Low-latency parameter estimation of Massive Black Hole Binaries for LISA using flow matching

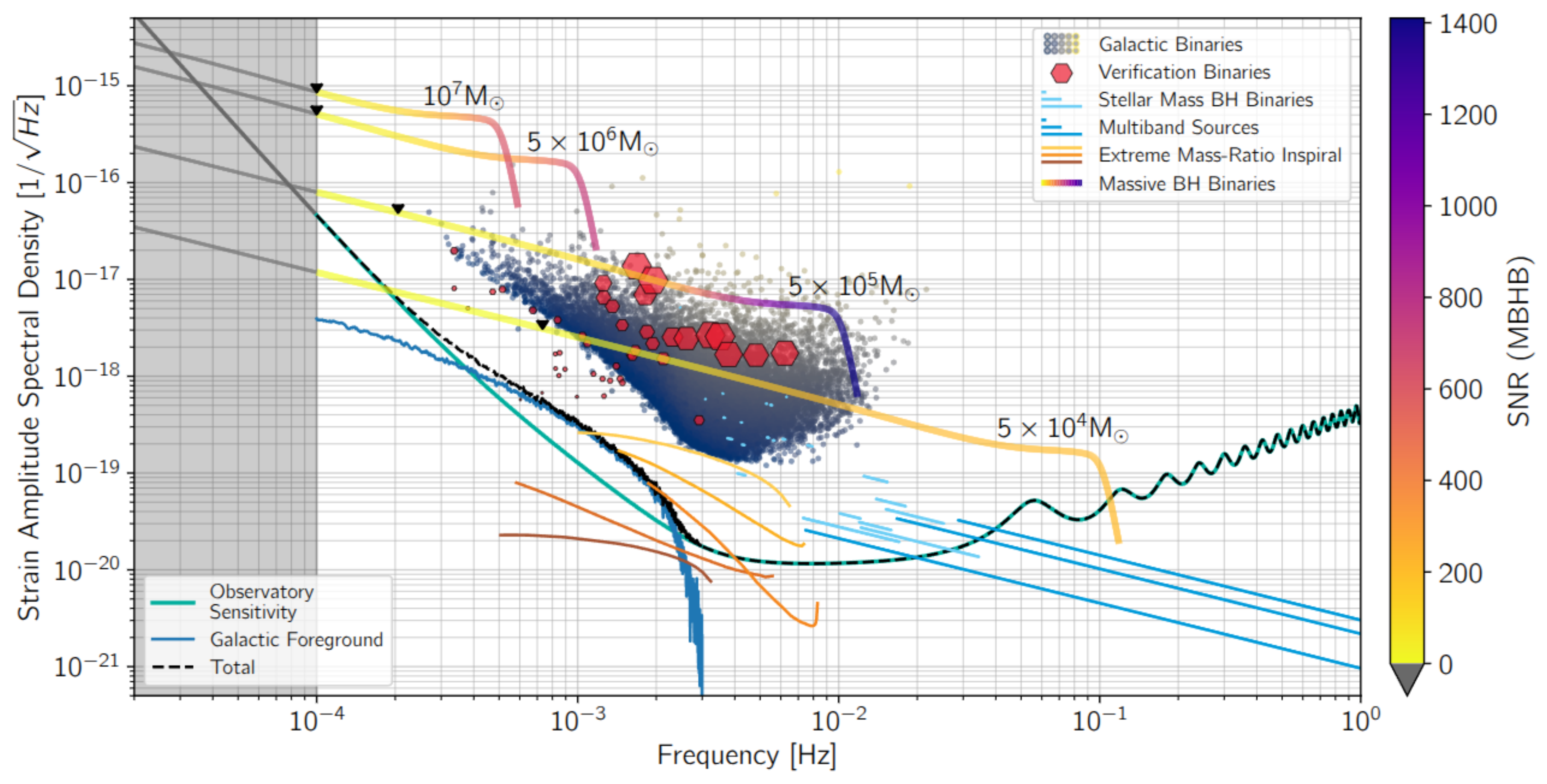
Malvina Bellotti, PhD student @ Observatoire de la Côte d'Azur, Nice

Supervisor: Natalia Korsakova (Observatoire de la Côte d'Azur, Nice)
Co-supervisor: Jess McIver (University of British Columbia, Vancouver)

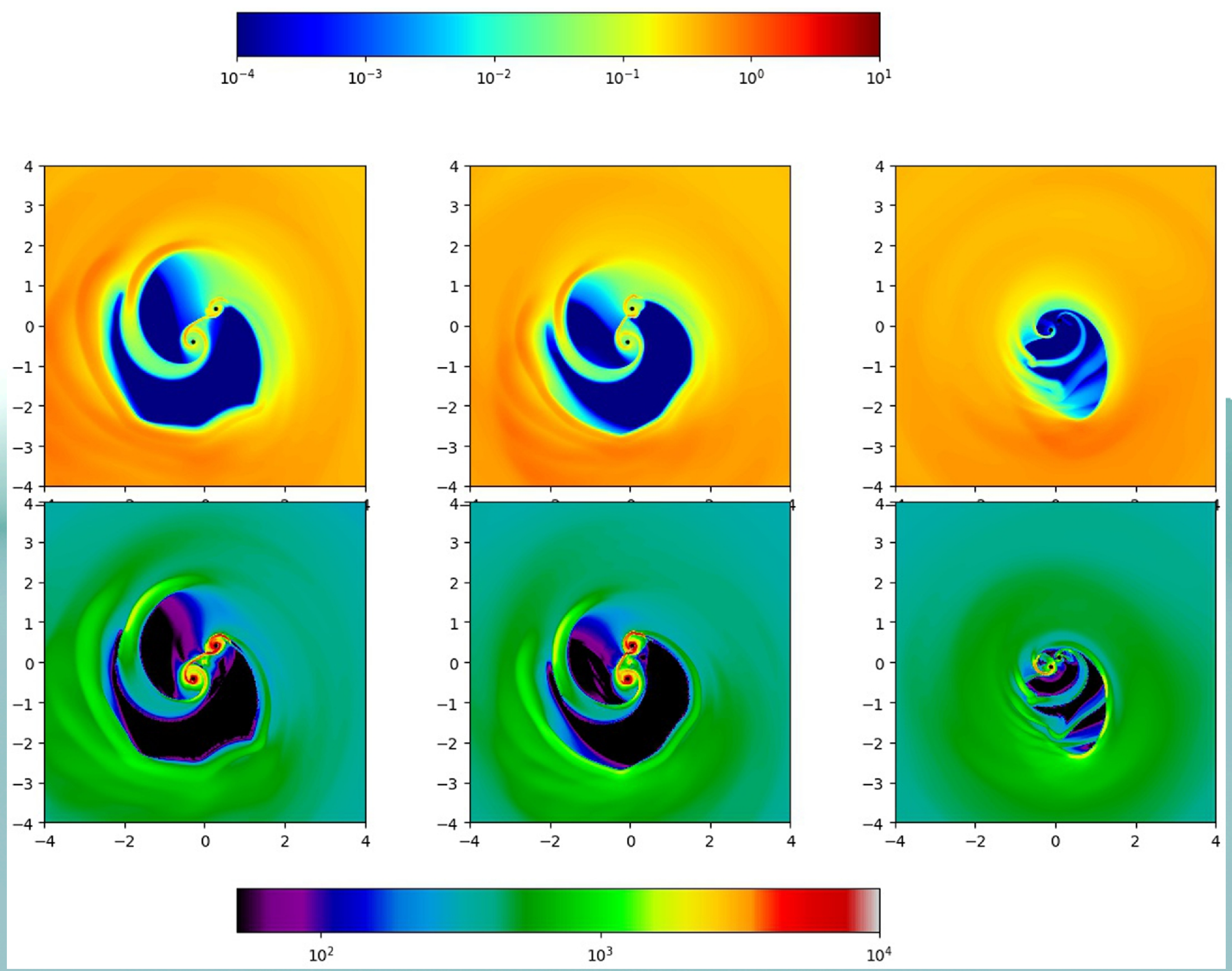
Massive Black Hole Binaries

$10^4\ M_\odot < M < 10^7\ M_\odot$

[Colpi+ 2024]



In presence of sufficient amount of gas an EM counterpart can be triggered



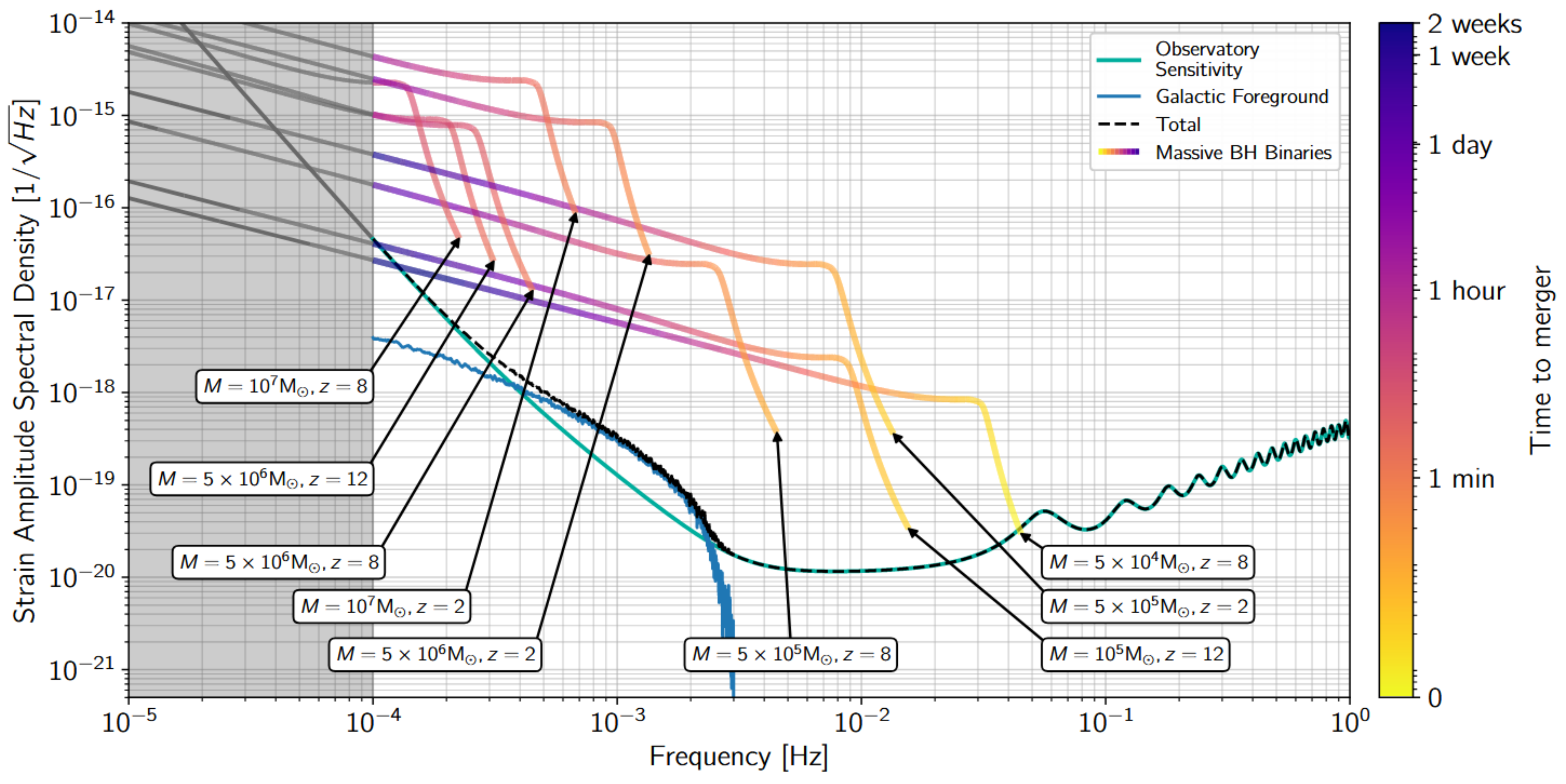
Tang +18

- Found at the centre of most galaxies
- Hierarchical galaxy mergers → binary systems
- Enter LISA band days to months before merger
- Loudest LISA sources → test GR, probe cosmic history, constrain cosmology

