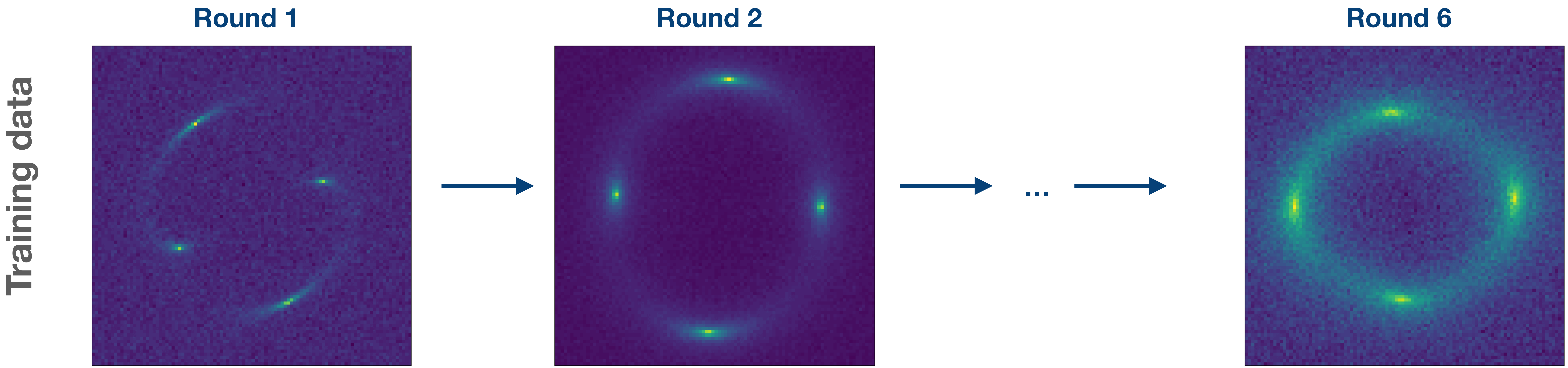


Image credit: Noemi Anau Montel

Marginal sequential inference

TMNRE

Key idea: Use a truncated version of the prior as proposal function.

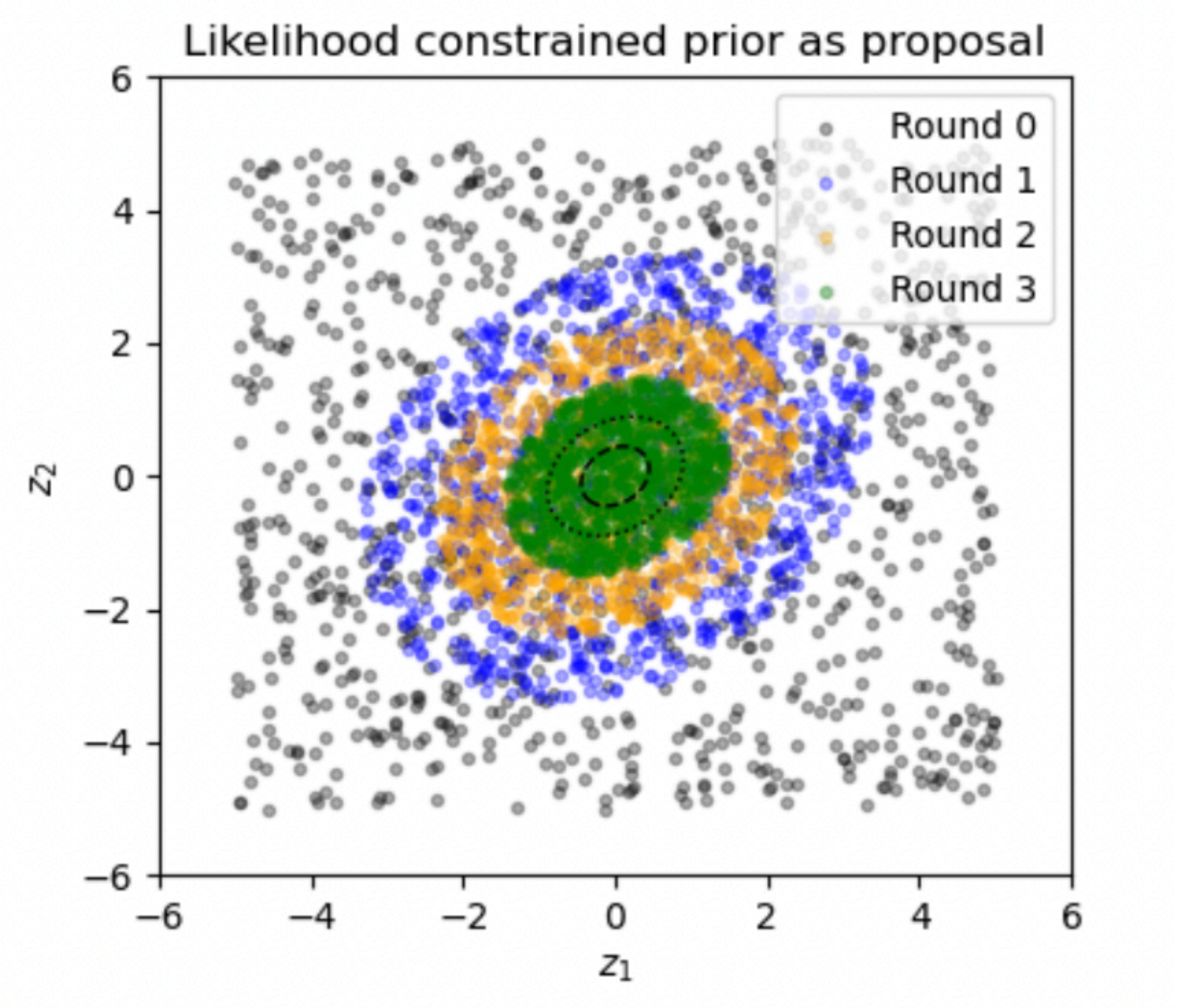
$$\tilde{p}^{(R)}(\mathbf{z}) = \frac{1}{Z}\mathbb{I}(\mathbf{z} \in \Gamma^{(R-1)})p(\mathbf{z})$$

Miller+ 2011.13951, 2107.01214 - swyft & TMNRE

We use a hard likelihood constrained prior truncation scheme, excluding low likelihood regions estimated in previous rounds.

$$\Gamma^{(R)} = \{\mathbf{z} \in \mathbb{R}^N : \tilde{r}^{(R)}(\mathbf{x};\mathbf{z}) > \epsilon\} \qquad \tilde{r}^{(R)}(\mathbf{x};\mathbf{z}) \simeq \frac{p(\mathbf{x}\,|\,\mathbf{z})}{p(\mathbf{x})}$$

Doing this leaves the learned ratio unaffected, and marginal estimation becomes possible



$$q_\phi^{(R)}\left(z_1 \mid \mathbf{x}\right) \simeq \int dz_2 \ldots dz_N\; p(\mathbf{x} \mid \mathbf{z})\,\frac{1}{Z}\mathbb{I}(\mathbf{z} \in \Gamma^{(R-1)})p(z_2, \ldots, z_N) = \int dz_2 \ldots dz_N\; p(\mathbf{x} \mid \mathbf{z})p(z_2, \ldots, z_N) + \mathcal{O}(\epsilon)$$

See also Greenberg+ 1905.07488, Durkan+ 2002.03712 for sequential methods
Applied to NPE in Deistler+ 2210.04815

