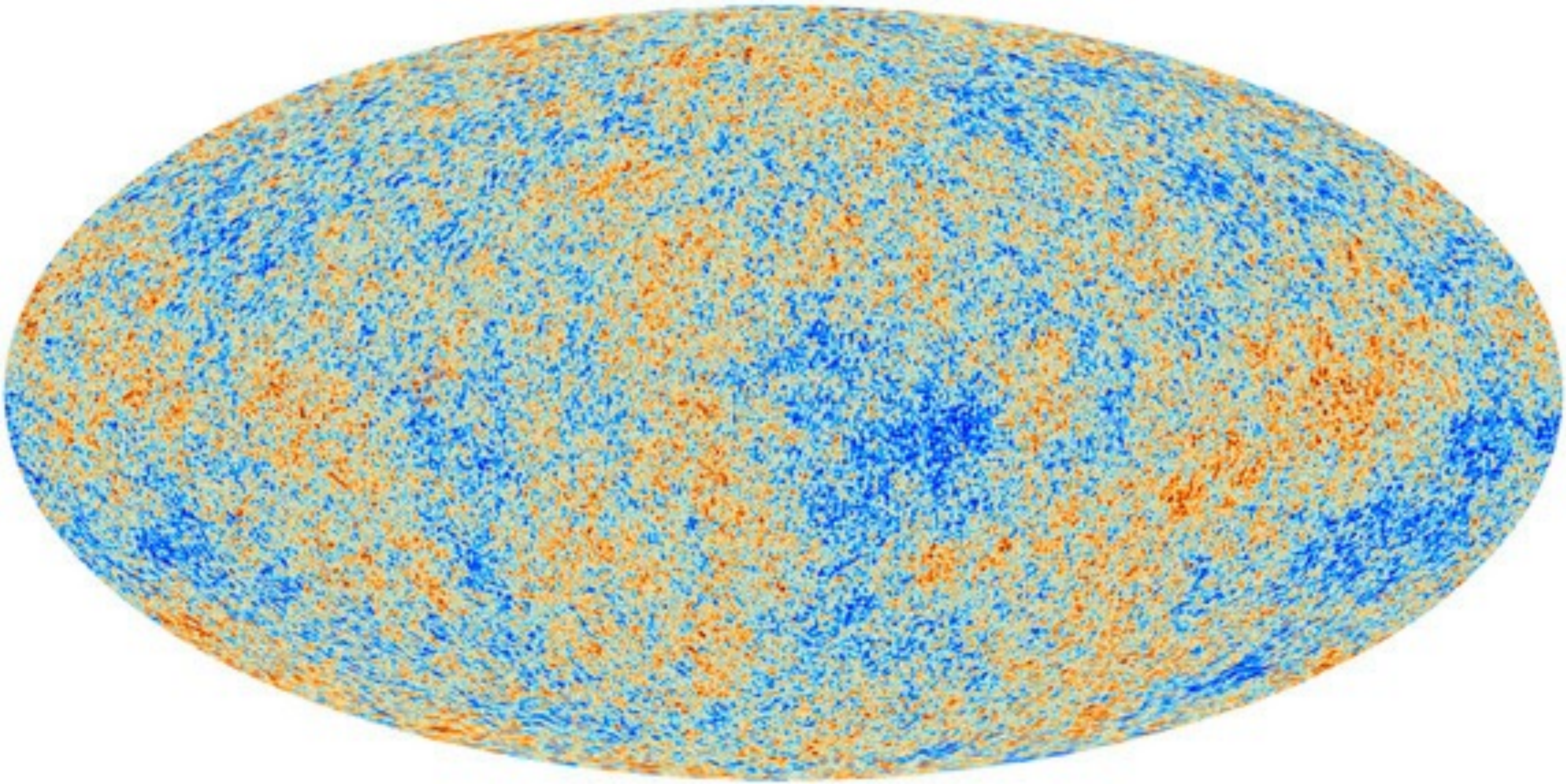
Cosmological Data Analysis for the 21st Century

