Accounting for Selection Effects in Supernova Cosmology with Simulation-Based Inference and Hierarchical Bayesian Modelling

UNIVERSITY OF CAMBRIDGE

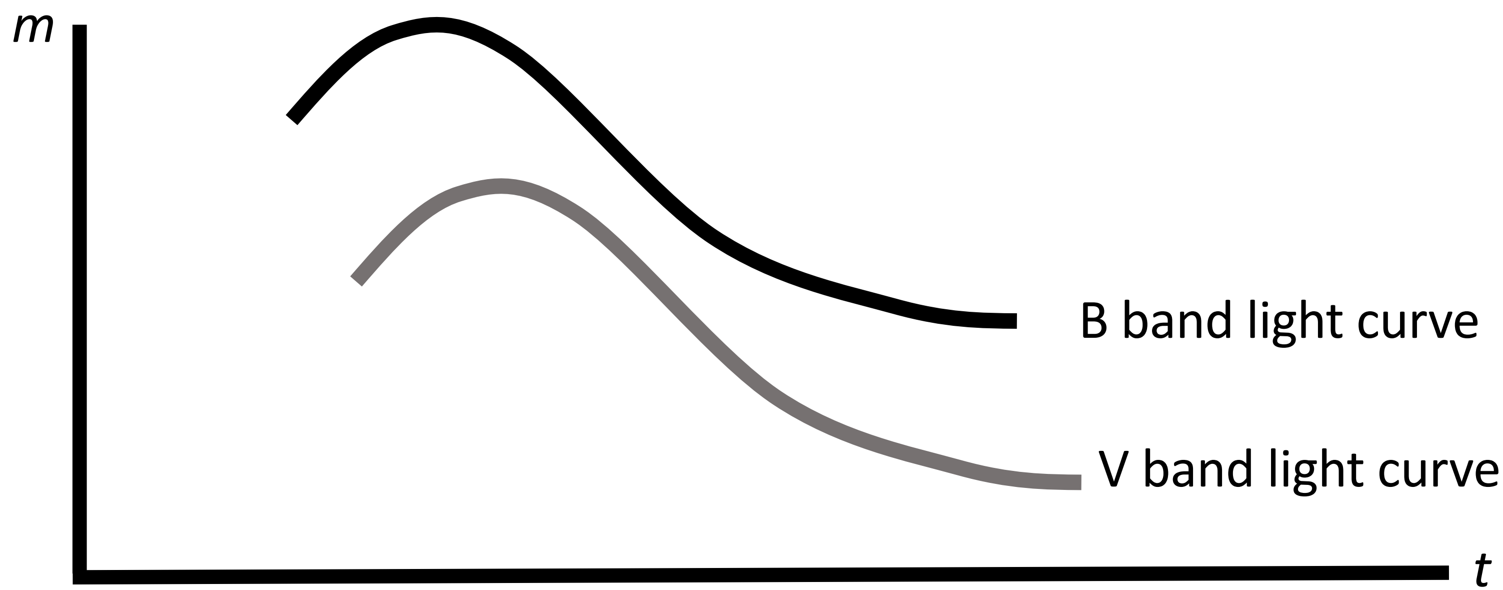
Ben Boyd

Supervisors: Kaisey Mandel & Matthew Grayling

in collaboration with: Ayan Mitra, Gautham Narayan, Richard Kessler & Stephen Thorp

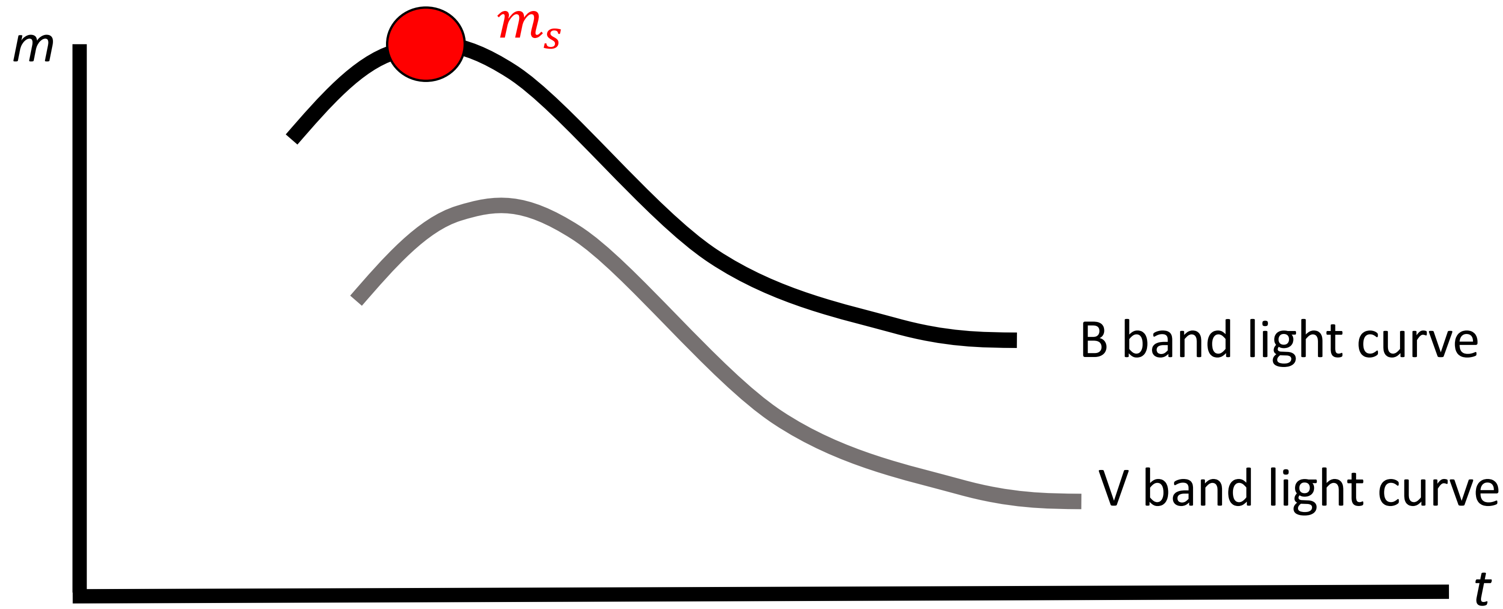
How do we standardise SNe Ia to get their distance?

- We have to use observable properties:



How do we standardise SNe Ia to get their distance?

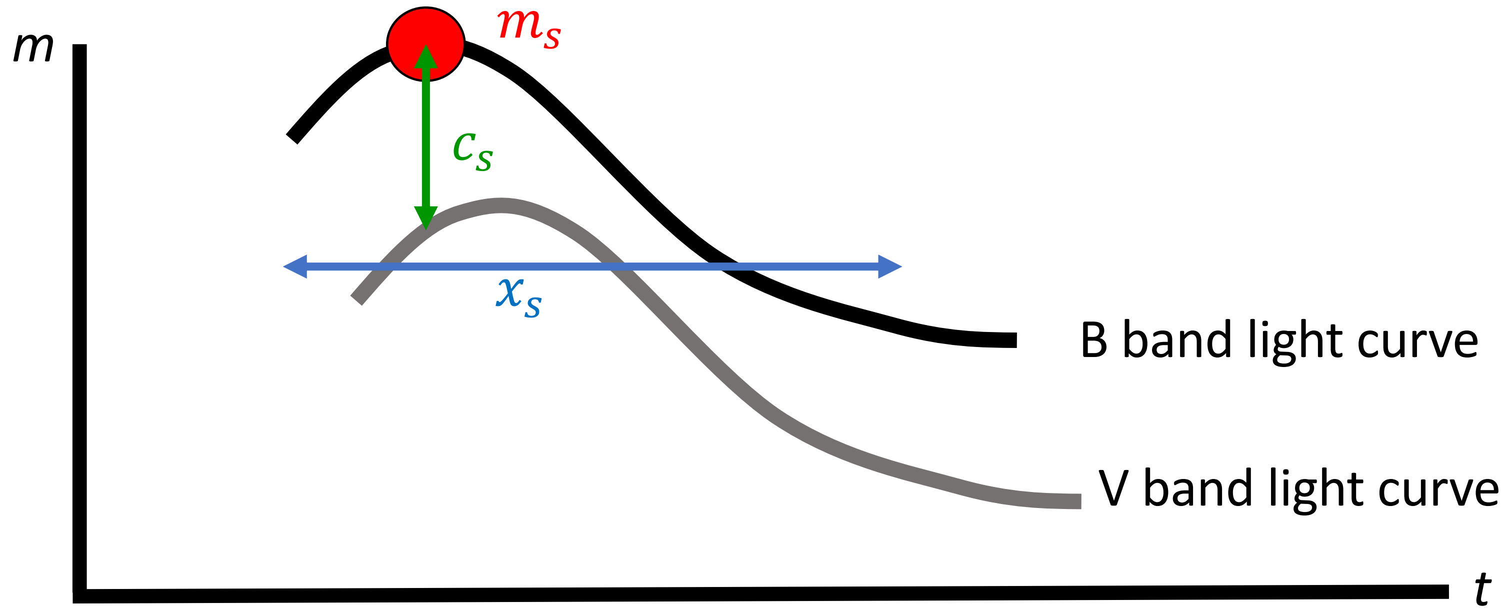
- We have to use observable properties:



Tripp Formula: $m_s = \mu_s(z_s|\Omega_{m0}, w_0) + M_0 +$

How do we standardise SNe Ia to get their distance?

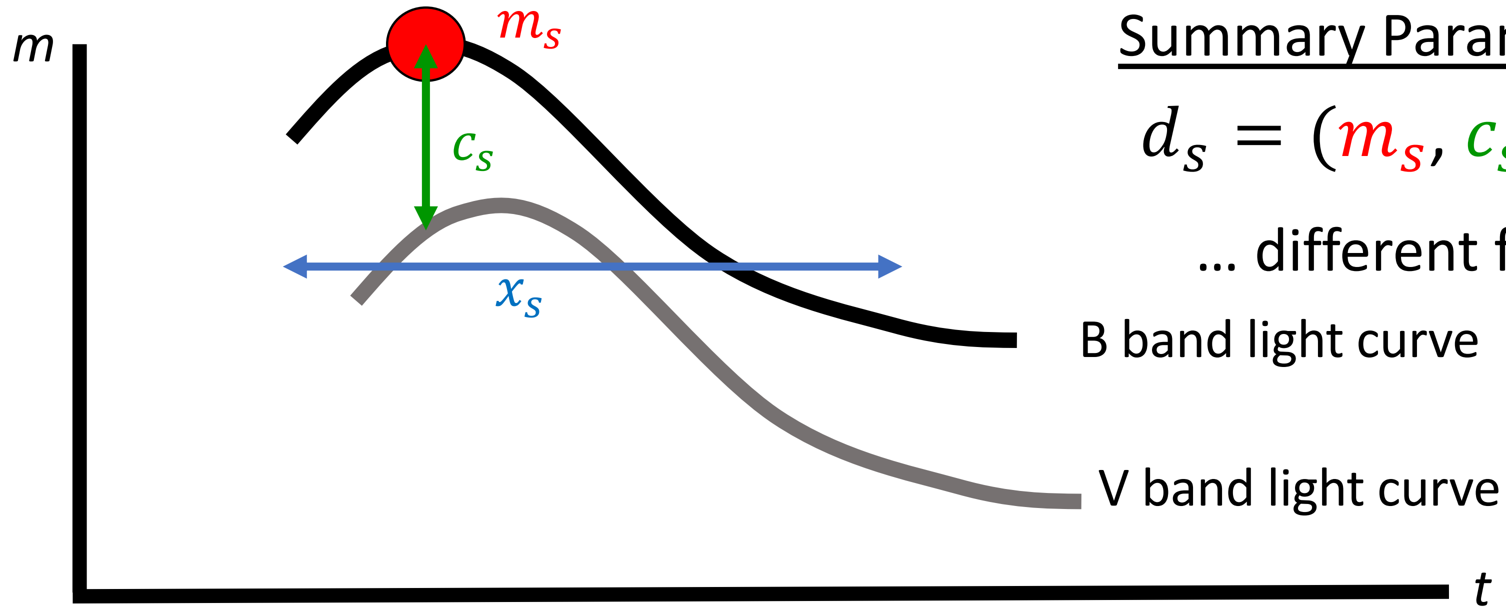
- We have to use observable properties:



Tripp Formula: $m_s = \mu_s(z_s|\Omega_{m0}, w_0) + M_0 + \alpha\, x_s + \beta\, c_s$

How do we standardise SNe Ia to get their distance?

- We have to use observable properties:



Summary Parameters

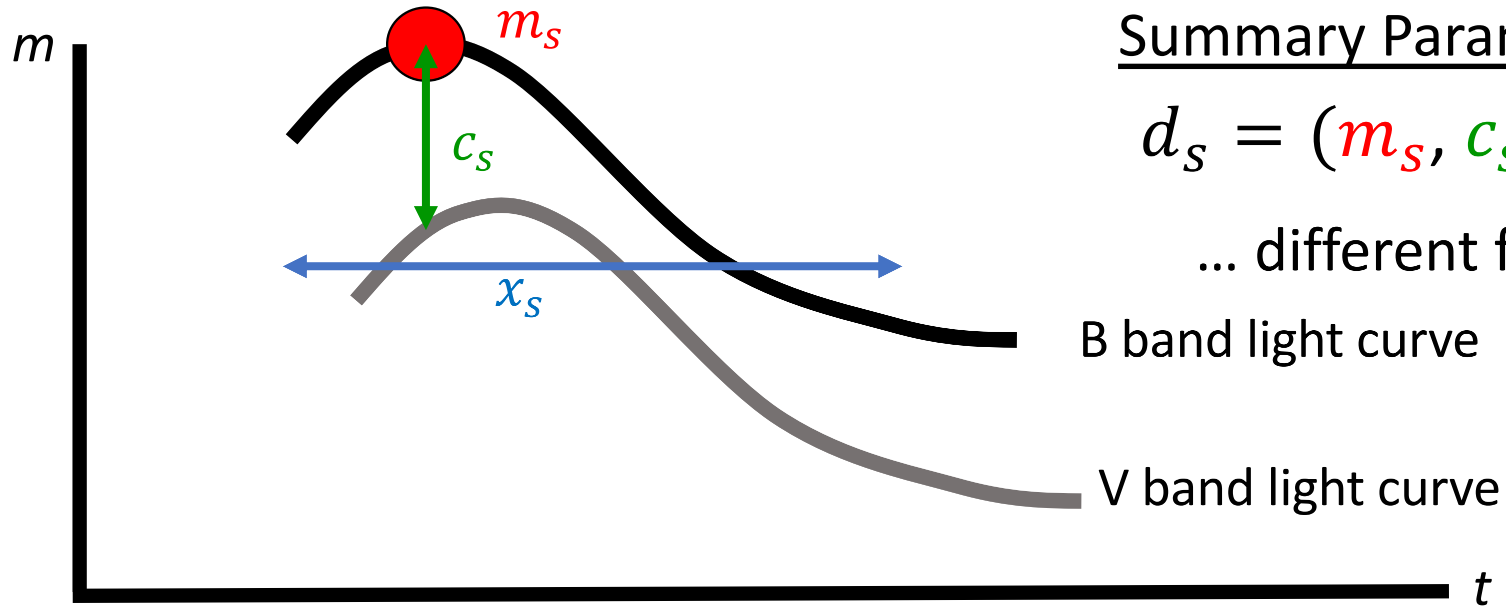
$$d_s = (m_s, c_s, x_s)$$

... different for each SN

Tripp Formula: $m_s = \mu_s(z_s | \Omega_{m0}, w_0) + M_0 + \alpha\, x_s + \beta\, c_s$

How do we standardise SNe Ia to get their distance?

- We have to use observable properties:



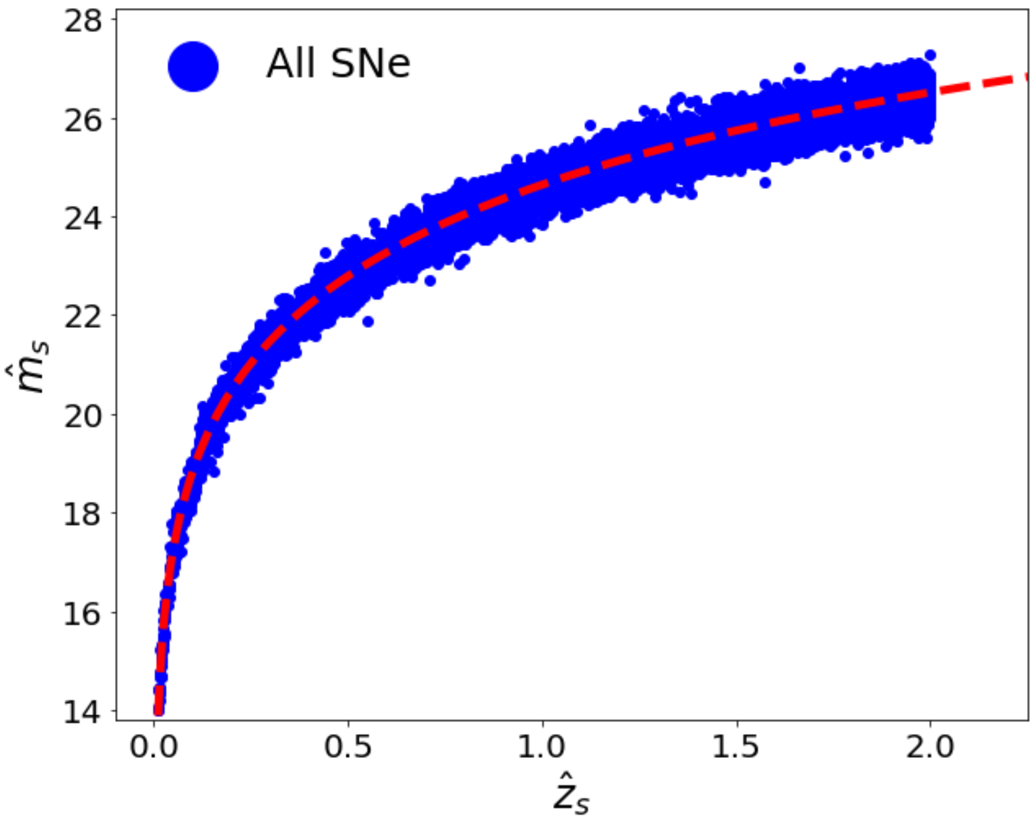
<u>Summary Parameters</u>

$$d_s = (m_s, c_s, x_s)$$

… different for each SN

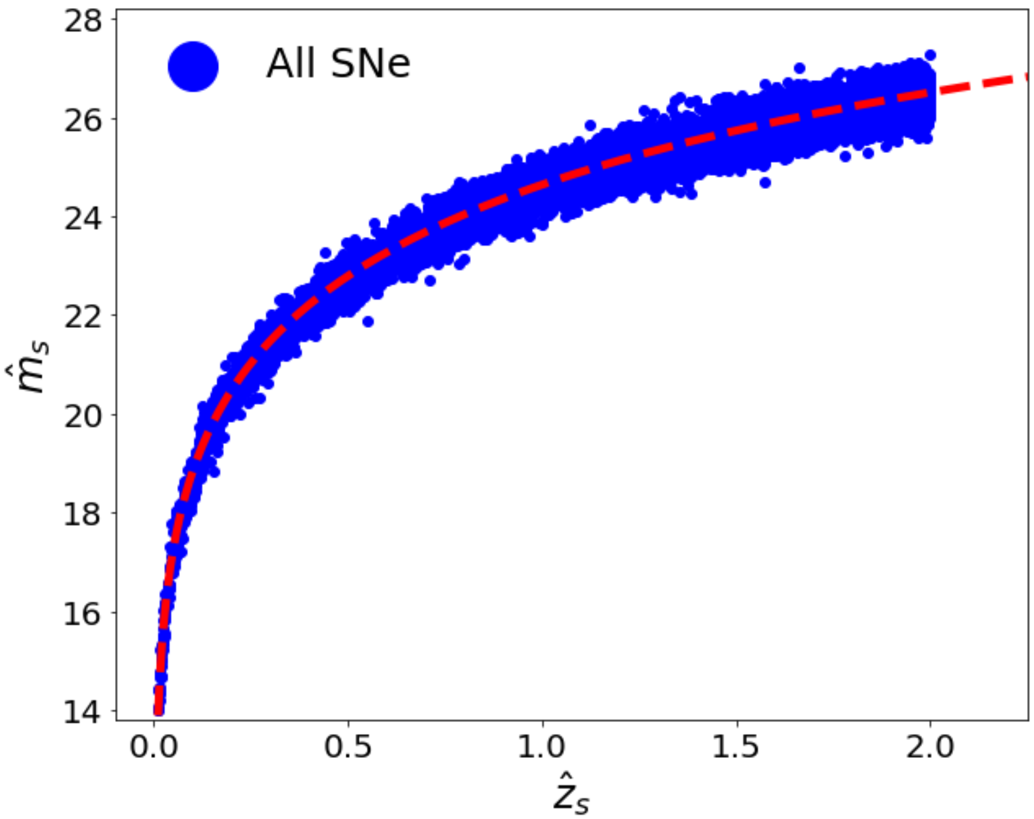
Tripp Formula: $m_s = \mu_s(z_s | \boxed{\Omega_{m0}, w_0}) + \boxed{M_0} + \boxed{\alpha}\, x_s + \boxed{\beta}\, c_s$

