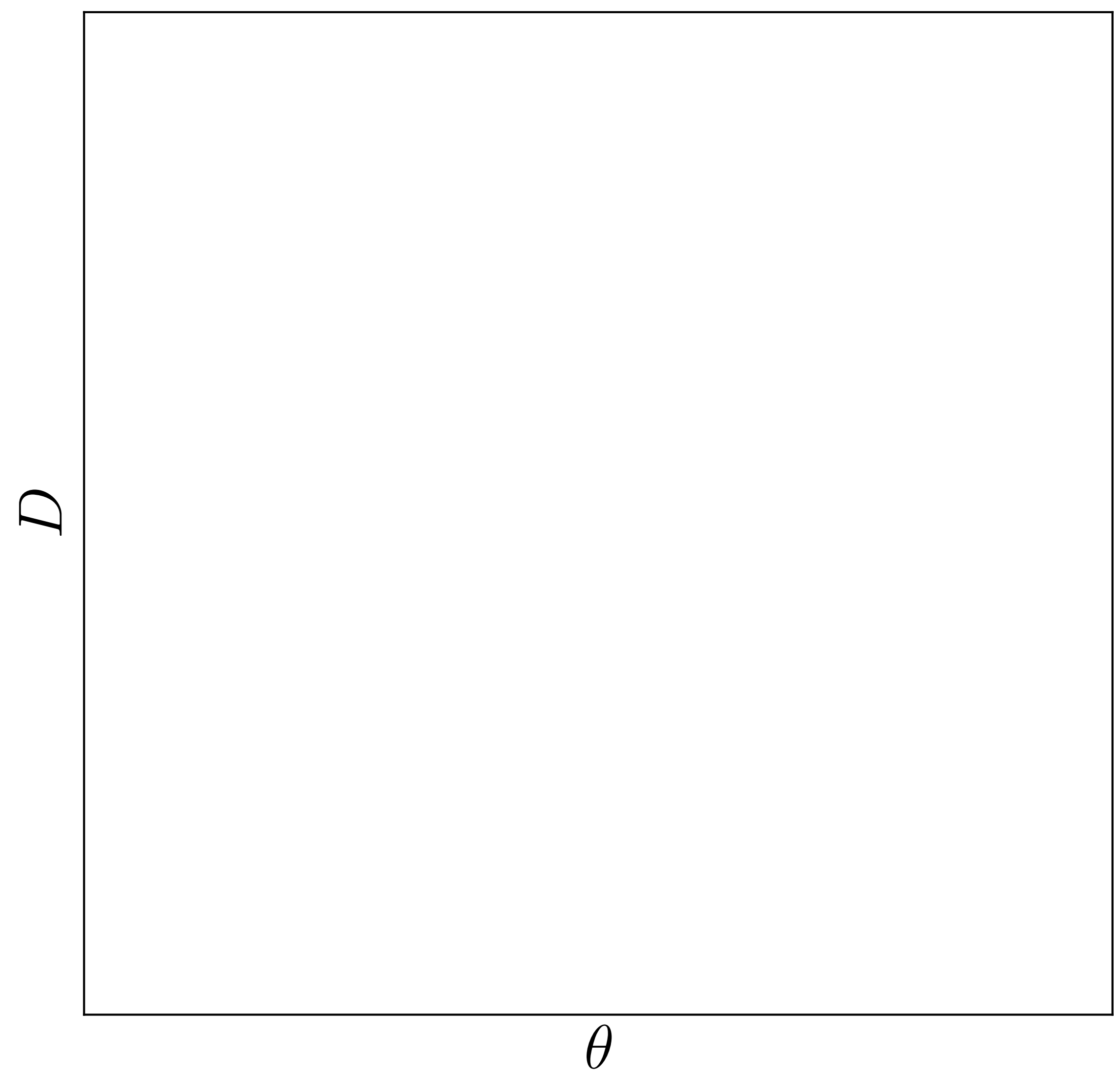
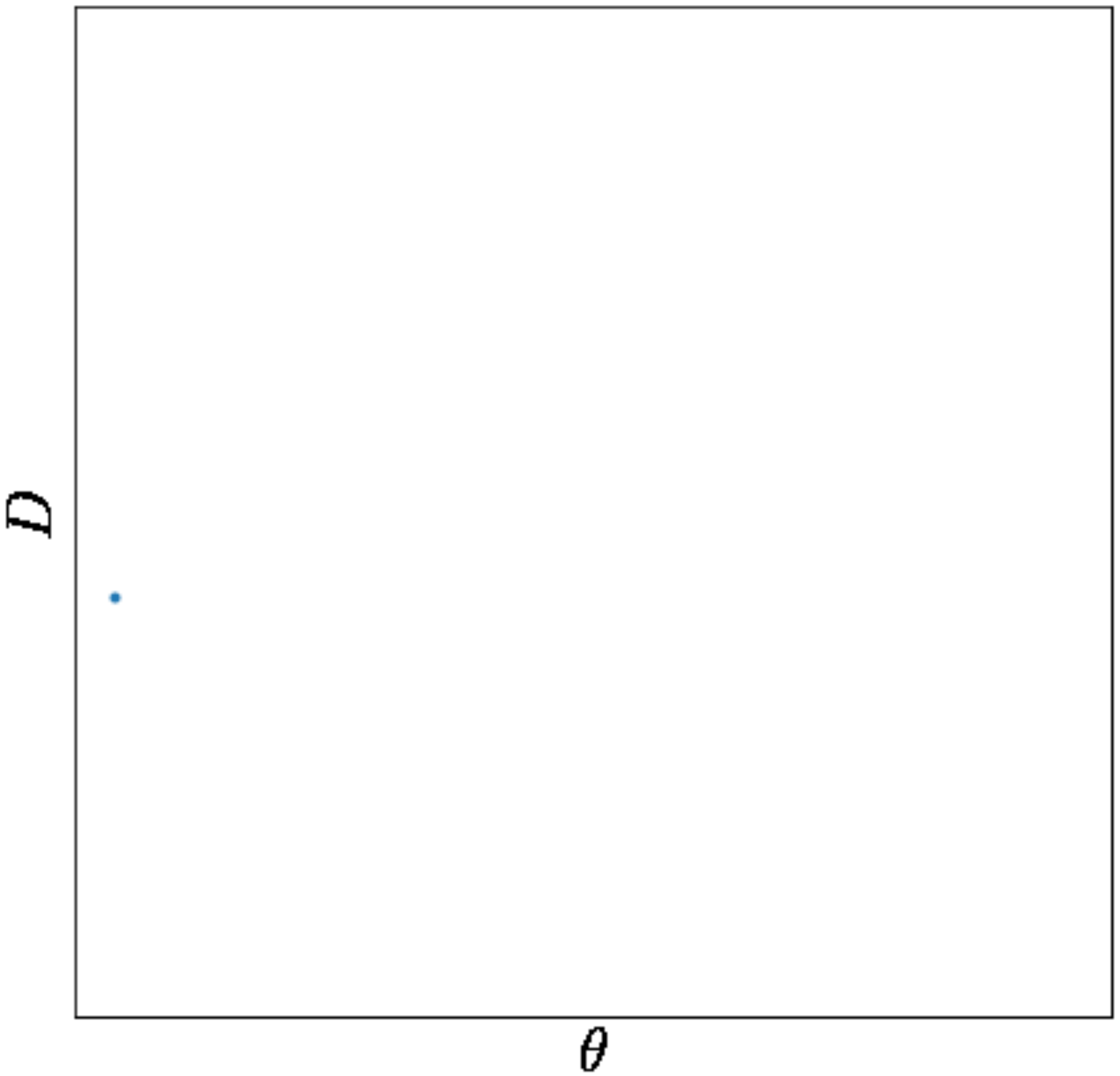
Robust Simulation-Based Inference with Bayesian Neural Networks

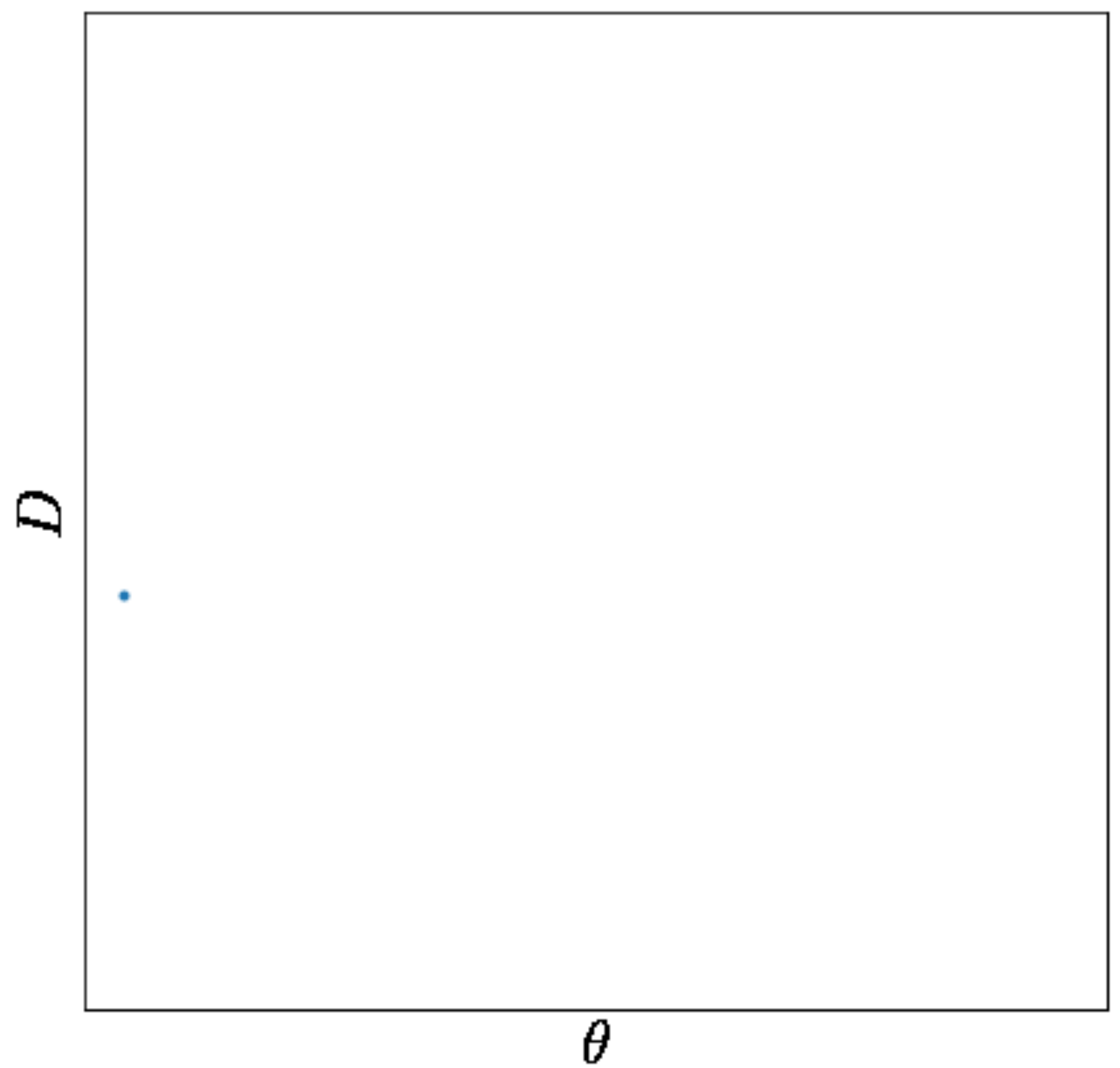
Will Handley, Miles Cranmer, Shirley Ho, Muntazir Abidi, Chang Hoon Hahn, Michael Eickenberg, Elena Massara, David Yallup

