Consistency and tension

What is consistency?

Data

Cosmic shear measurements
- Kilo-Degree Survey (KiDS)

Model

ΛCDM

What is consistency?

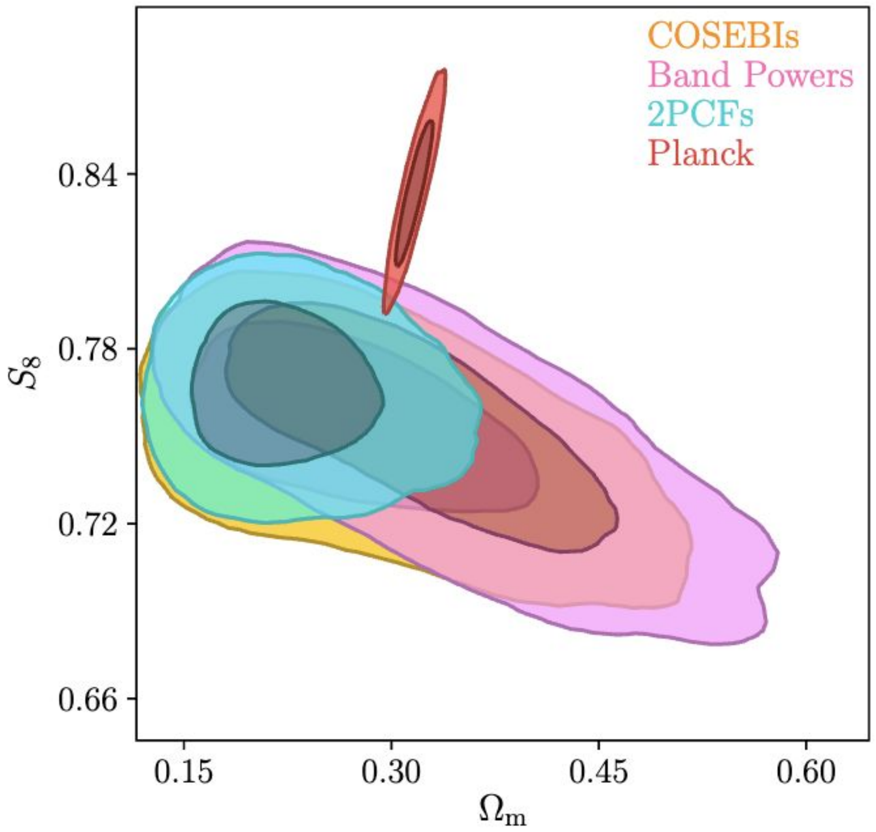
Data

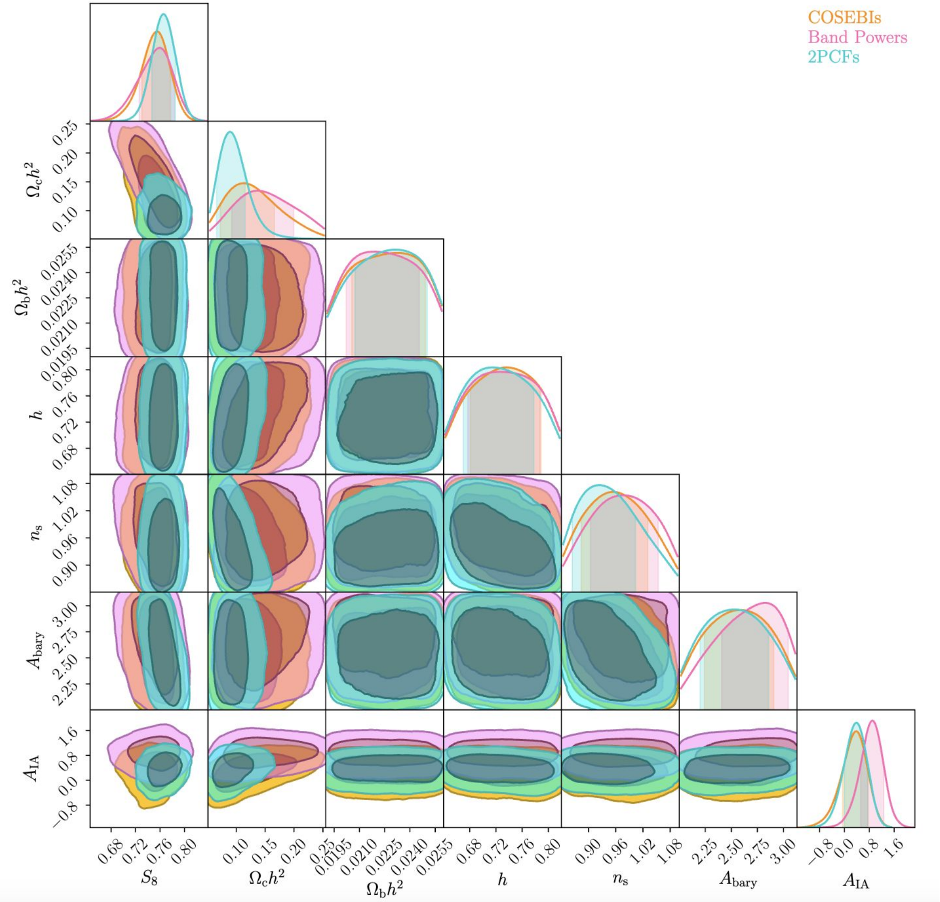
Cosmic shear measurements
- Kilo-Degree Survey (KiDS)

Model

ΛCDM

Question:
Assuming a ΛCDM cosmological model, what are the constraints on the cosmological parameters given the observed data?





Asgari et al. 2020, [arXiv:2105.09545](https://arxiv.org/abs/2105.09545)

What is consistency?

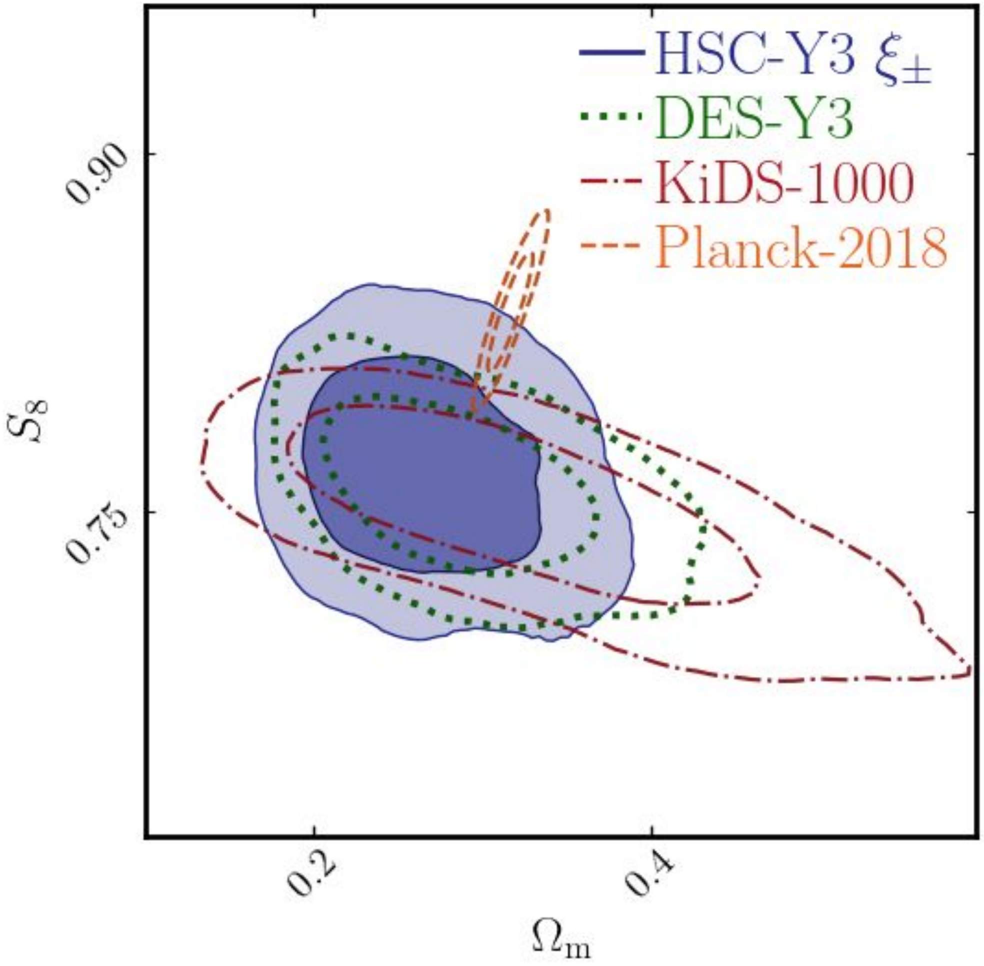
Data

Cosmic shear measurements
- Kilo-Degree Survey (KiDS)
- Dark Energy Survey (DES)
- Hyper Suprime Cam (HSC)

Model

ΛCDM

Question:
Assuming a ΛCDM cosmological model, are the constraints on the cosmological parameters from KiDS and DES consistent with each other?