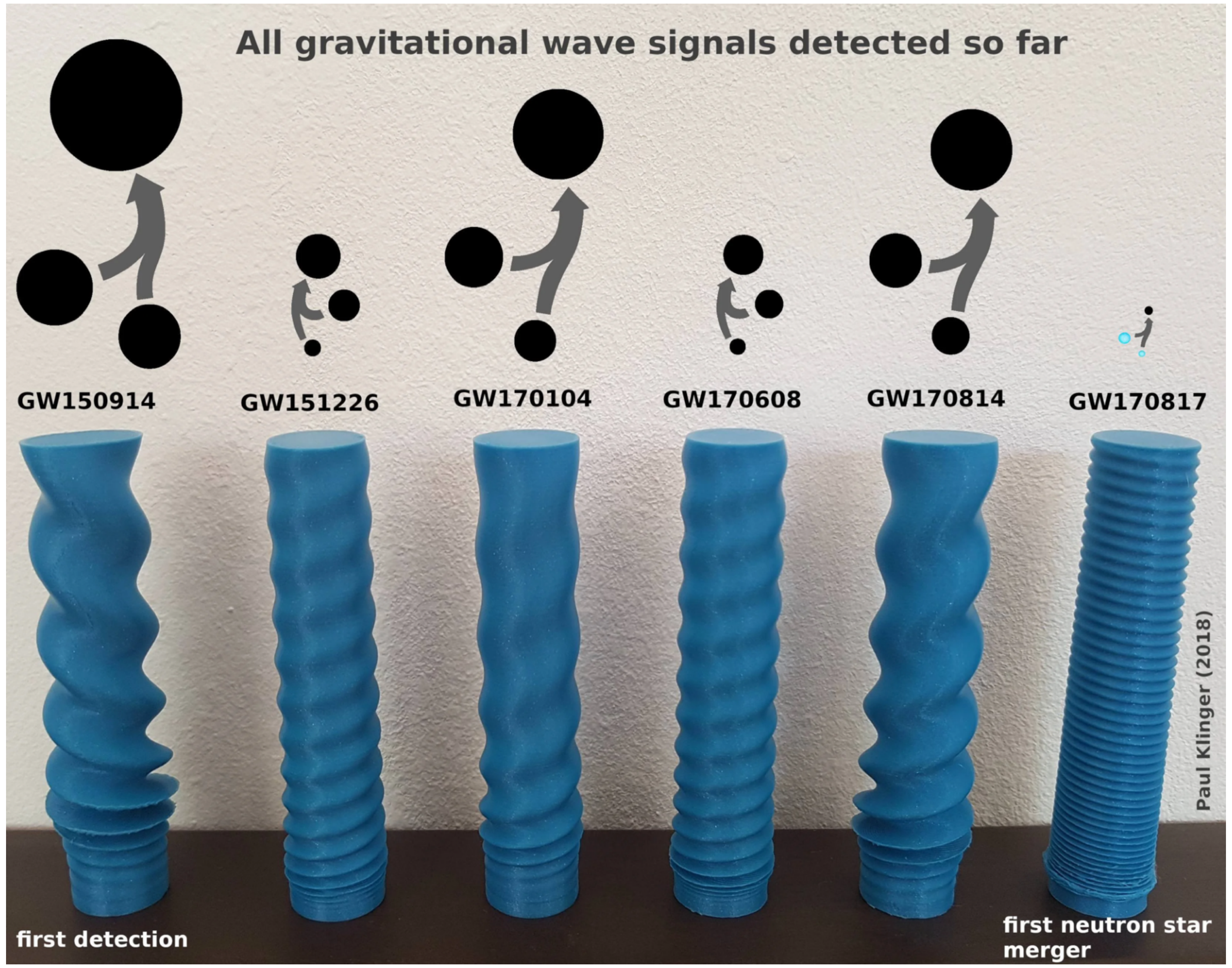
Applications

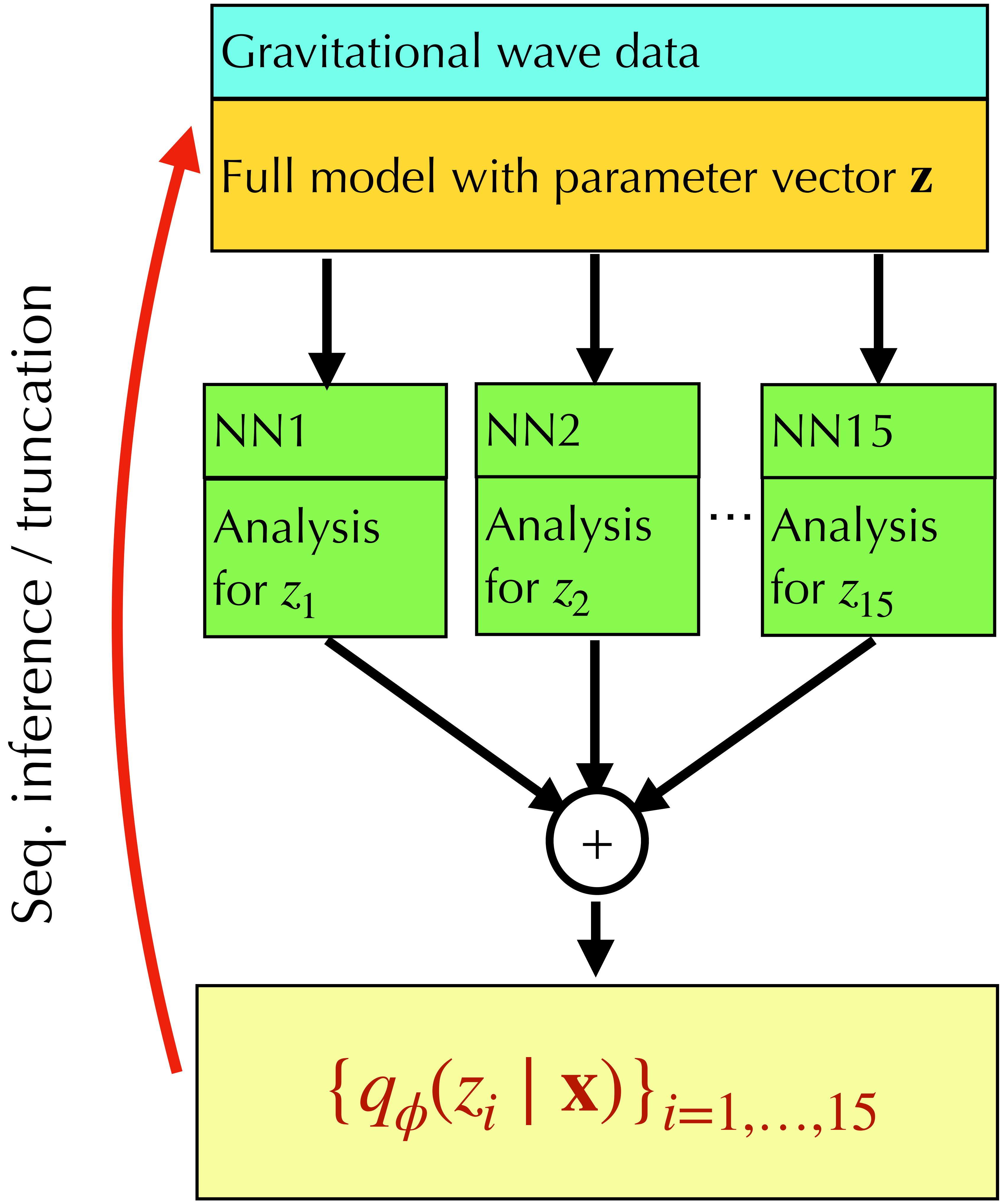
Overview

- 1) Gravitational waves (time series)
- 2) Strong lensing (image analysis)
- 3) Source population analysis (object detection, hierarchical models)
- 4) Image analysis (denoising)