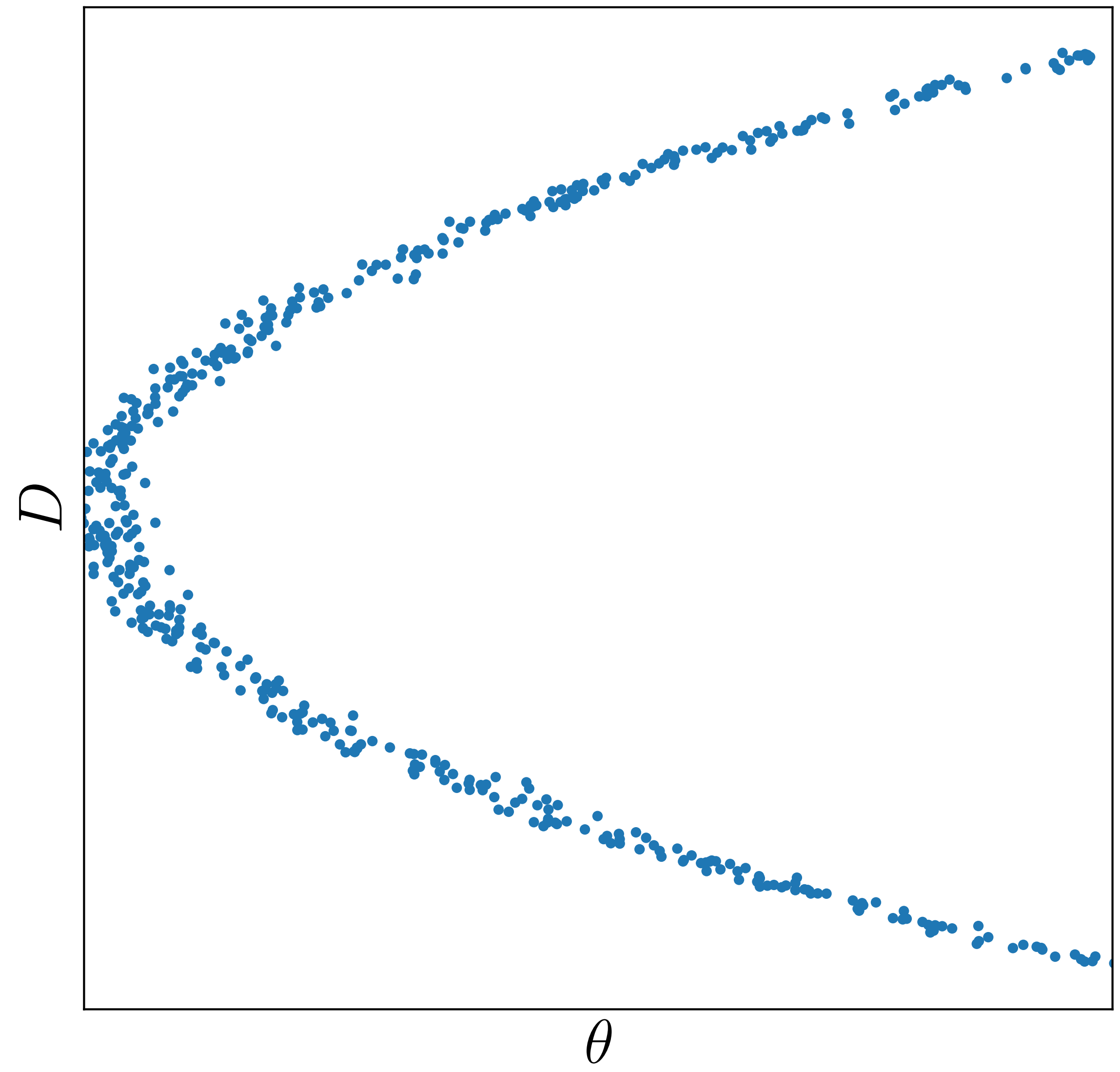
p.lemos@sussex.ac.uk

DELFI

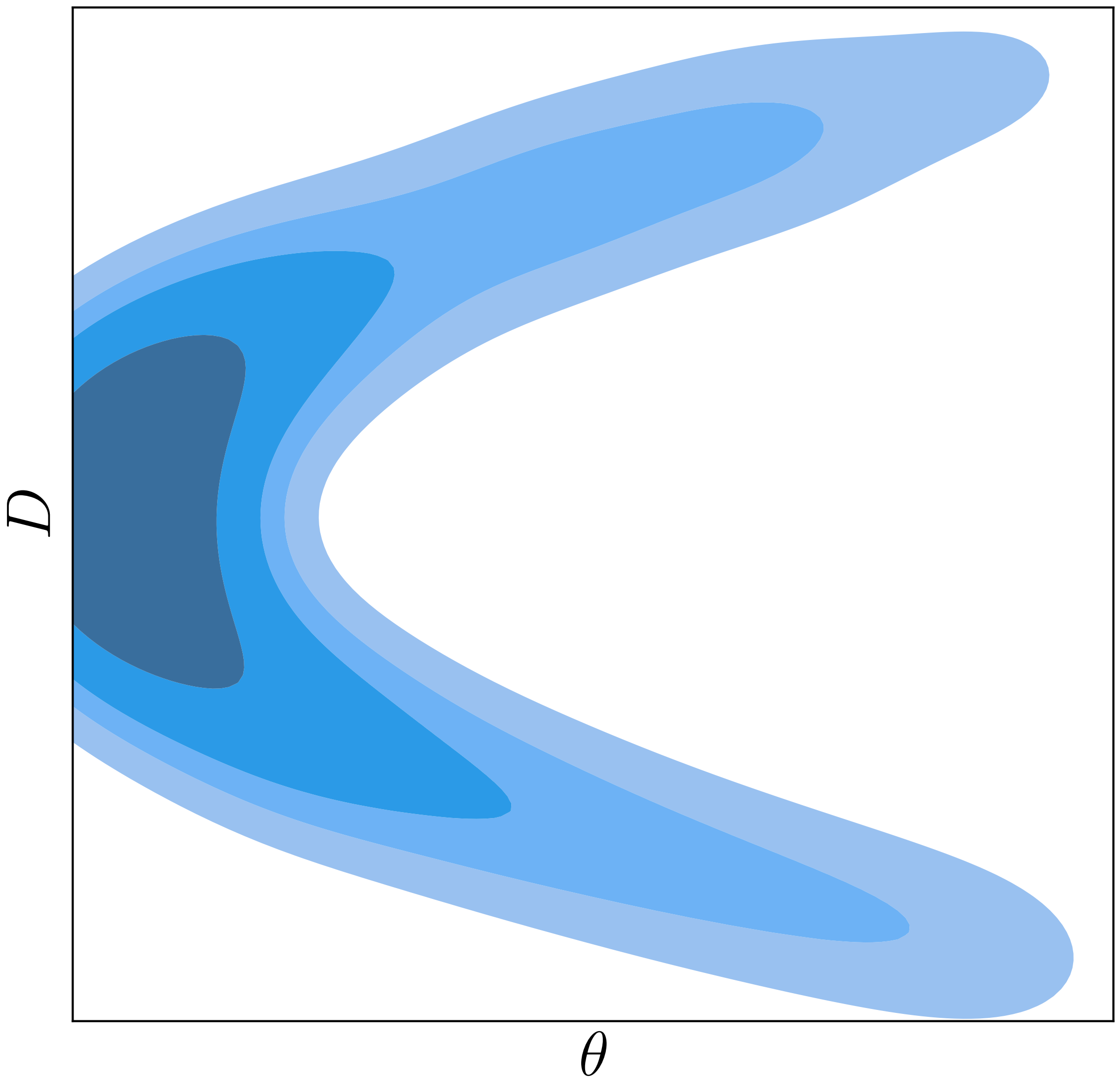

D
θ


D
θ

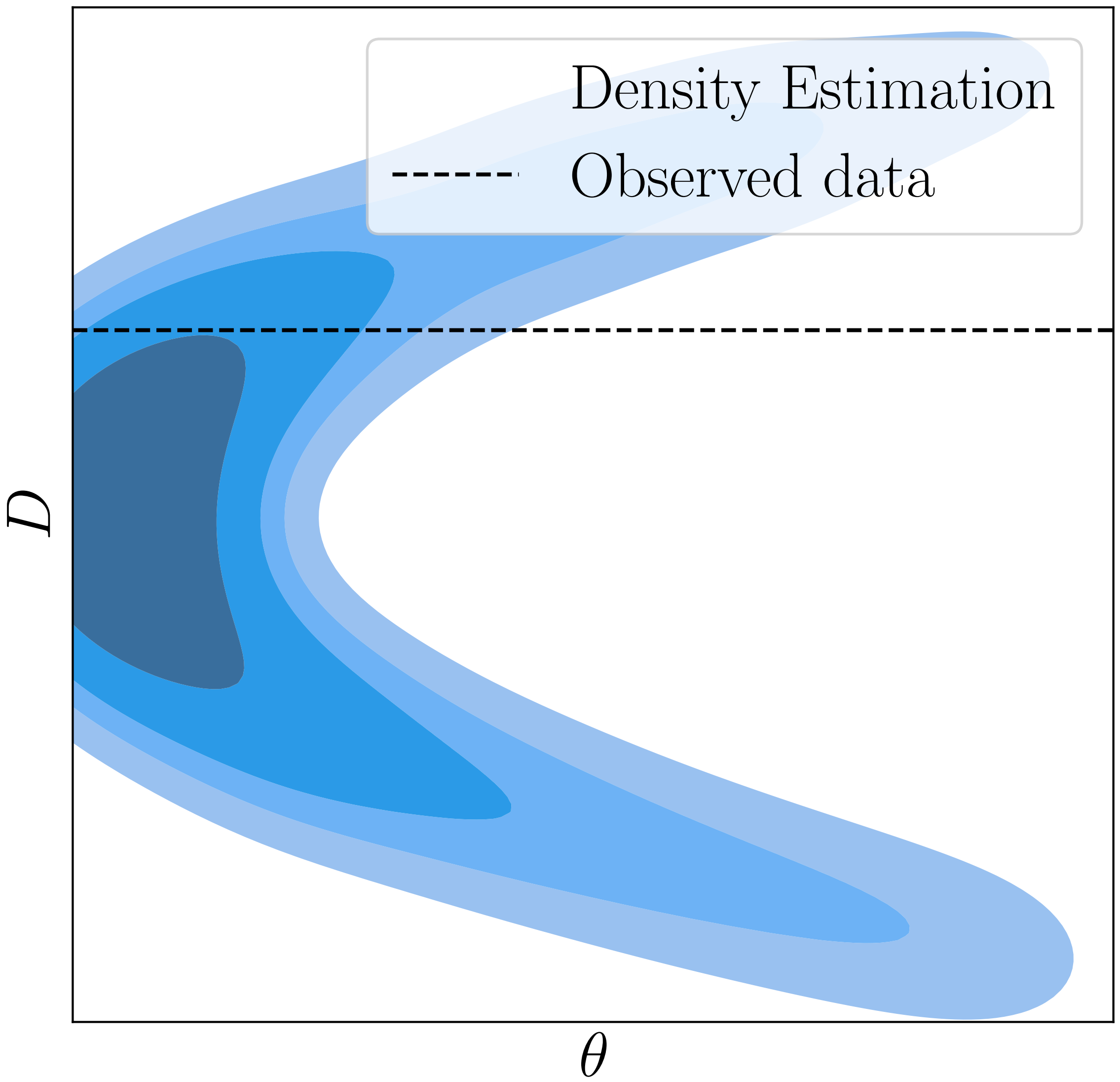

D
θ

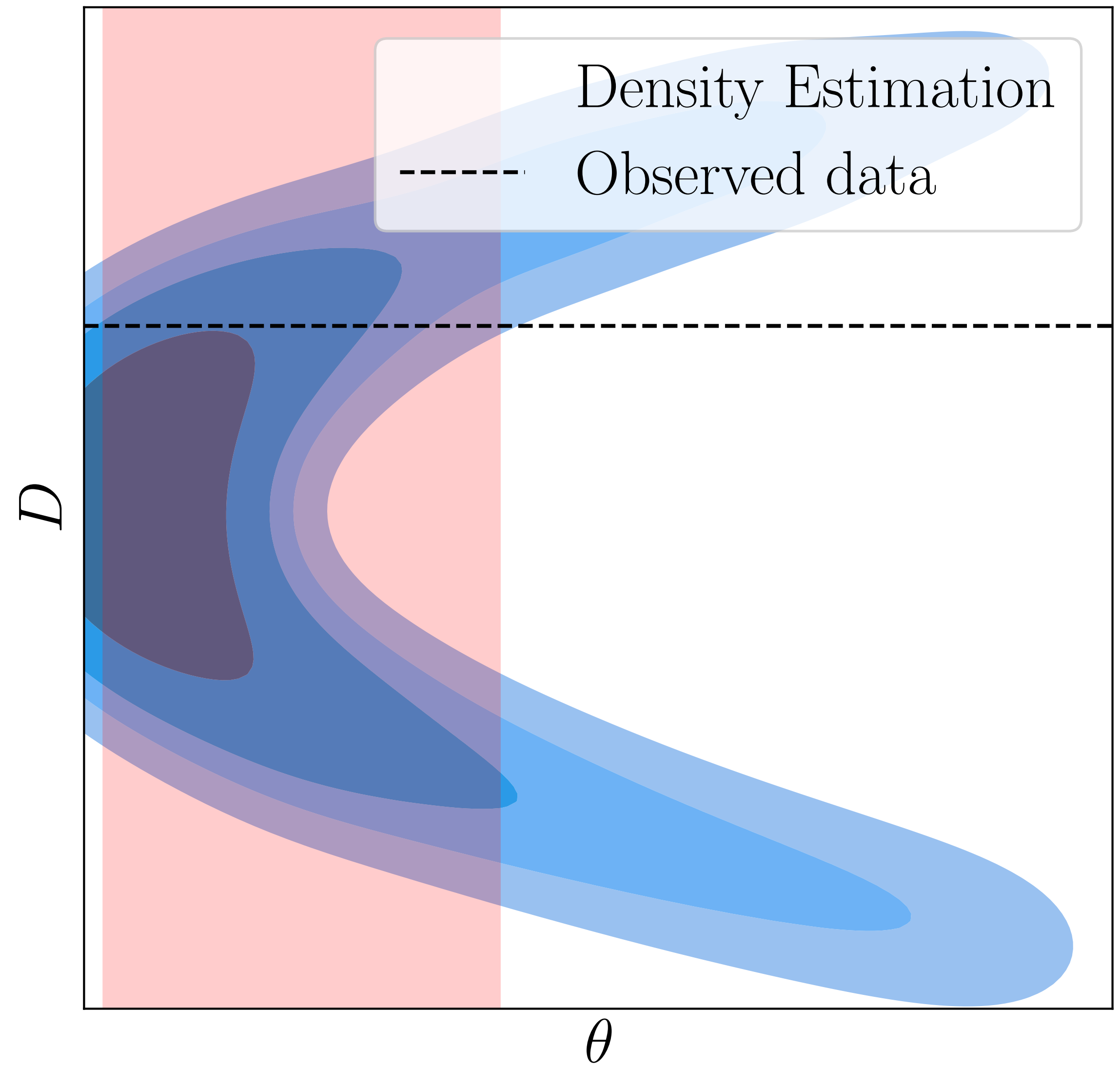


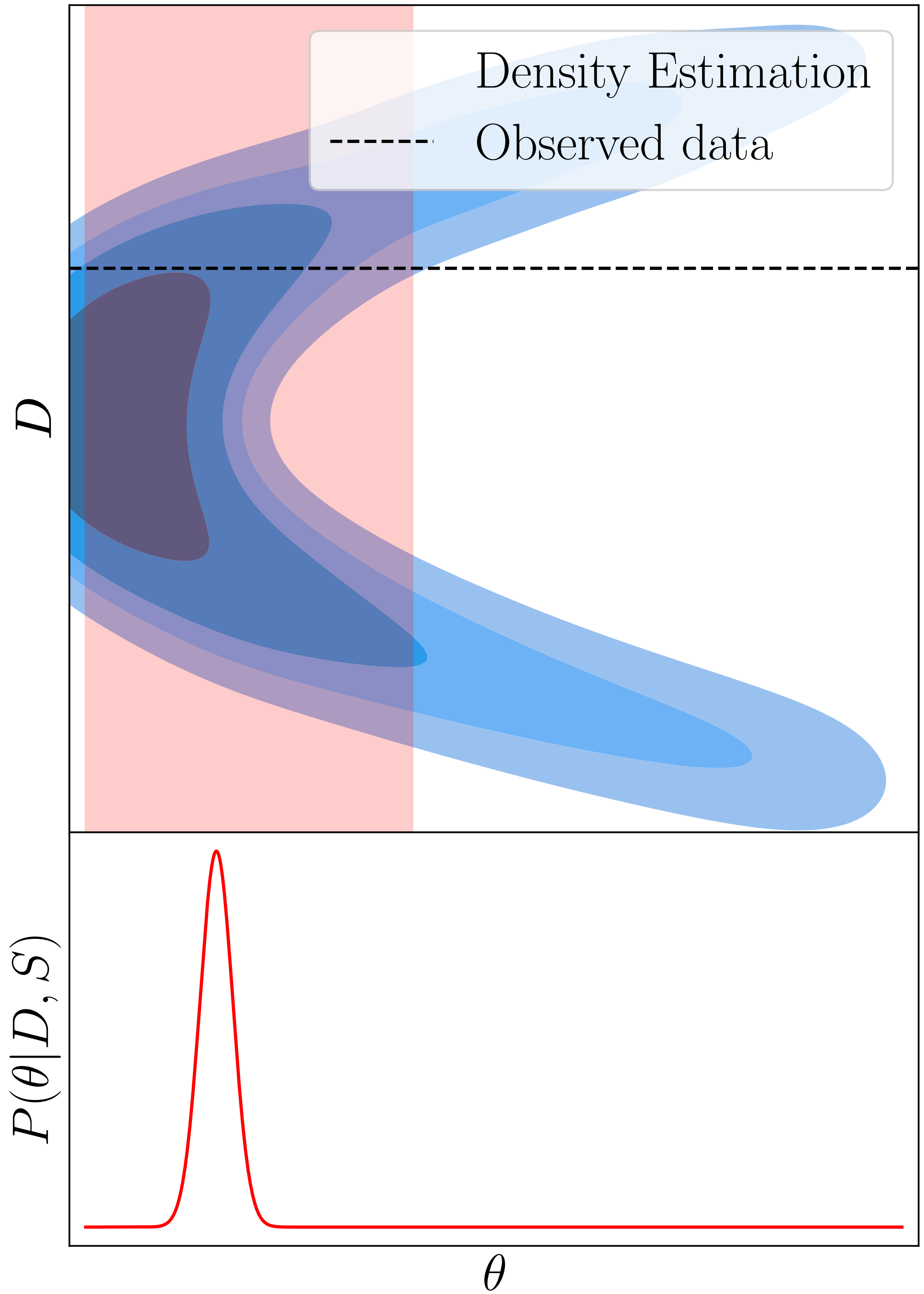
D
θ



D
θ

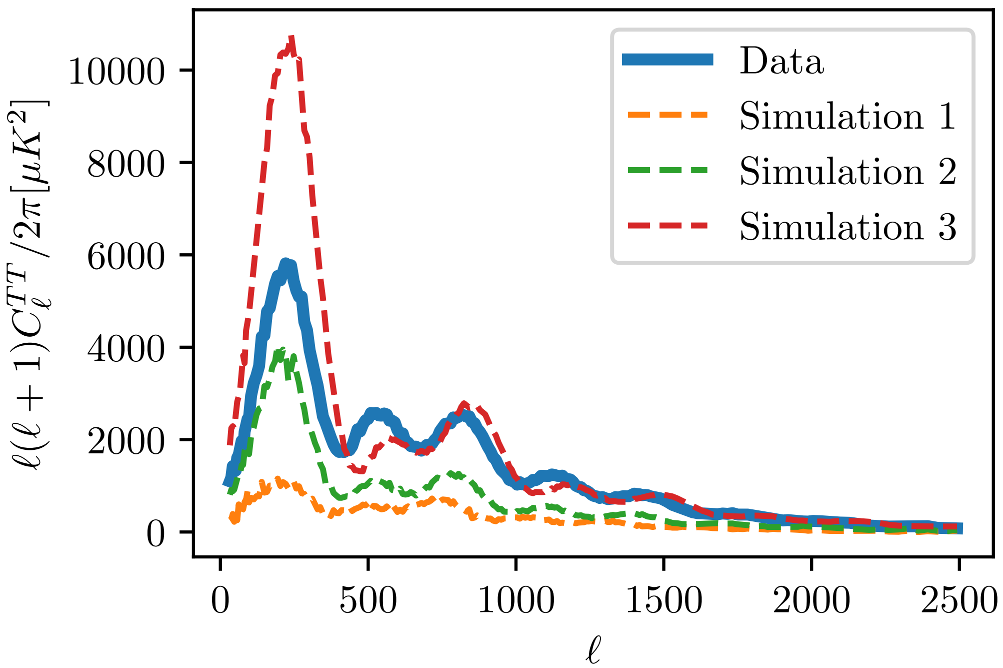

Density Estimation
Observed data
D
θ


Density Estimation
Observed data
D
θ

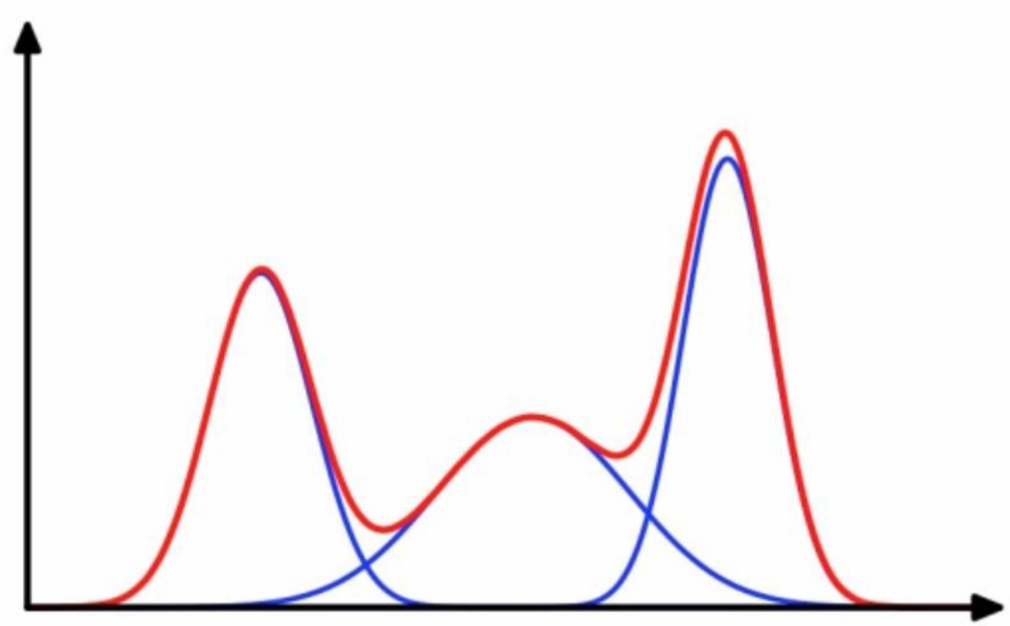

Density Estimation
Observed data
D
$P(\theta|D,S)$
θ

Density estimation

Mixture Density Network

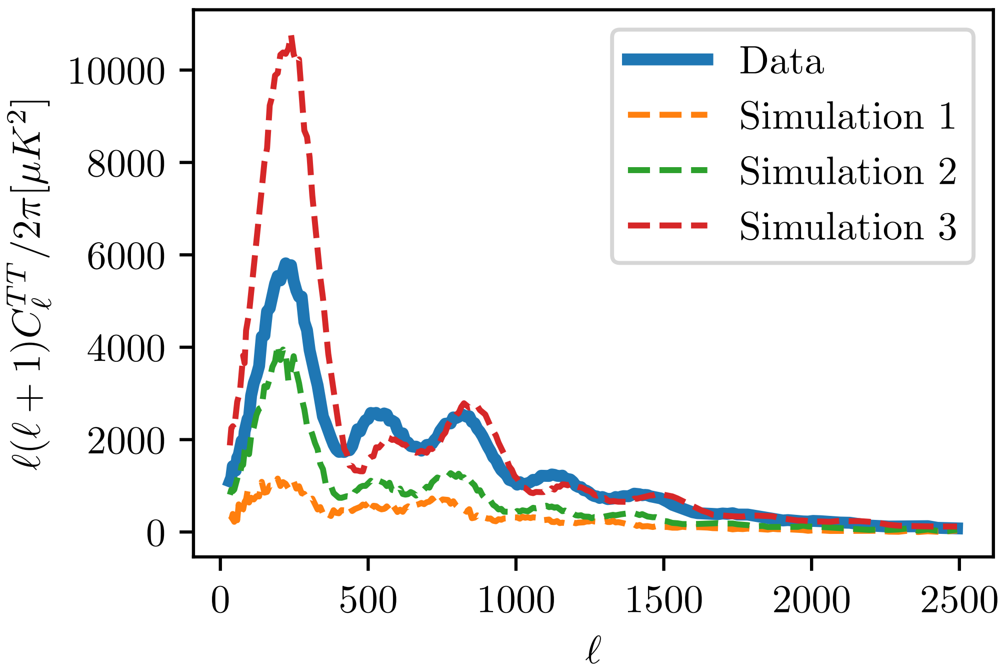


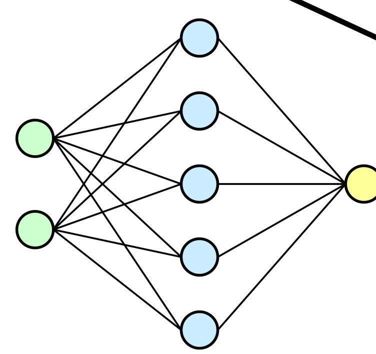
$$\boxed{\theta}$$

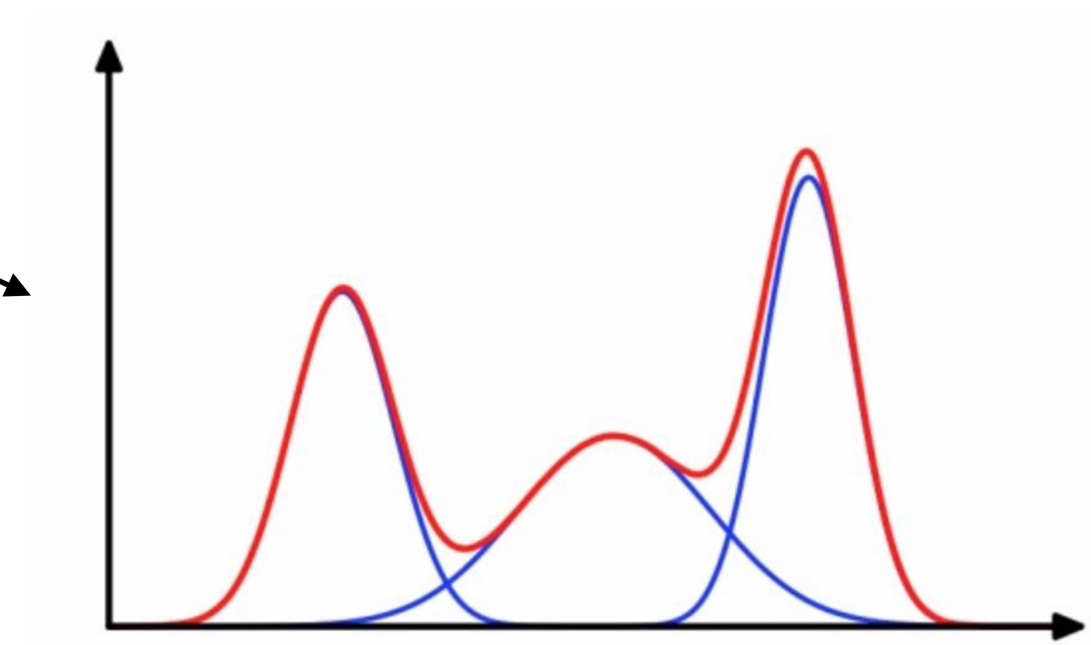


$$\boxed{\left\{\alpha_i(\theta), \mu_i(\theta), \Sigma_i(\theta)\right\}}$$

Mixture Density Network







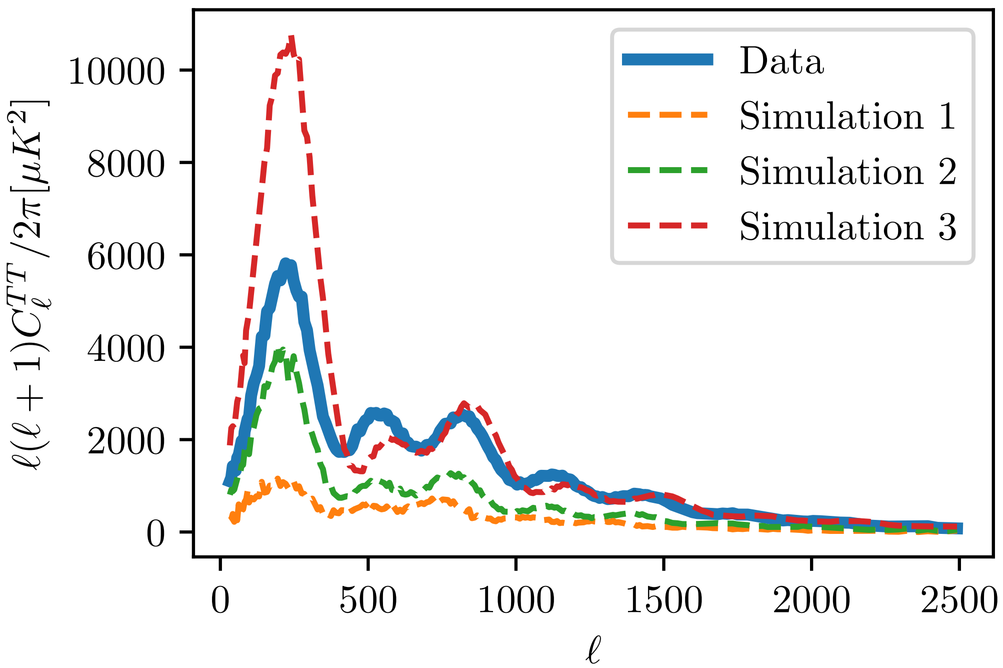
$\boxed{\theta}$

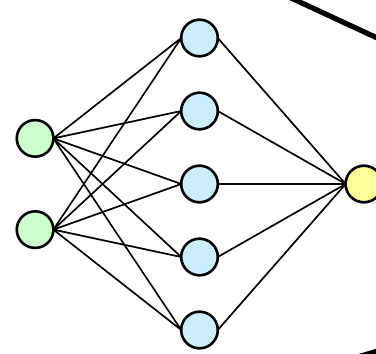
$$\boxed{\{\alpha_i(\theta), \mu_i(\theta), \Sigma_i(\theta)\}}$$

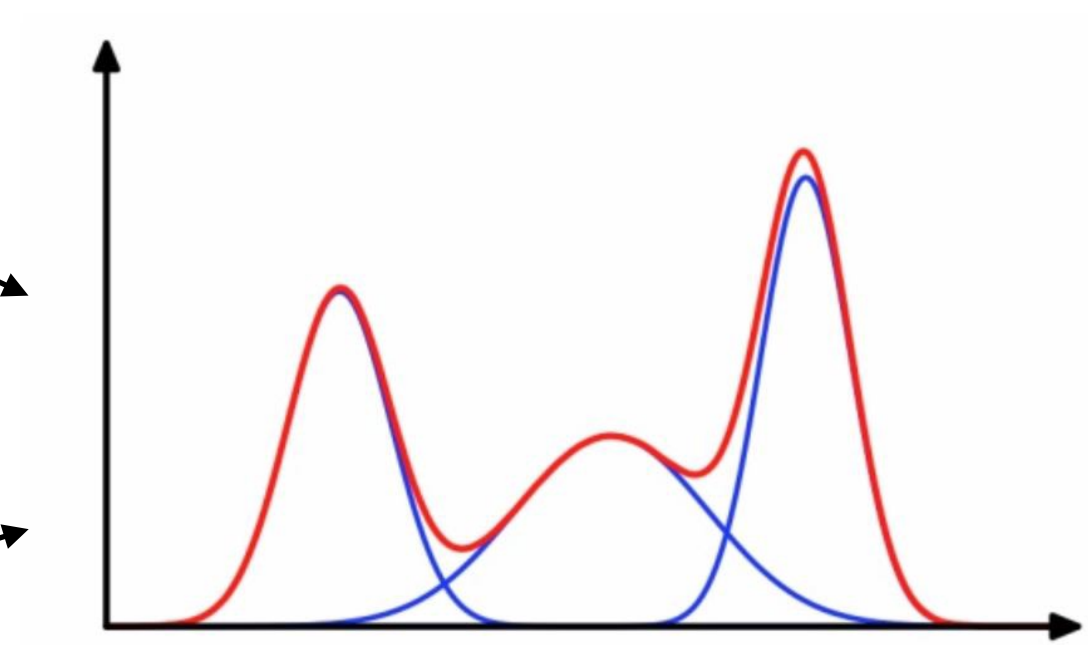
UNIVERSITY OF SUSSEX

UCL

Mixture Density Network



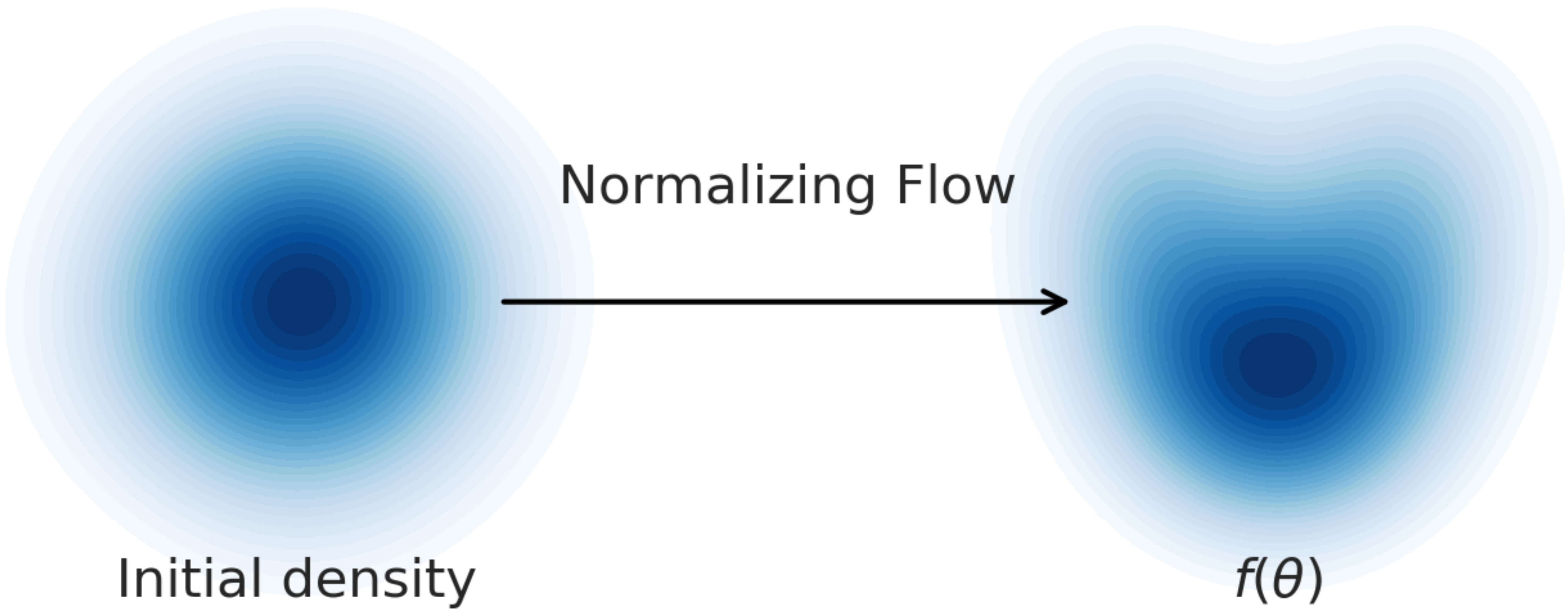




$$\boxed{\theta}$$

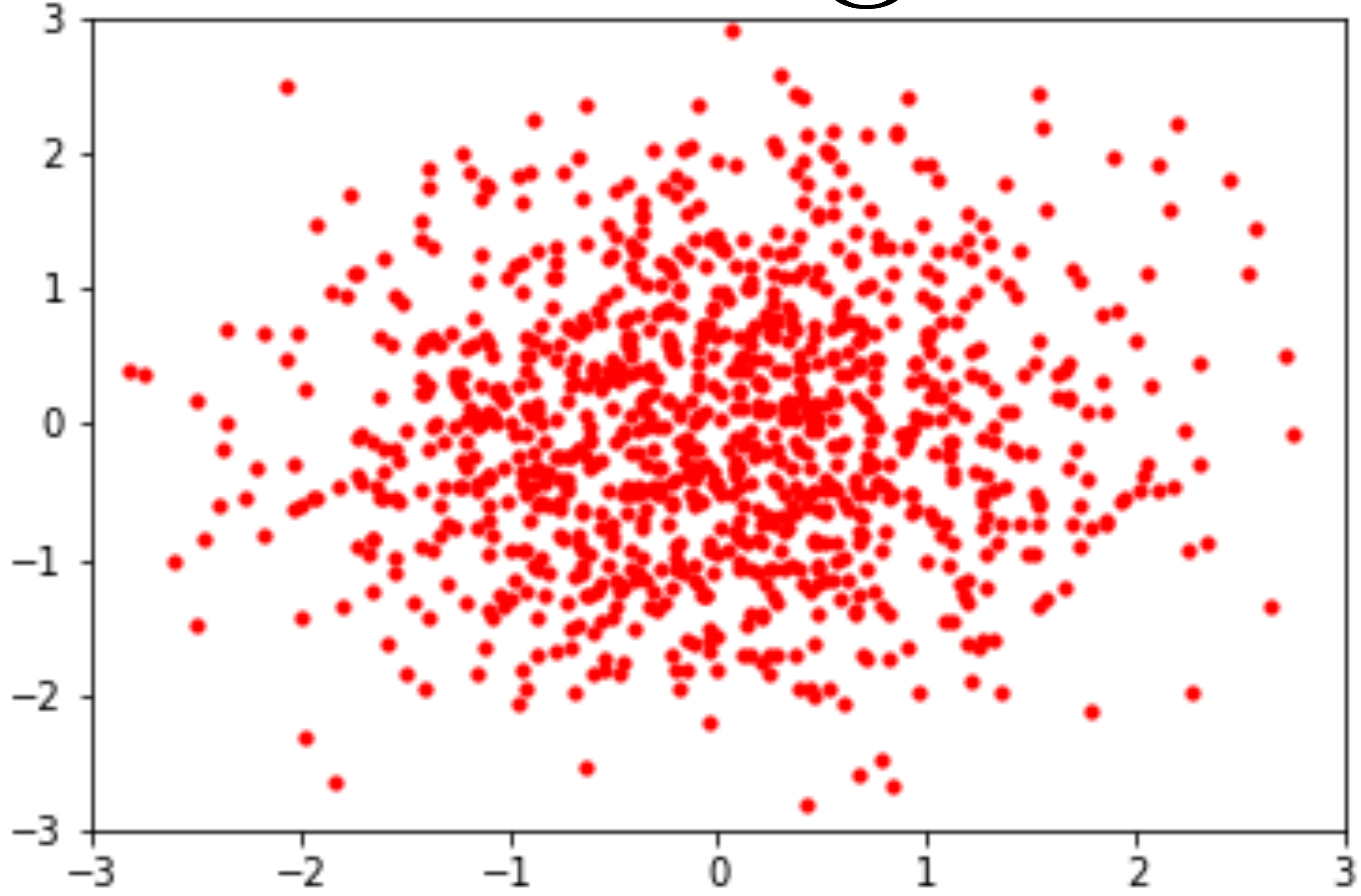
$$\boxed{\left\{\alpha_i(\theta), \mu_i(\theta), \Sigma_i(\theta)\right\}}$$

Normalizing Flows



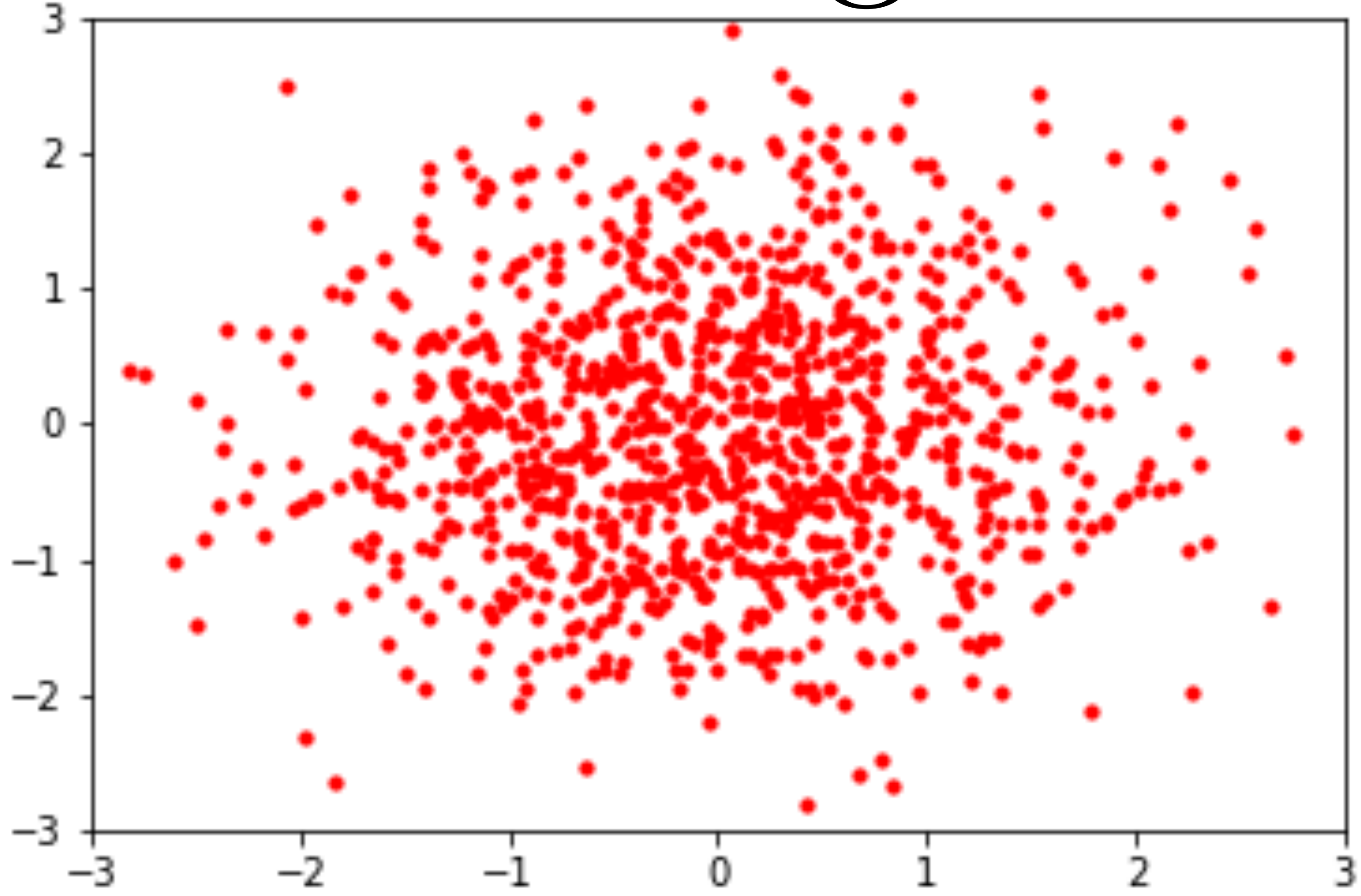
https://astroautomata.com/blog/simulation-based-inference/

Normalizing Flows



https://astroautomata.com/blog/simulation-based-inference/

Normalizing Flows

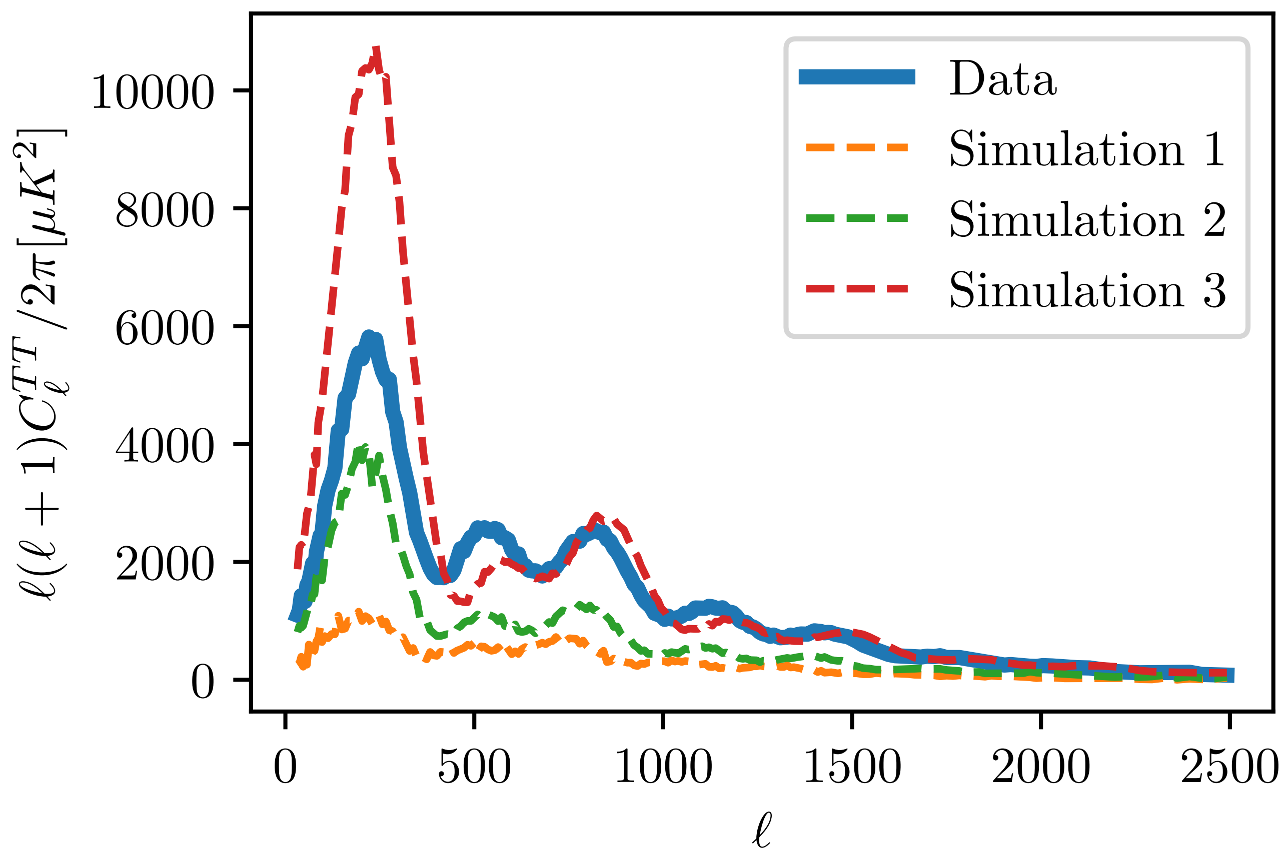


https://astroautomata.com/blog/simulation-based-inference/

Data compression

$\boxed{\theta}$

Parameters

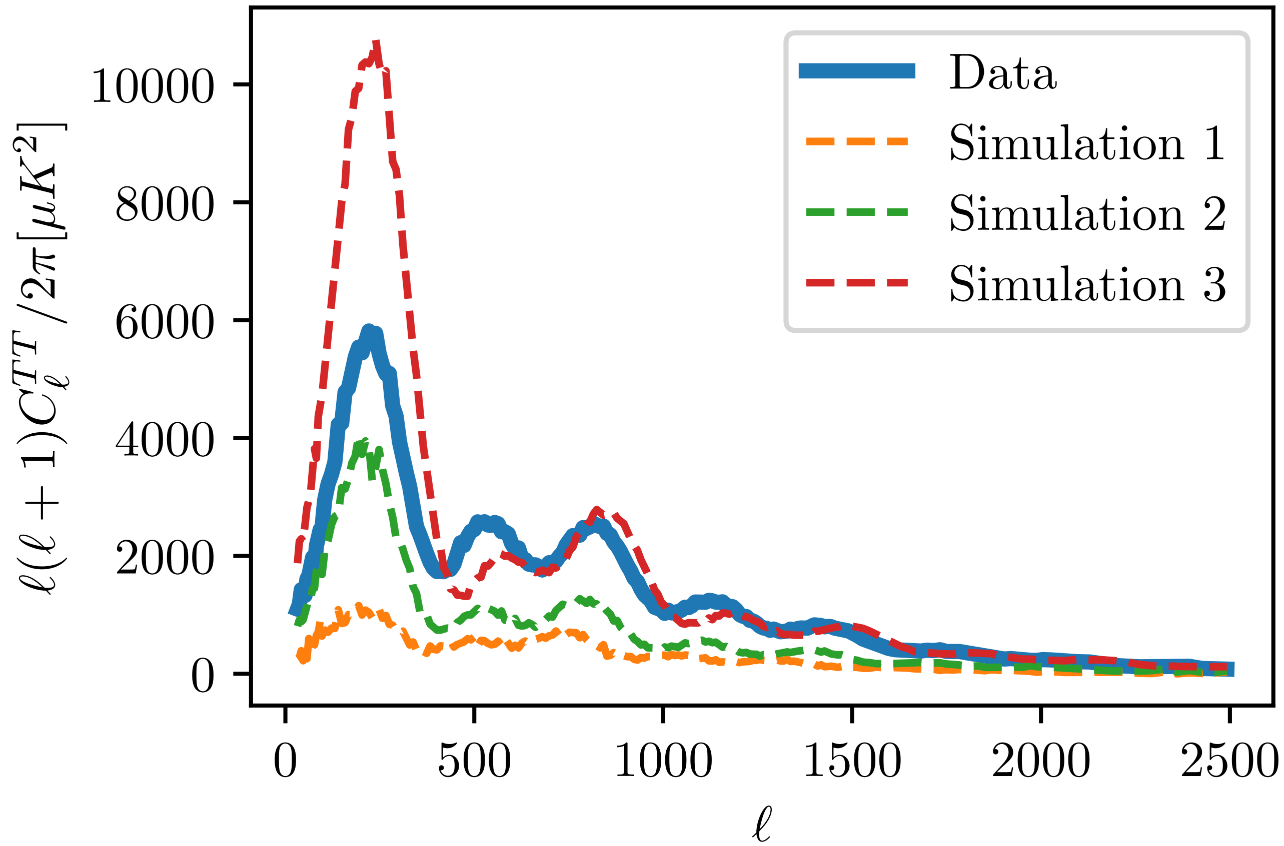


Simulated data

$\boxed{\hat{\theta}}$

Compressed Parameters

$$\boxed{\theta} \longrightarrow$$

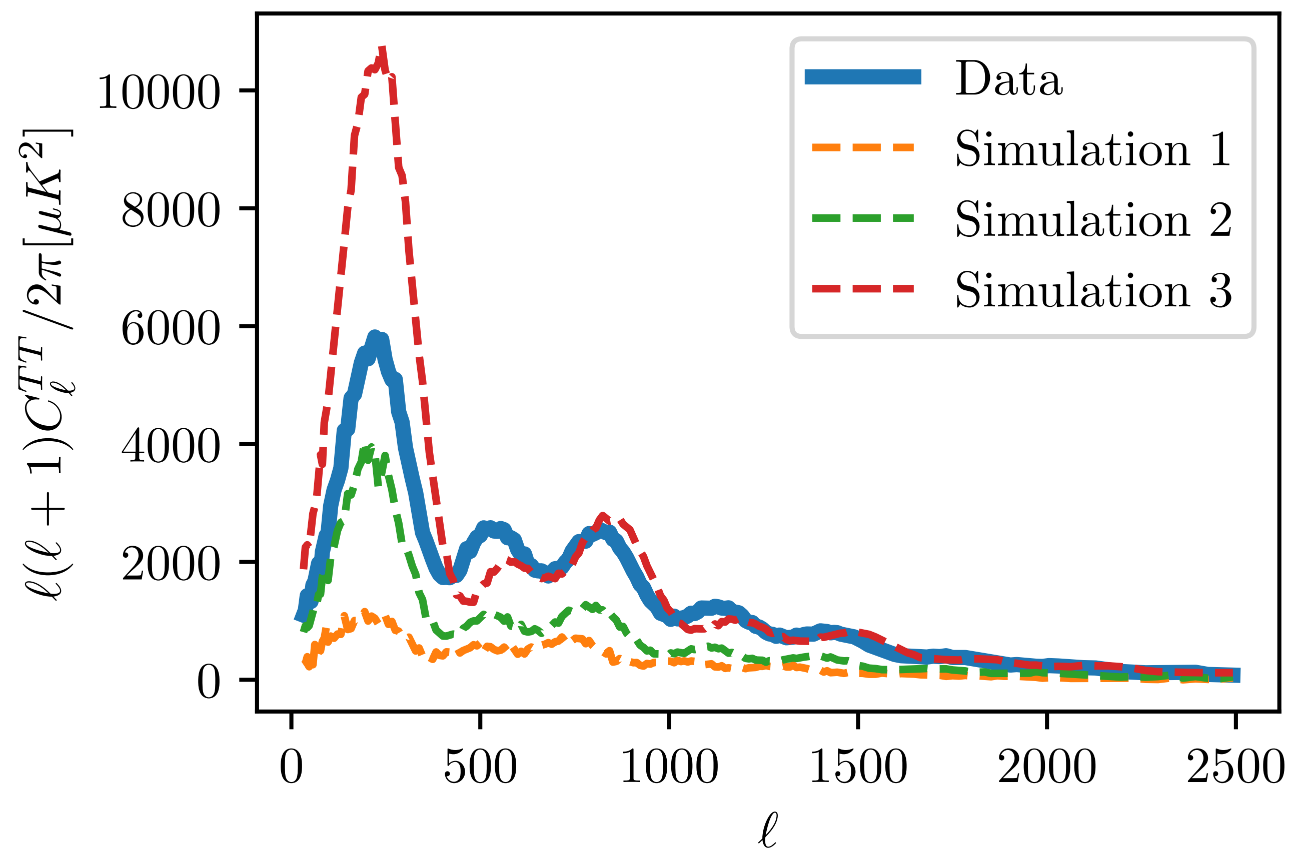


$$\boxed{\hat{\theta}}$$

Parameters

Simulated data

Compressed Parameters

$\boxed{\theta}$ $\longrightarrow$ 

$\longrightarrow$ $\boxed{\hat{\theta}}$

Parameters

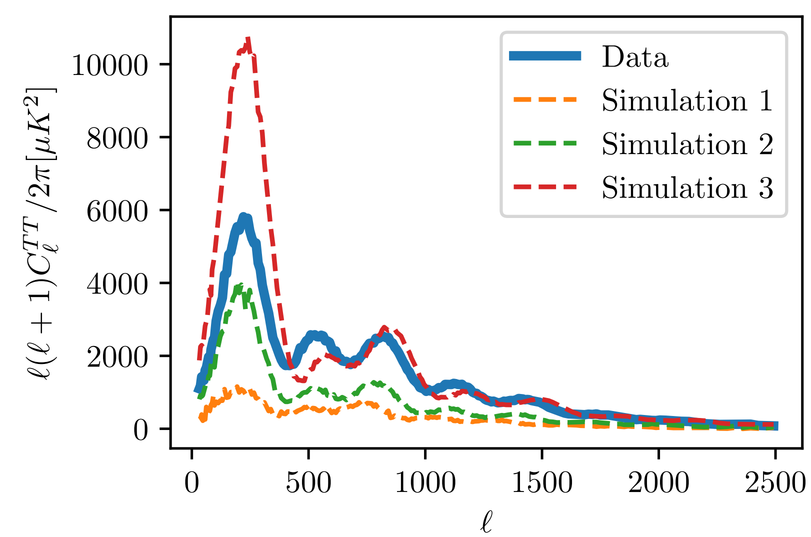
Simulated data

Compressed Parameters

Compression methods

- MOPED (Heavens, Jimenez & Lahav, 1999)
- IMNN (Charnock, Lavaux & Wandelt, 2018)

$\boxed{\theta}$ $\longrightarrow$



$\longrightarrow$ $\boxed{\hat{\theta}}$

Parameters

Simulated data

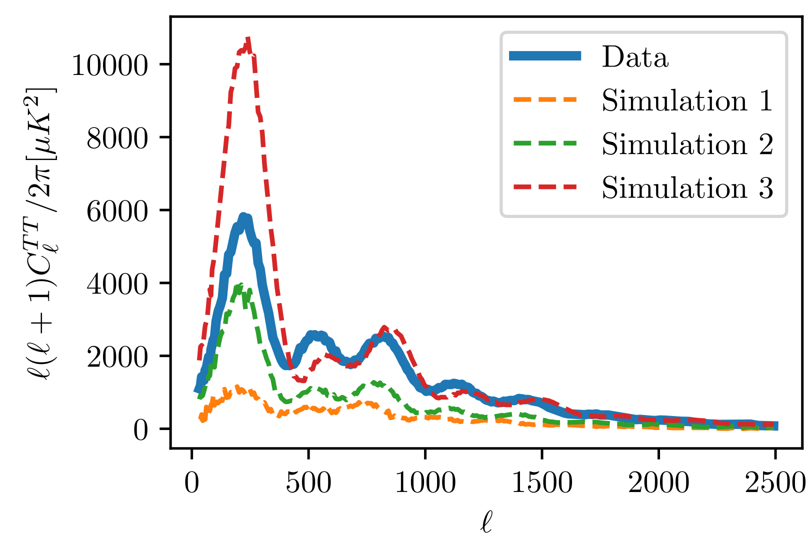
Compressed Parameters

Compression methods

- MOPED (Heavens, Jimenez & Lahav, 1999)
- IMNN (Charnock, Lavaux & Wandelt, 2018)

$\boxed{\theta} \longrightarrow$

Parameters



Simulated data

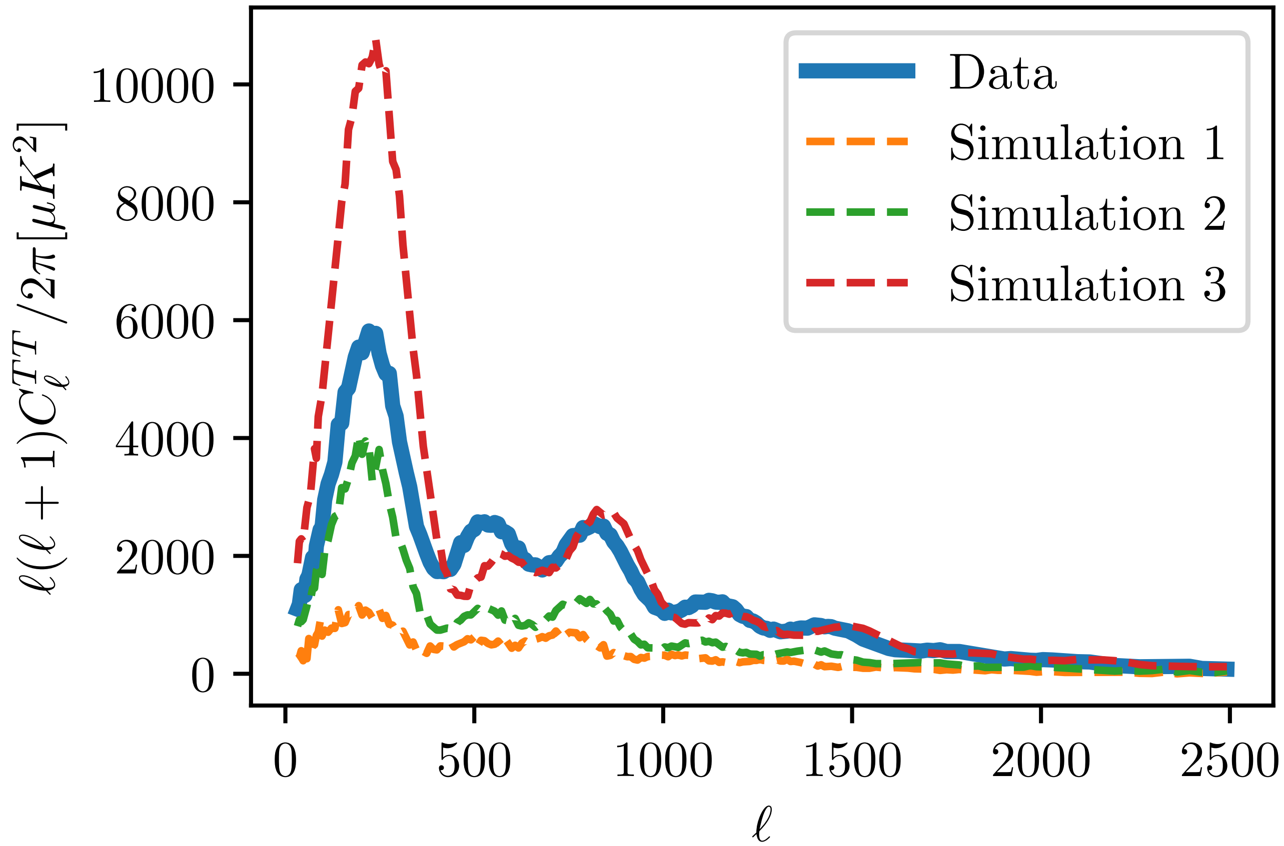
$\longrightarrow \boxed{\hat{\theta}}$

Compressed Parameters

All of them require a covariance matrix, or very fast simulations.