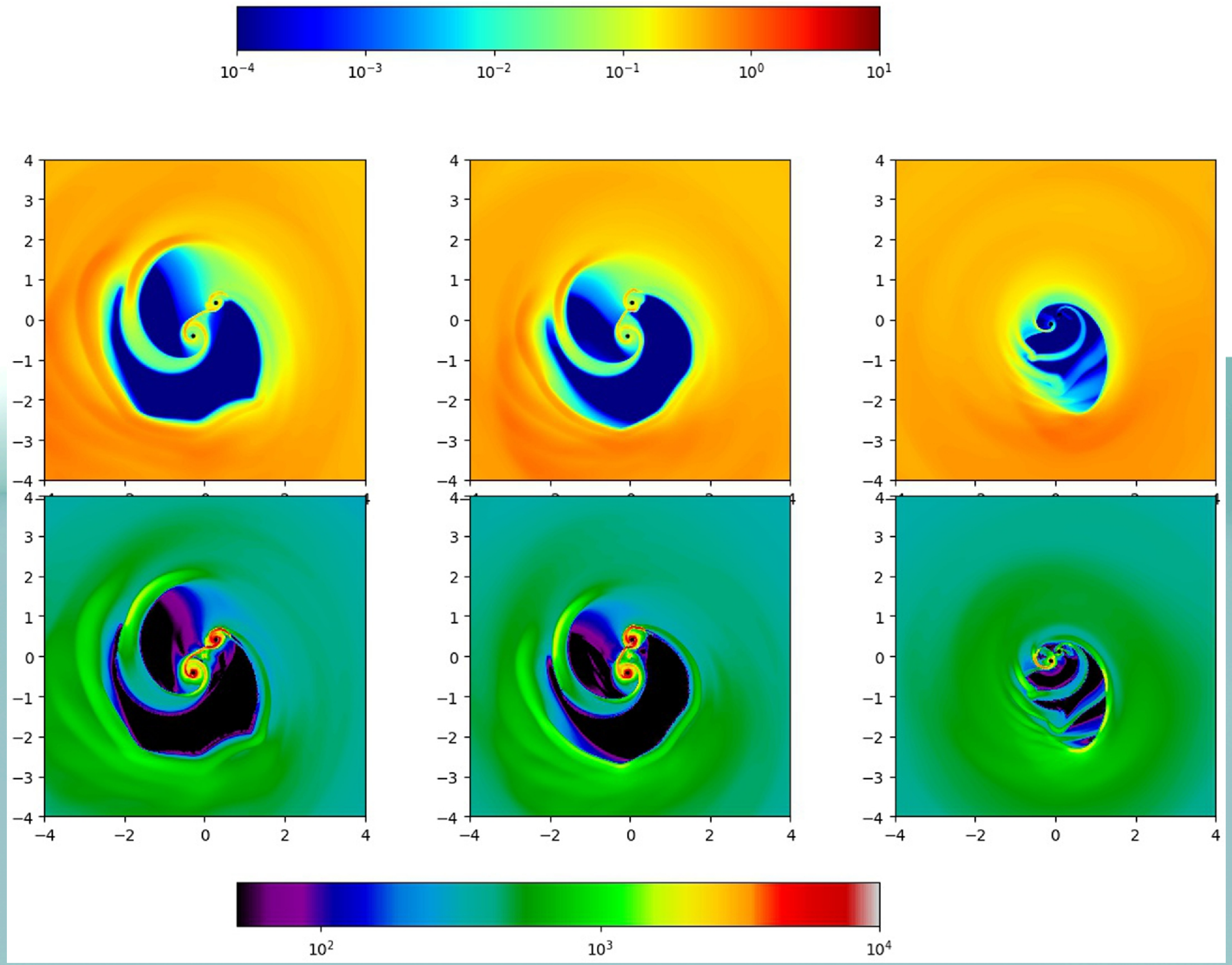
Massive Black Hole Binaries

10^4 M⊙ < M < 10^7 M⊙

[Colpi+ 2024]



In presence of sufficient amount of gas an EM counterpart can be triggered



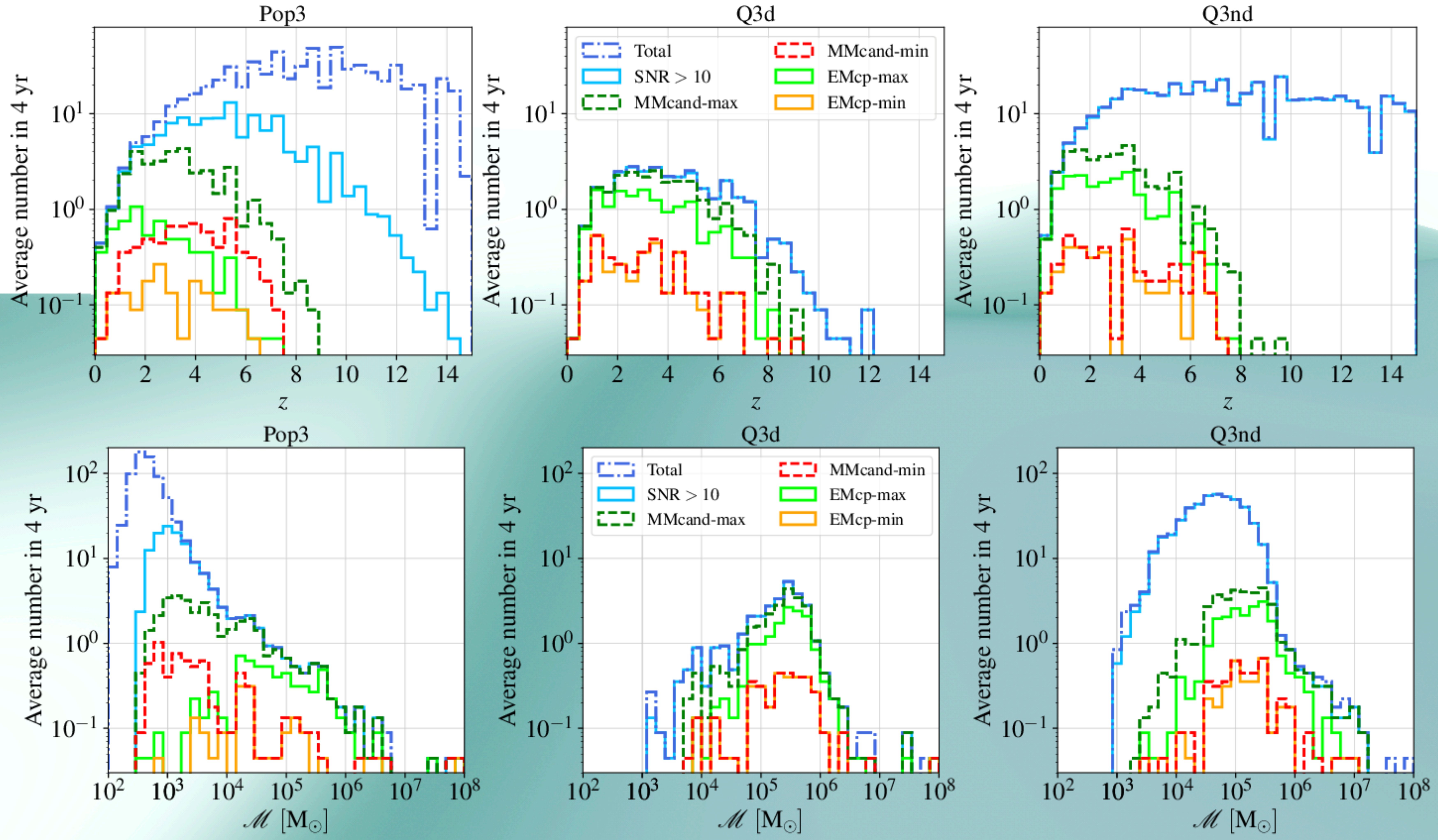
Tang +18

- Found at the centre of most galaxies
- Hierarchical galaxy mergers → binary systems
- Enter LISA band days to months before merger
- Loudest LISA sources → test GR, probe cosmic history, constrain cosmology

Population of MBHB we are focusing on

We focus on a sub-population of MBHBs, with masses between 10^4 and 10^5 solar masses, astrophysically motivated by the possibility of observing an EM signature for such systems