1) Gravitational wave parameter inference

Marginal inference of 15 waveform parameters



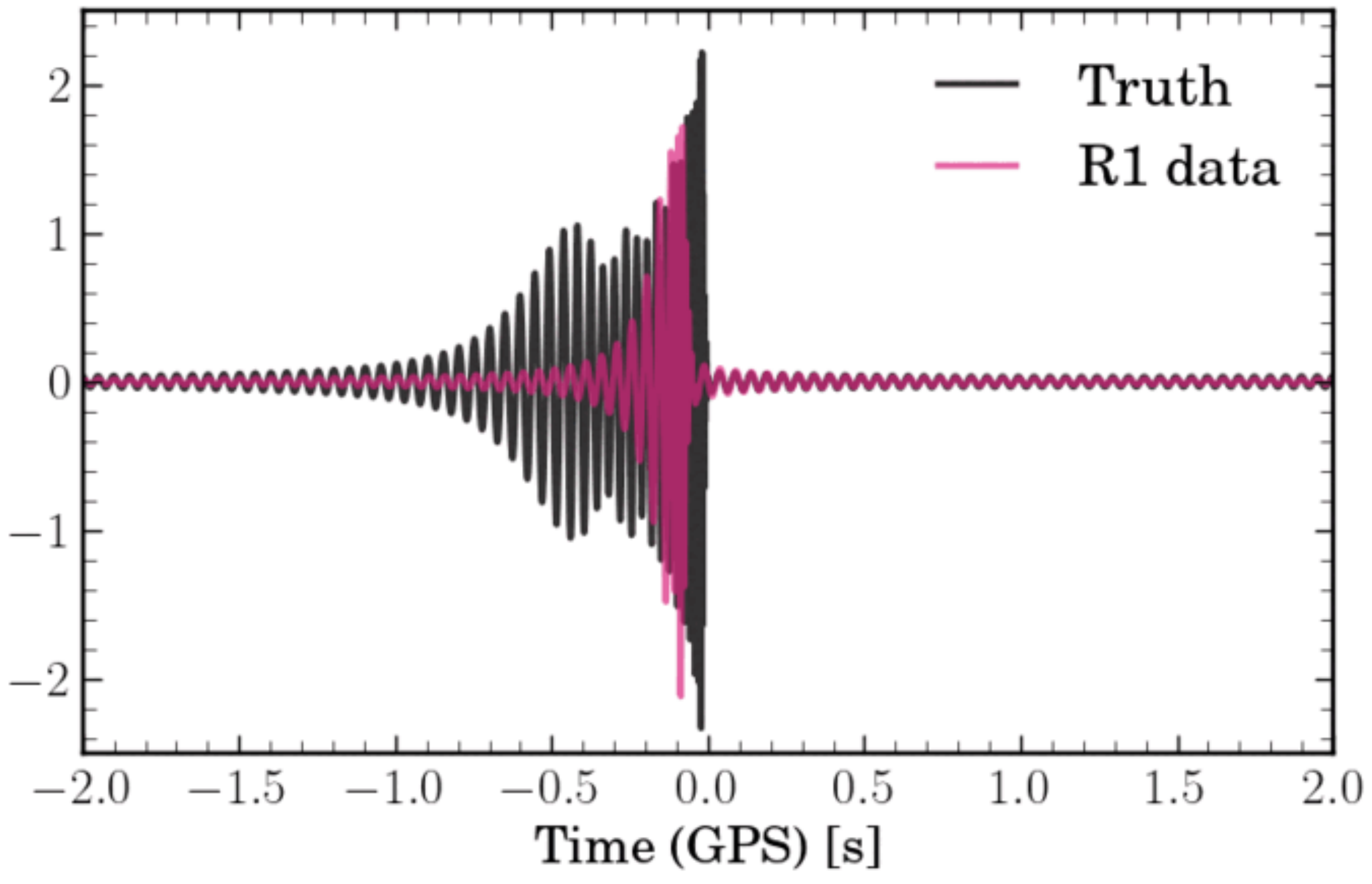


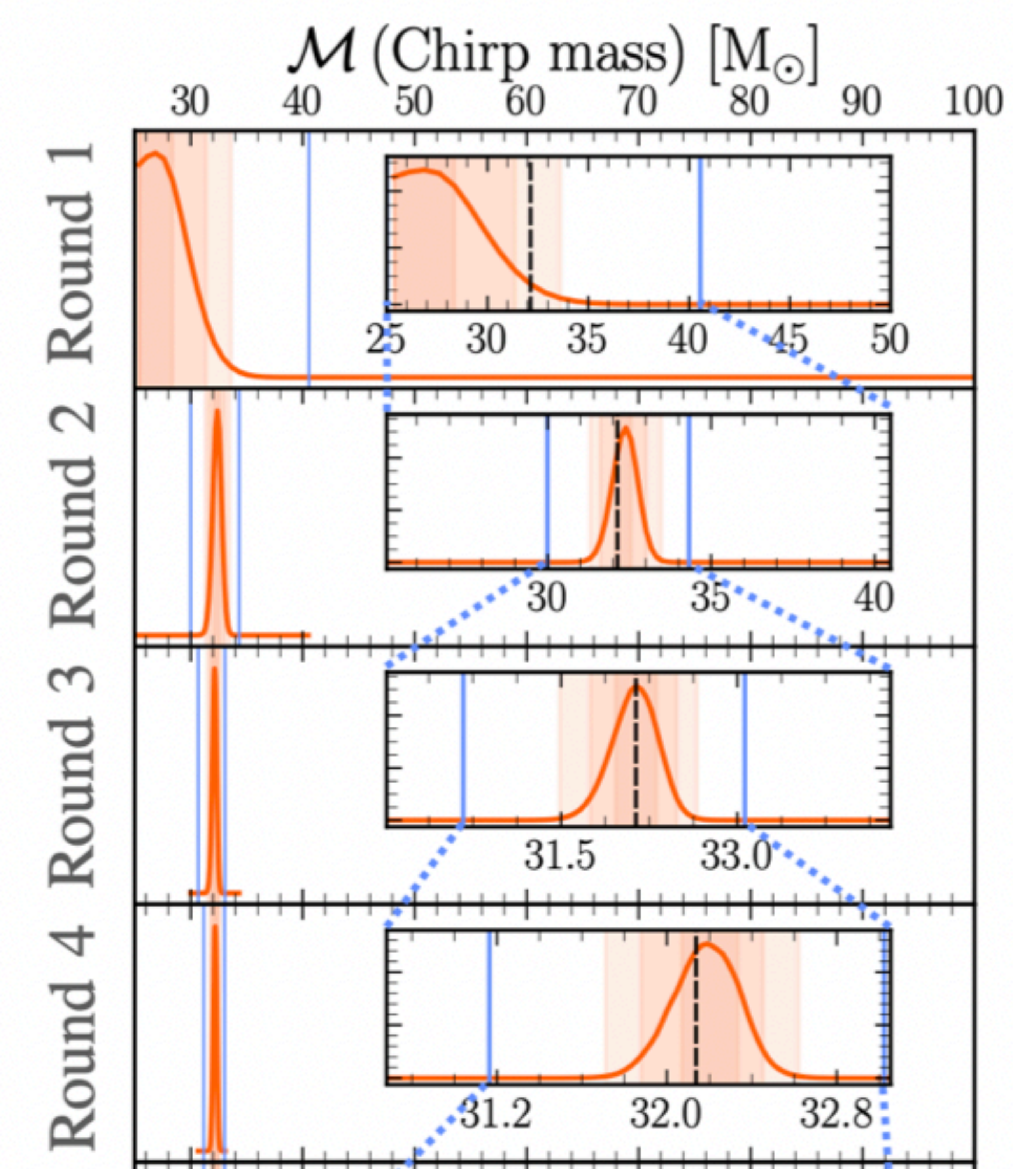
Bhardwaj+ 2304.02035

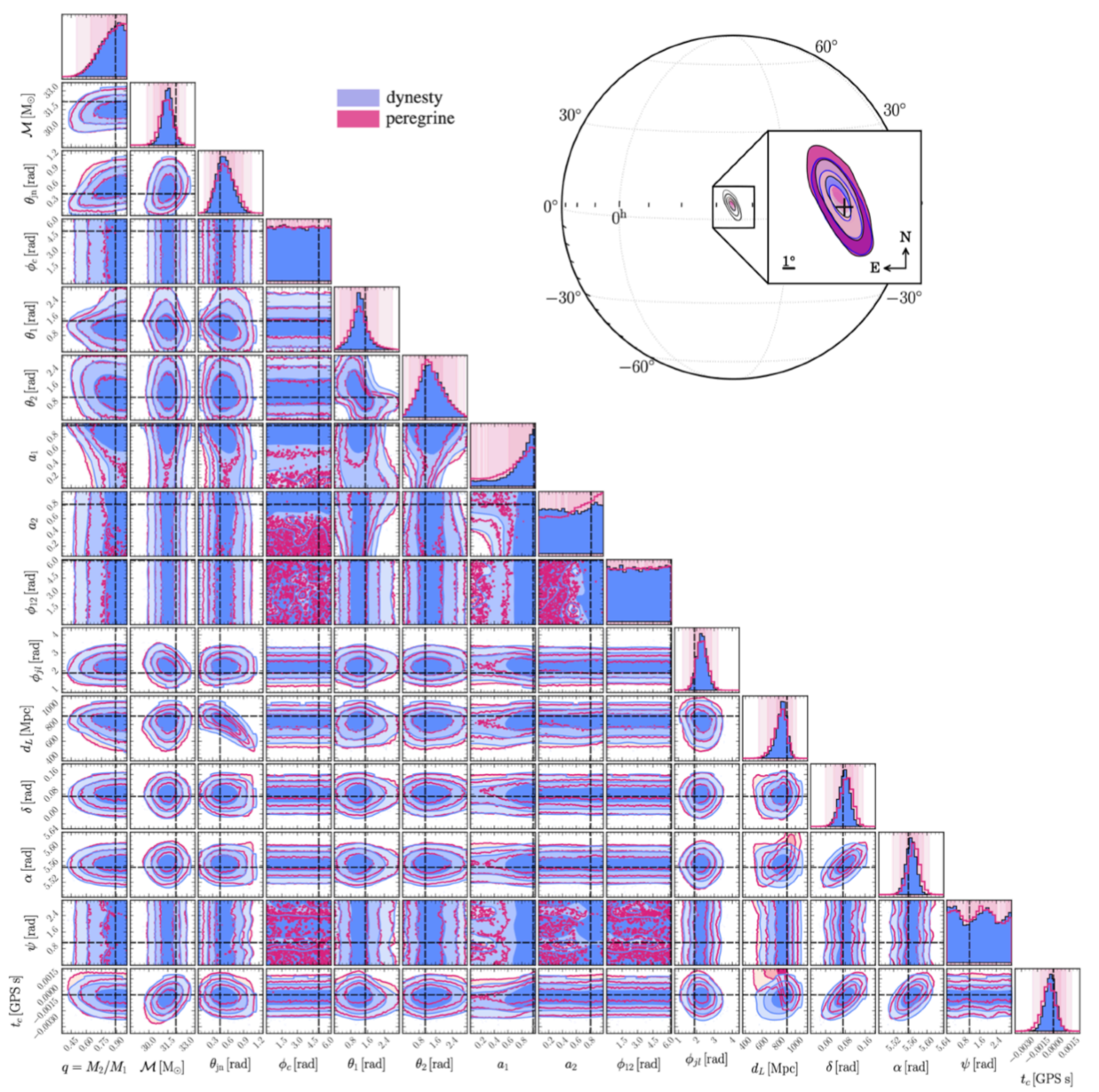
Related work: Dax+ 2106.12594

1) Gravitational wave parameter inference

Parameter correlations can be recovered at the end







The initial parameter scan is done using 15 1-dim marginals

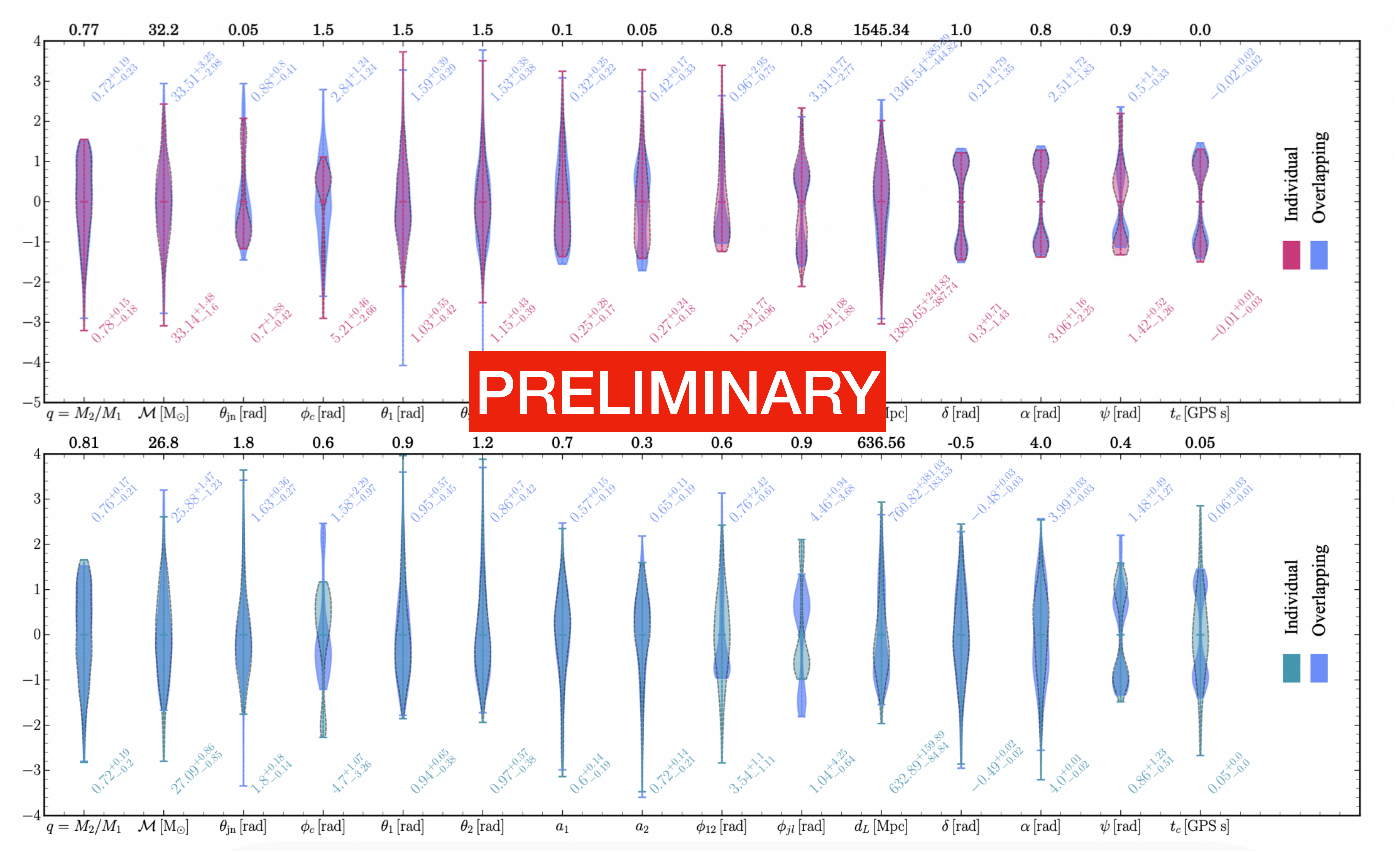
We can recover parameter correlations by estimating $(N-1)N/2$ 2-dim posteriors in the final round.

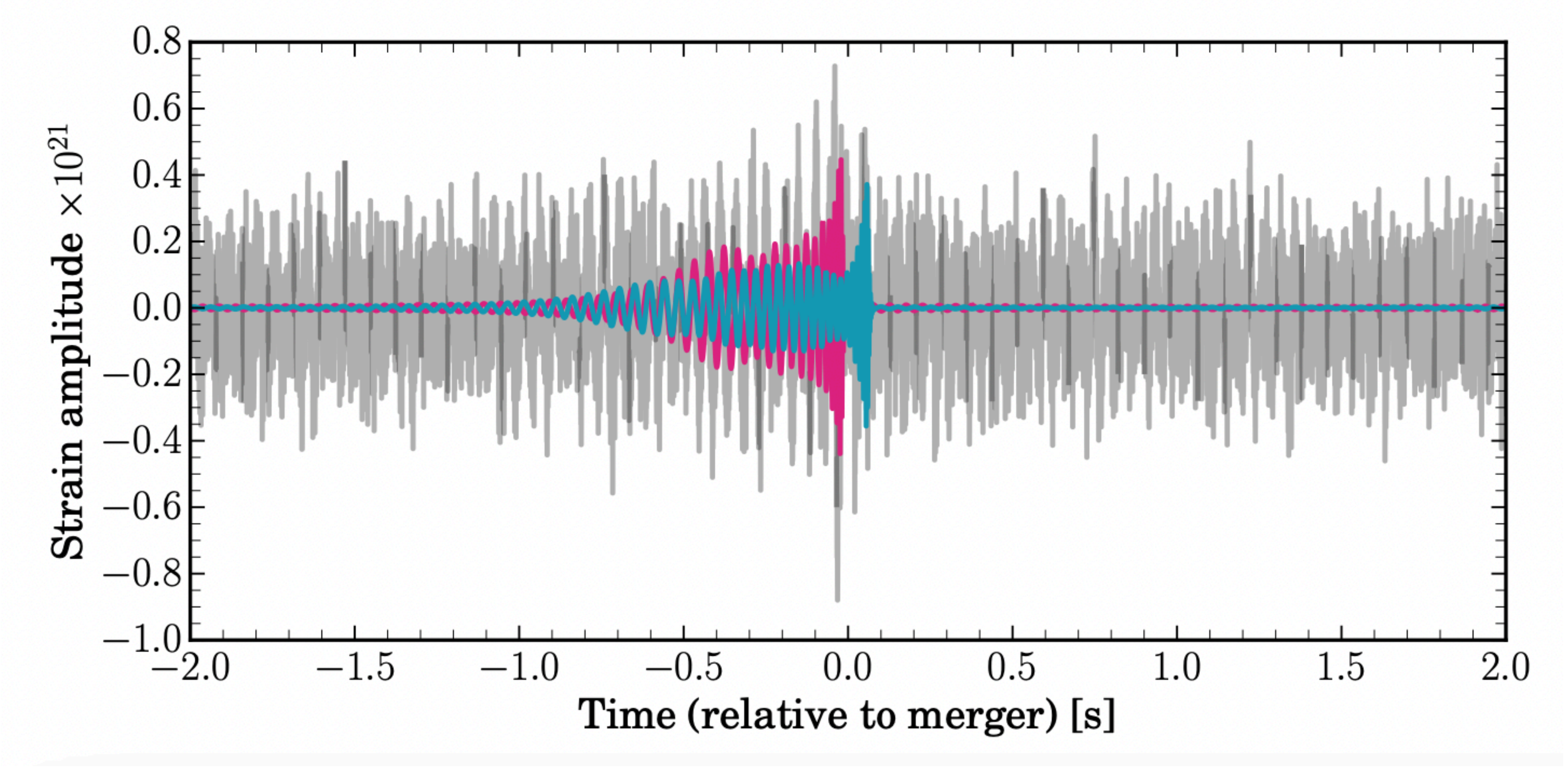
We need 10-100x less simulations than MCMC methods or fully amortised methods.

Bhardwaj+ 2304.02035

1) Gravitational wave parameter inference

Overlapping GW signals!