Li et al. 2023, [arXiv:2304.00702](https://arxiv.org/abs/2304.00702)

What is consistency?

Data

Cosmic shear measurements
- Kilo-Degree Survey (KiDS)
- Dark Energy Survey (DES)
- Hyper Suprime Cam (HSC)

Cosmic Microwave Background
- Planck

any other experiment

Model

ΛCDM

Question:
Assuming a specific model, are the constraints on the model parameters from dataset A and dataset B consistent with each other?

⇒ **Consistency between datasets**

What is consistency?

Data

Cosmic shear measurements
- Kilo-Degree Survey (KiDS)
- Dark Energy Survey (DES)
- Hyper Suprime Cam (HSC)

Cosmic Microwave Background
- Planck

any other experiment

Model

ΛCDM
wCDM

Question:
Does the data from a given experiment prefer wCDM over ΛCDM?

⇒ **Model comparison**

What is consistency?

Data

Cosmic shear measurements
- Kilo-Degree Survey (KiDS)
- Dark Energy Survey (DES)
- Hyper Suprime Cam (HSC)

Cosmic Microwave Background
- Planck

any other experiment

Model

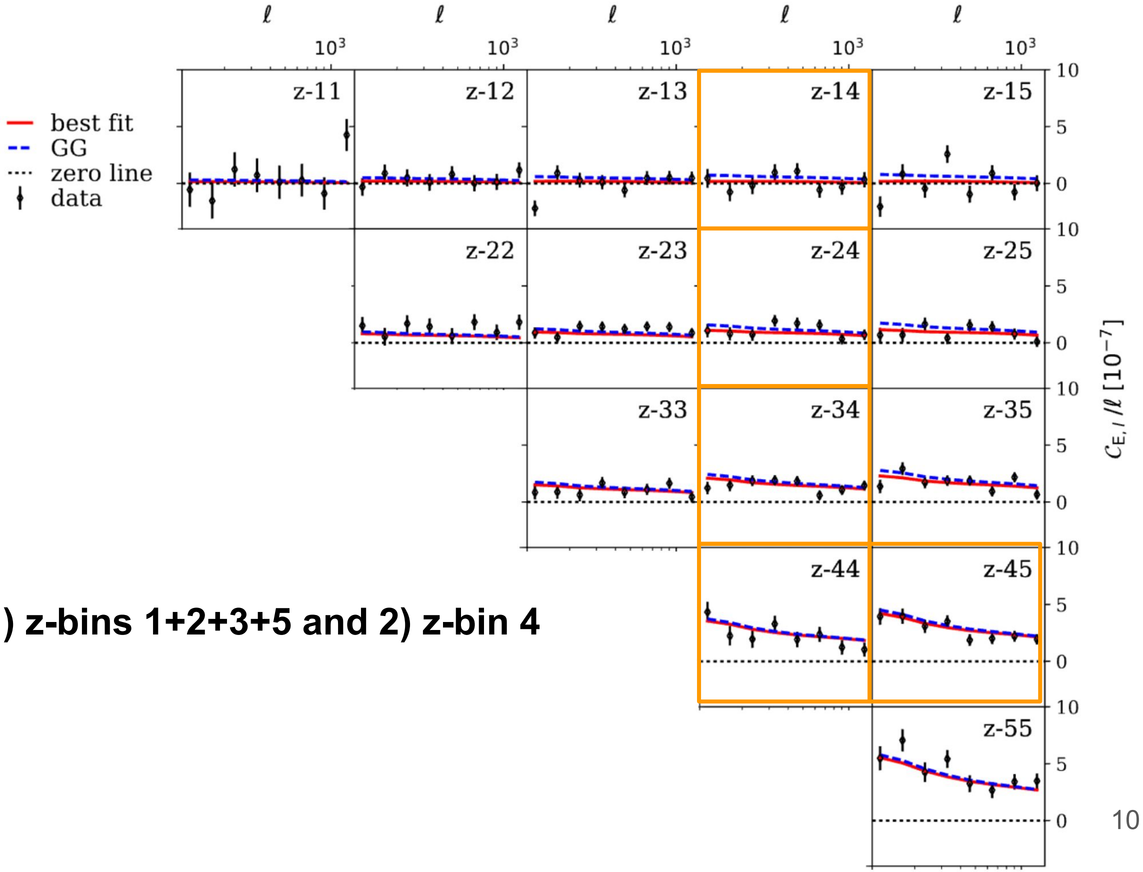
ΛCDM
wCDM
any extensions of ΛCDM you can think of

Consistency between datasets

Model comparison

Internal consistency of a given dataset

Question: If I split my dataset into two parts, are the constraints on the model parameters coming from both parts consistent with each other?



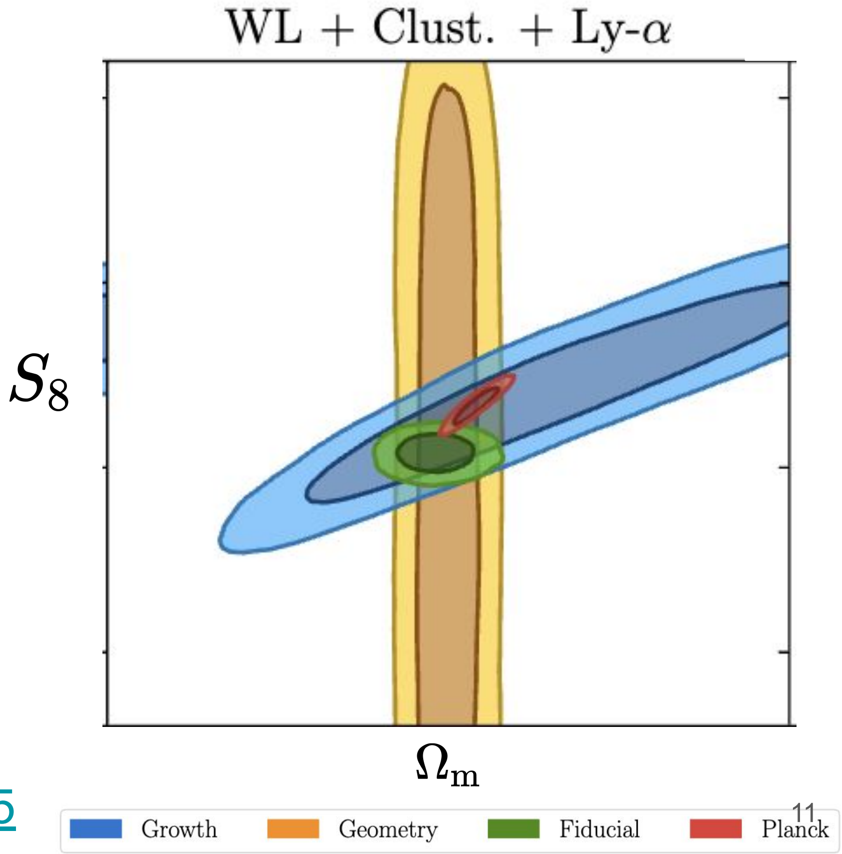
full datavector vs individual fits of 1) z-bins 1+2+3+5 and 2) z-bin 4

Internal consistency of a given model

Divide ΛCDM model into two regimes:

1. homogeneous background (geometry)
2. evolution of matter density fluctuations (growth)

Question: Are the constraints from the data on the two theory regimes consistent?