Mangiagli et al +22



Low Latency Alert Pipeline

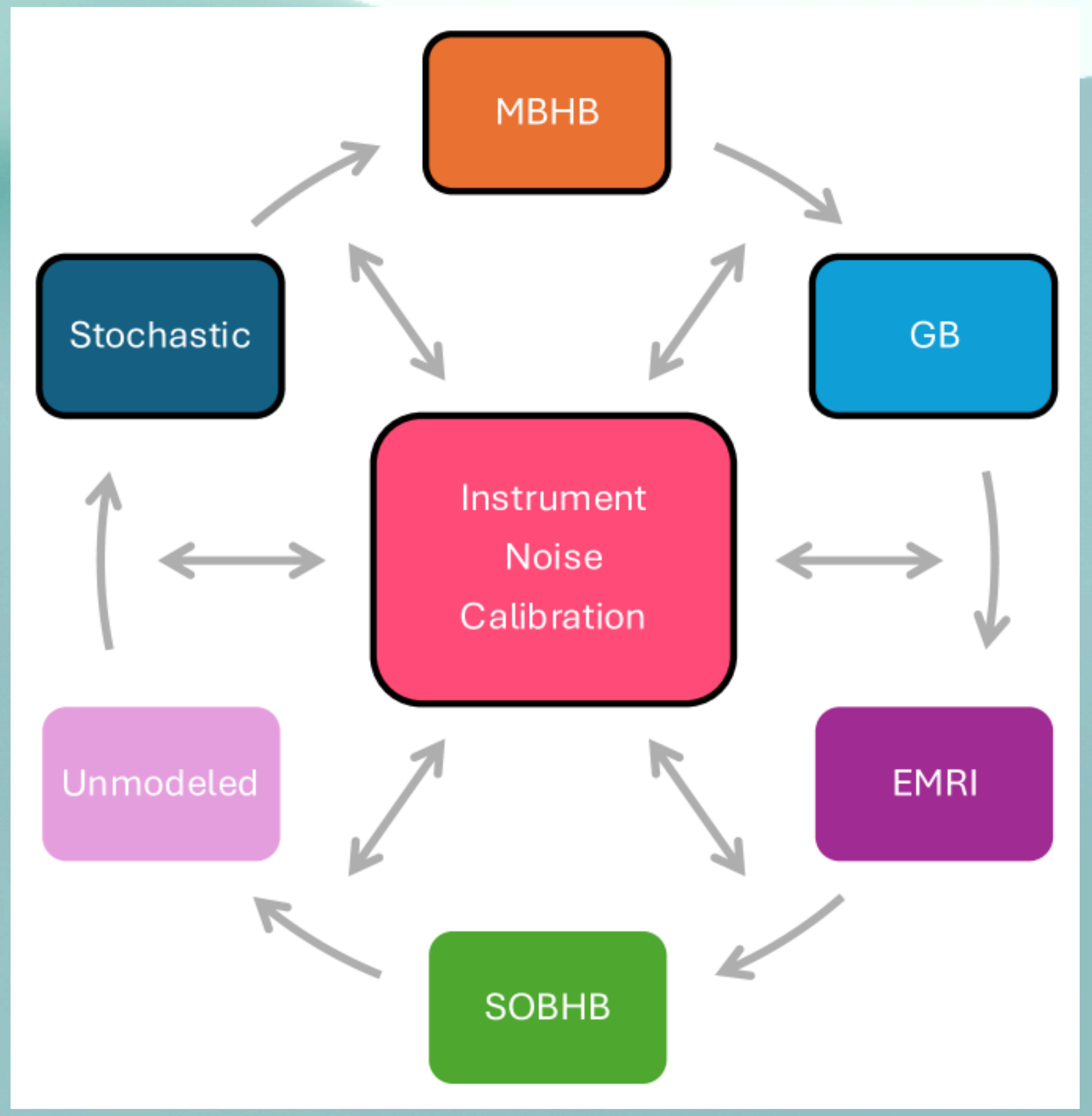
Real-time detection essential for:

- Trigger **protected periods** in LISA data stream
- **Multi-messenger alerts**: precursor EM emission detection + afterglow surveys
- Trigger **complete parameter estimation** pipelines

Goal:
analyze transients in **less than 1 hour** from data reaching the ground (focus on pre-merger phase for MBHBs)

continuously **update parameters** (masses, distance, merger time, sky position)

Global Fit is **too slow** for real time

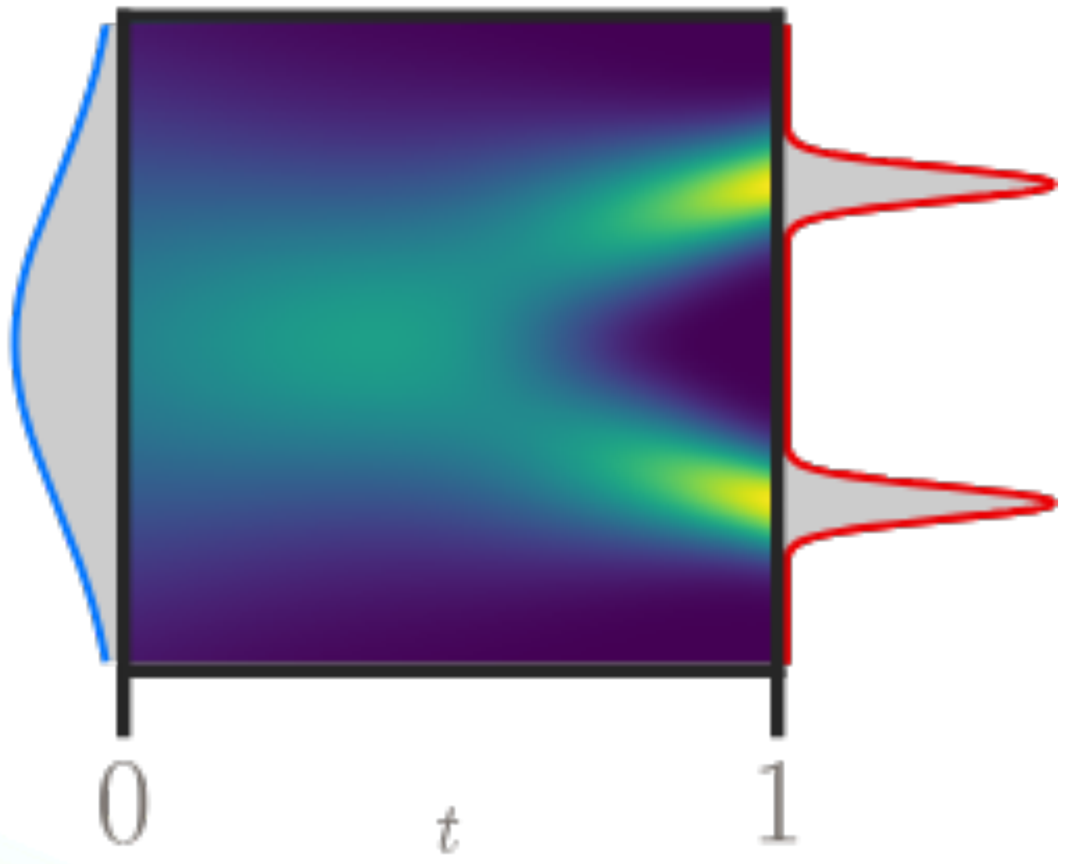


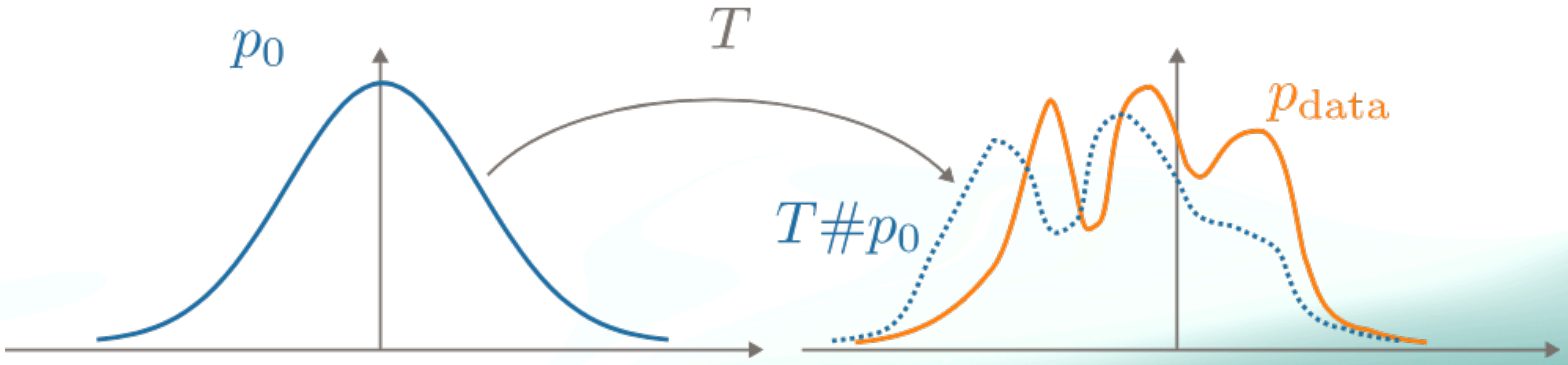
Katz+ 25

Focus on **MBHB**! We have to perform analysis in the presence of **other sources/realistic noise** but we can marginalise over them!

Our SBI framework: conditional flow matching

Flow Matching trains a neural network to predict the **velocity vector field** describing a the path from an easy distribution into a complex target distribution (parameters' posteriors)





https://dl.heeere.com/conditional-flow-matching/blog/conditional-flow-matching/

$$\mathscr{L}(\theta) = \mathbb{E}_{t,x_0,x_1}\left[||\, v_\theta(t, x_t \mid \mathbf{y}) - u_t \,||^2 \right]$$