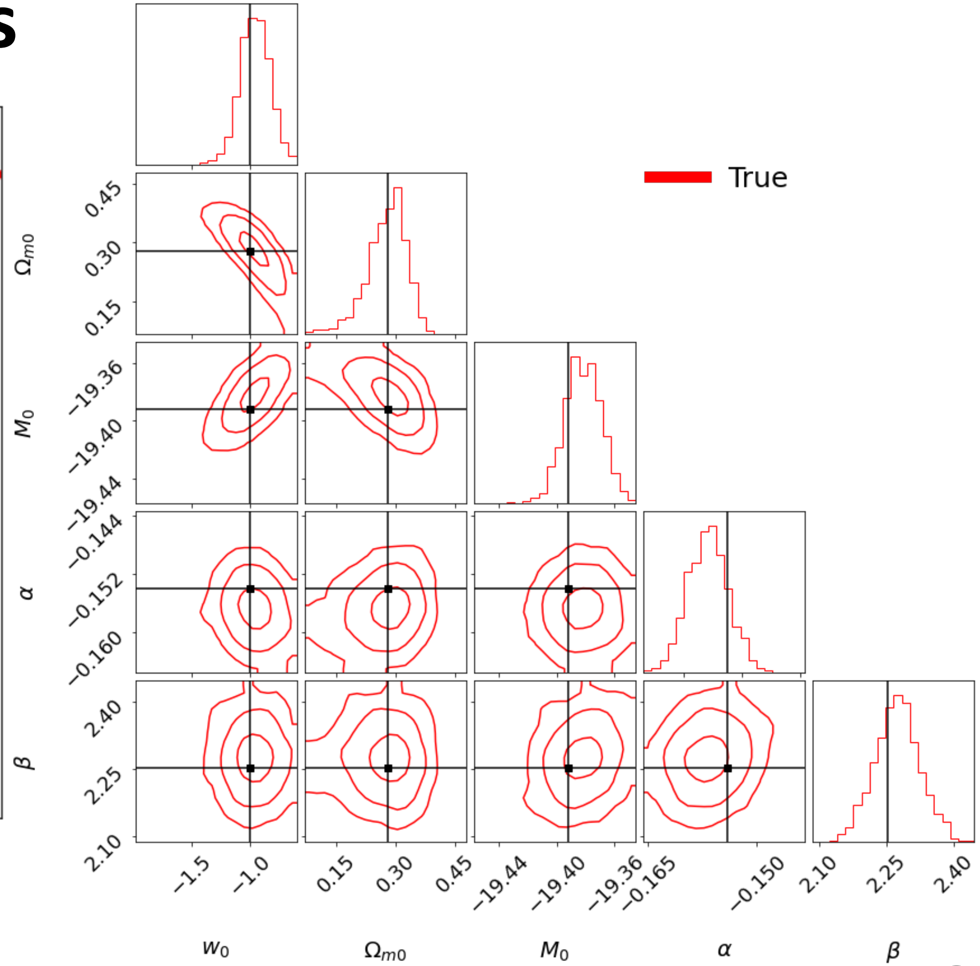
The effect of Malmquist Bias



Adapted from Boyd et al. (2024)

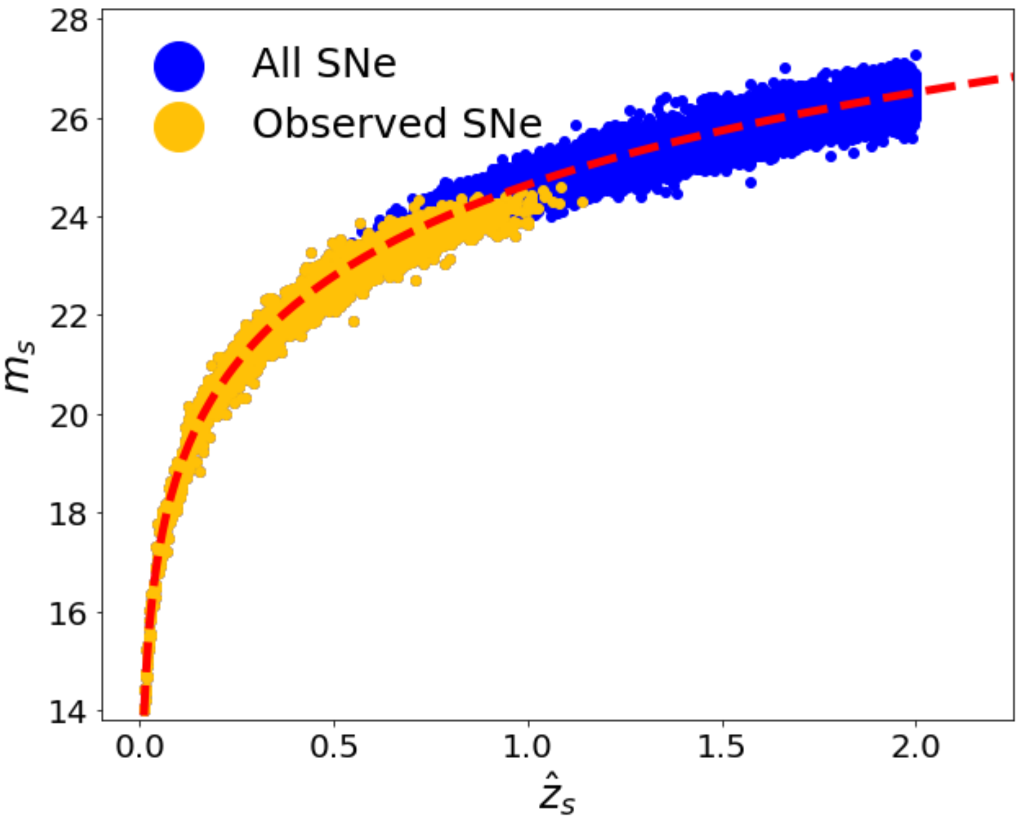
The effect of Malmquist Bias

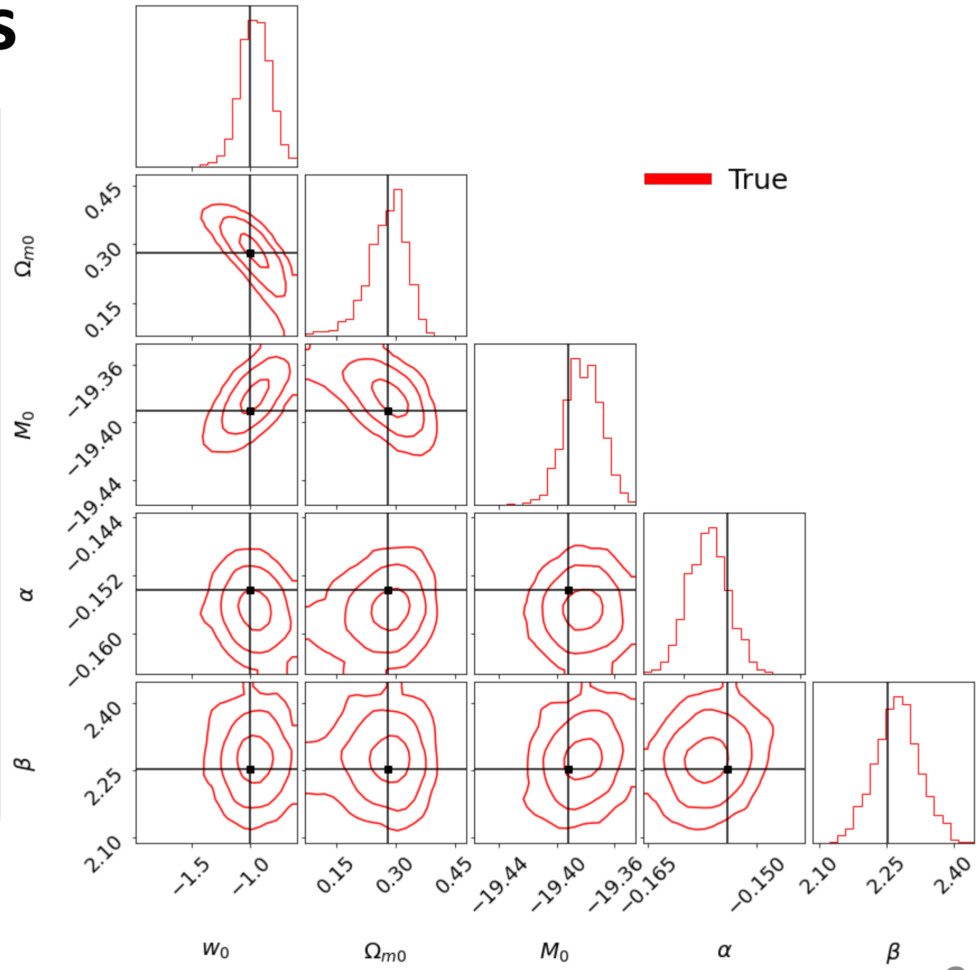




Adapted from Boyd et al. (2024)

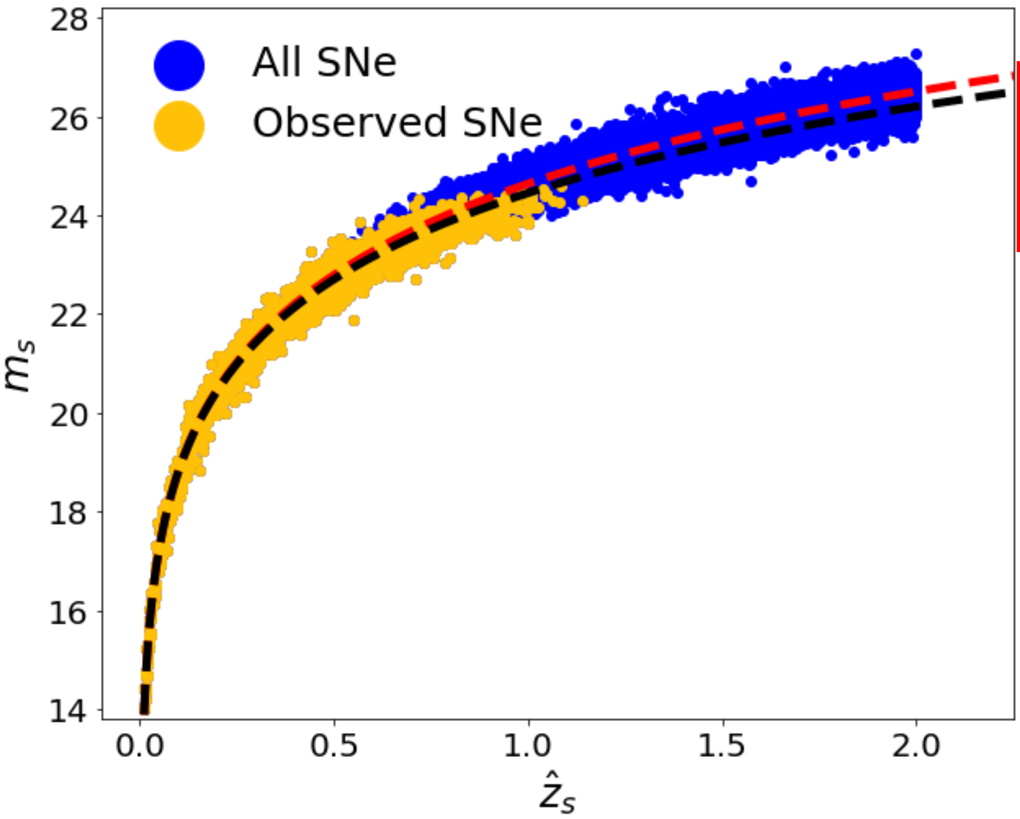
The effect of Malmquist Bias

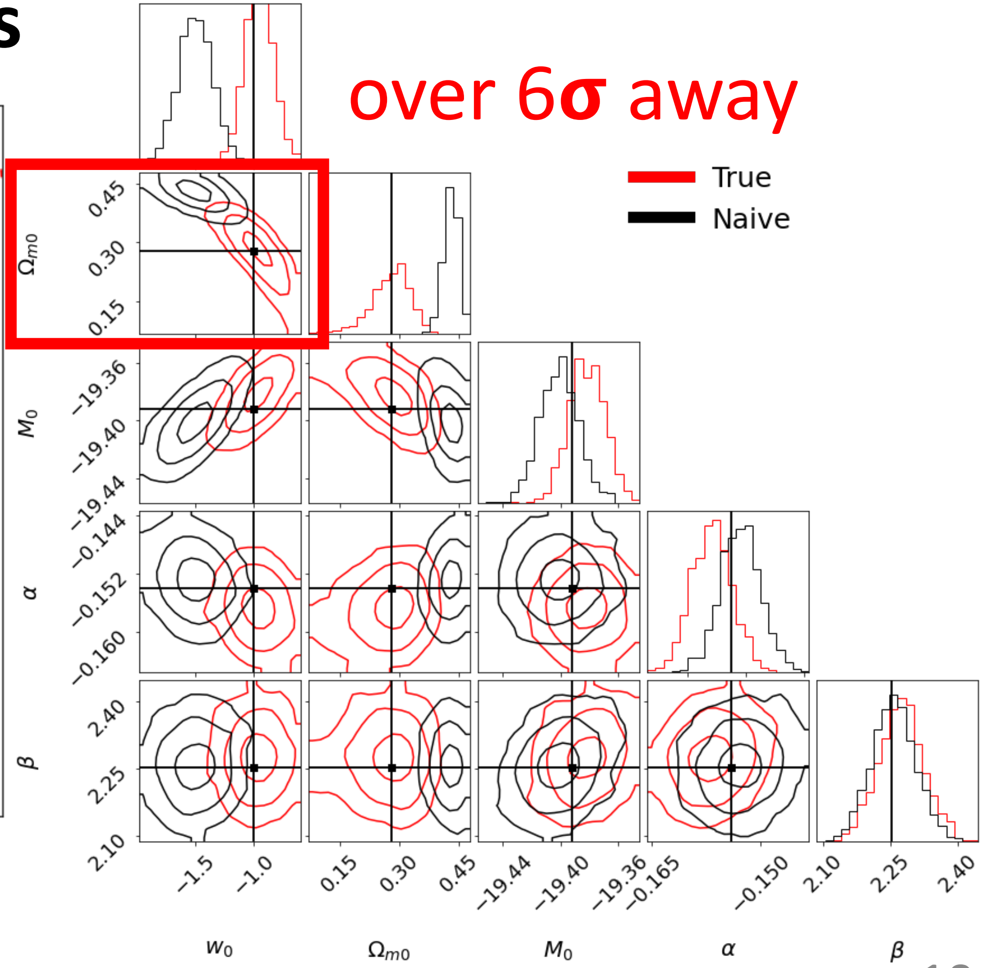




Adapted from Boyd et al. (2024)

The effect of Malmquist Bias





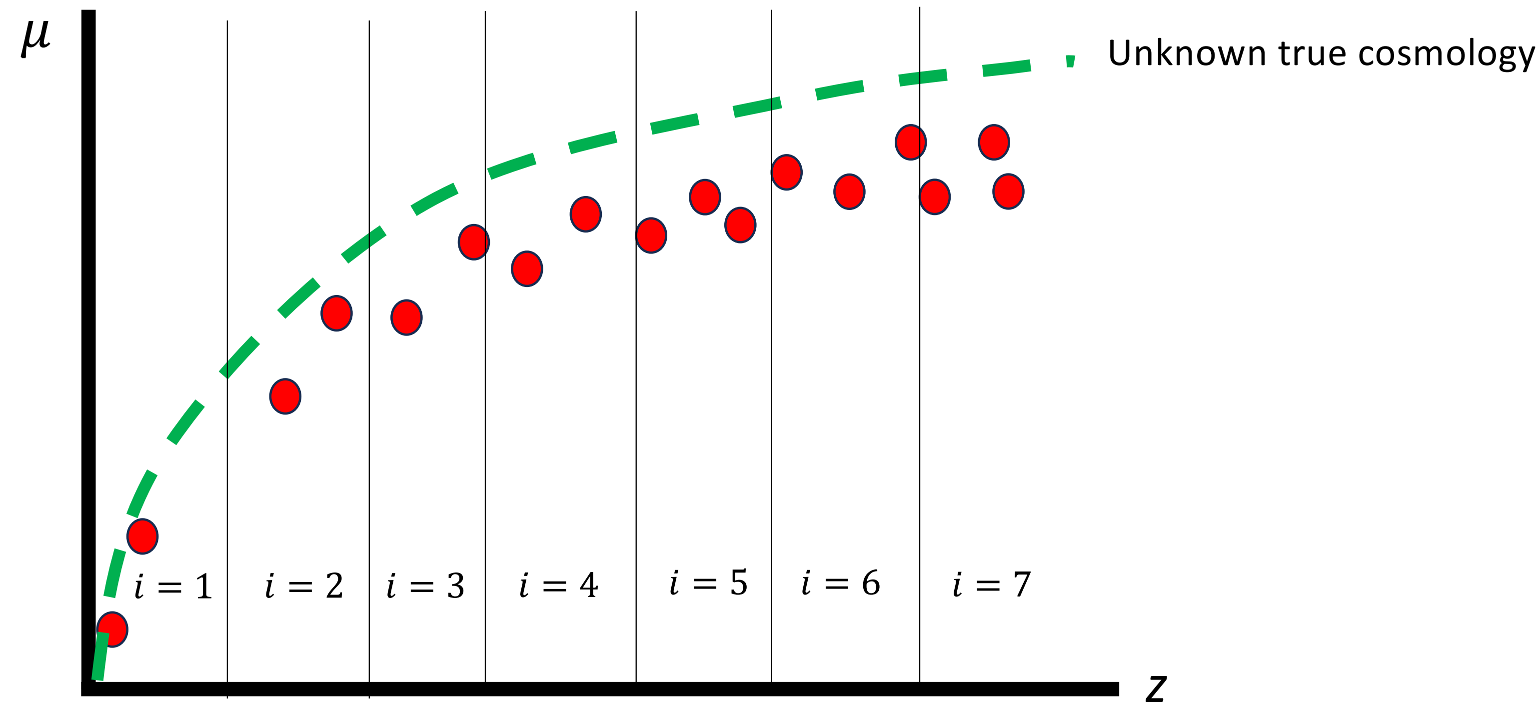
over 6**σ** away

Adapted from Boyd et al. (2024)

Method 1: Bias Corrections With Simulations

(Kessler & Scolnic 2017)

Real Data

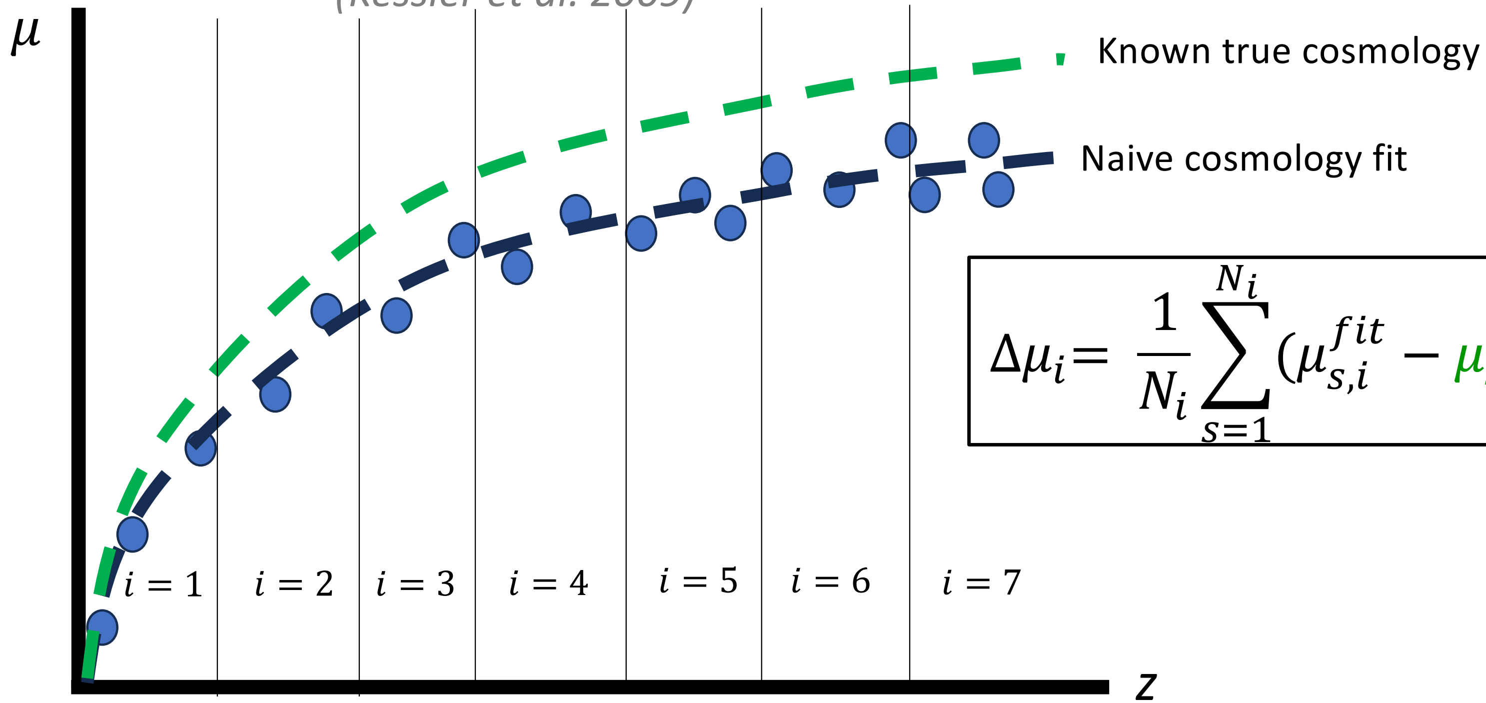


Method 1: Bias Corrections With Simulations

(Kessler & Scolnic 2017)