Pedro G. Ferreira

UNIVERSITY OF OXFORD

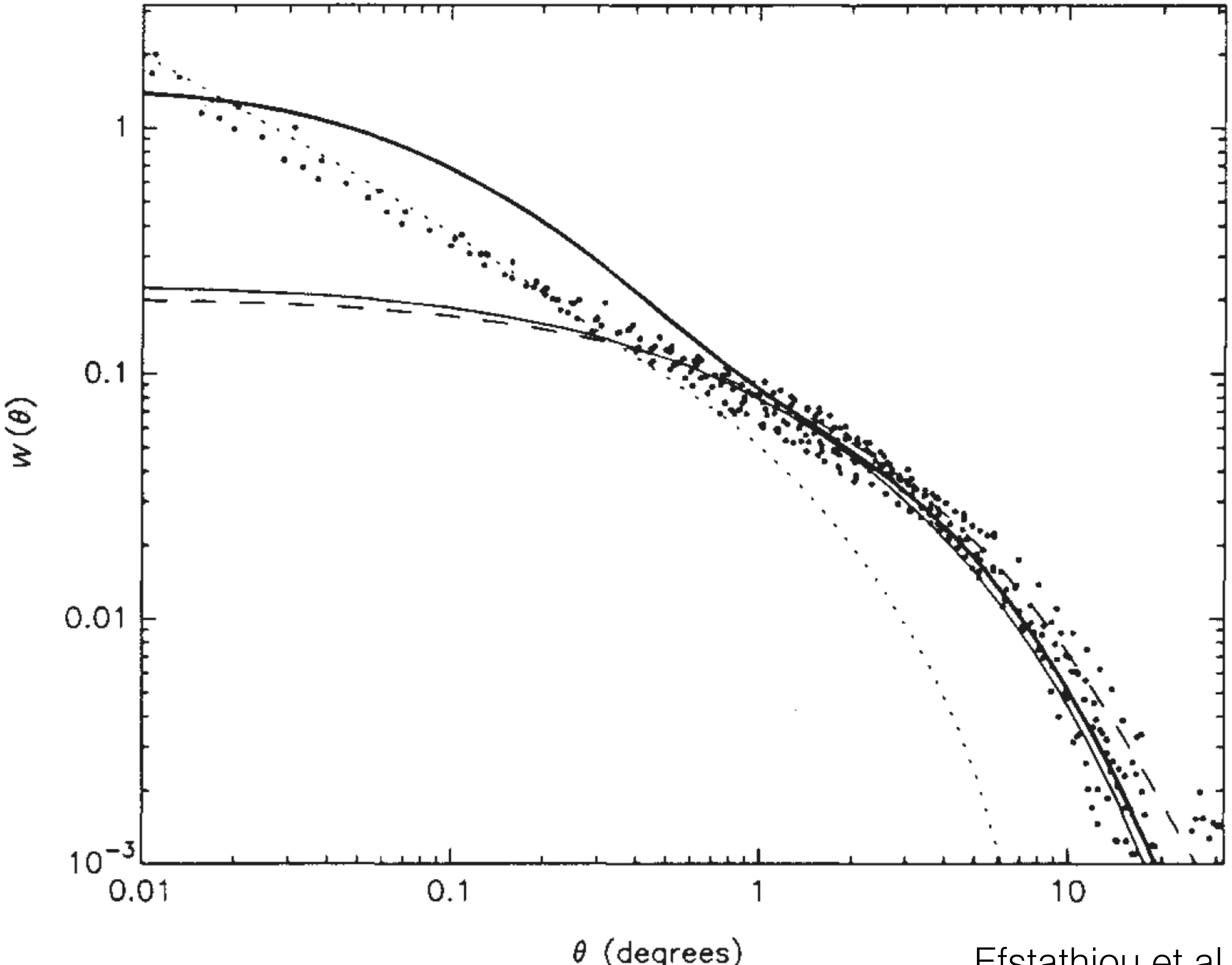
Prague, 2023

CMB today



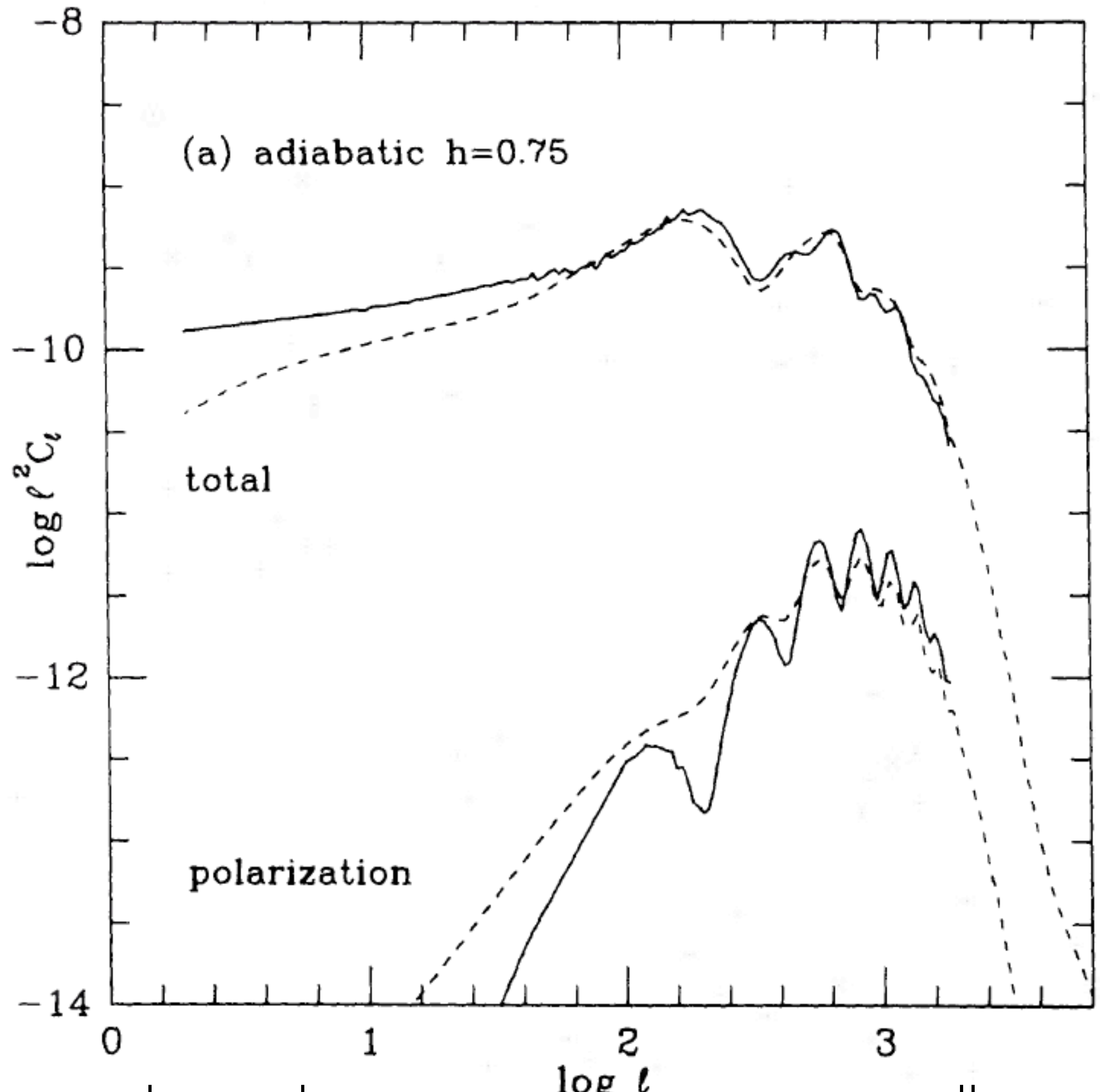
Planck 2015

The 1990s



Efstathiou et al 199

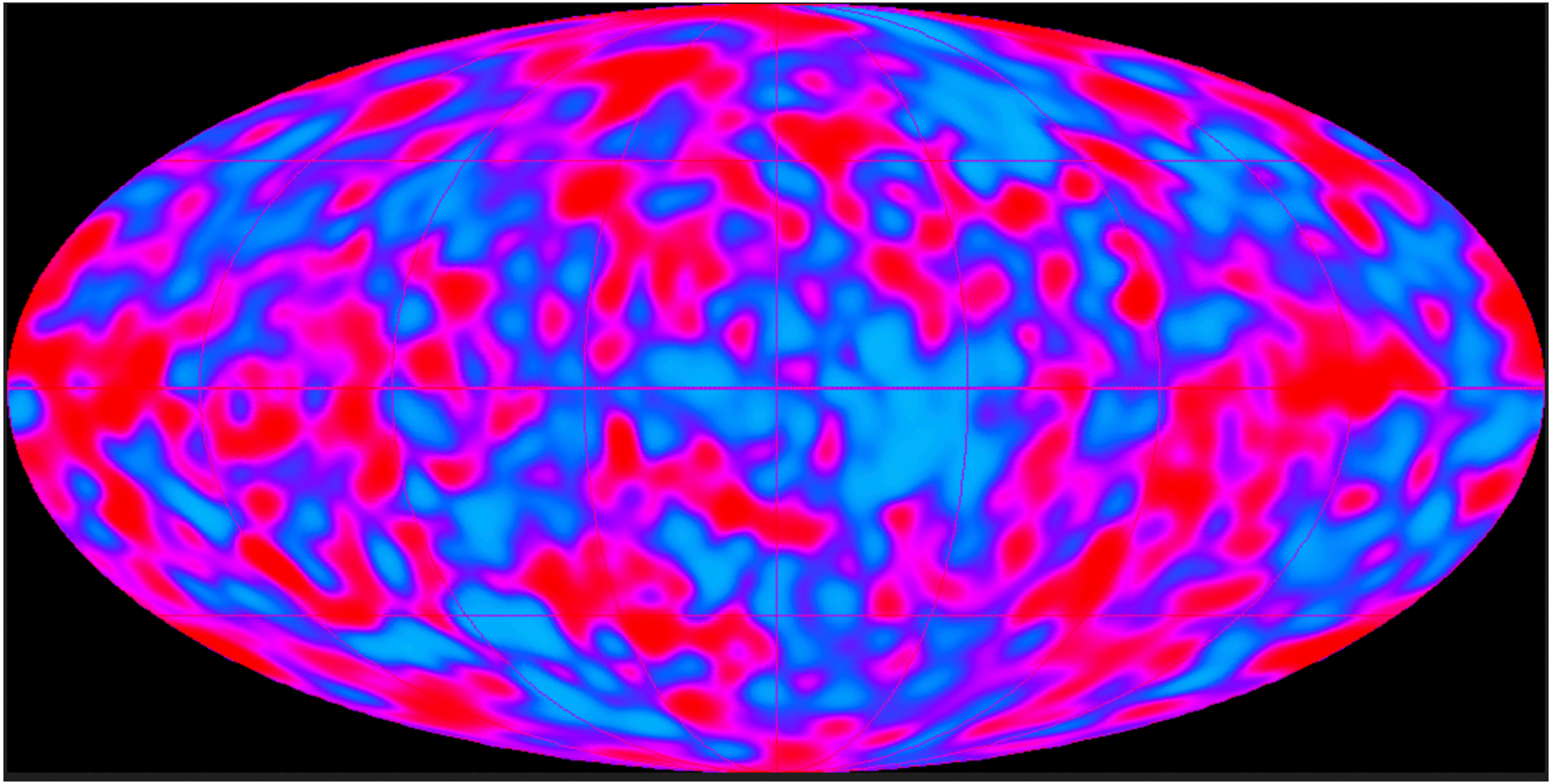
amount of fluctuations



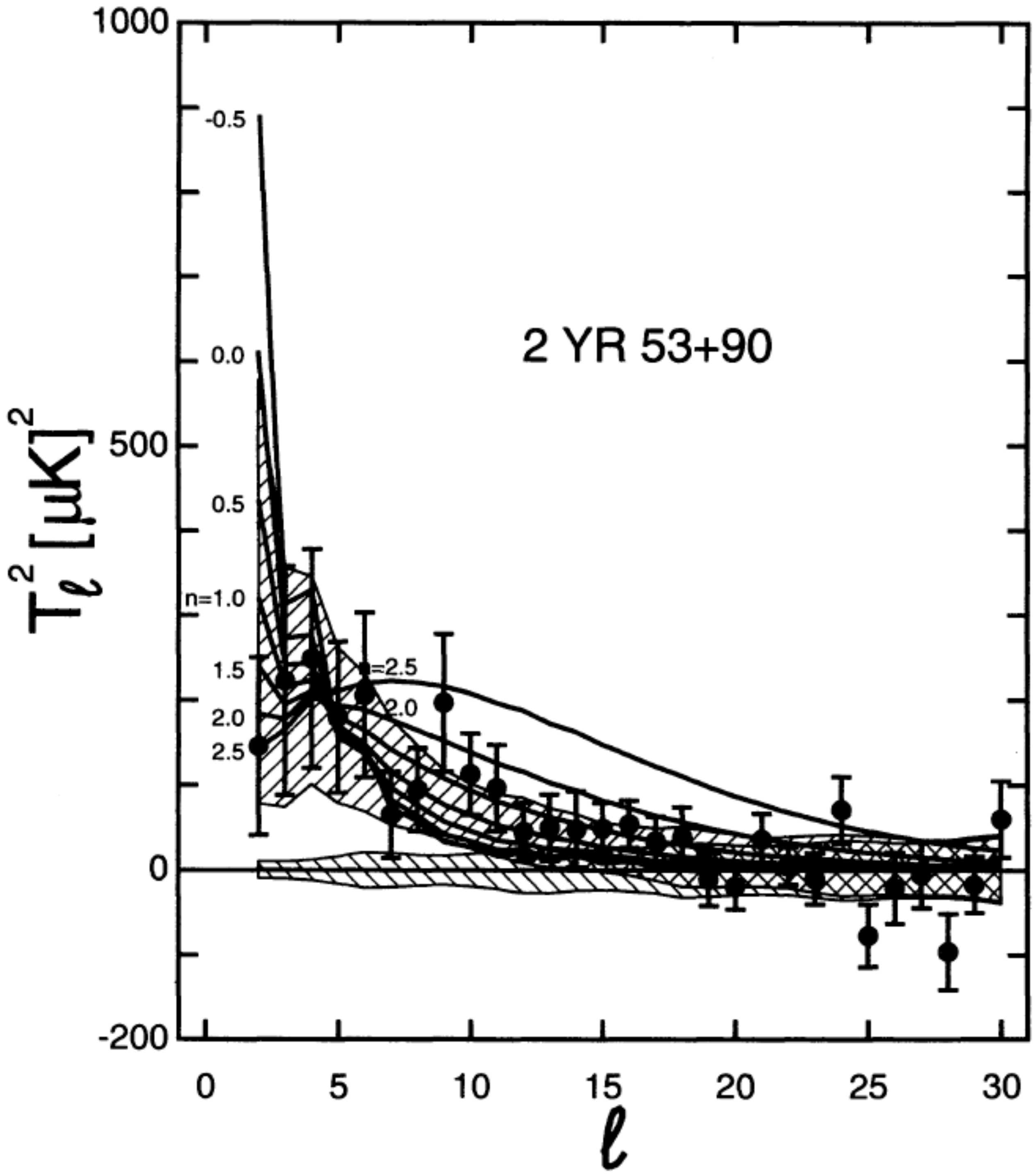
large angular scales

small angular scales

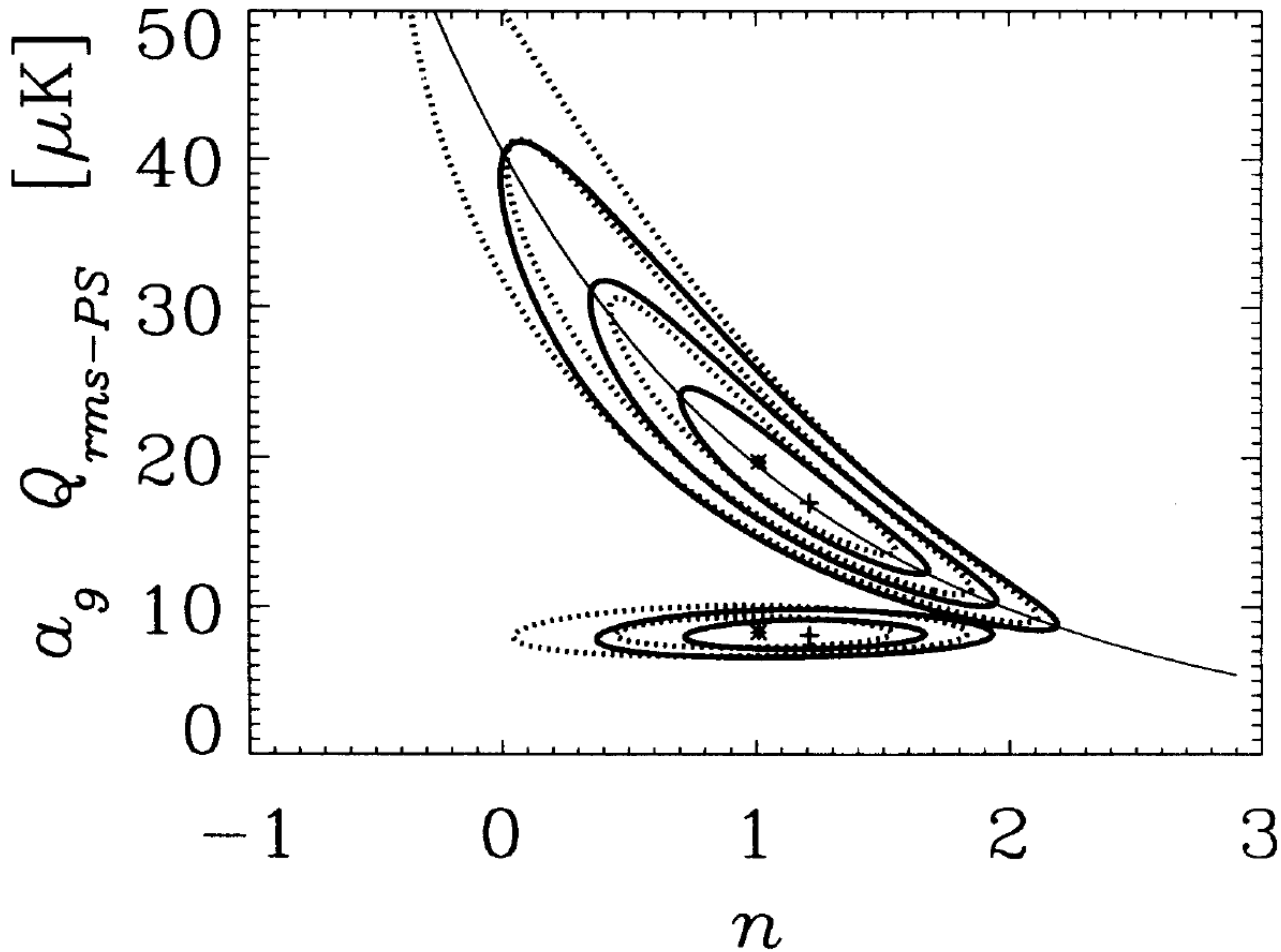
Bond & Efstathiou 1987



COBE 1992



Wright *et al* (COBE) 1994

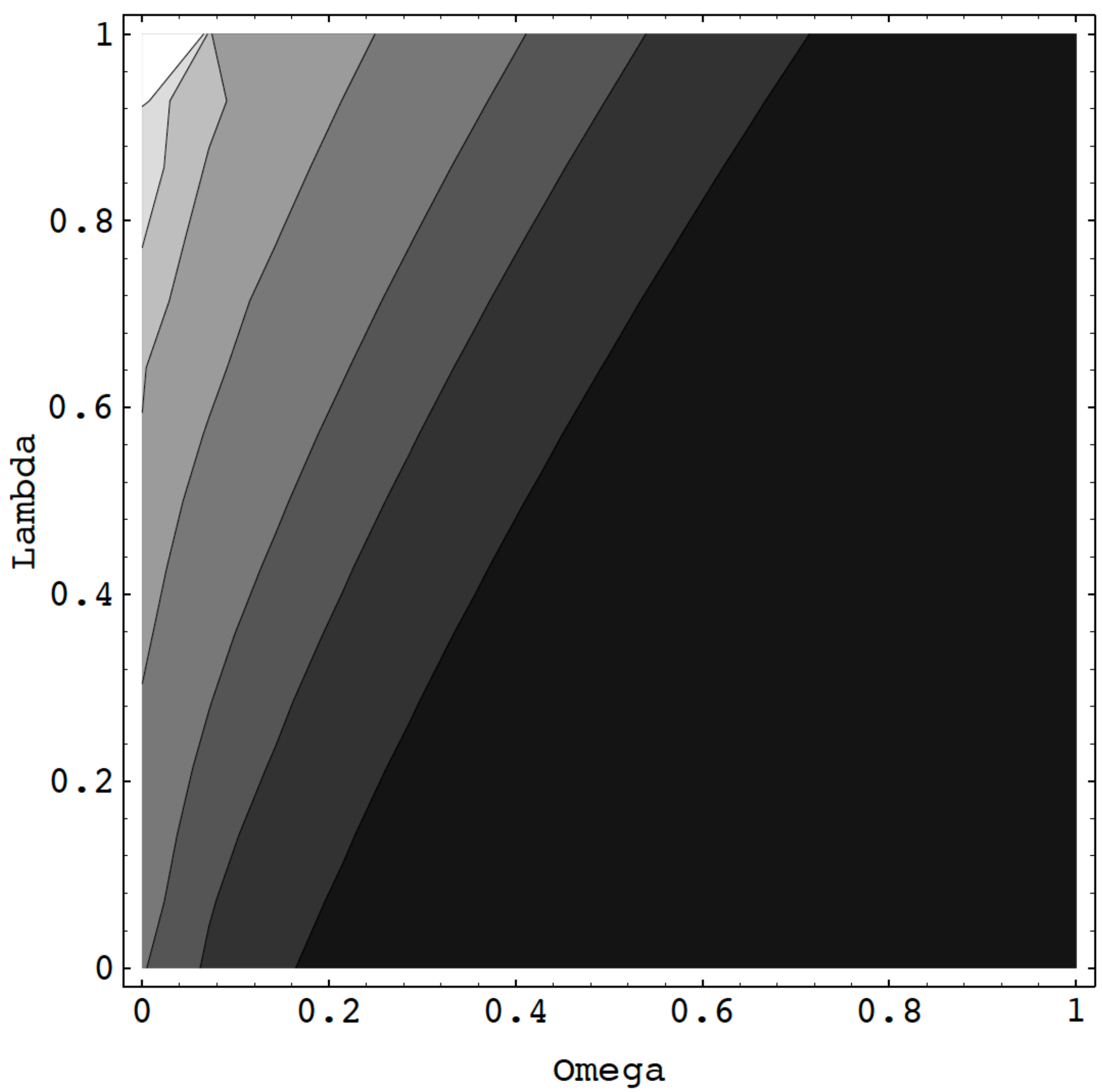


Gorski *et al* (COBE) 1994

Models and Evidence circa 1995

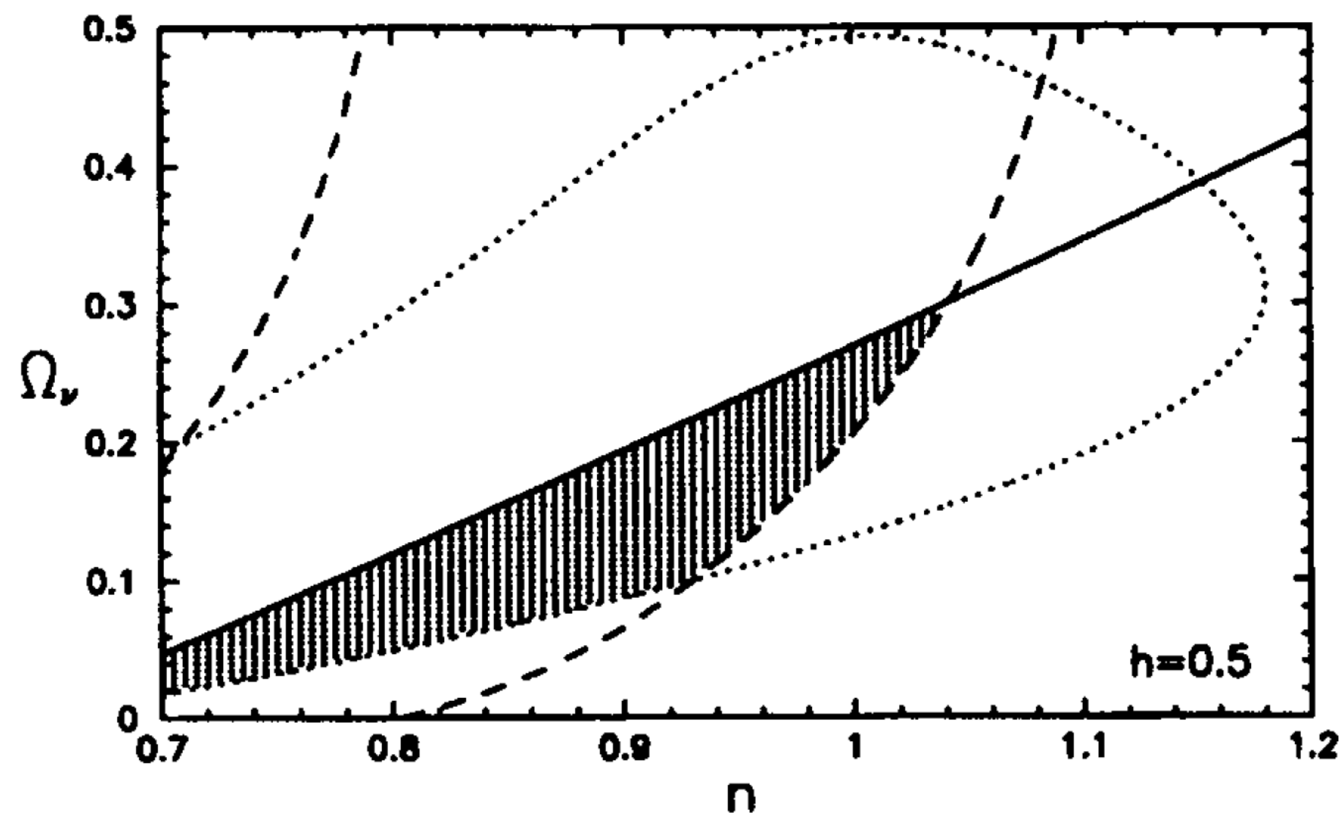
H_0 and Odds on Cosmology

Andrew Jaffe
Canadian Institute for Theoertical Astrophysics,