Neural compressor

$\boxed{\theta}$



$\boxed{\hat{\theta}}$

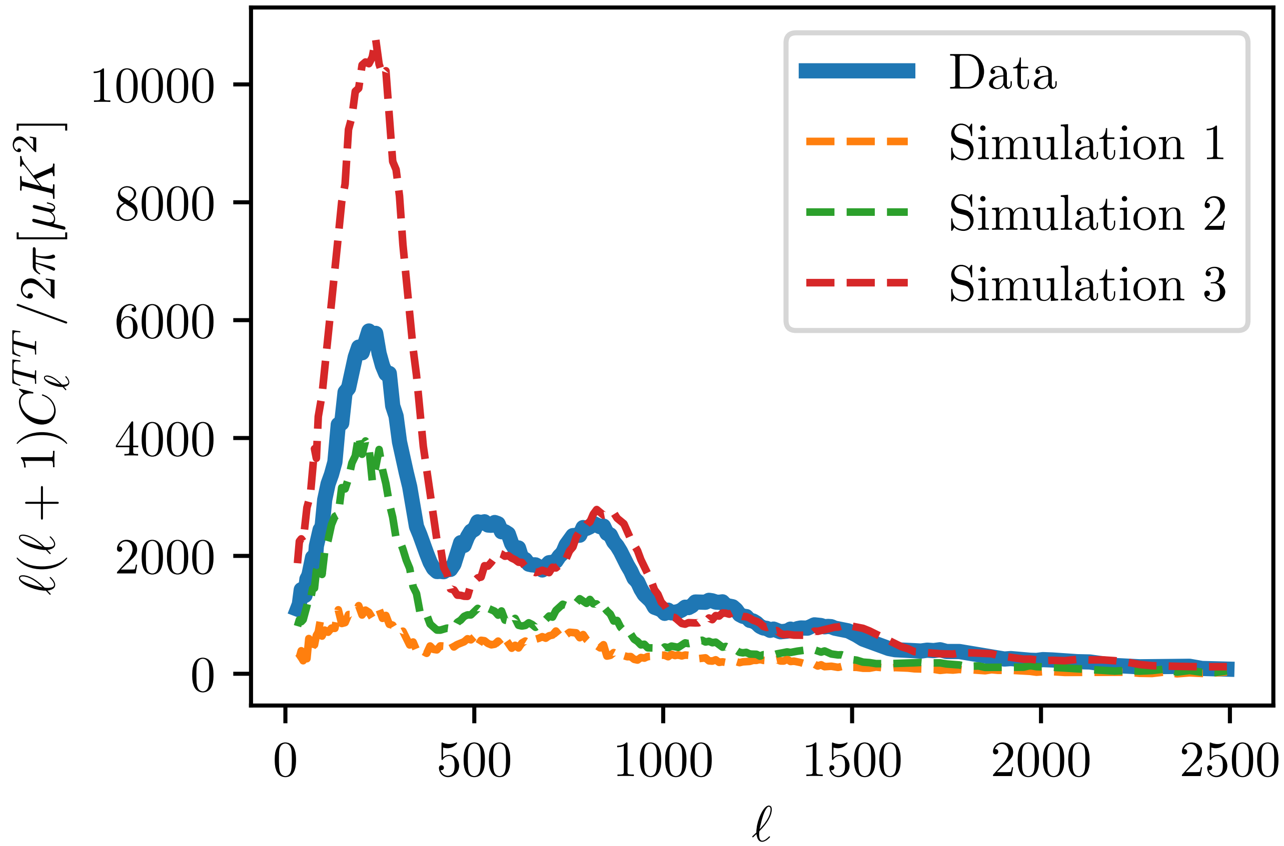
Parameters

Simulated data

Compressed Parameters

Neural compressor

$\boxed{\theta}$ $\longrightarrow$



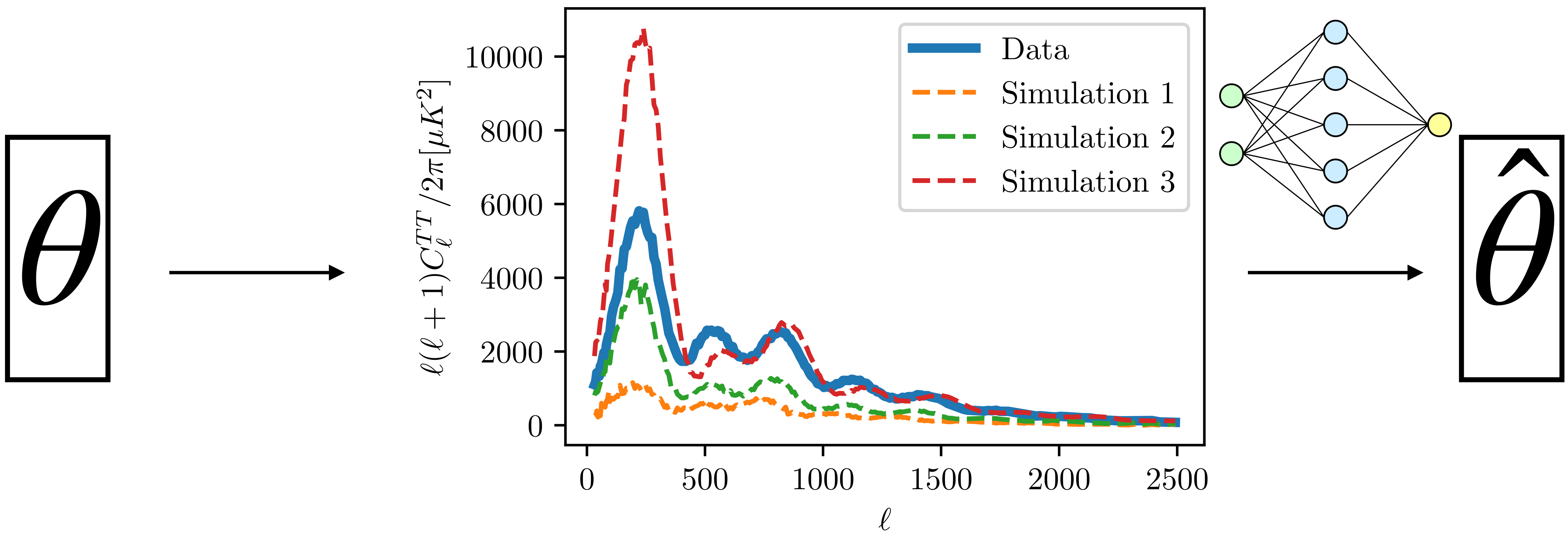
$\boxed{\hat{\theta}}$

Parameters

Simulated data

Compressed Parameters

Neural compressor



Parameters

Simulated data

Compressed Parameters

Our simulator

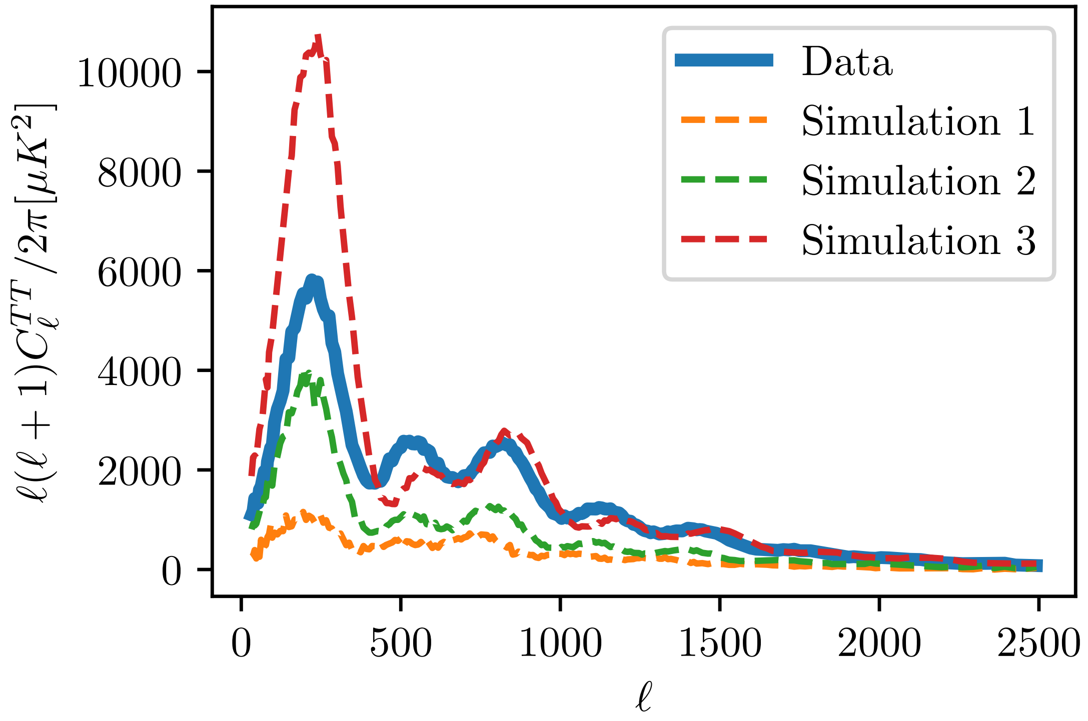
Realistic SBI has

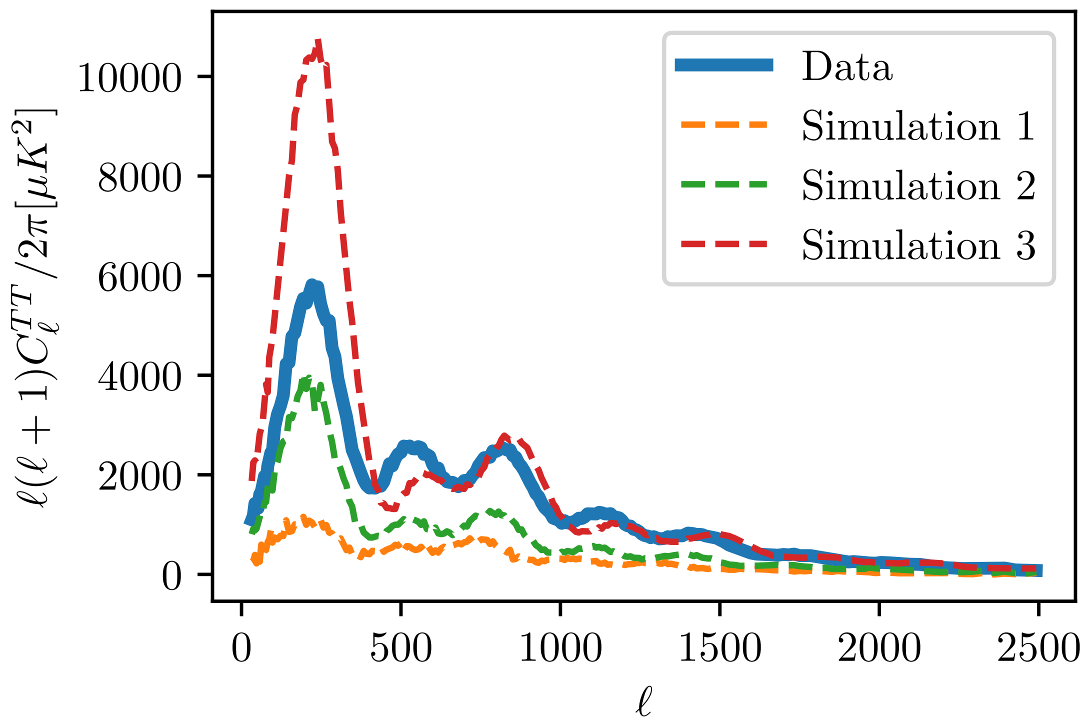
- Expensive simulations, often a fixed number of them (e.g. QUIJOTE, CAMELS...).
- Imperfect forward models, that do not capture every aspect of the observed data.

Example: CMB Cls

A simple example to test things:

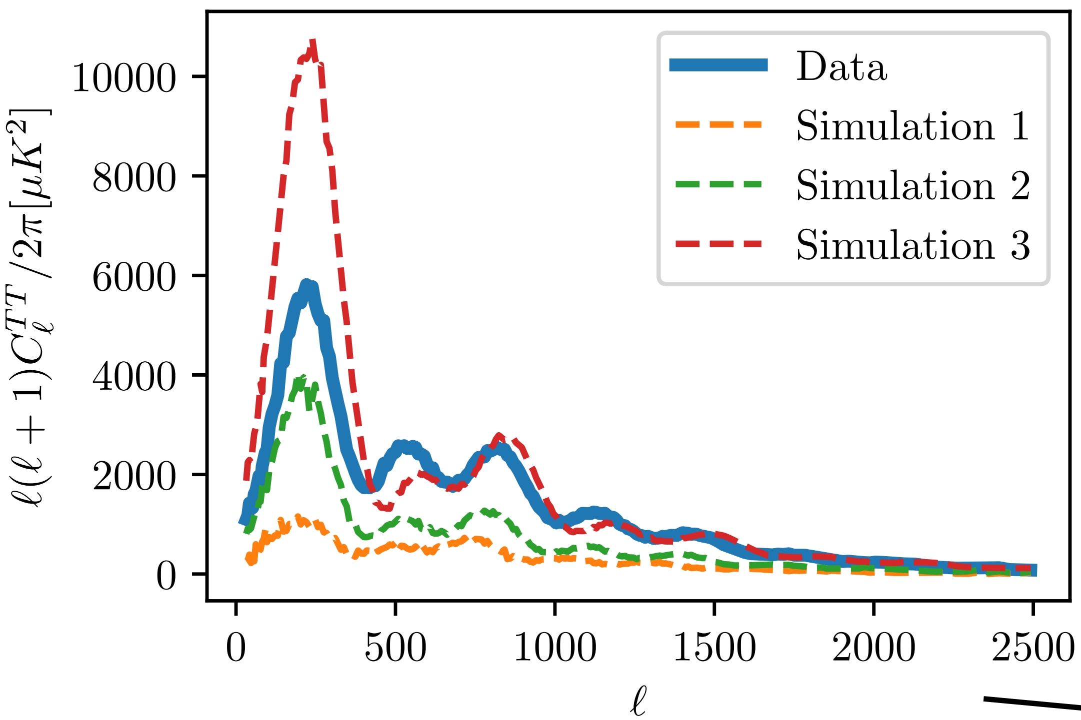
- We want to estimate the ΛCDM parameters from CMB power spectra.
- We do not have access to CAMB/CLASS
- Instead, we have access to 10.000 simulations of spectra for different parameters
- All the simulations have Planck noise added





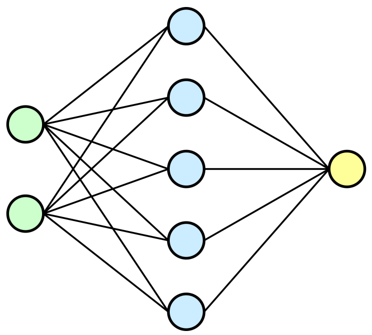
UNIVERSITY OF SUSSEX

UCL

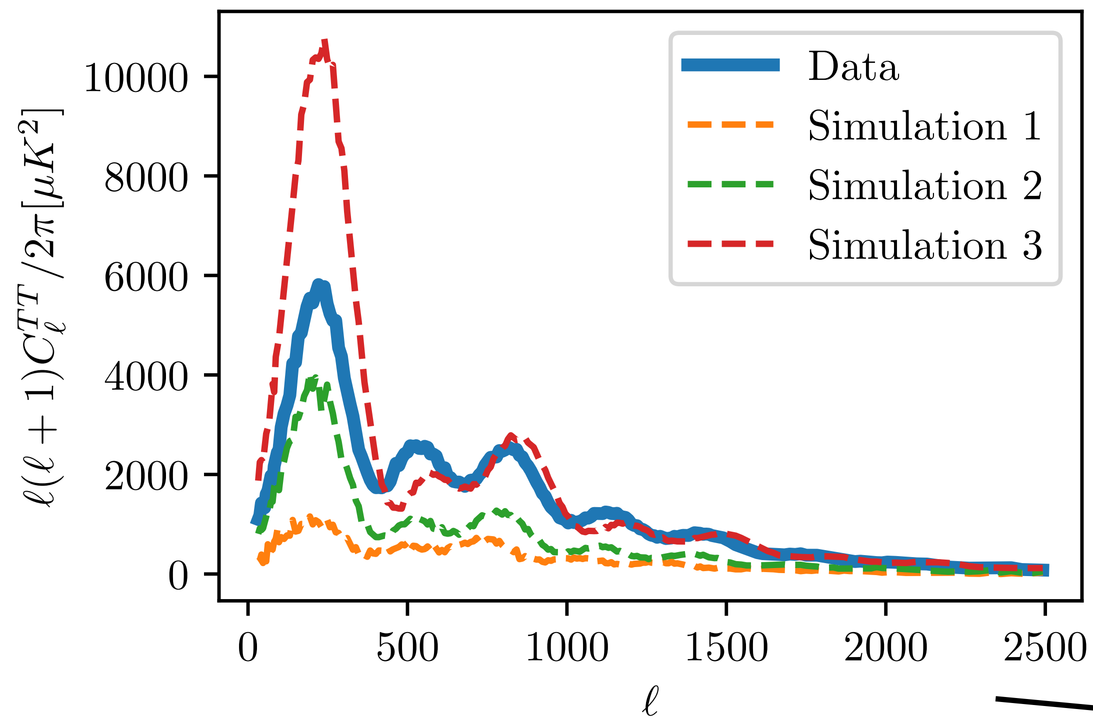


US
UNIVERSITY OF SUSSEX

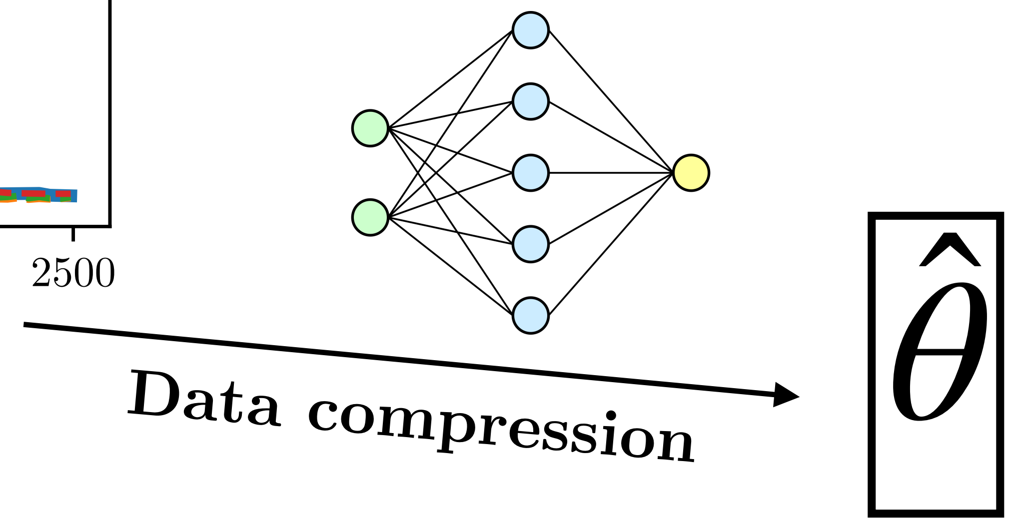
UCL



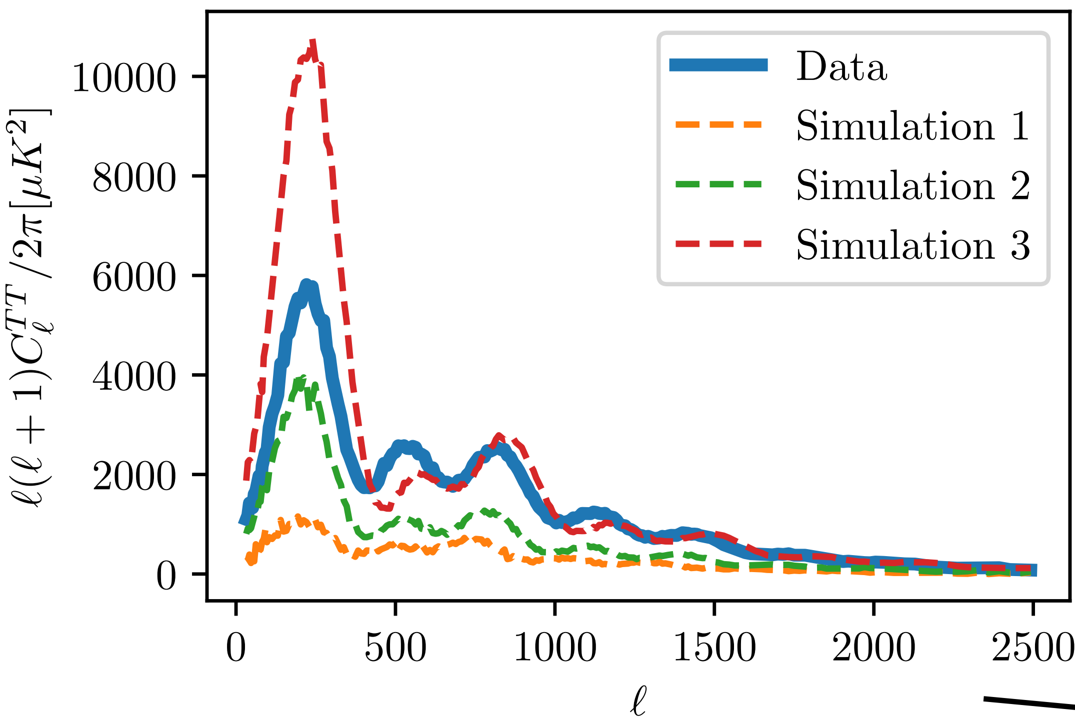
Data compression

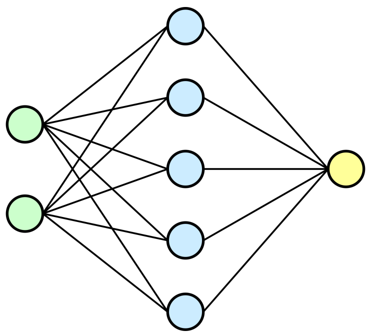


Data
Simulation 1
Simulation 2
Simulation 3
$\ell(\ell+1)C_\ell^{TT}/2\pi[\mu K^2]$
ℓ
0
2000
4000
6000
8000
10000
0
500
1000
1500
2000
2500



Data compression
$\hat{\theta}$

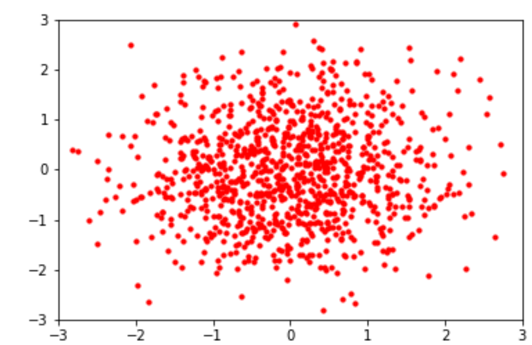


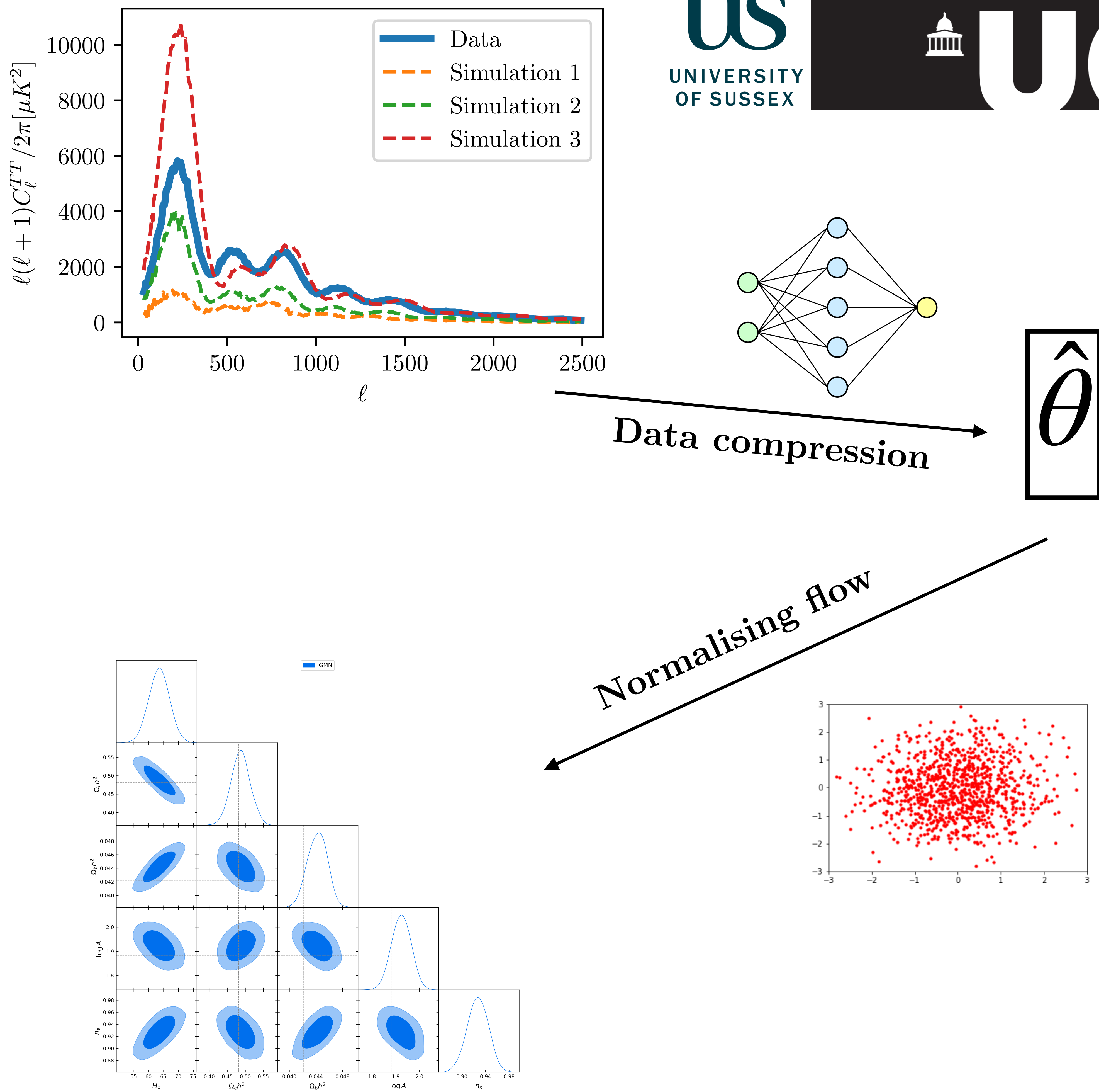


Data compression

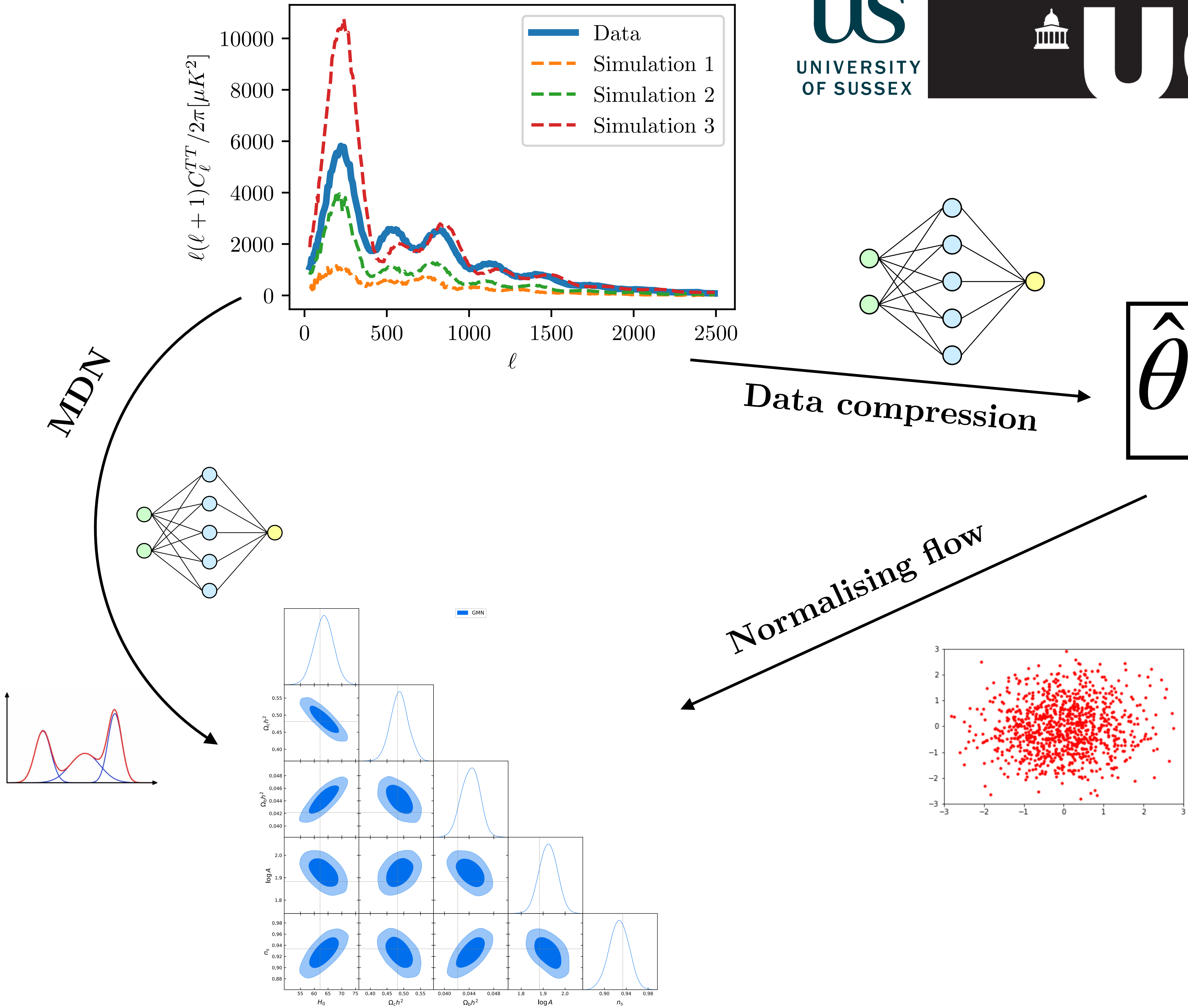
$$\boxed{\hat{\theta}}$$

Normalising flow




Data
Simulation 1
Simulation 2
Simulation 3
$\ell(\ell+1)C_\ell^{TT}/2\pi[\mu K^2]$
ℓ
Data compression
$\hat{\theta}$
Normalising flow
GMN
H_0
$\Omega_c h^2$
$\Omega_b h^2$
$\log A$
n_s

