Cosmological inference using gravitational waves and normalising flows

Federico Stachurski, **Martin Hendry & Chris Messenger**

University of Glasgow

Rencontres de Moriond 2025

PHYSICAL REVIEW D **109,** 123547 (2024)

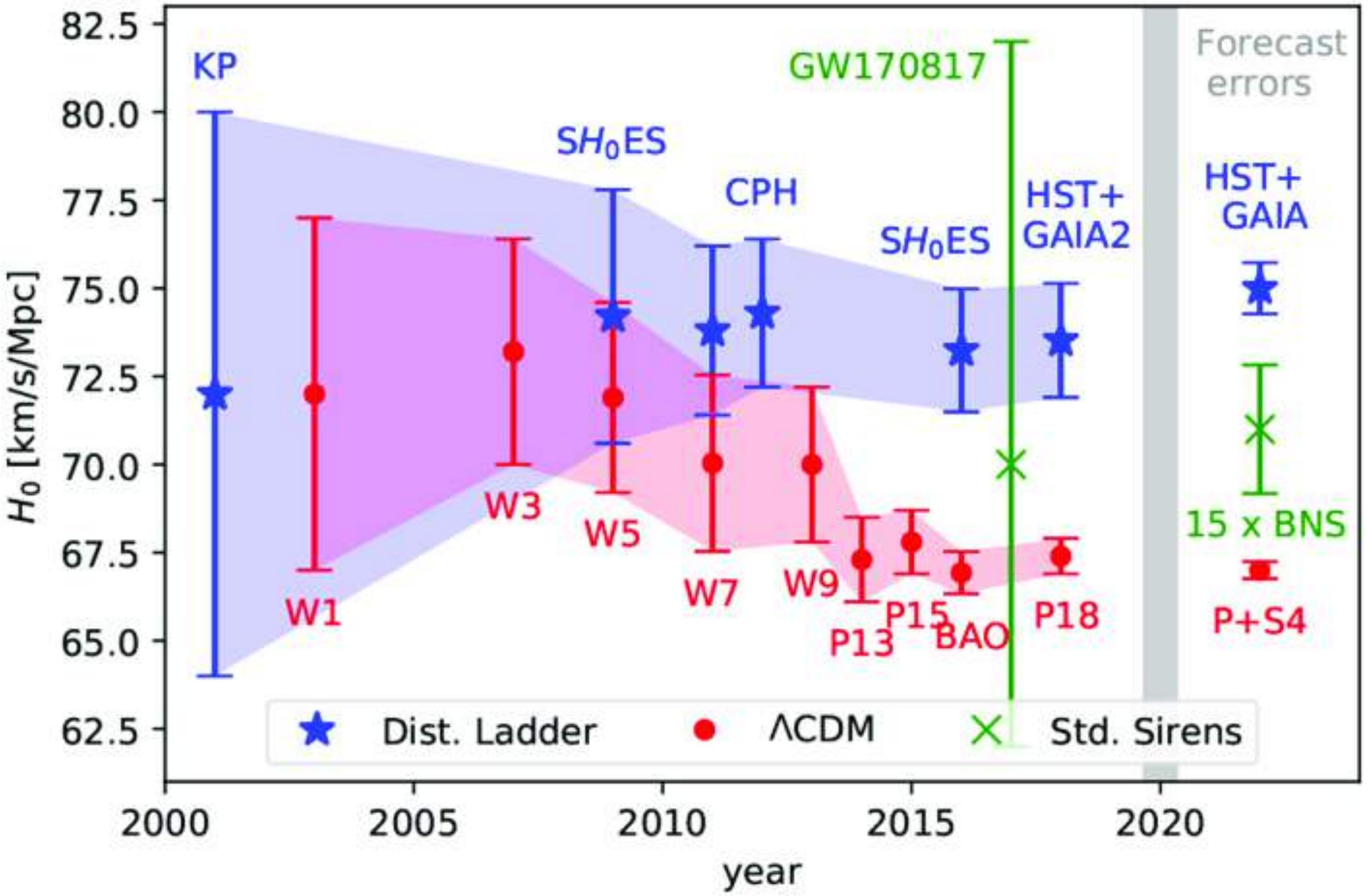
Cosmological inference using gravitational waves and normalizing flows

Federico Stachurski,[*] Christopher Messenger, and Martin Hendry

SUPA, School of Physics and Astronomy, University of Glasgow, Glasgow, United Kingdom

Motivation

Ezquiaga & Zumalacárregui 2018
Adapted from Beaton et al., 2016; Freedman, 2017



- In the long term - Independent local universe cosmological inference
- In the medium term - Hubble tension
- In the short term - Computational speed and flexibility

Gray *et al*, MNRAS 512, 2022
LVK Collaboration, ApJ 949, 2023

- Currently computationally limited - our baseline comparison is with gwcosmo
- Machine Learning (ML) can help hence we developed Cosmoflow

Stachurski *et al*, PRD 109, 2024

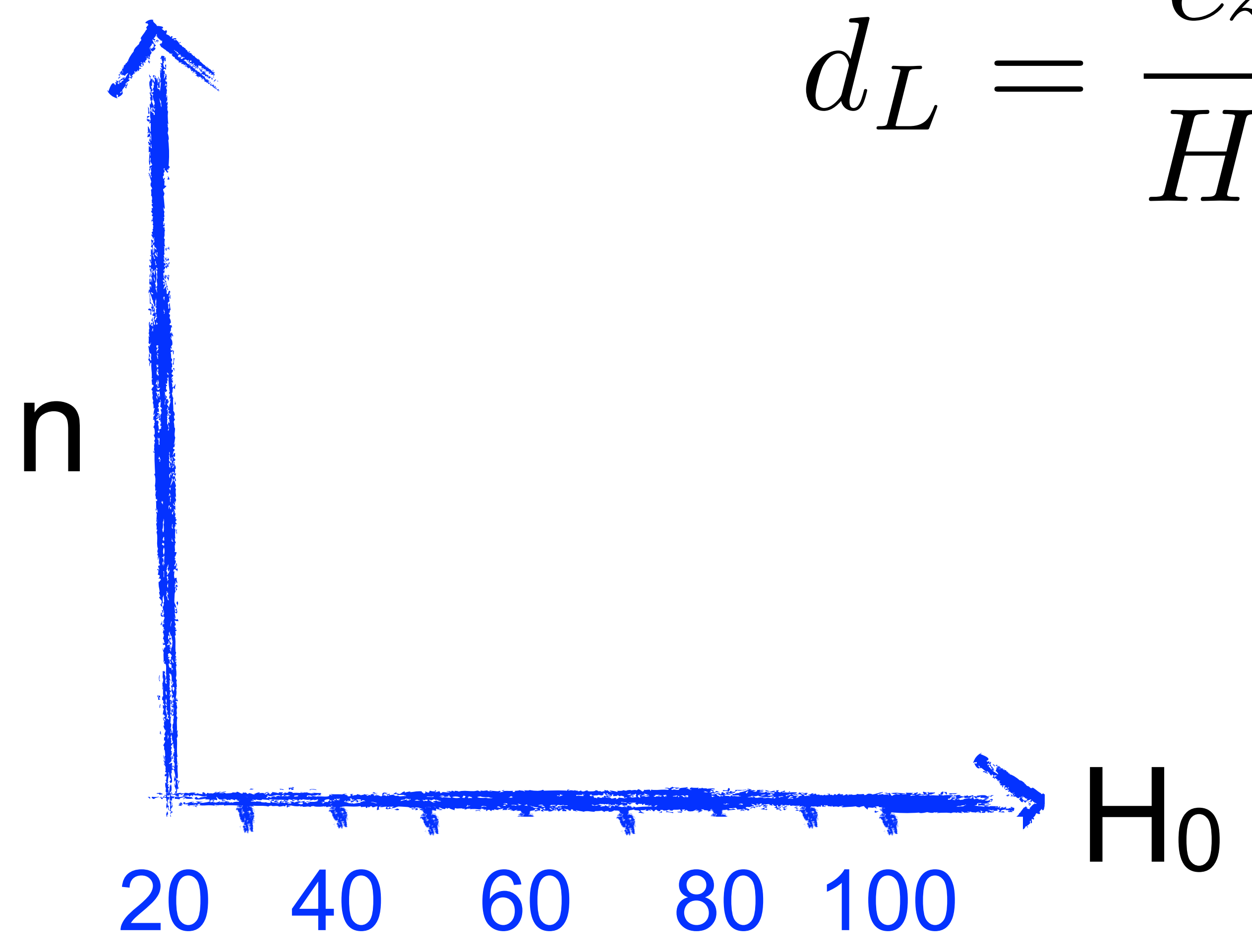
- We are not the only ones to try

Leyde et al, PRD 109, 2024

Dark Sirens

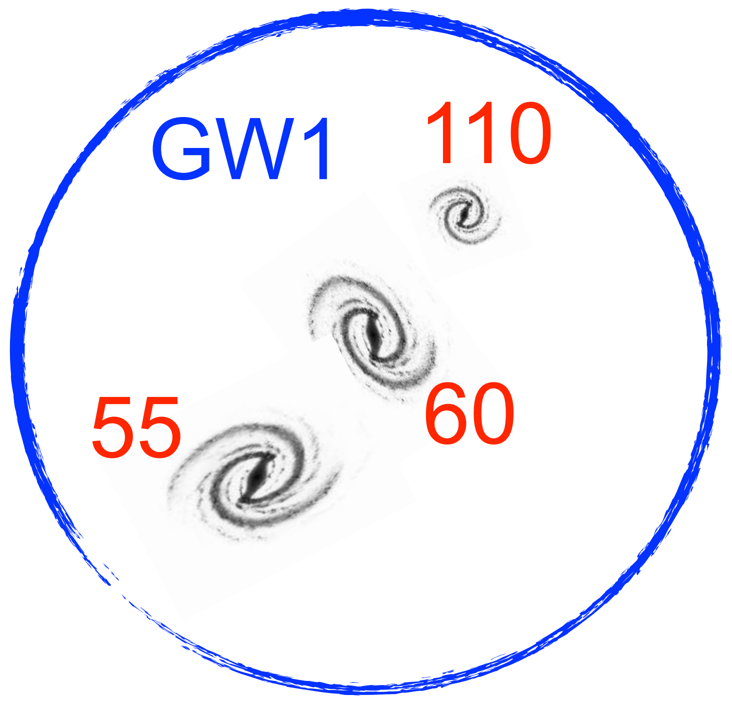
- Compact binary coalescence (CBC) events without EM transient counterparts
- The GW event provides information on the luminosity distance
- Galaxy catalogues represent the EM information on the redshift
- Average over galaxies within the GW error region
- Simple? What about selection effects?

$$d_L = \frac{cz}{H_0}$$



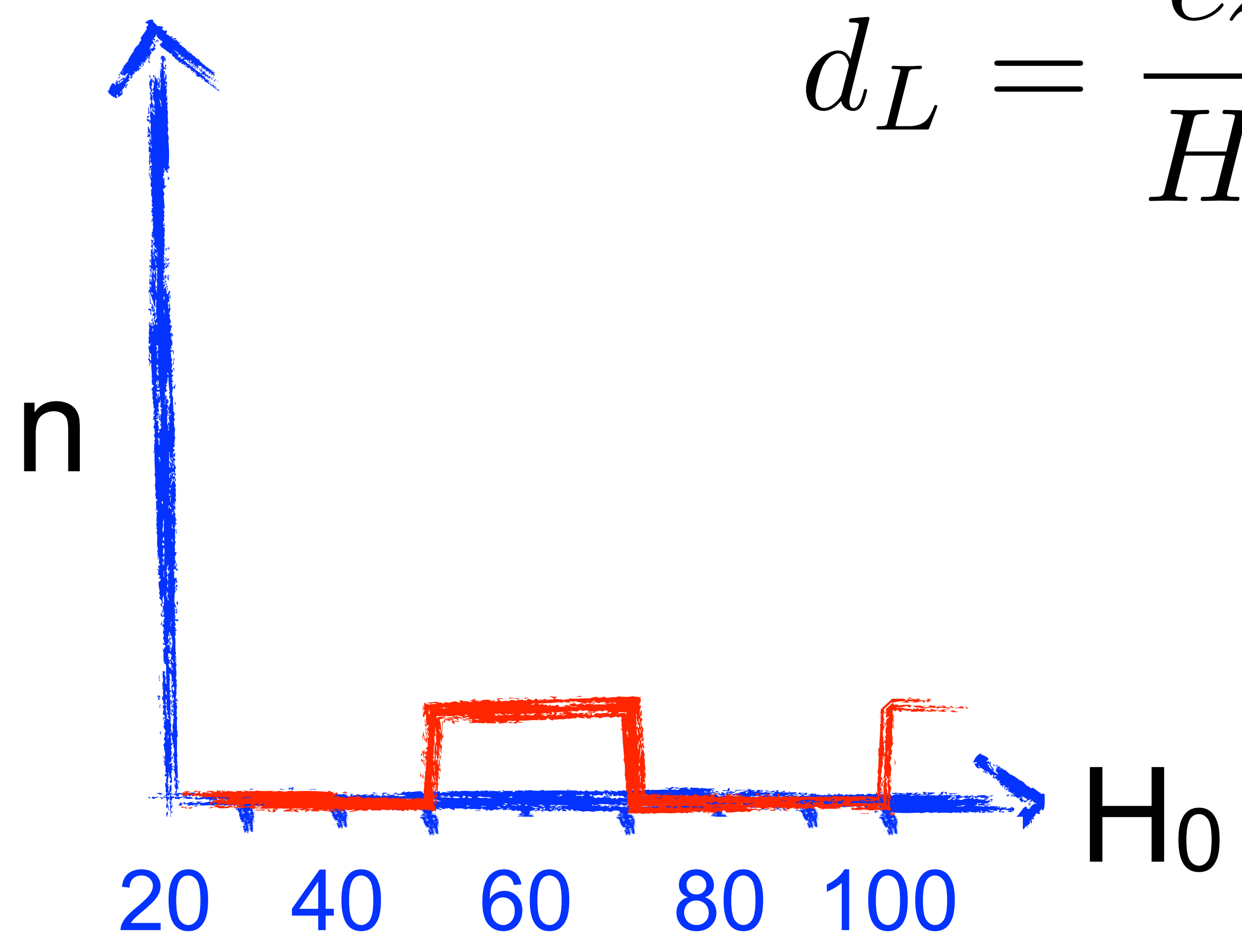
Dark Sirens

Schutz, Nature 323, 1986



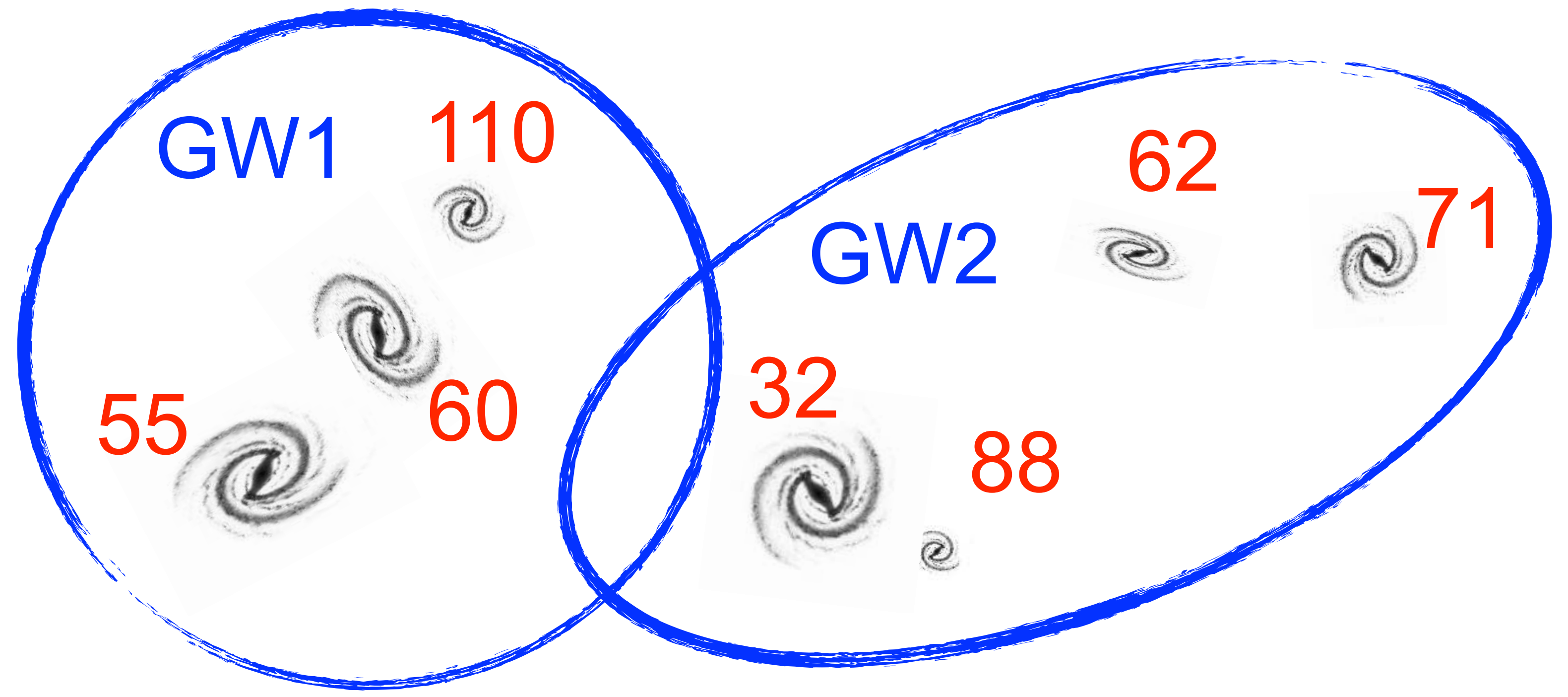
- Compact binary coalescence (CBC) events without EM transient counterparts
- The GW event provides information on the luminosity distance
- Galaxy catalogues represent the EM information on the redshift
- Average over galaxies within the GW error region
- Simple? What about selection effects?