Ruiz-Zapatero et al. 2021, [arXiv: 2105.09545]

Tension metrics:

1. Single summary statistics of the full likelihood or posterior
   a. Evidence
   b. Bayes factor
   c. Suspiciousness
2. Parameter space methods
   a. Differences of single or multiple model parameters
3. Data vector methods
   a. Difference in data vector space

Evidence (marginal likelihood):

posterior likelihood prior

Bayes' theorem: $P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} \Rightarrow \mathcal{P}_D(\theta) = \frac{\mathcal{L}_D(\theta)\pi(\theta)}{\mathcal{Z}}$

evidence

Evidence (marginal likelihood):

posterior

likelihood

prior

Bayes' theorem: $P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} \Rightarrow \mathcal{P}_D(\theta) = \frac{\mathcal{L}_D(\theta)\pi(\theta)}{\mathcal{Z}}$

evidence

Evidence: $\mathcal{Z} = \int \mathrm{d}\theta\, \mathcal{L}_D(\theta)\pi(\theta)$ ⇒ Likelihood, integrated over the prior probability of the parameters

- "Probability of the data given a model"
- ⇒ model evidence
- Useful for model comparison

Bayes factor

Model selection: Given some data D and two models M_1 and M_2 the Bayes factor is defined as: $R = \frac{\mathcal{Z}_1}{\mathcal{Z}_2} = \frac{P(D|M_1)}{P(D|M_2)} = \frac{P(M_1|D)P(M_2)}{P(M_2|D)P(M_1)}$

We usually use the logarithmic form: $\log R = \log \mathcal{Z}_1 - \log \mathcal{Z}_2$
and use **Jeffreys' scale** to interpret the values

Bayes factor

Model selection: Given some data D and two models M_1 and M_2 the Bayes factor is defined as: $R = \frac{\mathcal{Z}_1}{\mathcal{Z}_2} = \frac{P(D|M_1)}{P(D|M_2)} = \frac{P(M_1|D)P(M_2)}{P(M_2|D)P(M_1)}$

We usually use the logarithmic form: $\log R = \log \mathcal{Z}_1 - \log \mathcal{Z}_2$
and use **Jeffreys' scale** to interpret the values

Consistency test between two (uncorrelated) datasets:

$$\log R = \log \mathcal{Z}_{AB} - \log \mathcal{Z}_A - \log \mathcal{Z}_B$$

with:

$\mathcal{Z}_{AB}$: evidence of the combined sampling using datasets A and B with a single parameter set

$\mathcal{Z}_A, \mathcal{Z}_B$: evidence of the individual sampling with data A/B

Jeffreys' scale

$$\log R = \log \mathcal{Z}_1 - \log \mathcal{Z}_2$$

log R	**strength of evidence**
<0	negative (supports model 2)
0 - 0.5	barely worth mentioning
0.5 - 1	substantial
1 - 1.5	strong
1.5 - 2	very strong
>2	decisive

Nested sampling

[John Skilling, 2004]

> “The evidence Z is often the single most important number in the [Bayesian] problem and I think every effort should be devoted to calculating it” (MacKay 2003). Nested sampling does this by giving a direct estimate of the density of states. Posterior samples are an optional by-product.

Incomplete list of implementations:

- MultiNest
- PolyChord
- dynesty
- UltraNest
- Nautilus
- …

Nested sampling

$$\mathcal{Z} = \int \mathrm{d}\theta \, \mathcal{L}_D(\theta)\pi(\theta)$$

- high-dimensional integral
- only a small region in parameter space contributes significantly to the integral
- ⇒ need to find this region to calculate the evidence

Nested sampling

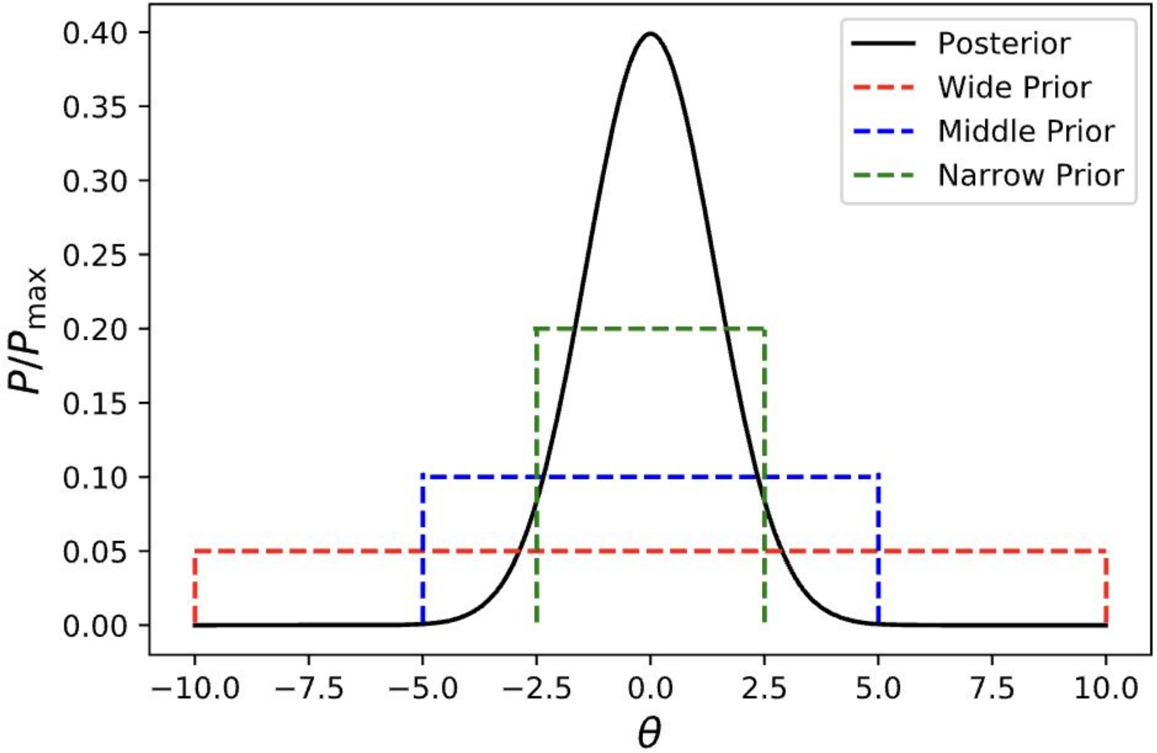
$$\mathcal{Z} = \int \mathrm{d}\theta \, \mathcal{L}_D(\theta)\pi(\theta)$$

- high-dimensional integral
- only a small region in parameter space contributes significantly to the integral
- ⇒ need to find this region to calculate the evidence

Traditional MCMC:

- usually focuses on generating samples around the peak of the posterior
- doesn’t generate a lot of samples in the tails of the posterior distribution