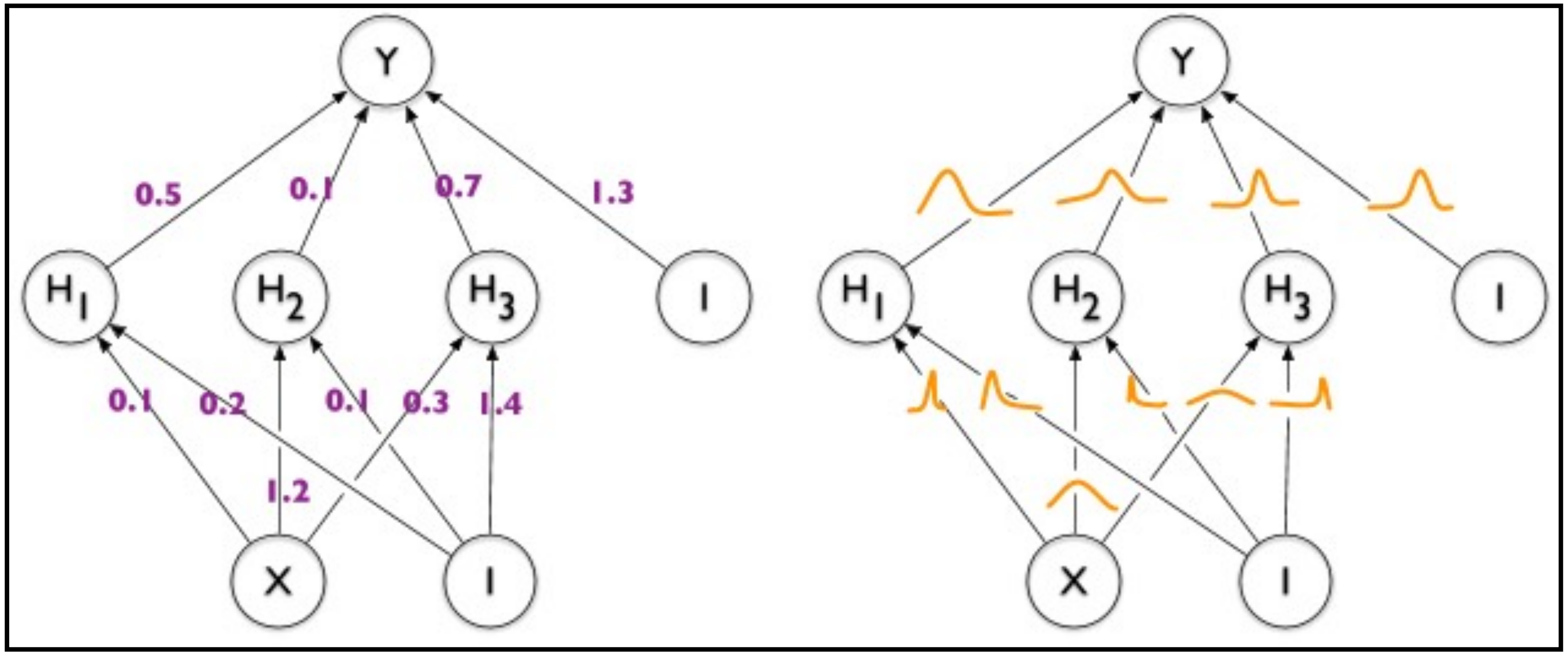
US
UNIVERSITY OF SUSSEX
UCL



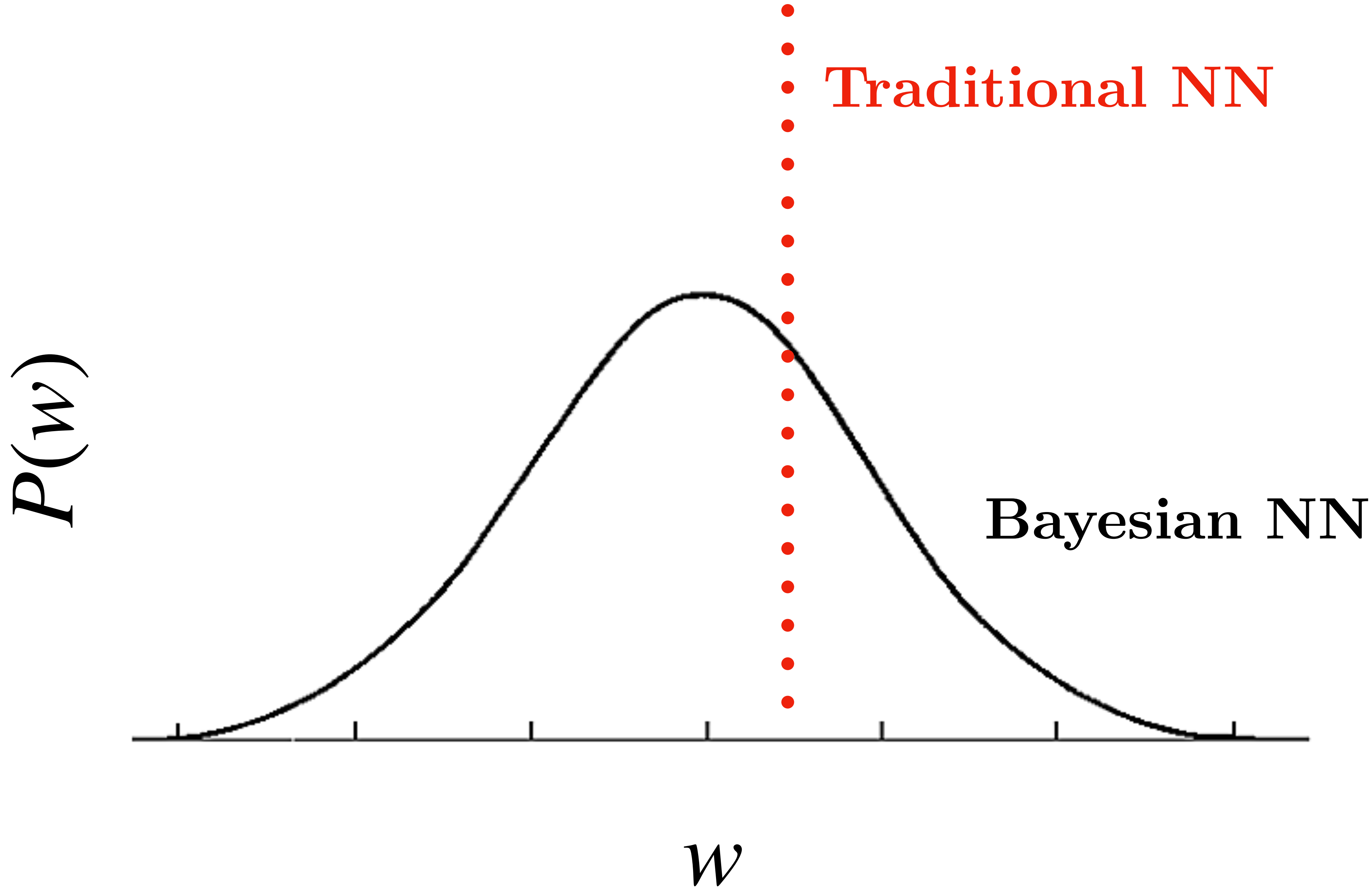
Bayesian Neural Networks

Bayesian Neural Networks

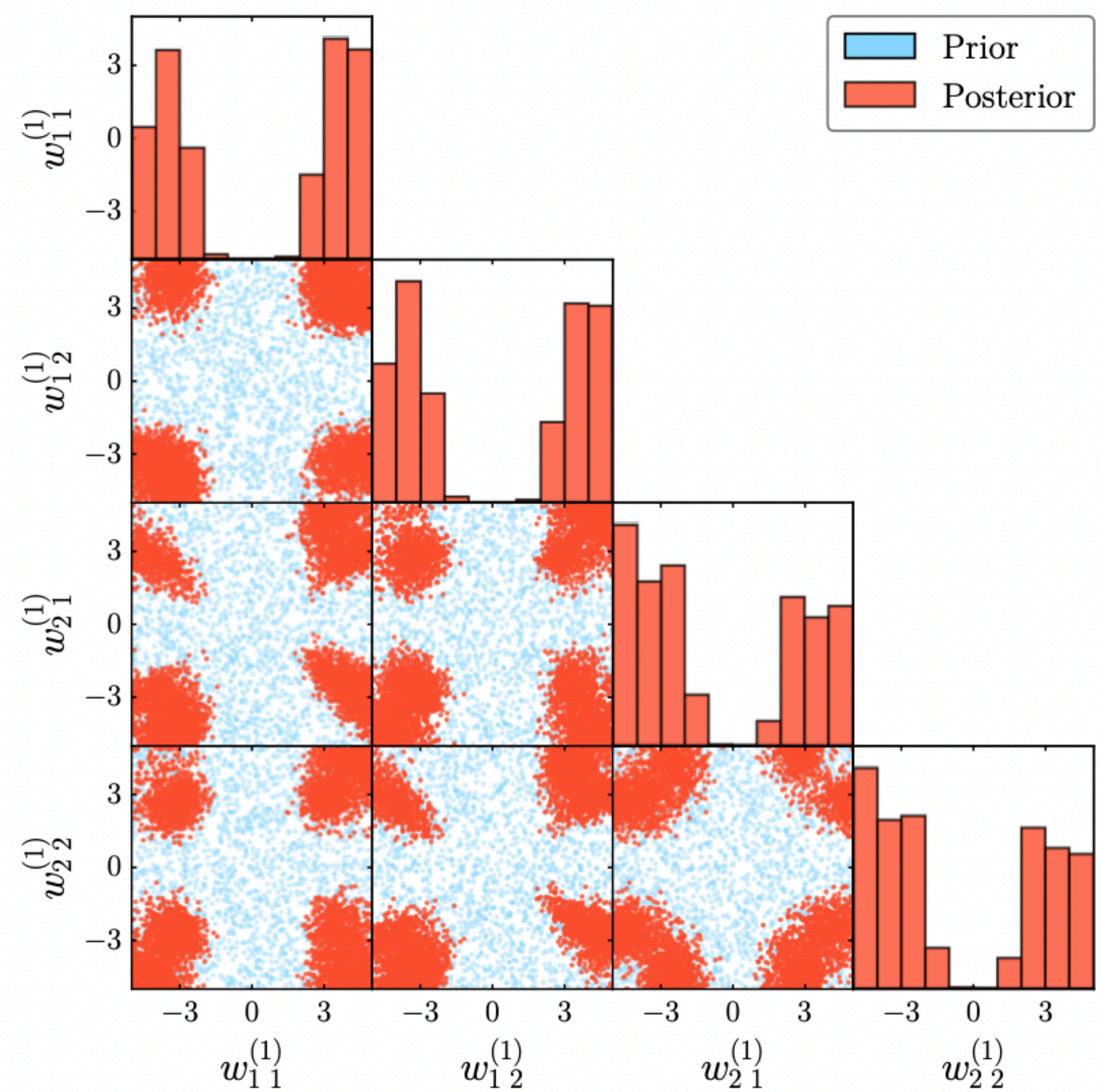
Bayesian Neural Networks



Bayesian Neural Networks



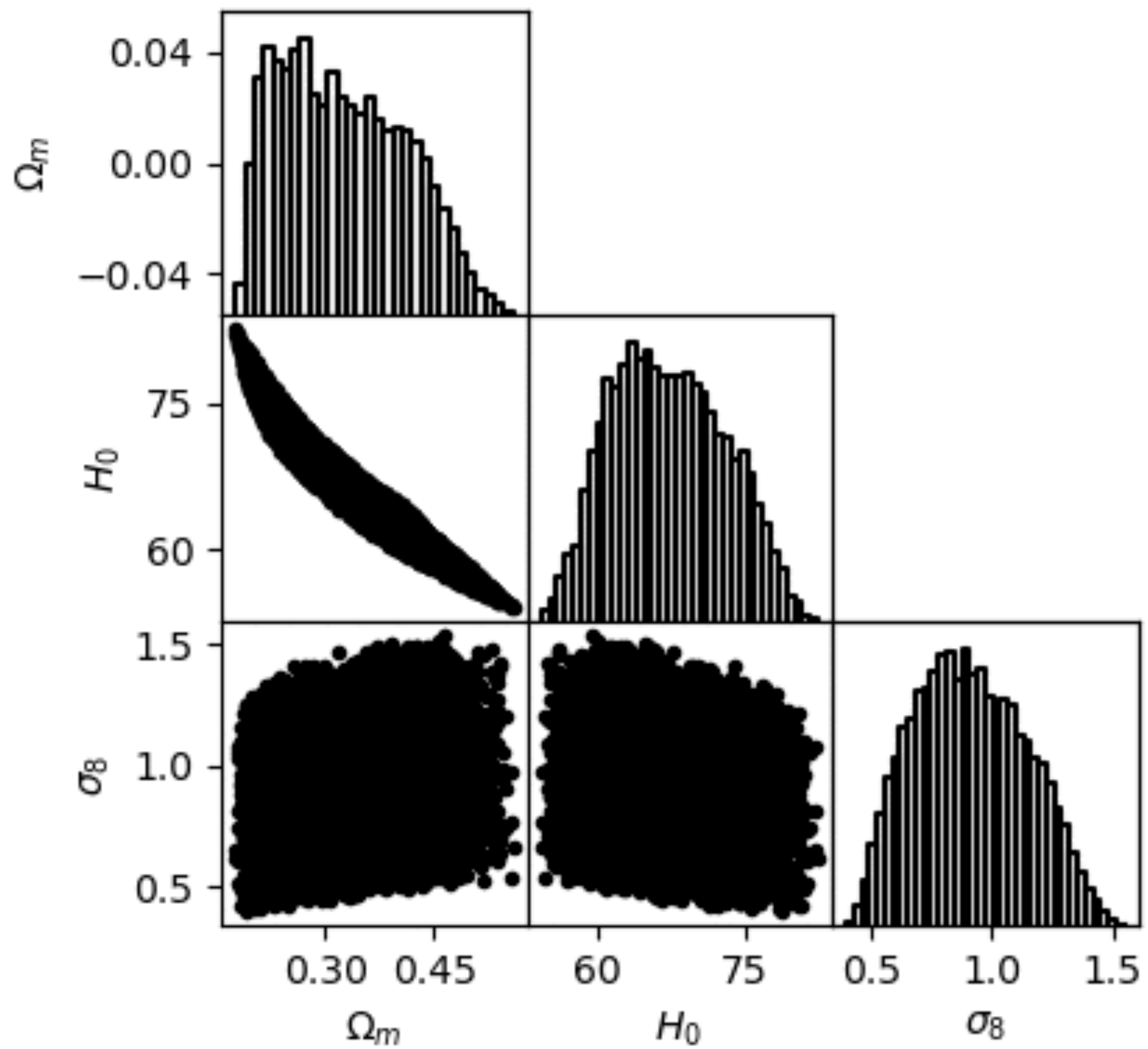
Split personalities in BNNs

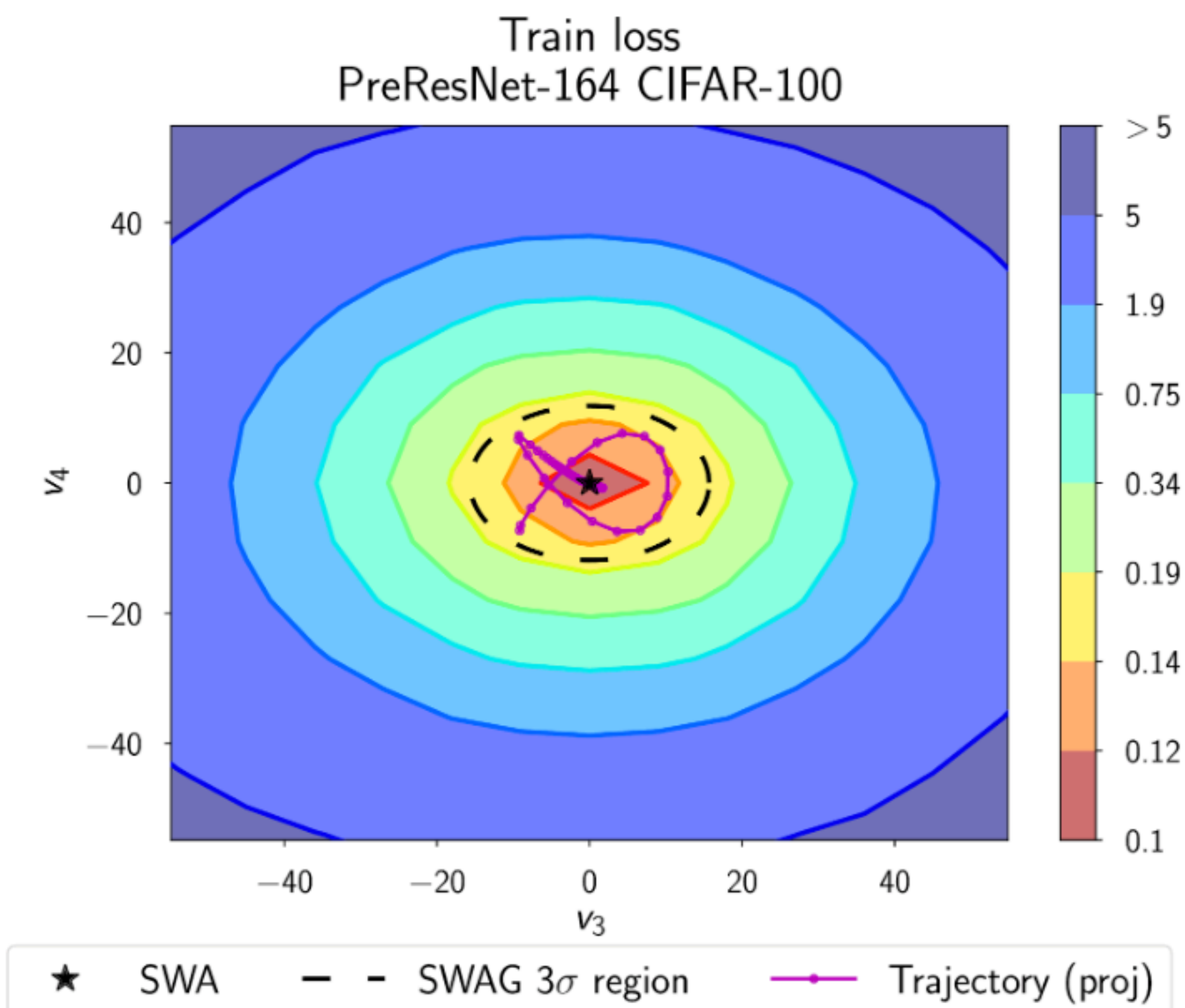


Yallup, Handley, Hobson, Lasenby & Lemos; 2022 — arXiv:2205.11151

How to marginalize?

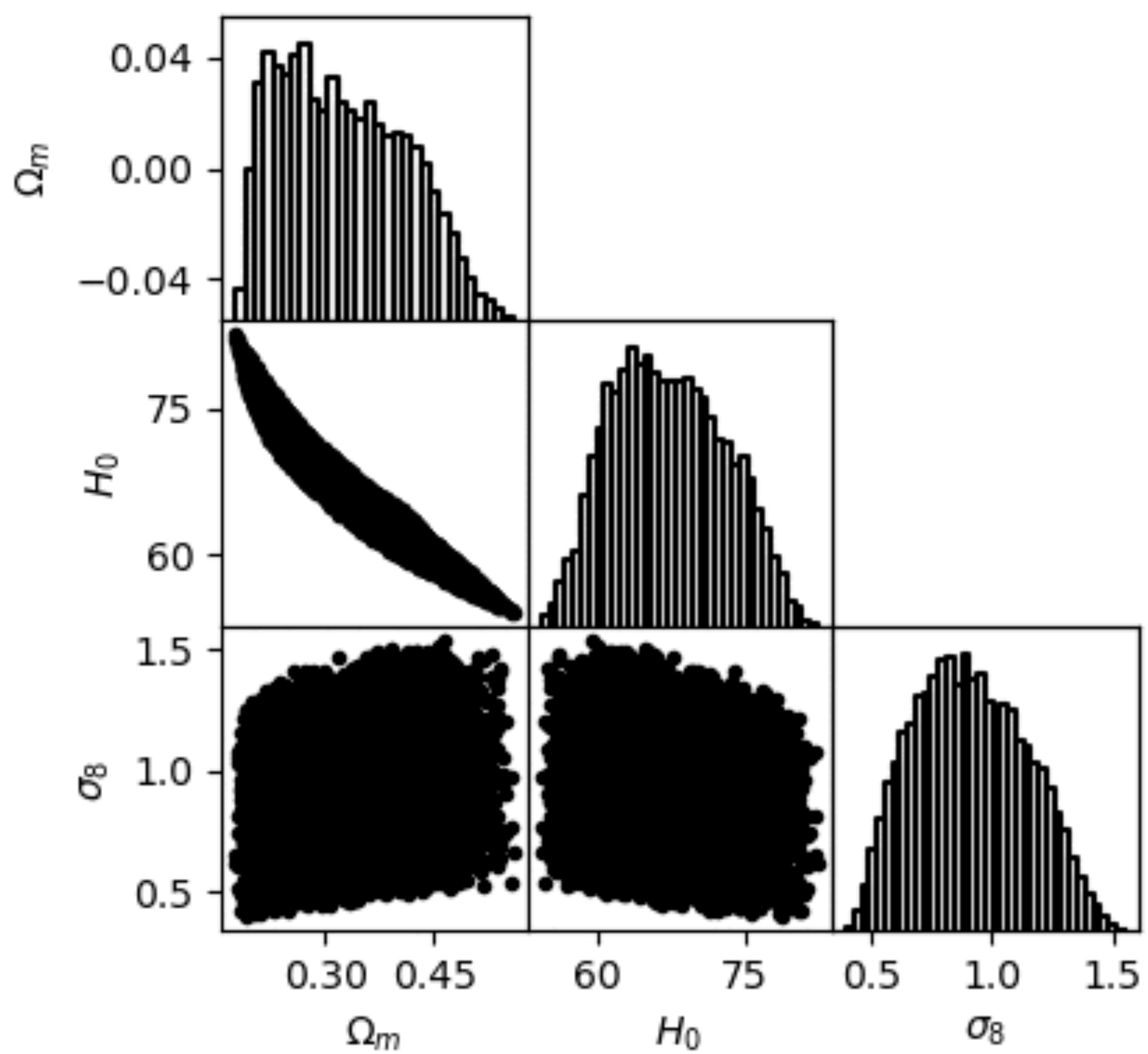
MCMC/ Nested Sampling	Stochastic Weighting Average



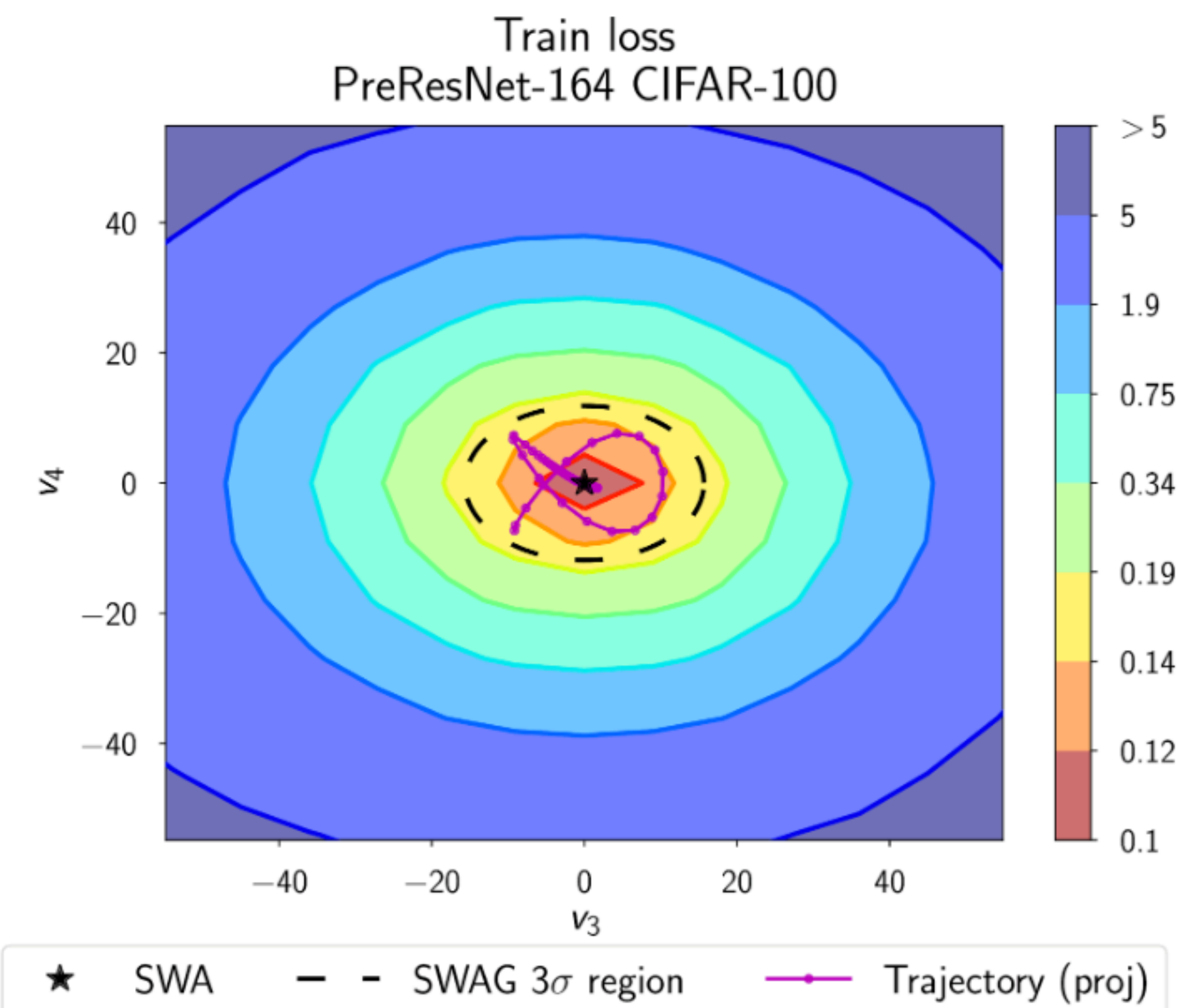


How to marginalize?

MCMC/
Nested Sampling

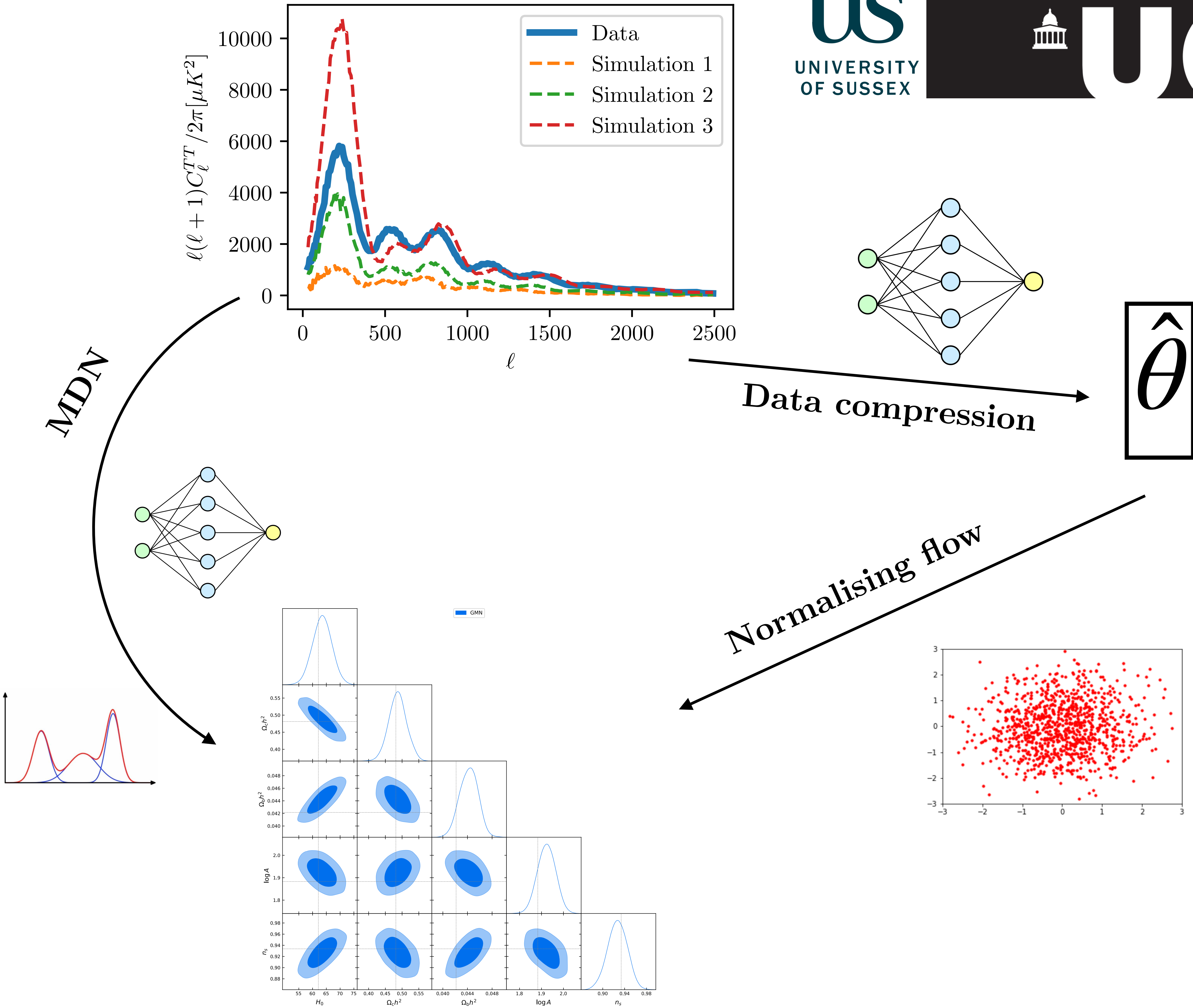


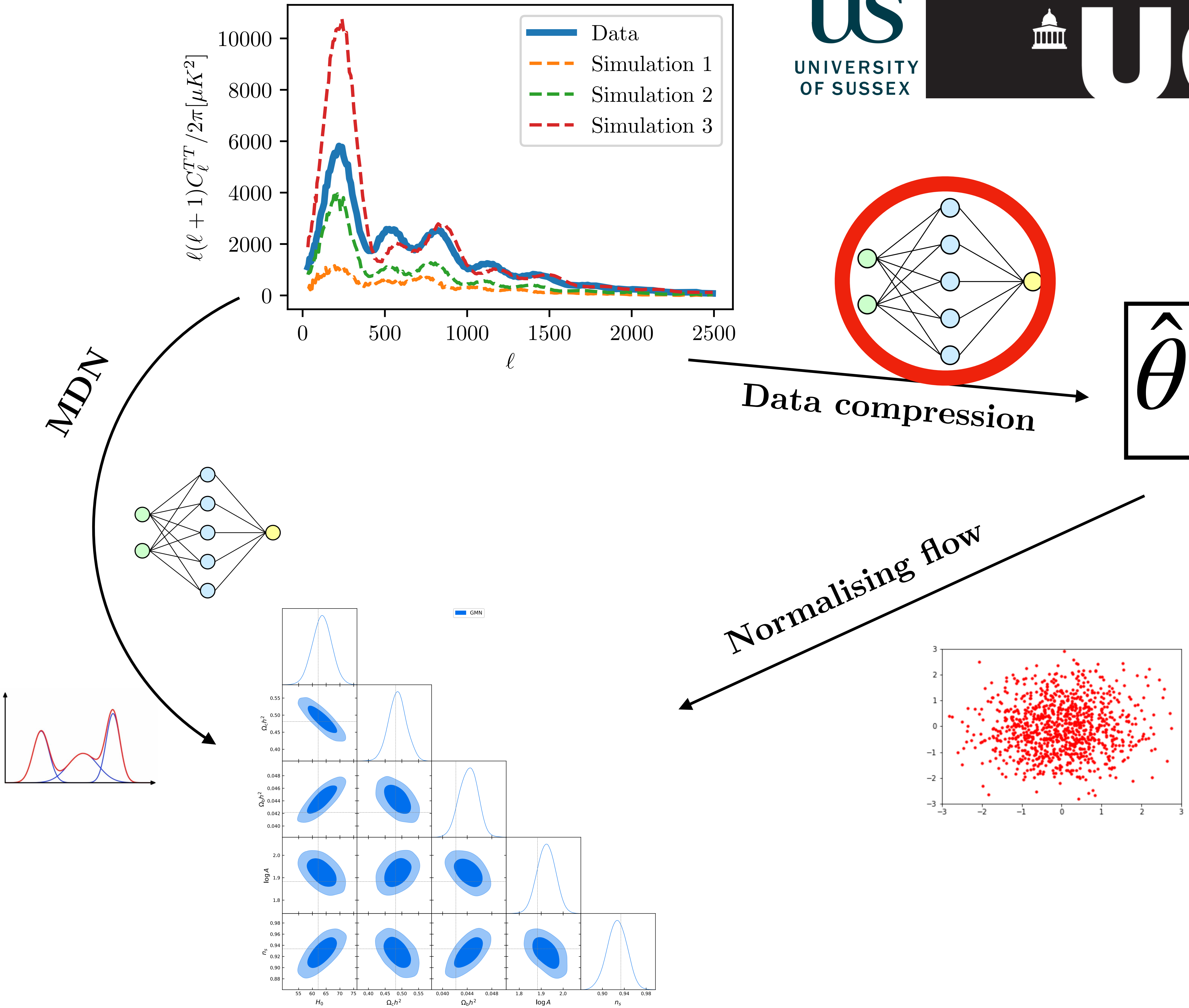
Stochastic
Weighting Average



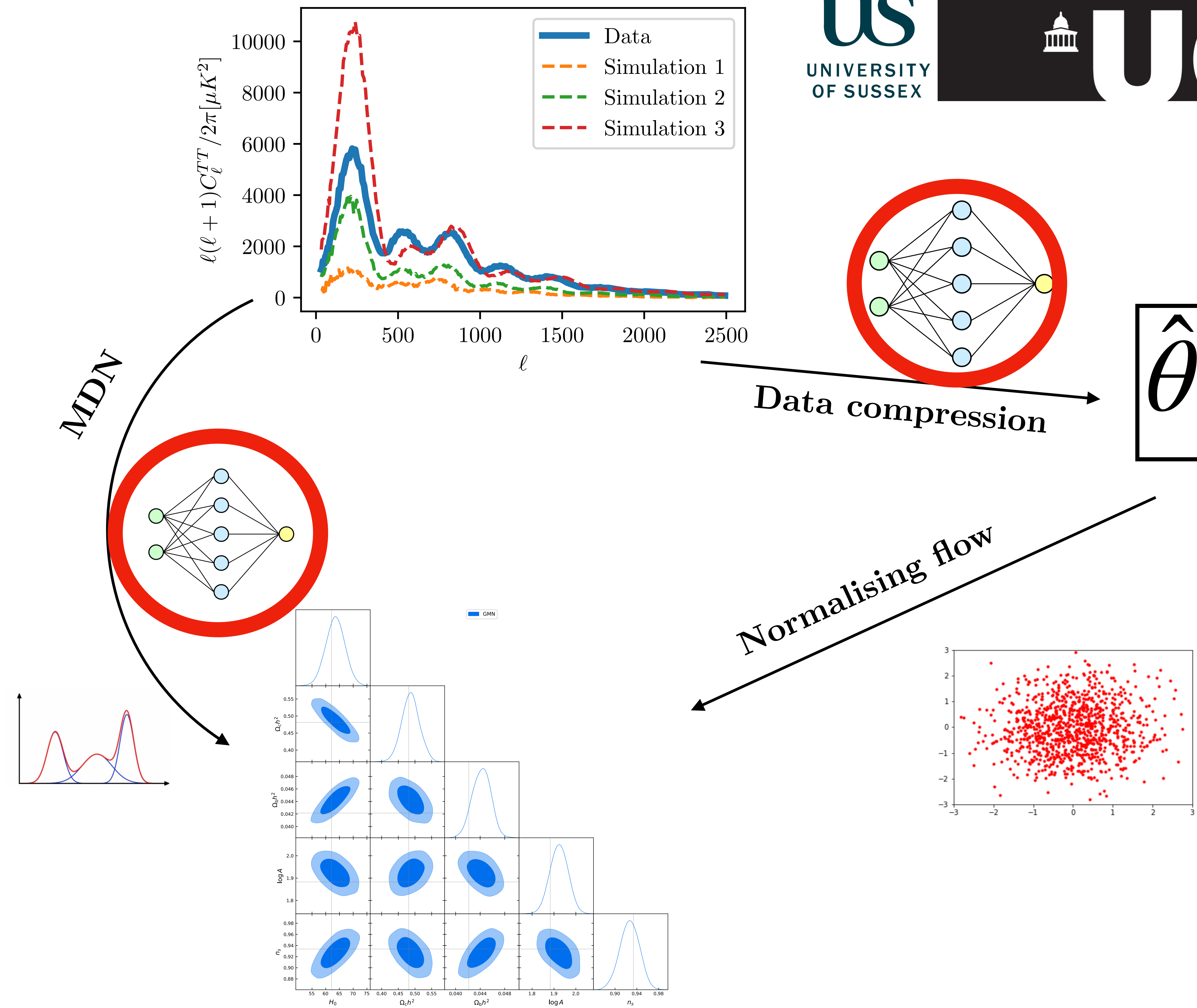
US
UNIVERSITY OF SUSSEX

UCL





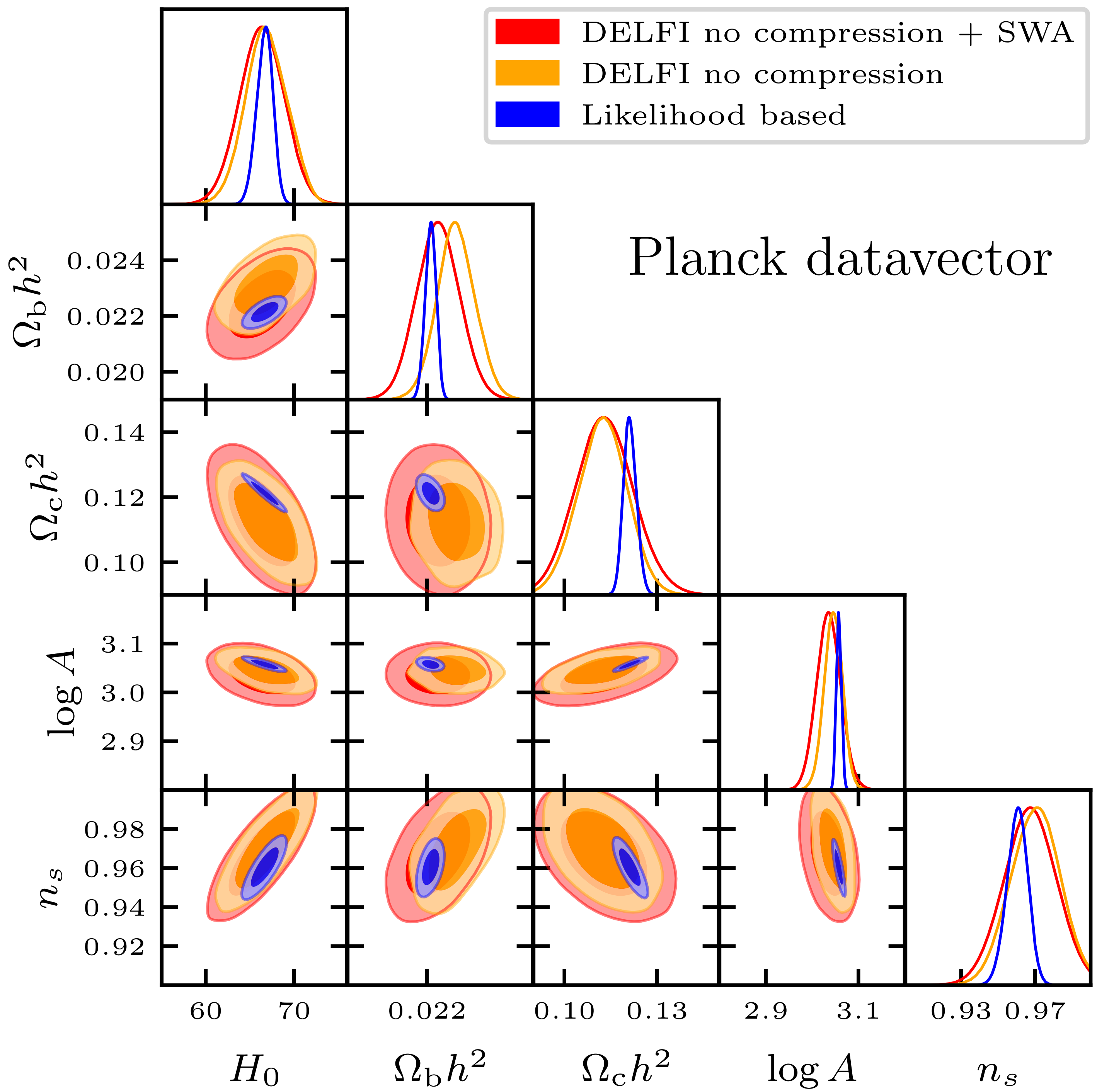
Data
Simulation 1
Simulation 2
Simulation 3
$\ell(\ell+1)C_\ell^{TT}/2\pi[\mu K^2]$
ℓ
MDN
Data compression
$\hat{\theta}$
Normalising flow
GMN
H_0
$\Omega_c h^2$
$\Omega_b h^2$
$\log A$
n_s