$$d_L = \frac{cz}{H_0}$$



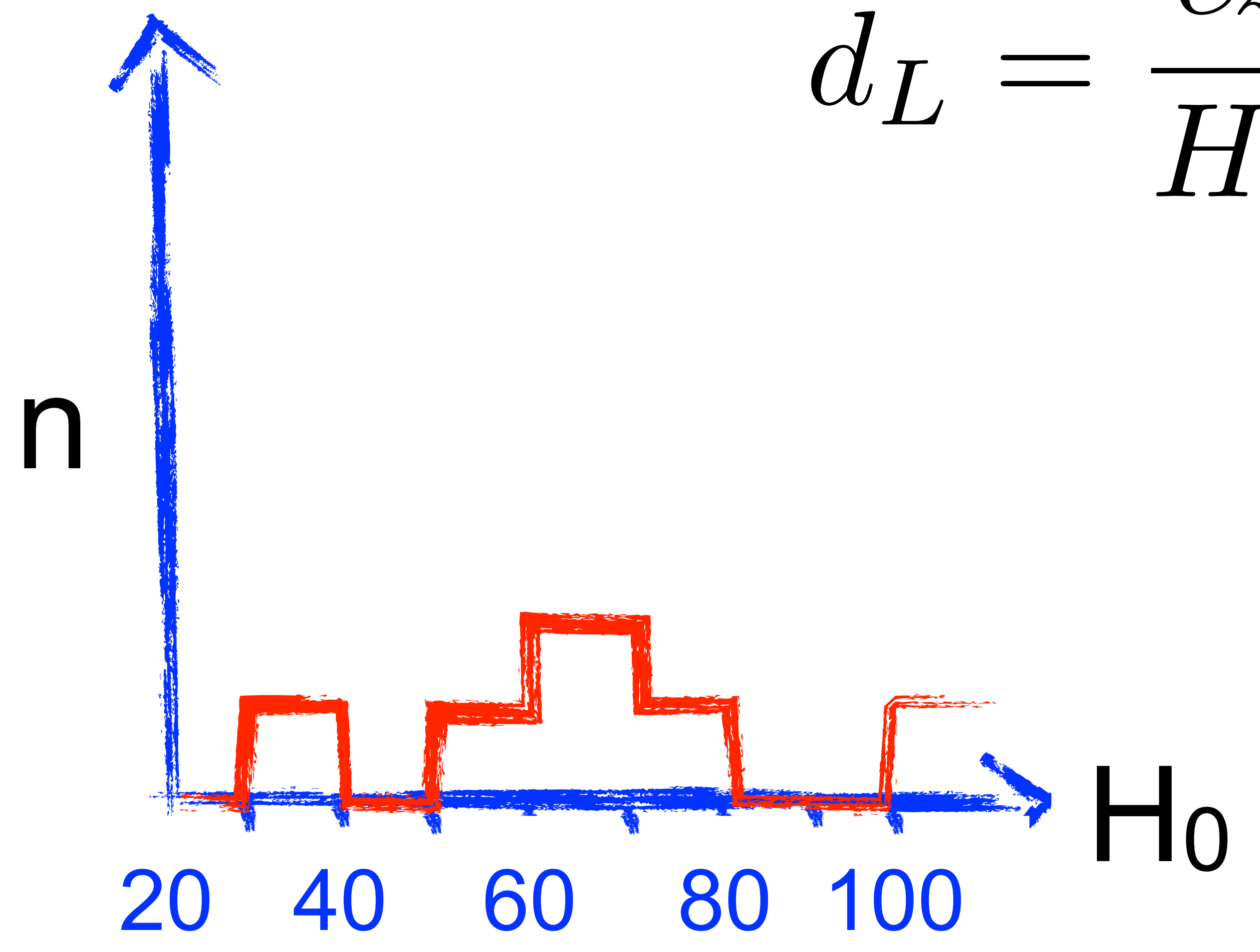
Dark Sirens

Schutz, Nature 323, 1986



- Compact binary coalescence (CBC) events without EM transient counterparts
- The GW event provides information on the luminosity distance
- Galaxy catalogues represent the EM information on the redshift
- Average over galaxies within the GW error region
- Simple? What about selection effects?

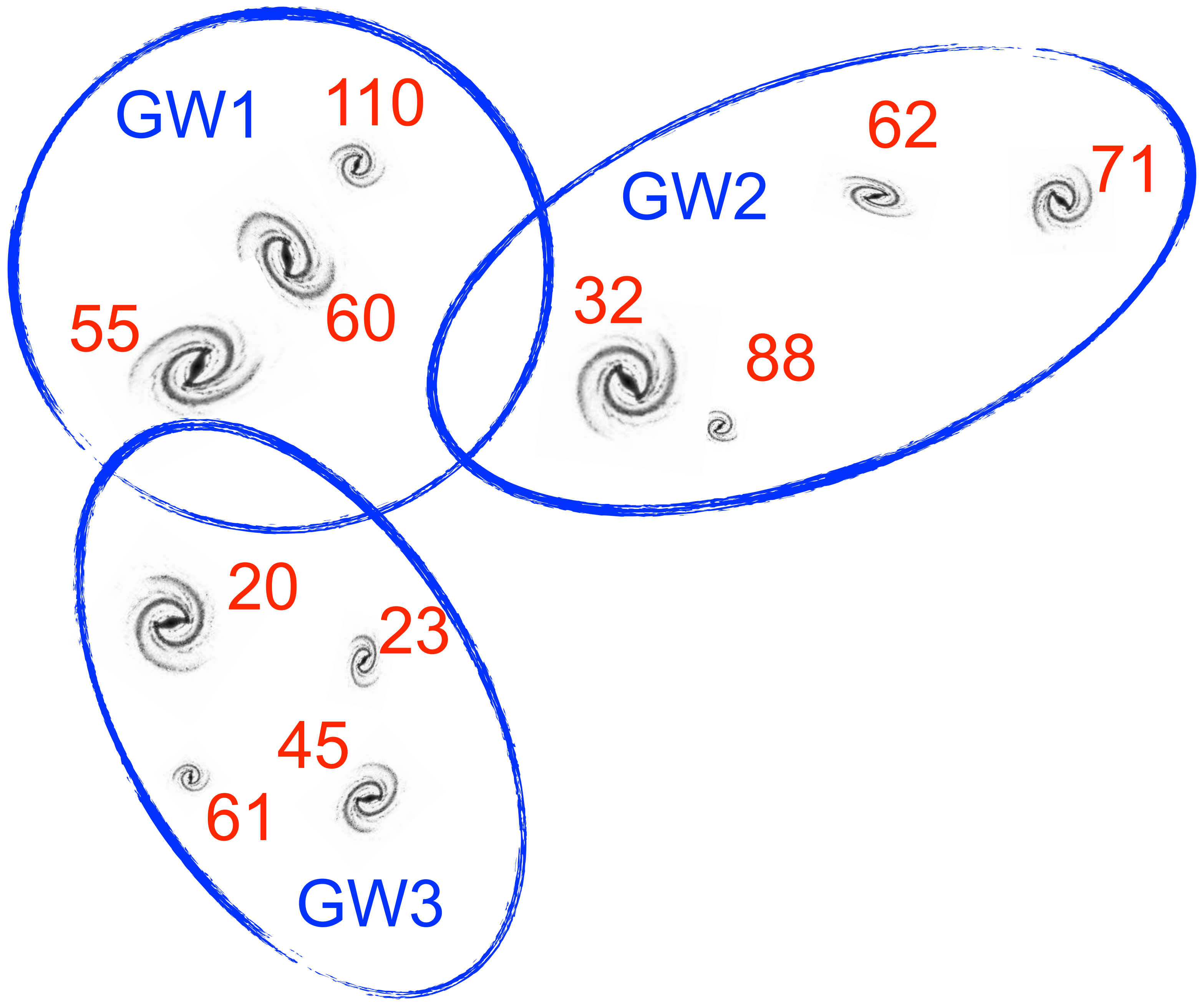
$$d_L = \frac{cz}{H_0}$$



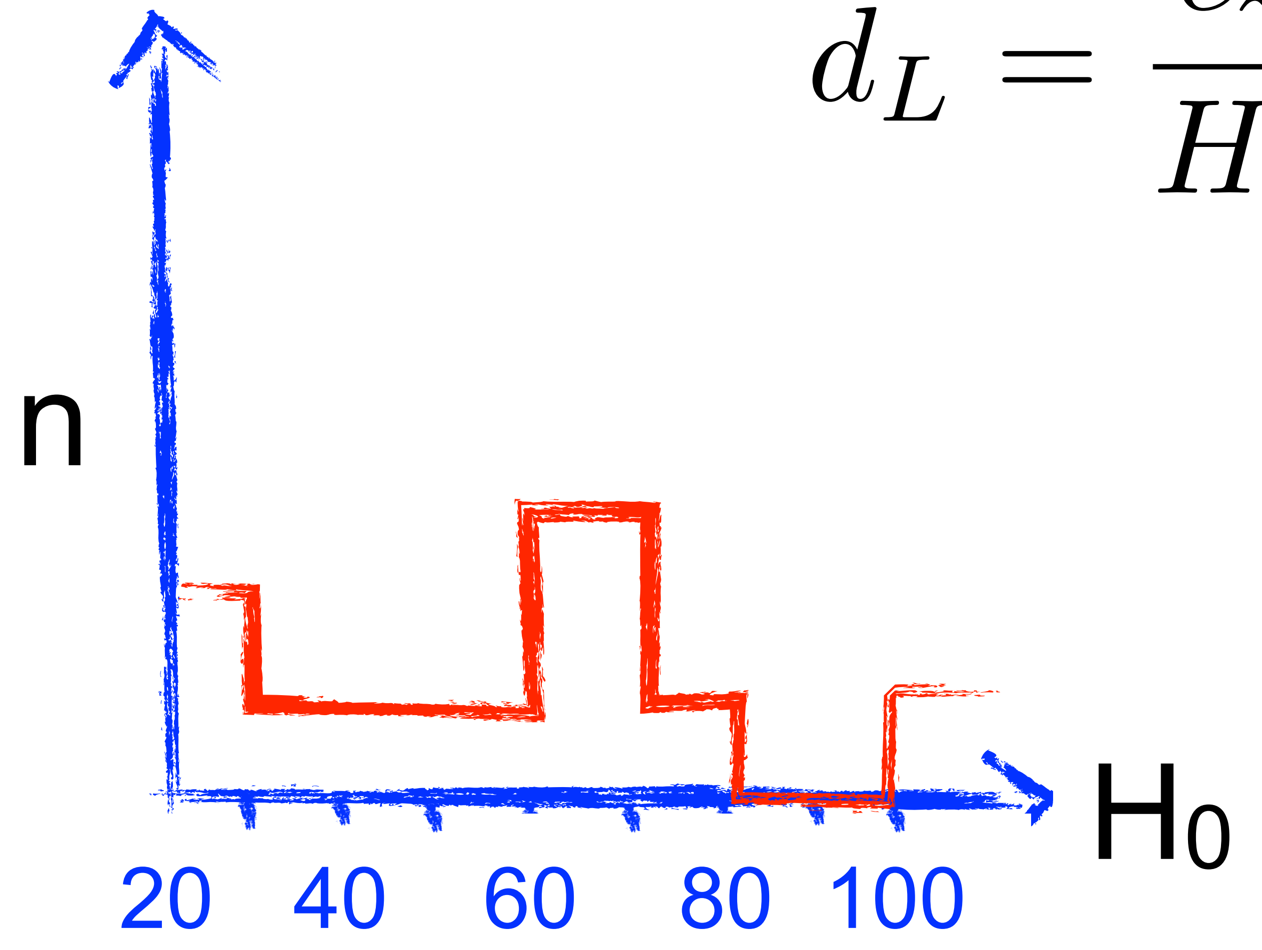
Schutz, Nature 323, 1986

Dark Sirens

- Compact binary coalescence (CBC) events without EM transient counterparts
- The GW event provides information on the luminosity distance
- Galaxy catalogues represent the EM information on the redshift
- Average over galaxies within the GW error region
- Simple? What about selection effects?