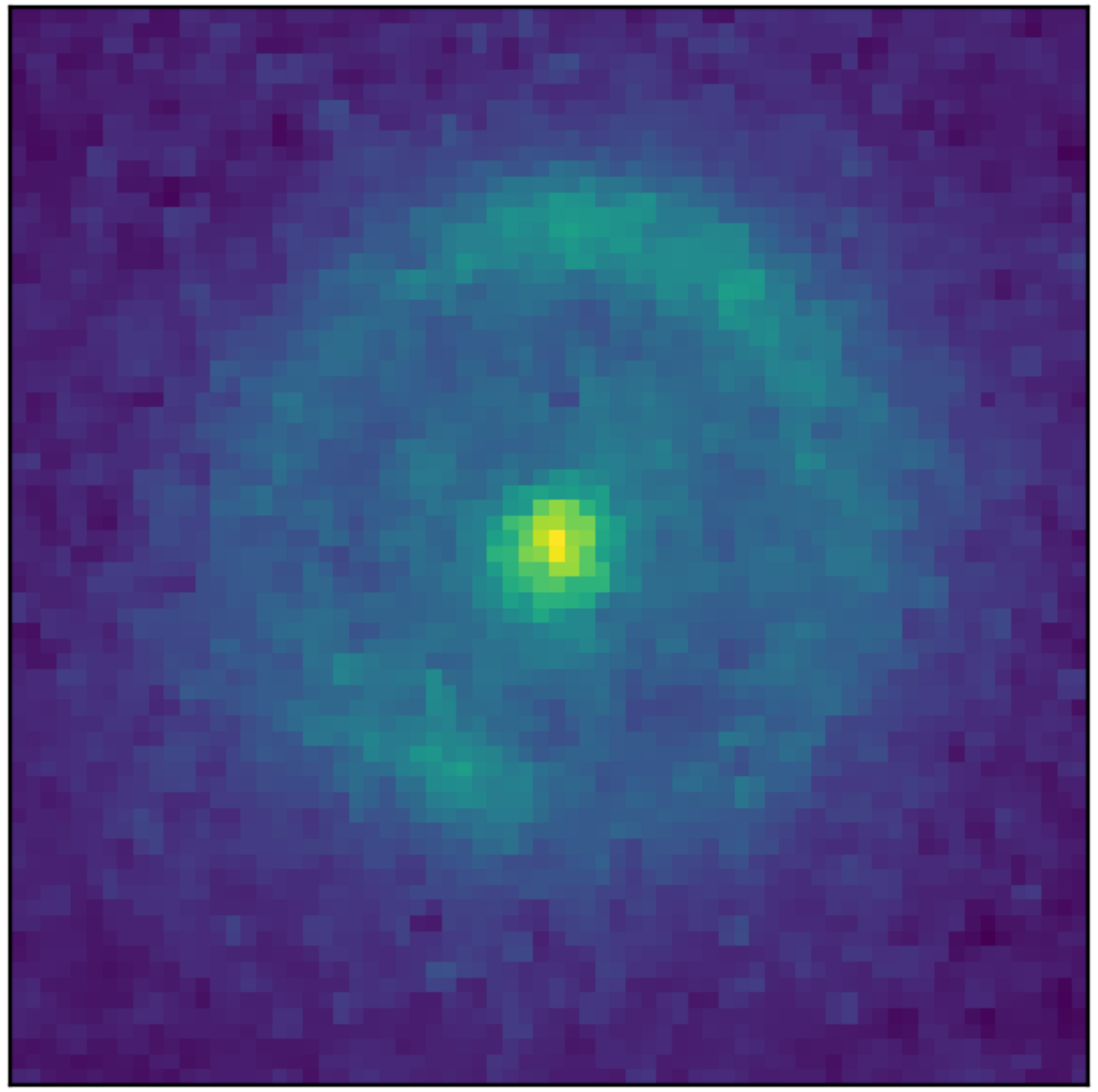
- IMRPhenomXPHM
- 36 hours (instead of >20 days)
- Faster than MCMC
- Much more precise than previous SBI attempts
- Precision only mildly degraded w.r.t. fits in absence of a second signal

Alvey+ 23xx.yyyyy

2) Strong lensing image analysis

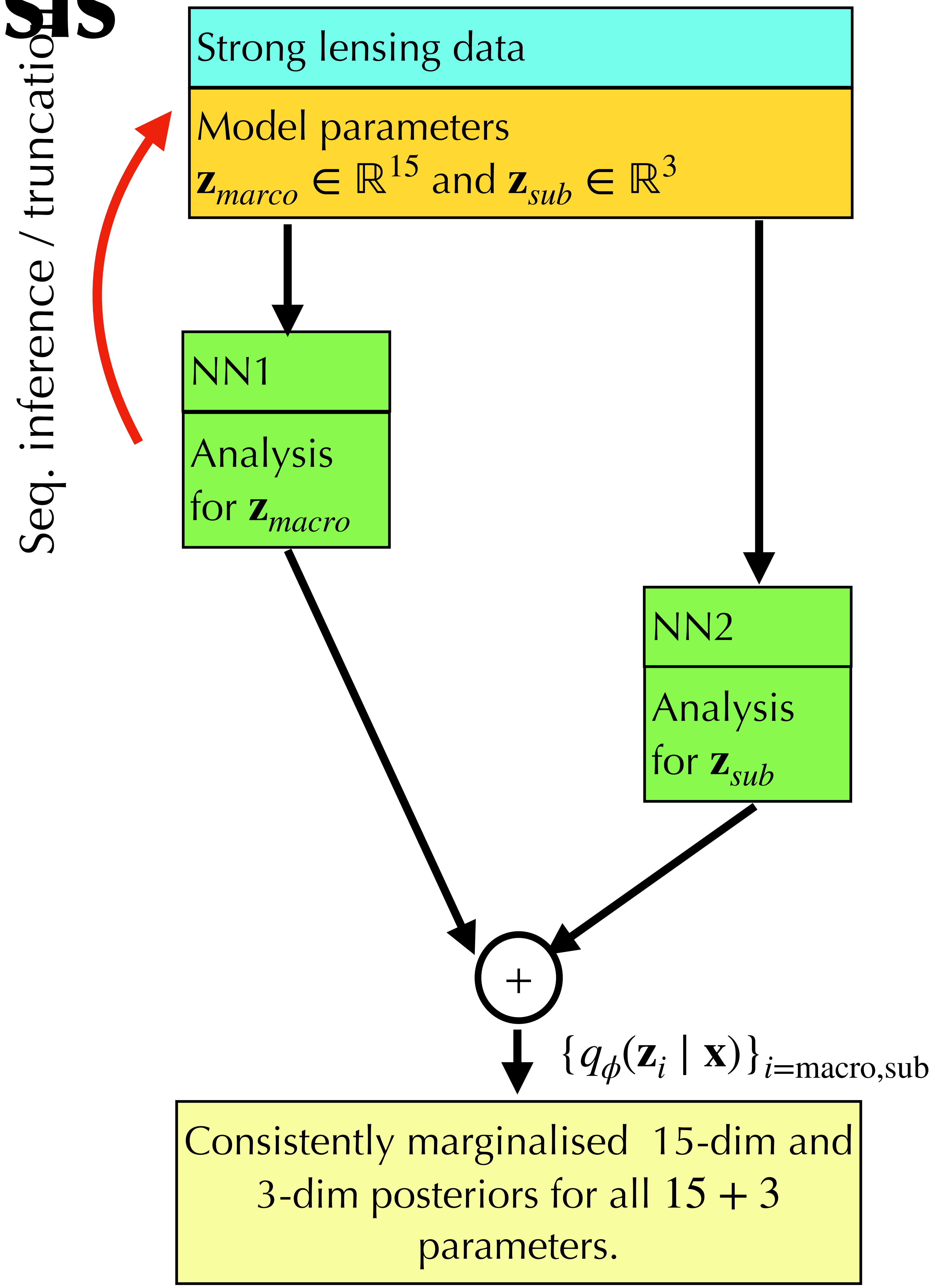
General method

Target image: JVAS B1938+666



Vegetti et al. (2012) - subhalo detection claim

Şengül et al. (2021) - detection reanalysed



2) Strong lensing image analysis

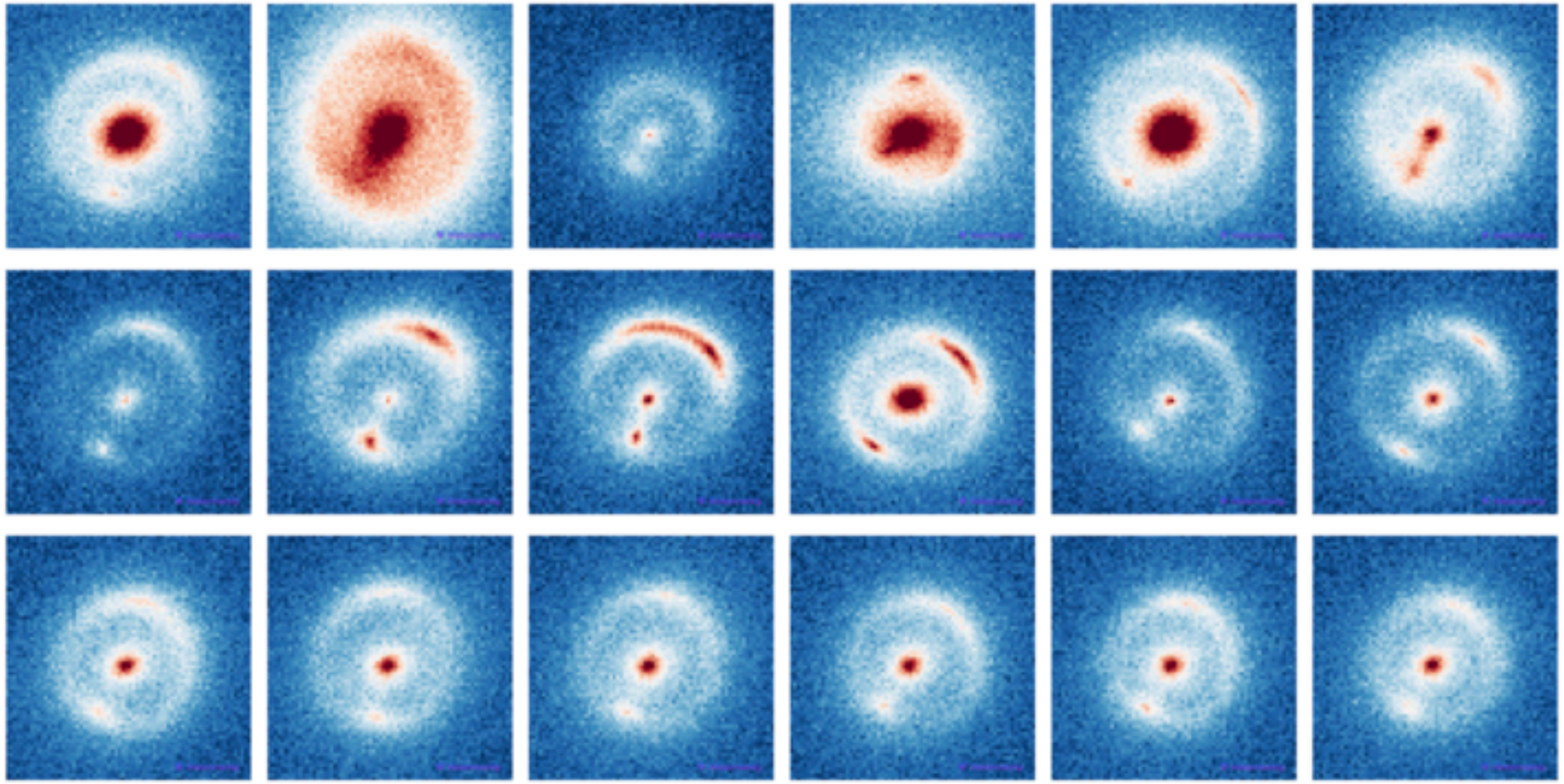
Truncation including correlations

PRELIMINARY

Untruncated

Box truncation
(Based on 14 1-dim posteriors)

Correlated truncation
(14 dimensional joined posterior, explored with **slice sampling**)