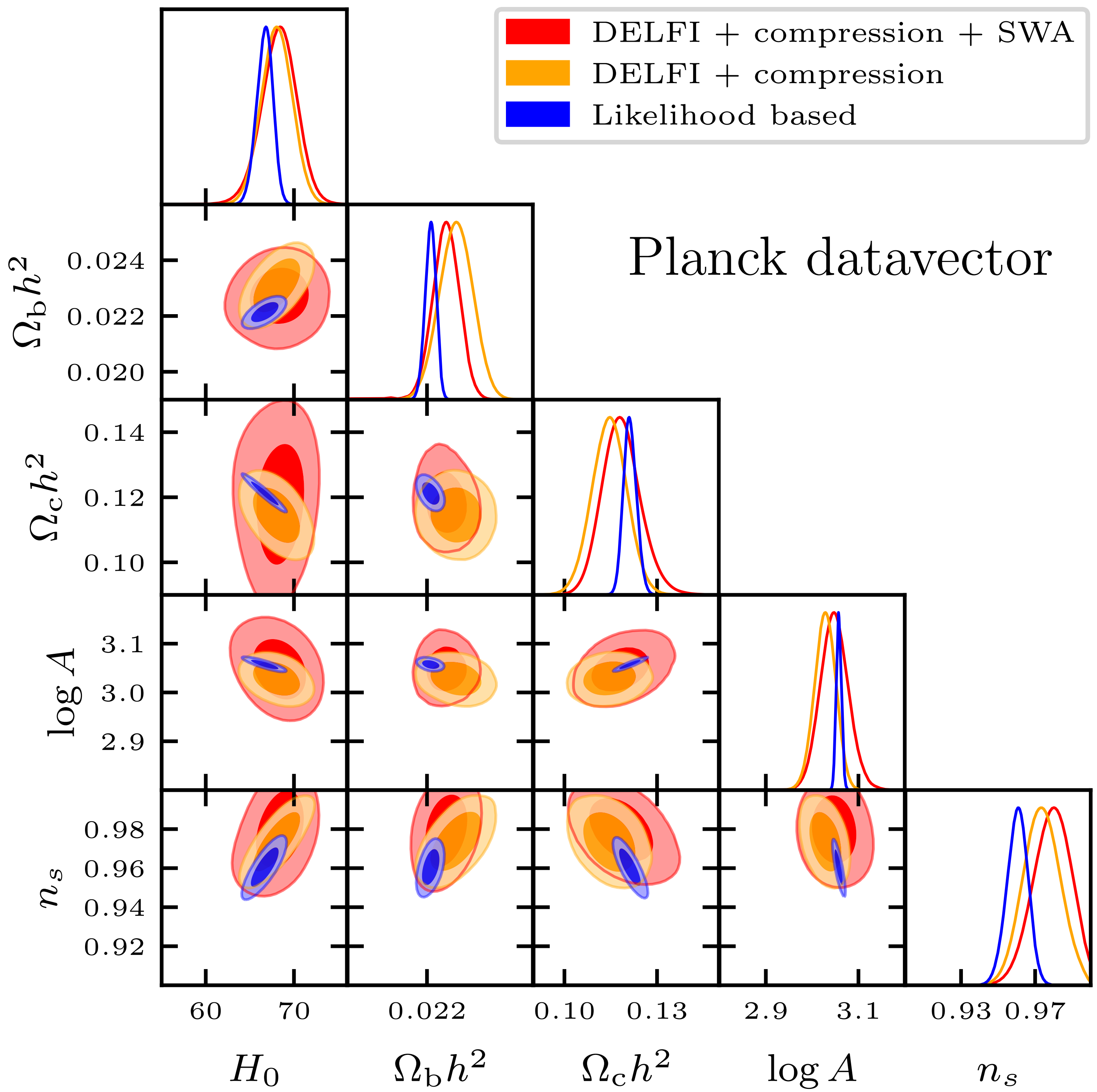
Data
Simulation 1
Simulation 2
Simulation 3
$\ell(\ell+1)C_\ell^{TT}/2\pi[\mu K^2]$
ℓ
Data compression
$\hat{\theta}$
Normalising flow
MDN
GMN
H_0
$\Omega_c h^2$
$\Omega_b h^2$
$\log A$
n_s

Results

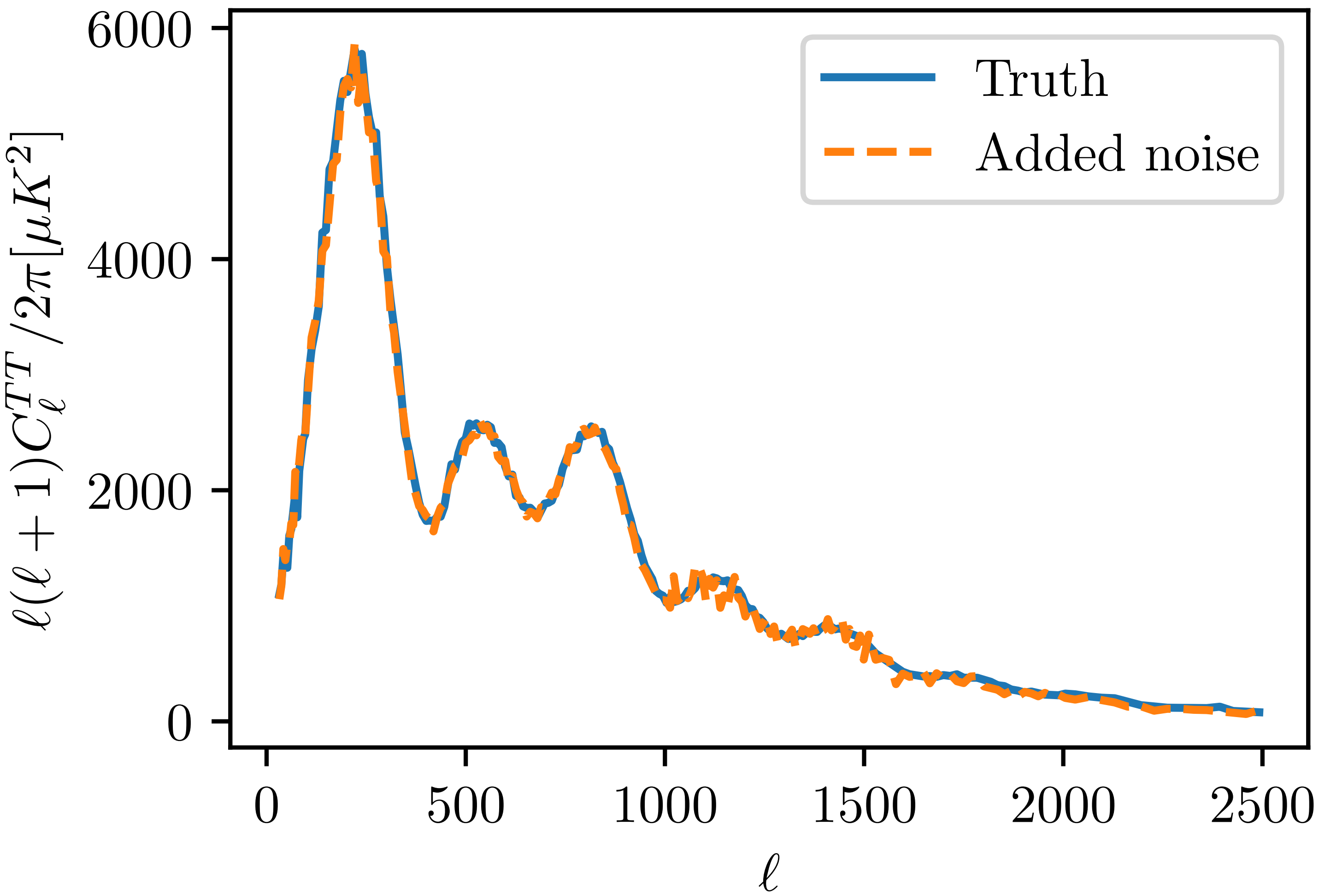
MDN - No compression



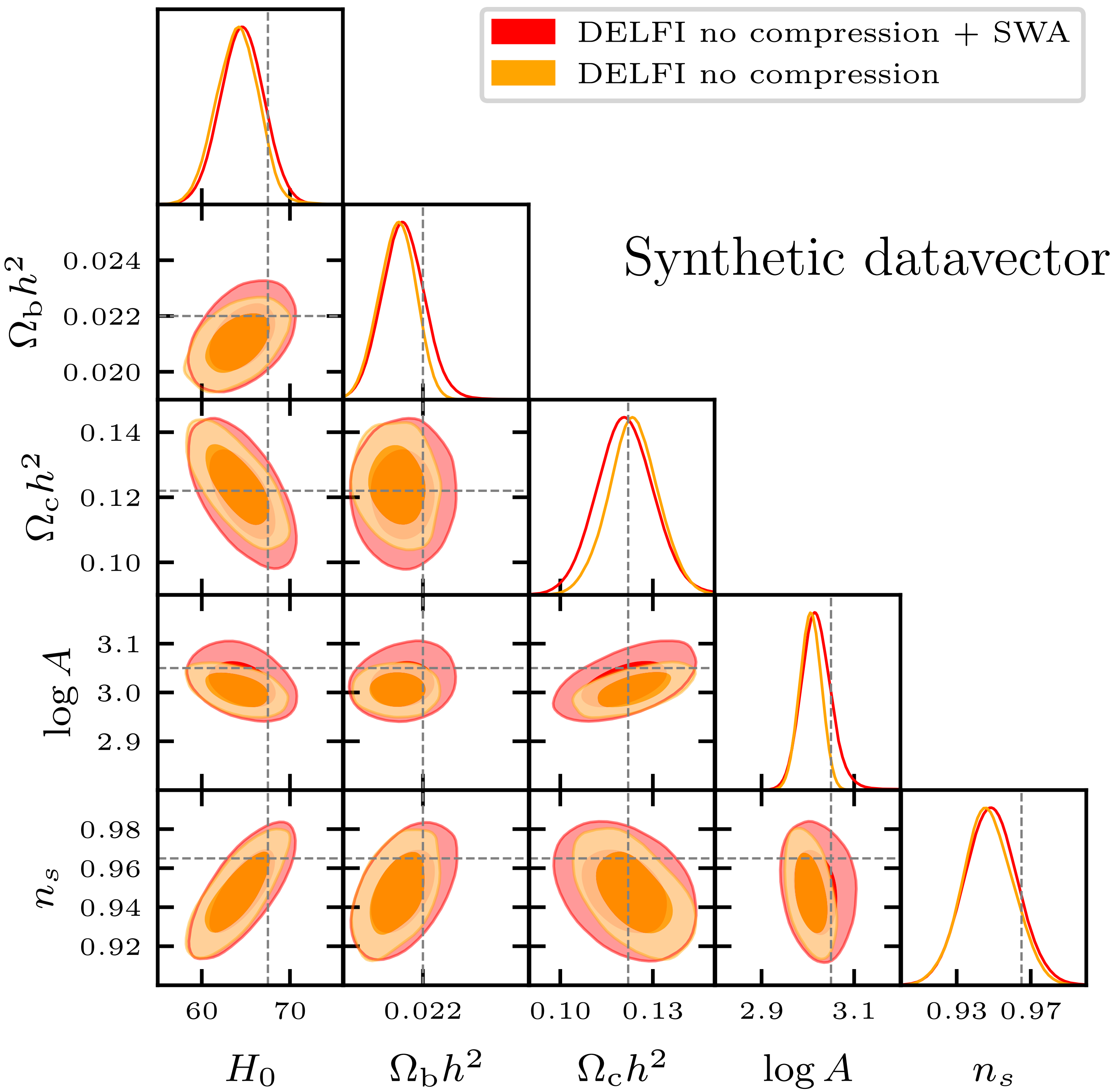
Normalizing flow - compression



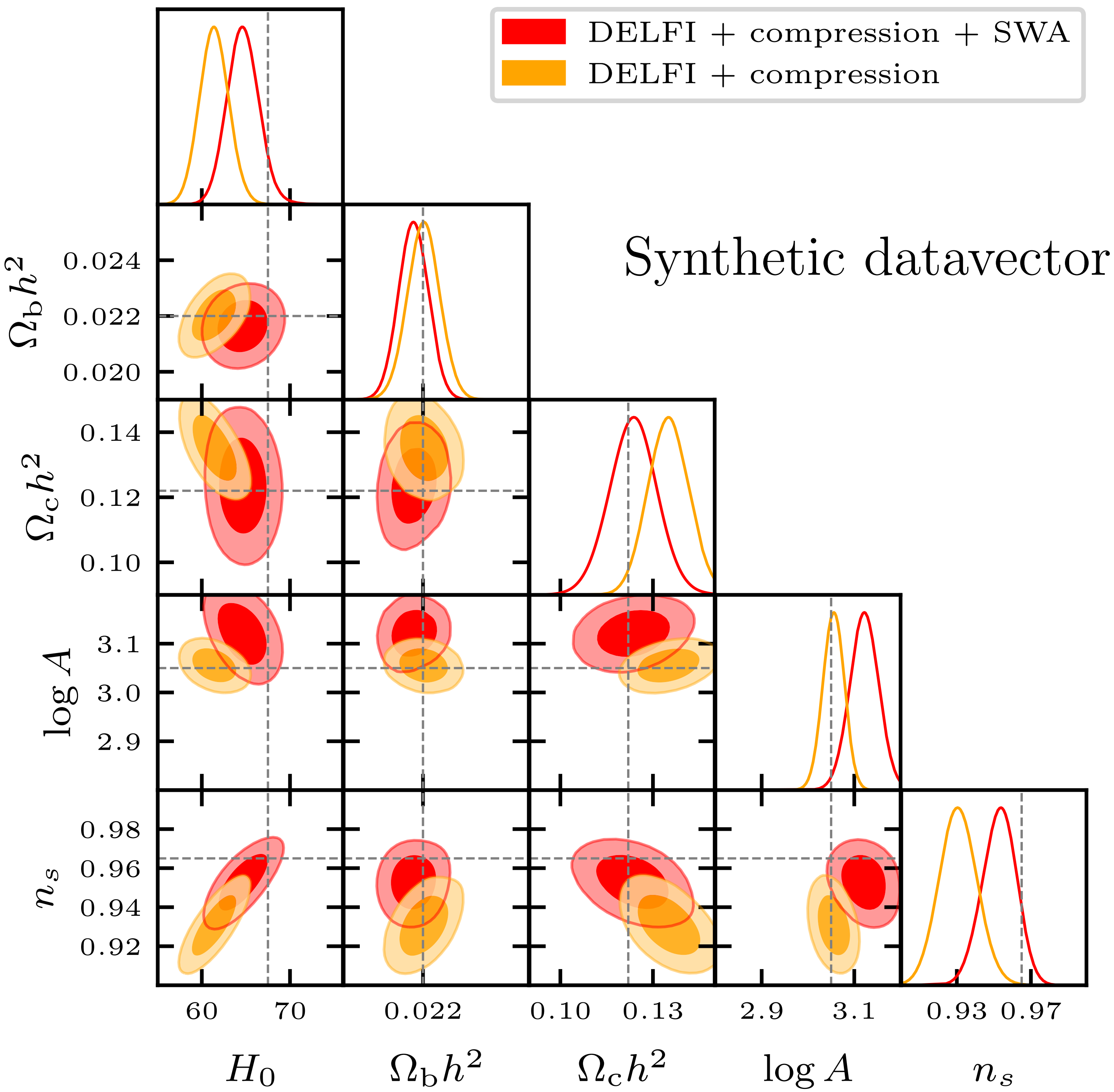
Noisy datavector



MDN - No compression



Normalizing flow - compression



Conclusions

- Realistic SBI has limited number of simulations, and imperfect forward models.

- BNNs produce more robust SBI in realistic conditions.

- Approximate marginalisation with SWA is easy (our code will be public soon

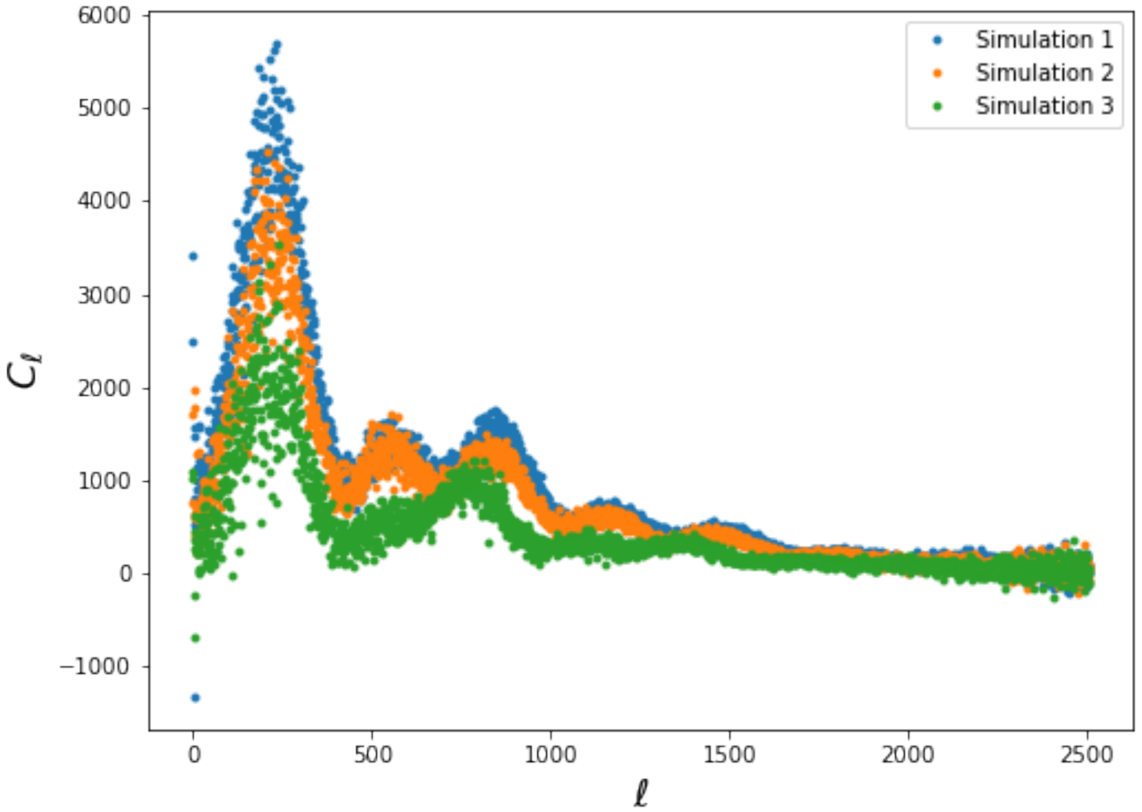
Thanks!

p.lemos@sussex.ac.uk

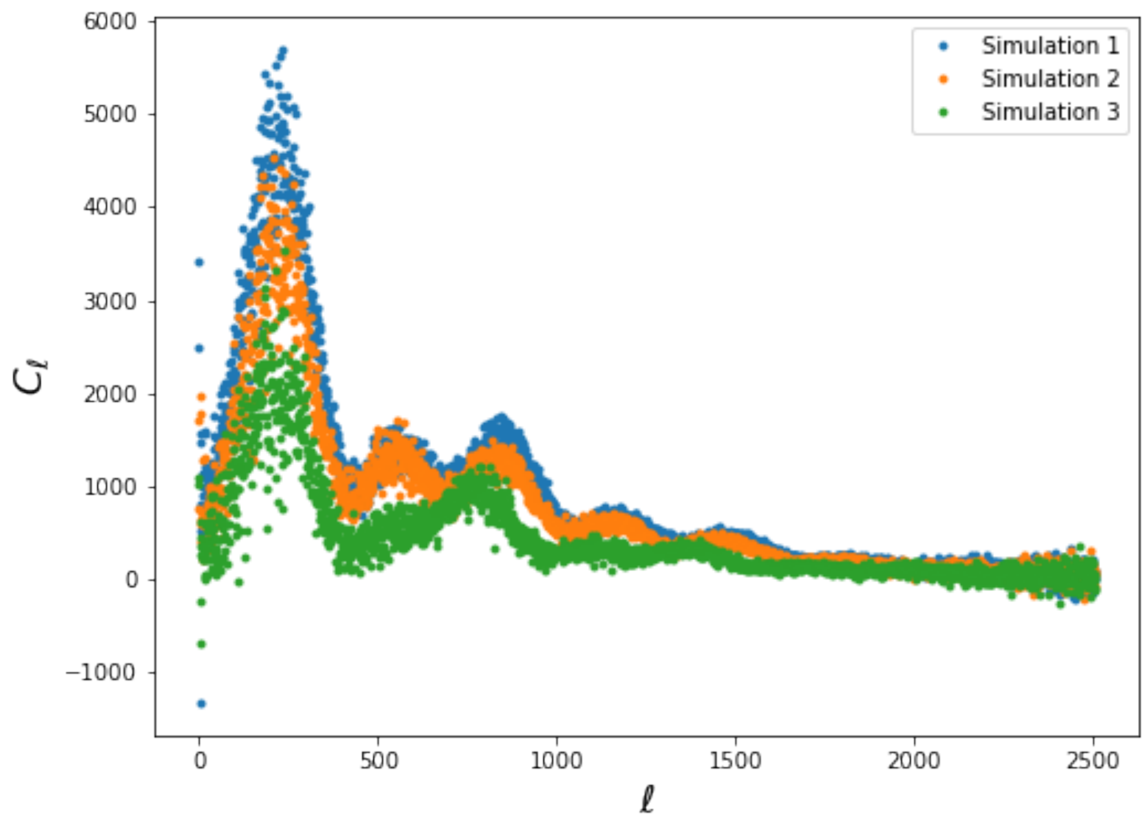
PabloLemosP

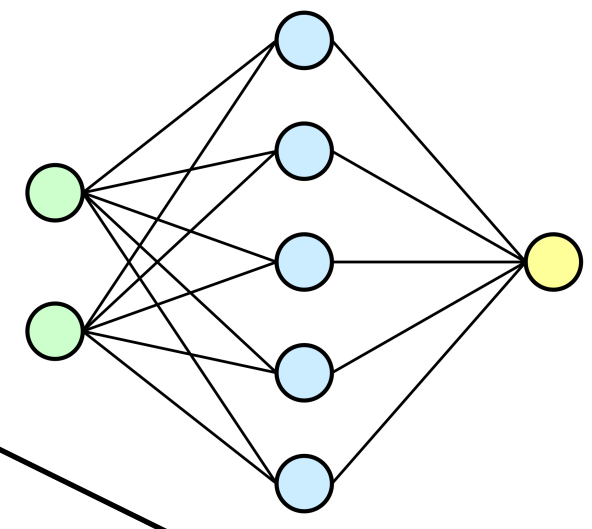
https://pablo-lemos.github.io

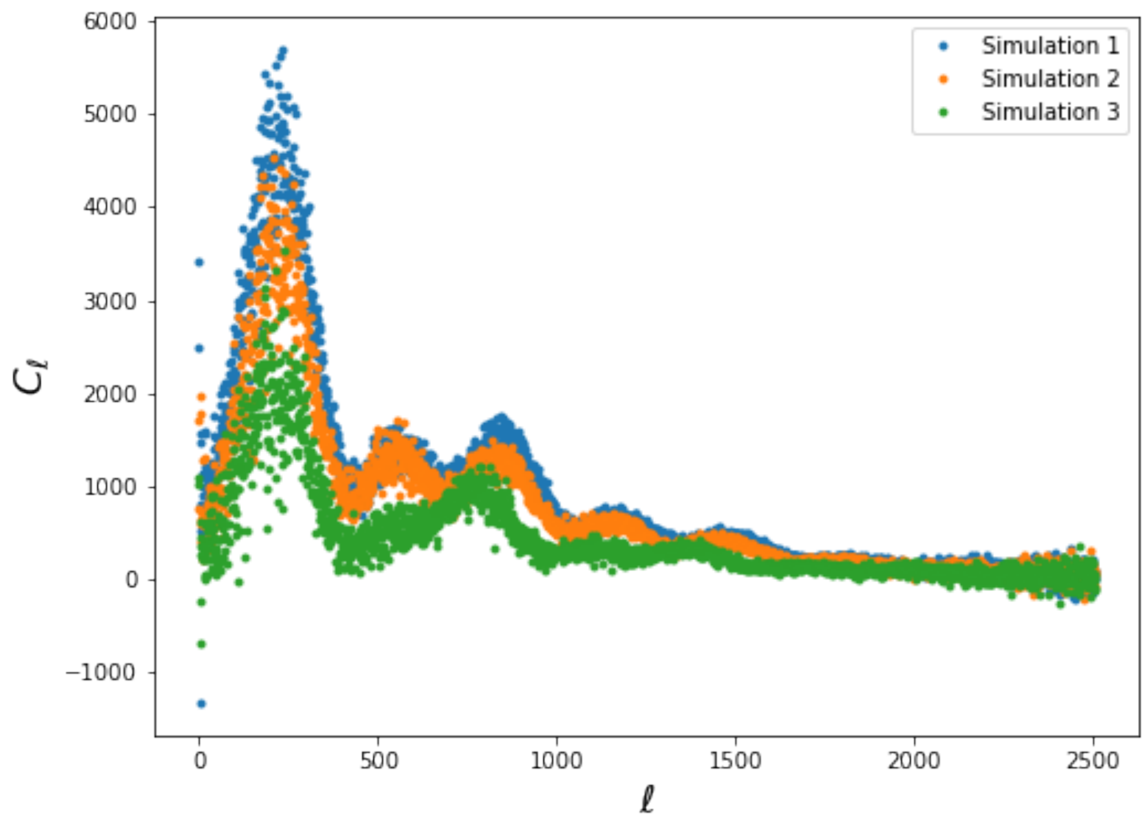
OLD

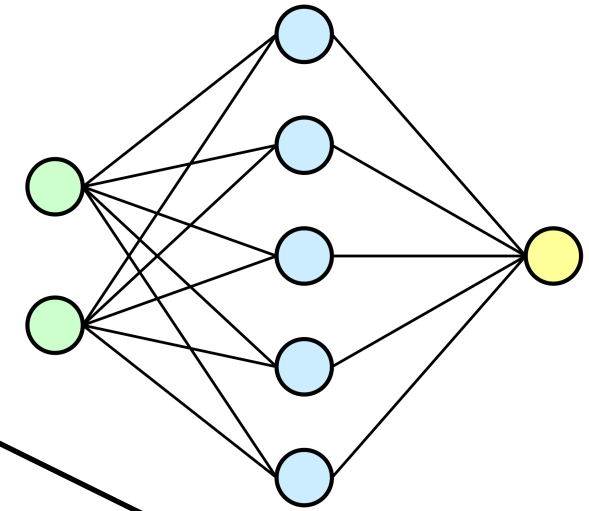


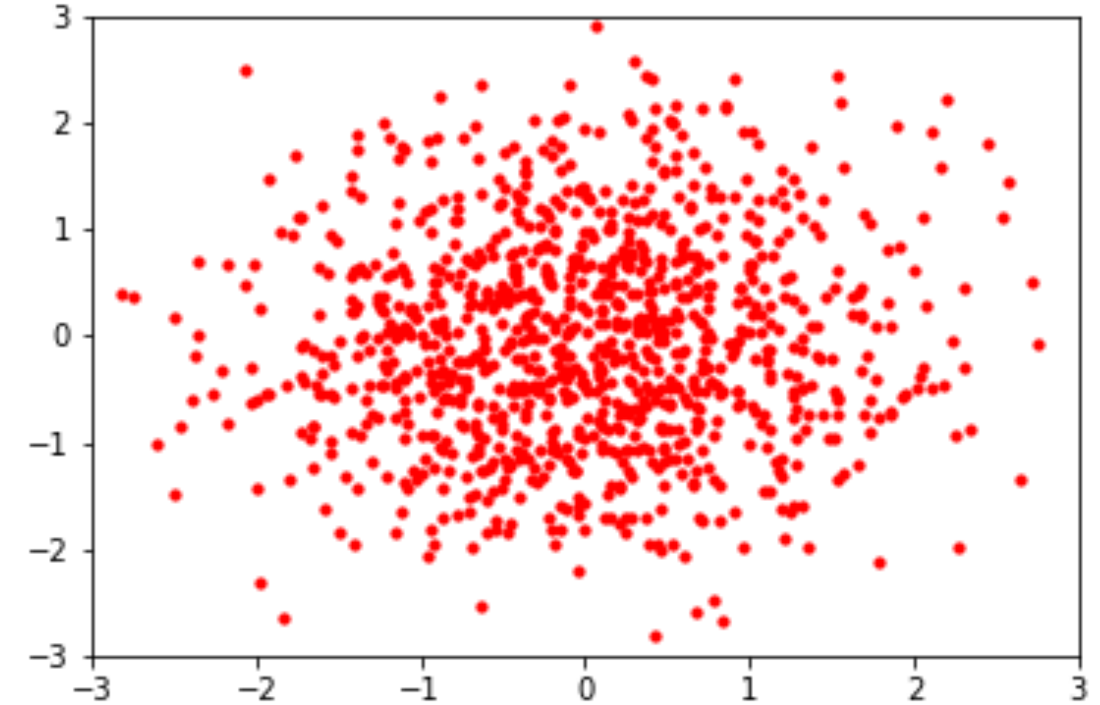
Simulation 1
Simulation 2
Simulation 3
6000
5000
4000
3000
2000
1000
0
−1000
C_ℓ
0
500
1000
1500
2000
2500
ℓ

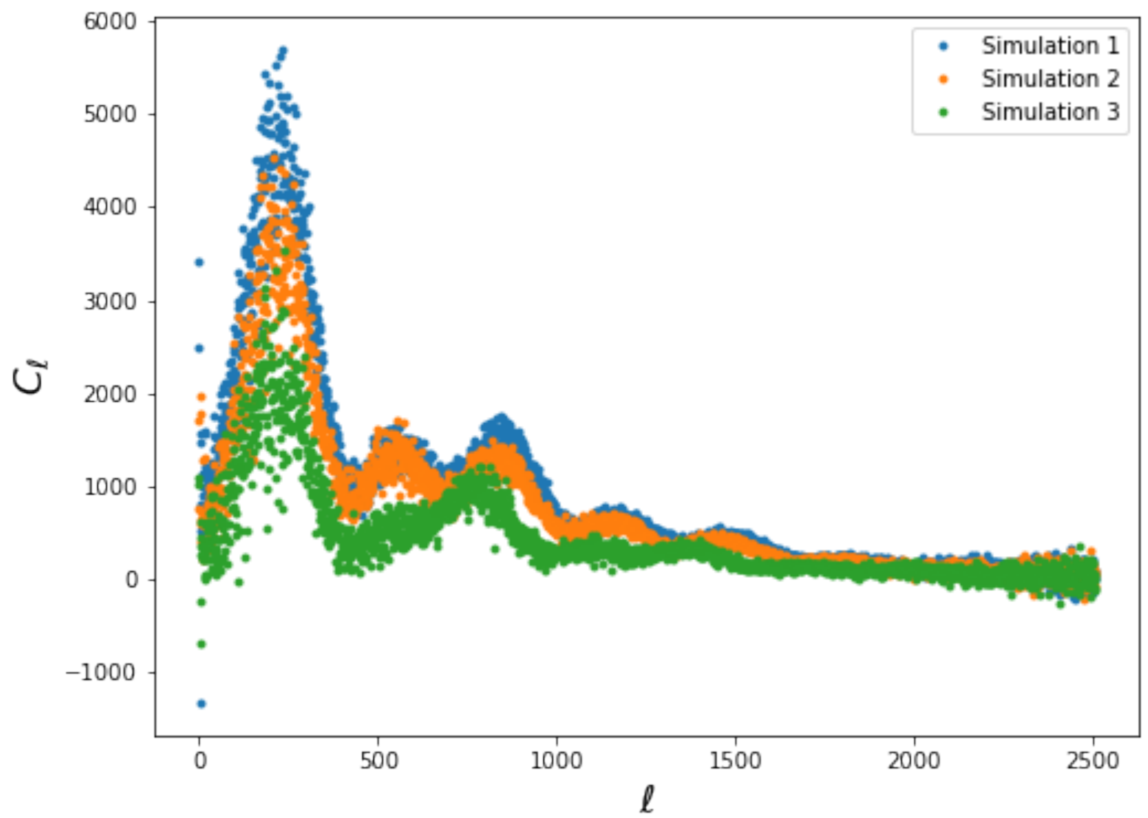

Simulation 1
Simulation 2
Simulation 3
6000
5000
4000
3000
2000
1000
0
−1000
C_ℓ
0
500
1000
1500
2000
2500
ℓ