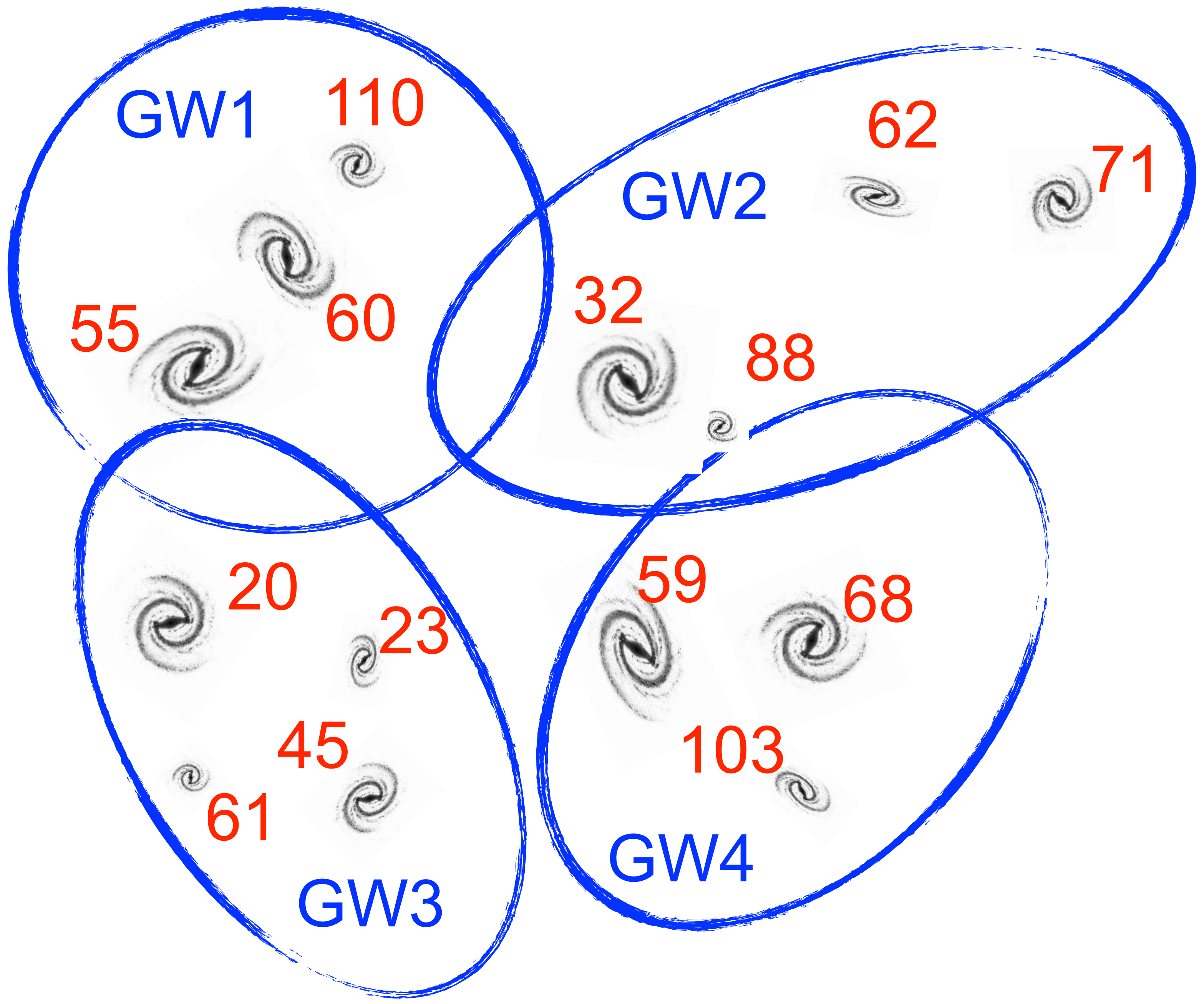
$$d_L = \frac{cz}{H_0}$$



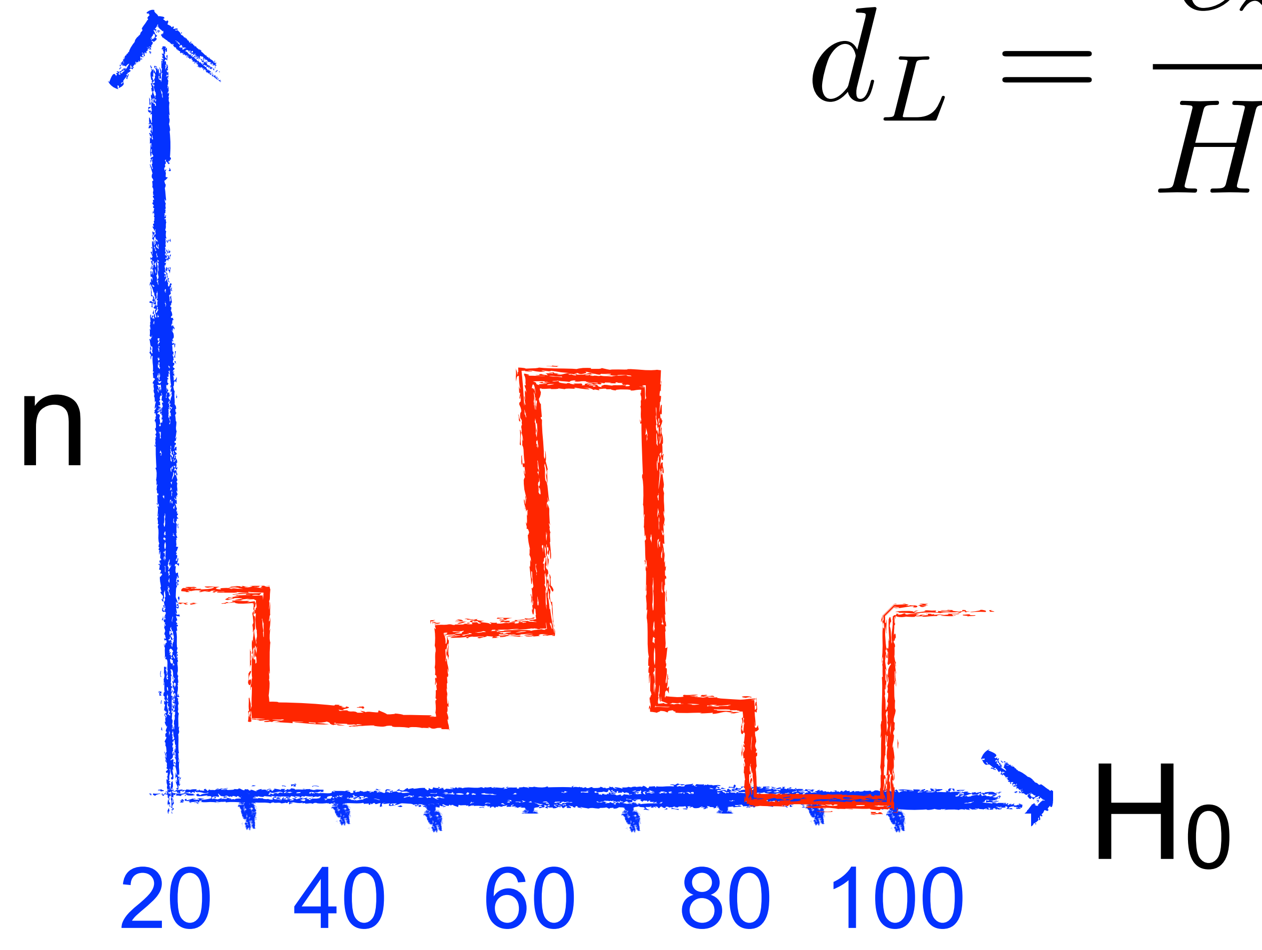
Dark Sirens

Schutz, Nature 323, 1986

- Compact binary coalescence (CBC) events without EM transient counterparts
- The GW event provides information on the luminosity distance
- Galaxy catalogues represent the EM information on the redshift
- Average over galaxies within the GW error region
- Simple? What about selection effects?



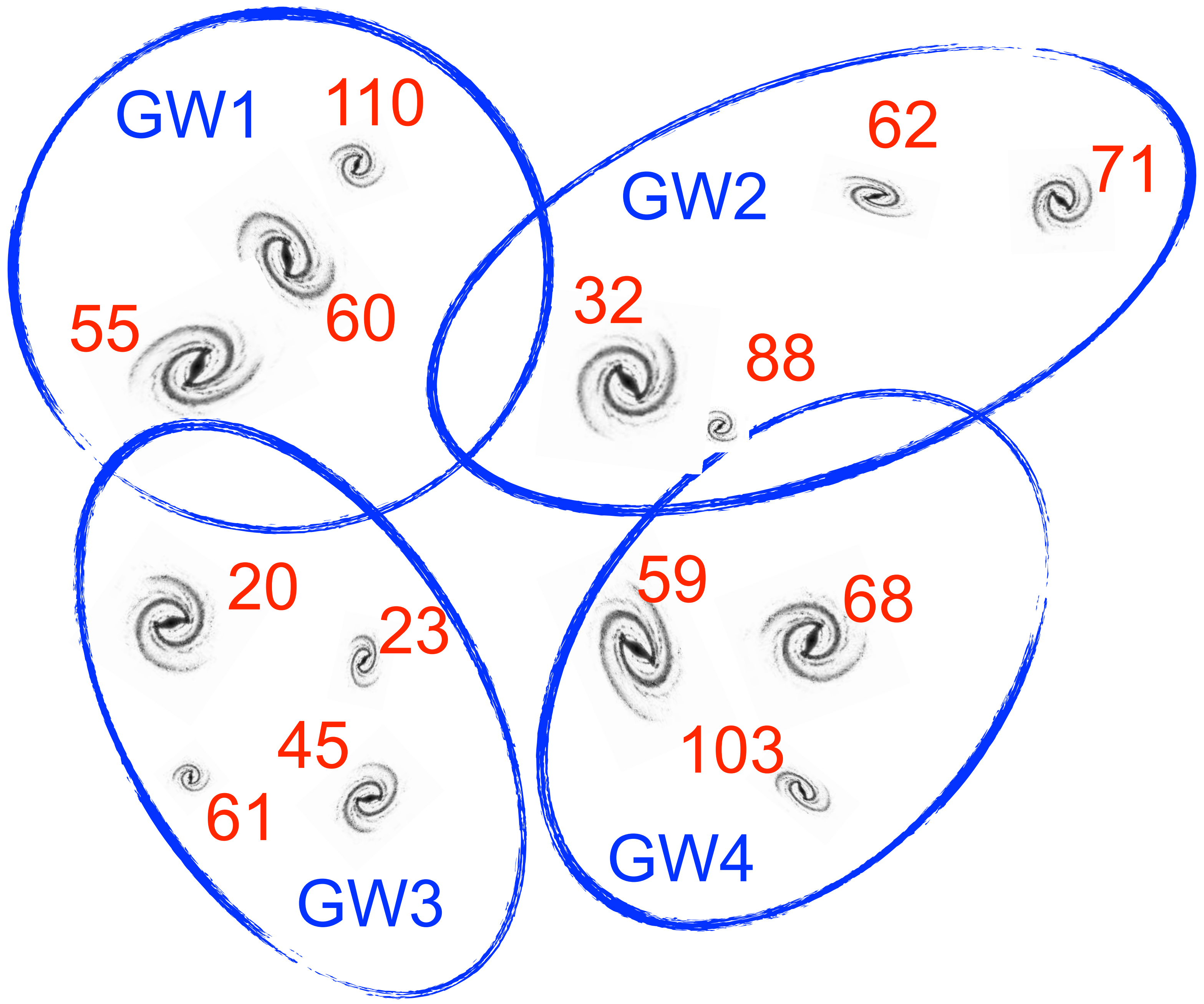
$$d_L = \frac{cz}{H_0}$$



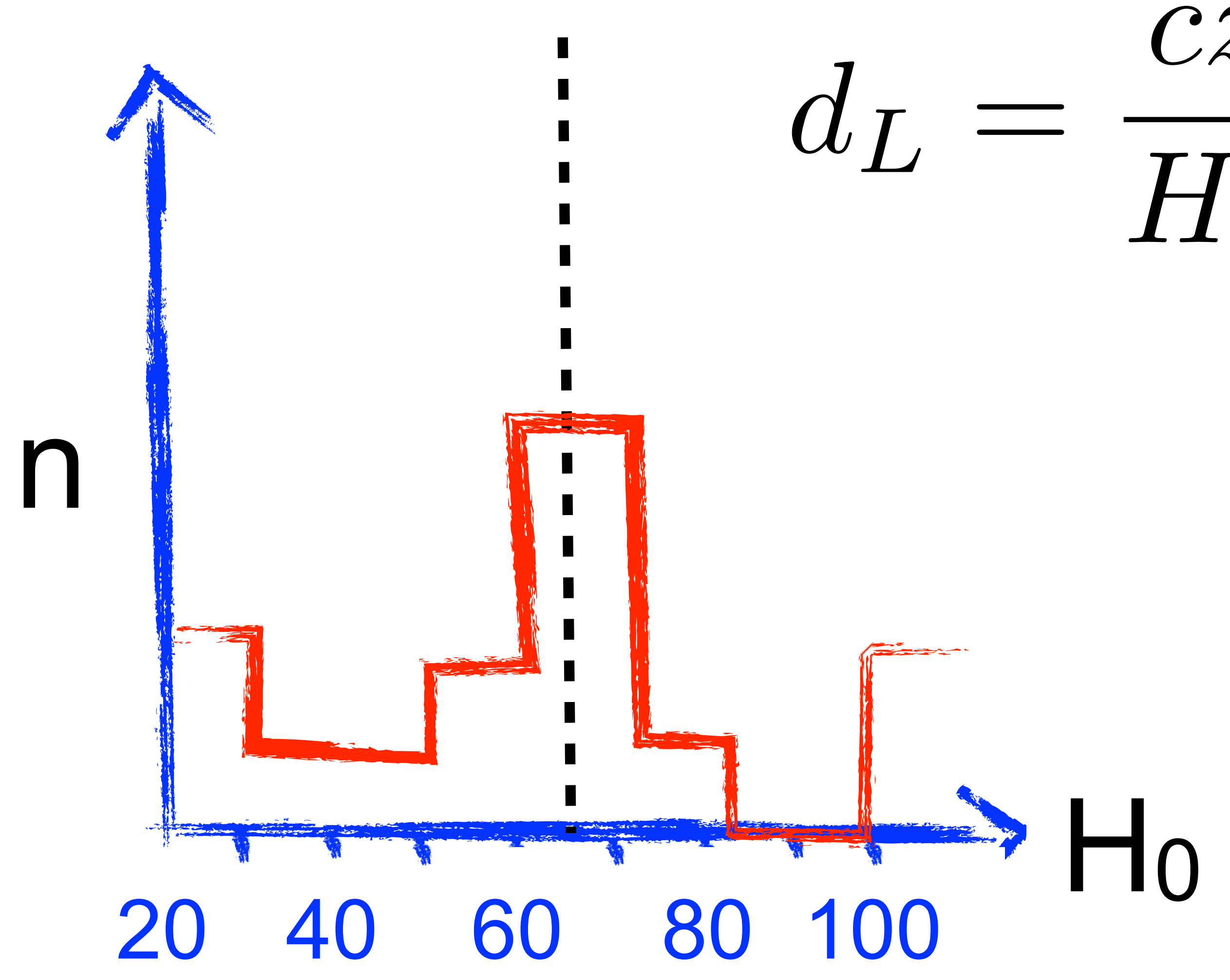
Dark Sirens

Schutz, Nature 323, 1986

- Compact binary coalescence (CBC) events without EM transient counterparts
- The GW event provides information on the luminosity distance
- Galaxy catalogues represent the EM information on the redshift
- Average over galaxies within the GW error region
- Simple? What about selection effects?



$$d_L = \frac{cz}{H_0}$$



The strategy

Modelled using the normalising flow

The H_0 posterior combined over events

Posterior samples from the i'th event

$$p(H_0 \mid \vec{h}, \vec{D}, I) \propto p(H_0 \mid I) \prod_{i}^{N_{\mathrm{det}}} \left\langle \frac{p(\vec{\theta}_i \mid H_0, D_i, I)}{p(D_i \mid \vec{\theta}_i, I)\, p(\vec{\theta}_i \mid H_0', I)} \right\rangle_{\theta_i \sim p(\theta_i | h_i, I)}$$

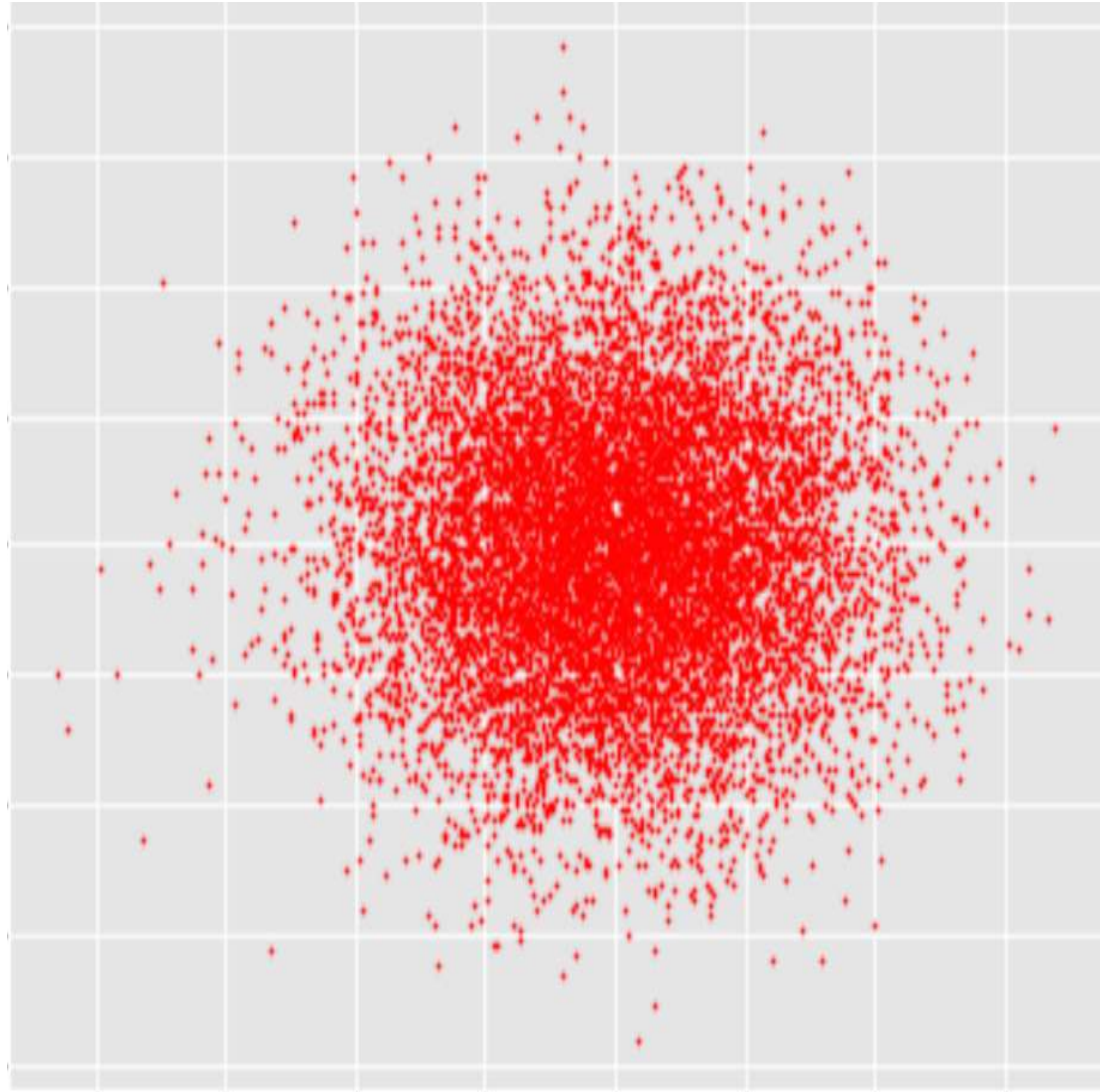
The probability of detection (accounting for selection effects)