SNANA Simulations

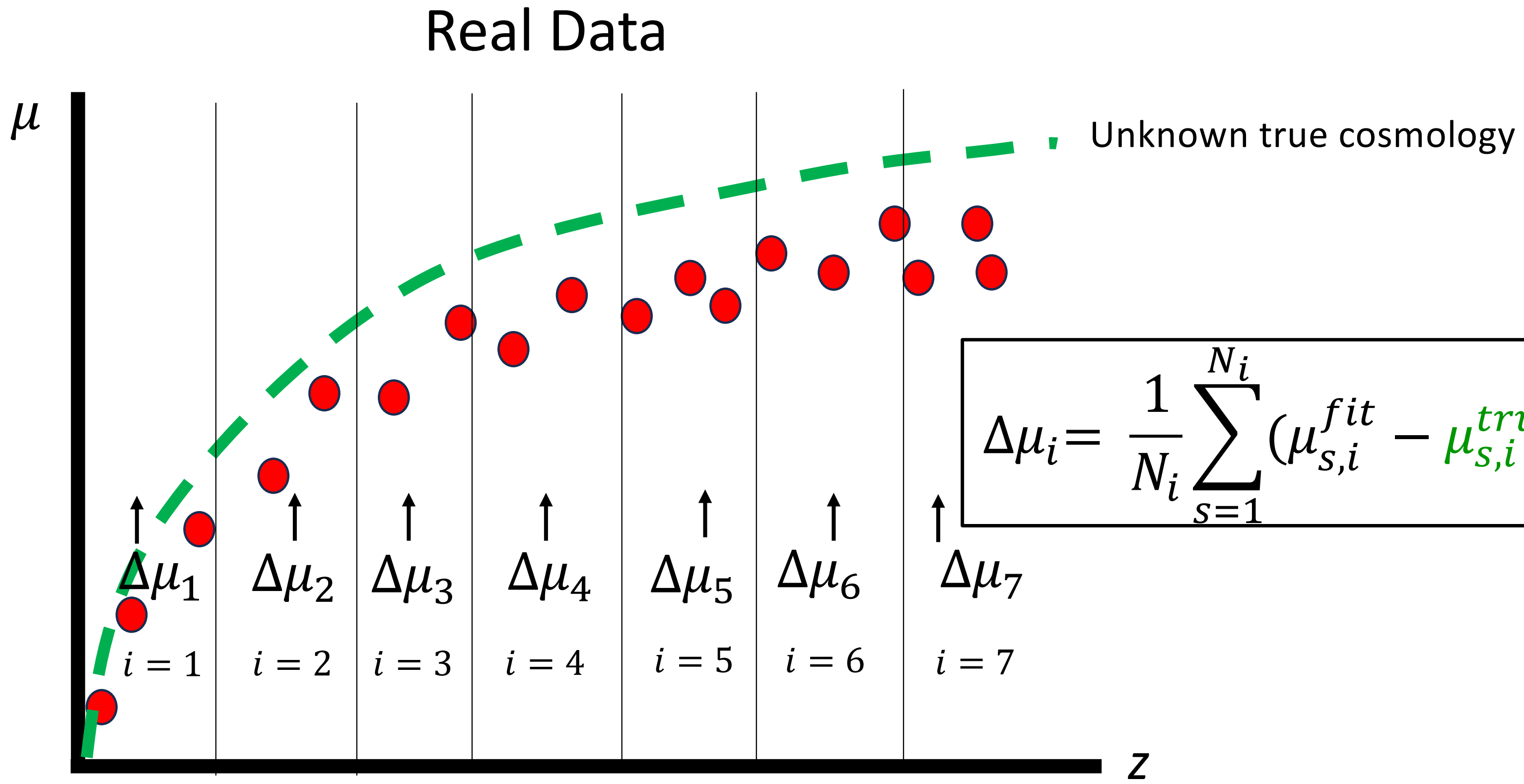
(Kessler et al. 2009)



$$\Delta\mu_i = \frac{1}{N_i}\sum_{s=1}^{N_i}\left(\mu_{s,i}^{fit} - \mu_{s,i}^{true}\right)$$

Method 1: Bias Corrections With Simulations

(Kessler & Scolnic 2017)

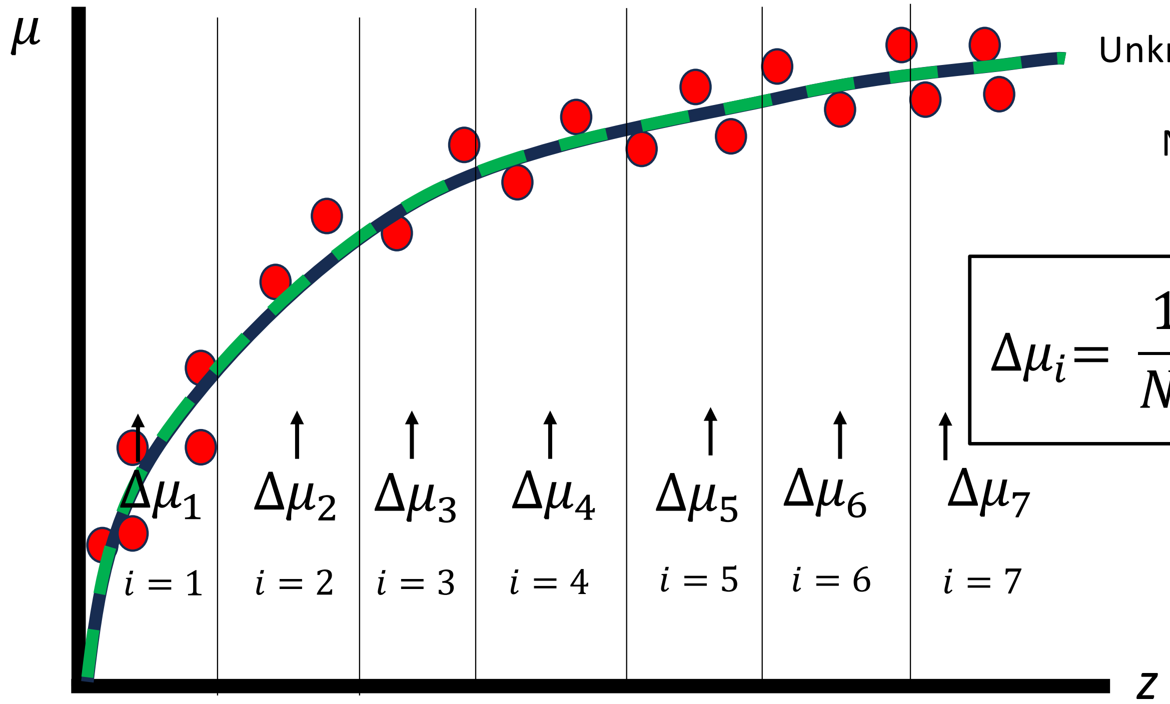


$$\Delta\mu_i = \frac{1}{N_i}\sum_{s=1}^{N_i}(\mu_{s,i}^{fit} - \mu_{s,i}^{true})$$

Method 1: Bias Corrections With Simulations

(Kessler & Scolnic 2017)

Real Data



Unknown true cosmology

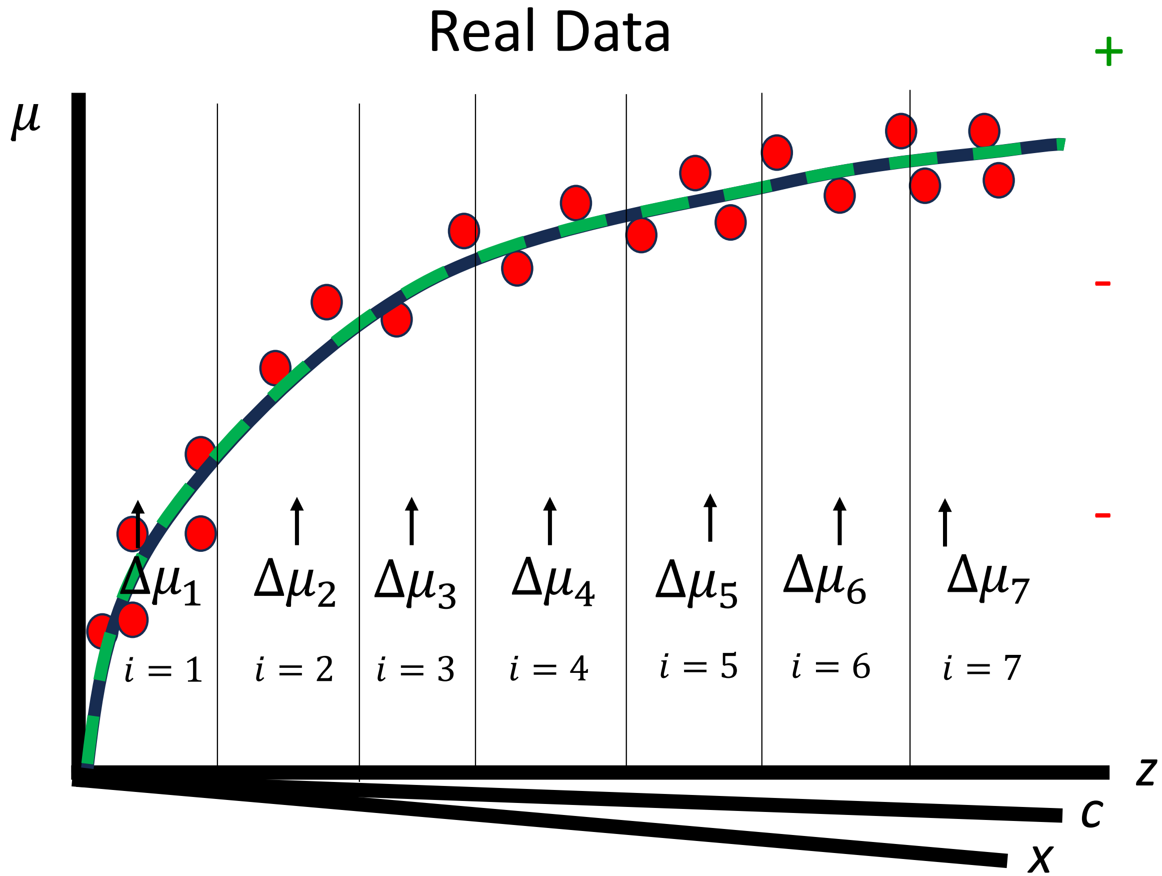
=

Naive cosmology fit

$$\Delta\mu_i = \frac{1}{N_i}\sum_{s=1}^{N_i}\left(\mu_{s,i}^{fit} - \mu_{s,i}^{true}\right)$$

Method 1: Bias Corrections With Simulations

(Kessler & Scolnic 2017)

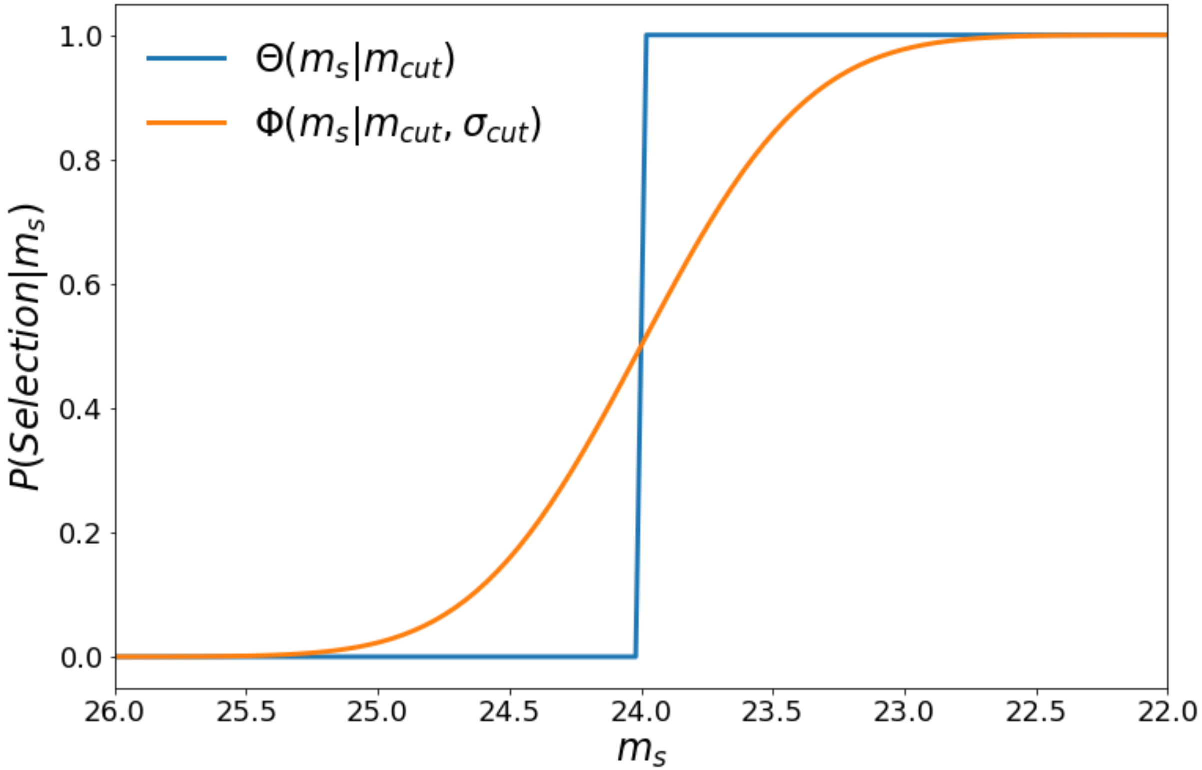


- \+ Uses realistic survey simulations
- \- Assumes fiducial cosmology to correct distances
- \- Uses unphysical binning

Method 2: Analytical Function Approach

(Rubin et al. 2015, March et al. 2018, Hinton et al. 2019)

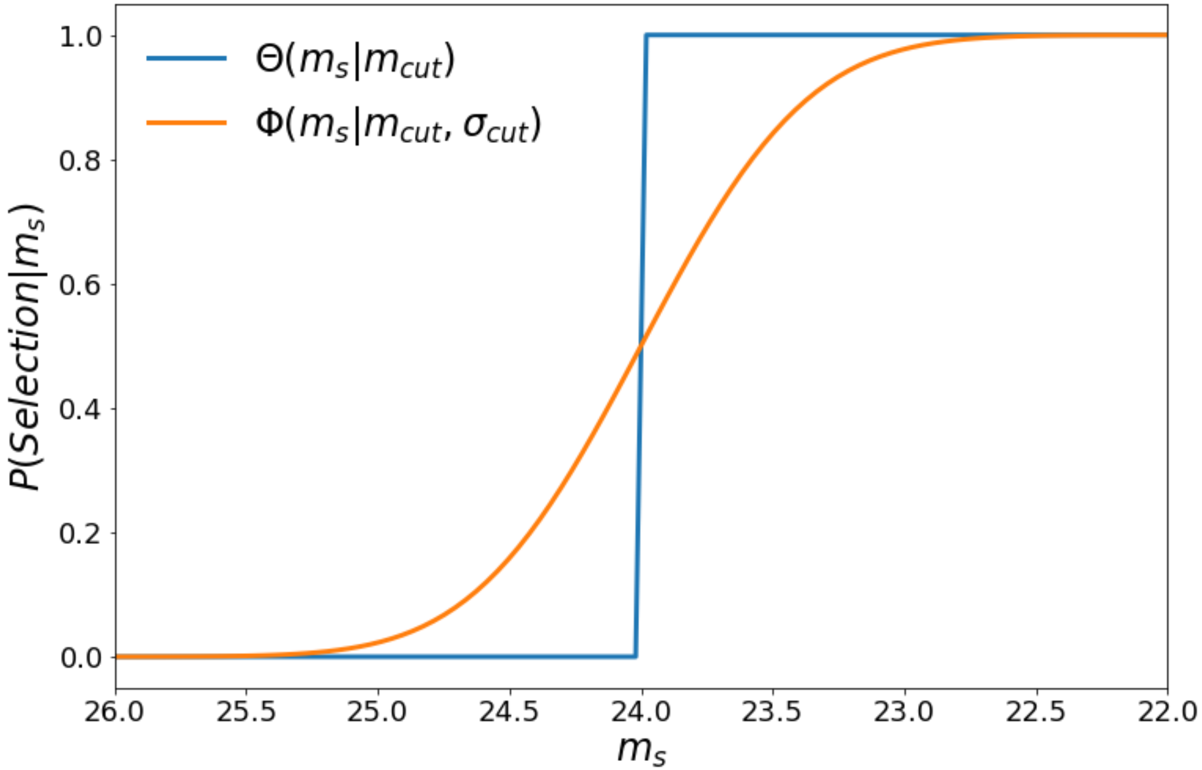
Assume the probability of selection is a known analytical function:



Method 2: Analytical Function Approach

(Rubin et al. 2015, March et al. 2018, Hinton et al. 2019)

Assume the probability of selection is a known analytical function:



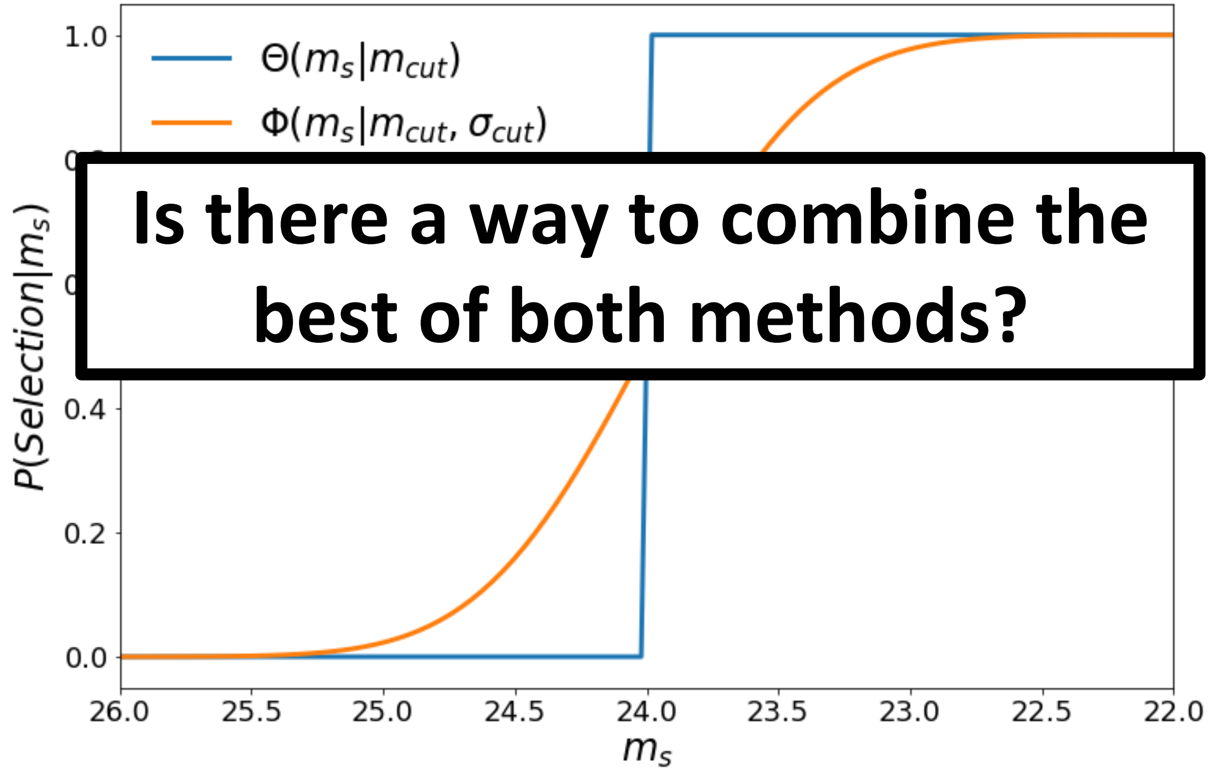
\+ One fully hierarchical Bayesian model

\- Assumes selection function is simple

Method 2: Analytical Function Approach

(Rubin et al. 2015, March et al. 2018, Hinton et al. 2019)

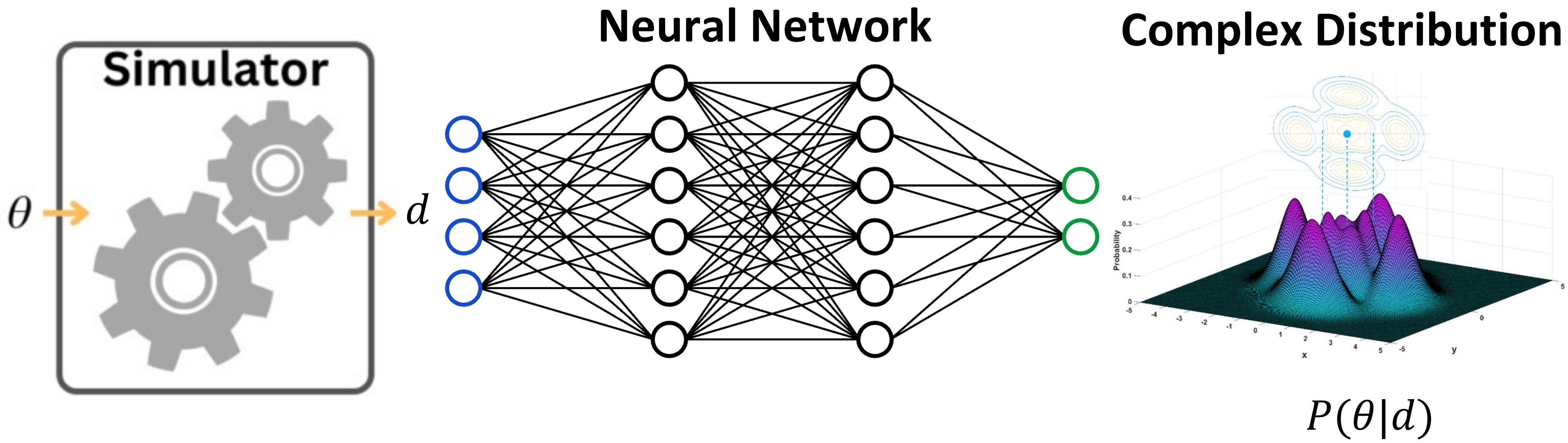
Assume the probability of selection is a known analytical function:



Is there a way to combine the best of both methods?