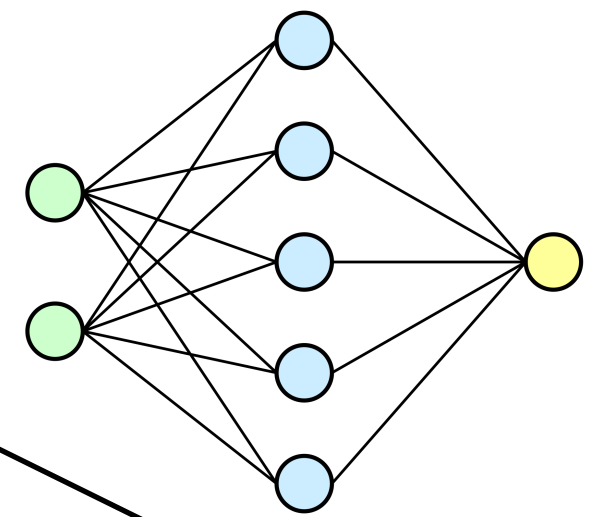


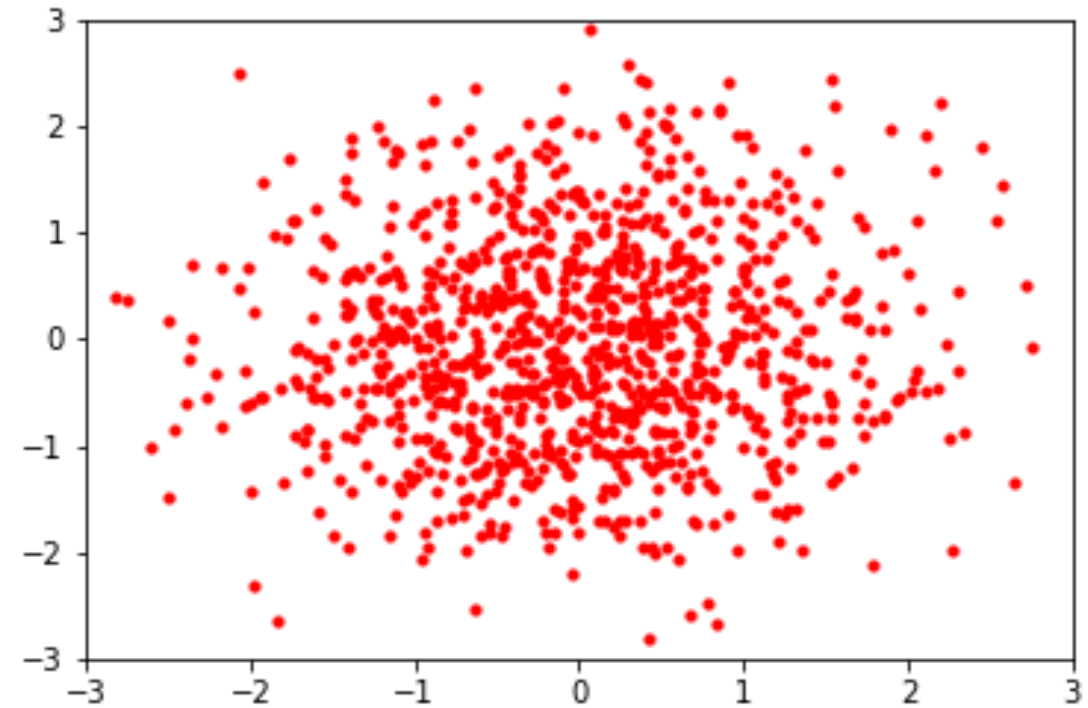

Simulation 1
Simulation 2
Simulation 3
6000
5000
4000
3000
2000
1000
0
−1000
C_ℓ
0
500
1000
1500
2000
2500
ℓ




3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3

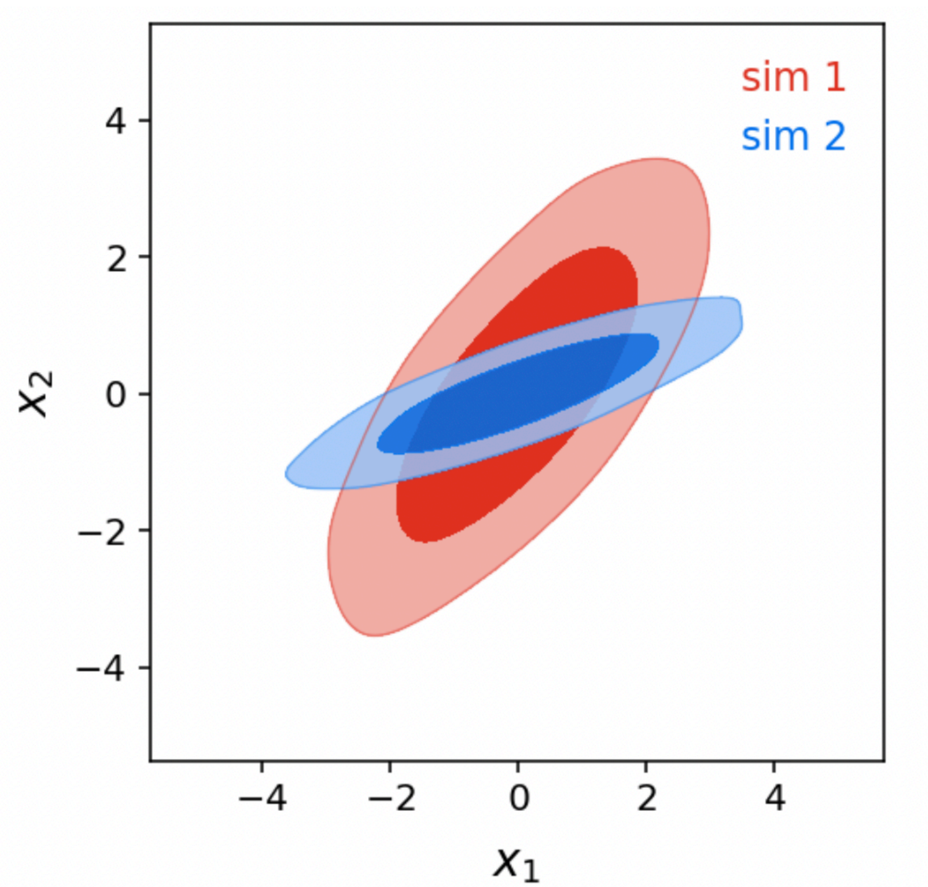




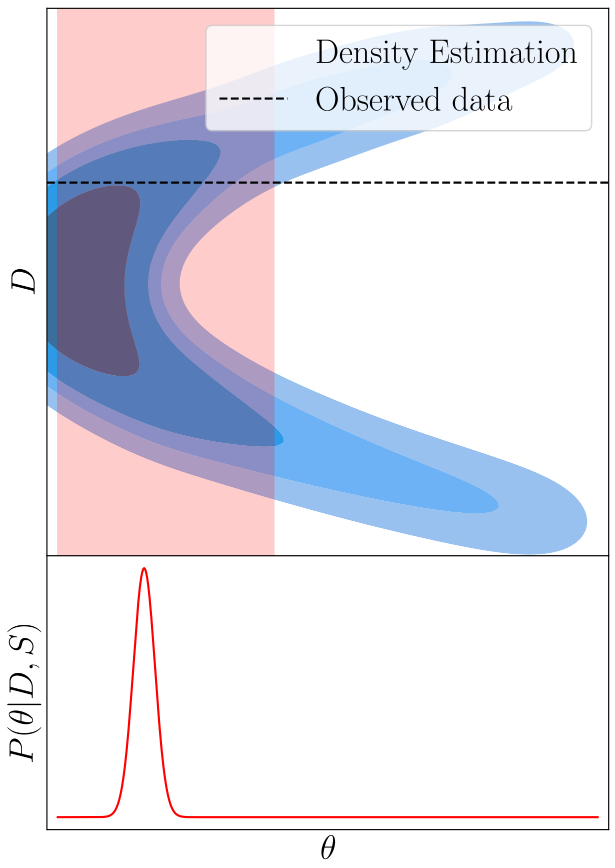


Observed
Data

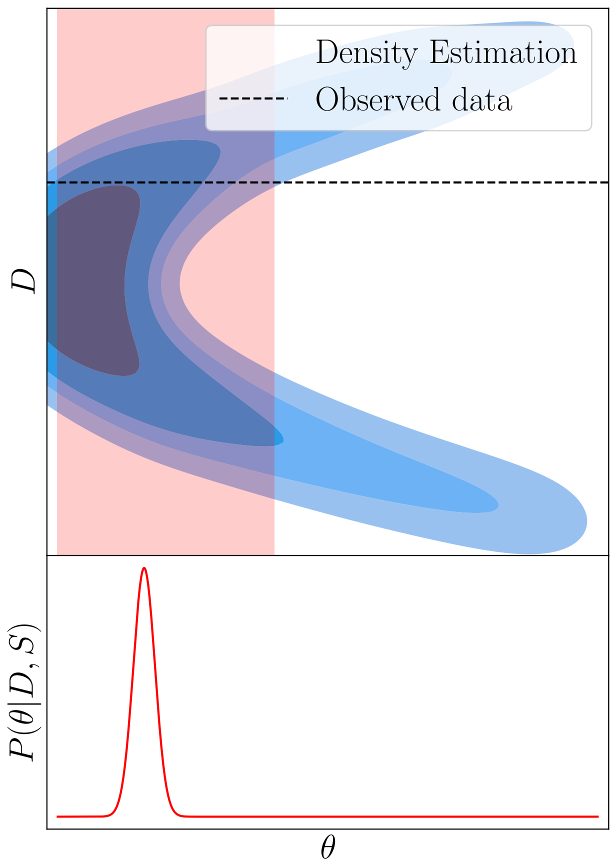
$$P(\theta \,|\, D, \mathrm{Sims})$$



$$P(\theta \,|\, D, S)$$

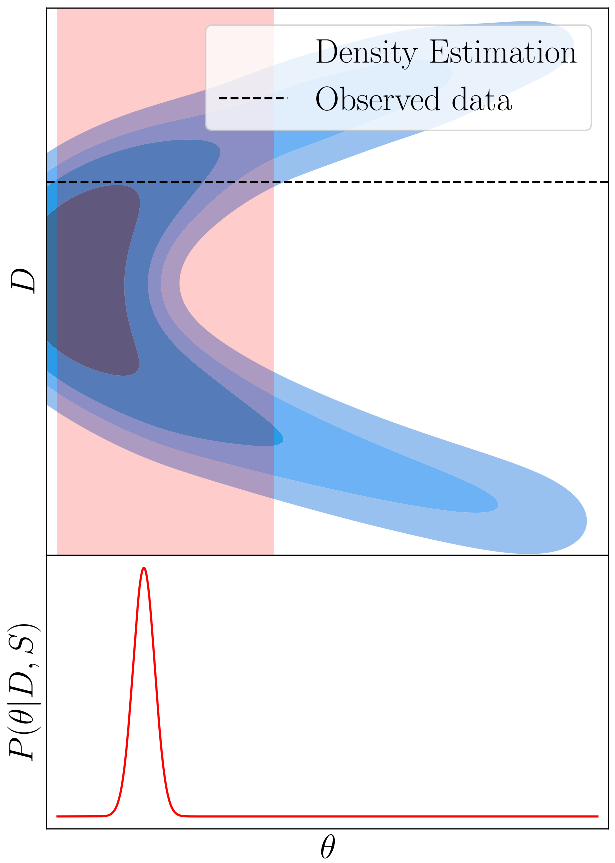


$$P(\theta \mid D, S) = \int \mathrm{d}\alpha \; P(\theta, \alpha \mid D, S) P(\alpha \mid D, S)$$



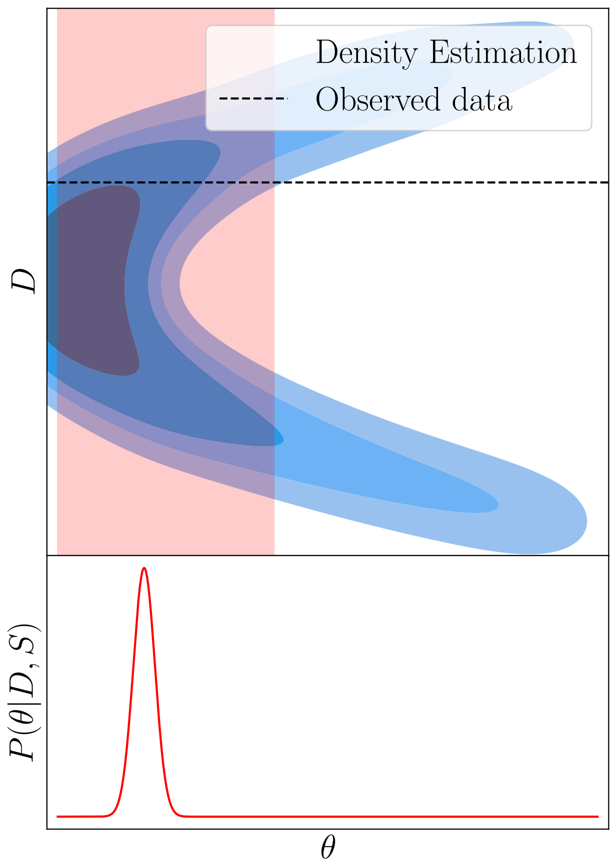
$$P(\theta \,|\, D, S) = \int \mathrm{d}\alpha \; P(\theta, \alpha \,|\, D, S) P(\alpha \,|\, D, S)$$

$$= \int \mathrm{d}\alpha \; P(\theta, \alpha \,|\, D) P(\alpha \,|\, S)$$



$$P(\theta\,|\,D,S) = \int \mathrm{d}\alpha\; P(\theta,\alpha\,|\,D,S)P(\alpha\,|\,D,S)$$

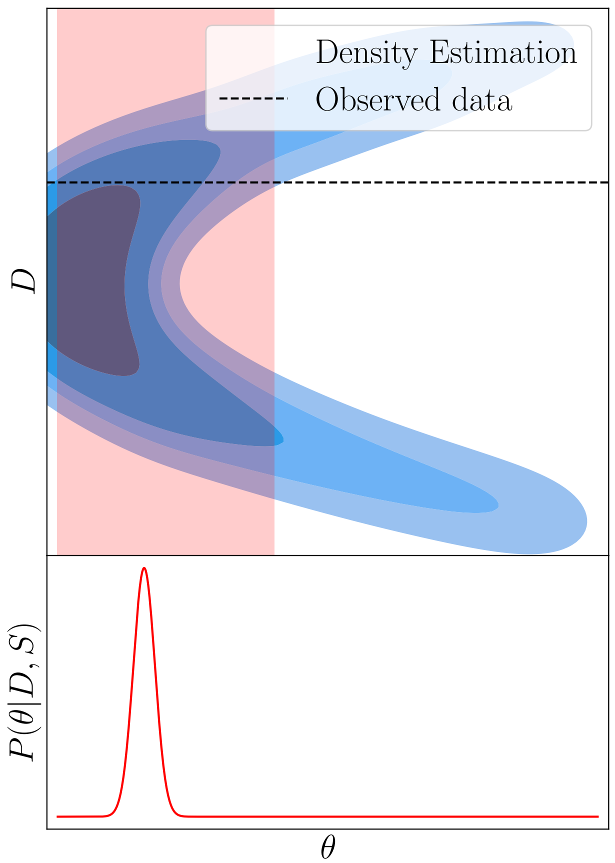
$$= \int \mathrm{d}\alpha\; P(\theta,\alpha\,|\,D)P(\alpha\,|\,S)$$



$$P(\alpha\,|\,S) = \delta(\alpha - \alpha_{\mathrm{BF}})$$

$$P(\theta \,|\, D, S) = \int \mathrm{d}\alpha \; P(\theta, \alpha \,|\, D, S) P(\alpha \,|\, D, S)$$

$$= \int \mathrm{d}\alpha \; P(\theta, \alpha \,|\, D) P(\alpha \,|\, S)$$

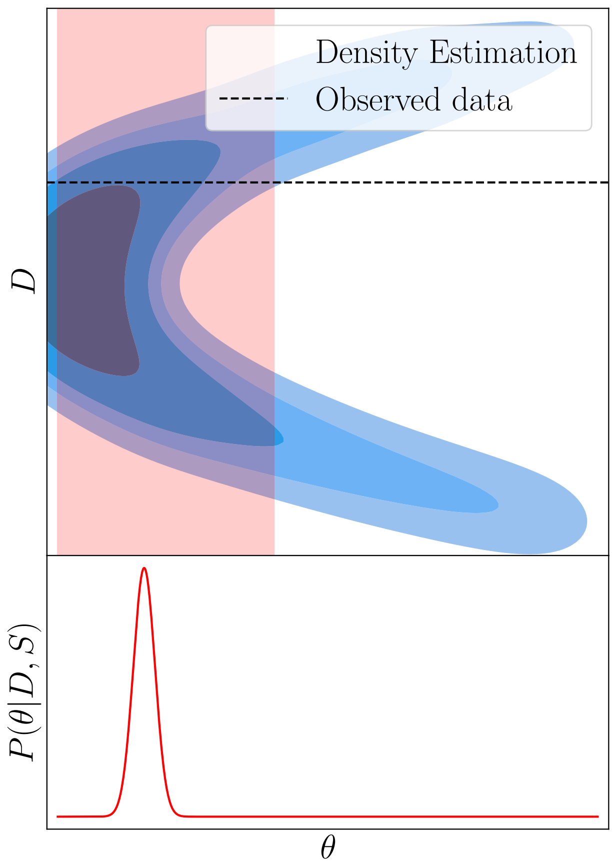


$$P(\alpha \,|\, S) = \delta(\alpha - \alpha_{\mathrm{BF}})$$

Marginalization

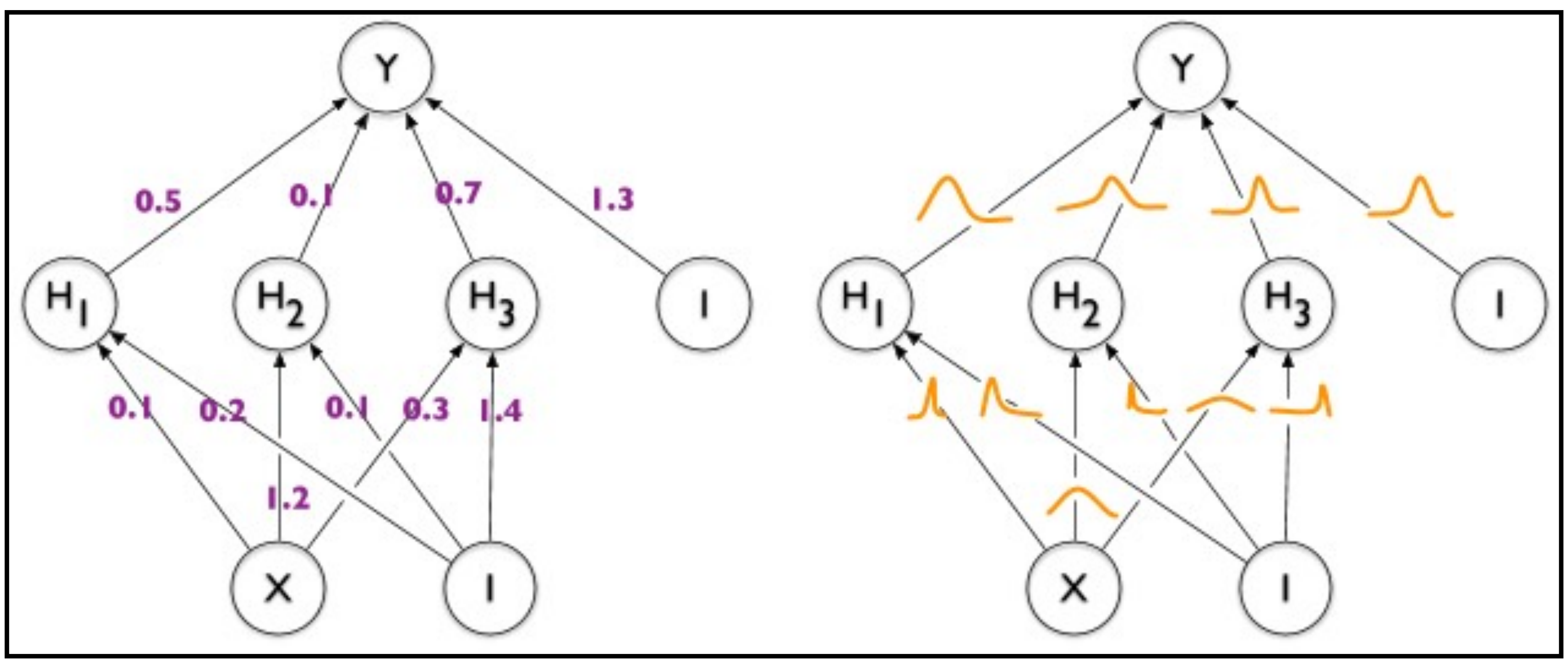
$$P(\theta\,|\,D,S) = \int \mathrm{d}\alpha\; P(\theta,\alpha\,|\,D,S)P(\alpha\,|\,D,S)$$

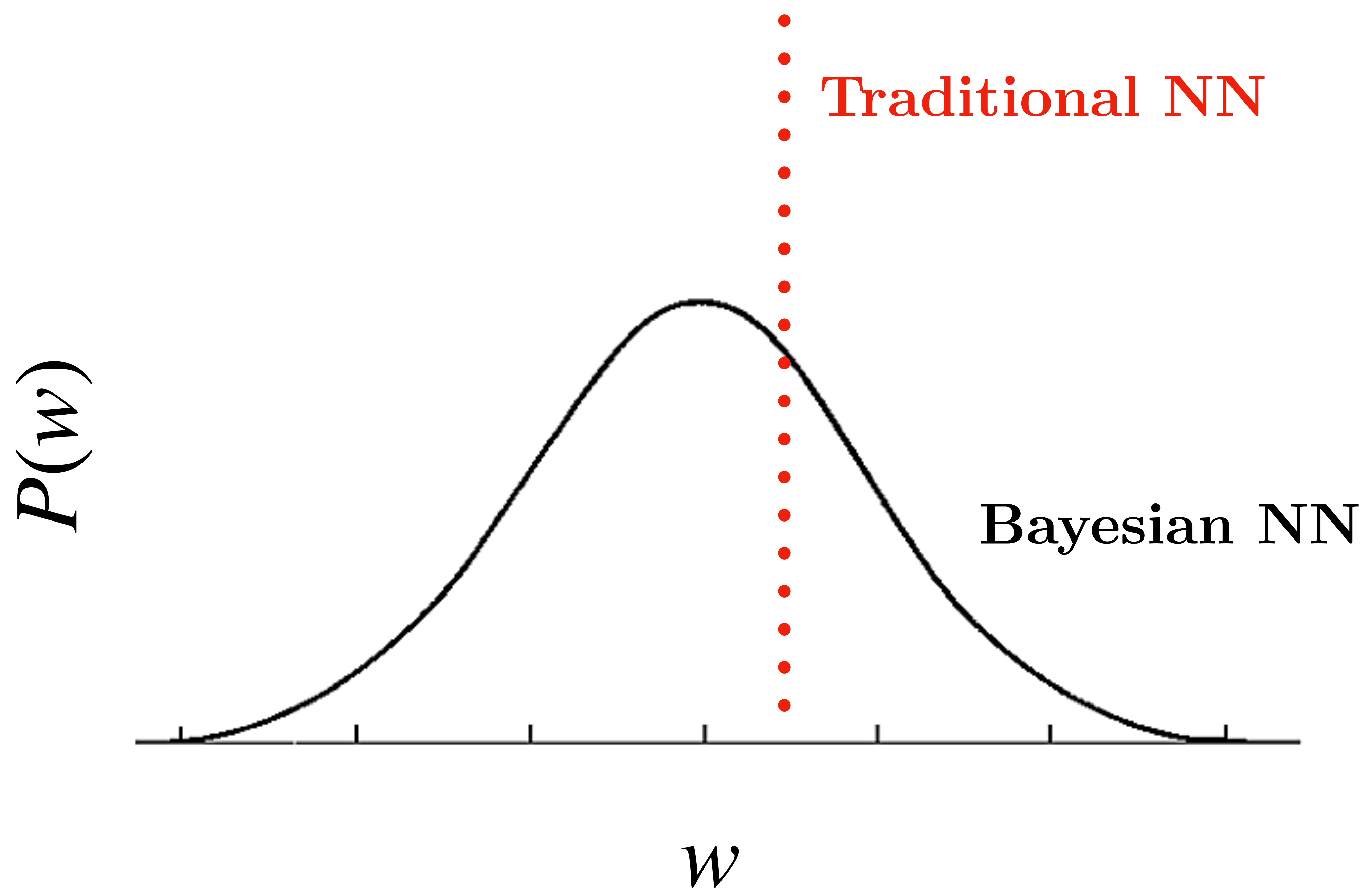
$$= \int \mathrm{d}\alpha\; P(\theta,\alpha\,|\,D)P(\alpha\,|\,S)$$



$$P(\alpha\,|\,S) = \delta(\alpha - \alpha_{\mathrm{BF}})$$

Marginalization



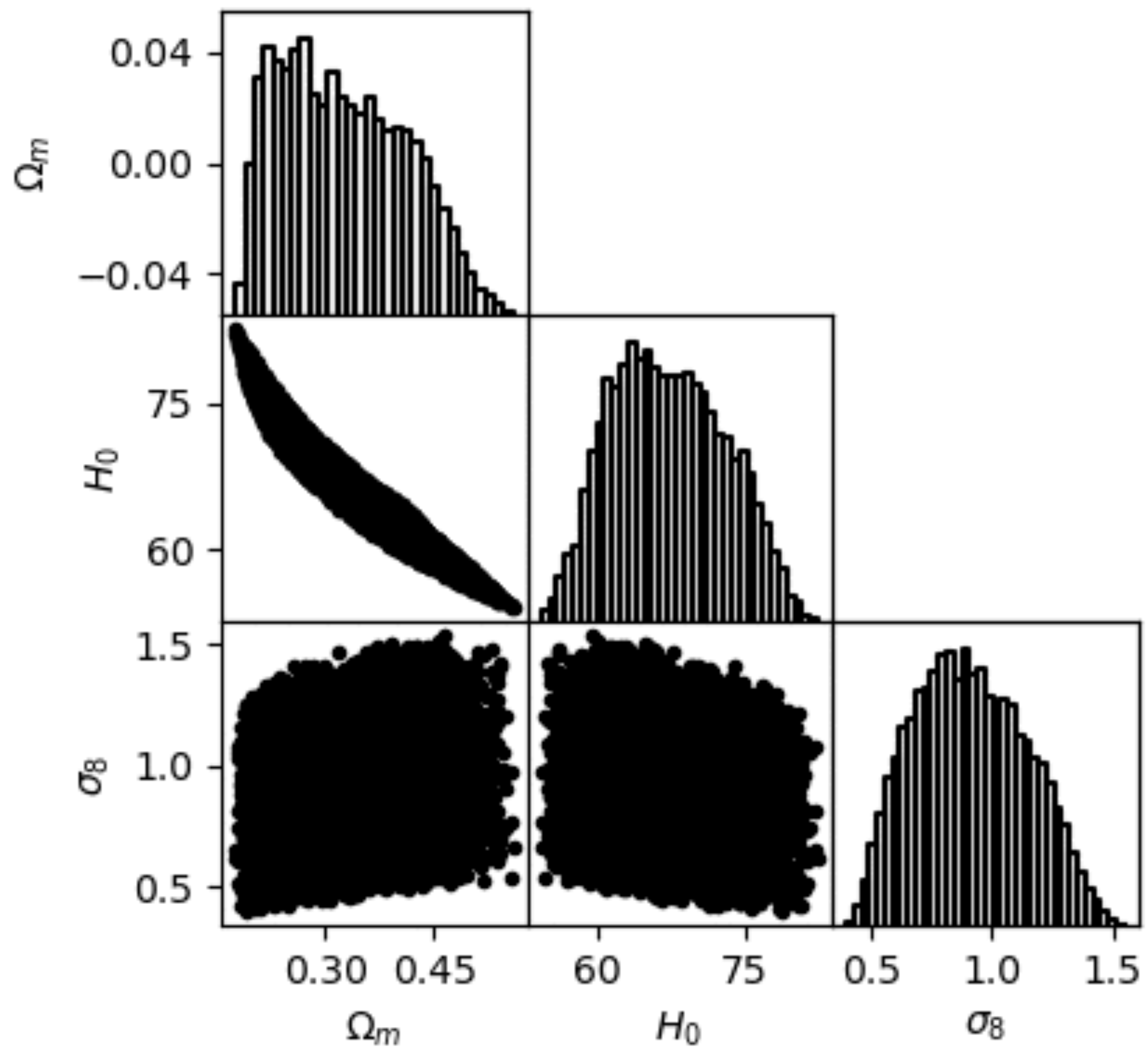


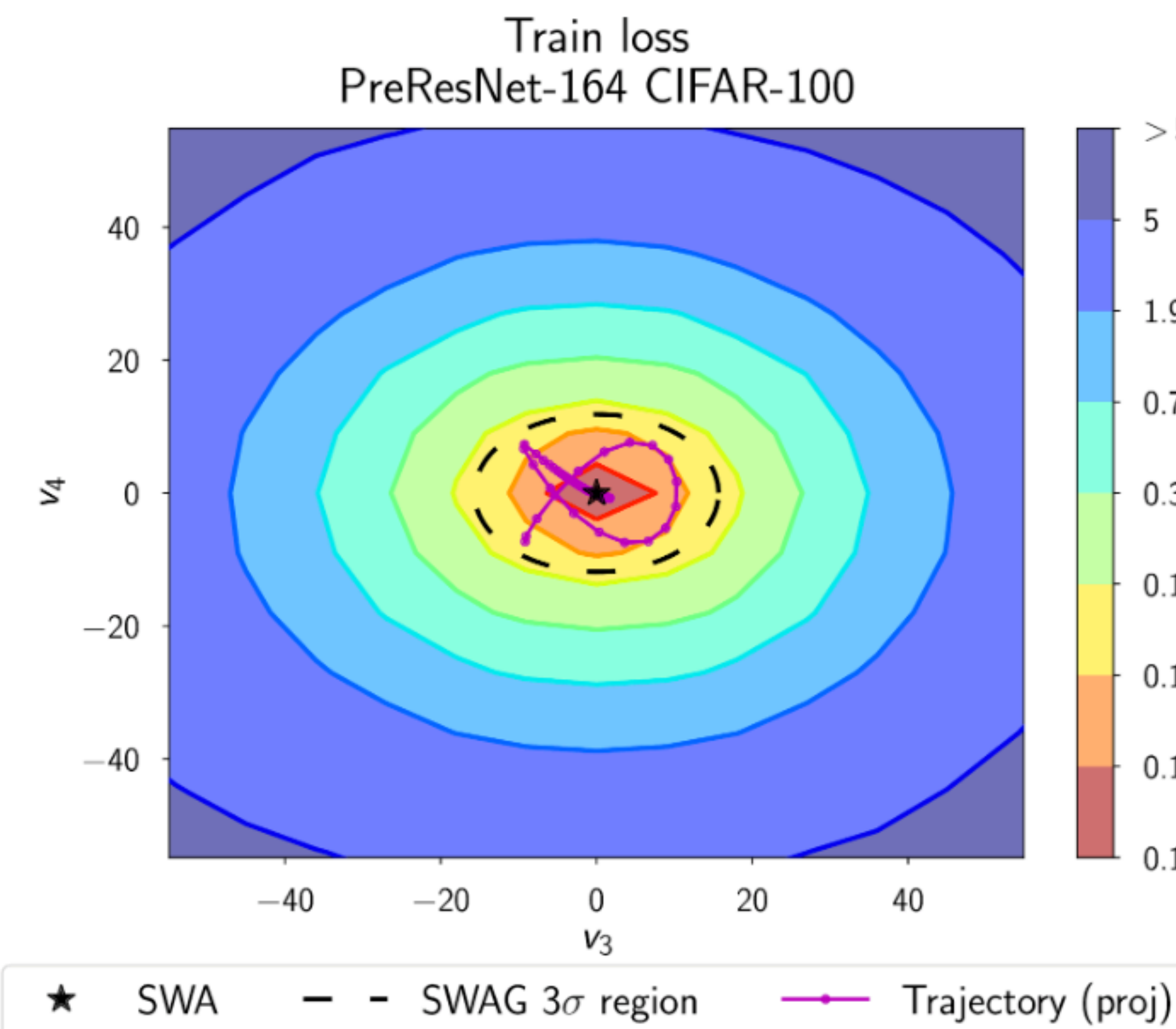
Traditional NN
Bayesian NN
$P(w)$
w

How to marginalize?

MCMC/
Nested Sampling

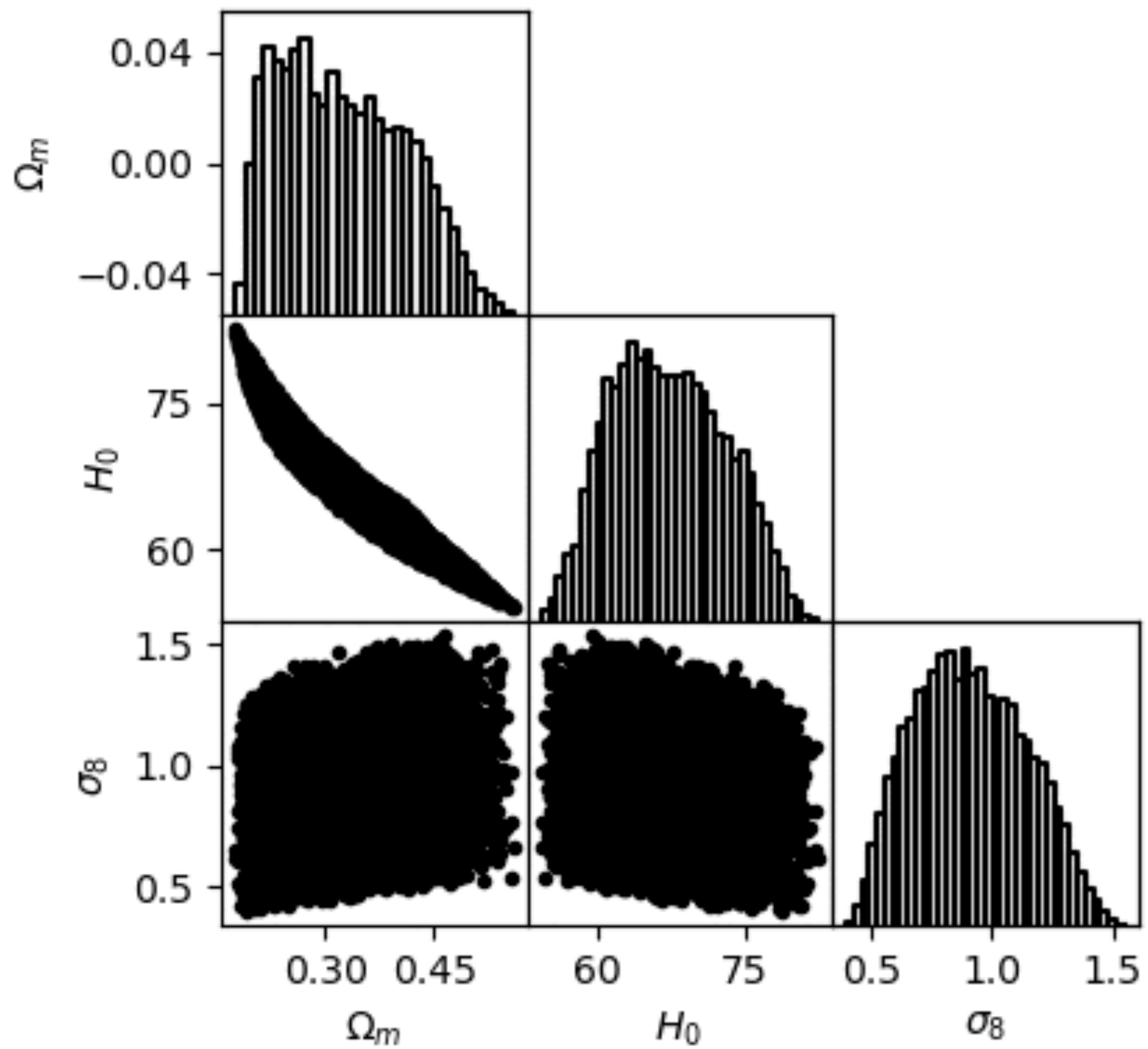
Stochastic
Weighting Average



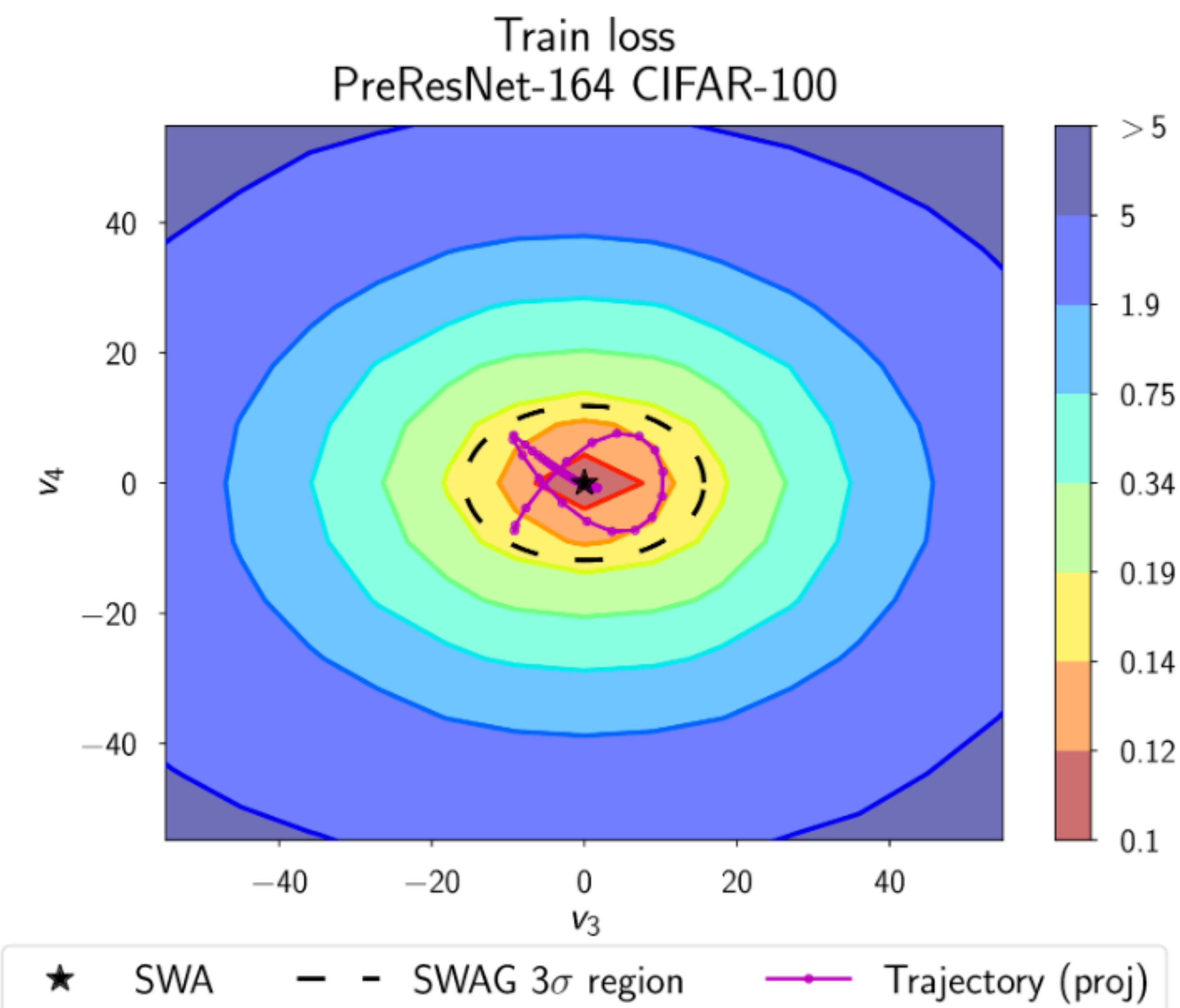


How to marginalize?