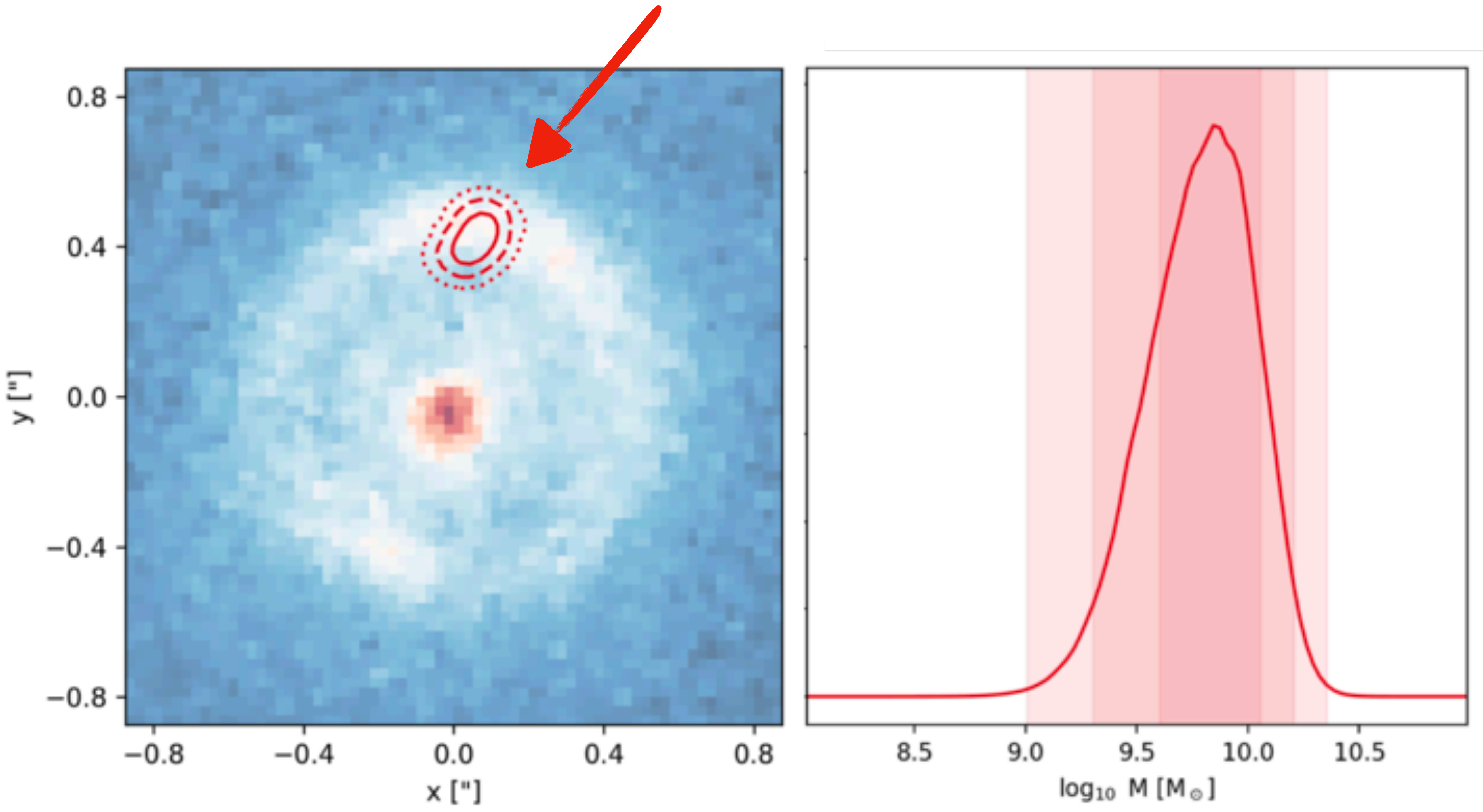
$\mathbf{z}_{sub} = (x, y, M)$

Accounting for parameter correlations in main lens massively reduces training data variance.
Likelihoods are automatically marginalised over other parameters correctly.

2) Strong lensing image analysis

Sensitivity to very faint signals

Training an inference network (here U-Net) on highly targeted training data enables
faint signal detection with recognition networks*!



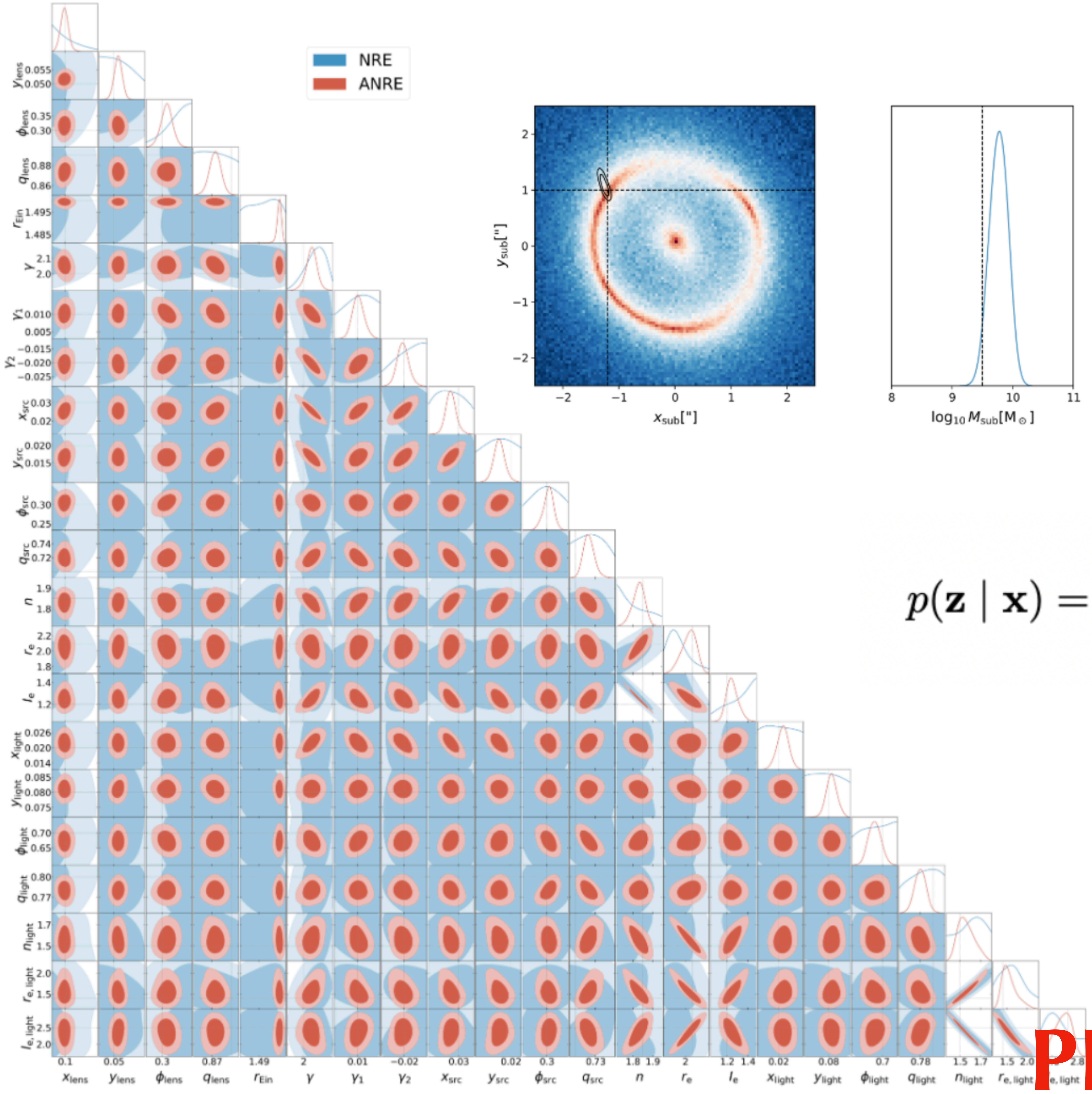
*the gravitational lensing effect of a small dark matter subhalo that distorts the image at the few percent level

PRELIMINARY

Anau Montel+ 23🤞🤞.yyyyy

2) Strong lensing image analysis

Corner plot



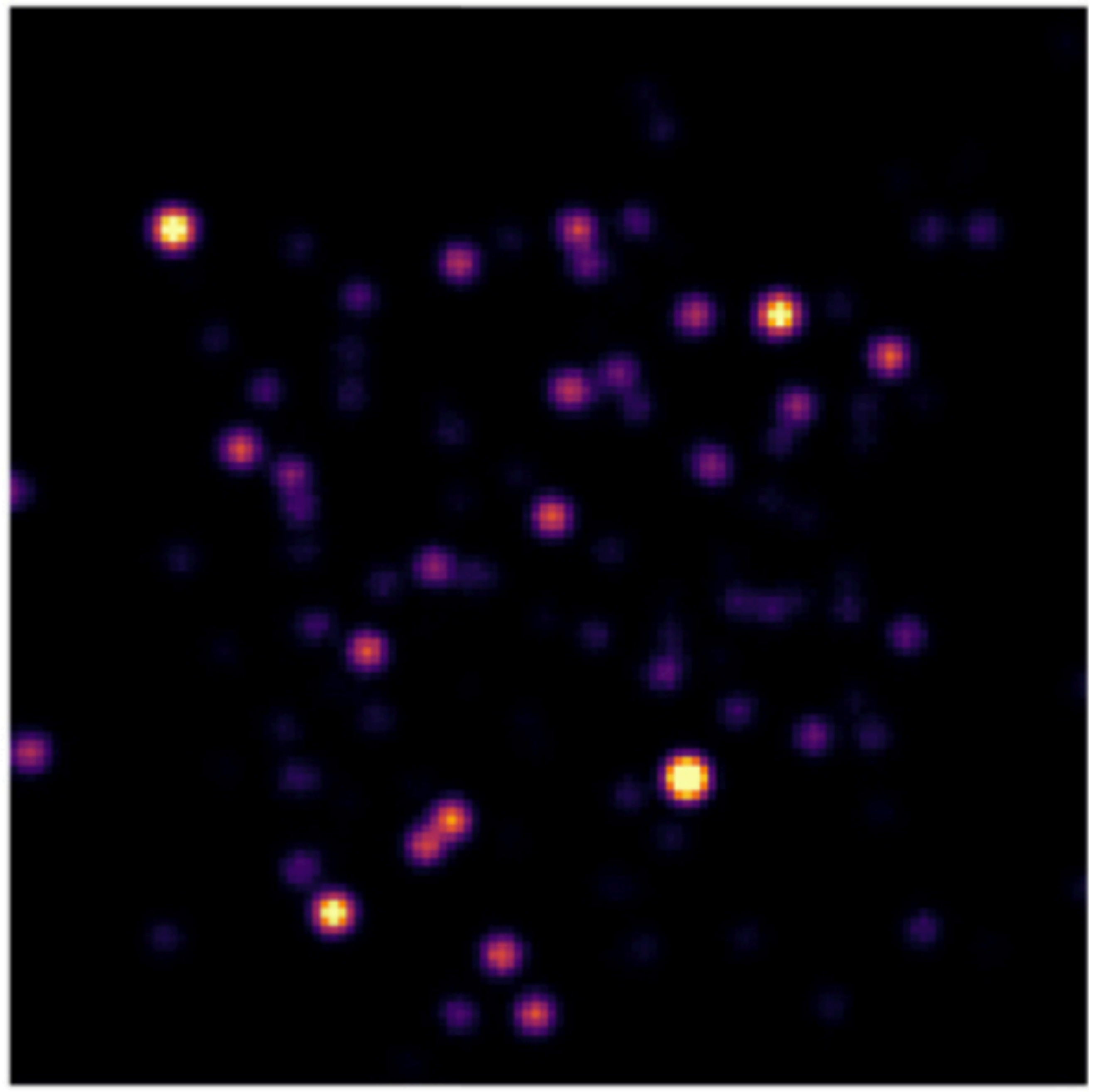
$$p(\mathbf{z} \mid \mathbf{x}) = p(z_1 \mid \mathbf{x}) \prod_{i=2}^{N} p(z_i \mid \mathbf{x}, z_{1:i-1})$$