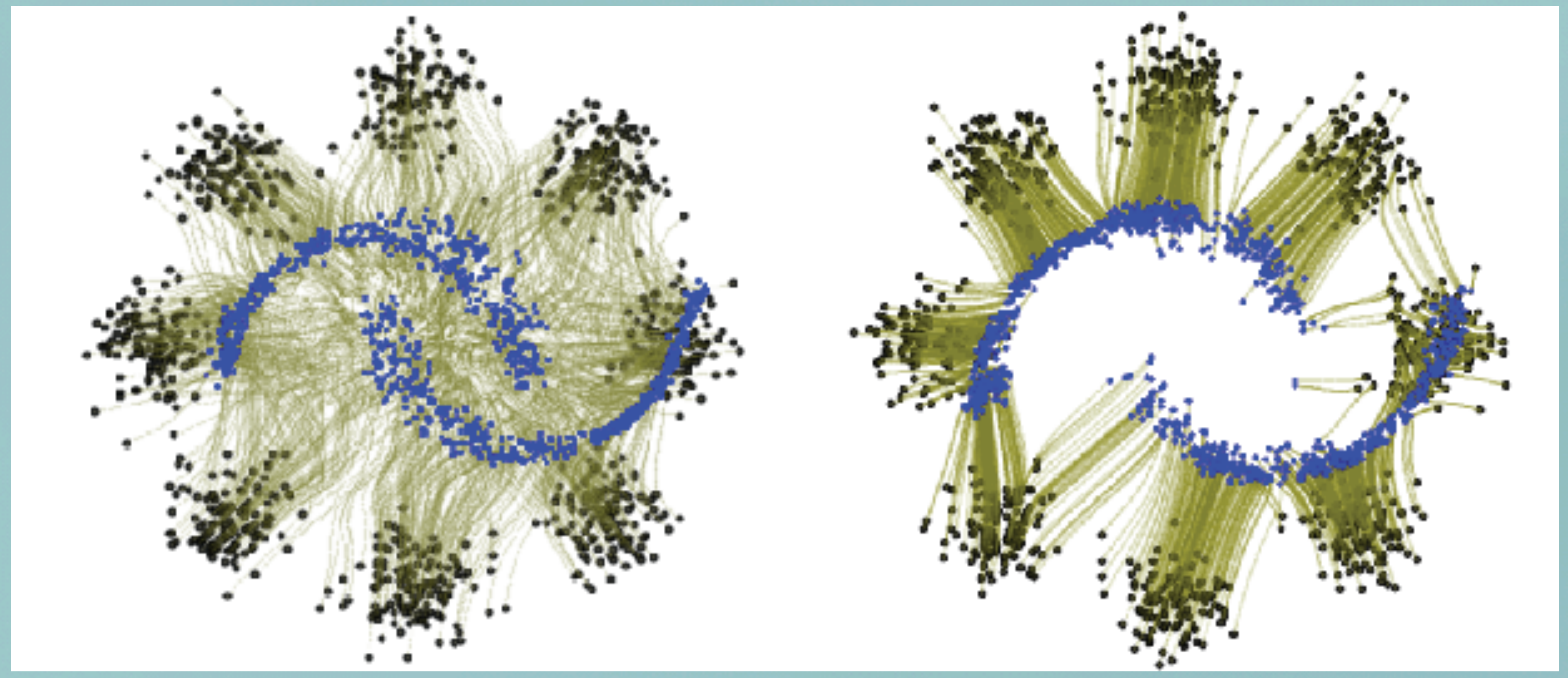
Loss: Mean Squared Error regression

prediction

target velocity

Optimal transport:

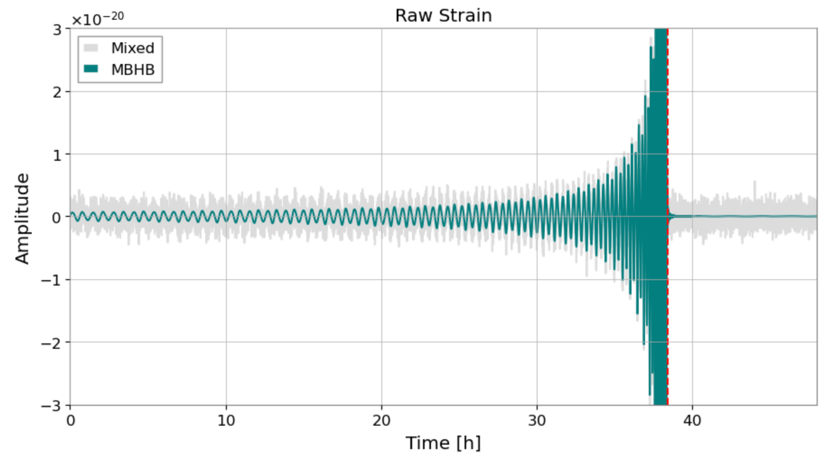
forces the paths to be straight lines with constant speed.



Tong et al + 23

Parameter estimation on MBHB

We train and evaluate our network on two data representations:

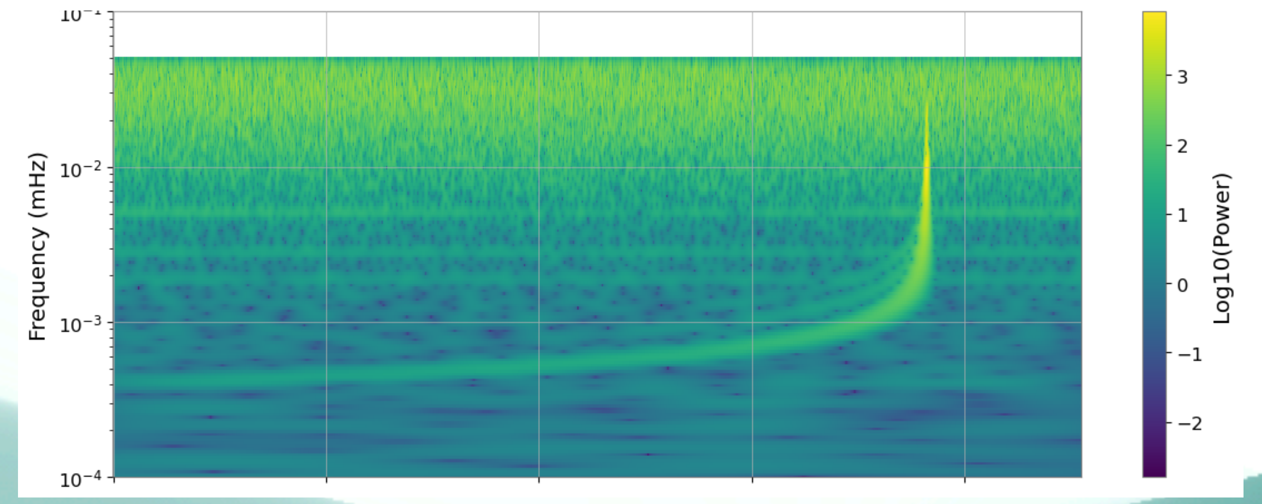


time series

Training data produced
on-the-fly

but

NNs could struggle to extract long temporal features from 1D sequences

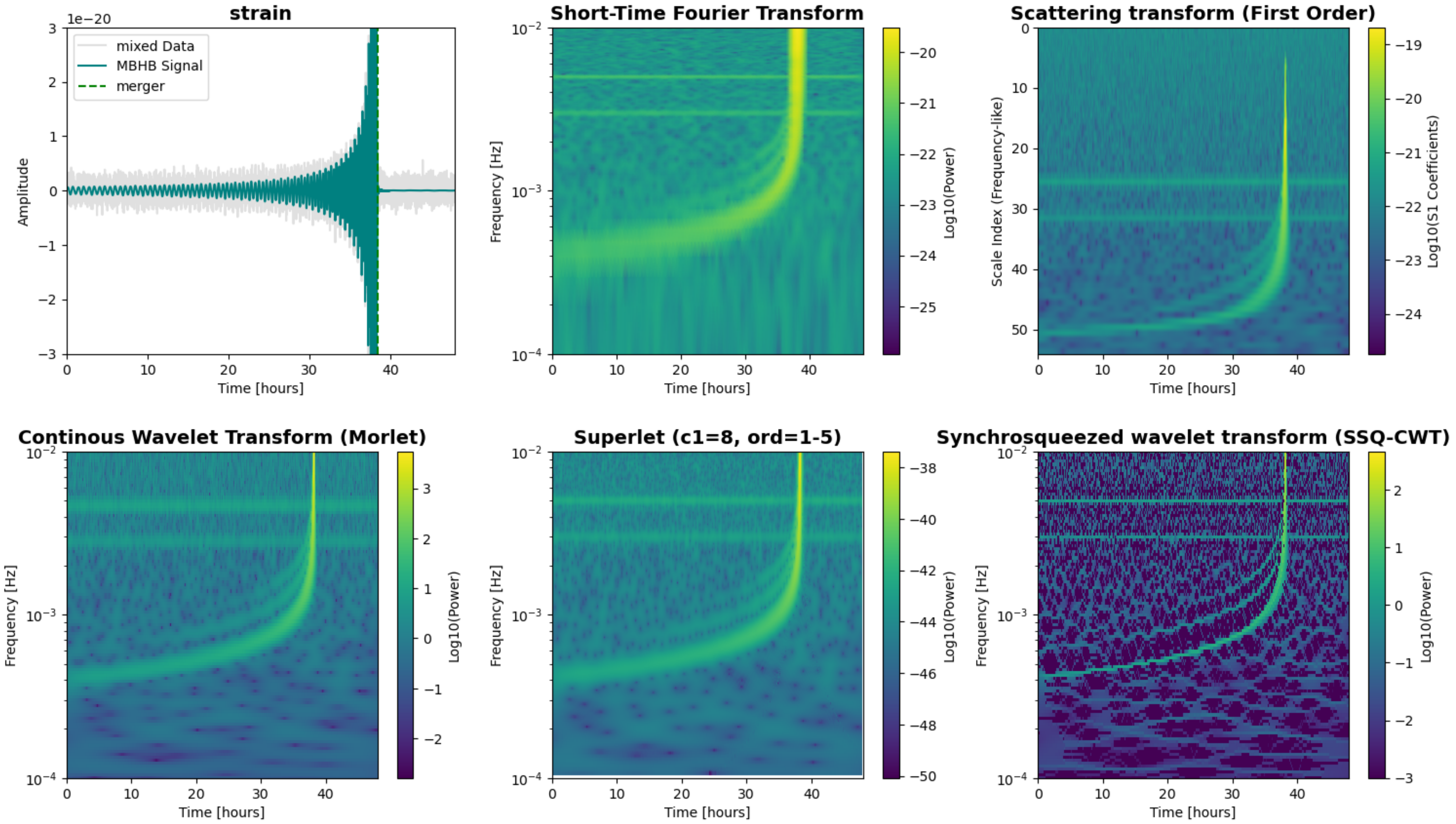


time-frequency

Offline production of dataset (memory-heavy, size-limited), easier for UNet 2D pattern recognition