60 St. George St., Toronto, Ontario M5S 1A1, Canada

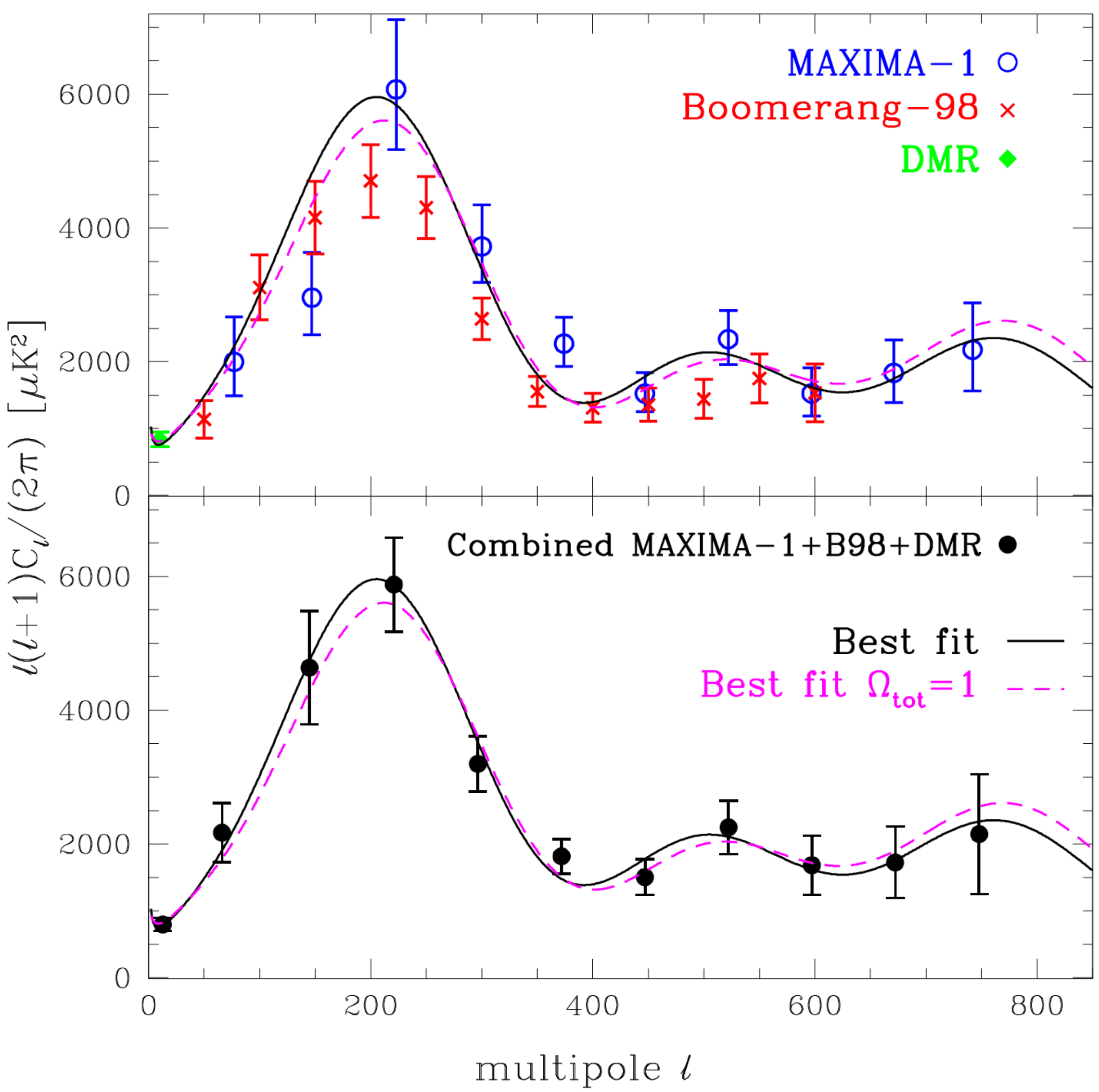


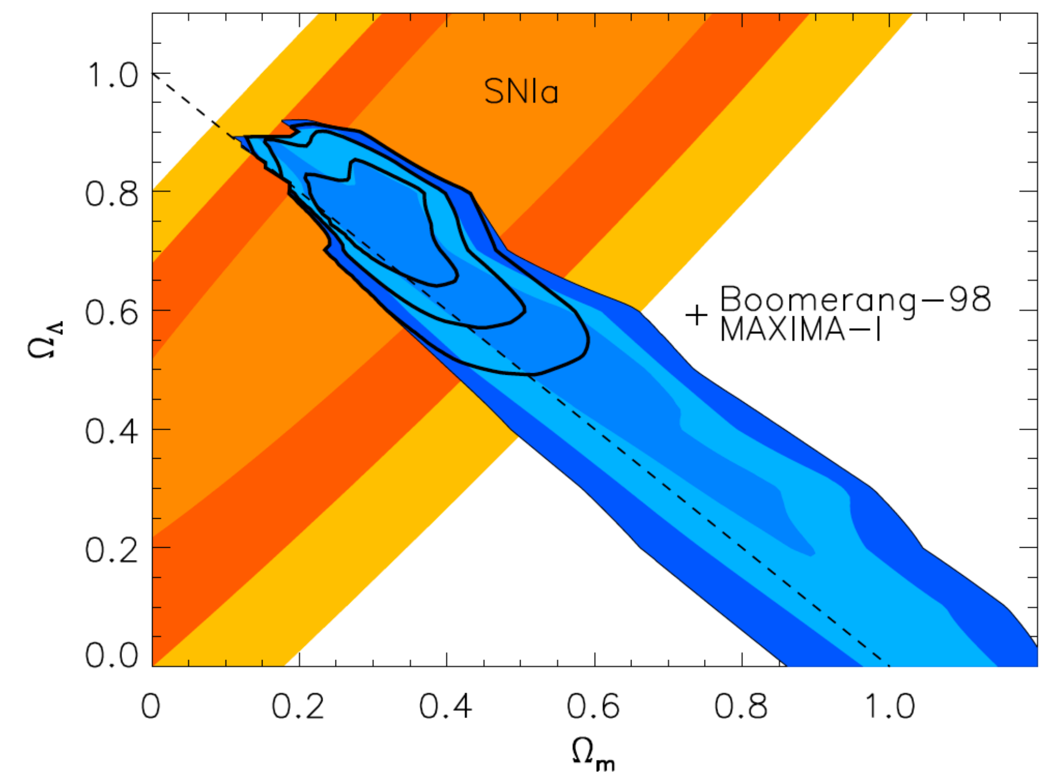
Pursuing parameters for critical-density dark matter models

Andrew R. Liddle,[1] David H. Lyth,[2] R. K. Schaefer,[3] Q. Shafi[3] and Pedro T. P. Viana[1]
[1]*Astronomy Centre, University of Sussex, Falmer, Brighton BN1 9QH*
[2]*School of Physics and Chemistry, University of Lancaster, Lancaster LA1 4YB*
[3]*Bartol Research Institute, University of Delaware, Newark, Delaware 19716, USA*



Models and Evidence circa 2000





	Ω_{tot}	$\Omega_b h^2$	n_s	$\Omega_c h^2$
B98+DMR	$1.15^{+0.10}_{-0.09}$	$0.036^{+0.006}_{-0.005}$	$1.04^{+0.10}_{-0.09}$	$0.24^{+0.08}_{-0.09}$