Prior dependence of the Bayes factor



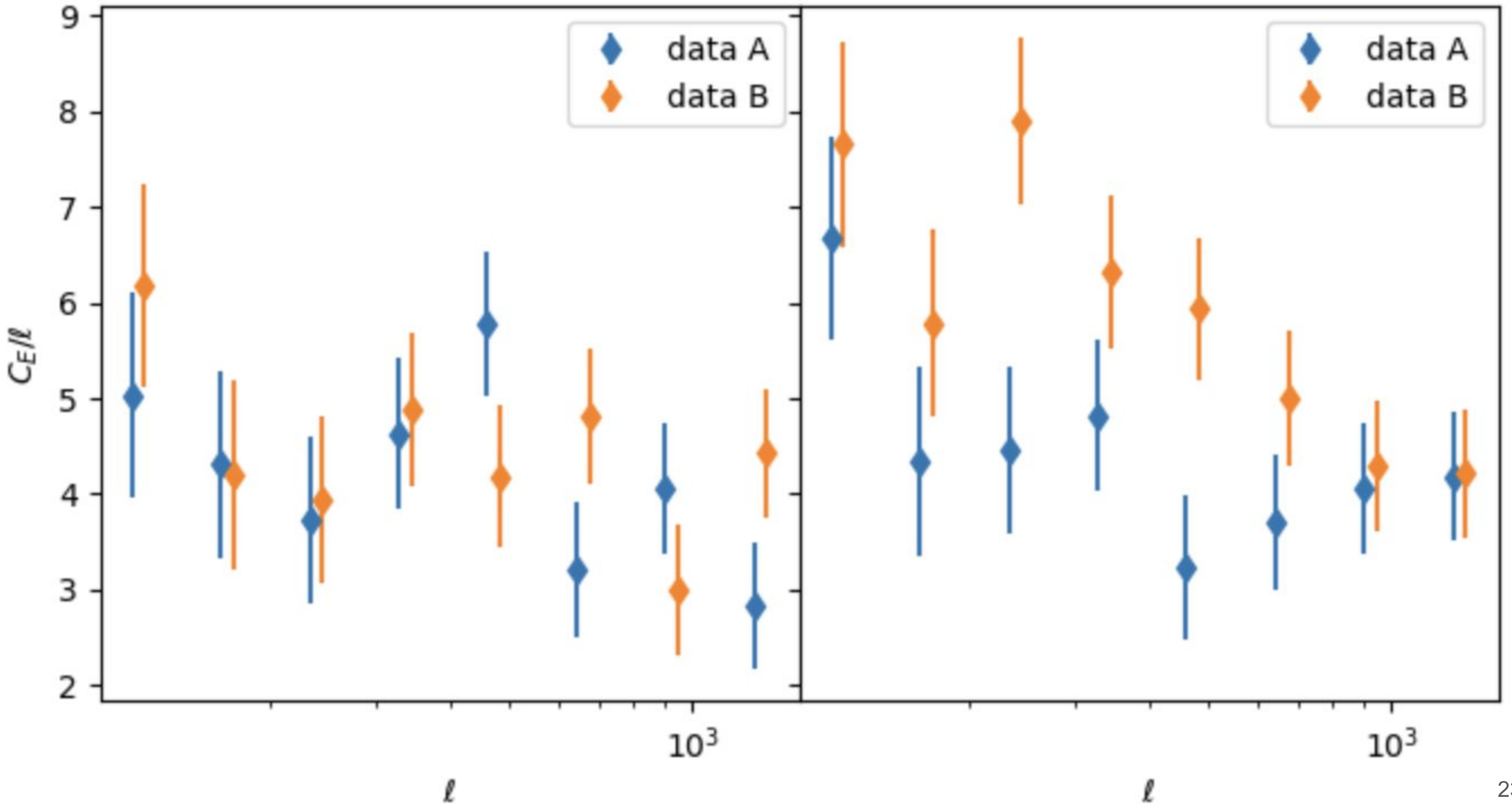
Lemos et al. 2020

https://github.com/BStoelzner/PreciseStatisticalAnalysis/

Install pymultinest:

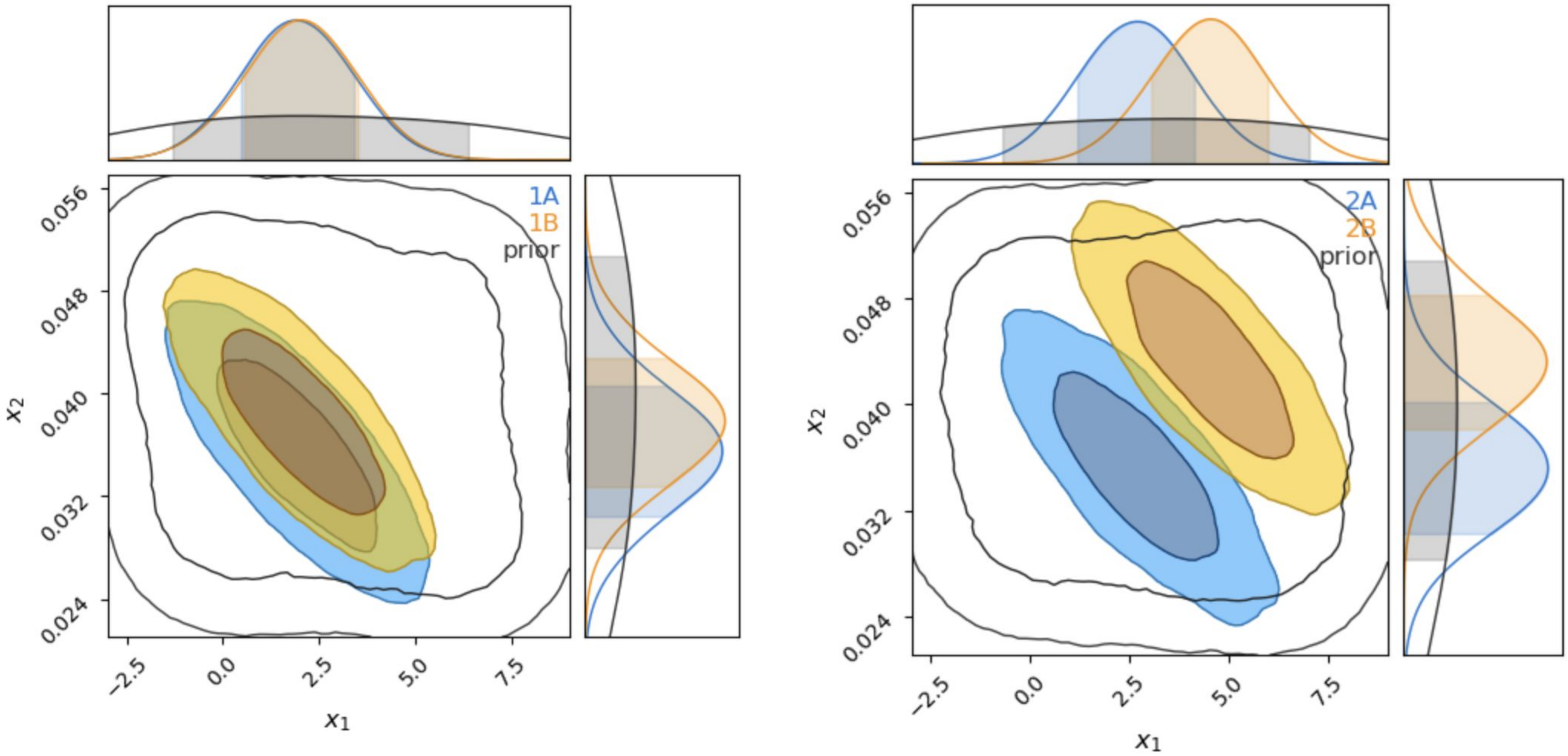
conda install -c conda-forge pymultinest

(unless you have an M1/M2 Mac)