- \+ One fully hierarchical Bayesian model
- \- Assumes selection function is simple

Solution: Simulation-Based Inference (SBI)!



Which distribution to model with SBI?

$$d_s = (m_s, c_s, x_s)$$
$$\theta = (\Omega_{m0}, w_0, M_0, \alpha, \beta, ..)$$

Bayes' Theorem

$$P(\theta|\{d_s\}) \propto P(\{d_s\}|\theta)P(\theta)$$

Which distribution to model with SBI?

$$d_s = (m_s, c_s, x_s)$$

$$\theta = (\Omega_{m0}, w_0, M_0, \alpha, \beta, ..)$$

Bayes' Theorem

$$P(\theta|\{d_s\}) \propto P(\{d_s\}|\theta)P(\theta)$$

Posterior

Likelihood

Prior

Which distribution to model with SBI?

$$d_s = (m_s, c_s, x_s)$$

$$\theta = (\Omega_{m0}, w_0, M_0, \alpha, \beta, ..)$$

Bayes' Theorem

$$P(\theta|\{d_s\}) \propto P(\{d_s\}|\theta)P(\theta)$$

We could model these

Which distribution to model with SBI?

$$d_s = (m_s, c_s, x_s)$$
$$\theta = (\Omega_{m0}, w_0, M_0, \alpha, \beta, ..)$$

Bayes' Theorem

$$P(\theta|\{d_s\}) \propto P(\{d_s\}|\theta)P(\theta)$$

We could model these

But:

- High dimensional so difficult to learn
- Would need to train a new neural network for each cosmology tested

Which distribution to model with SBI?

$$d_s = (m_s, c_s, x_s)$$

$$\theta = (\Omega_{m0}, w_0, M_0, \alpha, \beta, ..)$$

Bayes' Theorem

$$P(\theta|\{d_s\}) \propto P(\{d_s\}|\theta)P(\theta)$$

$$= \left[\prod_{s=1}^{N_{SN}} P(d_s|\theta)\right]P(\theta)$$

Which distribution to model with SBI?

$$d_s = (m_s, c_s, x_s)$$
$$\theta = (\Omega_{m0}, w_0, M_0, \alpha, \beta, ..)$$

Bayes' Theorem

$$P(\theta|\{d_s\}) \propto P(\{d_s\}|\theta)P(\theta)$$

$$= \left[\prod_{s=1}^{N_{SN}} \boxed{P(d_s|\theta)}\right]P(\theta)$$

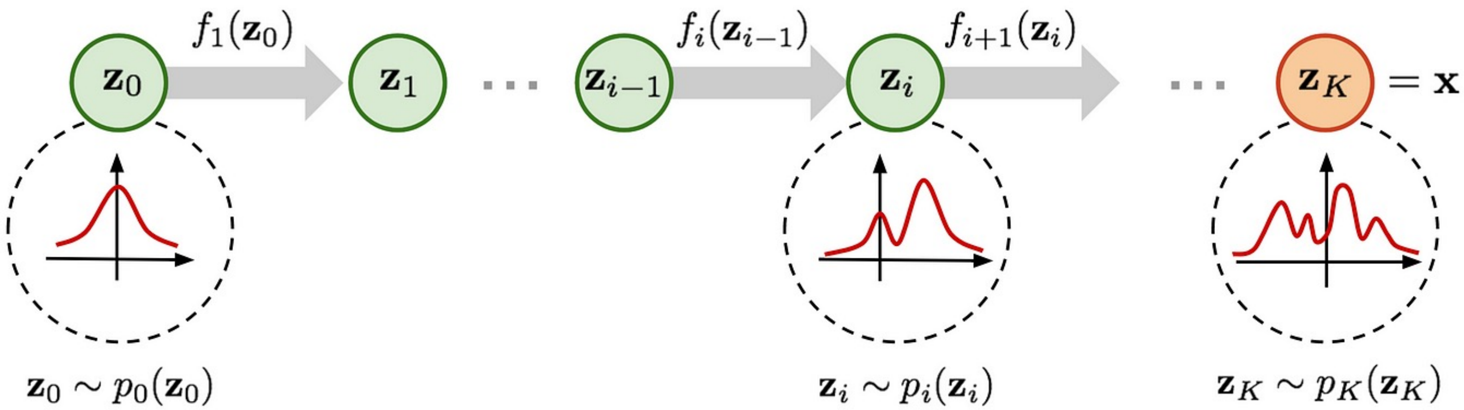
Benefits:

- Can be trained independent of cosmological model
- Only three dimensions
- Quick - we use Hamiltonian MC

What will we use to learn the distribution?

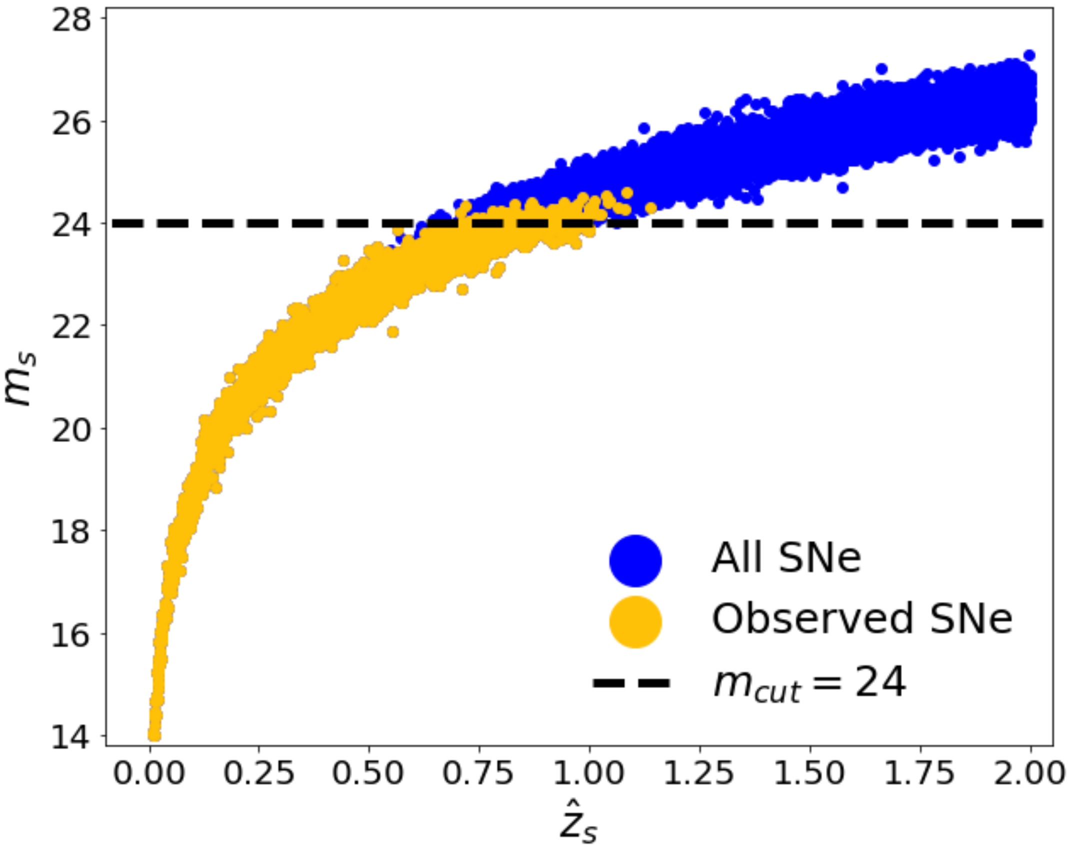
Normalising Flows:

- Transforms a simple Gaussian distribution into a complex distribution.
- The transforms are learnt with a neural network by maximising the log likelihood.



Source: Lilian Weng

Validation Test



Adapted from Boyd et al. (2024)

$$\hat{d}_s = (\hat{m}_s, \hat{c}_s, \hat{x}_s)$$

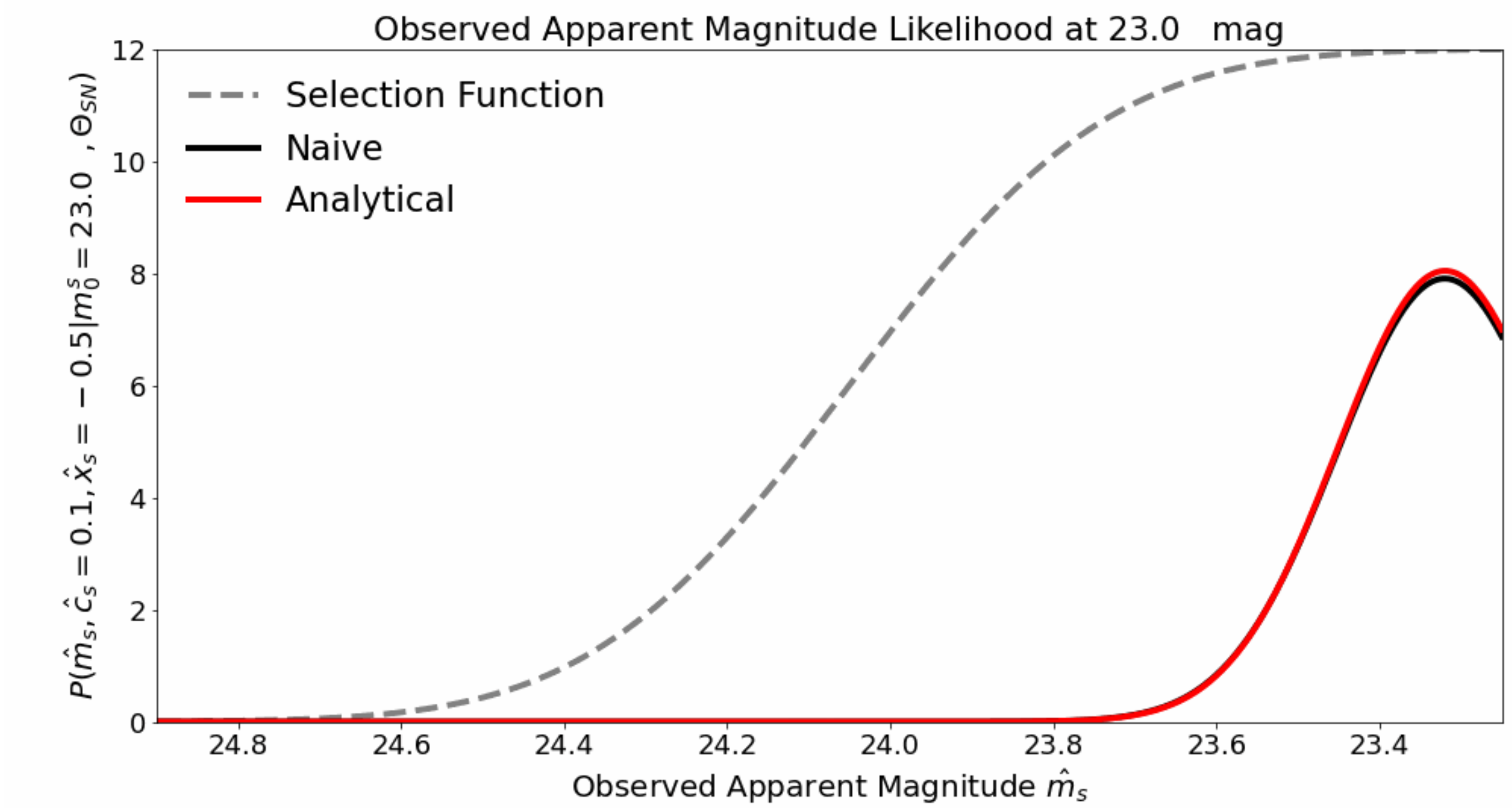
$$\theta = (\Omega_{m0}, w_0, M_0, \alpha, \beta, ..)$$

$$P(Selection|\hat{d}_s) =$$

$$\Phi\left(\frac{m_{cut} - (\hat{m}_s + a\hat{x}_s + b\hat{c}_s)}{\sigma_{cut}}\right)$$

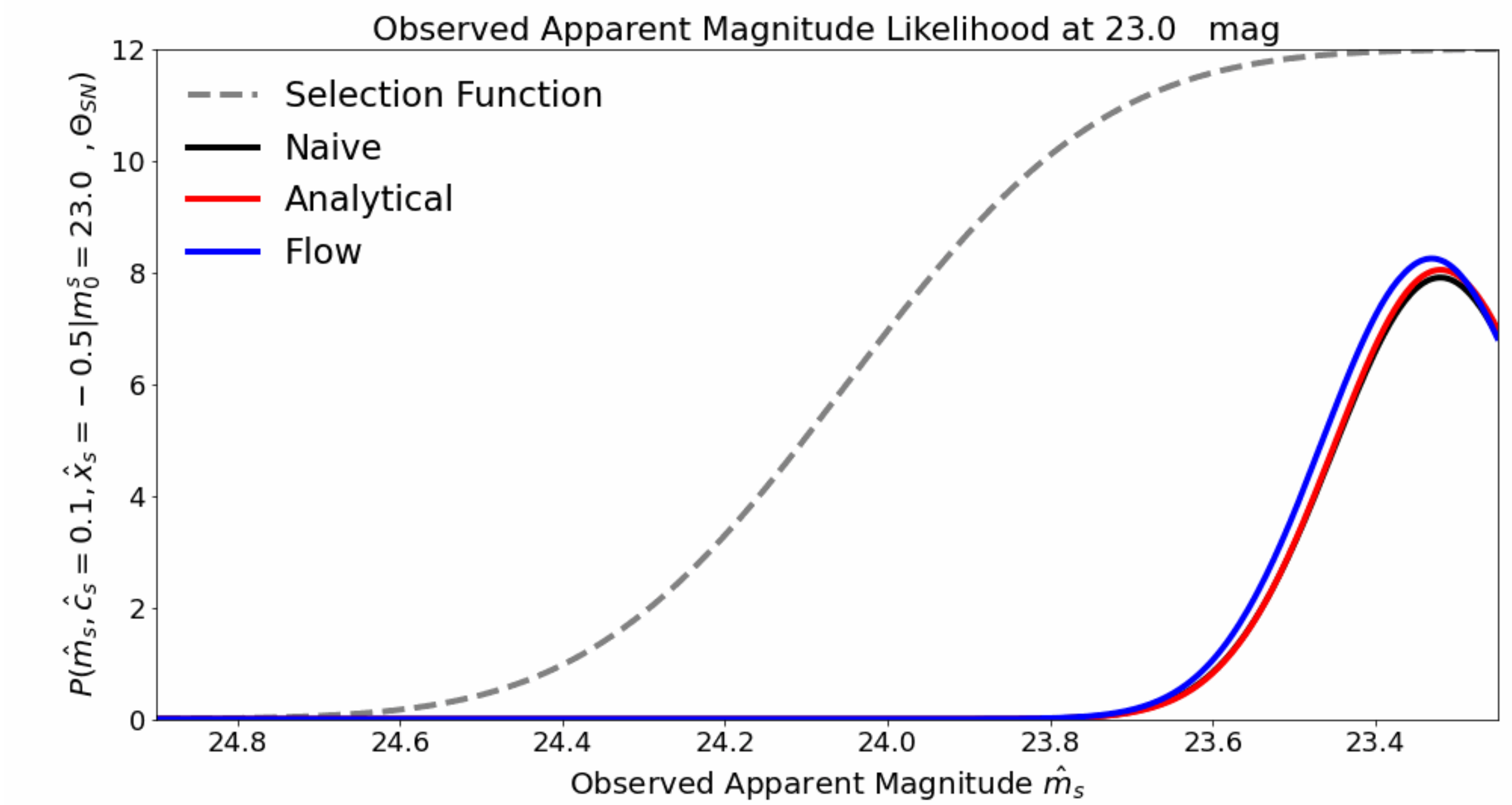
We have the analytical solution for this!

What do we need to learn?



Adapted from Boyd et al. (2024)

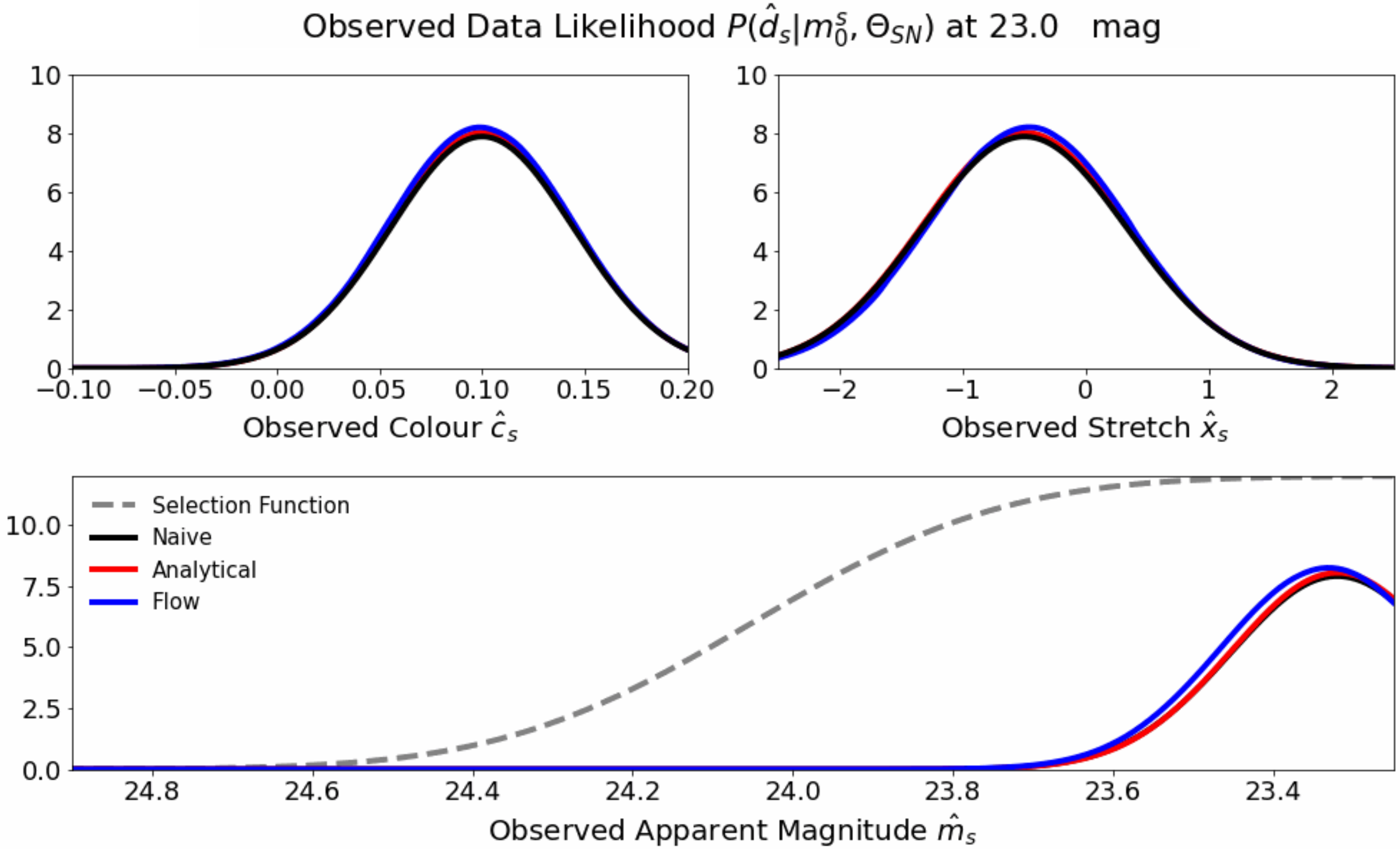
What do we learn?