MAXIMA-1+DMR	$1.01^{+0.09}_{-0.09}$	$0.031^{+0.007}_{-0.006}$	$1.06^{+0.10}_{-0.09}$	$0.18^{+0.07}_{-0.06}$
B98+MAXIMA-1+DMR	$1.11^{+0.07}_{-0.07}$	$0.032^{+0.005}_{-0.005}$	$1.01^{+0.09}_{-0.08}$	$0.14^{+0.06}_{-0.05}$
$+(\Omega_{tot}=1)$	1	$0.030^{+0.004}_{-0.004}$	$0.99^{+0.07}_{-0.06}$	$0.19^{+0.07}_{-0.06}$
CMB+LSS	$1.11^{+0.05}_{-0.05}$	$0.032^{+0.004}_{-0.004}$	$1.00^{+0.09}_{-0.06}$	$0.13^{+0.02}_{-0.01}$
CMB+SNIa	$1.09^{+0.06}_{-0.05}$	$0.032^{+0.005}_{-0.005}$	$1.00^{+0.09}_{-0.08}$	$0.10^{+0.04}_{-0.04}$
CMB+LSS+SNIa	$1.06^{+0.04}_{-0.04}$	$0.033^{+0.005}_{-0.004}$	$1.03^{+0.09}_{-0.07}$	$0.14^{+0.03}_{-0.02}$

MAXIMA & BOOMERANG 2001

MCMC arrives in 2001!