data A
data B
C_E/ℓ
ℓ
10^3
9
8
7
6
5
4
3
2

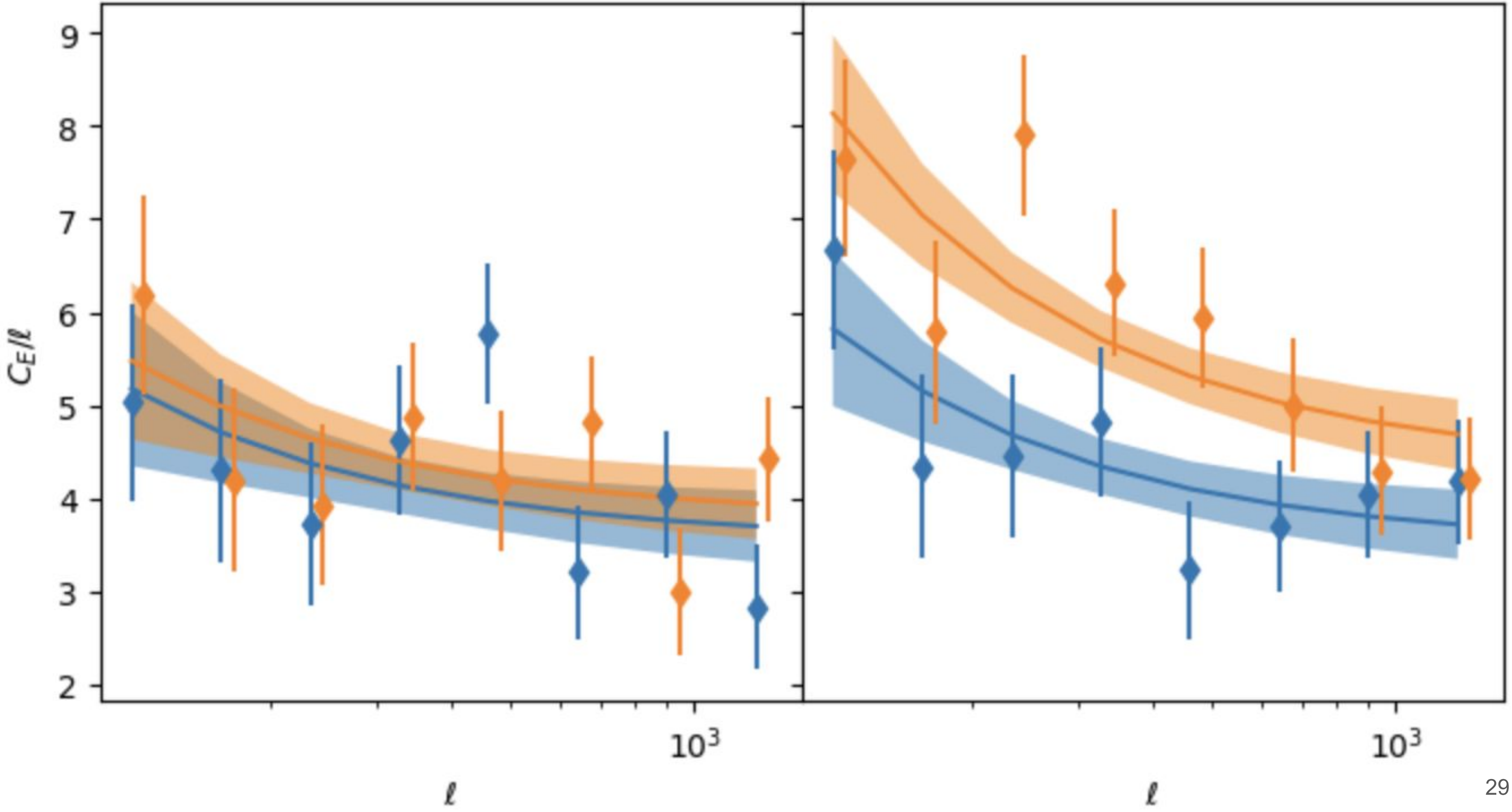
```python
def prior(cube, ndim, nparams):
    # define a linear transformation from [0,1] to the prior interval
    boundaries = np.array([[-3,9],[0.021,0.057]])
    for i in range(ndim):
        cube[i] = boundaries[i][0] + (boundaries[i][1]-boundaries[i][0])*cube[i]
    return(transformed_cube)

def log_lkl(cube, ndim, nparams):
    # Define a Gaussian log-likelihood with the data that you loaded from the text file
    diff = model(ell, cube[0], cube[1]) - data_A
    cholesky_transform = scipy.linalg.cholesky(covmat, lower=True)
    y = scipy.linalg.solve_triangular(cholesky_transform, diff, lower=True)
    chi2 = y.dot(y)
    # return the log-likelihood
    return(-0.5*chi2)
```

```
pymultinest.run(log_lkl_1A, prior_1, nDims, resume = False, outputfiles_basename='out_multinest/chain_1A',
                n_live_points=nLive, sampling_efficiency=eff, evidence_tolerance=tol)
```


1A
1B
prior
0.056
0.048
0.040
0.032
0.024
x_2
−2.5
0.0
2.5
5.0
7.5
x_1
2A
2B
prior
0.056
0.048
0.040
0.032
0.024
x_2
−2.5
0.0
2.5
5.0
7.5
x_1

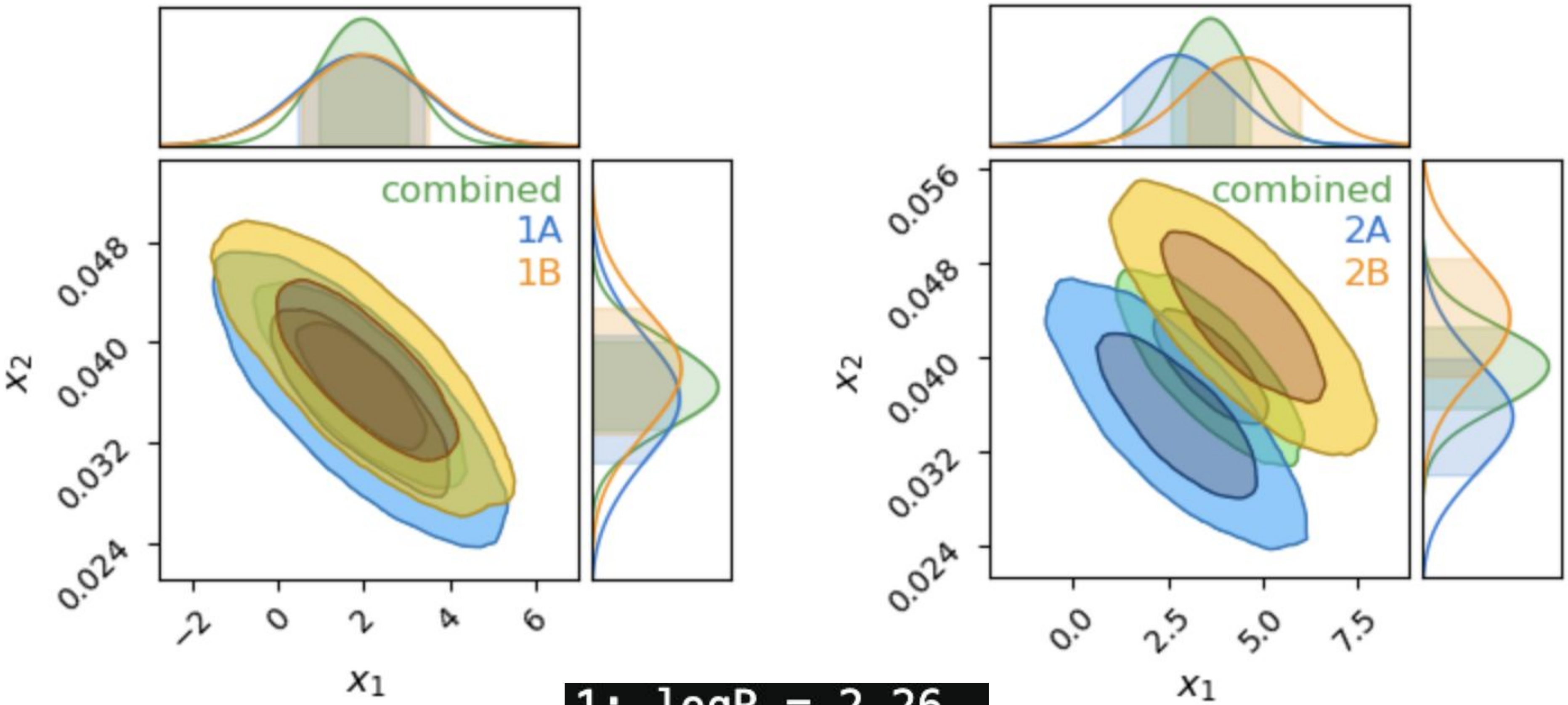
```
PPD_1A = np.array([model(ell, chain_1A[i,2], chain_1A[i,3]) for i in range(len(chain_1A))])
PPD_1B = np.array([model(ell, chain_1B[i,2], chain_1B[i,3]) for i in range(len(chain_1B))])

PPD_2A = np.array([model(ell, chain_2A[i,2], chain_2A[i,3]) for i in range(len(chain_2A))])
PPD_2B = np.array([model(ell, chain_2B[i,2], chain_2B[i,3]) for i in range(len(chain_2B))])
```

```
from statsmodels.stats.weightstats import DescrStatsW
weighted_stats_1A = DescrStatsW(PPD_1A, weights=chain_1A[:,0], ddof=0)
weighted_stats_1B = DescrStatsW(PPD_1B, weights=chain_1B[:,0], ddof=0)
weighted_stats_2A = DescrStatsW(PPD_2A, weights=chain_2A[:,0], ddof=0)
weighted_stats_2B = DescrStatsW(PPD_2B, weights=chain_2B[:,0], ddof=0)
```