The prior used to obtain the posterior samples

- We condition on detection (D) to minimise uninformative training data

Footnote: We use H_0 here for clarity but this can be replaced with a vector of cosmological and population parameters Ω

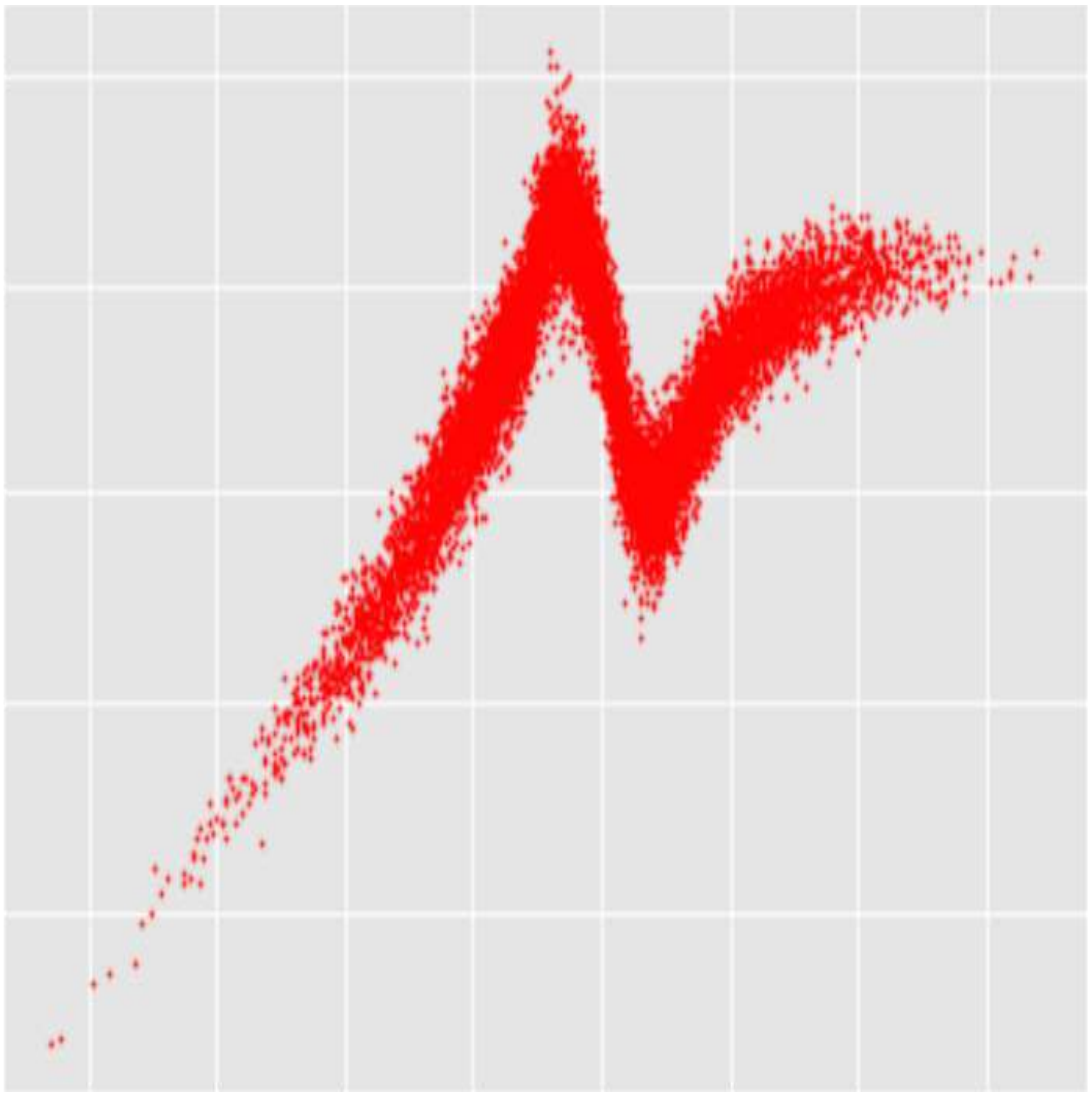
What is a Normalising Flow?



- A generative ML model that learns a conditional distribution
- $p_z(z)$ is a Normal distribution in a latent space
- $p_x(x \mid \omega)$ is the conditional data distribution
- $x = f(z, \phi \mid \omega)$ is a bijective function

$$\text{Loss}(\phi) = \text{KL}(p_x(x) \mid q_x(x)) = \text{const.} - \mathbb{E}_{p_x}\left[\log p_z(f^{-1}(x, \phi \mid \omega)) + \log\left(\det\left|\frac{df^{-1}(x, \phi \mid \omega)}{px}\right|\right)\right]$$

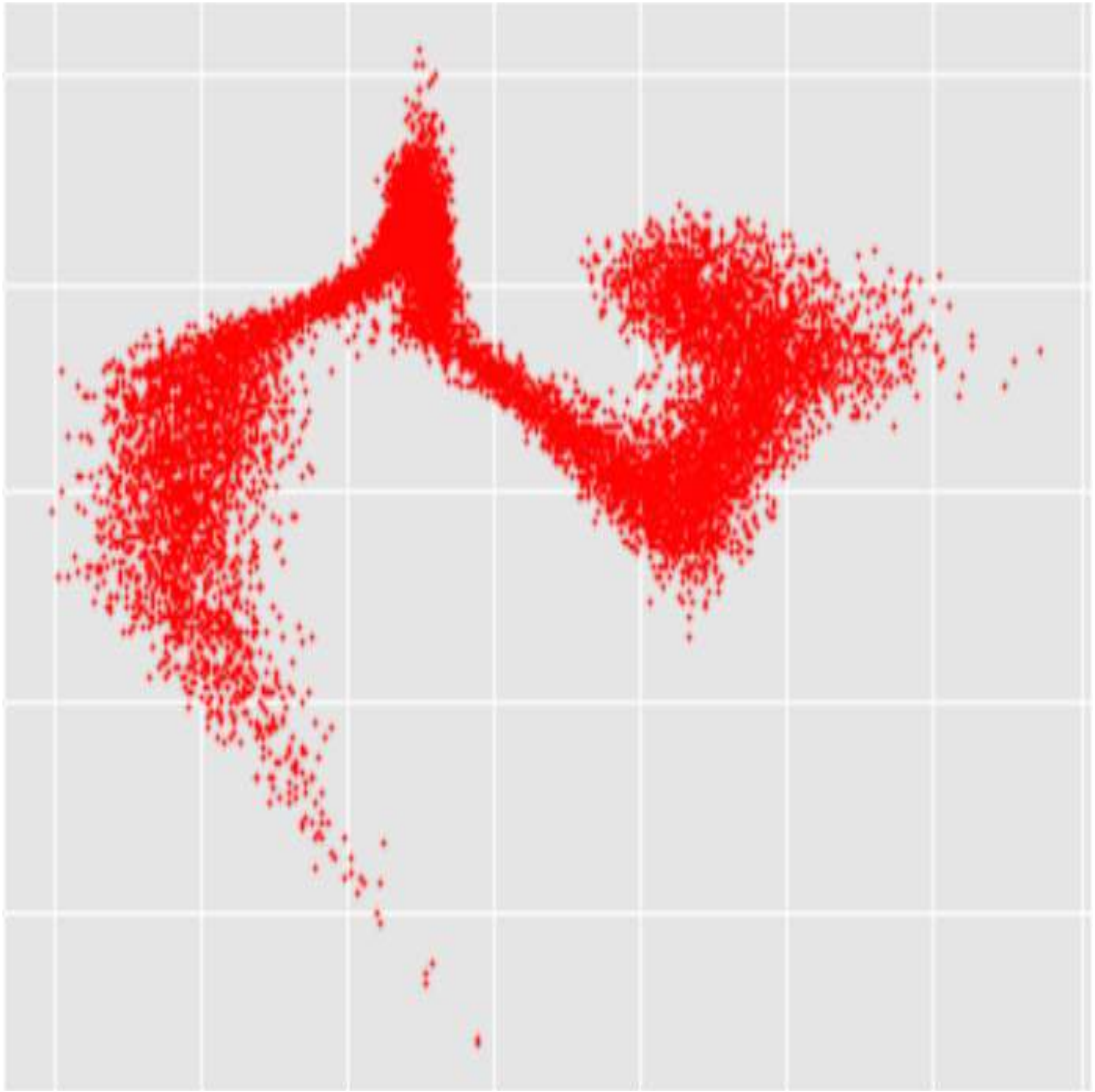
What is a Normalising Flow?



- A generative ML model that learns a conditional distribution
- $p_z(z)$ is a Normal distribution in a latent space
- $p_x(x \mid \omega)$ is the conditional data distribution
- $x = f(z, \phi \mid \omega)$ is a bijective function

$$\text{Loss}(\phi) = \text{KL}(p_x(x) \mid q_x(x)) = \text{const.} - \mathbb{E}_{p_x}\left[\log p_z(f^{-1}(x, \phi \mid \omega)) + \log\left(\det\left|\frac{df^{-1}(x, \phi \mid \omega)}{px}\right|\right)\right]$$

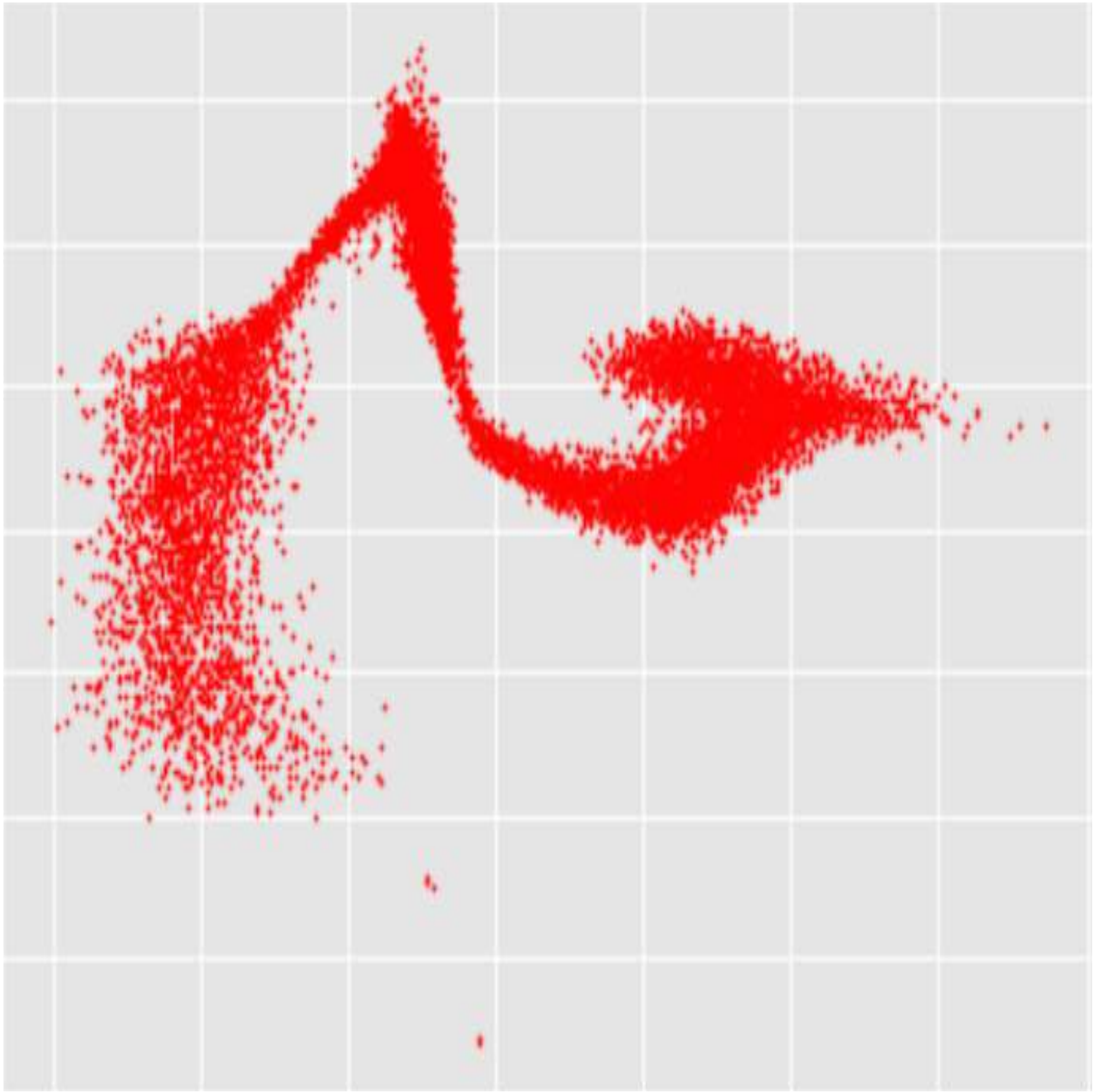
What is a Normalising Flow?



- A generative ML model that learns a conditional distribution
- $p_z(z)$ is a Normal distribution in a latent space
- $p_x(x \mid \omega)$ is the conditional data distribution
- $x = f(z, \phi \mid \omega)$ is a bijective function

$$\text{Loss}(\phi) = \text{KL}(p_x(x) \mid q_x(x)) = \text{const.} - \mathbb{E}_{p_x}\left[\log p_z(f^{-1}(x, \phi \mid \omega)) + \log\left(\det\left|\frac{df^{-1}(x, \phi \mid \omega)}{px}\right|\right)\right]$$

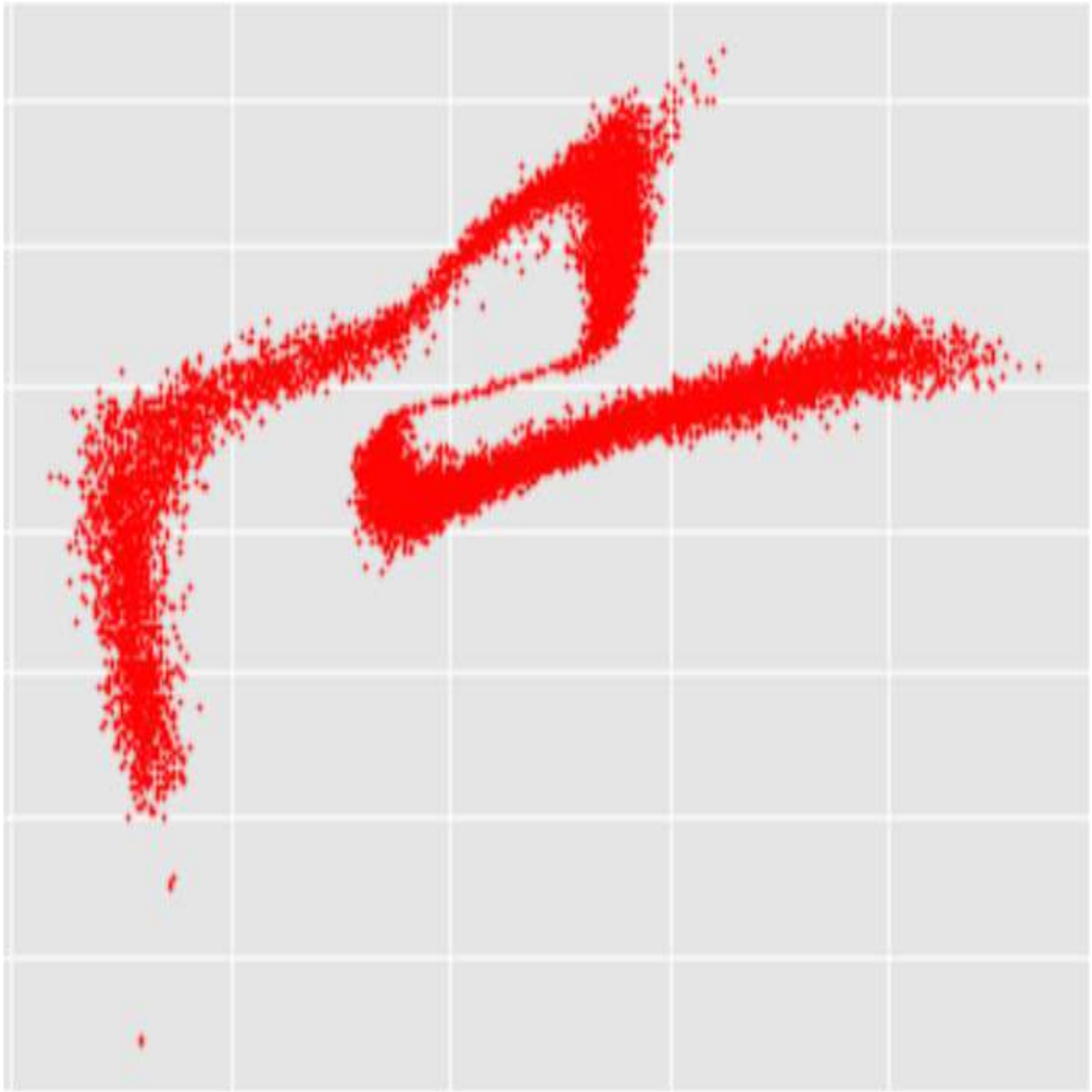
What is a Normalising Flow?