Bayesian Methods for Cosmological Parameter Estimation from Cosmic Microwave Background Measurements

Nelson Christensen[1], Renate Meyer[2], Lloyd Knox[3], Ben Luey[1]
[*]

[1]Physics and Astronomy, Carleton College, Northfield, MN 55057,USA
[2]Department of Statistics, The University of Auckland, Auckland, New Zealand
[3]Department of Physics, University of California - Davis, Davis, CA 95616, USA

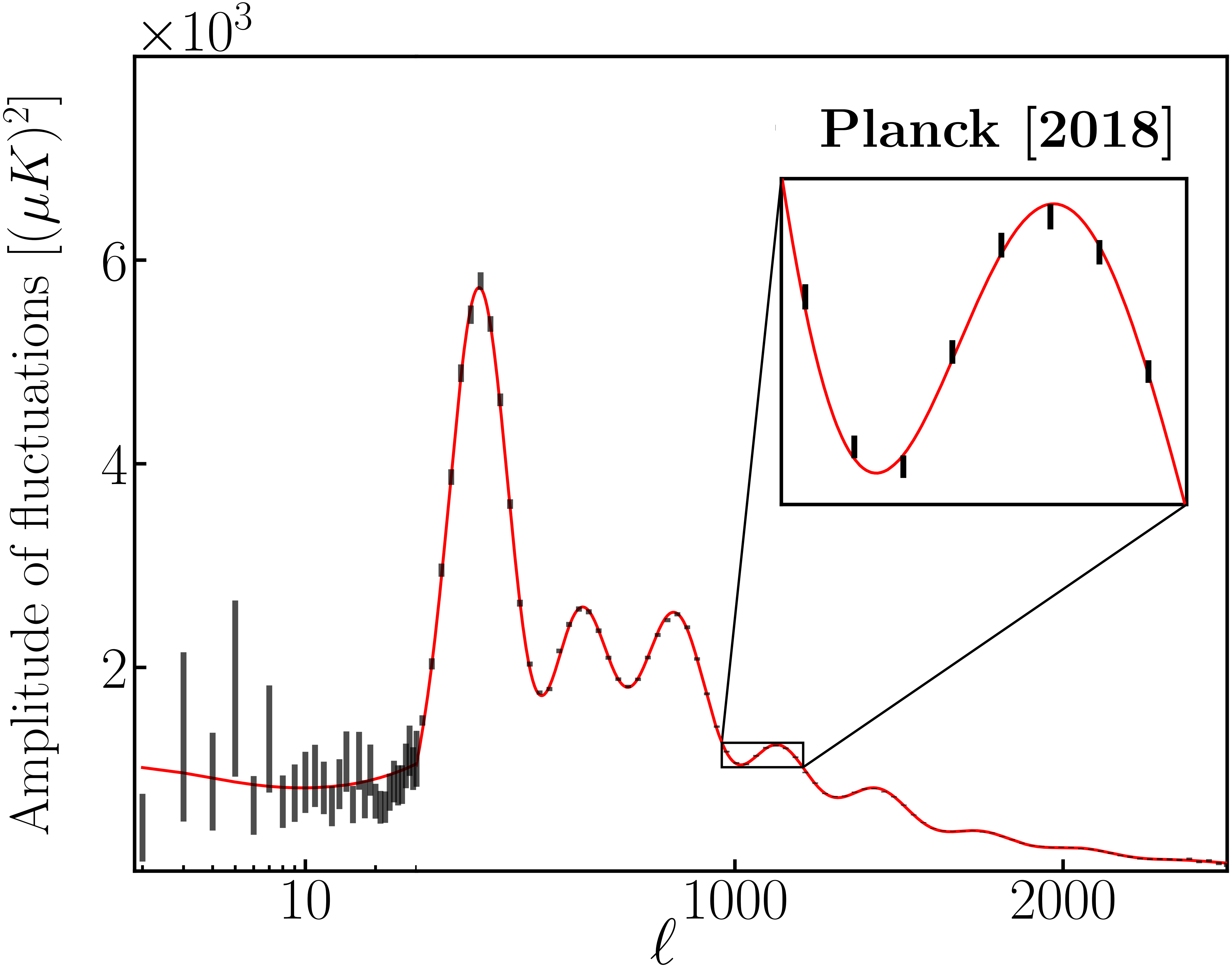
Abstract. We present a strategy for a statistically rigorous Bayesian approach to the problem of determining cosmological parameters from the results of observations of anisotropies in the cosmic microwave background. Our strategy relies on Markov chain Monte Carlo methods, specifically the Metropolis-Hastings algorithm, to perform the necessary high–dimensional integrals. We describe the Metropolis–Hastings algorithm in detail and discuss the results of our test on simulated data.

The Methodological Revolution

1990s: The rise of Bayesian statistics

- Likelihoods
- Evidence
- MCMC
- Hierarchical Models
- etc …

Now

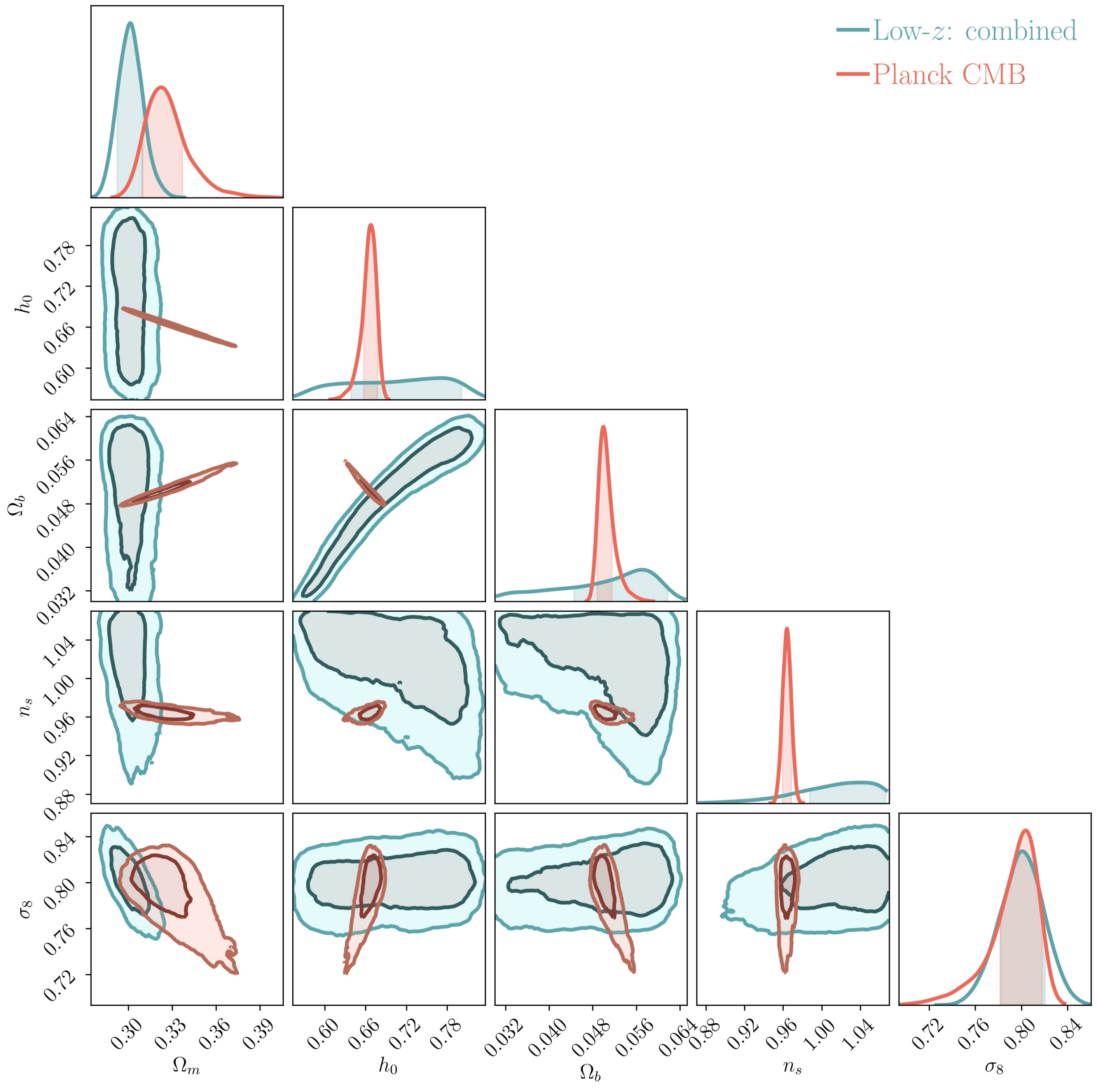


$\times 10^3$
Amplitude of fluctuations $[(\mu K)^2]$
6
4
2
10
1000
2000
ℓ
Planck [2018]