Exploring different time-frequency representations

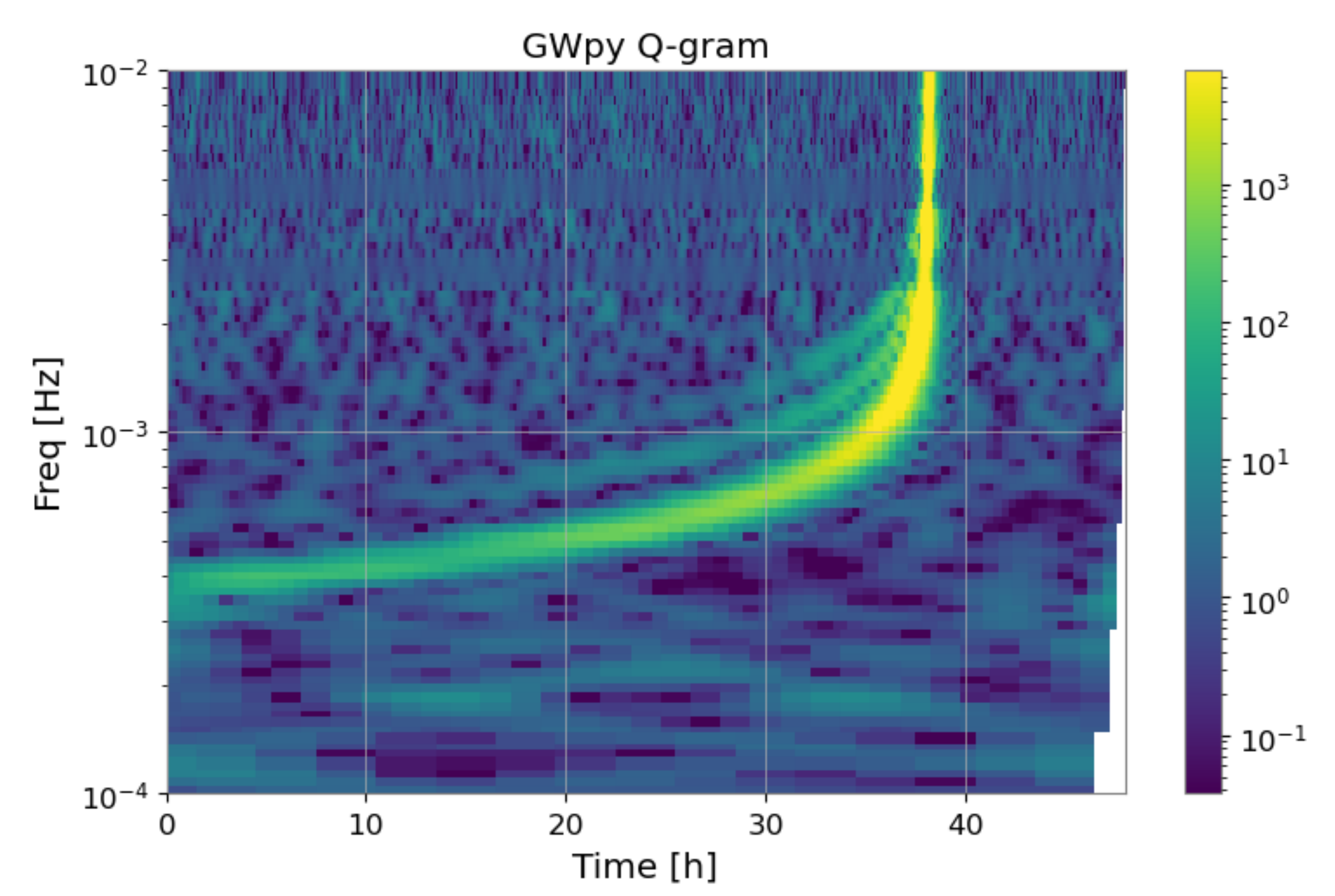
Choosing the optimal time-frequency representation



BENCHMARK RESULTS:

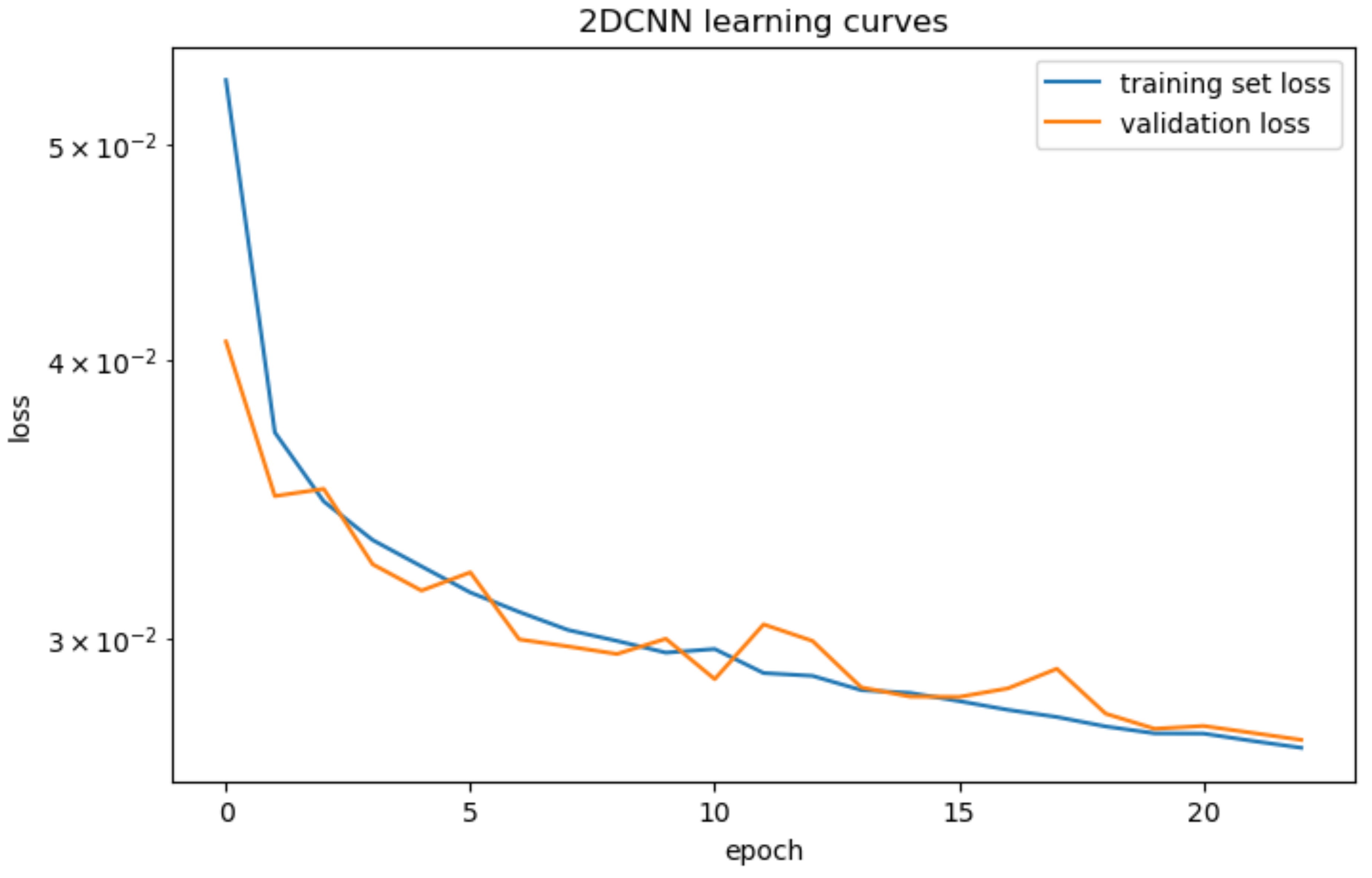
Method	Time (s)	RAM (MB)
STFT	0.001	0.673
Scattering1D	0.635	7.119
Superlets	6.547	225.571
CWT / SSQ	0.125	23.730
GWpy Q-gram	0.046	0.222

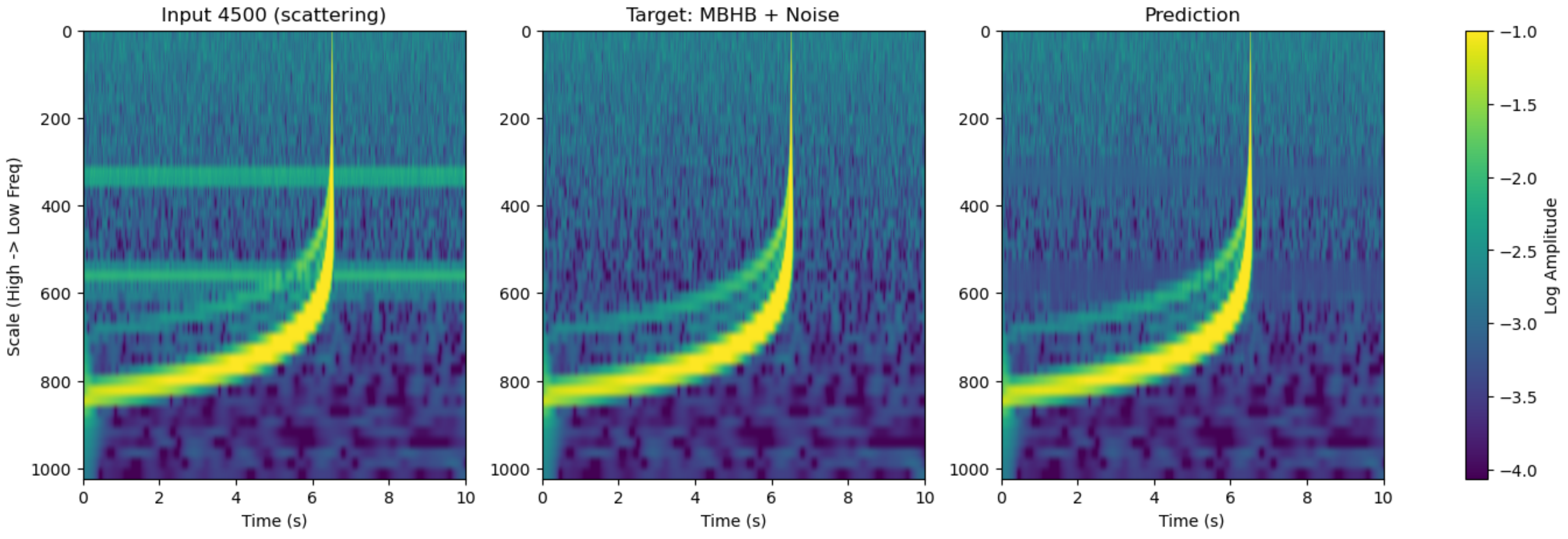
+ 1s to build the plot!