Adapted from Boyd et al. (2024)

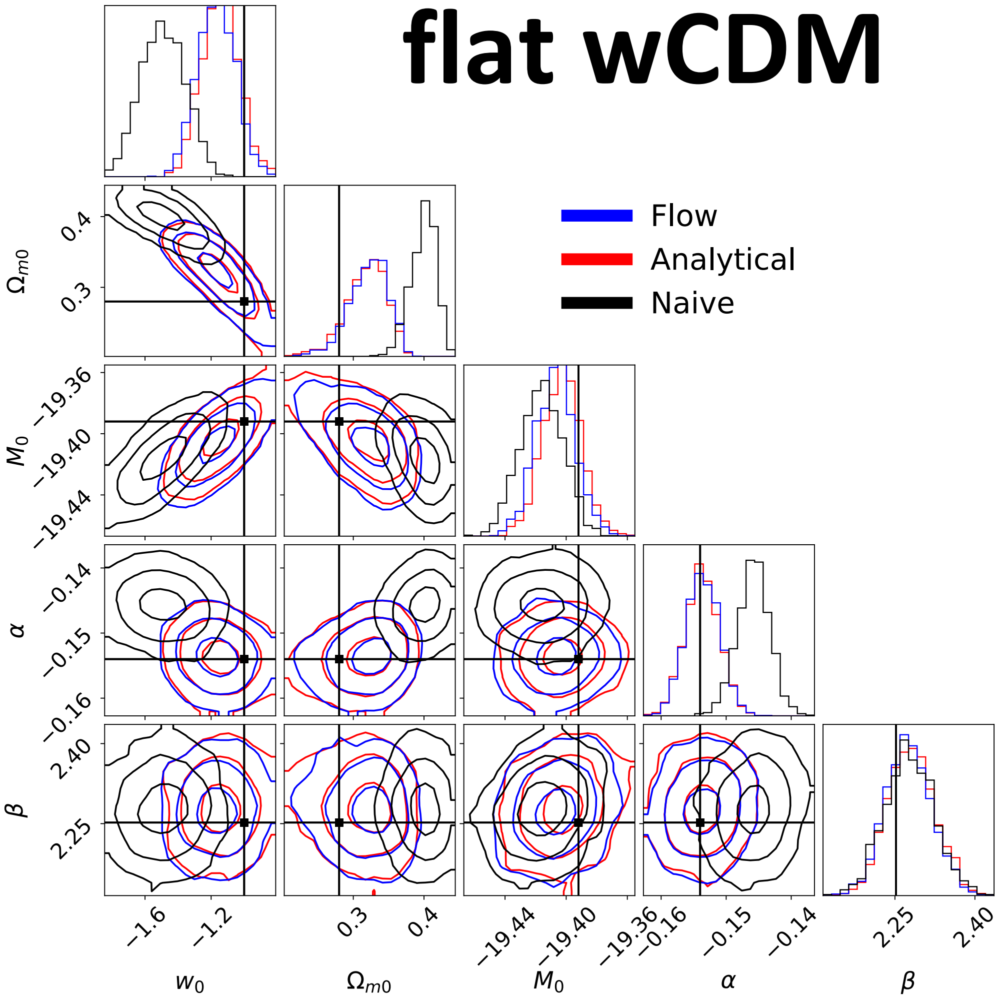
What do we learn?

Adapted from Boyd et al. (2024)



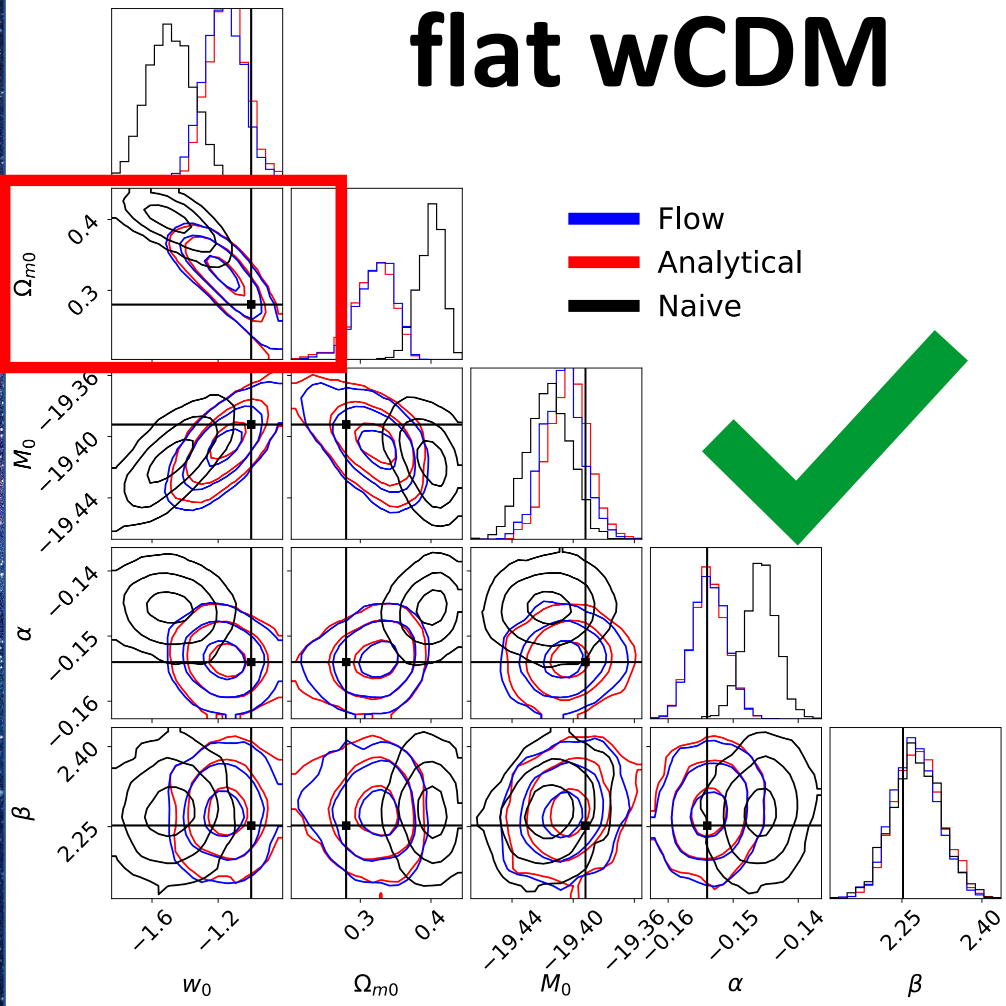
Boyd et al. (2024)

flat wCDM

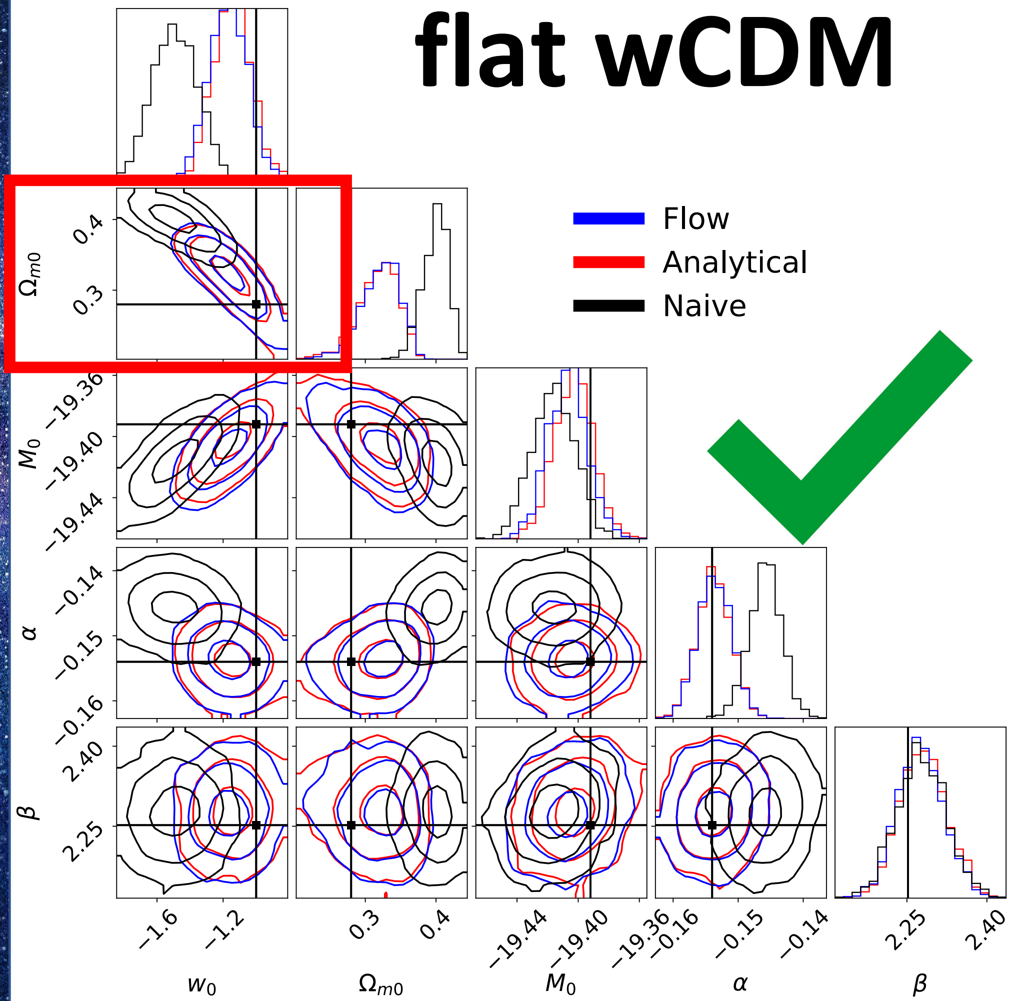


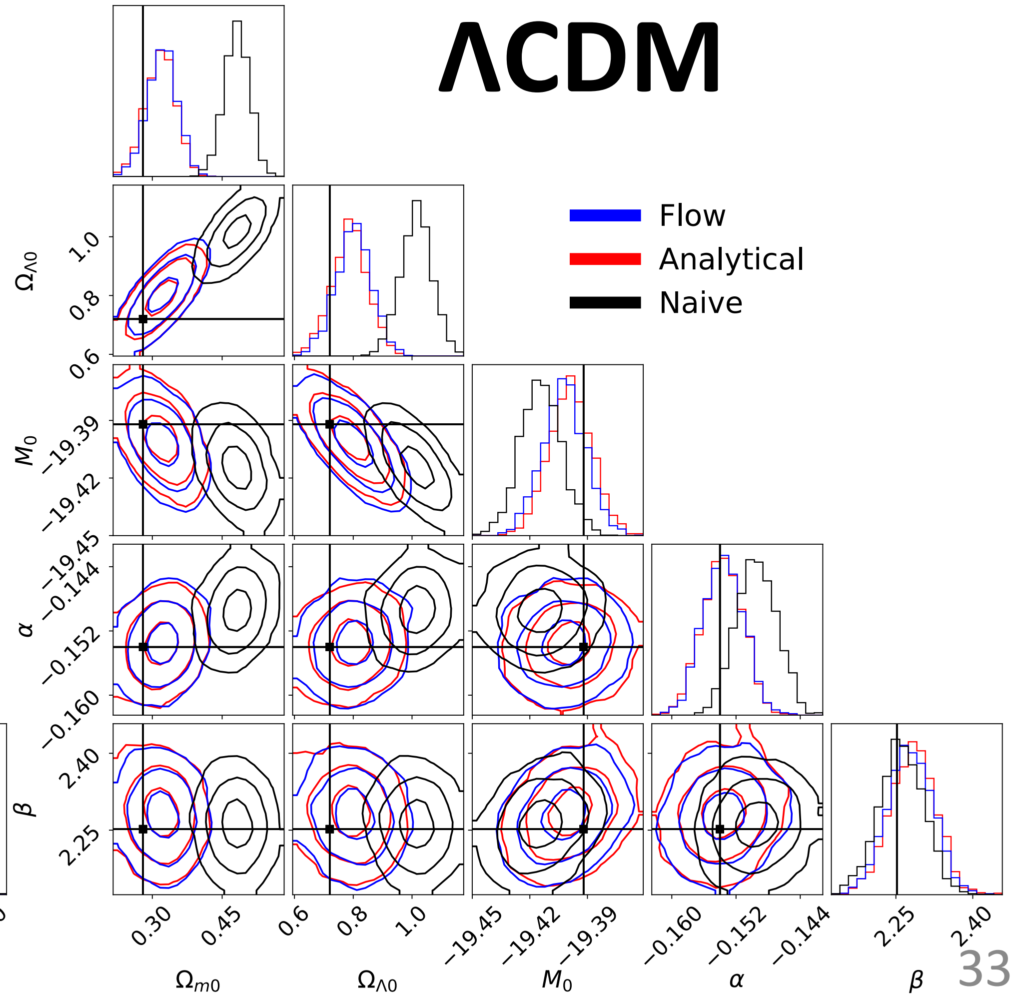
Boyd et al. (2024)

flat wCDM

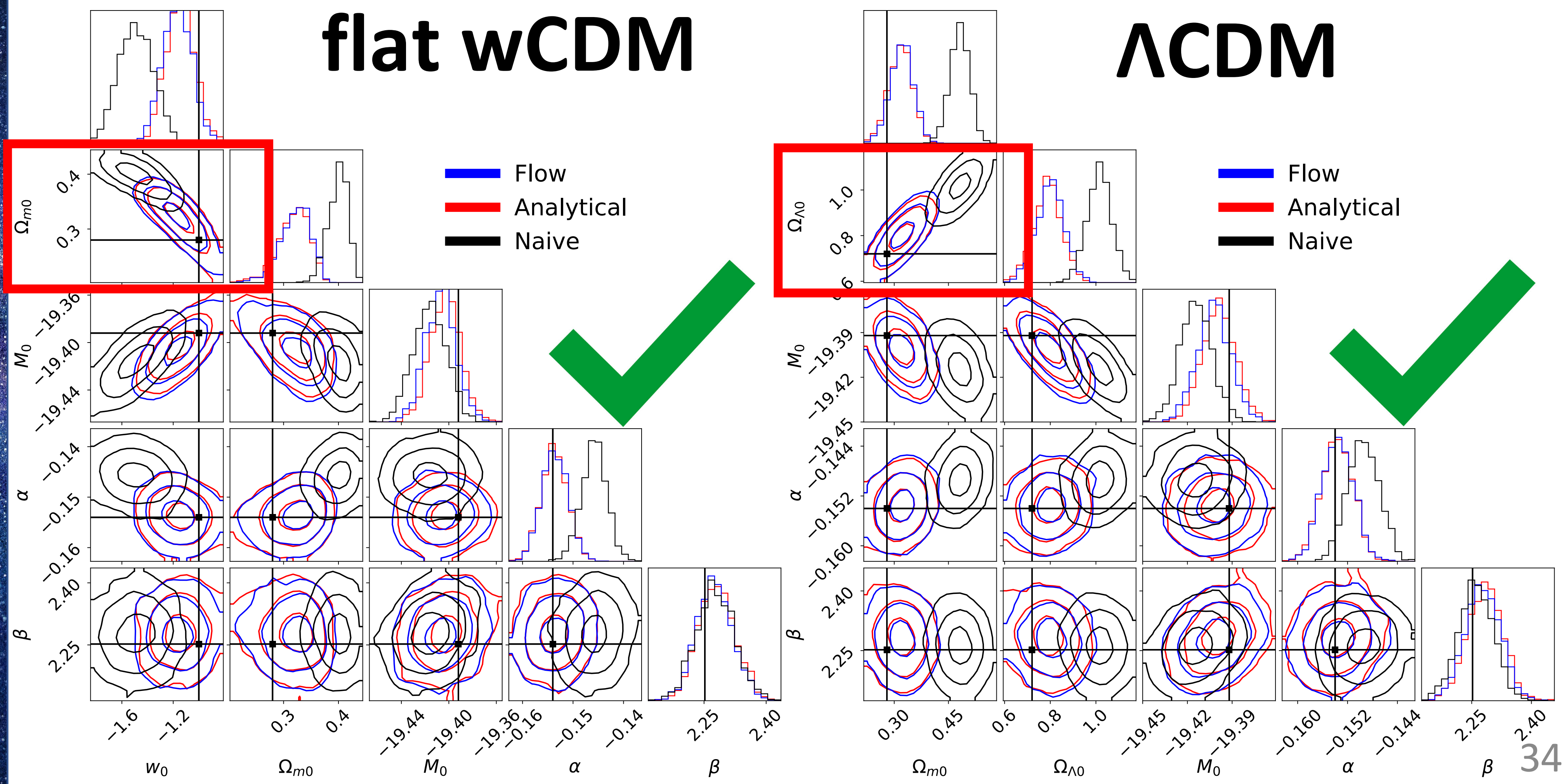


Boyd et al. (2024)



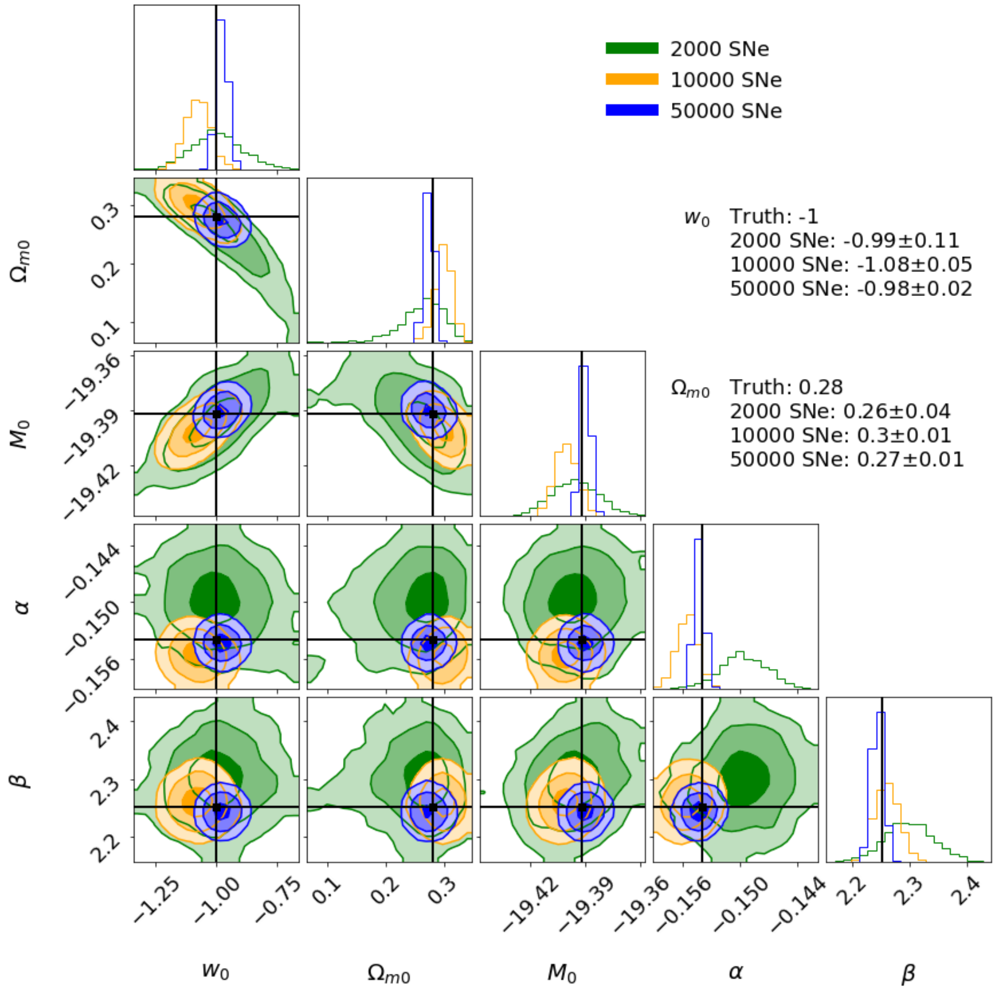


Boyd et al. (2024)



Scalable to Future Spectroscopic Samples

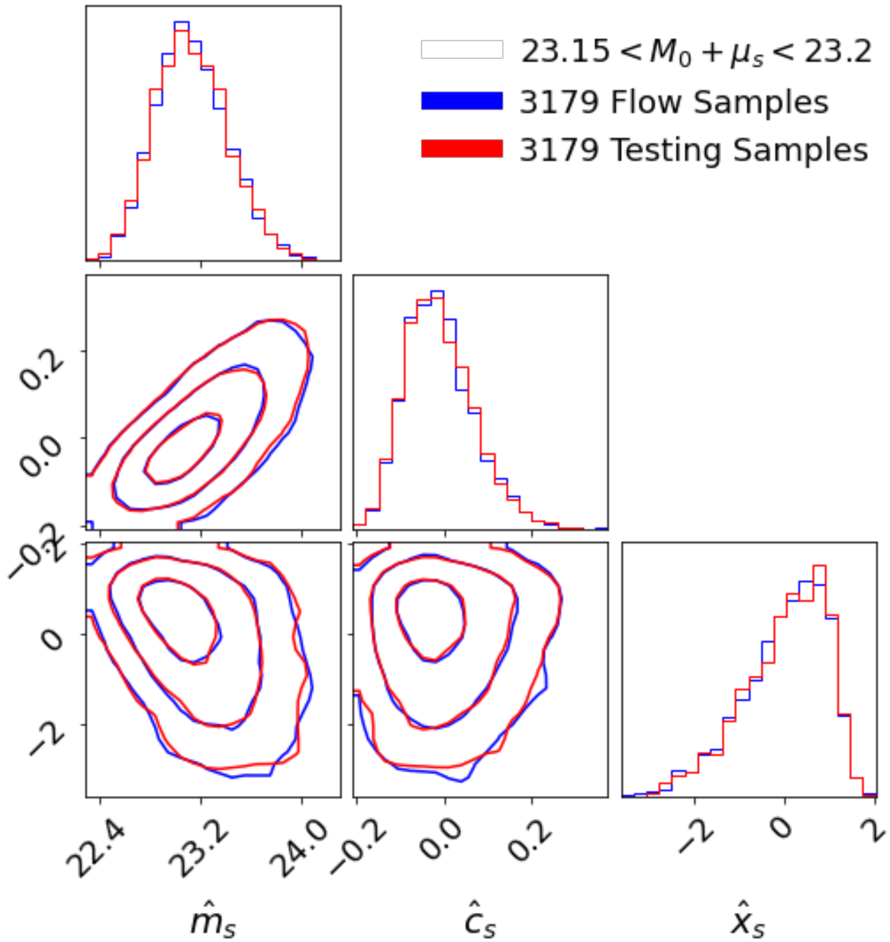
- The normalising flow can marginalise over latent variables allowing is to scale better than regular HBM approaches
- Consistent results with 50,000 SN takes under an hour on GPU