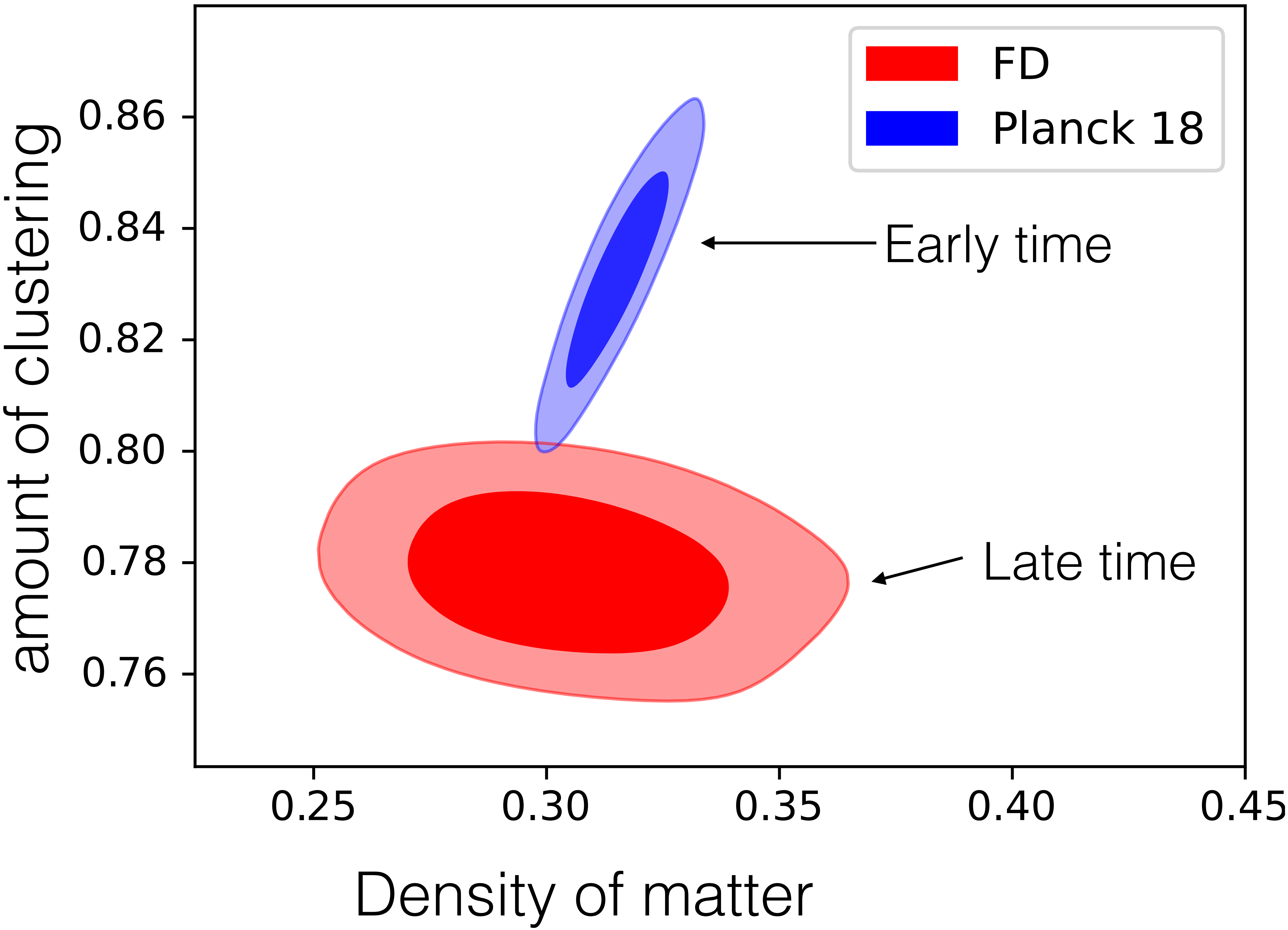
Parameter	TT+lowE 68% limits	TE+lowE 68% limits	EE+lowE 68% limits	TT,TE,EE+lowE 68% limits	TT,TE,EE+lowE+lensing 68% limits	TT,TE,EE+lowE+lensing+BAO 68% limits
$\Omega_b h^2$	0.02212 ± 0.00022	0.02249 ± 0.00025	0.0240 ± 0.0012	0.02236 ± 0.00015	0.02237 ± 0.00015	0.02242 ± 0.00014
$\Omega_c h^2$	0.1206 ± 0.0021	0.1177 ± 0.0020	0.1158 ± 0.0046	0.1202 ± 0.0014	0.1200 ± 0.0012	0.11933 ± 0.00091
$100\theta_{MC}$	1.04077 ± 0.00047	1.04139 ± 0.00049	1.03999 ± 0.00089	1.04090 ± 0.00031	1.04092 ± 0.00031	1.04101 ± 0.00029
τ	0.0522 ± 0.0080	0.0496 ± 0.0085	0.0527 ± 0.0090	$0.0544^{+0.0070}_{-0.0081}$	0.0544 ± 0.0073	0.0561 ± 0.0071
$\ln(10^{10}A_s)$	3.040 ± 0.016	$3.018^{+0.020}_{-0.018}$	3.052 ± 0.022	3.045 ± 0.016	3.044 ± 0.014	3.047 ± 0.014
n_s	0.9626 ± 0.0057	0.967 ± 0.011	0.980 ± 0.015	0.9649 ± 0.0044	0.9649 ± 0.0042	0.9665 ± 0.0038
H_0 [km s^{-1} Mpc^{-1}] . .	66.88 ± 0.92	68.44 ± 0.91	69.9 ± 2.7	67.27 ± 0.60	67.36 ± 0.54	67.66 ± 0.42
Ω_Λ	0.679 ± 0.013	0.699 ± 0.012	$0.711^{+0.033}_{-0.026}$	0.6834 ± 0.0084	0.6847 ± 0.0073	0.6889 ± 0.0056
Ω_m	0.321 ± 0.013	0.301 ± 0.012	$0.289^{+0.026}_{-0.033}$	0.3166 ± 0.0084	0.3153 ± 0.0073	0.3111 ± 0.0056
$\Omega_m h^2$	0.1434 ± 0.0020	0.1408 ± 0.0019	$0.1404^{+0.0034}_{-0.0039}$	0.1432 ± 0.0013	0.1430 ± 0.0011	0.14240 ± 0.00087
$\Omega_m h^3$	0.09589 ± 0.00046	0.09635 ± 0.00051	$0.0981^{+0.0016}_{-0.0018}$	0.09633 ± 0.00029	0.09633 ± 0.00030	0.09635 ± 0.00030
σ_8	0.8118 ± 0.0089	0.793 ± 0.011	0.796 ± 0.018	0.8120 ± 0.0073	0.8111 ± 0.0060	0.8102 ± 0.0060
$S_8 \equiv \sigma_8(\Omega_m/0.3)^{0.5}$.	0.840 ± 0.024	0.794 ± 0.024	$0.781^{+0.052}_{-0.060}$	0.834 ± 0.016	0.832 ± 0.013	0.825 ± 0.011
$\sigma_8\Omega_m^{0.25}$	0.611 ± 0.012	0.587 ± 0.012	0.583 ± 0.027	0.6090 ± 0.0081	0.6078 ± 0.0064	0.6051 ± 0.0058
z_{re}	7.50 ± 0.82	$7.11^{+0.91}_{-0.75}$	$7.10^{+0.87}_{-0.73}$	7.68 ± 0.79	7.67 ± 0.73	7.82 ± 0.71
$10^9 A_s$	2.092 ± 0.034	2.045 ± 0.041	2.116 ± 0.047	$2.101^{+0.031}_{-0.034}$	2.100 ± 0.030	2.105 ± 0.030
$10^9 A_s e^{-2\tau}$	1.884 ± 0.014	1.851 ± 0.018	1.904 ± 0.024	1.884 ± 0.012	1.883 ± 0.011	1.881 ± 0.010
Age [Gyr]	13.830 ± 0.037	13.761 ± 0.038	$13.64^{+0.16}_{-0.14}$	13.800 ± 0.024	13.797 ± 0.023	13.787 ± 0.020
z_*	1090.30 ± 0.41	1089.57 ± 0.42	$1087.8^{+1.6}_{-1.7}$	1089.95 ± 0.27	1089.92 ± 0.25	1089.80 ± 0.21
r_* [Mpc]	144.46 ± 0.48	144.95 ± 0.48	144.29 ± 0.64	144.39 ± 0.30	144.43 ± 0.26	144.57 ± 0.22
$100\theta_*$	1.04097 ± 0.00046	1.04156 ± 0.00049	1.04001 ± 0.00086	1.04109 ± 0.00030	1.04110 ± 0.00031	1.04119 ± 0.00029
z_{drag}	1059.39 ± 0.46	1060.03 ± 0.54	1063.2 ± 2.4	1059.93 ± 0.30	1059.94 ± 0.30	1060.01 ± 0.29
r_{drag} [Mpc]	147.21 ± 0.48	147.59 ± 0.49	146.46 ± 0.70	147.05 ± 0.30	147.09 ± 0.26	147.21 ± 0.23
k_D [Mpc^{-1}]	0.14054 ± 0.00052	0.14043 ± 0.00057	0.1426 ± 0.0012	0.14090 ± 0.00032	0.14087 ± 0.00030	0.14078 ± 0.00028
z_{eq}	3411 ± 48	3349 ± 46	3340^{+81}_{-92}	3407 ± 31	3402 ± 26	3387 ± 21
k_{eq} [Mpc^{-1}]	0.01041 ± 0.00014	0.01022 ± 0.00014	$0.01019^{+0.00025}_{-0.00028}$	0.010398 ± 0.000094	0.010384 ± 0.000081	0.010339 ± 0.000063
$100\theta_{s,eq}$	0.4483 ± 0.0046	0.4547 ± 0.0045	0.4562 ± 0.0092	0.4490 ± 0.0030	0.4494 ± 0.0026	0.4509 ± 0.0020

Planck 2015

Large Scale Structure (DES)