PRELIMINARY

3) Source population analysis

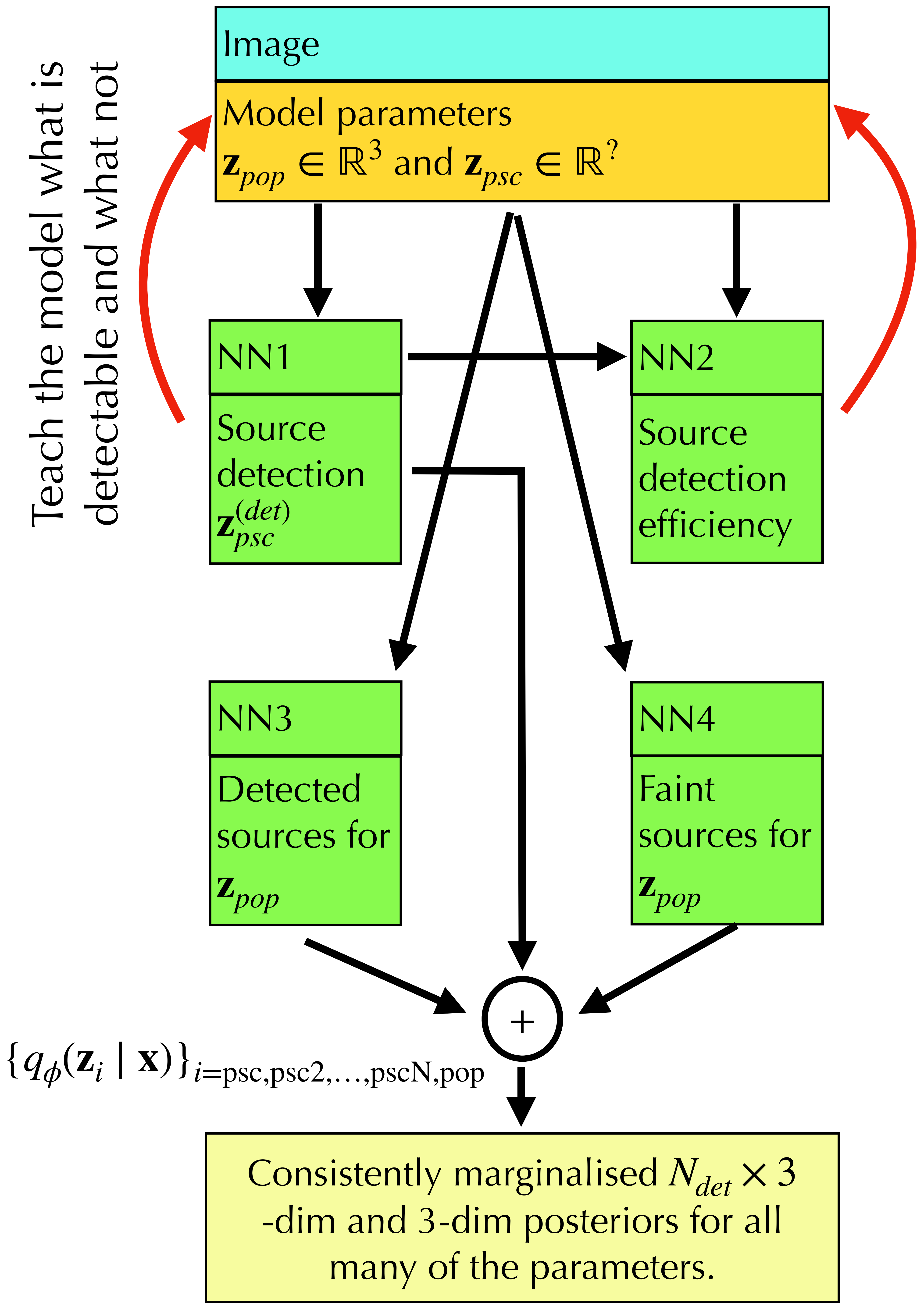
General method

Target data is an image of point sources



Simulation code

- Sample popualtion parameters
- Sample number of point sources (Poisson distribution)
- Sample properties of point sources (position, luminosity)
- Put all sources on the sky

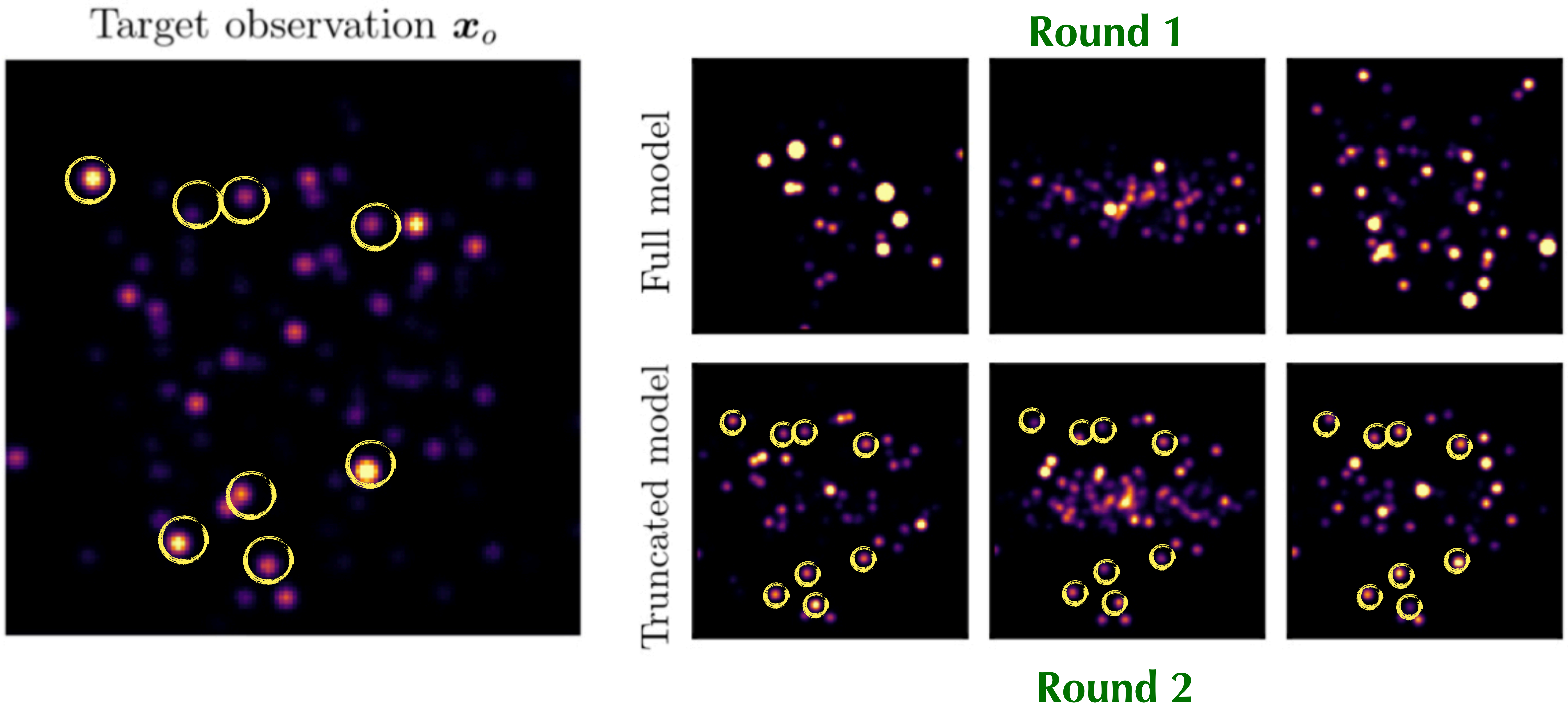


$$p(\mathbf{x} \mid z) = \sum_{N=1}^{\infty} p(\mathbf{x} \mid z_{psc}^{(1)}, \ldots, z_{psc}^{(N)}) \prod_{i=1}^{N} p(z_{psc}^{(i)} \mid z_{pop}) p(N \mid z_{pop}) p(z_{pop})$$