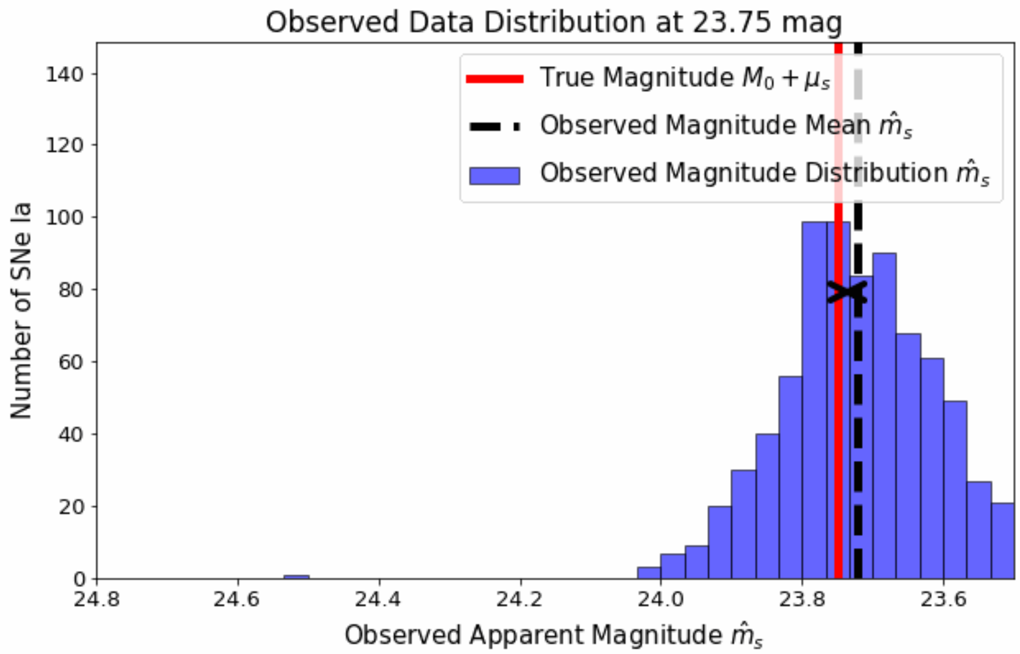


Boyd et al. (2025, in prep)

Train on State-of-the-art SNANA Simulations

(Kessler et al. 2009 SNANA)



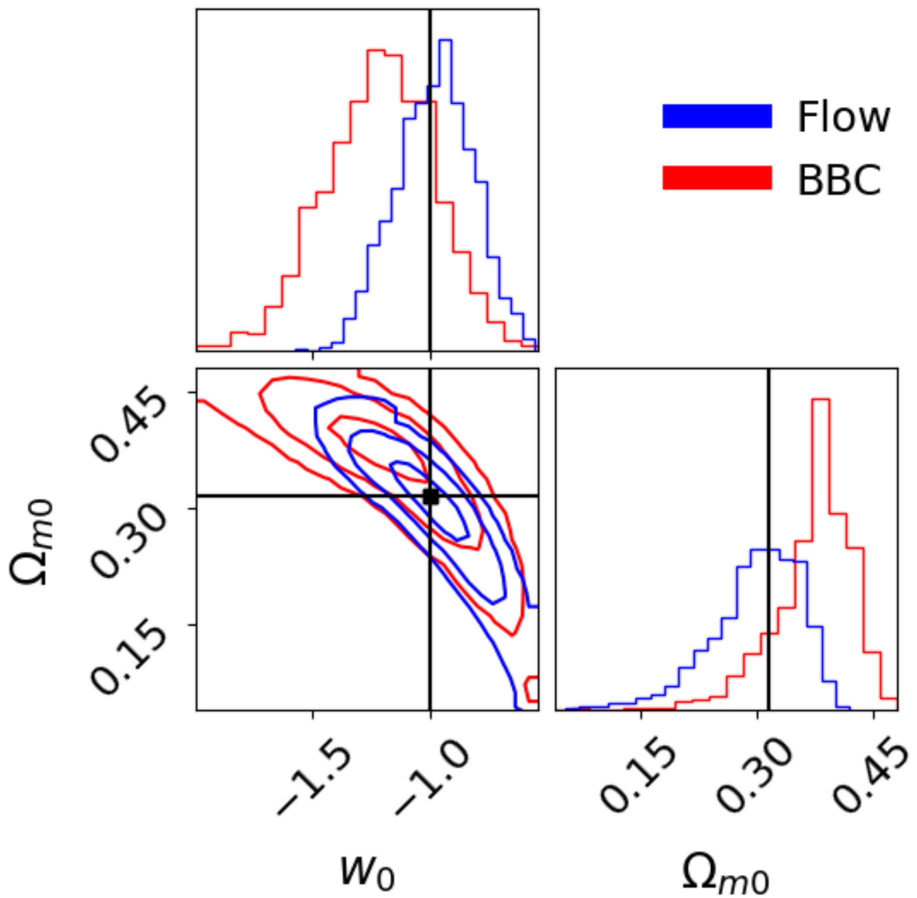


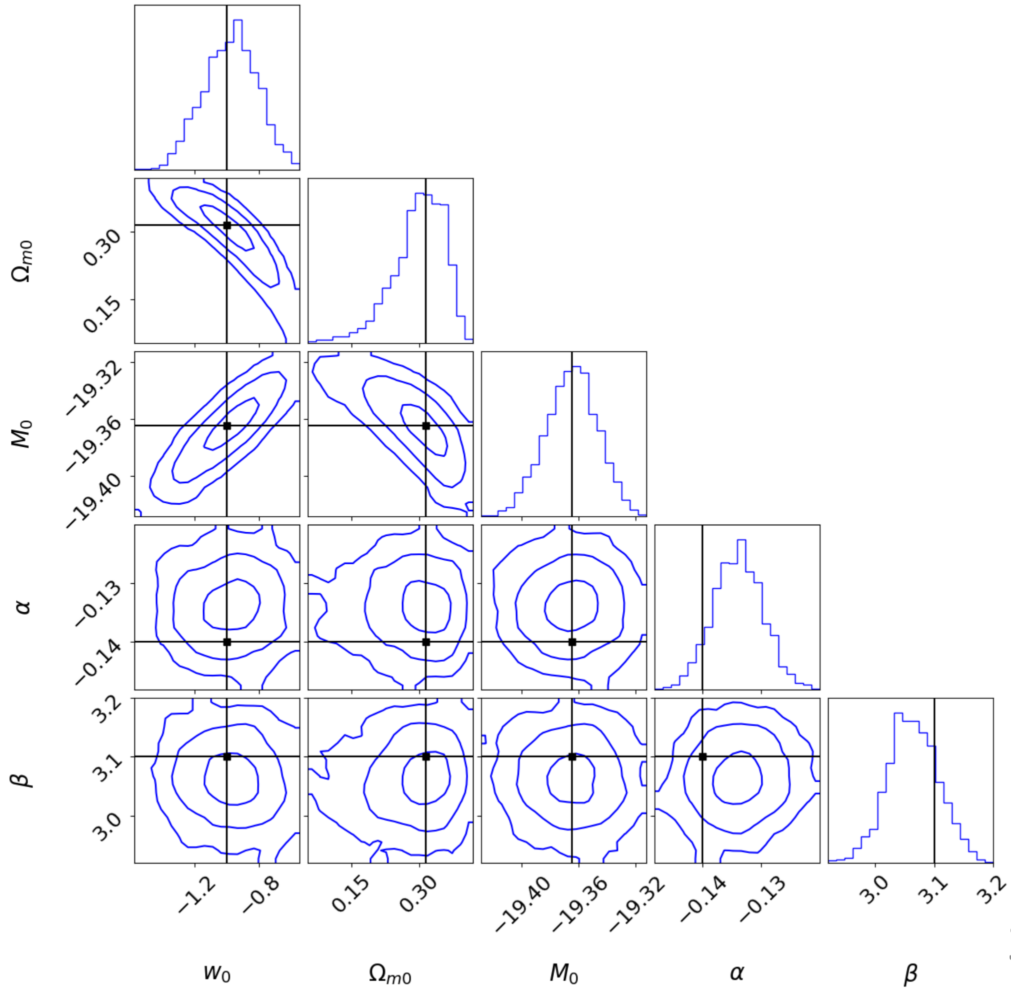
Boyd et al. (2025, in prep.)

Test on State-of-the-art SNANA Simulations

Boyd et al. (2025, in prep)

(Kessler et al. 2009 SNANA)

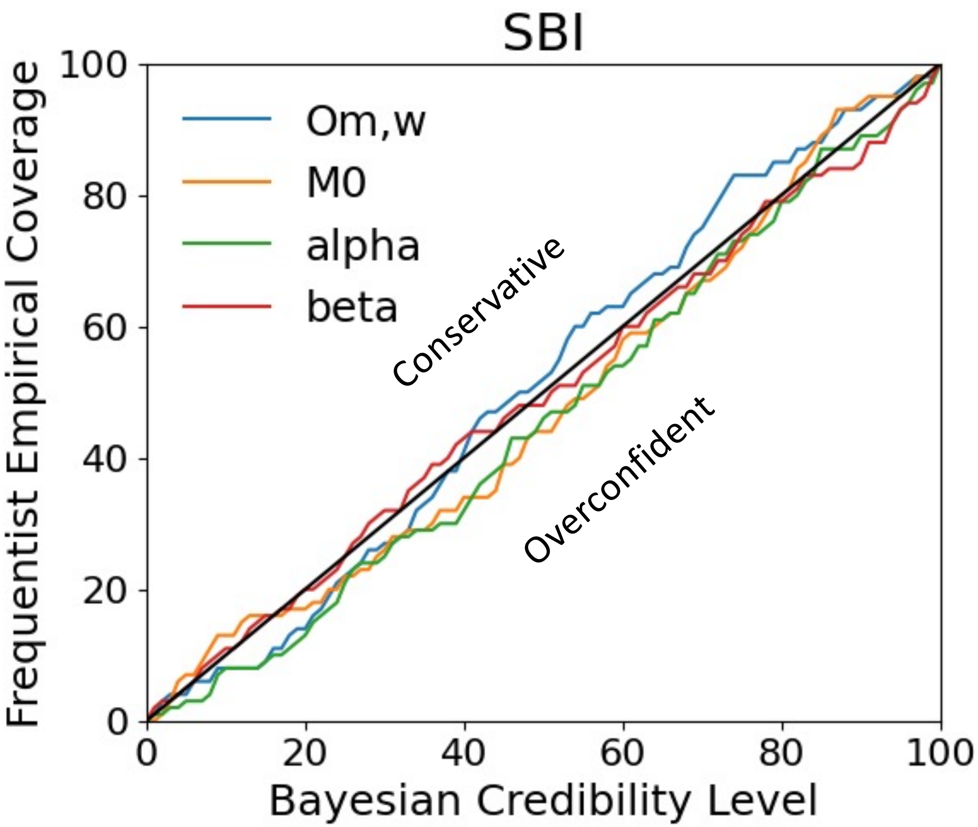


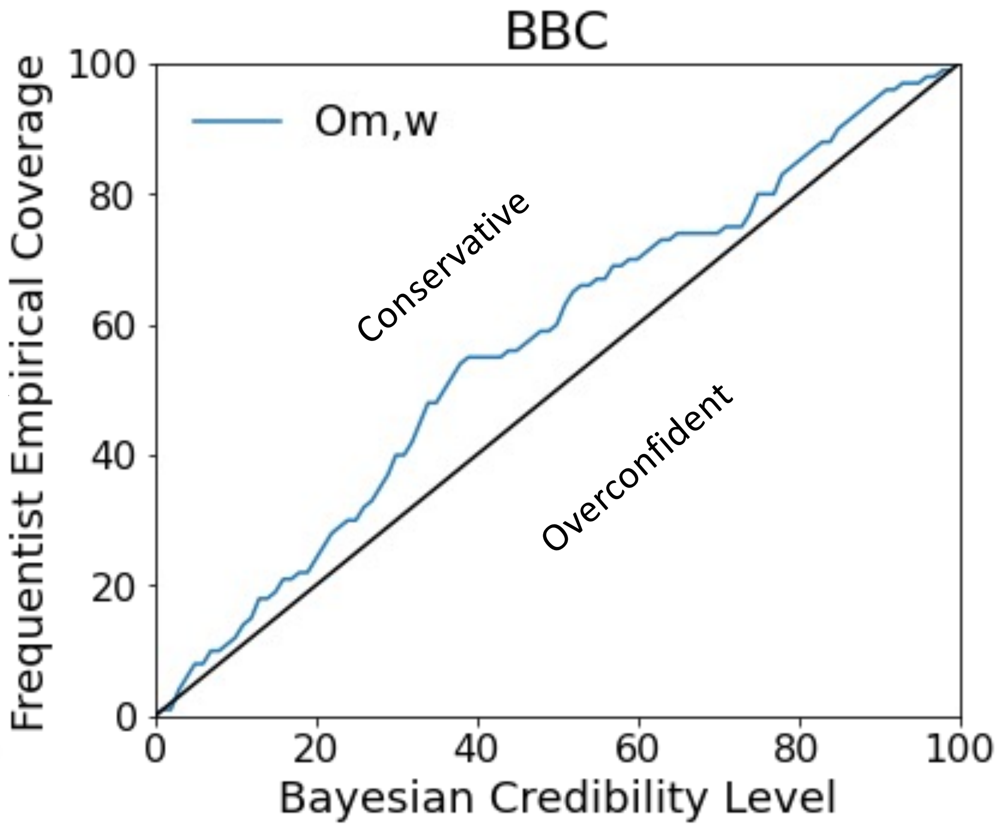


Test on State-of-the-art SNANA Simulations

Boyd et al. (2025, in prep)

(Kessler et al. 2009 SNANA)





Future Work: Dust and Mass-Step

$$P(\hat{\boldsymbol{d}}_s |\ z_s, \boldsymbol{\Theta}_{\mathrm{SN}}, \Omega_{m0}, w_0)$$

We can include dust/mass-step parameters in the likelihood:

$$\boldsymbol{\Theta}_{\mathrm{SN}} = (\alpha, \beta, E_s, R_V, \Delta M, ...)$$

Modular Extensions: Tomographic Cosmology

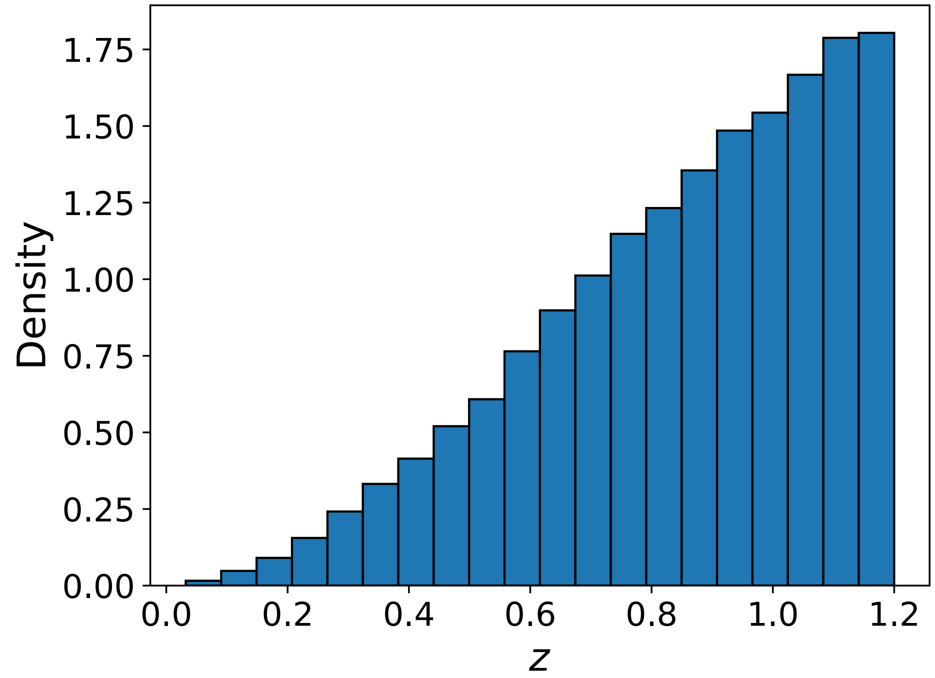
(see Karchev & Trotta 2024)

$$P(\hat{\boldsymbol{d}}_s | \; z_s, \boldsymbol{\Theta}_{\mathrm{SN}}, \Omega_{m0}, w_0) \boxed{\times P(z_s | \Omega_{m0}, w_0, \zeta_{\mathrm{rate}})}$$

Modular Extensions: Tomographic Cosmology

(see Karchev & Trotta 2024)

$$P(\hat{\boldsymbol{d}}_s|\ z_s, \boldsymbol{\Theta}_{\rm SN}, \Omega_{m0}, w_0) \boxed{\times\ P(z_s|\Omega_{m0}, w_0, \zeta_{\rm rate})}$$

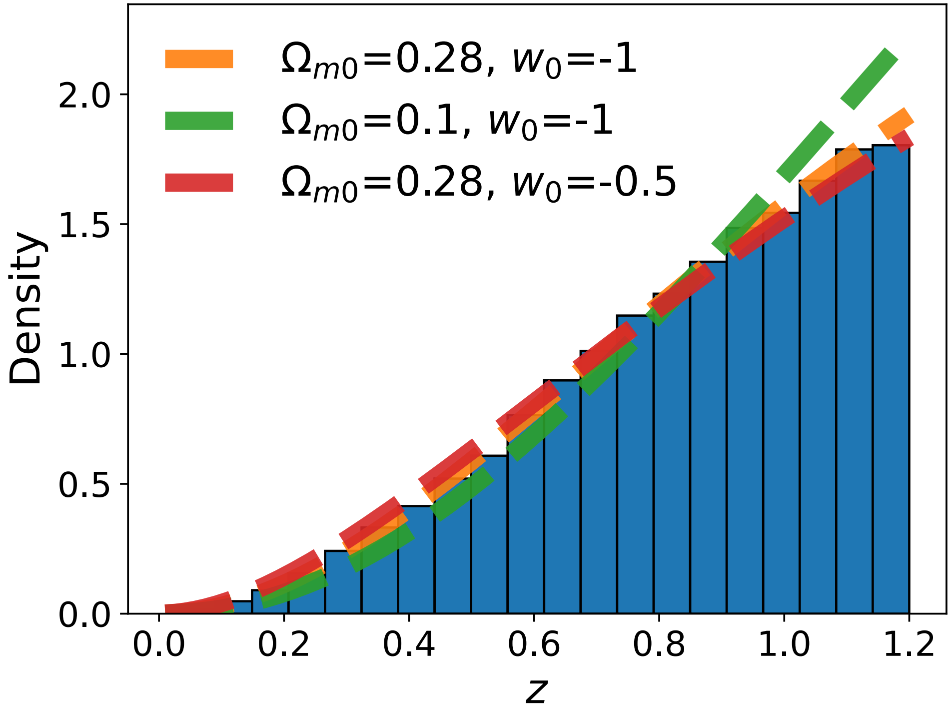


Boyd et al. (2025, in prep.)

Modular Extensions: Tomographic Cosmology

(see Karchev & Trotta 2024)

$$P(\hat{\boldsymbol{d}}_s|\ z_s, \boldsymbol{\Theta}_{\mathrm{SN}}, \Omega_{m0}, w_0) \boxed{\times P(z_s|\Omega_{m0}, w_0, \zeta_{\mathrm{rate}})}$$

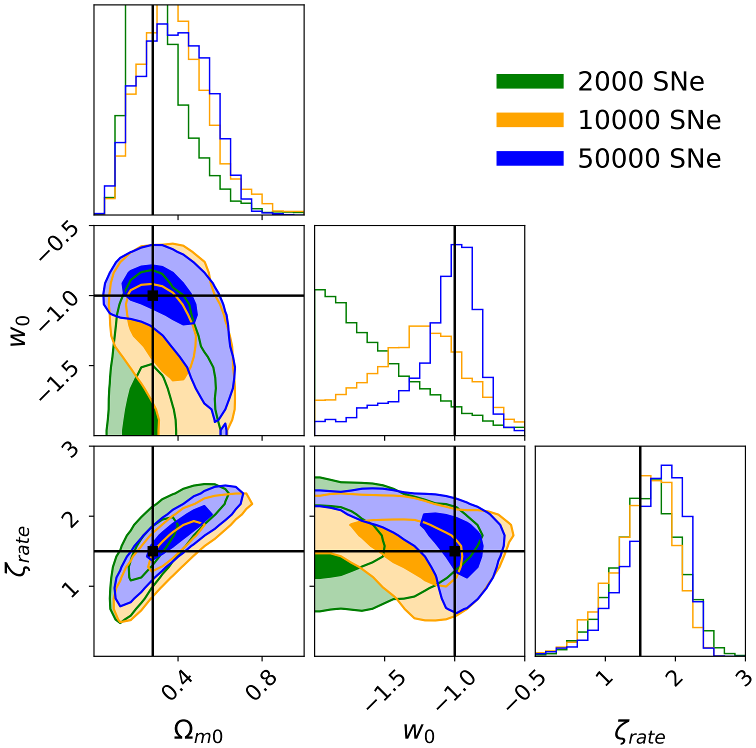


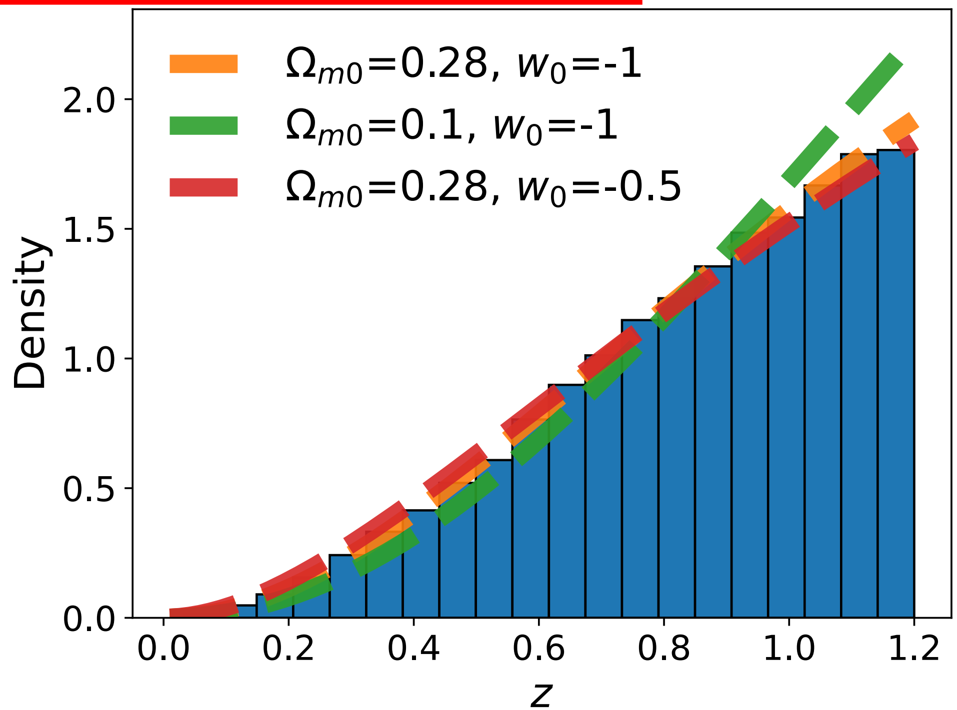
Boyd et al. (2025, in prep.)