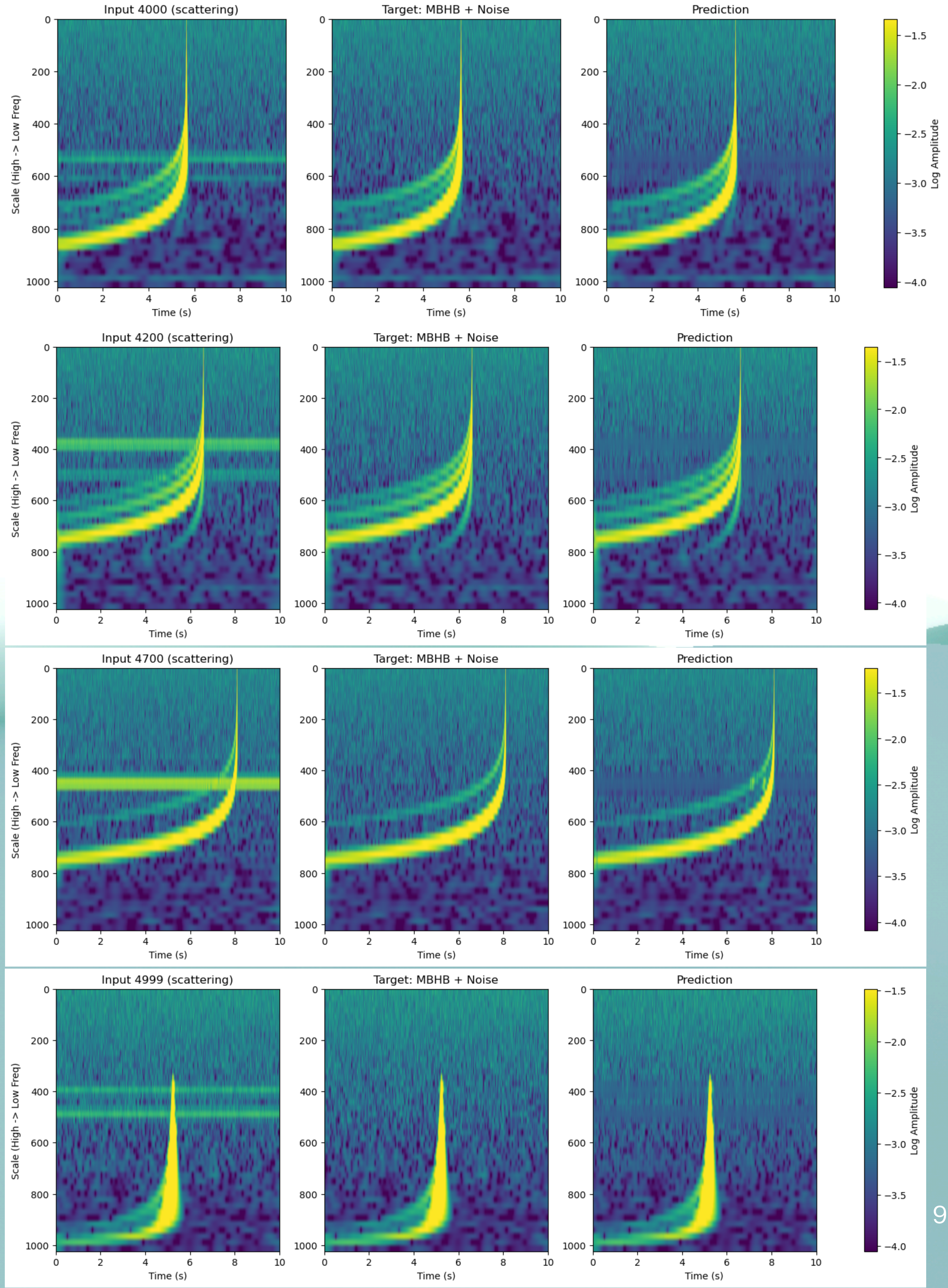


Ignoring galactic binaries using autoencoders

2D CNN scattering transform

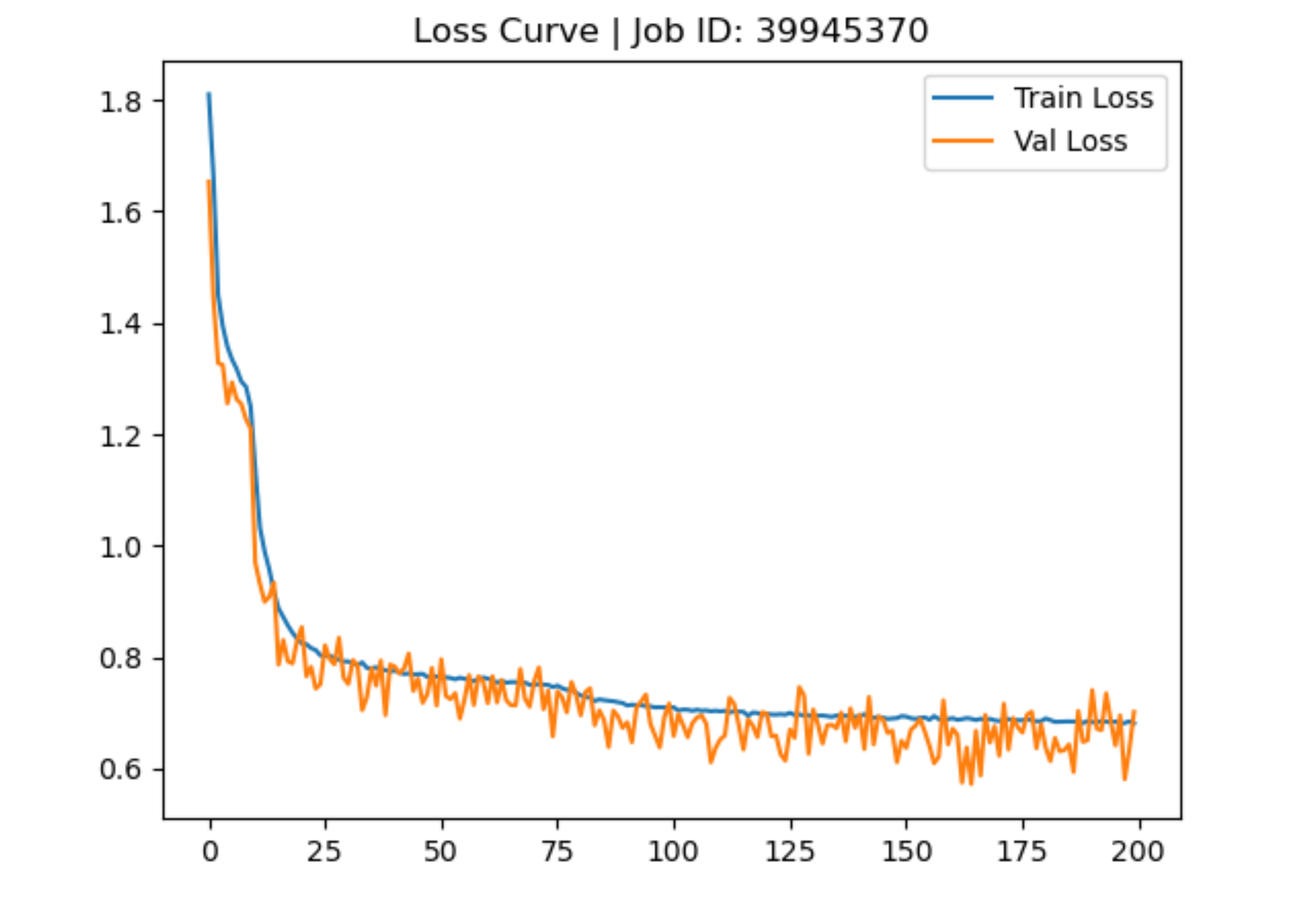


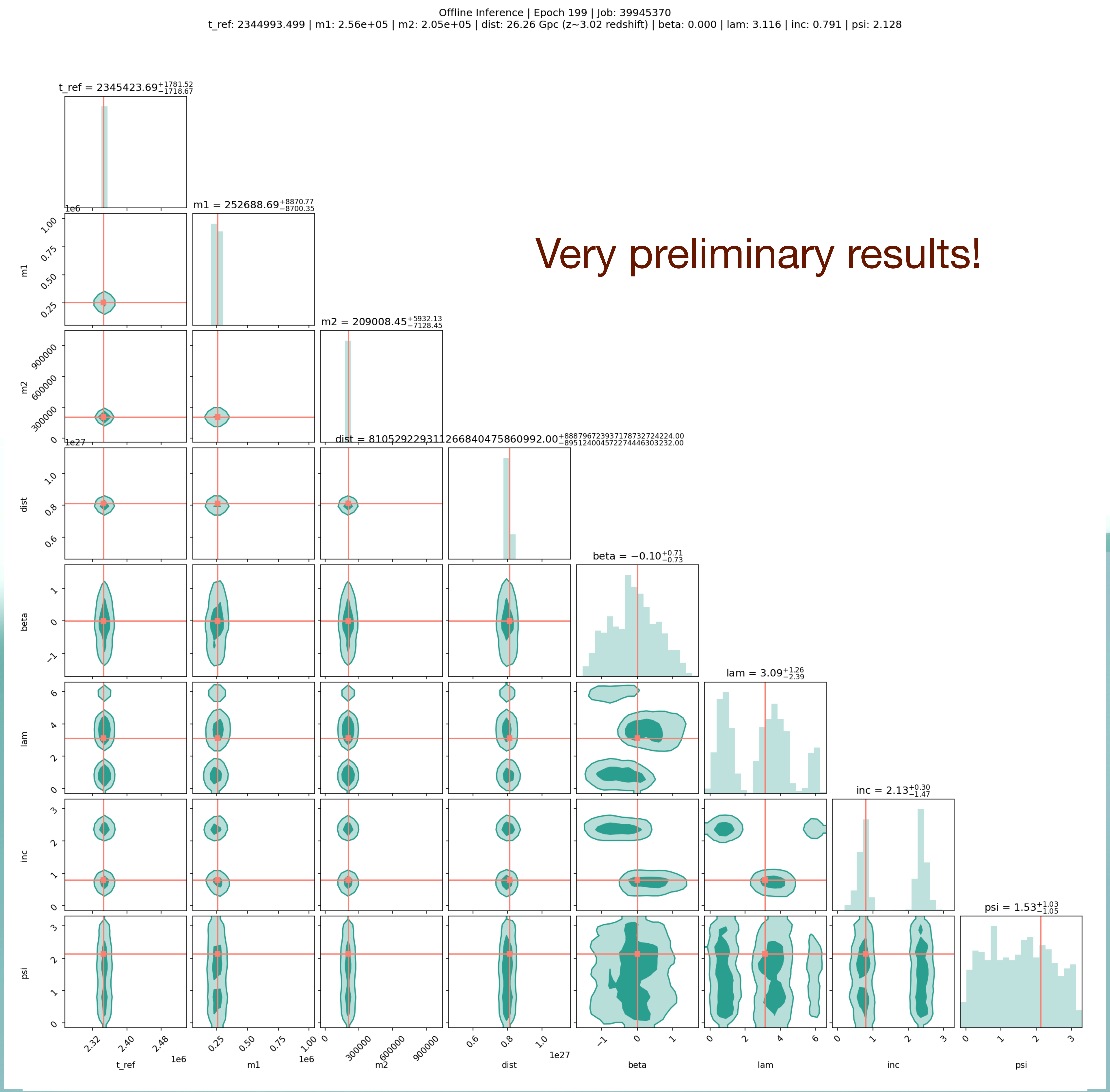




PE on time-frequency

Benchmarking against **standard Bayesian inference methods** (MCMC or Nested Sampling) is **necessary** to establish the posteriors that we expect.