Anau Montel & CW 2211.04291

3) Source population analysis

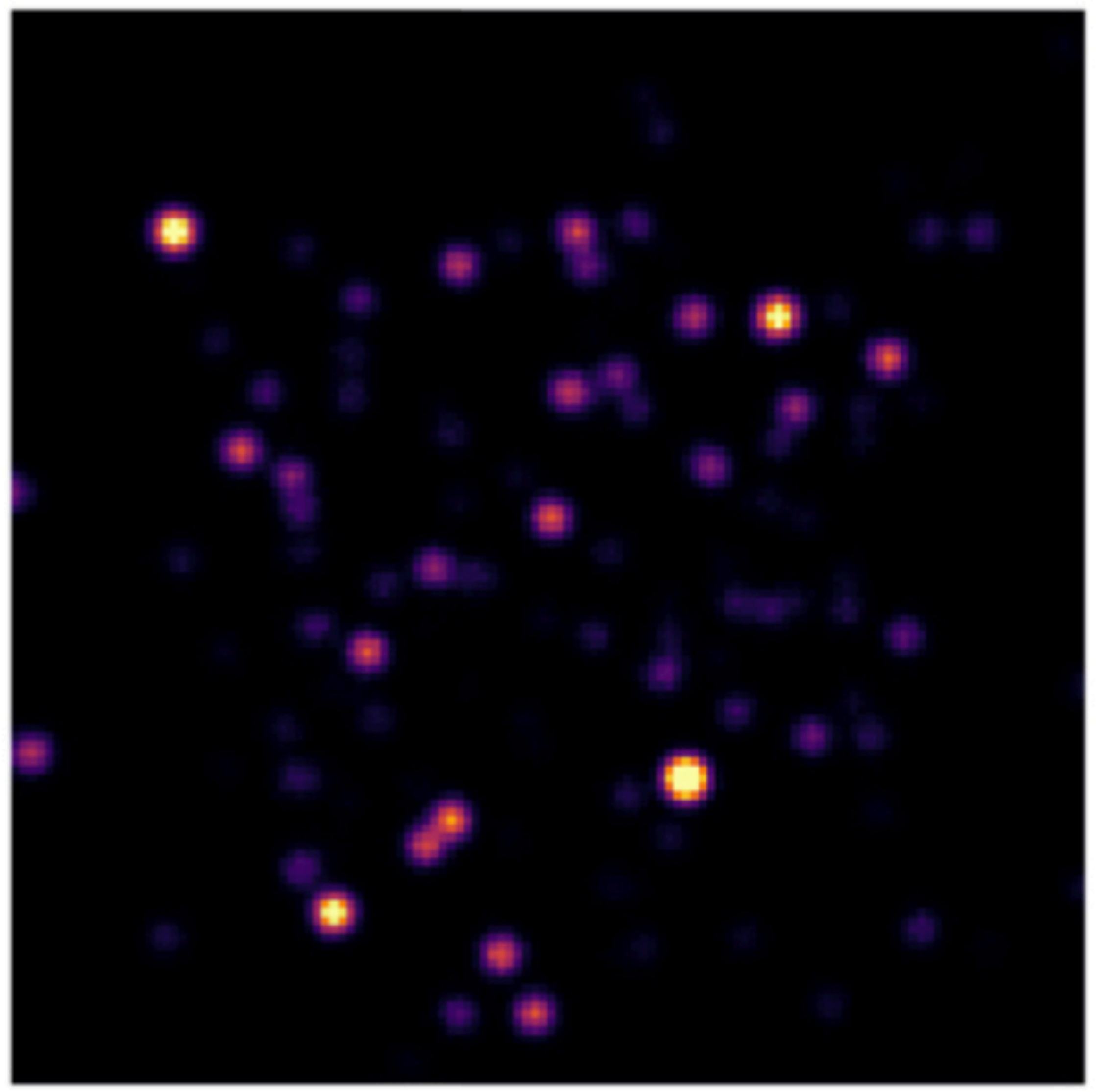
Truncation in practice



Anau Montel & CW 2211.04291

3) Source population analysis

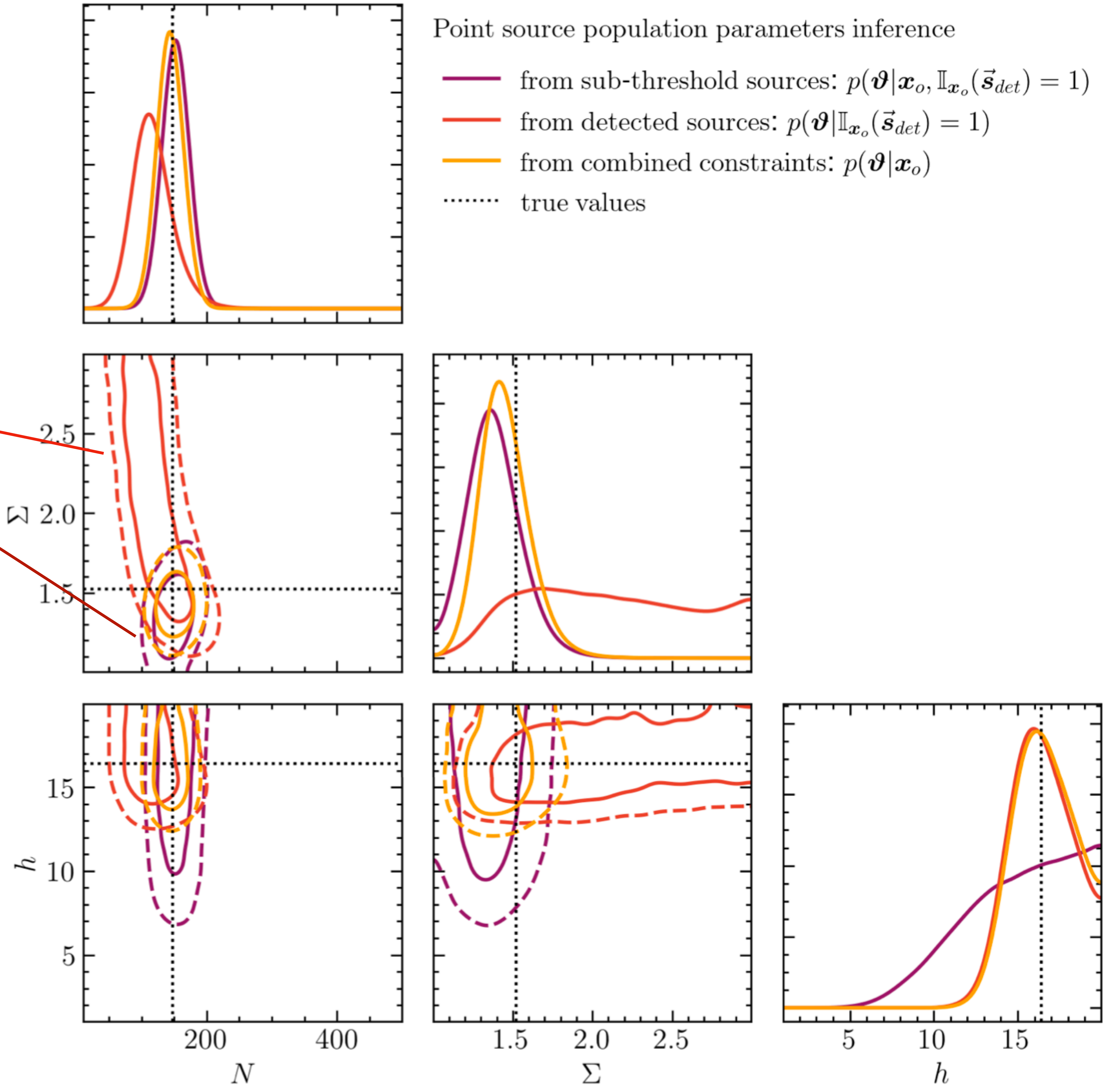
Consistent population parameter infer



Constraints on population parameters…

…based on **detected sources**

…based on **sub-threshold sources**

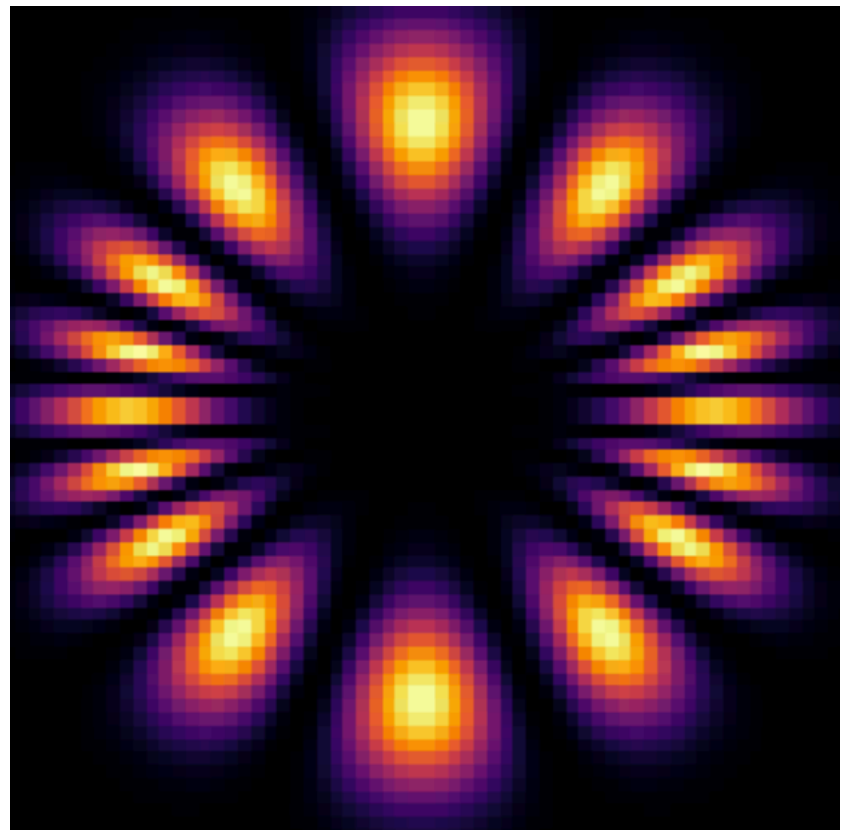


- **First method (AFAIK) to perform self-consistent determination of population parameters based on detected *and* undetected objects.**
- The trick is to make observational biases potentially related to point source detection part of the model itself.

Anau Montel & CW 2211.04291

4) Image analysis

Towards image analysis with SBI: Sequential inference is also possible for high-dimensional image analysis problems



- Toy model: Exponentiated Gaussian random field

$$x_i = e^{z_i} + \epsilon \ , \quad \mathbf{z} \sim \mathcal{GP}$$

- To this end, we train the joined likelihood

$$\frac{p(\mathbf{x}|\mathbf{z})}{p(\mathbf{x})}$$

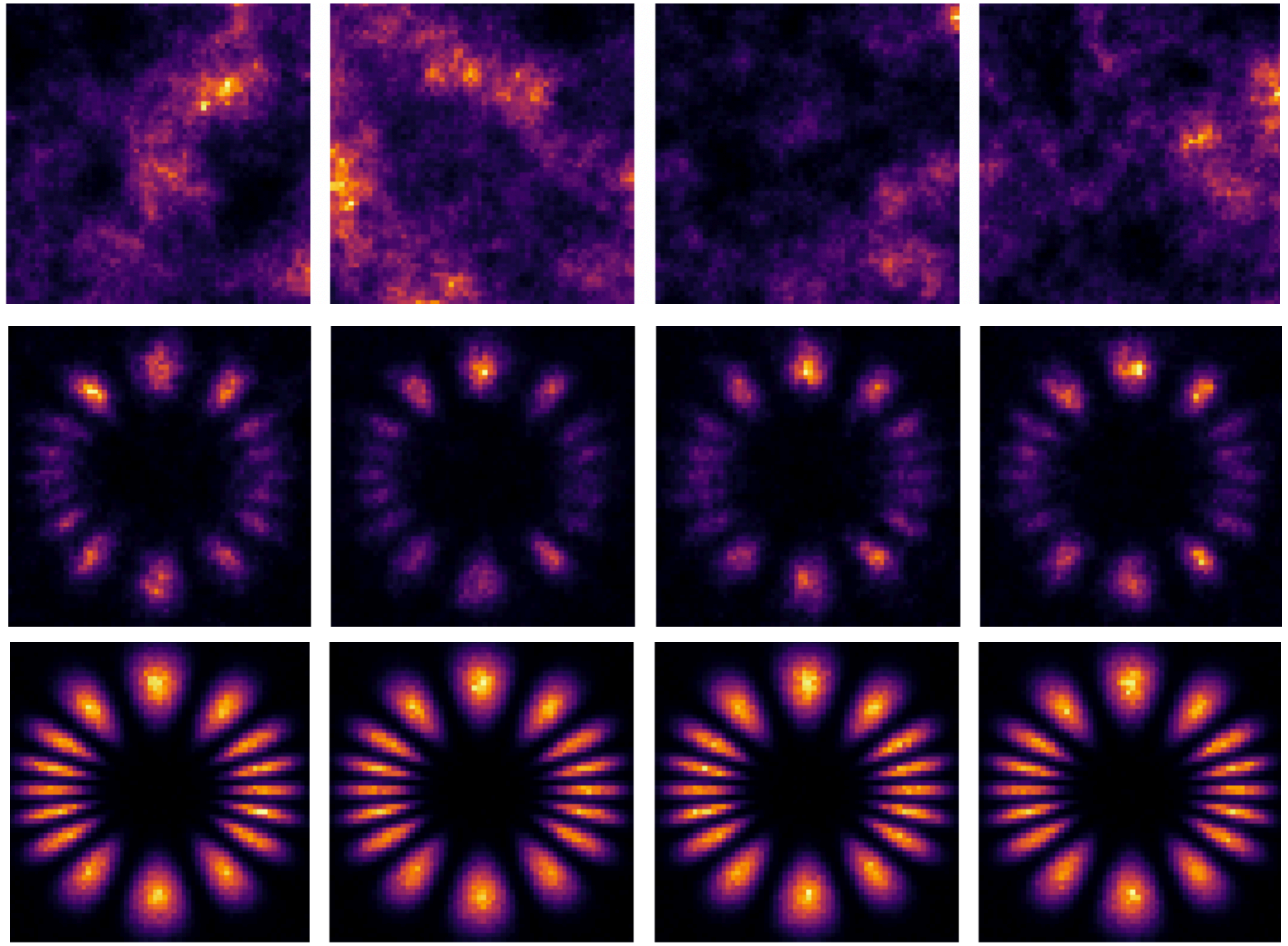
$\mathbf{z}$

(Gaussian approx)

Round 1

Round 2

Round 3



PRELIMINARY!