- A generative ML model that learns a conditional distribution
- $p_z(z)$ is a Normal distribution in a latent space
- $p_x(x \,|\, \omega)$ is the conditional data distribution
- $x = f(z, \phi \,|\, \omega)$ is a bijective function

$$\text{Loss}(\phi) = \text{KL}(p_x(x) \,|\, q_x(x)) = \text{const.} - \mathbb{E}_{p_x}\left[\log p_z(f^{-1}(x, \phi \,|\, \omega)) + \log\left(\det\left|\frac{df^{-1}(x, \phi \,|\, \omega)}{px}\right|\right)\right]$$

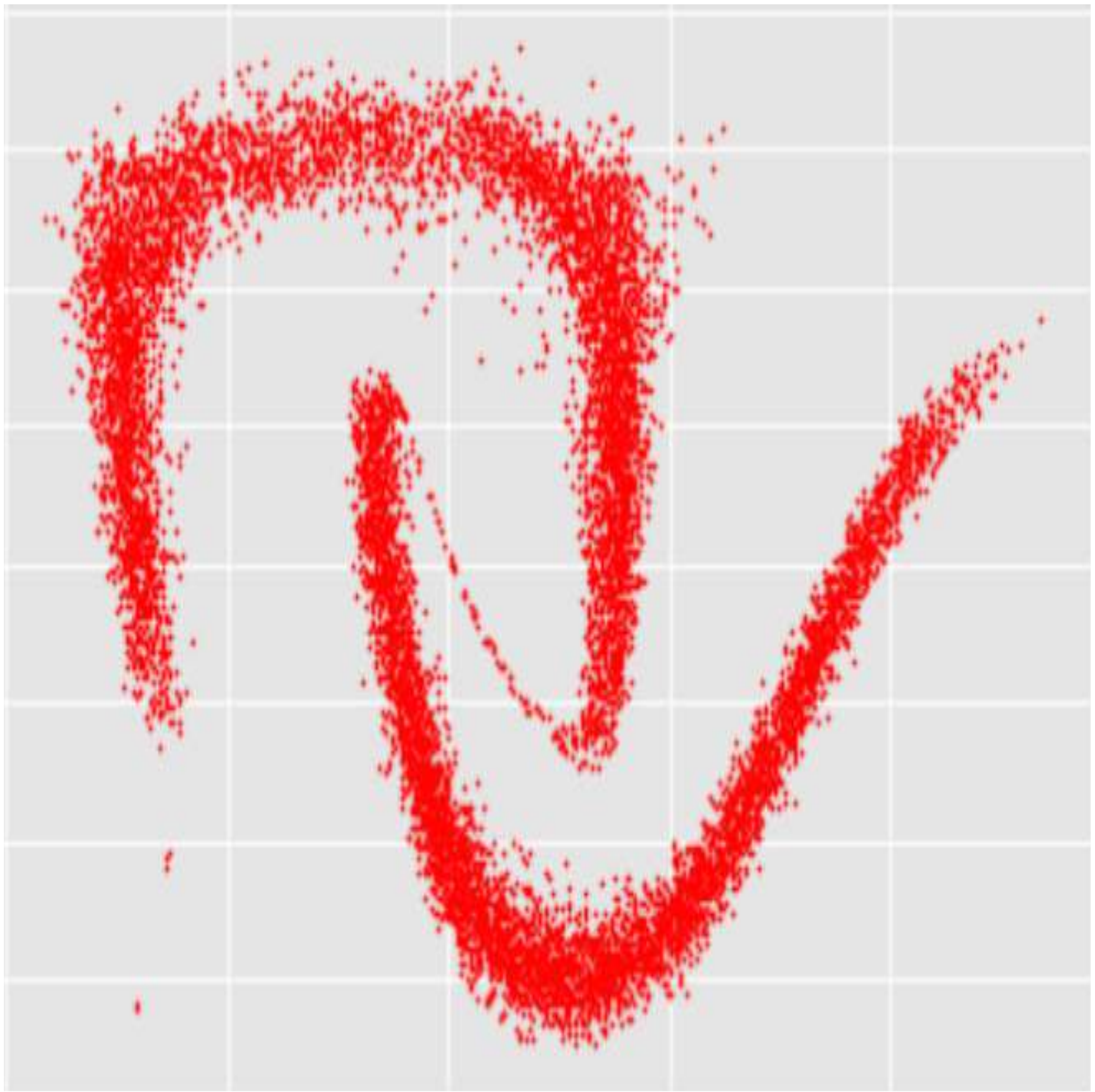
What is a Normalising Flow?



- A generative ML model that learns a conditional distribution
- $p_z(z)$ is a Normal distribution in a latent space
- $p_x(x \mid \omega)$ is the conditional data distribution
- $x = f(z, \phi \mid \omega)$ is a bijective function

$$\text{Loss}(\phi) = \text{KL}(p_x(x) \mid q_x(x)) = \text{const.} - \mathbb{E}_{p_x}\left[\log p_z(f^{-1}(x, \phi \mid \omega)) + \log\left(\det\left|\frac{df^{-1}(x, \phi \mid \omega)}{px}\right|\right)\right]$$

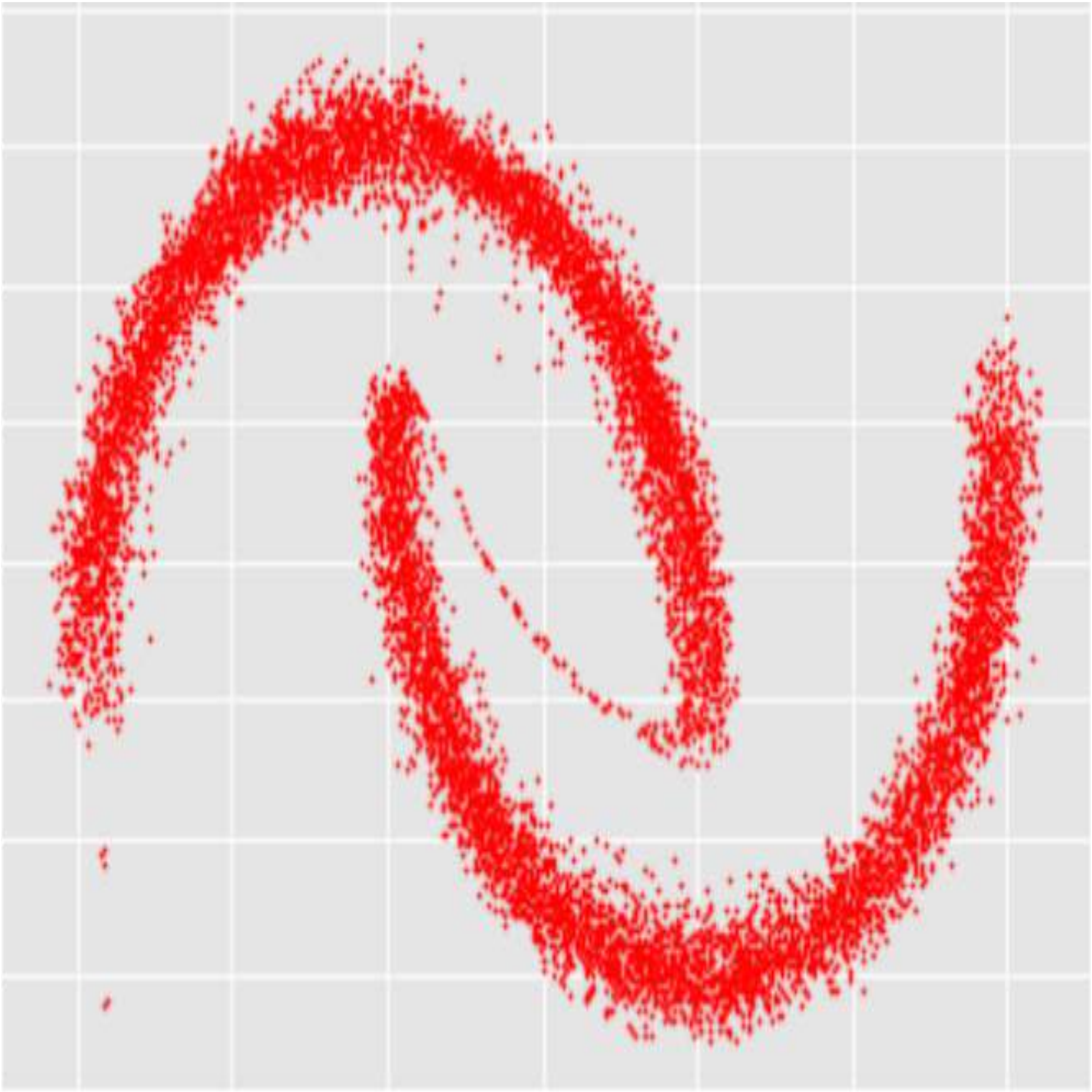
What is a Normalising Flow?



- A generative ML model that learns a conditional distribution
- $p_z(z)$ is a Normal distribution in a latent space
- $p_x(x \mid \omega)$ is the conditional data distribution
- $x = f(z, \phi \mid \omega)$ is a bijective function

$$\text{Loss}(\phi) = \text{KL}(p_x(x) \mid q_x(x)) = \text{const.} - \mathbb{E}_{p_x}\left[\log p_z(f^{-1}(x, \phi \mid \omega)) + \log\left(\det\left|\frac{df^{-1}(x, \phi \mid \omega)}{px}\right|\right)\right]$$

What is a Normalising Flow?



- A generative ML model that learns a conditional distribution
- $p_z(z)$ is a Normal distribution in a latent space
- $p_x(x \mid \omega)$ is the conditional data distribution
- $x = f(z, \phi \mid \omega)$ is a bijective function

$$\text{Loss}(\phi) = \text{KL}(p_x(x) \mid q_x(x)) = \text{const.} - \mathbb{E}_{p_x}\left[\log p_z(f^{-1}(x, \phi \mid \omega)) + \log\left(\det\left|\frac{df^{-1}(x, \phi \mid \omega)}{px}\right|\right)\right]$$

Data generation

- Our aim is to generate samples from $p(H_0, \vec{\theta} \,|\, D, I)$, so:
  - Sample from the H_0 prior
  - Sample a redshift from $\dfrac{p(z \,|\, H_0, I)}{(1+z)}$
  - Sample α, δ uniformly on the sky
  - Sample a Luminosity L from the weighted Schechter function $\propto L\, p(L \,|\, H_0)$
  - Determine if the galaxy "would" be inside our galaxy catalogue (GLADE+)

Dálya et al, MNRAS 514, 2022

Key points:

1. The Flow is learning the EM and detection conditioned GW prior

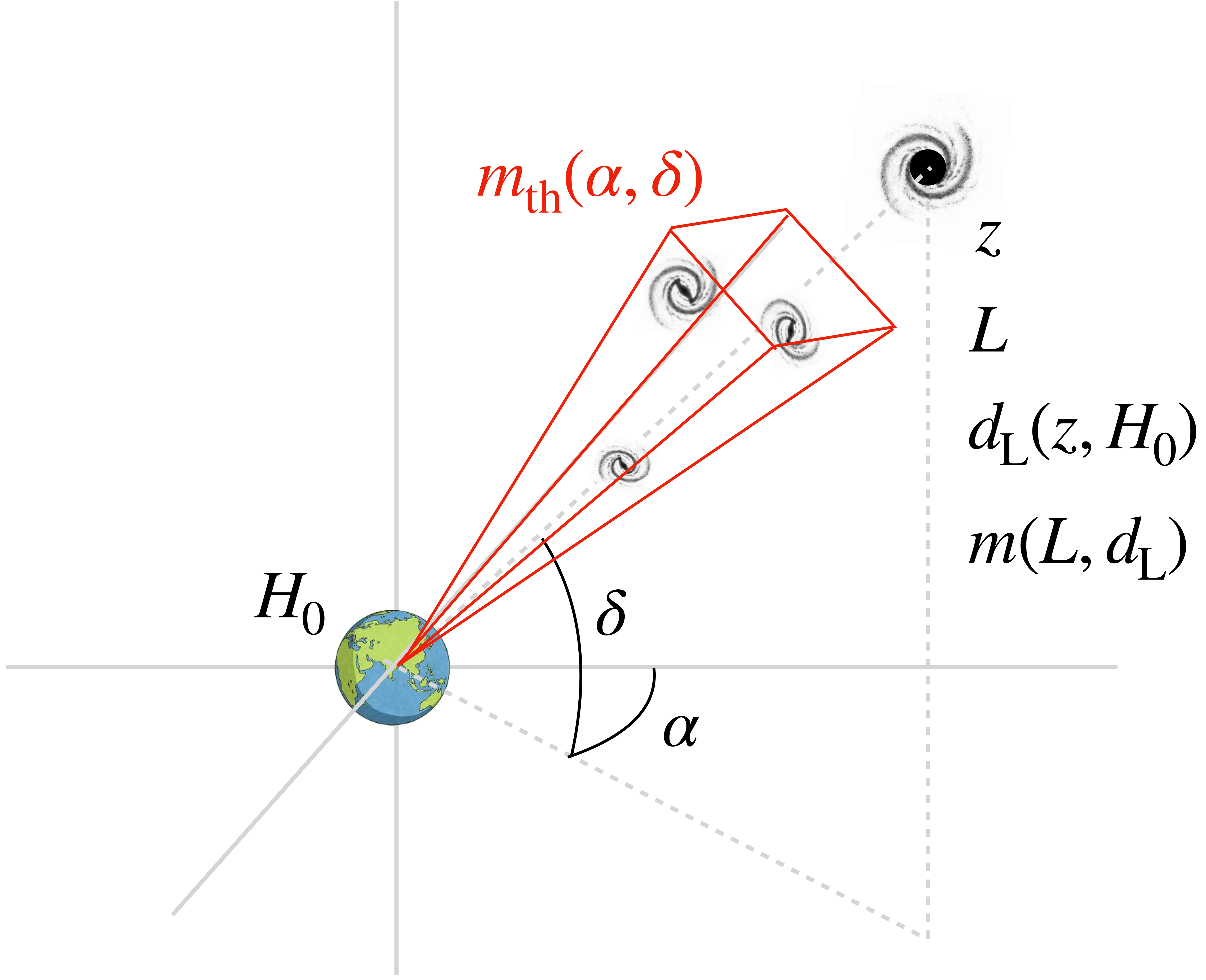
2. We generate GW events that come from the catalogue and beyond