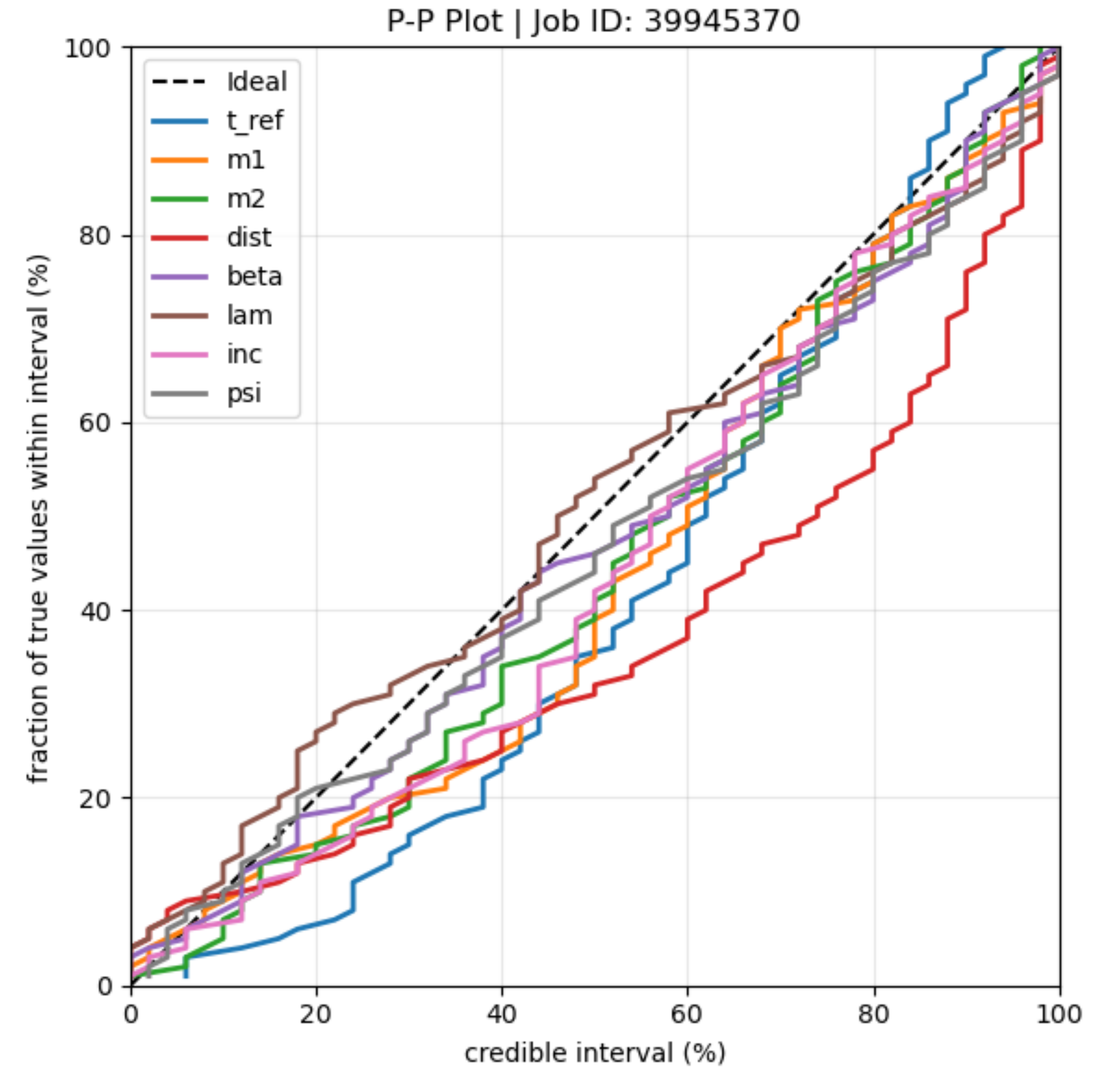
Very preliminary results!

Preliminary results: P-P plot

Compares the predicted credible intervals against the actual recovery rate of the true parameters.

true parameters fall outside the predicted credible intervals more frequently than expected

we need further training, or hyperparameter tuning

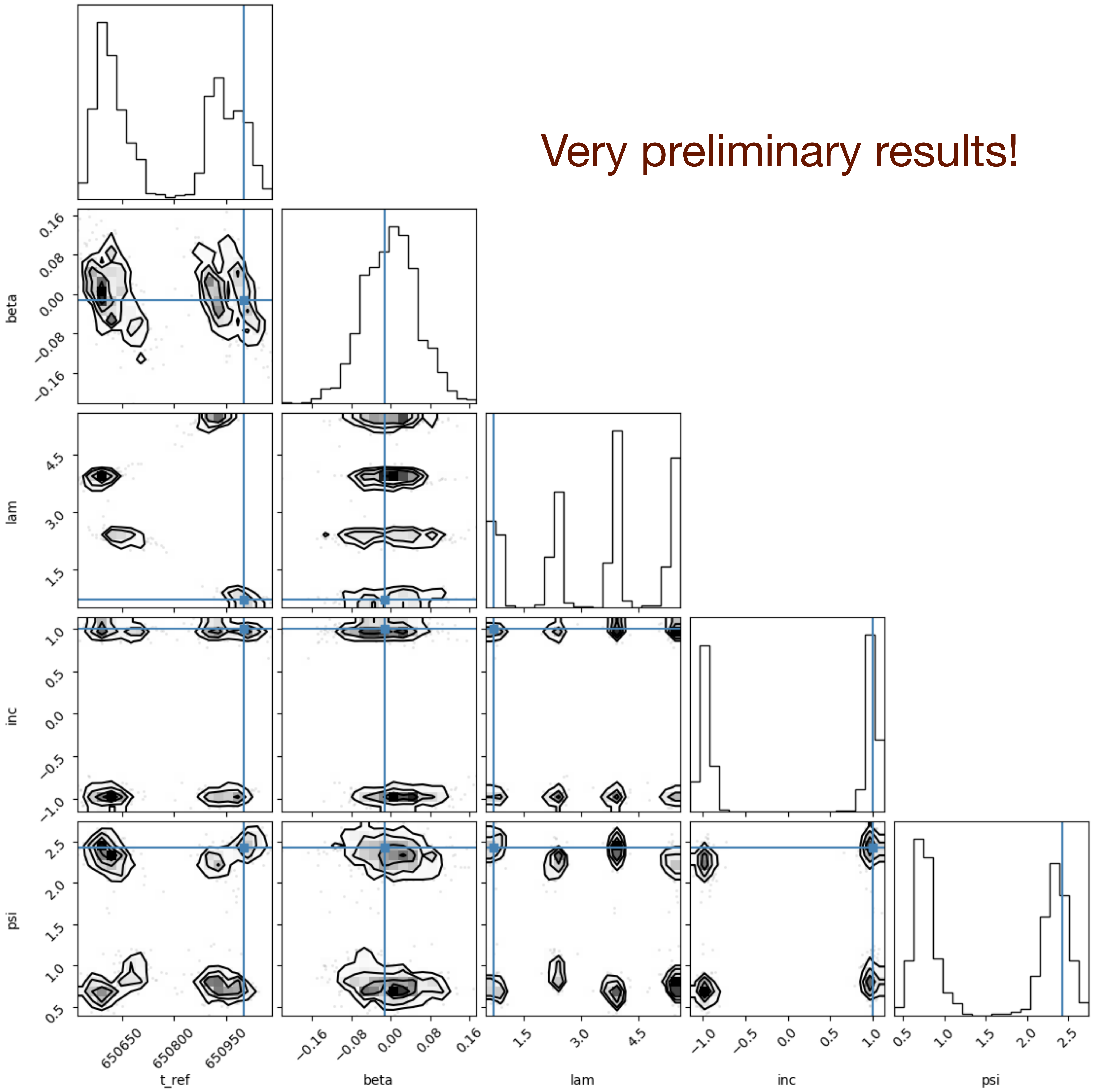


Parameter estimation

Time series

Benchmarking against **standard Bayesian inference methods** (MCMC or Nested Sampling) is **necessary** to establish the posteriors that we expect.

Very preliminary results!



To summarise:

Low-Latency parameter estimation: Simulation-Based Inference using Conditional Flow Matching and Normalizing Flows allows rapid, likelihood-free PE

SBI can naturally **marginalise** over other **signals** and **noise artefacts**

Data representation: time-frequency representations might better align with our architecture

Next steps:

- Compare to standard Bayesian inference methods
- Apply our method to data containing realistic noise and gaps

Thank you !

Backup slides

MBHB

MBHs need to reach the milliparsec distance to merge in $t < t_{\rm Hubble}$

$$t_{\rm GW}, \simeq 1Gyr\left(\frac{a}{mpc}\right)^4\left(\frac{10^6 M_\odot}{M}\right)^3$$

Dynamical friction $\longrightarrow$ binary hardening $\longrightarrow$ gravitational wave emission

kpc $\longrightarrow$ pc   pc $\longrightarrow$ mpc   mpc $\longrightarrow$ coalescence

EM emissions in gas rich environments:

- Circumbinary disks: accretion onto circumbinary disk and mini-disks around each BH
- Magnetized jets
- Post-merger shocks: shock waves induced by gravitational wave kick after coalescence
- Emission from host galaxy

Population of MBHB we are focusing on

We focus on a sub-population of MBHBs, with masses between 10^4 and 10^5 solar masses, astrophysically motivated by the possibility of observing an EM signature for such systems