=> Posterior draws

Ongoing work: CW, Anau Montel, List

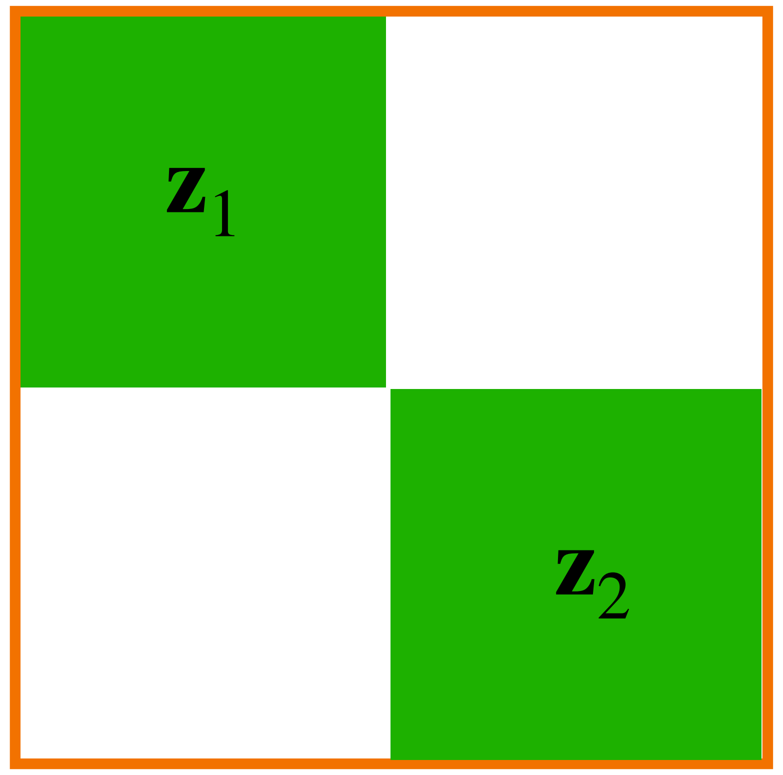
4) Image analysis

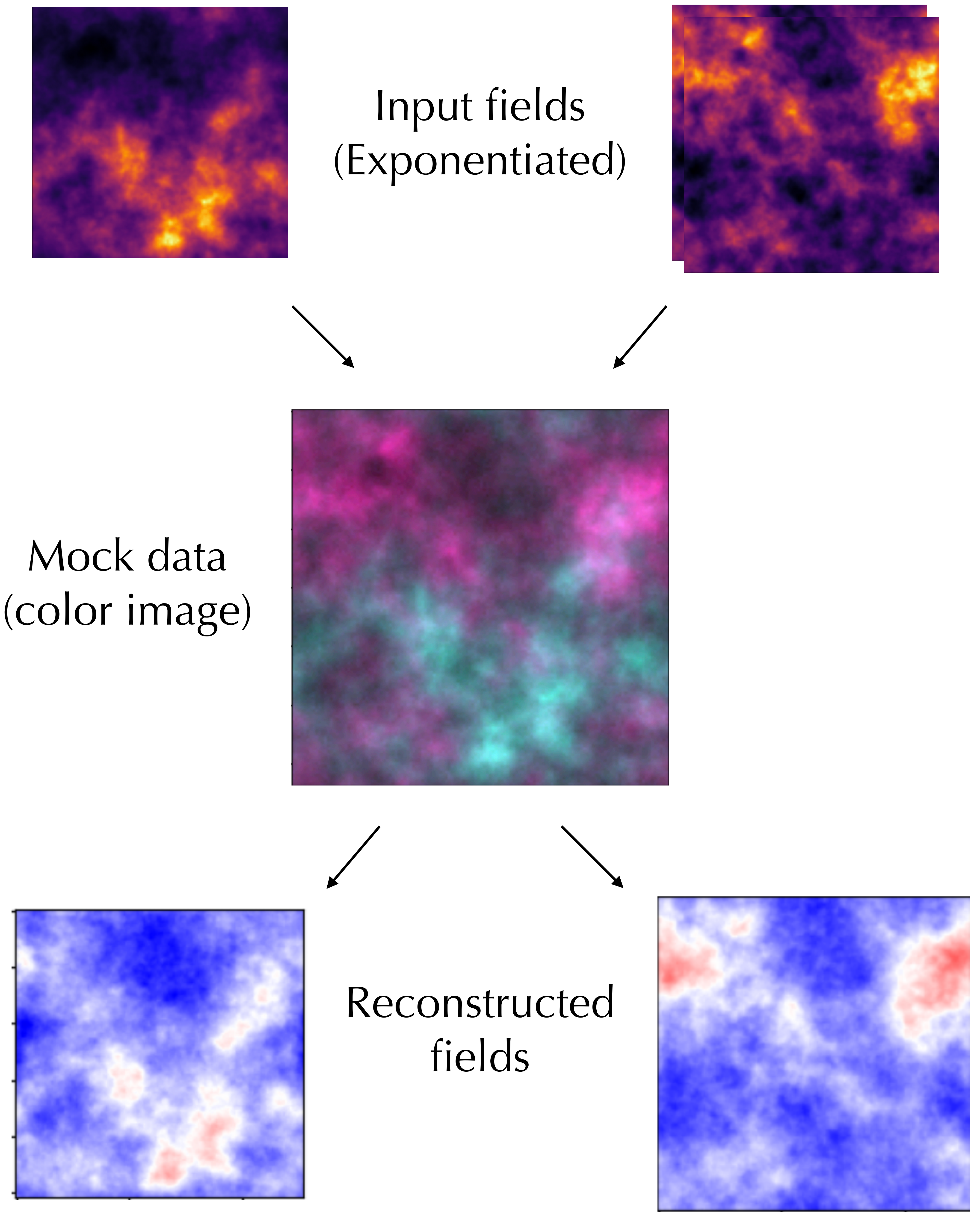
Maybe Proximal nested sampling? Cai+ 2106.03646

Example component separation
We learn the two high-dimensional likelihoods of each component, marginalised over the other components

$$\frac{p(\mathbf{x}\,|\,\mathbf{z}_1)}{p(\mathbf{x})} \qquad \frac{p(\mathbf{x}\,|\,\mathbf{z}_2)}{p(\mathbf{x})}$$

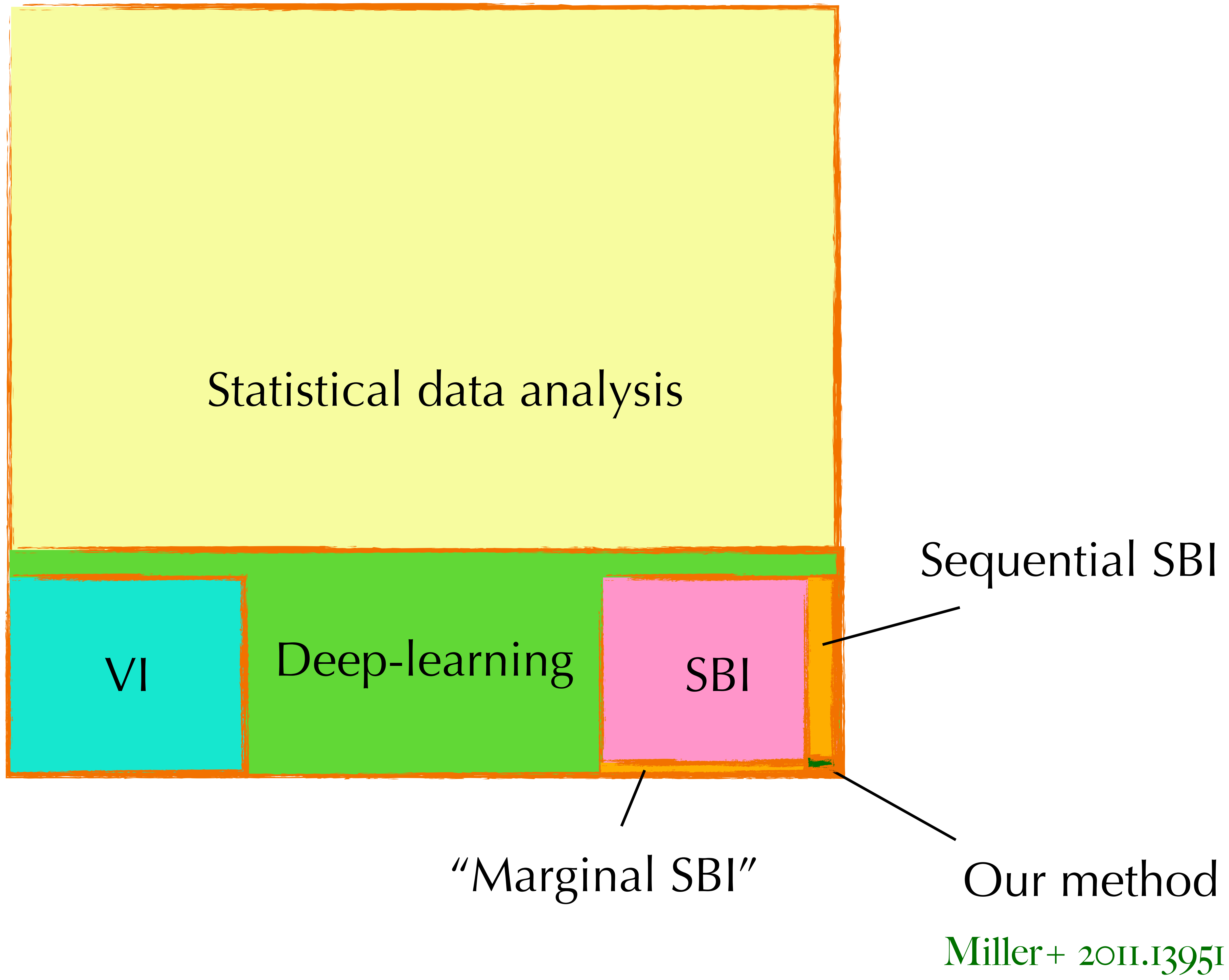
Truncation scheme





Ongoing work: CW, Anau Montel, List

Pretty niche, but growing exponentially



Rate of papers using TMNRE is growing exponentially

2021

1. "Fast and Credible Likelihood-Free **Cosmology** with Truncated Marginal Neural Ratio Estimation" Cole+ 2111.08030

2022

2. "Estimating the warm dark matter mass from **strong lensing** images with truncated marginal neural ratio estimation" Anau Montel+, 2205.09126
3. "SICRET: **Supernova Ia Cosmology** with truncated marginal neural Ratio EsTimation" Karchev+2209.06733
4. "One never walks alone: the effect of the perturber population on subhalo measurements in **strong gravitational lenses**" Coogan+ 2209.09918
5. "Detection is truncation: studying **source populations** with truncated marginal neural ratio estimation" Anau Montel+ 2211.04291

2023

6. "Debiasing **Standard Siren Inference** of the Hubble Constant with Marginal Neural Ratio Estimation" Gagnon-Hartman+ 2301.05241
7. "Constraining the X-ray heating and reionization using **21-cm power spectra** with Marginal Neural Ratio Estimation" Saxena+ 2303.07339
8. "Peregrine: Sequential simulation-based inference for **gravitational wave signals**", Bhardwaj+ 2304.02035
9. "Albatross: A scalable simulation-based inference pipeline for analysing **stellar streams** in the Milky Way", Alvey+ 2304.02032
10. …
11. …
12. …
13. …

Outlook & Conclusions

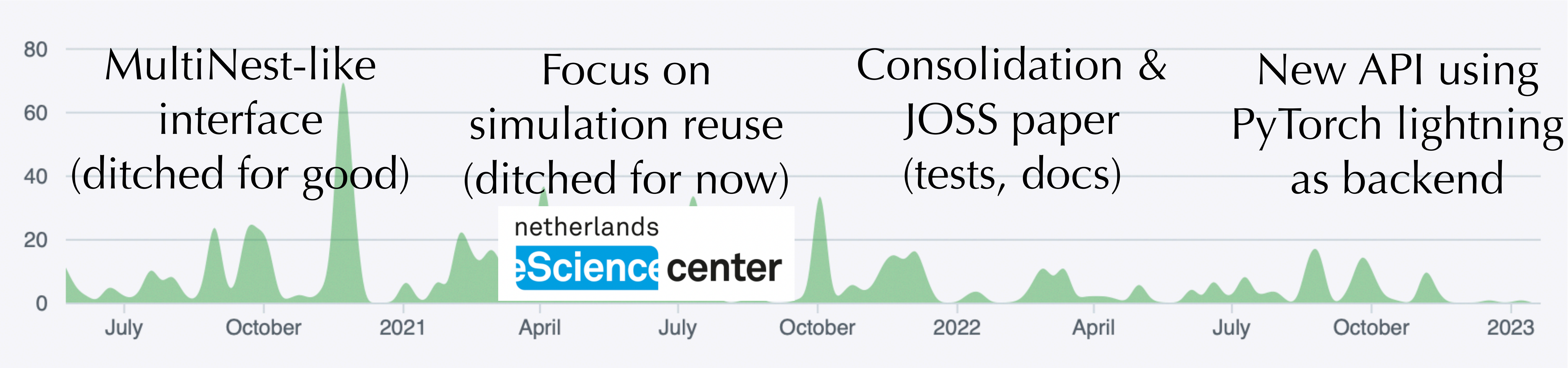
Swyft software package

Everybody being frustrated that code is changing all the time

"Are you using the latest Swyft version?"

~Stability

Initial plan:
"Hey, writing a python module for TMNRE would be cool and impactful, ***should take 2-3 weeks****"*



3 years later…

Miller+ 2011.13951

Miller+ 2107.01214

Miller+ JOSS 2022

Methods that we worked with in our group

Variational inference

PyKeOps

Normalising flows

HMC

TMNRE

Gaussian processes

Probabilistic programming

Differentiable simulators

Scalable TMNRE

Density TMNRE

Hierarchical TMNRE

Image analysis TMNRE

Open questions

- **Gradients**: How to exploit gradient information for TMNRE? Is there a way? Is it worth it?
- **Hard likelihood constraint prior samples:** How to most efficiently sample from constrained likelihood regions in very high dimensions (Langevin sampling, proximal optimisation methods?)
- **Automatisation**: Can the determination of truncation schemes and optimal network architectures be automatised? Can ChatGPT help?
- **Goodness-of-fit:** How to perform goodness-of-fit tests etc in the context of SBI? How to detect that the model is wrong?
- **Data volume:** How to handle situations with high volume data? Storing all simulation data seems infeasible in this case.
- **Fundamental limitations:** Are there inference tasks that *only* can be done with the joined posterior, and would not be accessible by TMNRE?

Conclusions

- Finding new physics in complex data is becoming increasingly challenging.
- Traditional data analysis techniques cannot recover the full statistical inference picture in many cases.
- SBI can provide accurate and precise *projections* of the full inference problem.
- Swyft/TMNRE is our attempt to make marginal inference possible.
- Lots of promising results, much more to come, stay tuned!

Thanks!