3. Shifts difficulty to sampling and not modelling (very flexible)

4. Your results are only as good as your training data

Data generation

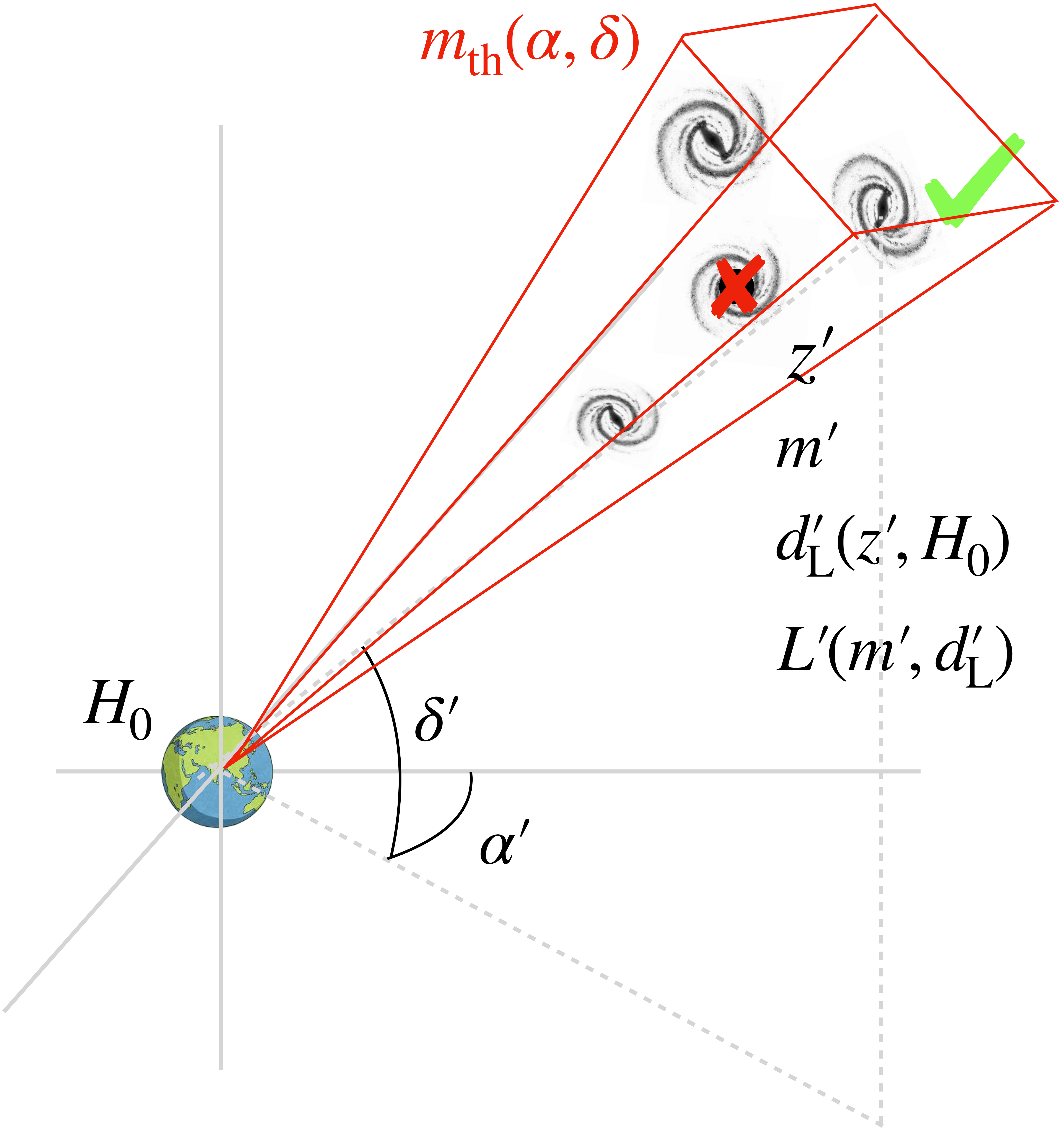
- Our aim is to generate samples from $p(H_0, \vec{\theta} \mid D, I)$, so:
  - Sample from the H_0 prior
  - Sample a redshift from $\dfrac{p(z \mid H_0, I)}{(1+z)}$
  - Sample α, δ uniformly on the sky
  - Sample a Luminosity L from the weighted Schechter function $\propto L\, p(L \mid H_0)$
  - Determine if the galaxy "would" be inside our galaxy catalogue (GLADE+)



Dálya et al, MNRAS 514, 2022

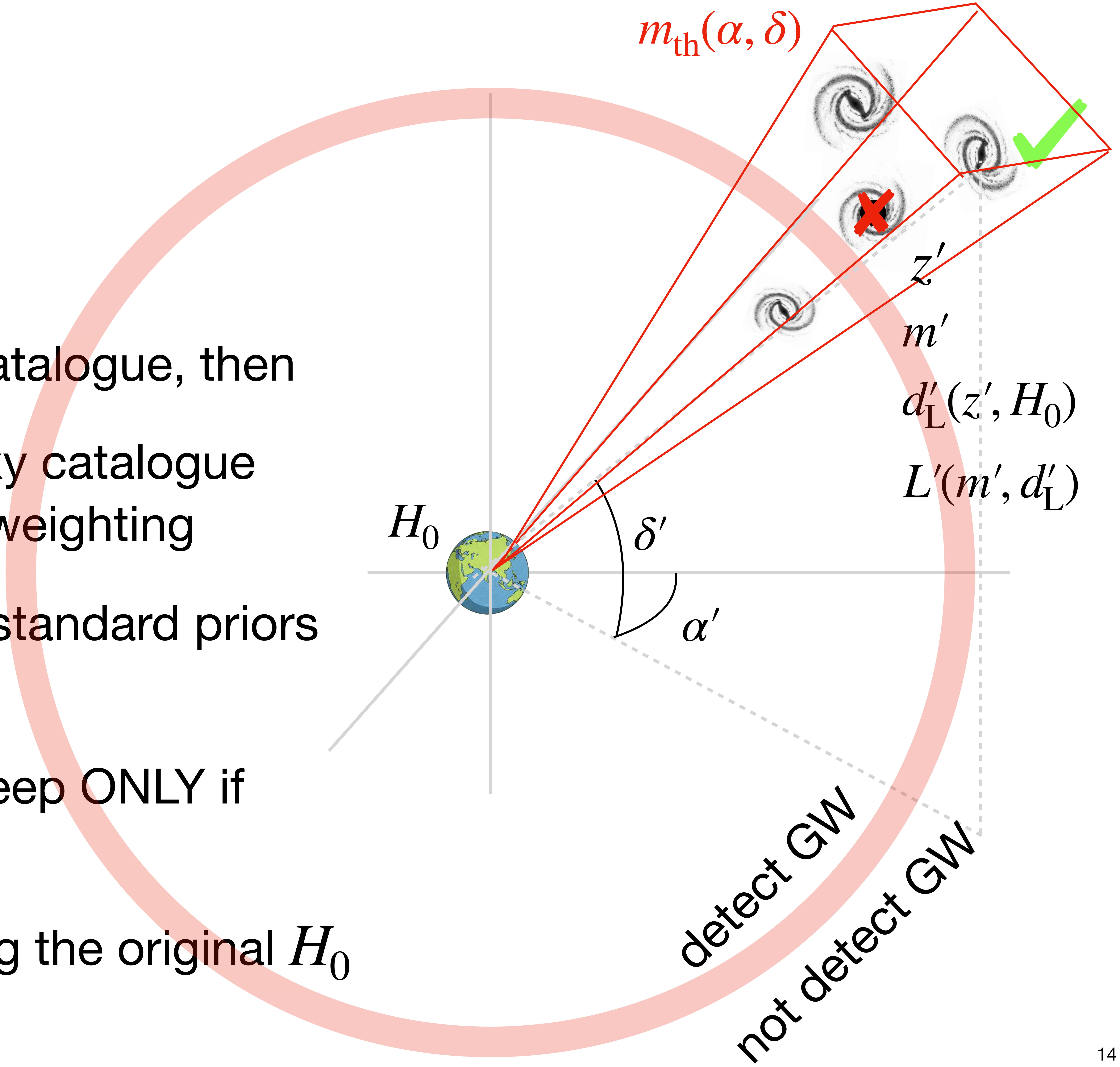
Data generation

- If it would HAVE been in the catalogue, then
  - Resample galaxy from galaxy catalogue pixel using L and $(1+z)^{-1}$weighting
- Sample GW parameters from standard priors (masses get redshifted)
- Sample from $p(D\,|\,\vec{\theta}, I)$ and keep ONLY if detected
- Otherwise start again - keeping the original H_0



Data generation

- If it would HAVE been in the catalogue, then
  - Resample galaxy from galaxy catalogue pixel using L and $(1+z)^{-1}$weighting
- Sample GW parameters from standard priors (masses get redshifted)
- Sample from $p(D\,|\,\vec{\theta}, I)$ and keep ONLY if detected
- Otherwise start again - keeping the original H_0