MCMC/
Nested Sampling

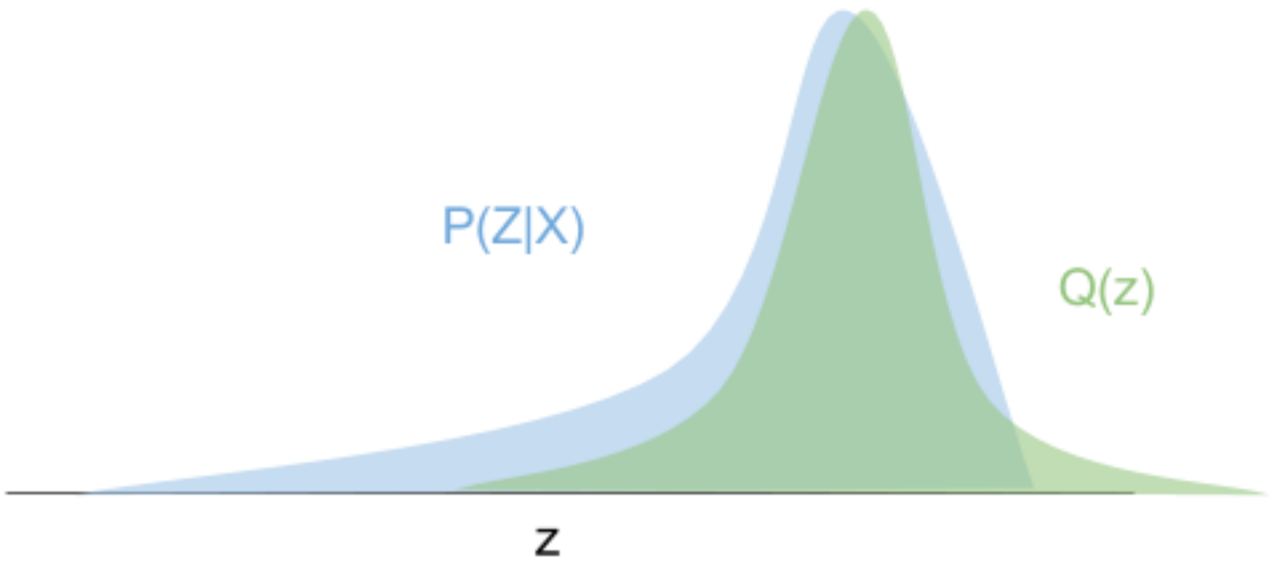


Stochastic
Weighting Average

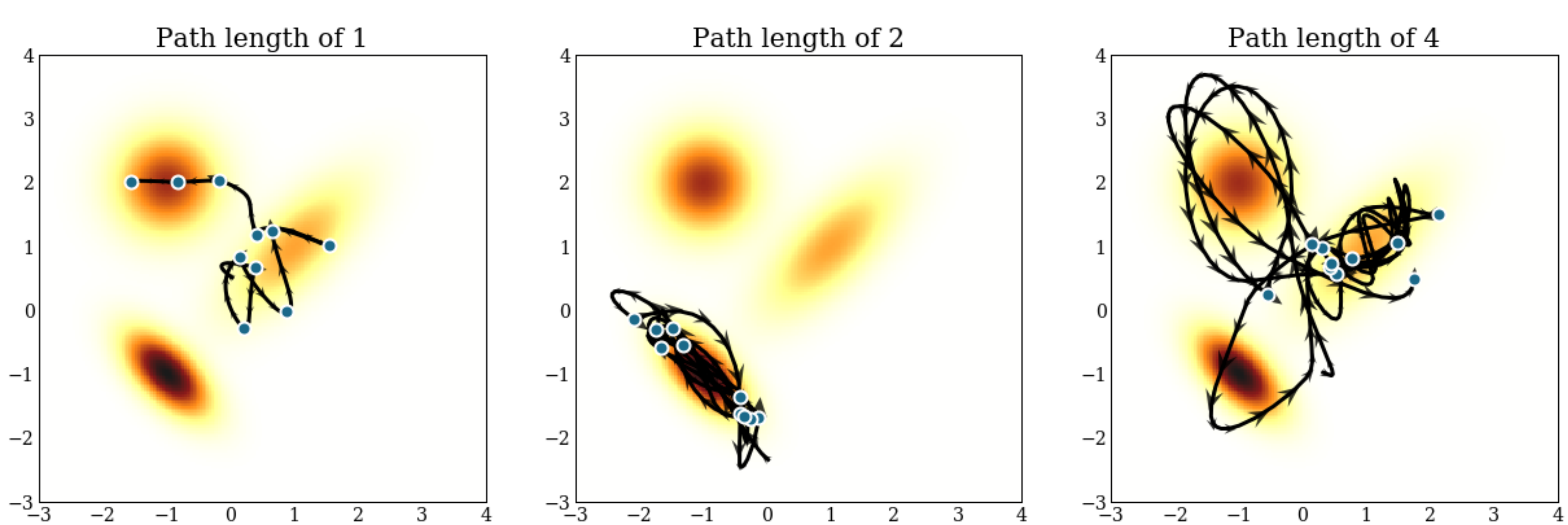


How to marginalize?

Variational Inference



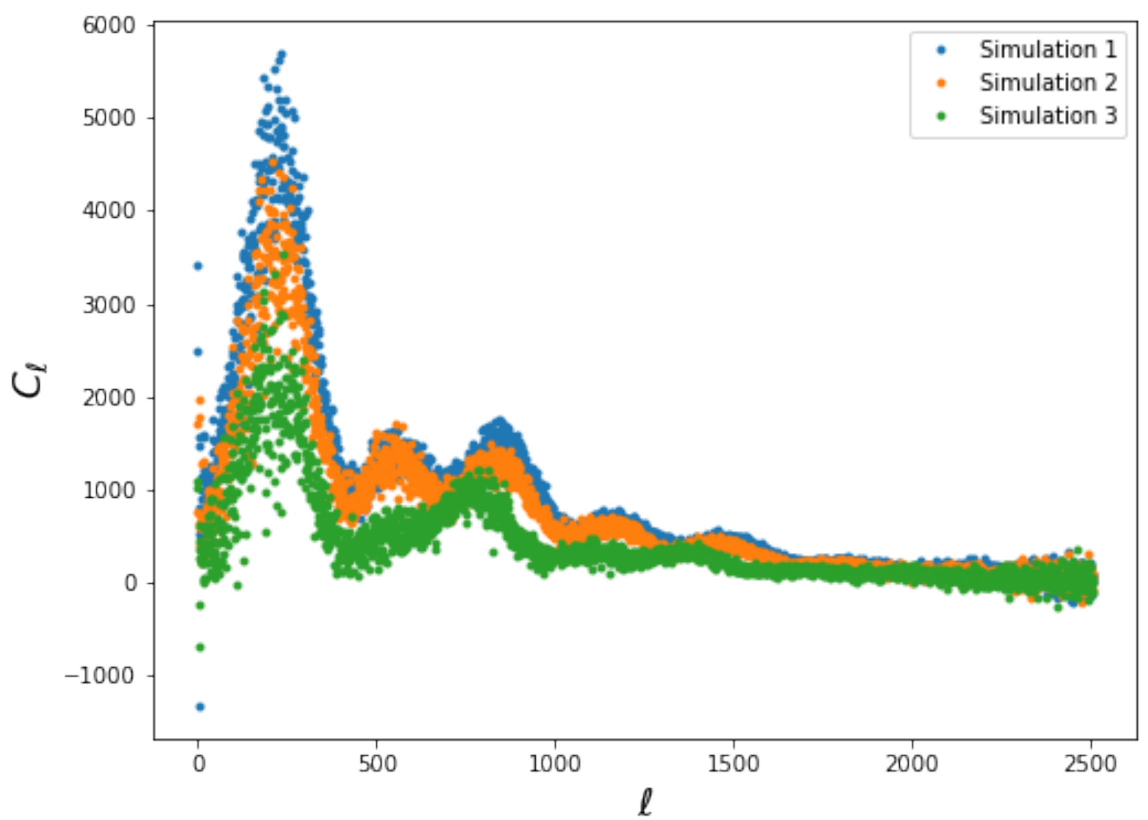
HMC

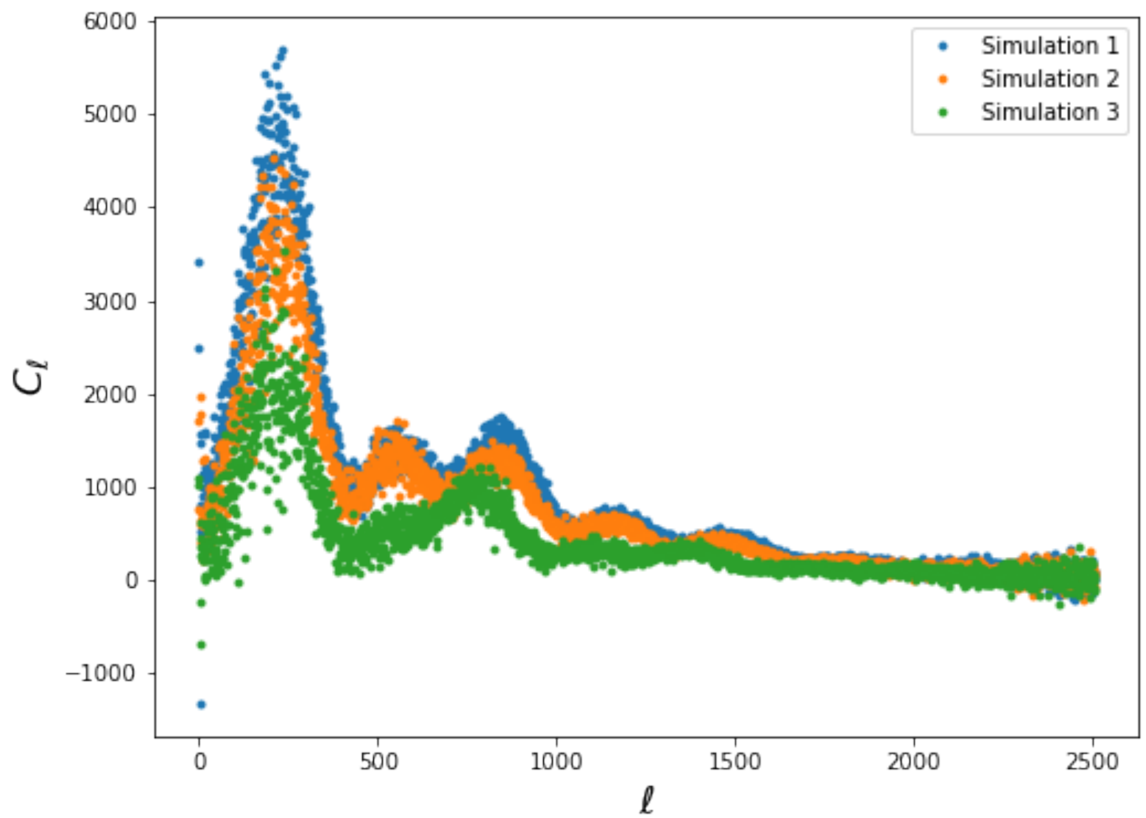


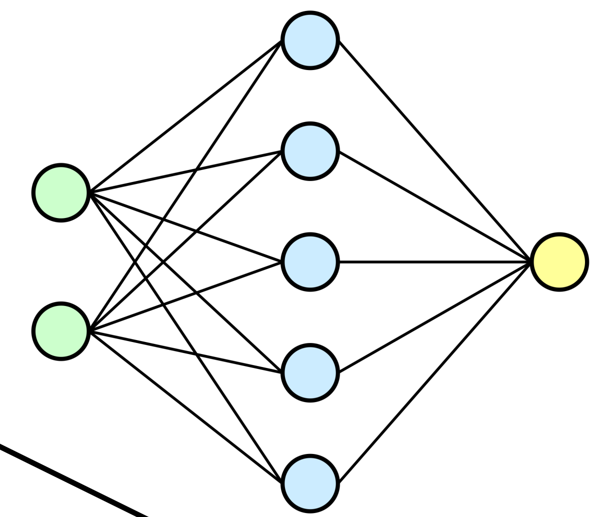
Example: CMB Cls

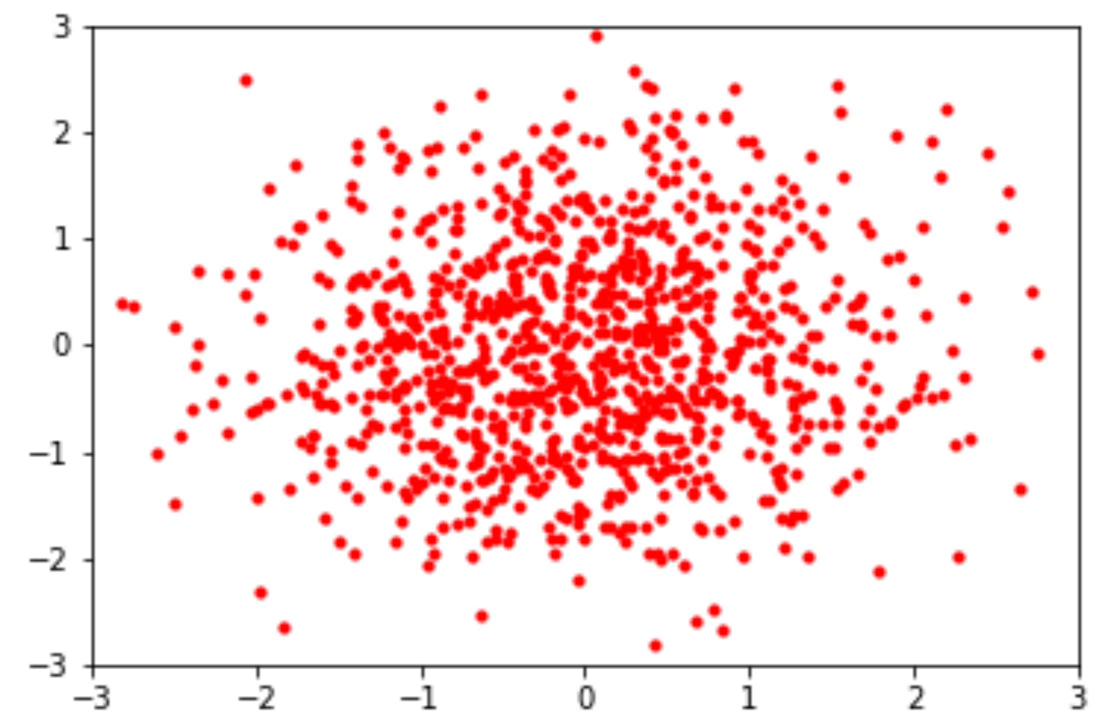
A simple example to test things:

- We want to estimate the ΛCDM parameters from CMB power spectra.
- We do not have access to CAMB/CLASS
- Instead, we have access to 10.000 simulations of spectra for different parameters
- All the simulations have Planck noise added