Mangiagli et al +22

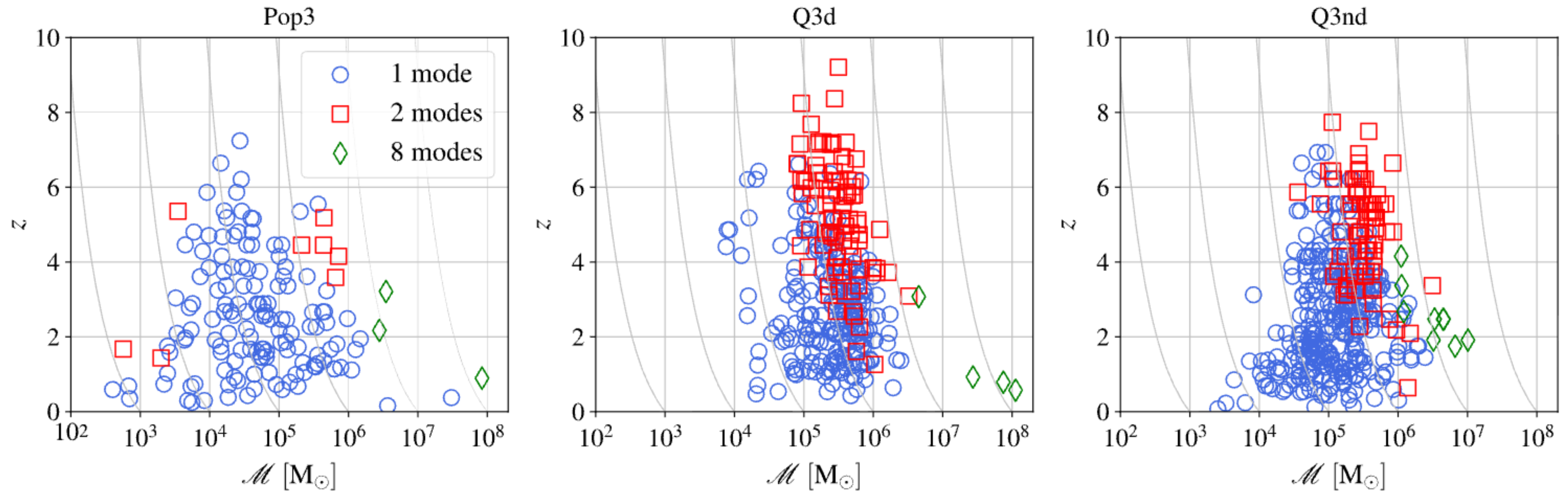


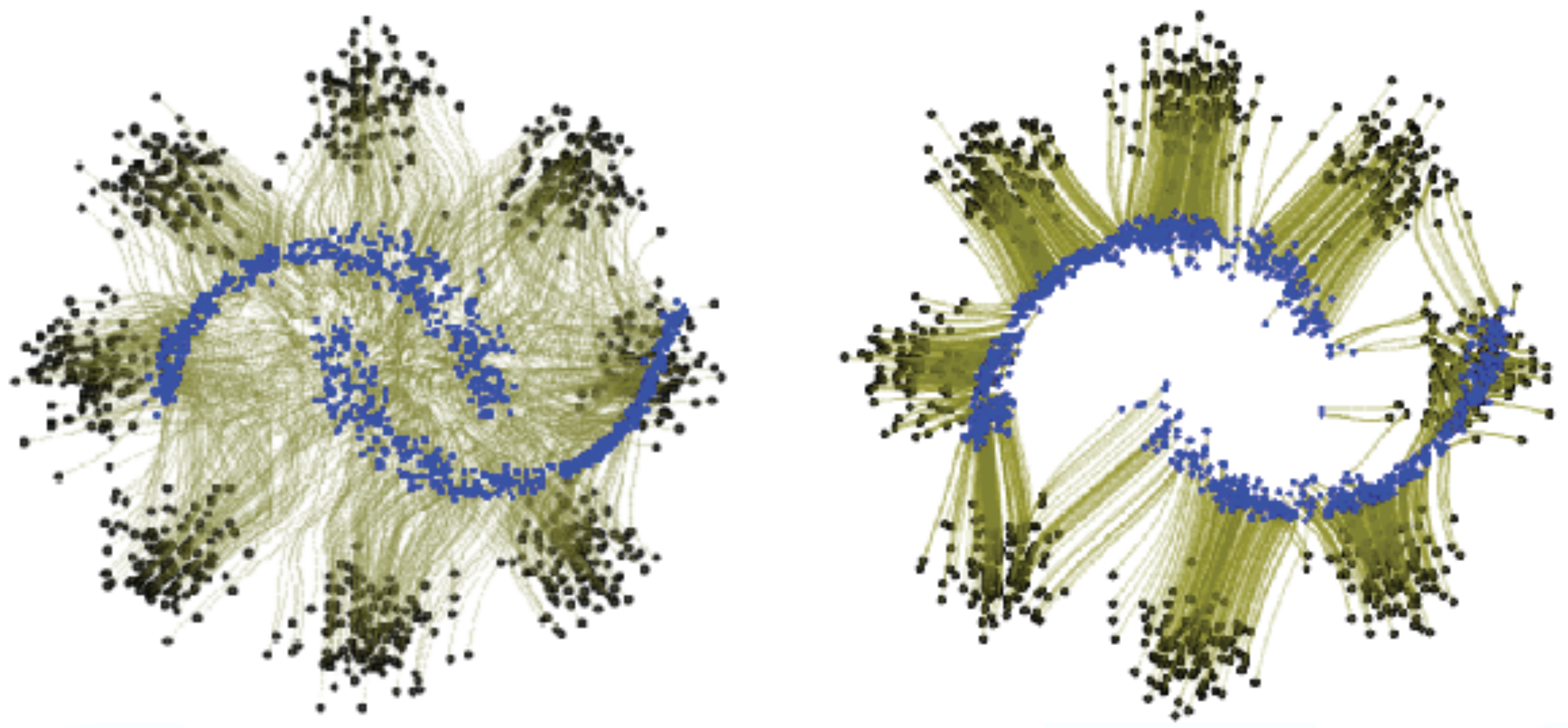
FIG. 13. Distribution of the *1mode*, *2modes*, and *8modes* EMcps in the $z - \mathcal{M}$ plane in the maximizing case for the three astrophysical scenarios. The gray solid curved lines in background correspond to constant redshifted chirp mass values.

How does conditional flow matching work?

$$
\begin{tikzcd}[column sep=huge, row sep=large]
p_0(x_0) \arrow[dd, "\text{Solve } \frac{\mathrm{d}x}{\mathrm{d}t} = u_t(x)"', "\text{with } x(0) = x_0"] & & p_{0|1}(x_0 \mid x_1) \arrow[ll, "\mathbb{E}_{x_1 \sim p_1}"'] \arrow[dd, "\text{Solve } \frac{\mathrm{d}x}{\mathrm{d}t} = u_t(x \mid x_1)"', "\text{with } x(0) = x_0"] \\
{} & & {} \arrow[ll, squiggly, "\mathbb{E}_{x_1 \sim {\color{red} p_{1|t}(\cdot|x)}}"'] \\
p_t(x_t) & & p_{t|1}(x_t \mid x_1) \arrow[ll, "\mathbb{E}_{x_1 \sim p_1}"']
\end{tikzcd}
$$

Our SBI framework: conditional flow matching

Tong et al + 23



Optimal transport

Optimal Transport forces the paths to be straight lines with constant speed. This makes the paths non-crossing, drastically simplifying what the neural network has to learn.

Could work on changing paths to adapt to our model (not straight path)

We construct a probability path that connects noise $x_0 \sim \mathcal{N}(0,I)$ to the true parameters $x_1 \sim q(x_1)$

Using Optimal Transport the path is defined as: $x_t = (1-t)x_0 + tx_1$

Derivative of this path —> target conditional vector field $u_t(x_t \mid x_1, x_0) = \frac{d}{dt}x_t = x_1 - x_0$

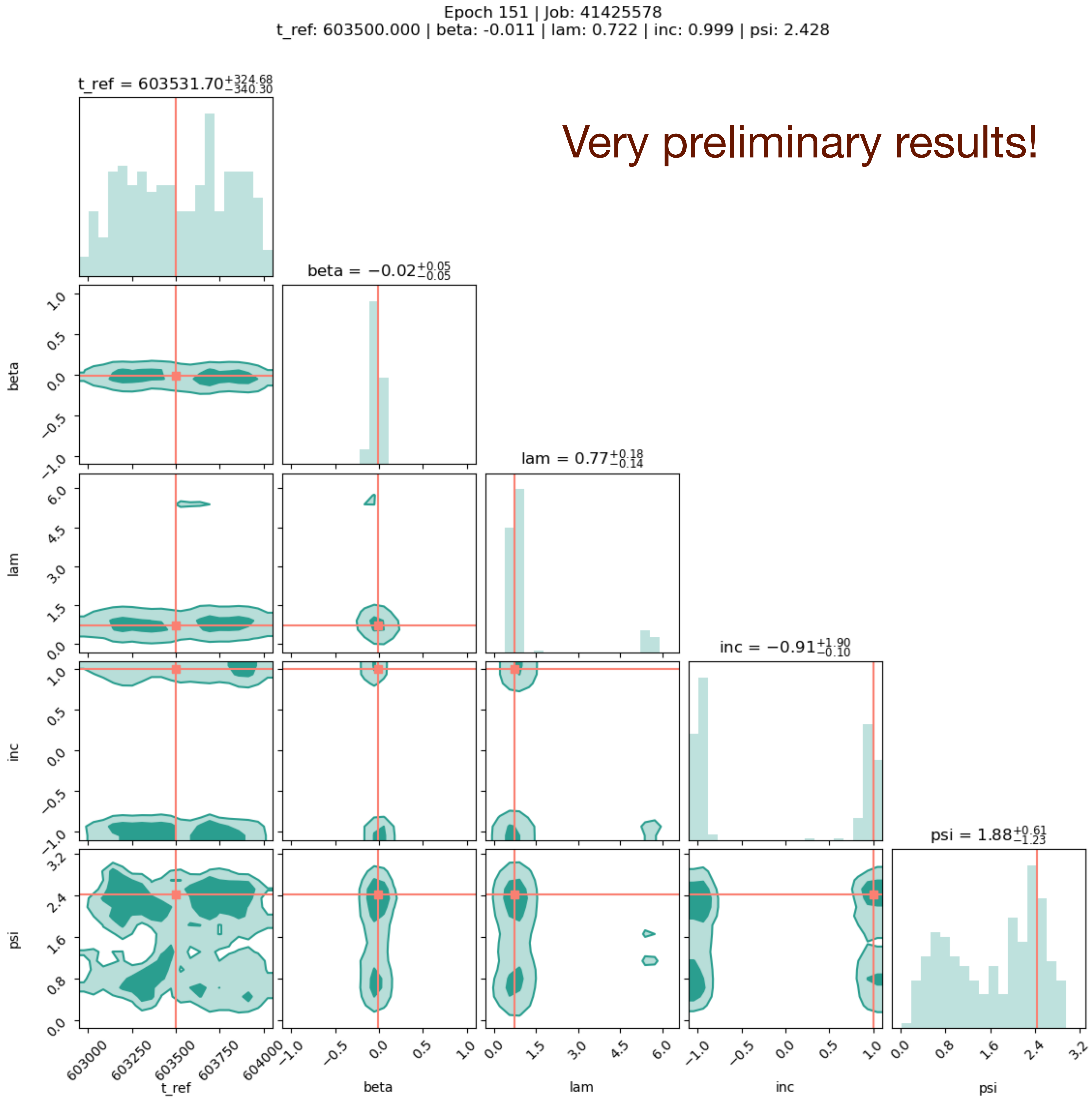
Thanks to the CFM theorem, if we train the network to match the *conditional* vector fields, it implicitly learns the true global vector field!

Parameter estimation

Time series

Benchmarking against **standard Bayesian inference methods** (MCMC or Nested Sampling) is **necessary** to establish the posteriors that we expect.

Very preliminary results!



Trying to recover phase information

CWT

Real and Imaginary part