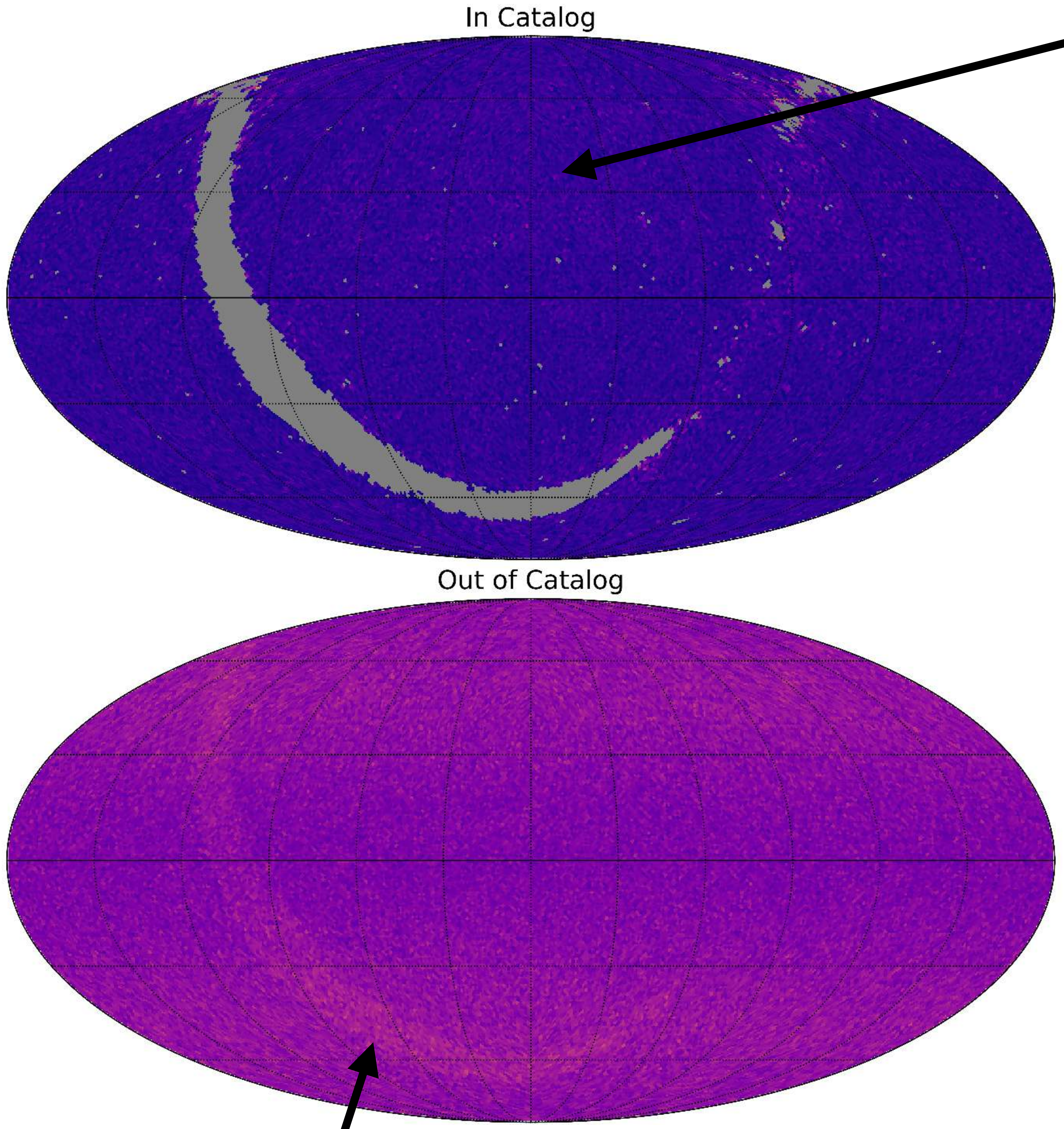


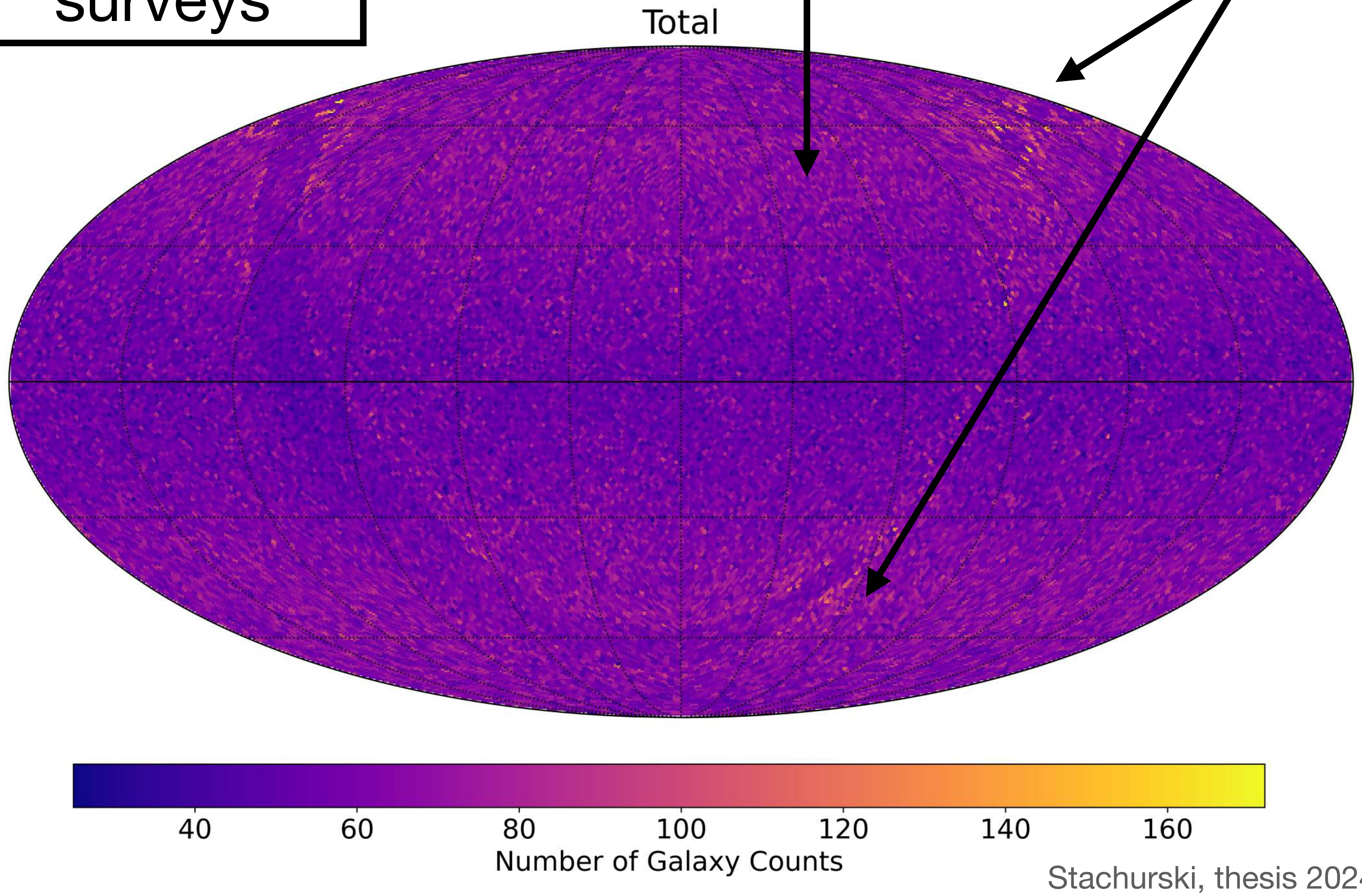
Data generation

Patches due to multiple surveys

Bias away from $\delta = 0$ due to antenna patterns

Some edge effects still present





Stachurski, thesis 2024

Uniform in comoving volume + extra in galactic plane

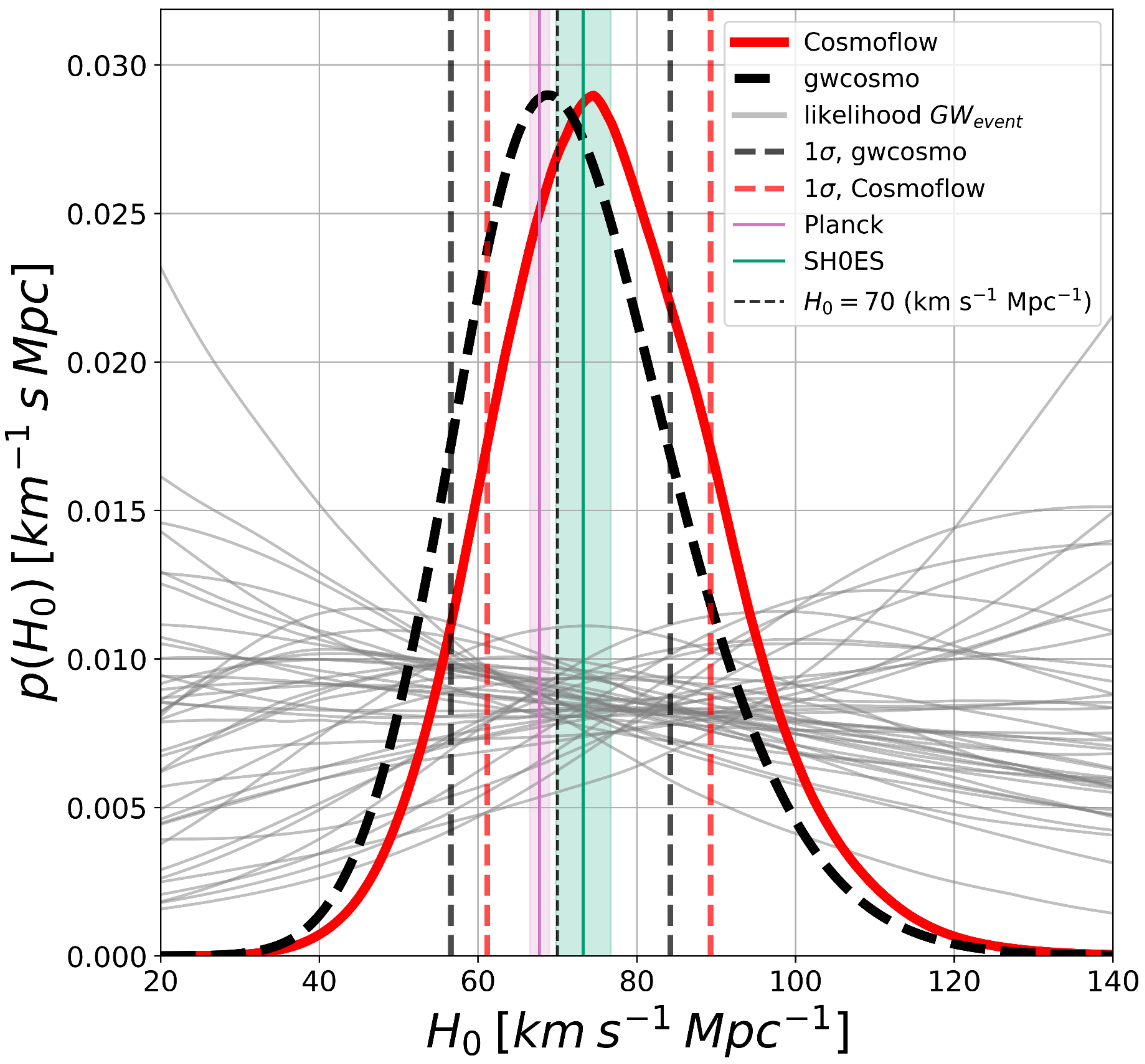
Initial results - O3 42 BBH events

Stachurski *et al*, PRD 109, 2024

Combined O3 result

- Data generation + Training in ~days
- Results obtained in <5 secs
- Significantly faster than original methods … but
- Not fair to compare to current GPU-powered analyses
- Currently re-formulating the problem to be more efficient in high dimensions

Stachurski *et al*, PRD 109, 2024

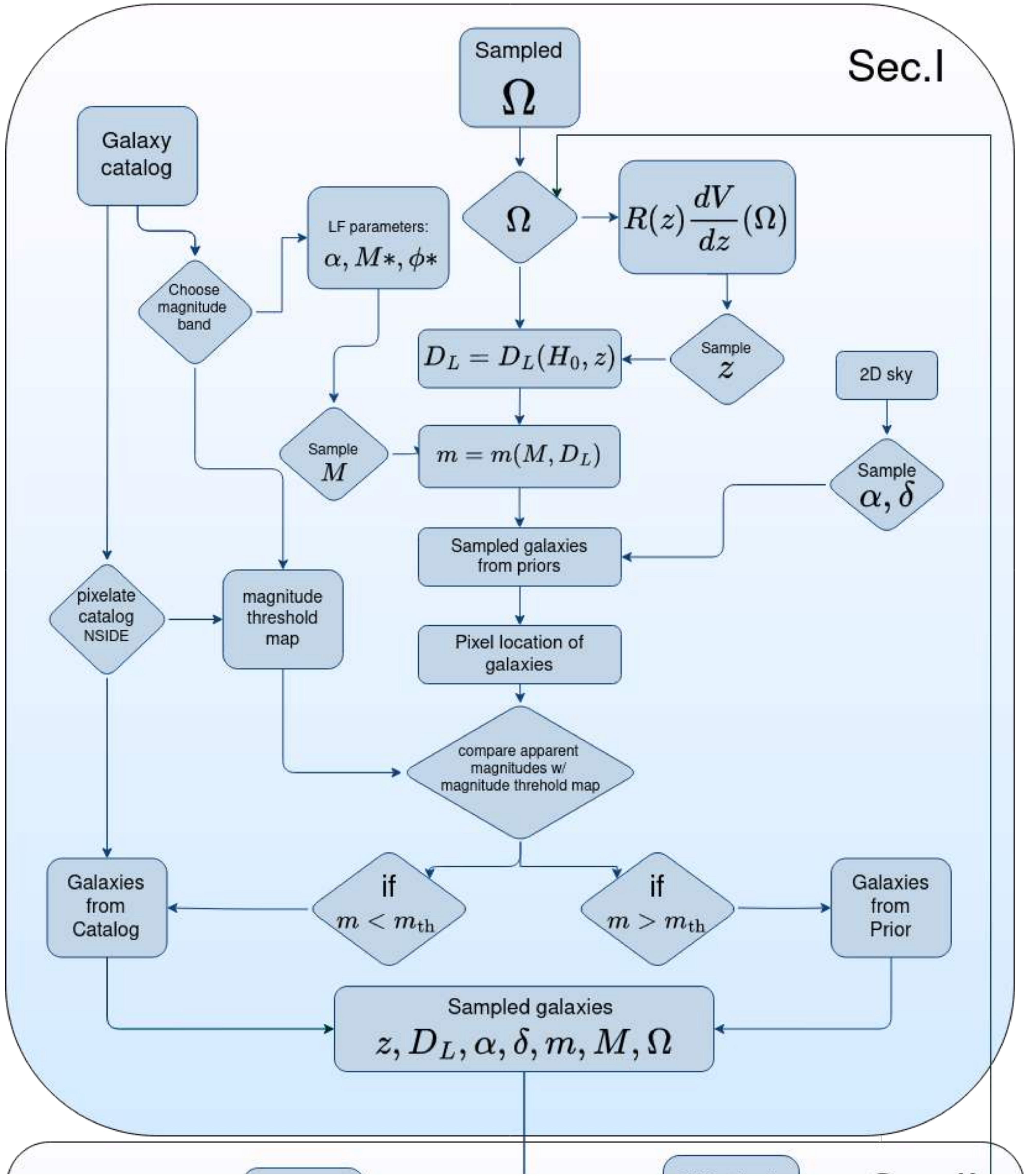


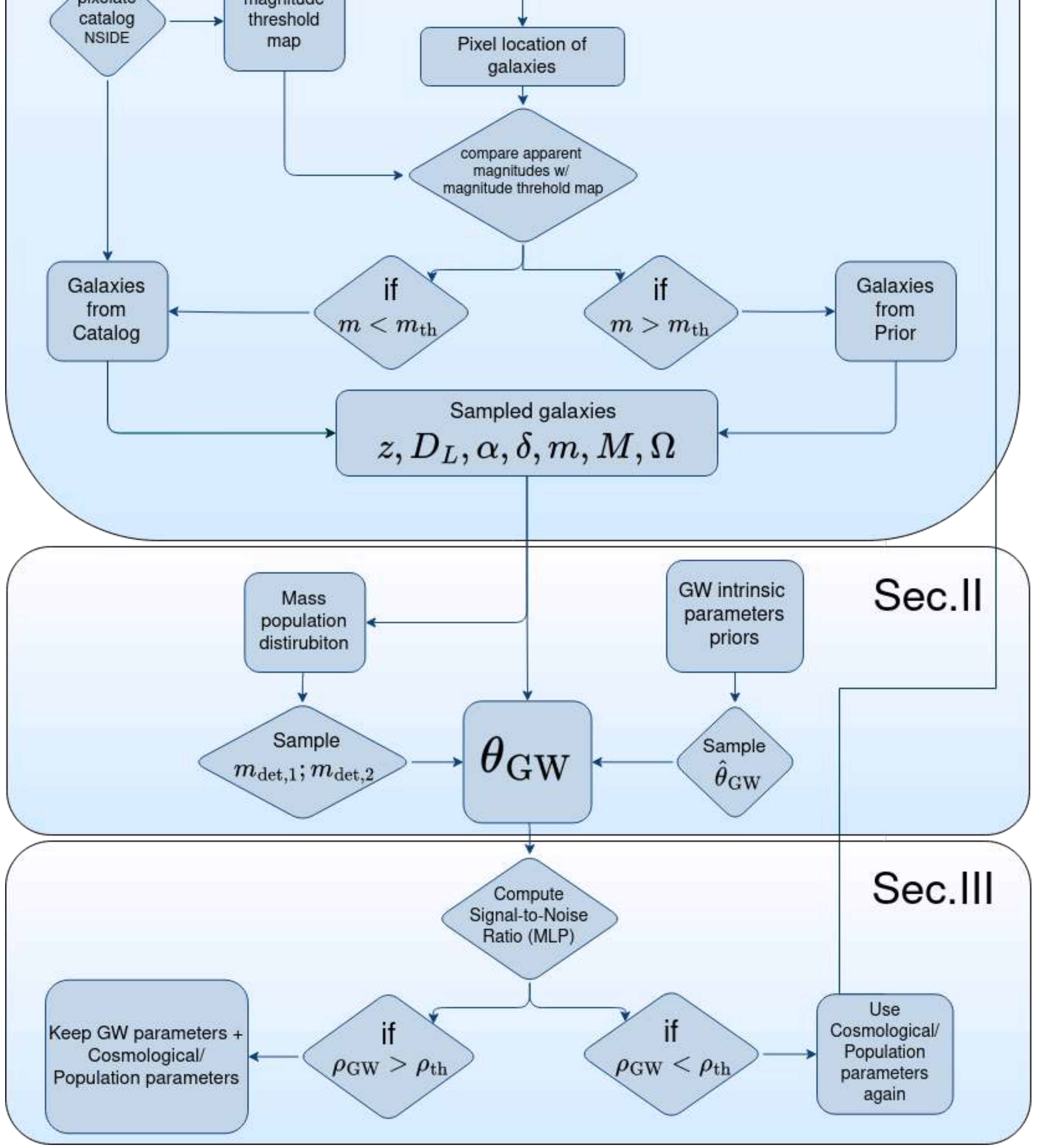
Summary

- GW standard sirens can provide insight into the Hubble tension
- Existing methods are becoming computationally costly
- We have shown that an ML Normalising Flow model can be trained to learn a galaxy catalogue driven GW prior
- This then leads to comparable results with deviations due to different model assumptions, ML model noise, + unknown?
- The 1-D case has been extended to 15-D but the issue of efficient combination of event posteriors is ongoing