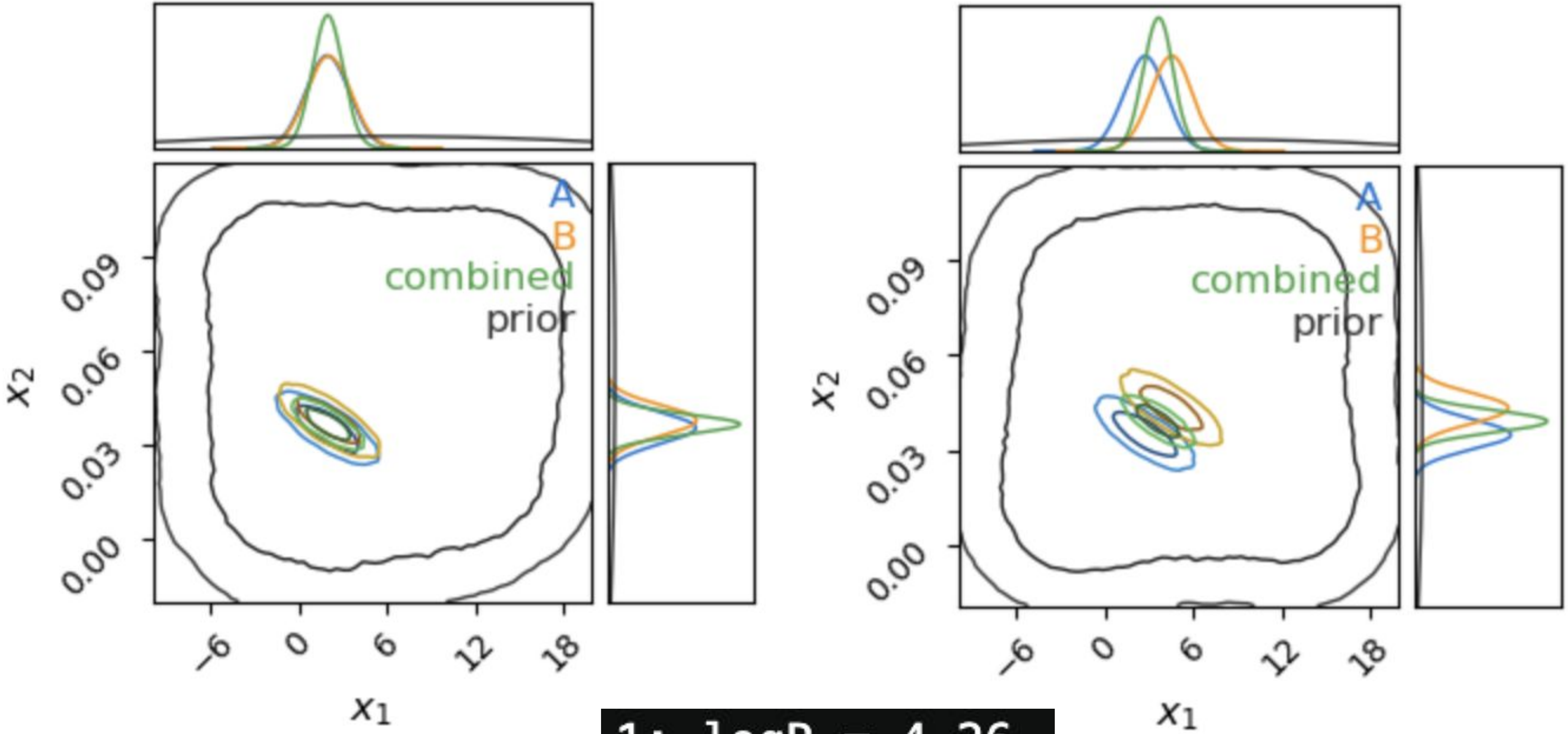


9
8
7
6
5
4
3
2
C_E/ℓ
10^3
ℓ
10^3
ℓ

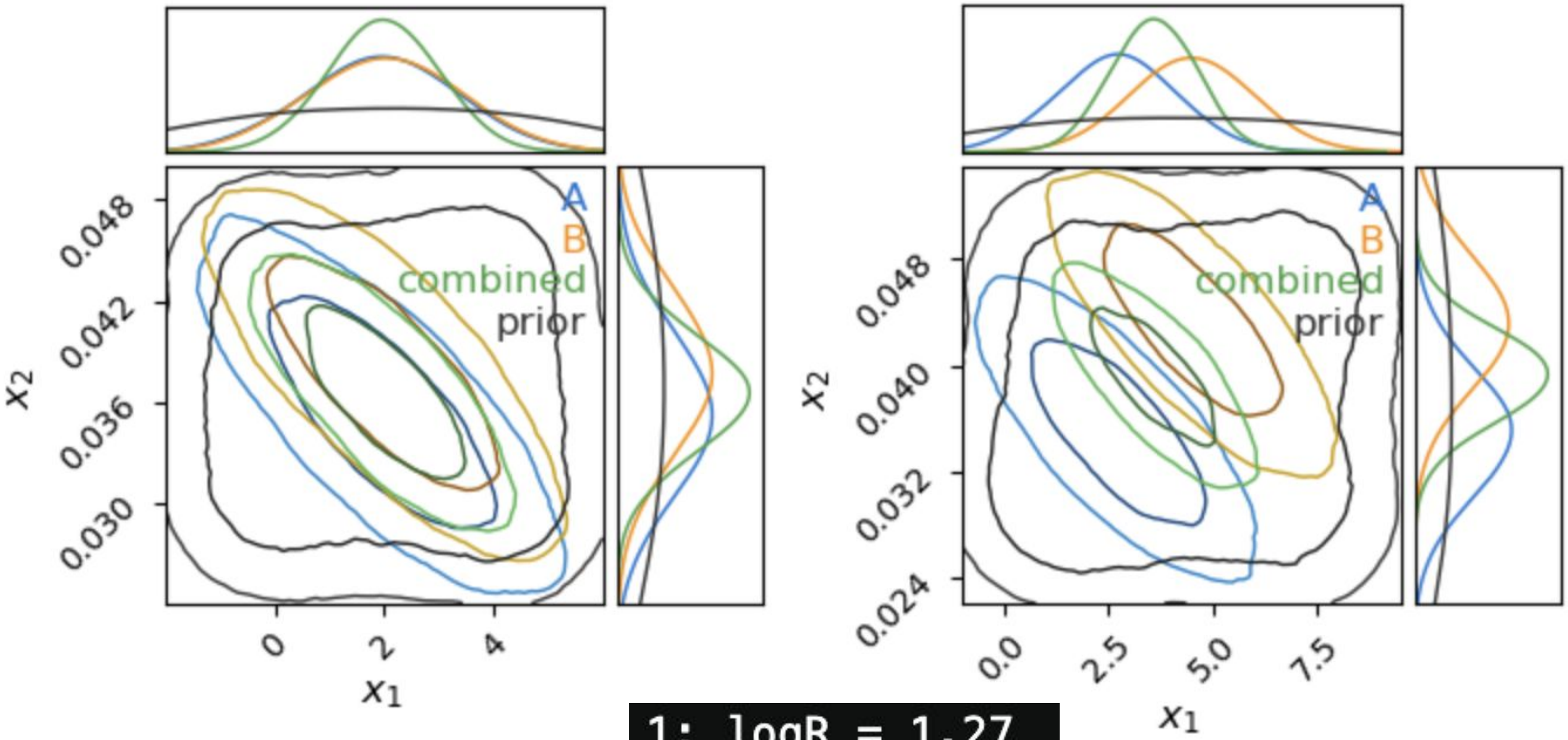
```
Z_1A = A1.get_stats()['global evidence']
Z_1B = B1.get_stats()['global evidence']
Z_1_combined = combined_1.get_stats()['global evidence']
Z_2A = A2.get_stats()['global evidence']
Z_2B = B2.get_stats()['global evidence']
Z_2_combined = combined_2.get_stats()['global evidence']
logR_2 = Z_2_combined - Z_2A - Z_2B
logR_1 = Z_1_combined - Z_1A - Z_1B
print('1: logR = %.2f'%logR_1)
print('2: logR = %.2f'%logR_2)
```




```
1: logR = 2.26
2: logR = -2.93
```




```
1: logR = 4.26
2: logR = −0.66
```