DES 2022

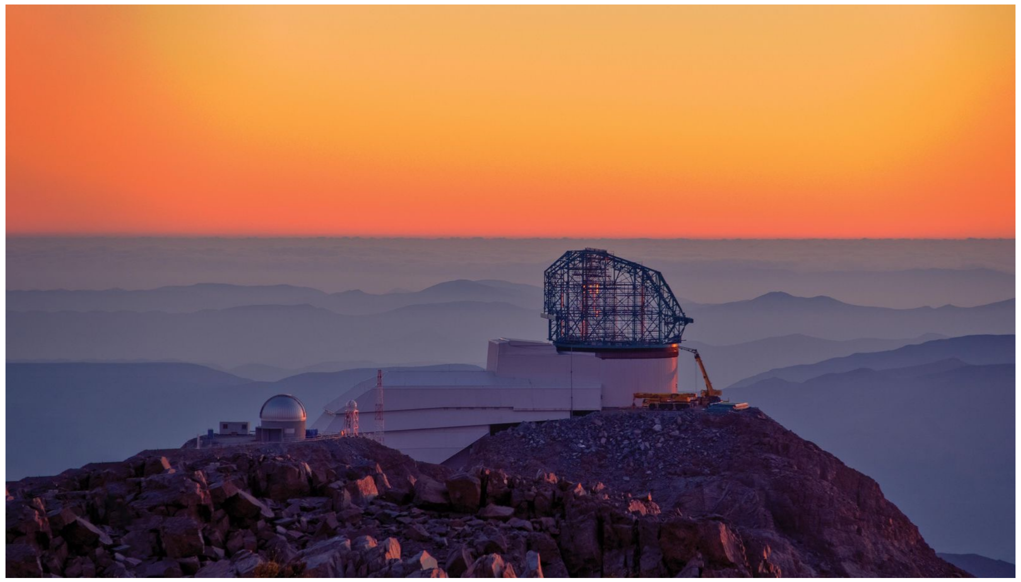


Garcia-Garcia et al 2021

But we can measure more …

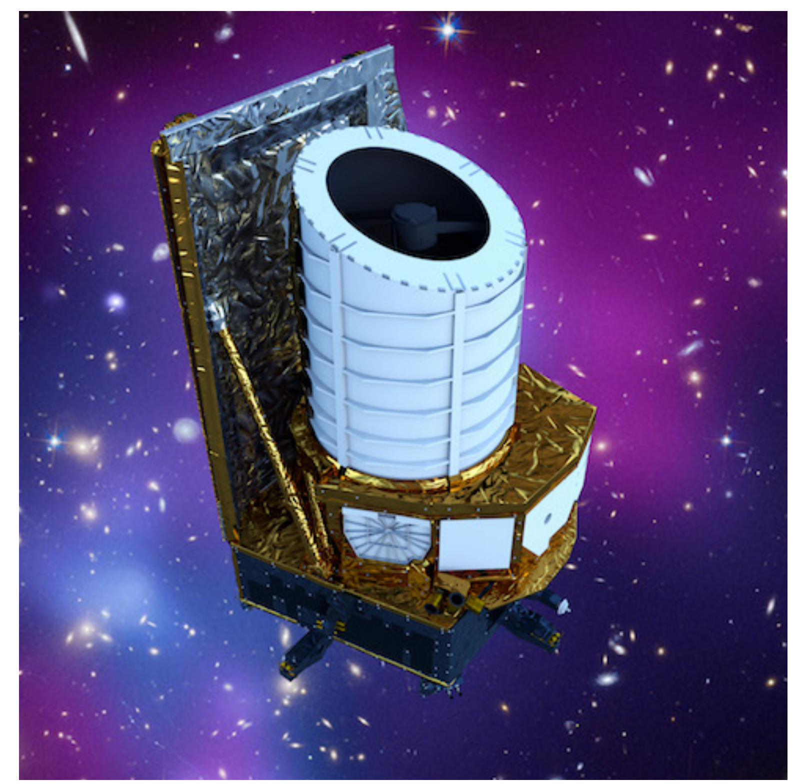
- Neutrino masses
- Gravitational parameters
- Coupling constants
- Magnetic fields
- Dark matter properties
- Fifth Forces
- Halo properties
- Bias
- Star Formation
- Feedback Mechanism
- …