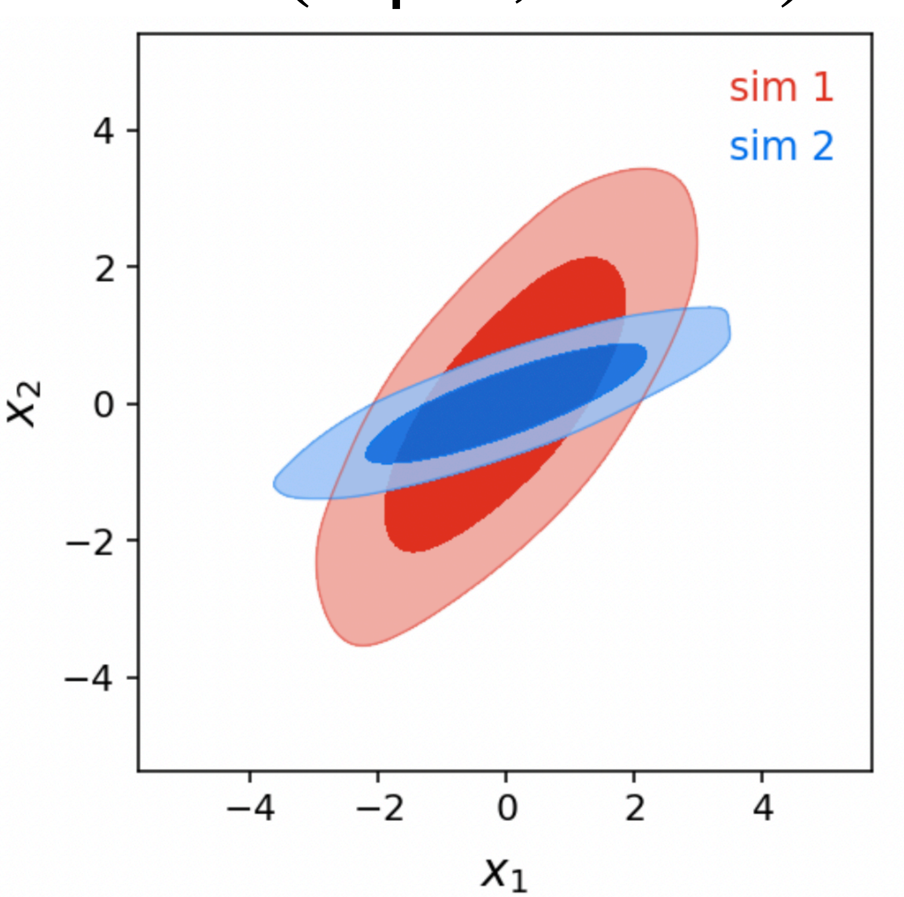


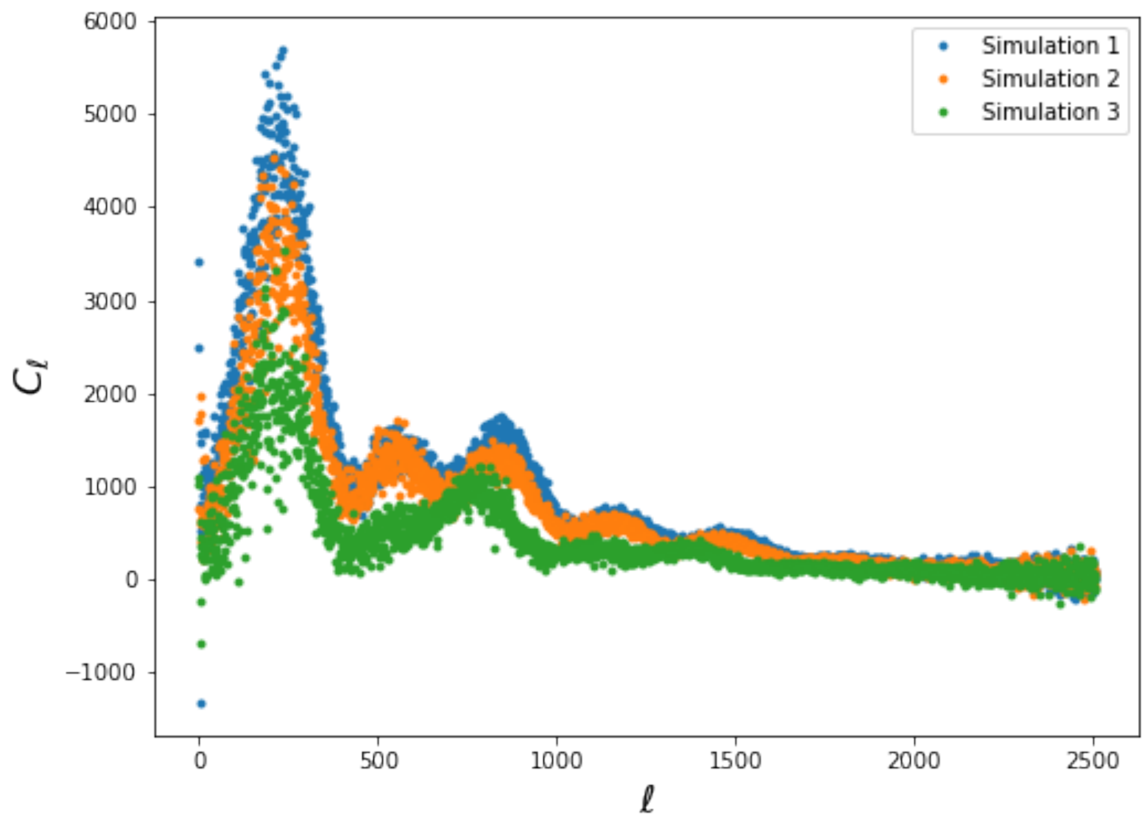


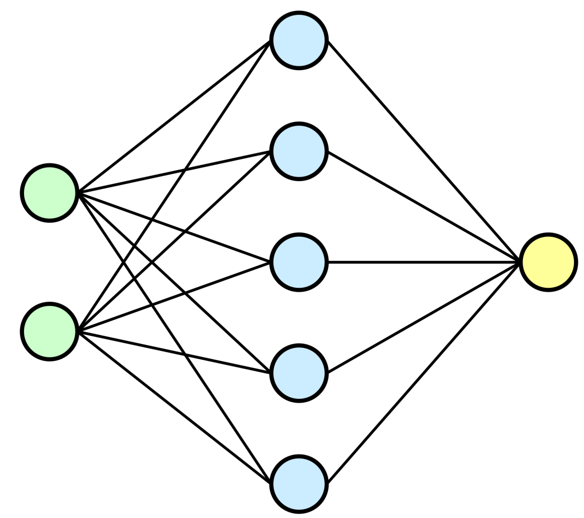




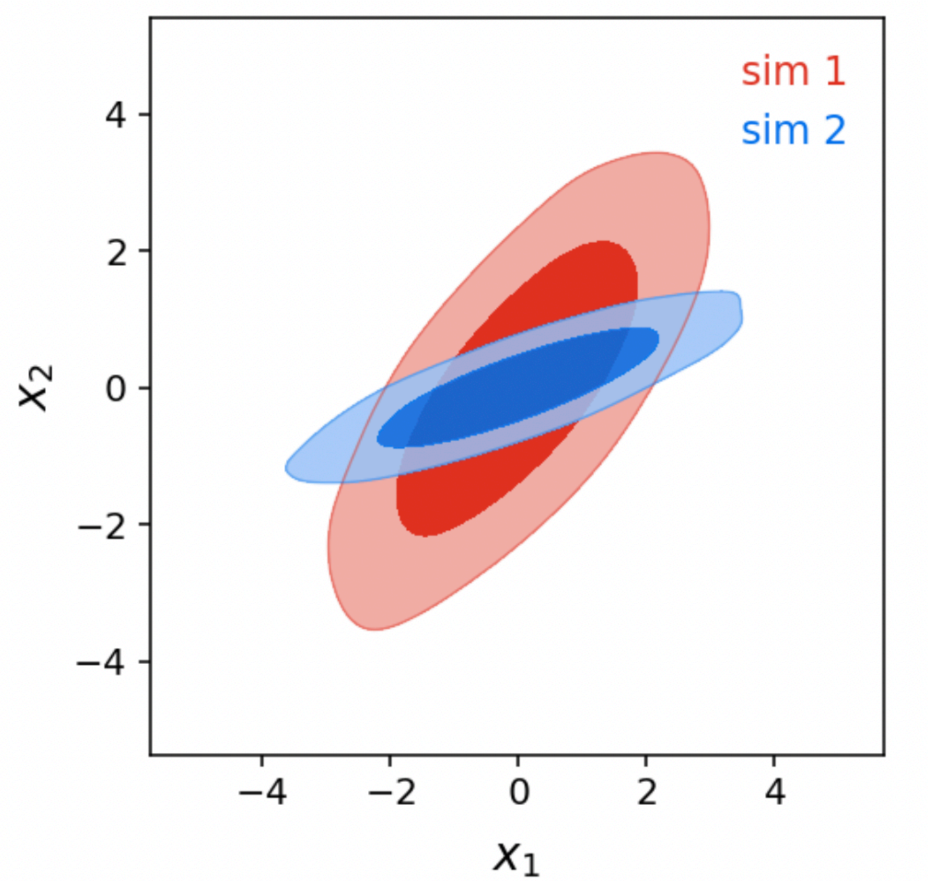
$$P(\theta \,|\, D, \mathrm{Sims})$$

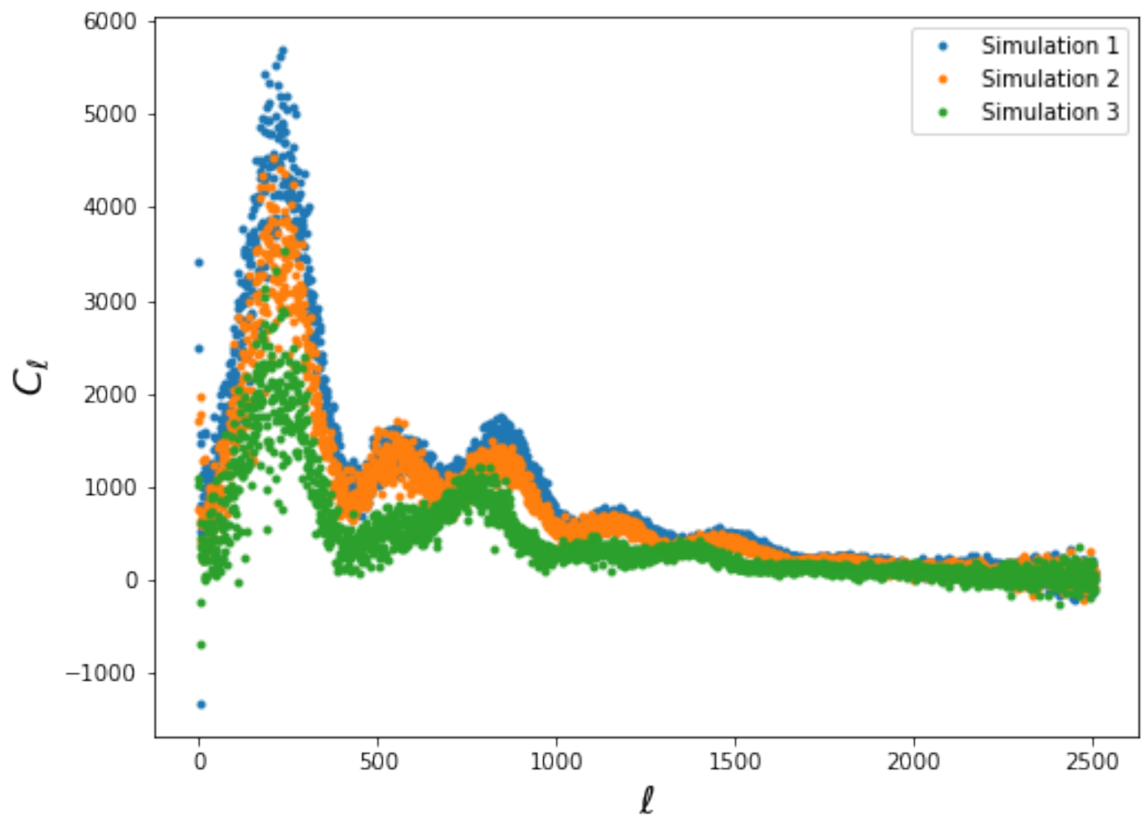


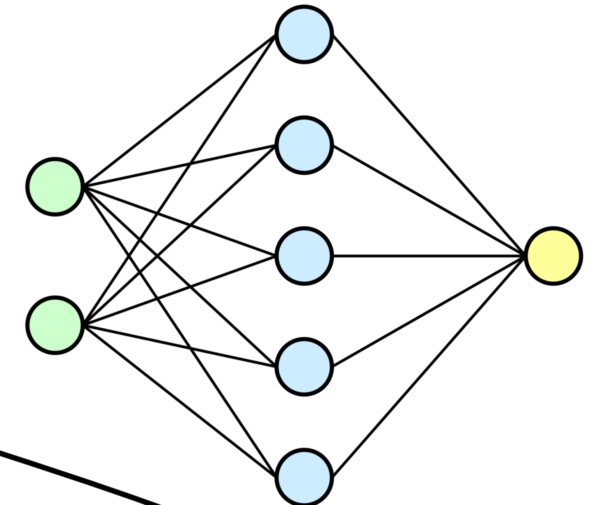




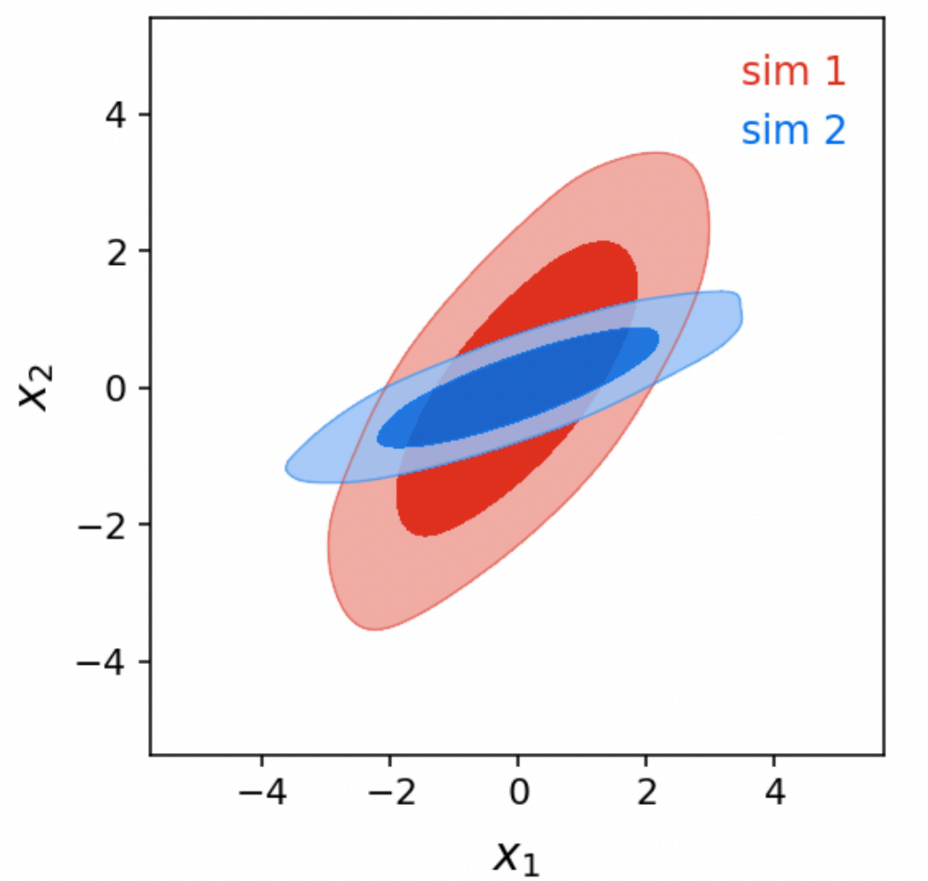
$$P(\theta \mid D, \mathrm{Sims})$$



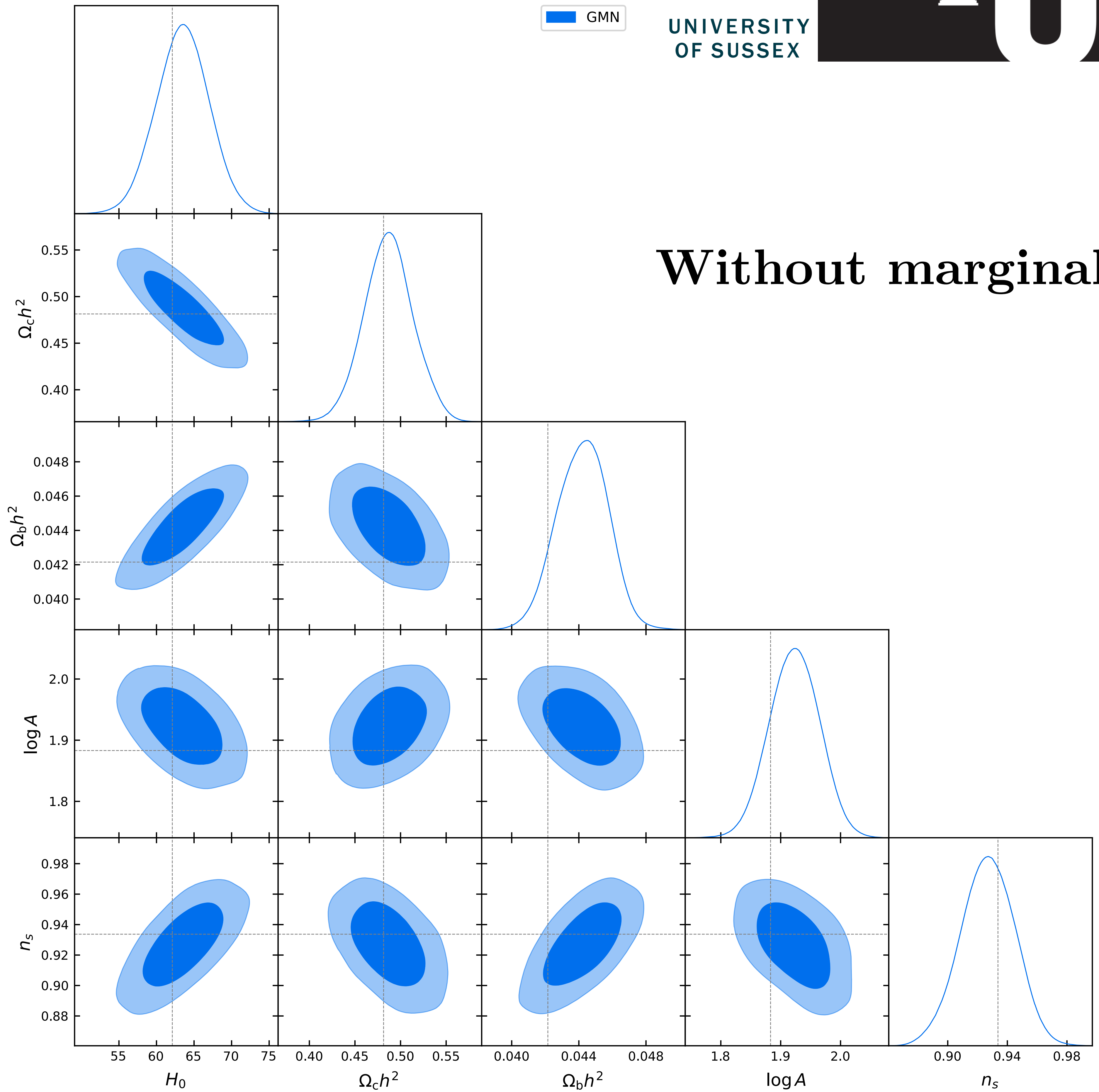


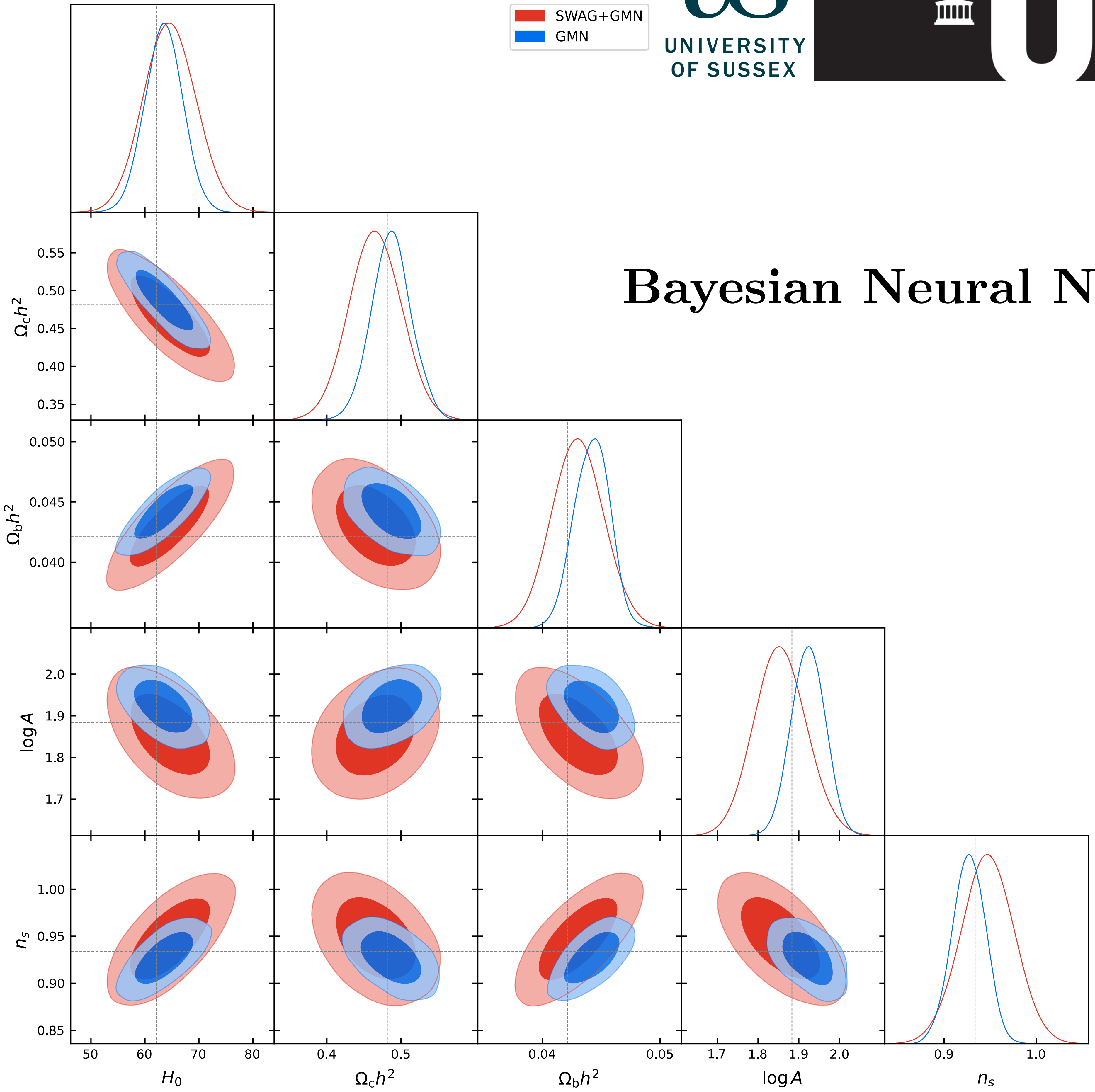


$$P(\theta \,|\, D, \mathrm{Sims})$$



Without marginalising





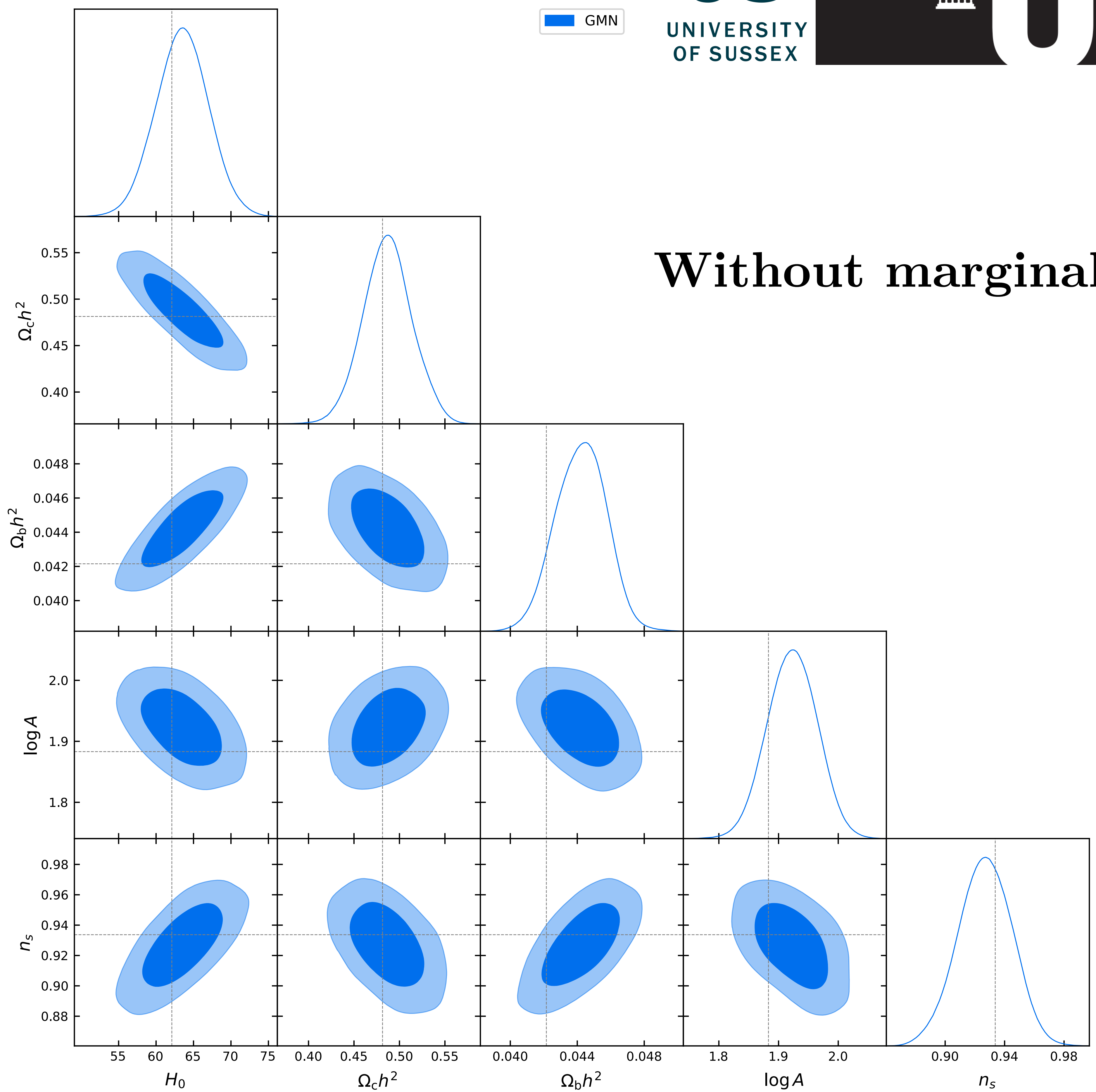
Bayesian Neural Network

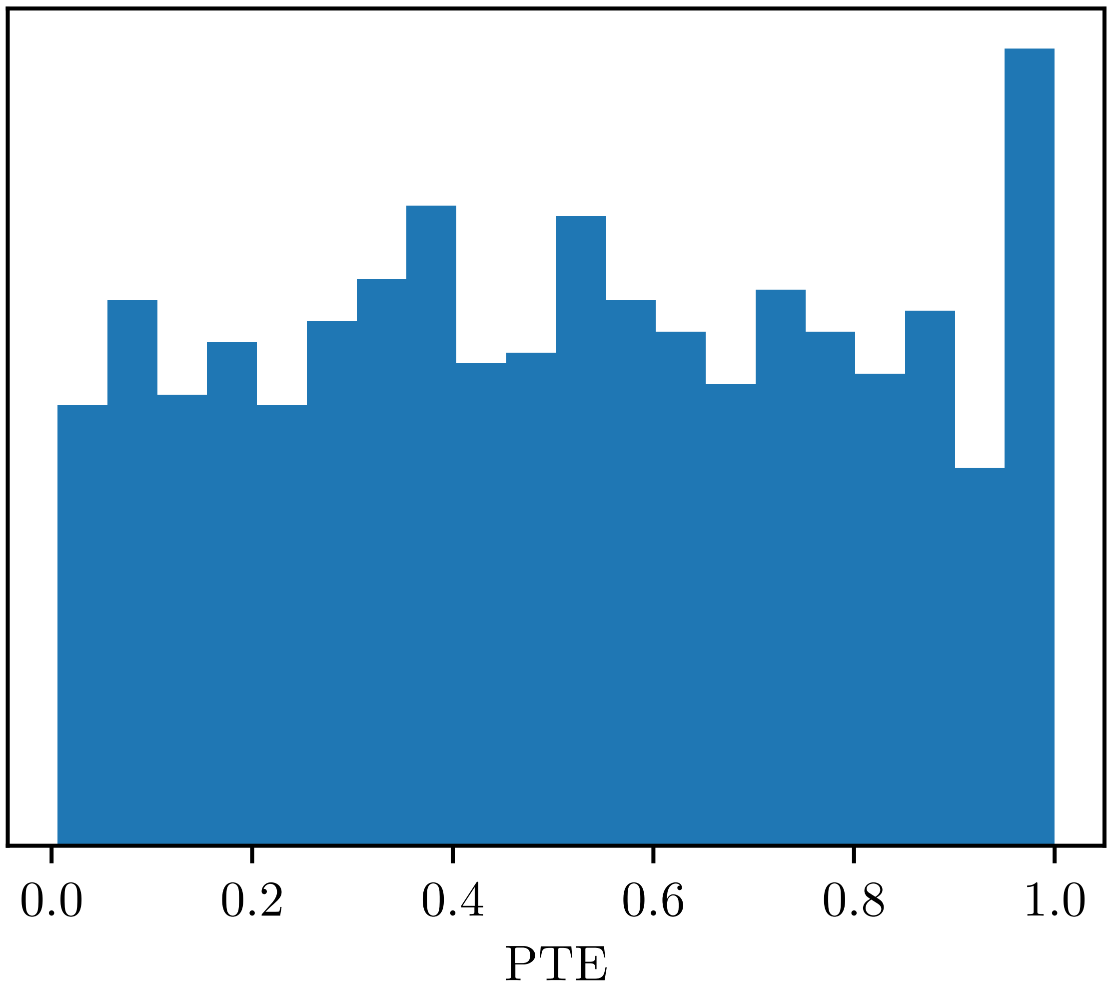
Bayesian Neural Networks

- More robustness
- Better generalisation properties
- Better control of overfitting

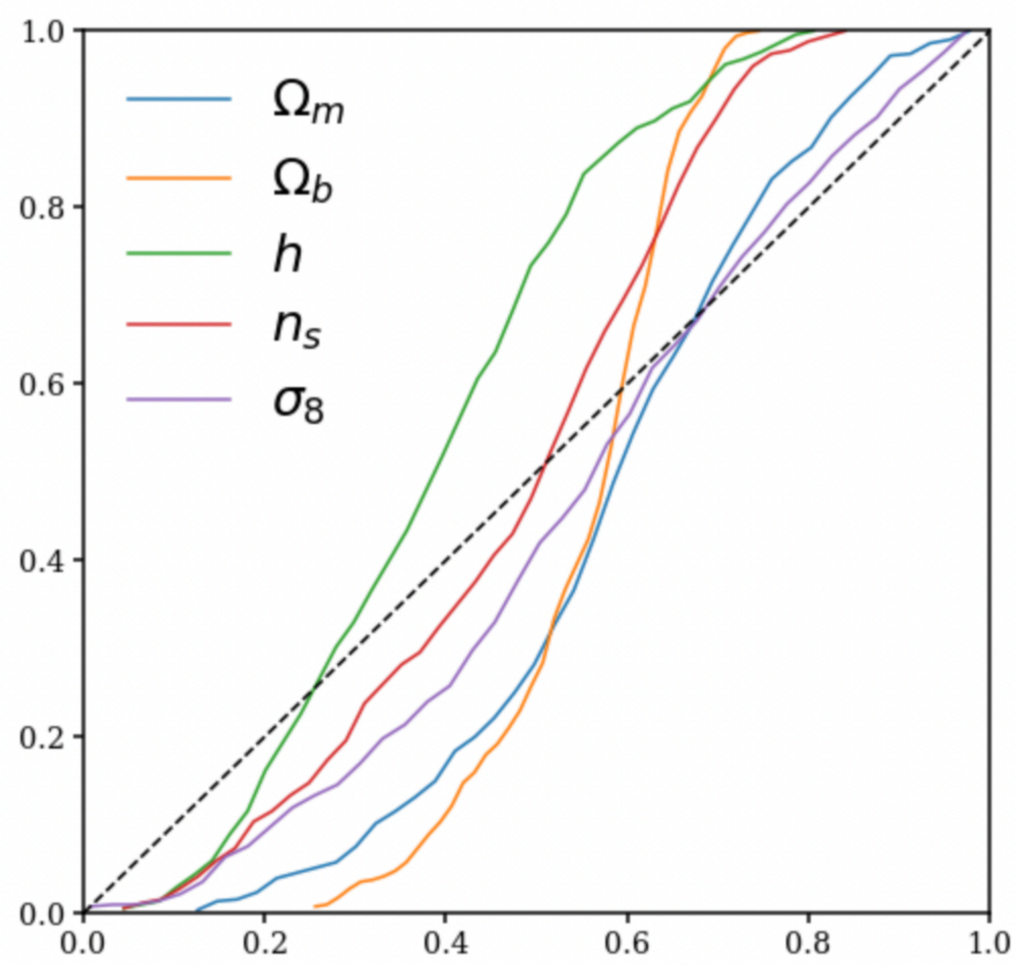
Bonus: Validation

Without marginalising

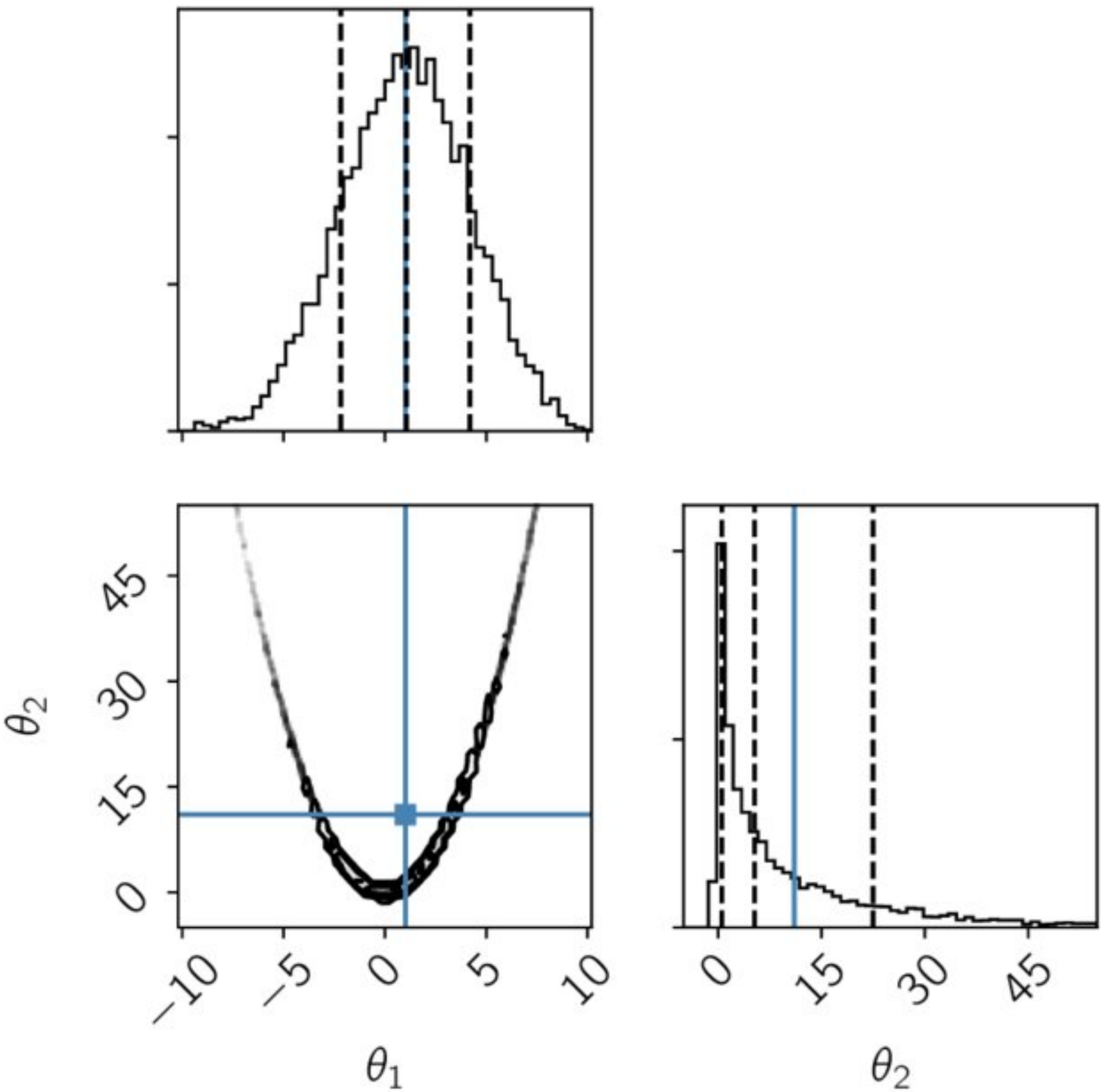




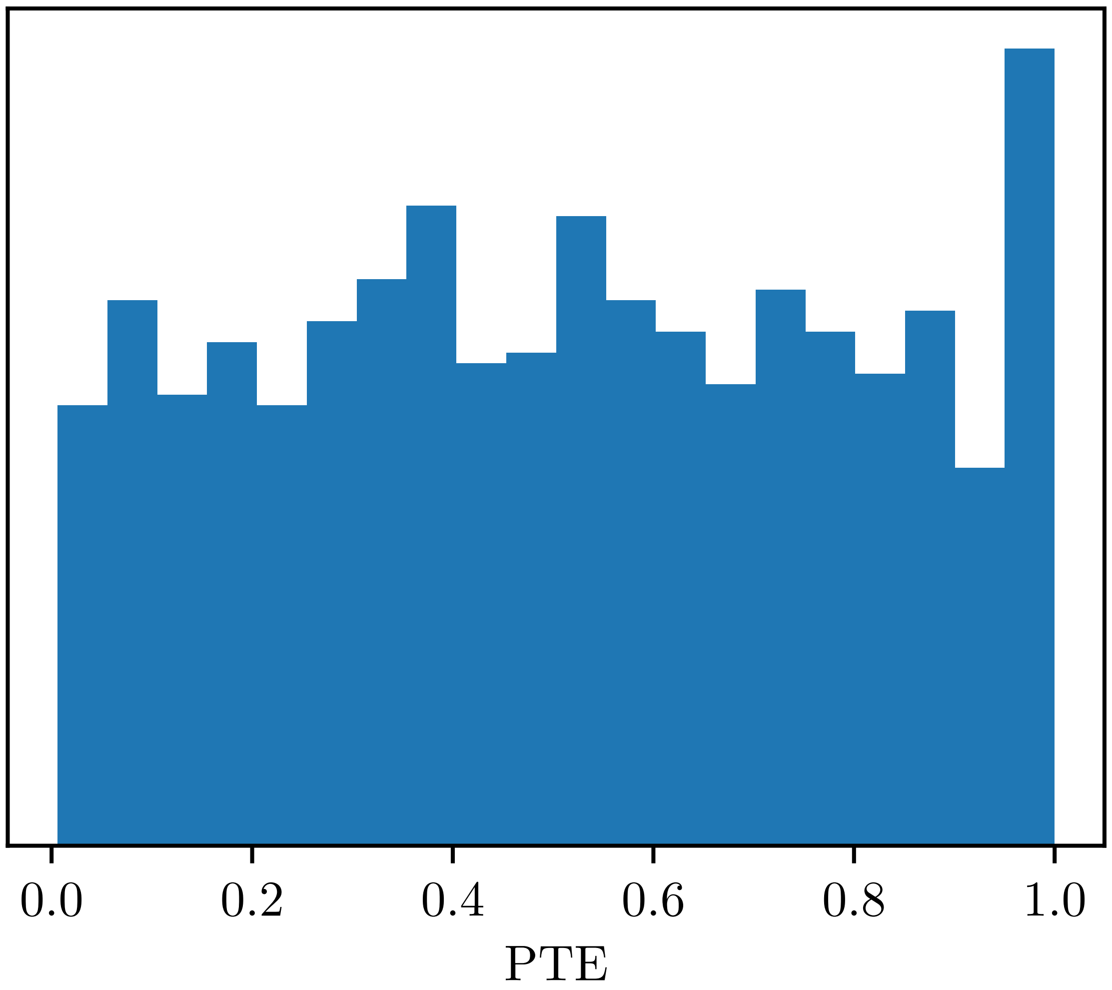
0.0
0.2
0.4
0.6
0.8
1.0
PTE

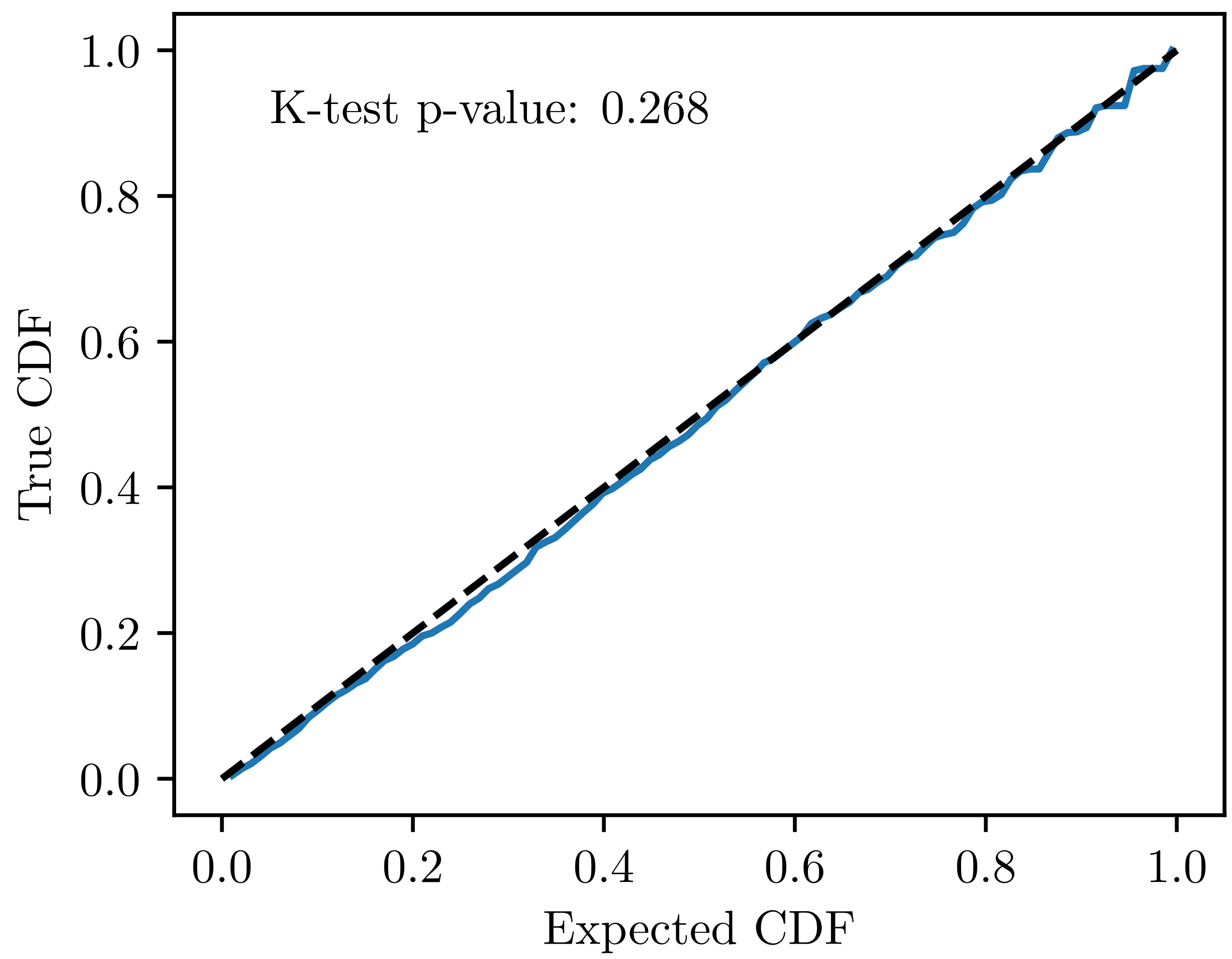


Ω_m
Ω_b
h
n_s
σ_8
0.0
0.2
0.4
0.6
0.8
1.0

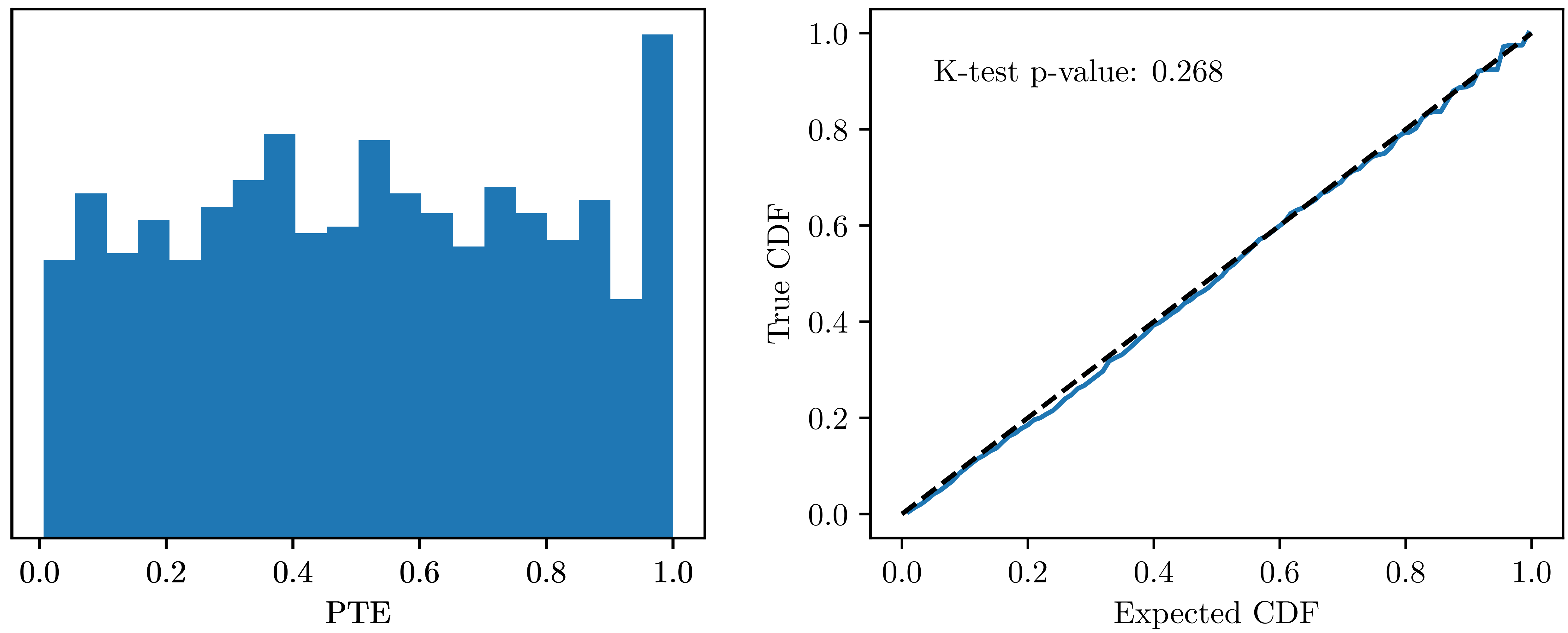


45
30
15
0
θ_2
−10
−5
0
5
10
θ_1
0
15
30
45
θ_2


0.0
0.2
0.4
0.6
0.8
1.0
PTE



0.0
0.2
0.4
0.6
0.8
1.0
PTE
K-test p-value: 0.268
True CDF
Expected CDF



In arXiv soon! (Unless someone has already done it!)

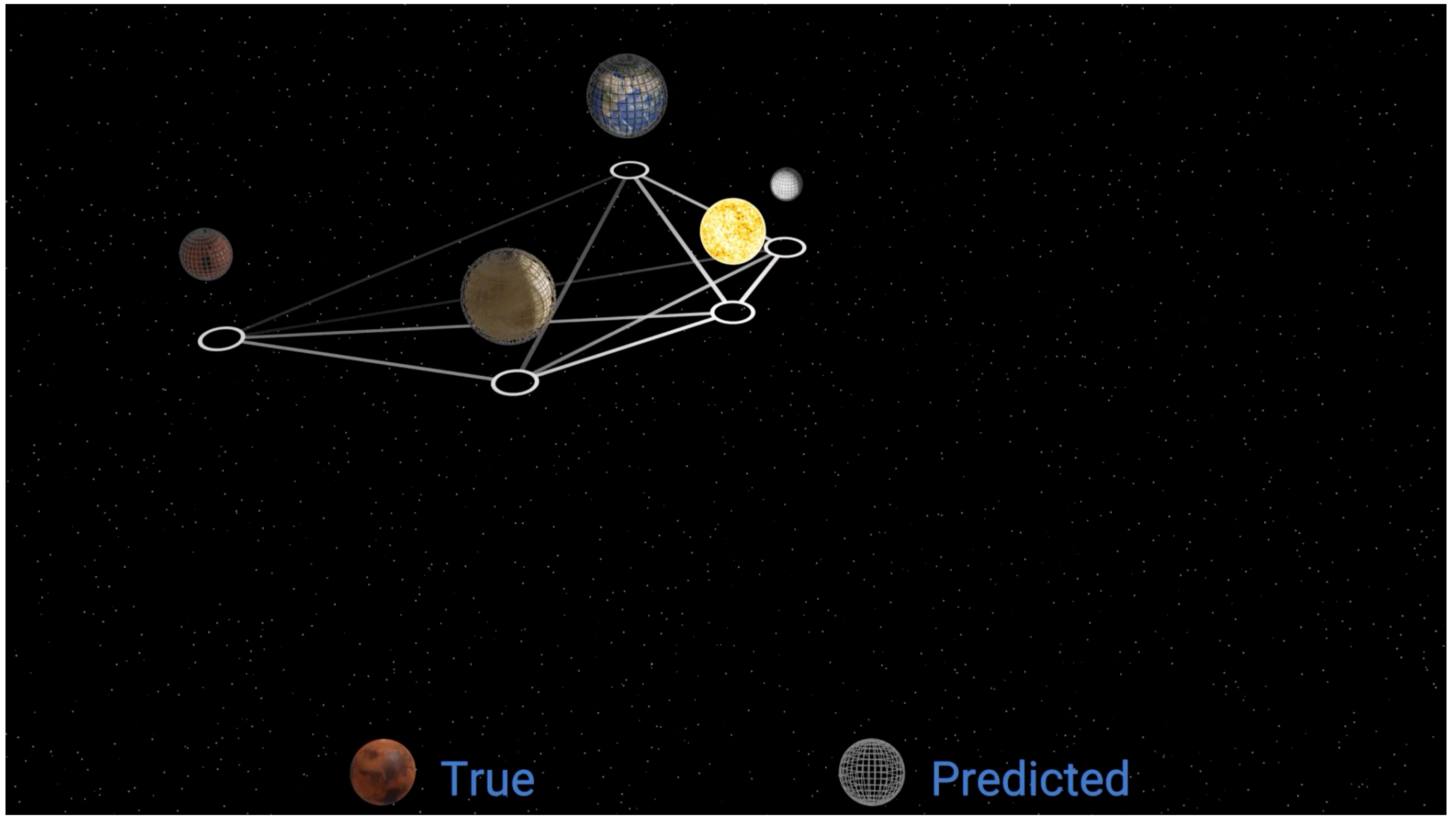
Other work

- Suspiciousness: Quantifying tension in cosmology
- The sum of the masses of MW and M31 with DELFI
- Cosmological analyses of Planck, DES and Simons Observatory.
- Automated scientist to re-discover Newtonian gravity from Solar System data

p.lemos@sussex.ac.uk

@PabloLemosP

https://pablo-lemos.github.io



Other work

- Suspiciousness: Quantifying tension in cosmology
- The sum of the masses of MW and M31 with DELFI
- Cosmological analyses of Planck, DES and Simons Observatory.
- Automated scientist to re-discover Newtonian gravity from Solar System data

p.lemos@sussex.ac.uk

@PabloLemosP

https://pablo-lemos.github.io