1: logR = 1.27
2: logR = -3.14

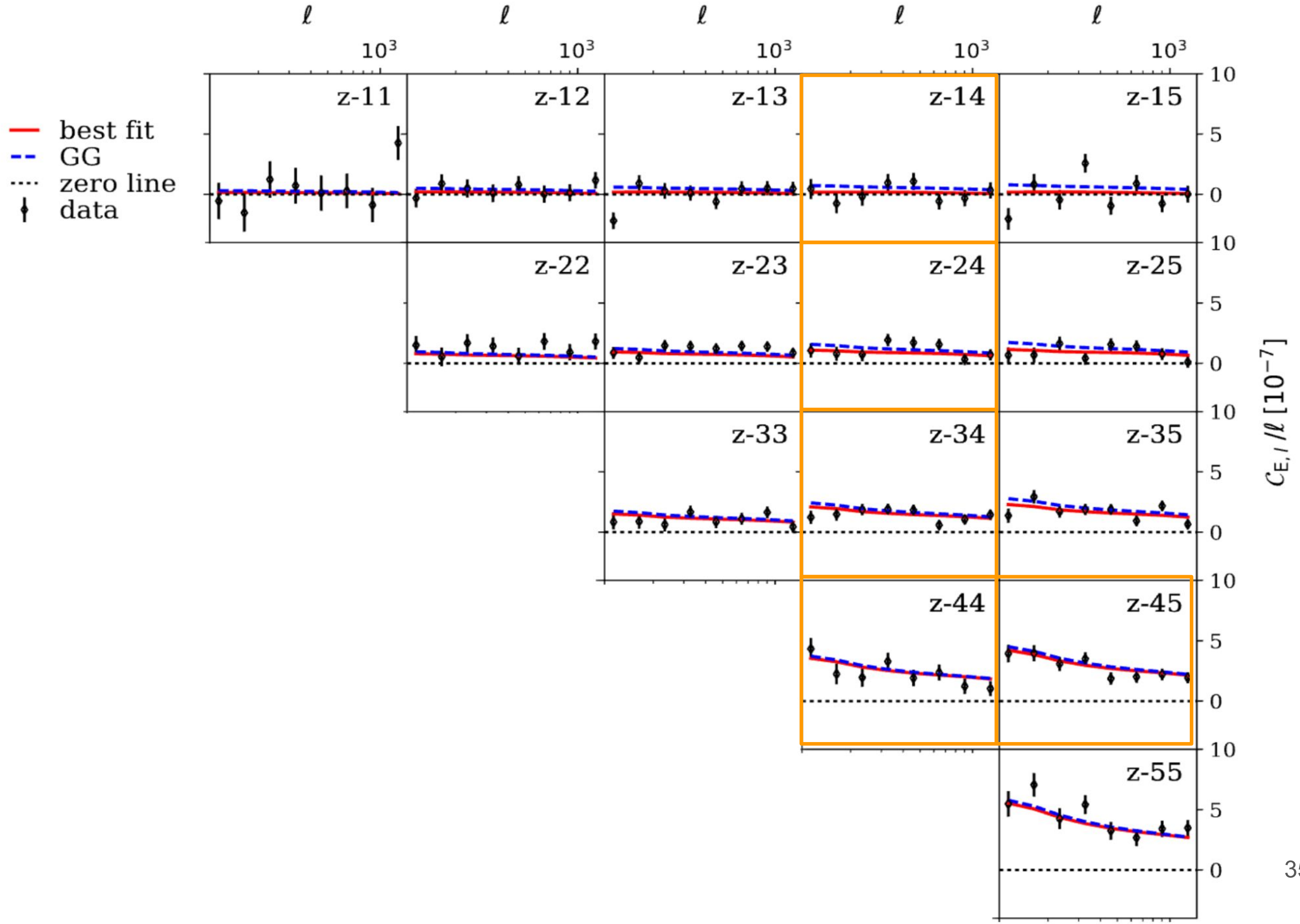
Suspiciousness

(Handley et al. 2019, Lemos et al. 2020)

"To address these concerns about prior volume, we can instead use the Suspiciousness S introduced in H19, which can be understood as the value of R that corresponds to the narrowest possible priors that do not significantly alter the shape of the posteriors"