Modular Extensions: Tomographic Cosmology

(see Karchev & Trotta 2024)

$$P(\hat{\boldsymbol{d}}_s|\ z_s, \boldsymbol{\Theta}_{\mathrm{SN}}, \Omega_{m0}, w_0) \boxed{\times\ P(z_s|\Omega_{m0}, w_0, \zeta_{\mathrm{rate}})}$$





Boyd et al. (2025, in prep.)

Modular Extensions: Photometric Redshifts

$$P(\hat{\boldsymbol{d}}_s|\ z_s, \boldsymbol{\Theta}_{\rm SN}, \Omega_{m0}, w_0) \times P(z_s|\Omega_{m0}, w_0, \zeta_{\rm rate}) \times P(\hat{z}_s|z_s)$$

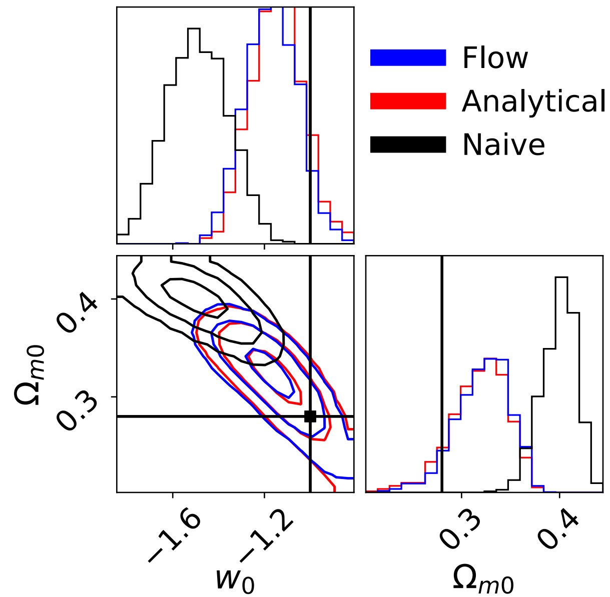
Modular Extensions: Photometric Redshifts

$$P(\hat{\boldsymbol{d}}_s|\ z_s, \boldsymbol{\Theta}_{\mathrm{SN}}, \Omega_{m0}, w_0) \times P(z_s|\Omega_{m0}, w_0, \zeta_{\mathrm{rate}}) \boxed{\times P(\hat{z}_s|z_s)}$$

Can also include photometric redshift likelihood as we move towards LSST

Summary:

- We have developed a flexible method that can learn intractable likelihoods that account for Malmquist bias.
- These are quick to learn and low dimensional.
- We can then use these learnt likelihoods in our hierarchical Bayesian analysis for constrains on different cosmological models.



Boyd et al. (2024)
SBI+HBM arXiv

What is a Hierarchical Bayesian Model?

$$\boldsymbol{d_s} = (d_1^s, d_2^s)$$

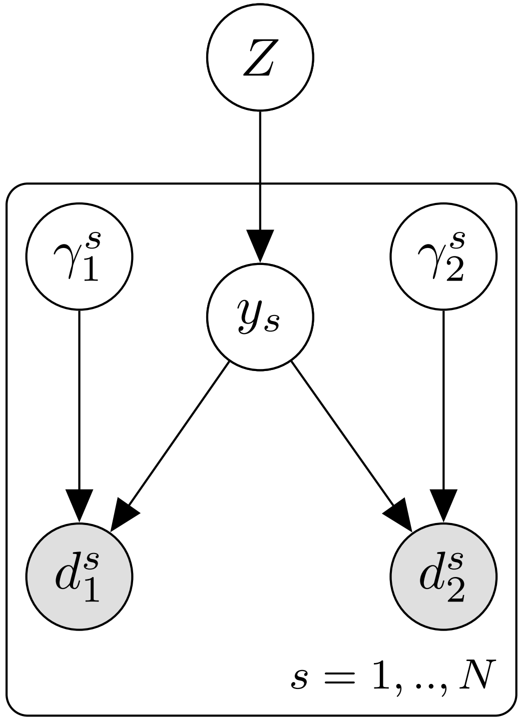
$$\boldsymbol{\theta} = (\{\gamma_1^s, \gamma_2^s, y_s\}, Z)$$

$$P(\boldsymbol{\theta}|\{\boldsymbol{d_s}\}) \propto P(\{\boldsymbol{d_s}\}|\boldsymbol{\theta})P(\boldsymbol{\theta})$$

$$= \Big[\prod_{s=1}^{N} P(d_s|\theta)\Big]P(\theta)$$

$$= \Big[\prod_{s=1}^{N} P(d_1^s|\gamma_1^s, y_s, Z)P(d_2^s|\gamma_2^s, y_s, Z)$$

$$\times P(\gamma_1^s)P(\gamma_2^s)P(y_s)\Big]P(Z)$$



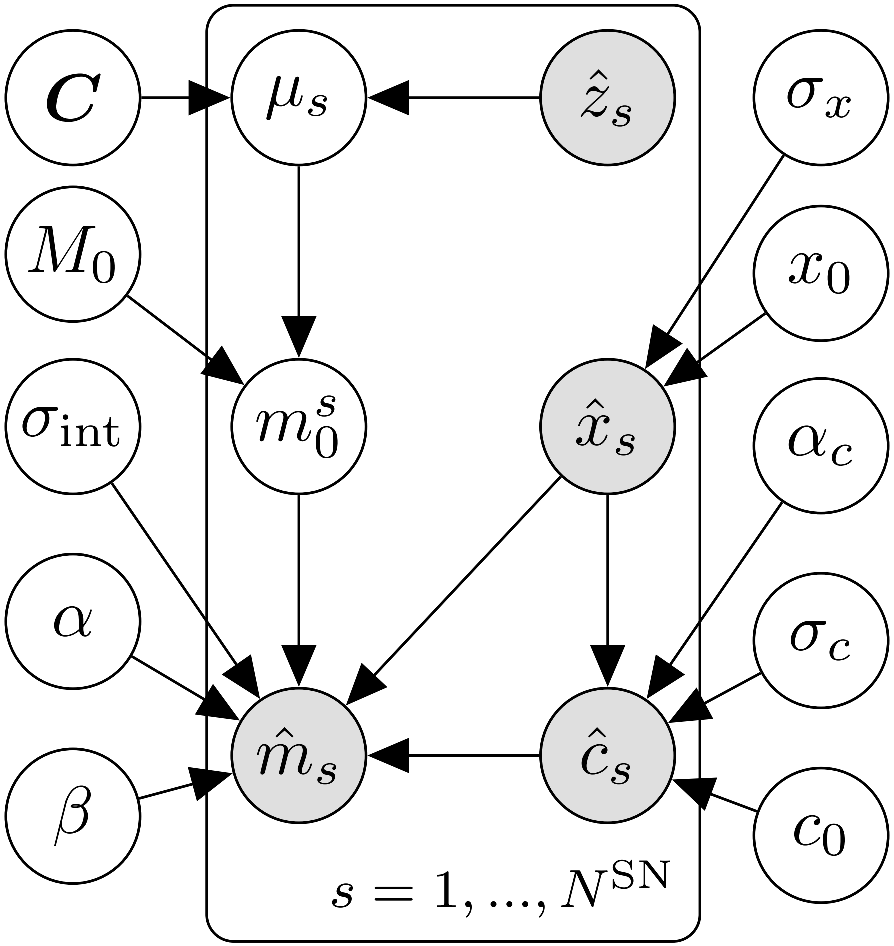
Graphical Model Example

Full Model likelihood:

$$P(\hat{d}_s|\theta) =$$
$$N(\hat{d}_s|d_s, W_s)$$
$$N(m_s\,|\mu_s(z_s|\Omega_{m0}, w_0) + M_0 + \alpha\, x_s\, + \beta\, c_s\,, \sigma_{int}^2)$$
$$N(c_s\,|c_0 + \alpha_c\, x_s\,, \sigma_c^2)\; N(x_s\,|x_0, \sigma_x^2)$$
$$\Phi\left(selection|\,\frac{m_{cut} - (\widehat{m}_s + a\hat{x}_s + b\hat{c}_s\,)}{\sigma_{cut}}\right)$$

$$Where$$
$$\hat{d}_s = (\widehat{m}_s, \hat{c}_s, \hat{x}_s)$$
$$d_s = (m_s, c_s, x_s)$$
W_s *is the LC summary measurement covariance matrix*

Hierarchical Model:



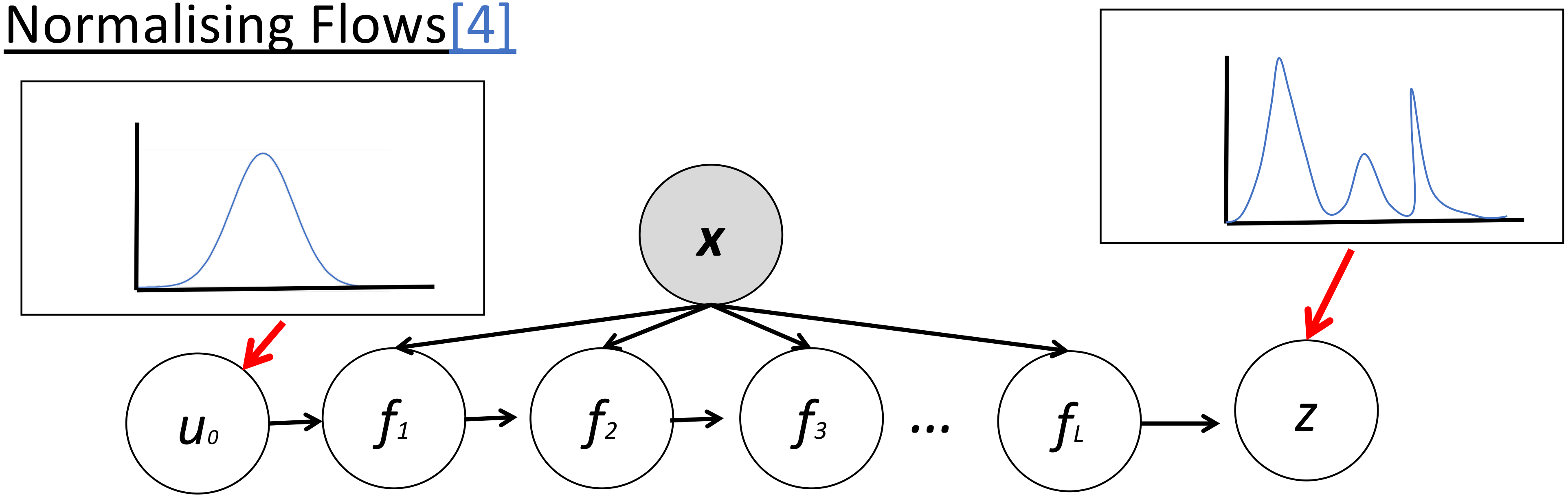
Using the same likelihood for different cosmological models

The reason we can use the same data likelihood for different cosmologies without retraining/redefining is because $\widehat{m}_s$ is conditionally independent from $\mu_s(z_s|\ C)$, so we can write:

$$P(\widehat{m}_s|\ \mu_s(z_s|\ C) + M_0) = P(\widehat{m}_s|\ m_0)\ P(m_0|\mu_s(z_s|\ C)+ M_0)$$

Where $P(m_0 = \mu_s(z_s|\ C)+ M_0)$=1 so is a deterministic function. This means we only need to give the likelihood the latent observed apparent magnitude and not the cosmological parameters directly.

Normalising Flows[4]



f needs to be*:*

- Bijective
- Easily invertible
- Continuously differentiable.

$$q_{\phi}(\mathbf{x} \mid \boldsymbol{\theta}) = \mathcal{N}(\mathbf{z}_0 \mid \mathbf{0}, \mathbf{I}) \prod_k \left| \det\left(\frac{\partial f_k}{\partial \mathbf{z}_{k-1}} \right) \right|^{-1}$$

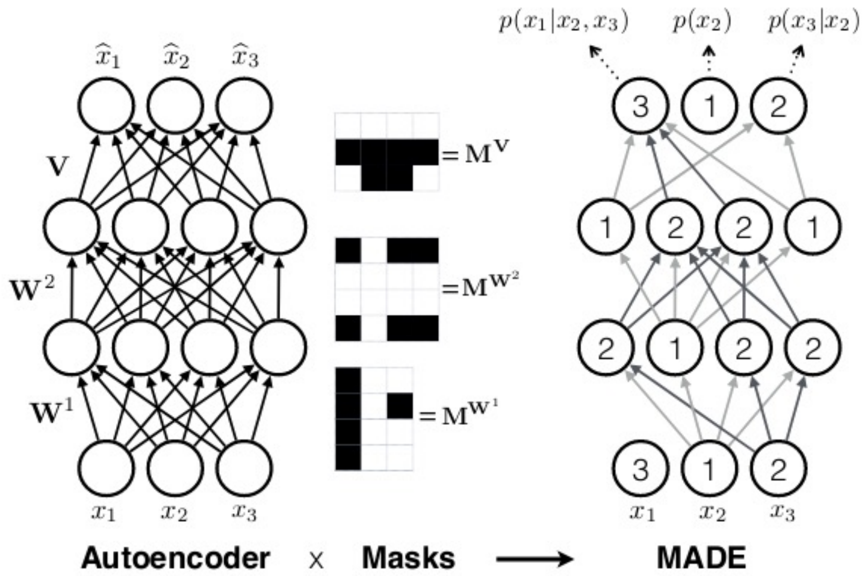
$$\hat{p}(\boldsymbol{\theta} \mid \mathbf{x}_o) \propto q_{\phi}(\mathbf{x}_o \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta}),$$

[4] C. E. G. Tabak and E. Vanden-Eijnden, "Density estimation by dual ascent of the log-likelihood," Communications in Mathematical Sciences, vol. 8, no. 1, pp. 217–233, 2010.

Masked Autoregressive Flow

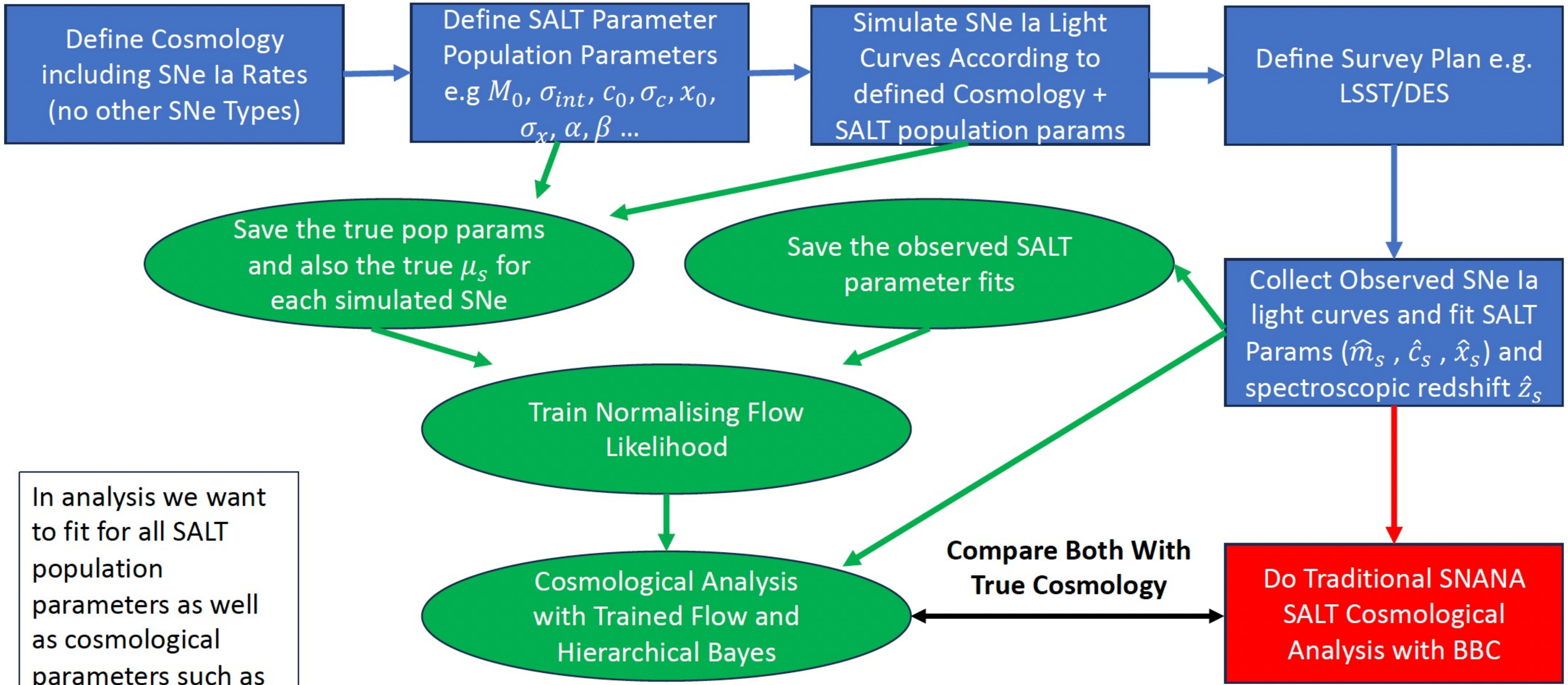
Papamakarios et al. (2017)

- Masked Autoregressive Flows (MAFs) chain together a series of MADE blocks that slowly transforms a simple multivariate distribution into a complex distribution.
- MADE blocks use clever architectures where each dimension is dependent on the previous dimension.
- This makes the Jacobian lower triangular, allowing the determinant of the transform to be be computed easily.