Extra slides

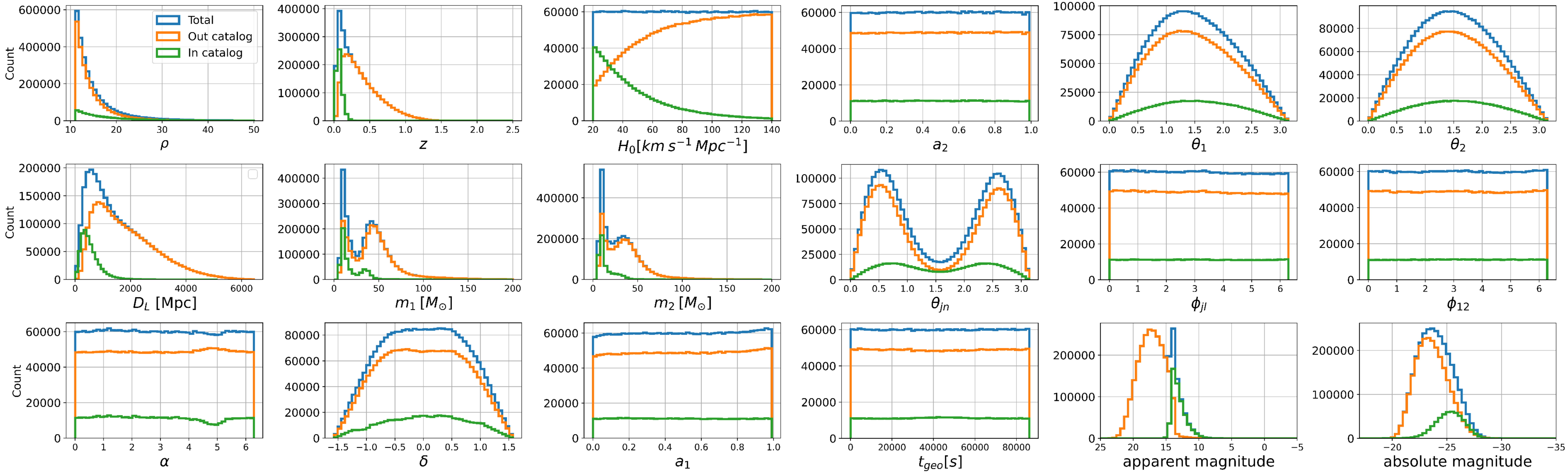
Flow diagram





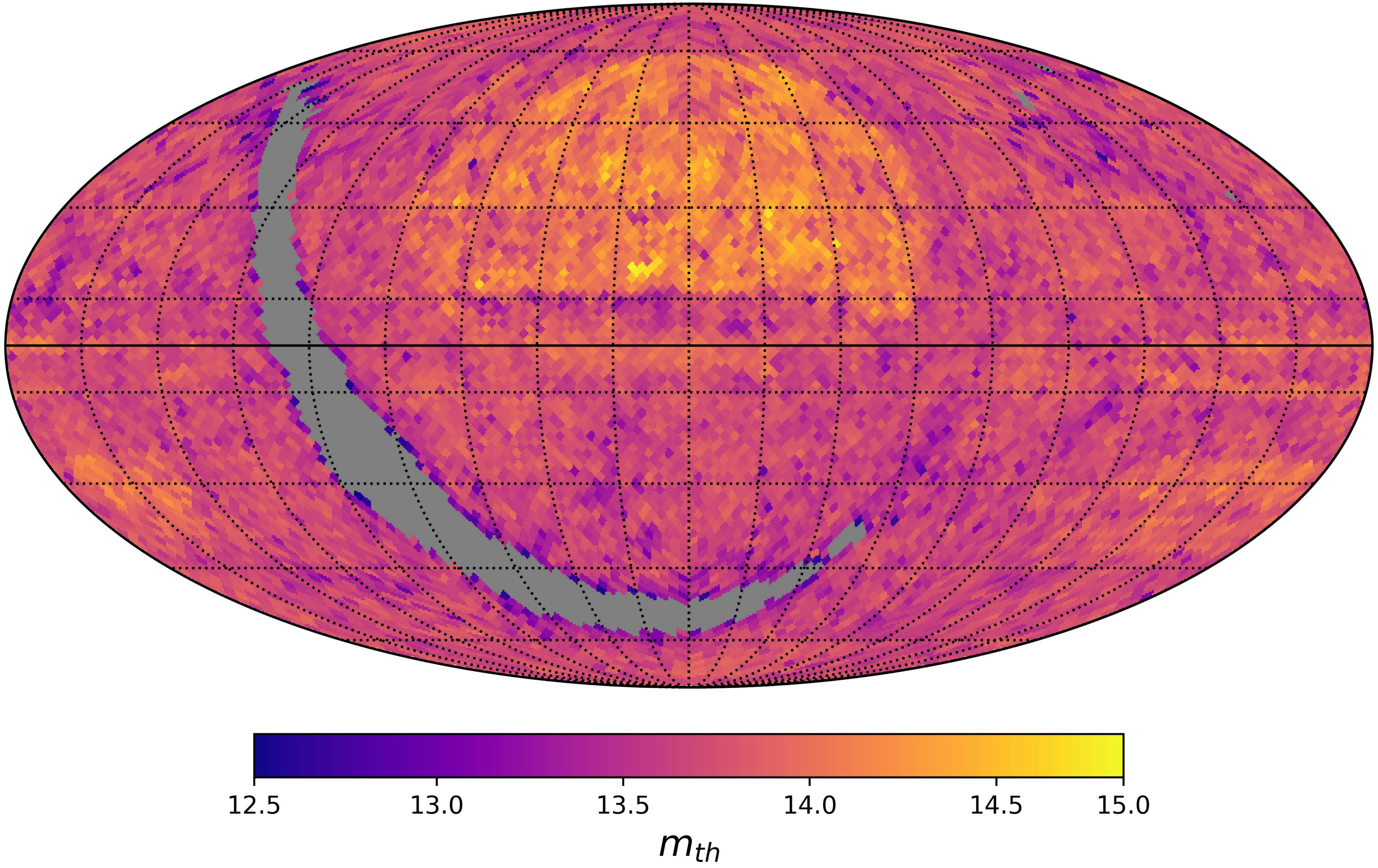
Data generation

Stachurski, thesis 2024

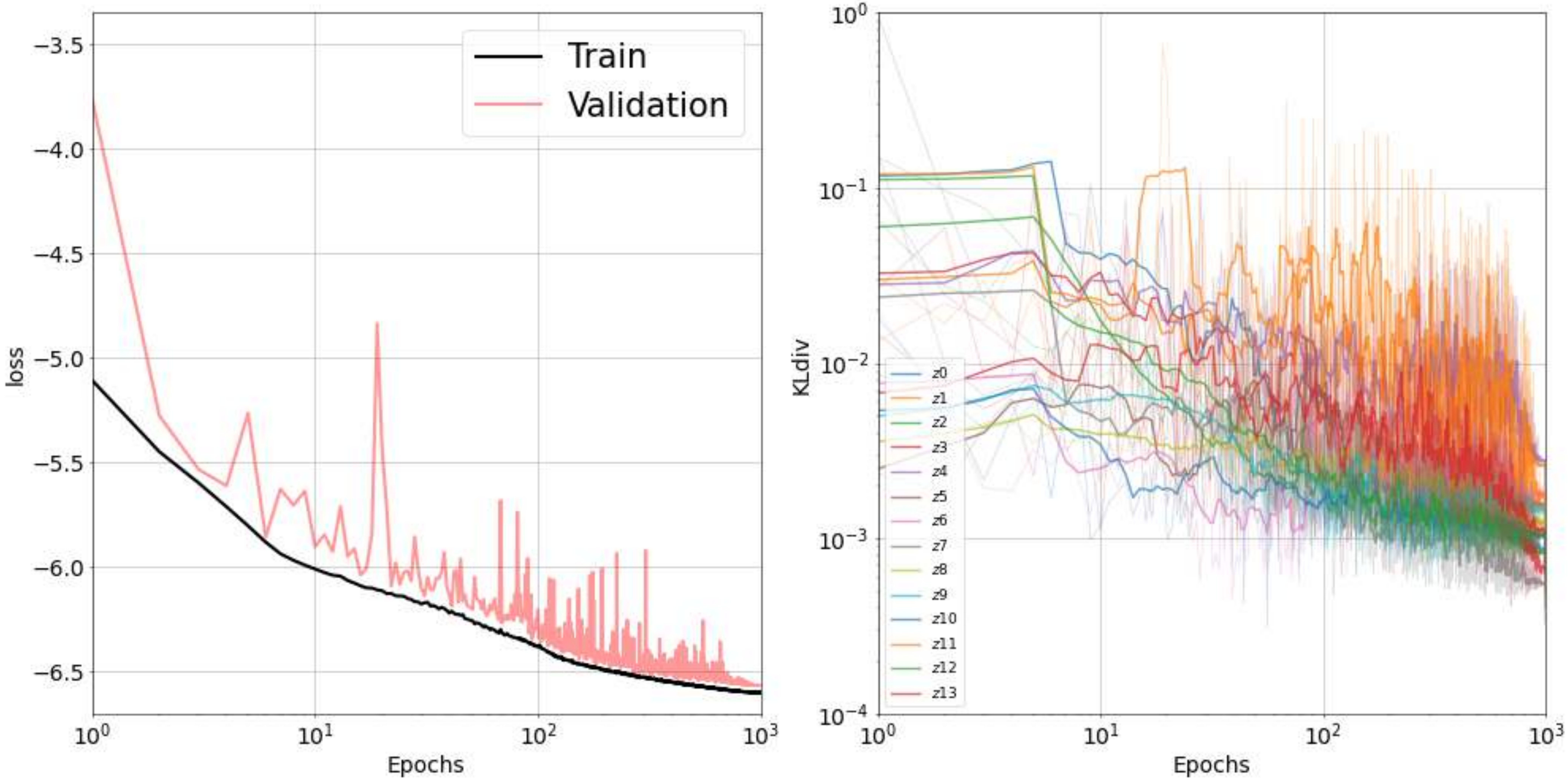


Magnitude threshold (GLADE+)

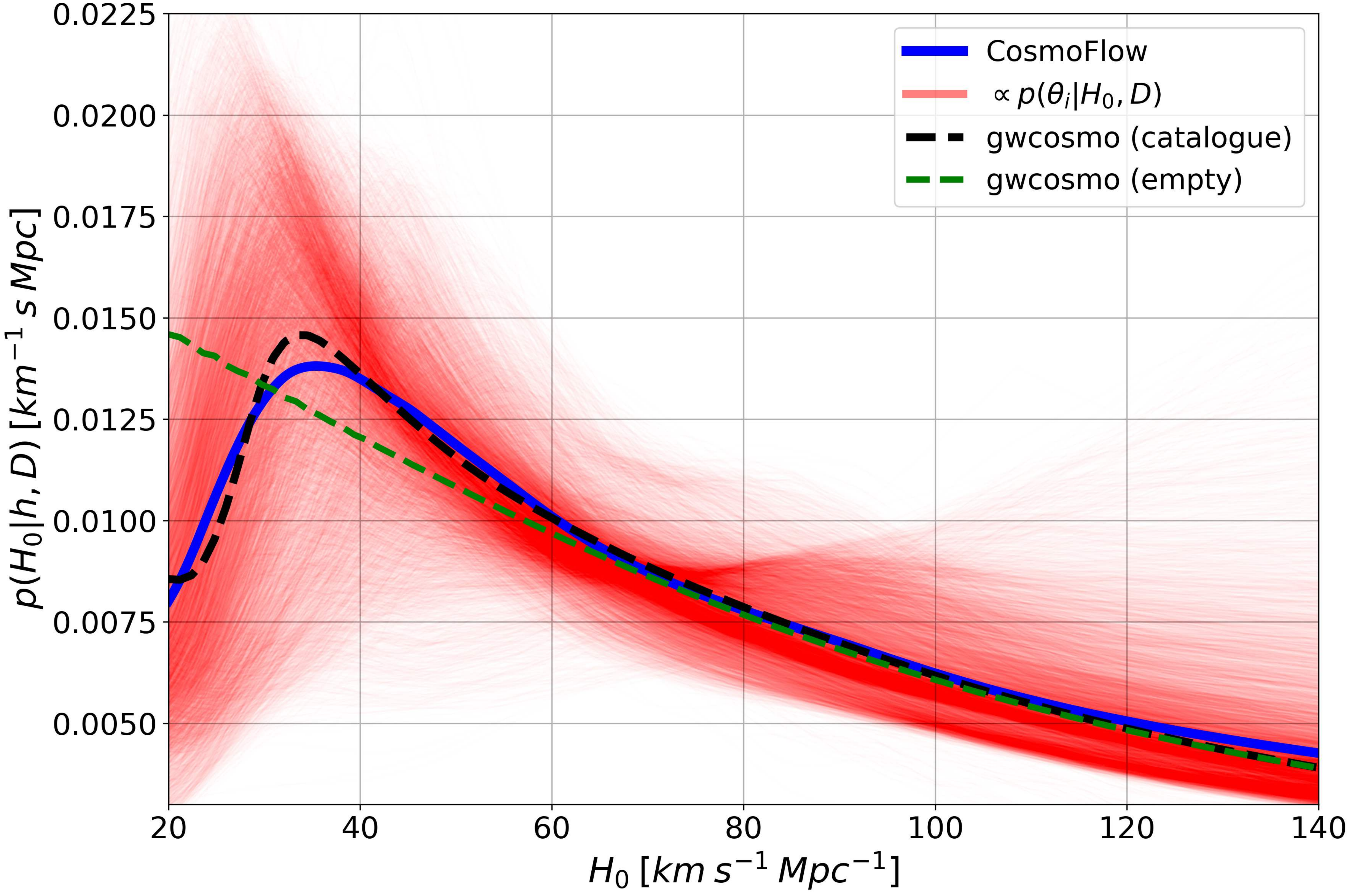
Dálya et al, MNRAS 514, 2022



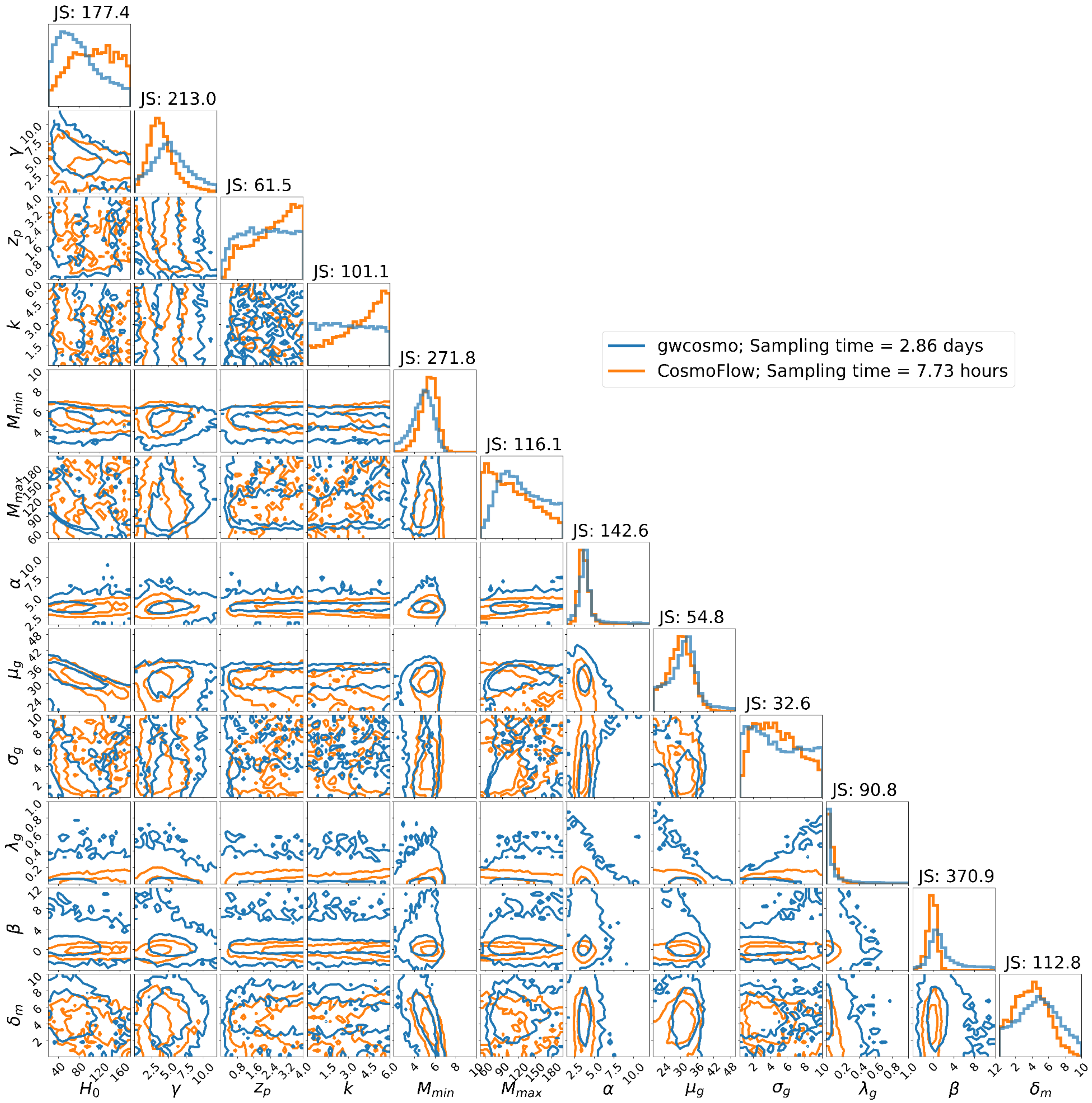
Training loss



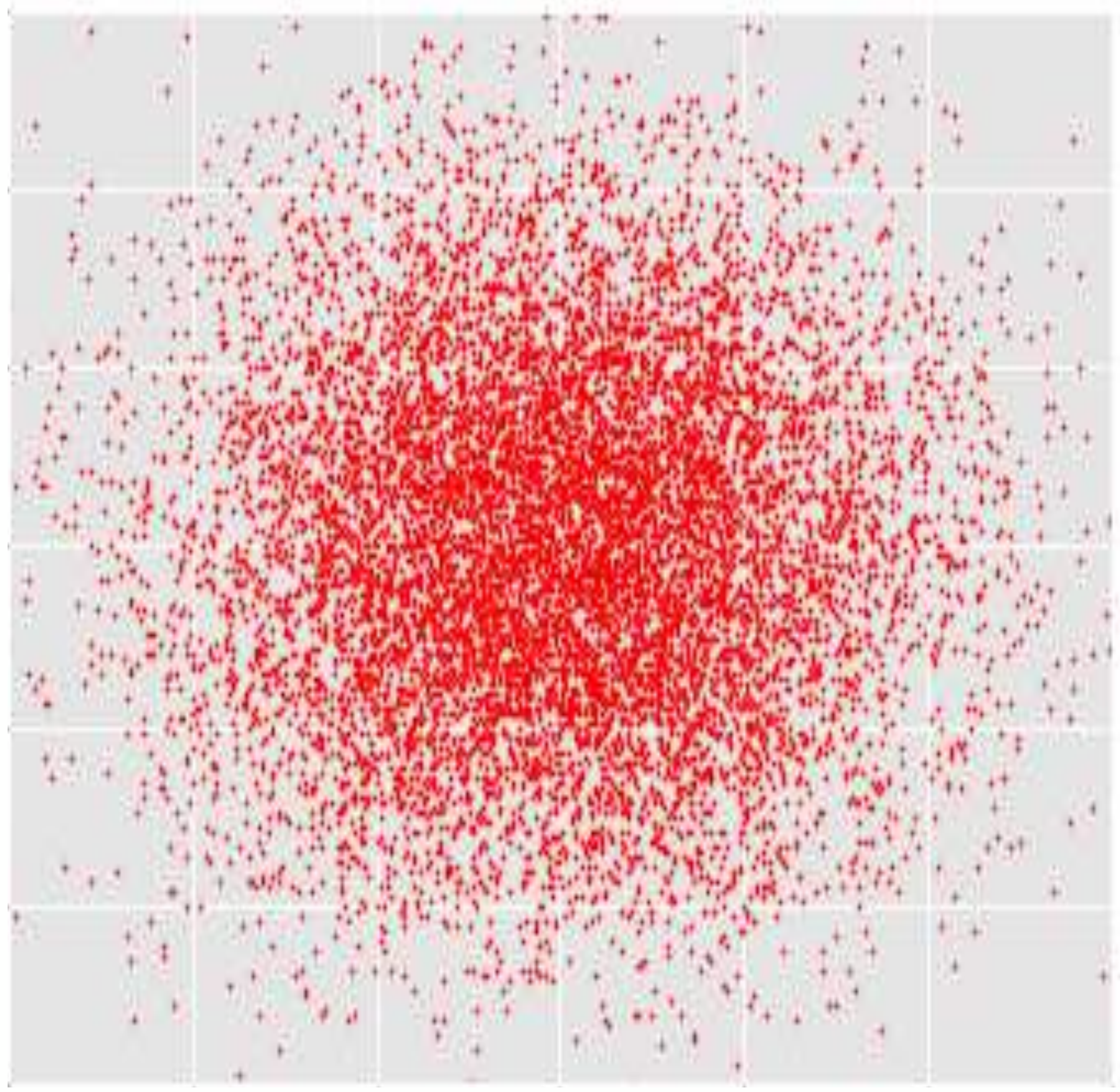
Sample marginalisation



Higher dimensions



What is a Normalising Flow?



- A generative ML model that learns a conditional distribution
- $p_z(z)$ is a Normal distribution in a latent space
- $p_x(x \mid \omega)$ is the conditional data distribution
- $x = f(z, \phi \mid \omega)$ is a bijective function

$$\text{Loss}(\phi) = \text{KL}(p_x(x) \mid q_x(x)) = \text{const.} - \mathbb{E}_{p_x}\left[\log p_z(f^{-1}(x, \phi \mid \omega)) + \log\left(\det\left|\frac{df^{-1}(x, \phi \mid \omega)}{px}\right|\right)\right]$$