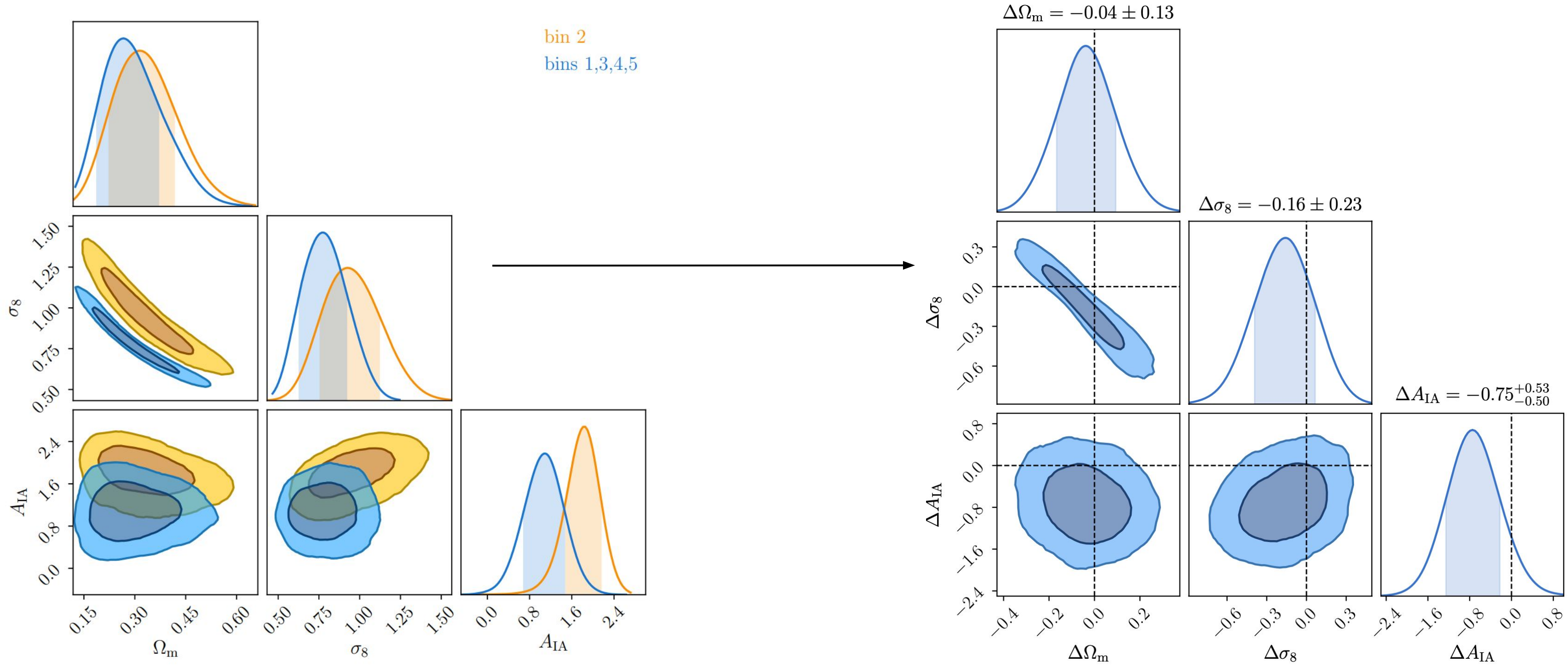
$\downarrow$

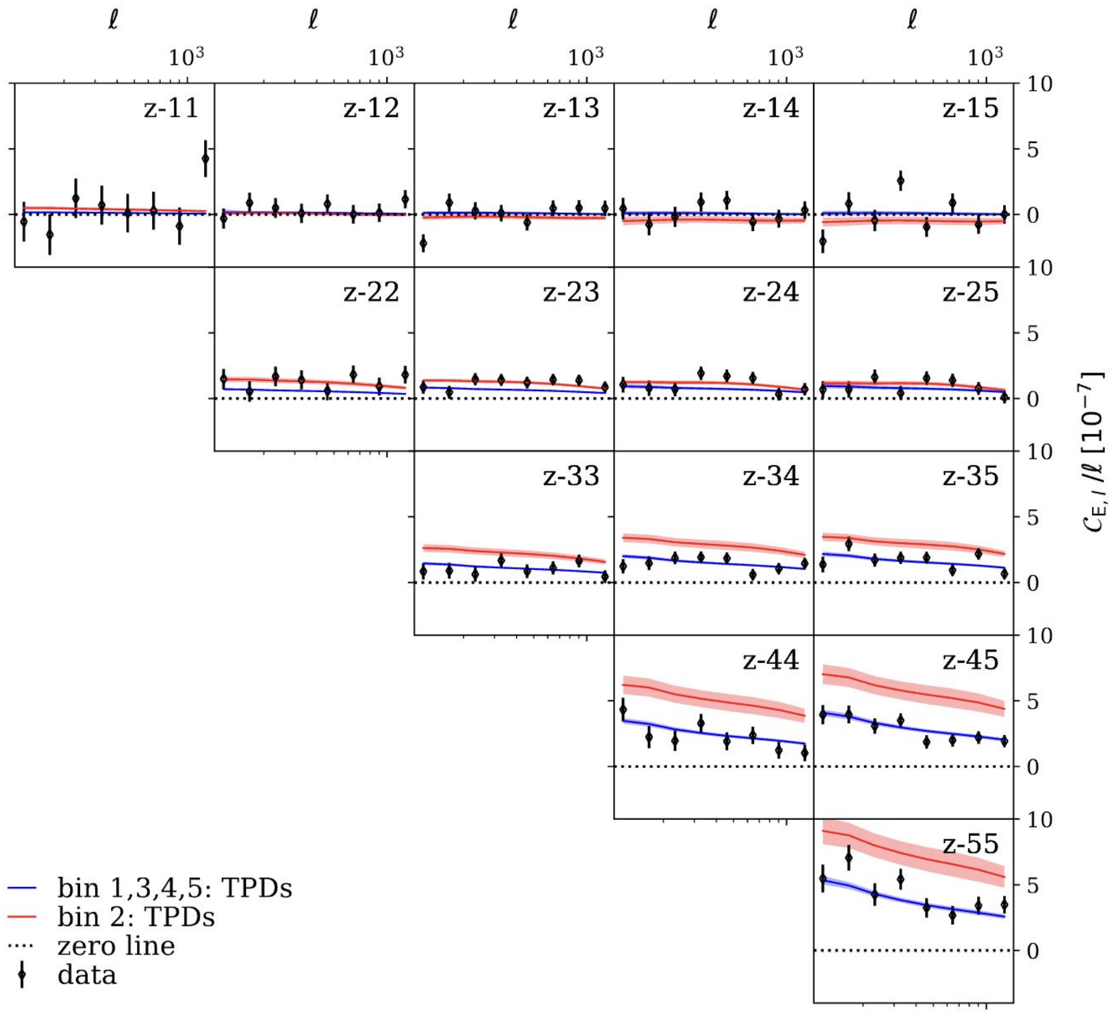
N σ


best fit
GG
zero line
data
z-11
z-12
z-13
z-14
z-15
z-22
z-23
z-24
z-25
z-33
z-34
z-35
z-44
z-45
z-55
ℓ
10^3
$C_{E,I}/\ell\ [10^{-7}]$
10
5
0

Parameter space methods: KiDS-1000

- Compute the posterior distribution of the difference between model parameter duplicates, quantify the deviation from zero





ℓ
10^3
z-11
z-12
z-13
z-14
z-15
z-22
z-23
z-24
z-25
z-33
z-34
z-35
z-44
z-45
z-55
10
5
0
$C_{E,I}/\ell\ [10^{-7}]$
bin 1,3,4,5: TPDs
bin 2: TPDs
zero line
data

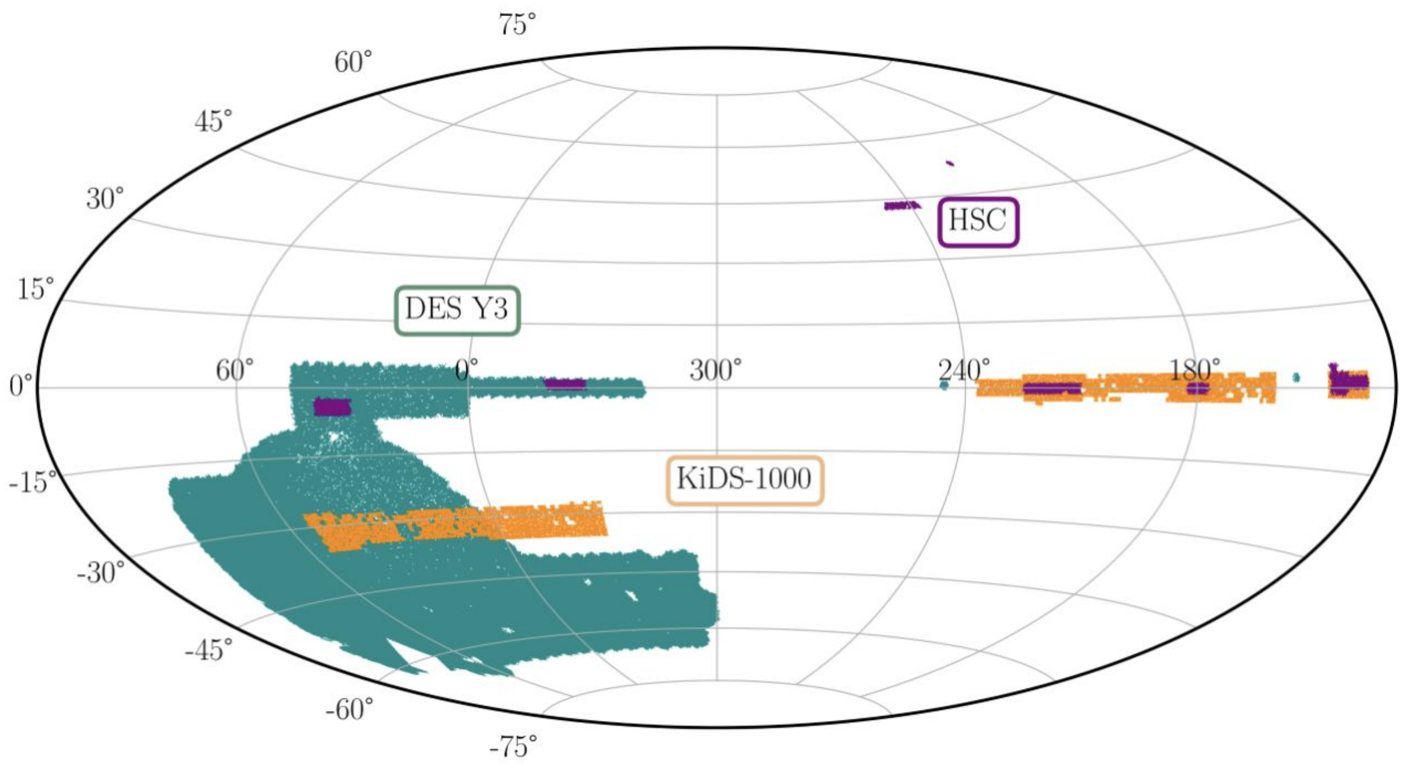


Figure 7. Survey footprints from DES Y3 (green) and KiDS-1000 (orange). The HSC-Y1 footprint (purple) overlaps KiDS in the North and DES in the South complicating the modelling of cross-survey covariance. For this reason, we limit our joint-survey analysis to DES and KiDS.

KiDS + DES et al., arXiv:2305.17173