Masked Autoencoder for Distribution Estimation (MADE).
Germain et al. (2015)

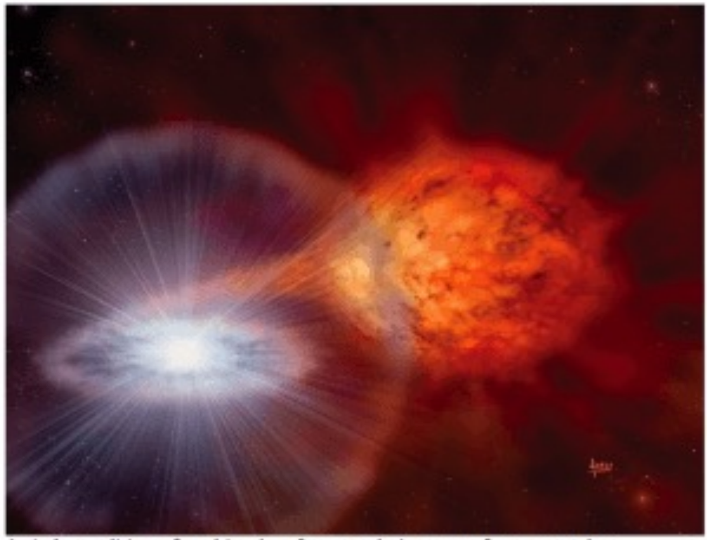
Survey Simulation Analysis Comparing SBI and BBC



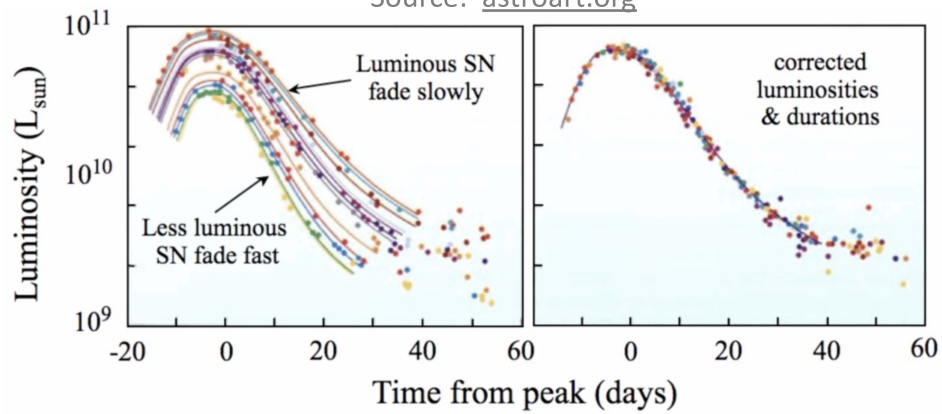
In analysis we want to fit for all SALT population parameters as well as cosmological parameters such as Ω_m and w

Type Ia Supernovae:

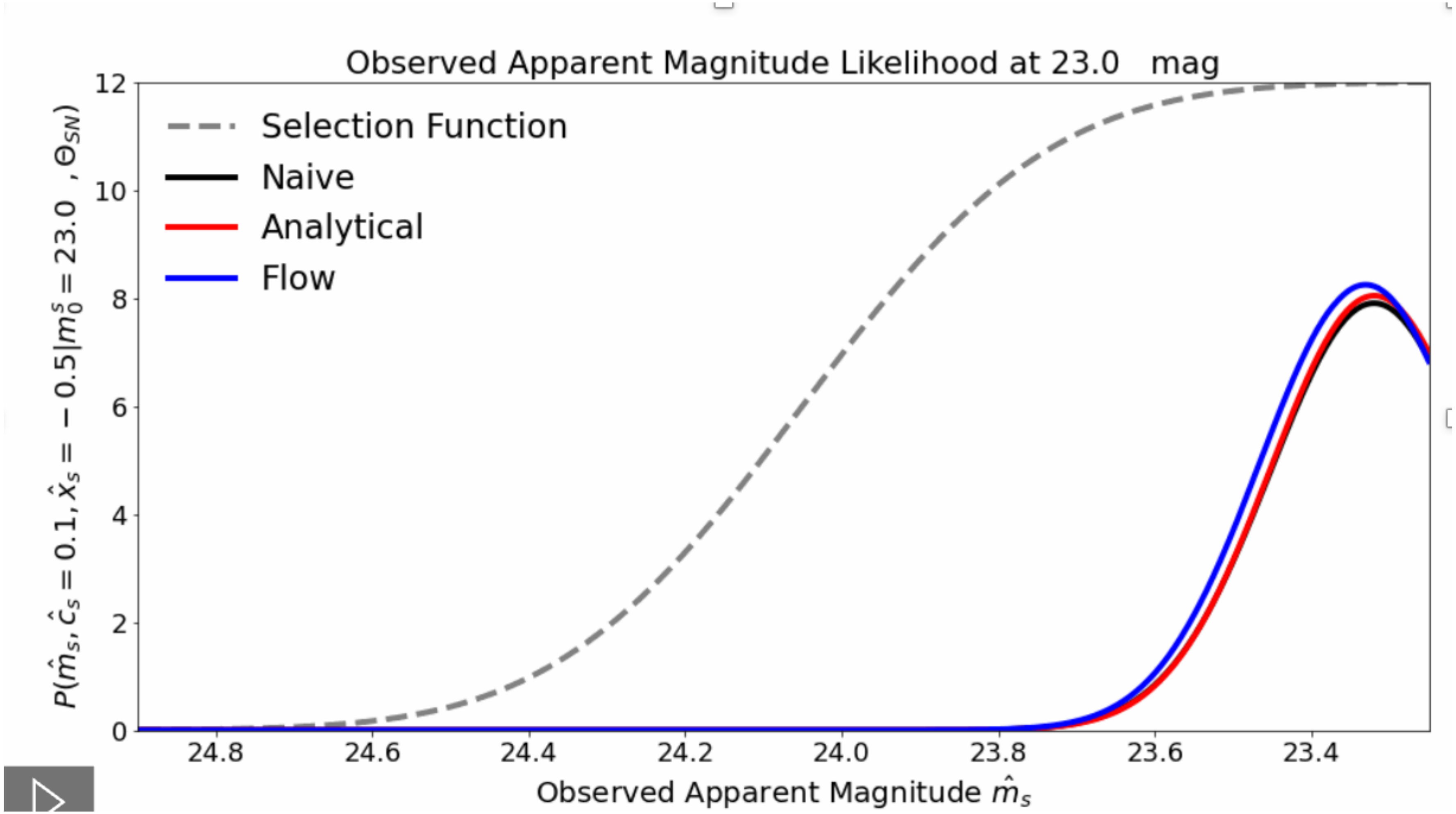
- Thermonuclear explosions from a binary system involving at least one white dwarf.
- Easy to identify (spectroscopically) with their distinctive silicon absorption feature.
- Light curves are remarkably consistent meaning we can standardise them to use for distances!

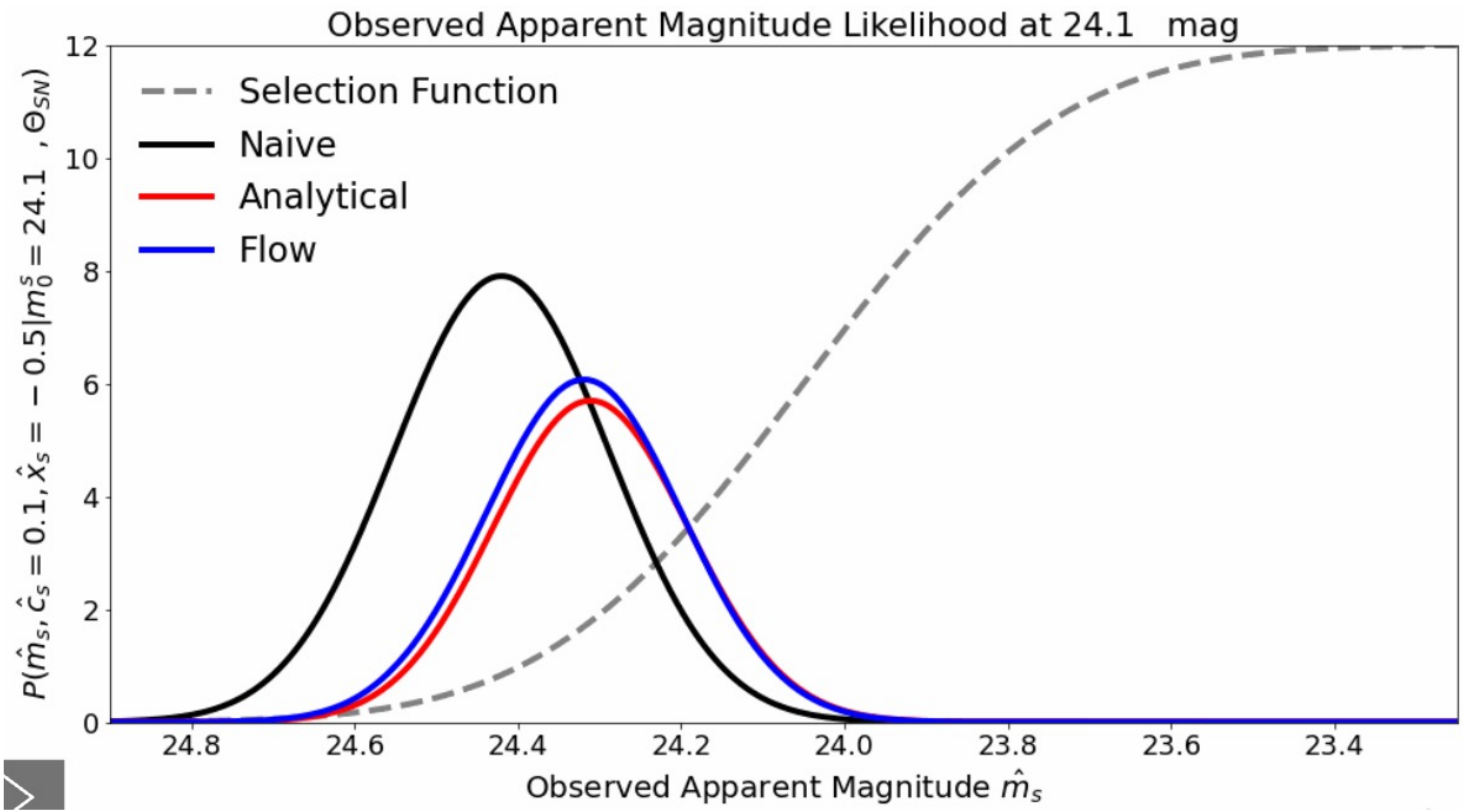


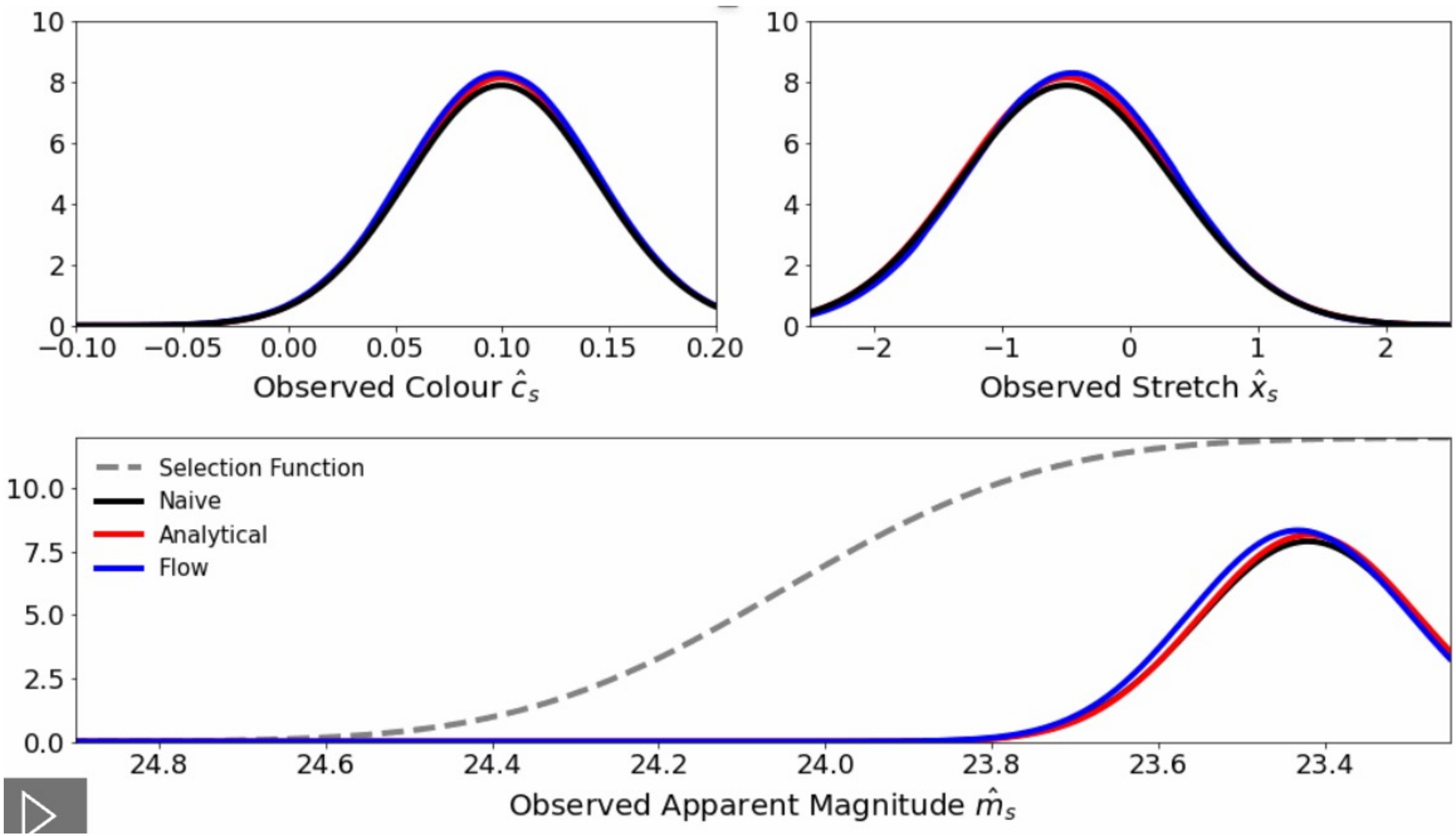
Source: astroart.org



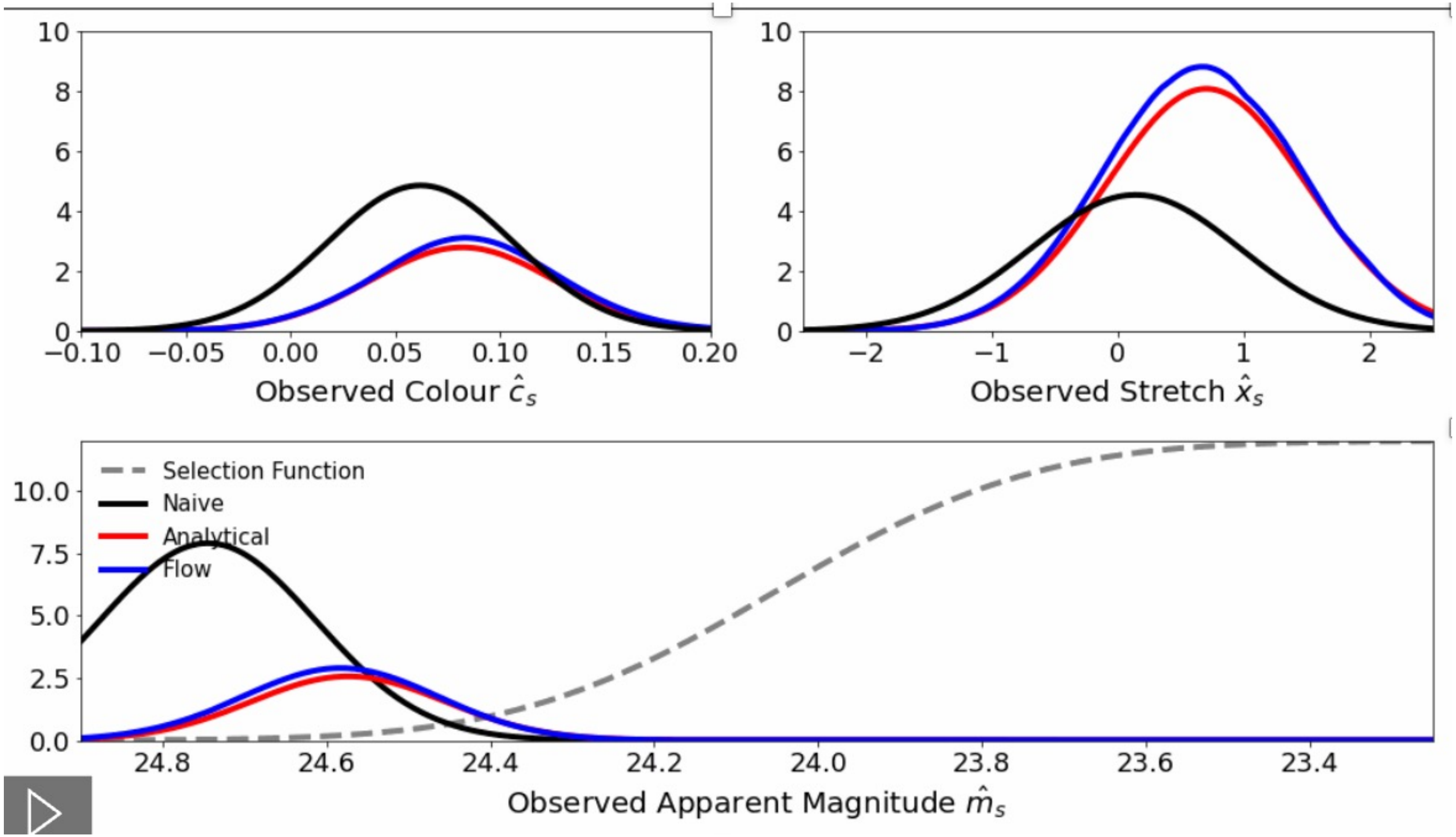
Source: 10.1109/EDUCON45650.2020.9125209


Observed Apparent Magnitude Likelihood at 23.0 mag
Selection Function
Naive
Analytical
Flow
$P(\hat{m}_s, \hat{c}_s = 0.1, \hat{x}_s = -0.5 | m_0^s = 23.0 \ , \Theta_{SN})$
0
2
4
6
8
10
12
24.8
24.6
24.4
24.2
24.0
23.8
23.6
23.4
Observed Apparent Magnitude $\hat{m}_s$


Observed Apparent Magnitude Likelihood at 24.1 mag
Selection Function
Naive
Analytical
Flow
$P(\hat{m}_s, \hat{c}_s = 0.1, \hat{x}_s = -0.5 | m_0^s = 24.1 \ , \Theta_{SN})$
0
2
4
6
8
10
12
24.8
24.6
24.4
24.2
24.0
23.8
23.6
23.4
Observed Apparent Magnitude $\hat{m}_s$



10
8
6
4
2
0
−0.10
−0.05
0.00
0.05
0.10
0.15
0.20
Observed Colour $\hat{c}_s$
10
8
6
4
2
0
−2
−1
0
1
2
Observed Stretch $\hat{x}_s$
Selection Function
Naive
Analytical
Flow
10.0
7.5
5.0
2.5
0.0
24.8
24.6
24.4
24.2
24.0
23.8
23.6
23.4
Observed Apparent Magnitude $\hat{m}_s$



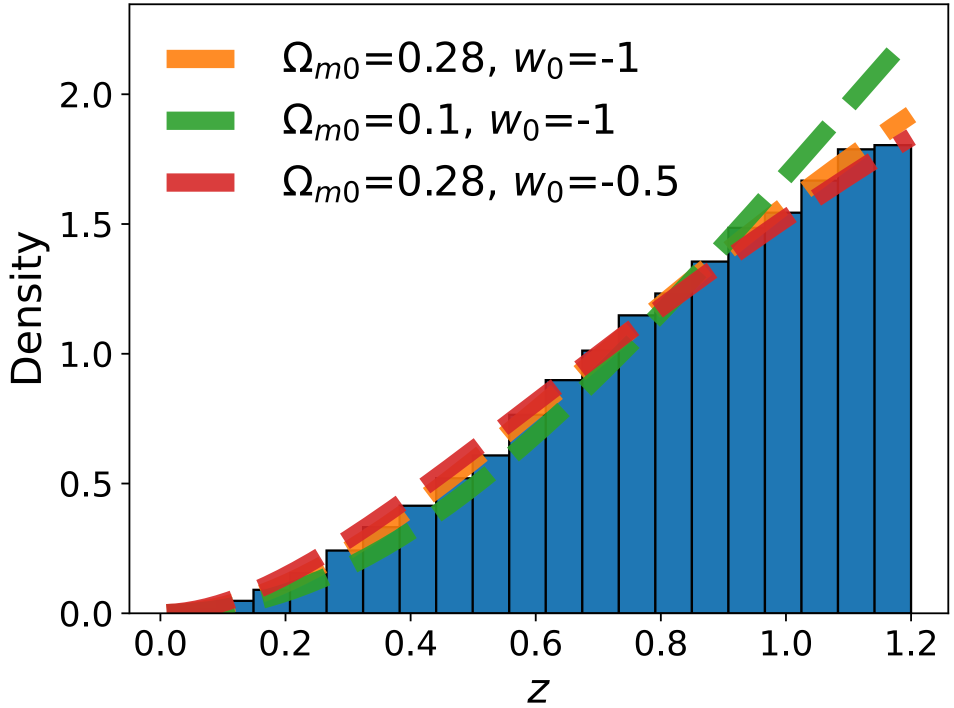
10
8
6
4
2
0
−0.10
−0.05
0.00
0.05
0.10
0.15
0.20
Observed Colour $\hat{c}_s$
10
8
6
4
2
0
−2
−1
0
1
2
Observed Stretch $\hat{x}_s$
Selection Function
Naive
Analytical
Flow
10.0
7.5
5.0
2.5
0.0
24.8
24.6
24.4
24.2
24.0
23.8
23.6
23.4
Observed Apparent Magnitude $\hat{m}_s$

Modular Extensions: Tomographic Cosmology

(see Karchev & Trotta 2024)

$$P(\hat{\boldsymbol{d}}_s|\ z_s, \boldsymbol{\Theta}_{\text{SN}}, \Omega_{m0}, w_0) \times P(z_s|\Omega_{m0}, w_0, \zeta_{\text{rate}})$$

$$P(z_s|\Omega_{m0}, w_0, \zeta_{\text{rate}}) \propto \frac{(z_s+1)^{\zeta_{\text{rate}}}}{1+z_s} \left| \frac{\mathrm{d}V_c(z_s, \Omega_{m0}, w_0)}{\mathrm{d}z_s} \right|$$



Modular Extensions: Tomographic Cosmology

(see Karchev & Trotta 2024)

$$P(\hat{\boldsymbol{d}}_s|\ z_s, \boldsymbol{\Theta}_{\rm SN}, \Omega_{m0}, w_0) \times P(z_s|\Omega_{m0}, w_0, \zeta_{\rm rate})$$

$$P(z_s|\Omega_{m0}, w_0, \zeta_{\rm rate}) \propto \frac{(z_s+1)^{\zeta_{\rm rate}}}{1+z_s}\left|\frac{\mathrm{d}V_c(z_s, \Omega_{m0}, w_0)}{\mathrm{d}z_s}\right|$$