The new observatories



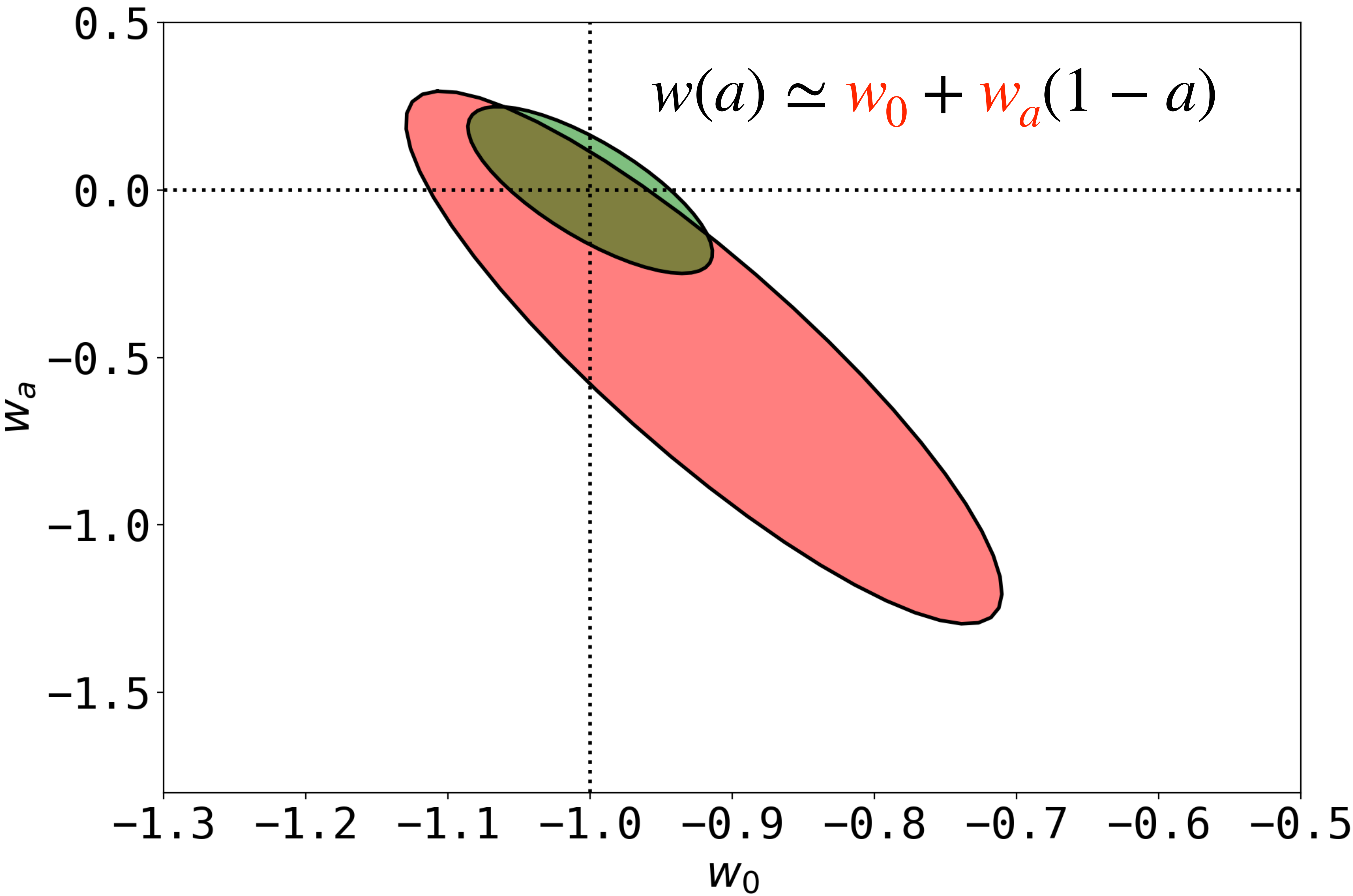
Vera Rubin Observatory

Simons Observatory



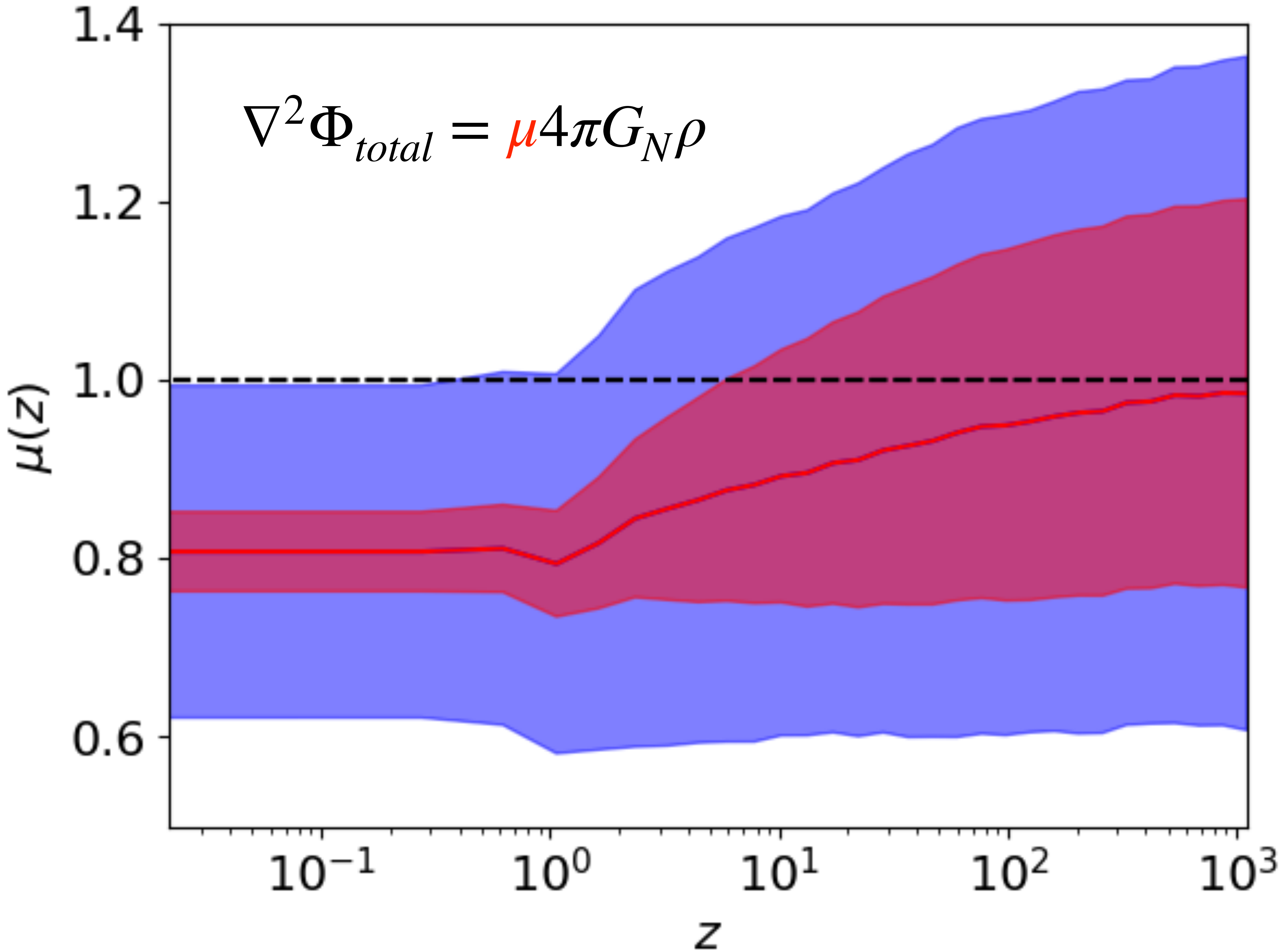
Euclid

Now + 10 years



Alam et al 2021

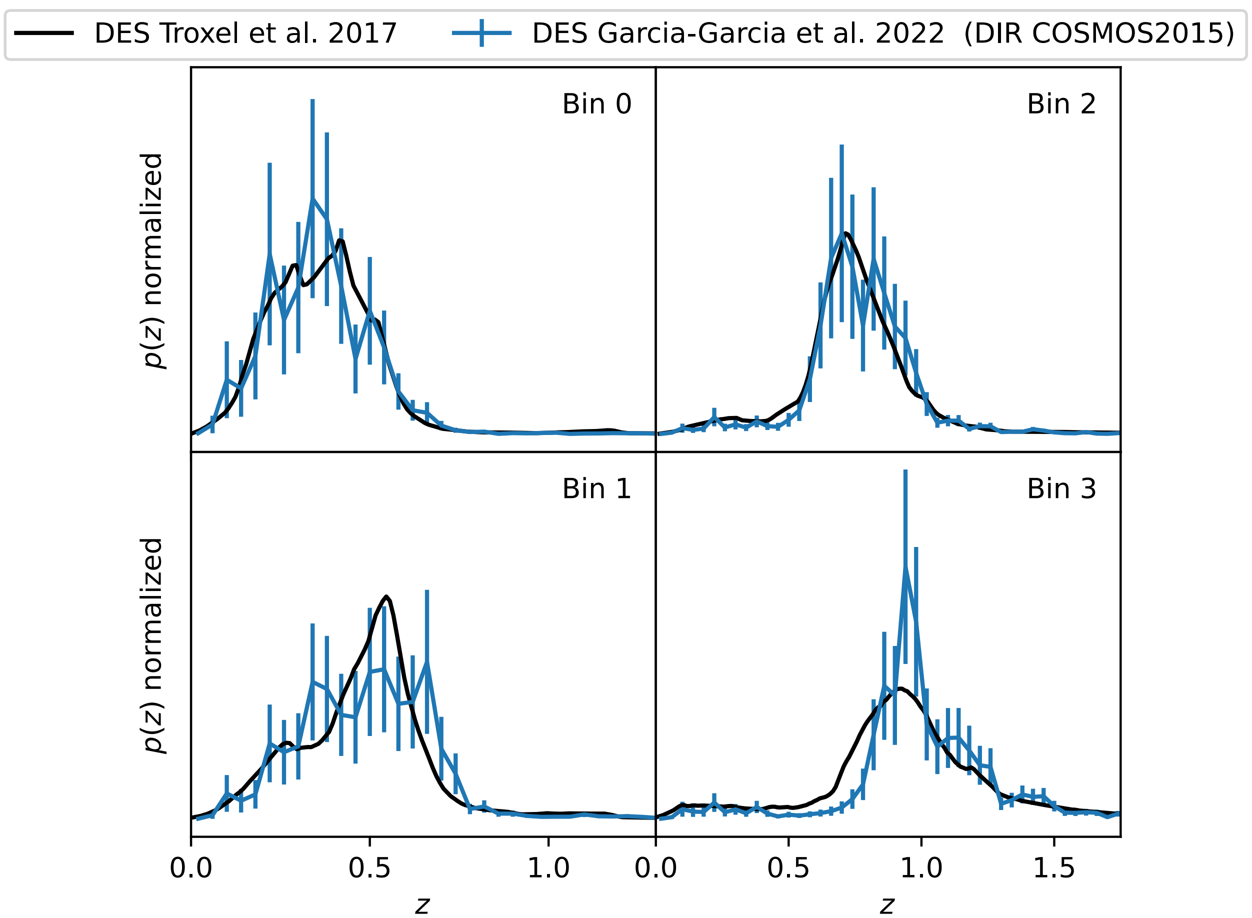
Now + 10 years

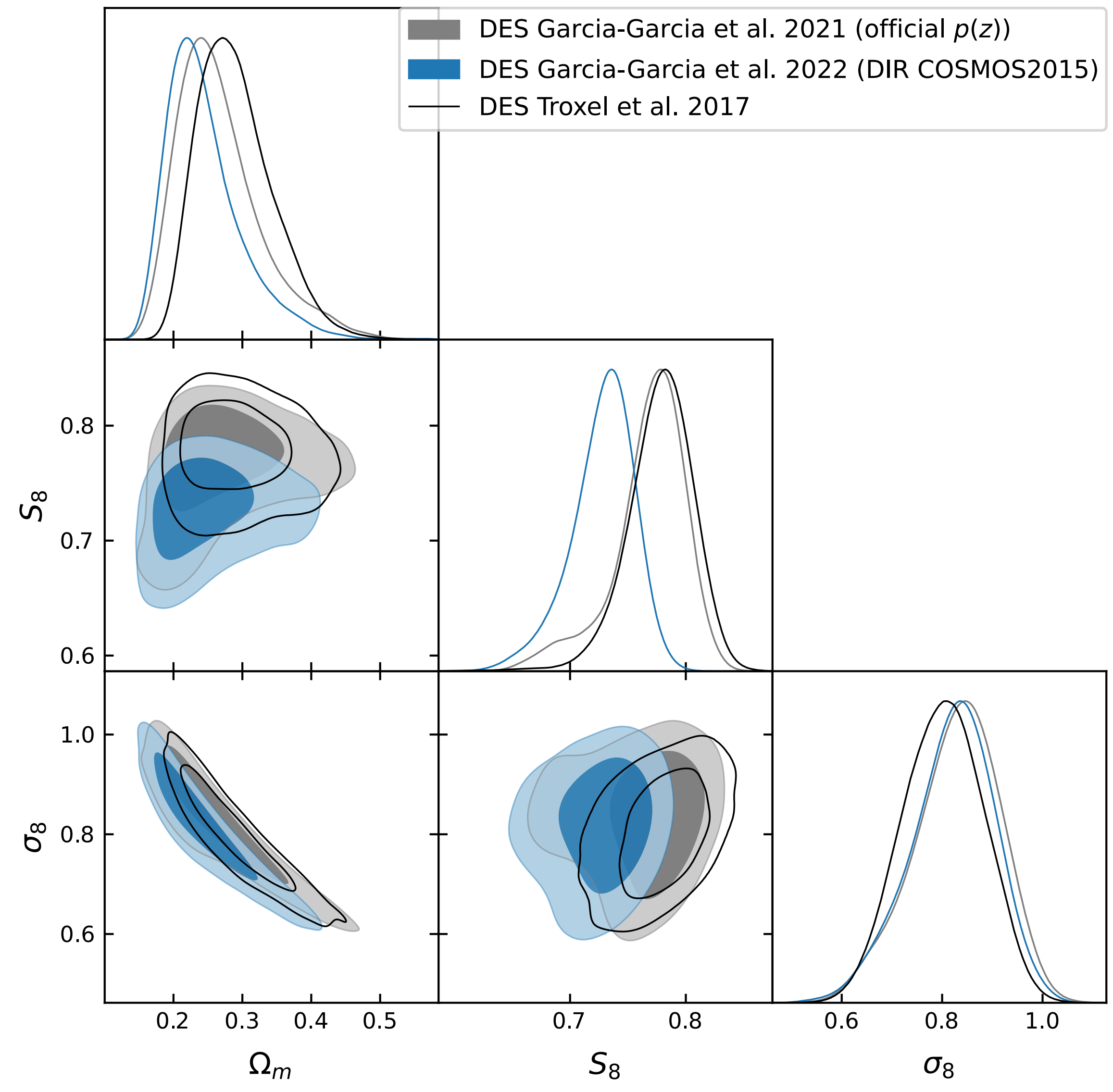


Ruiz-Zapatero et al 2021

Must Do Better!

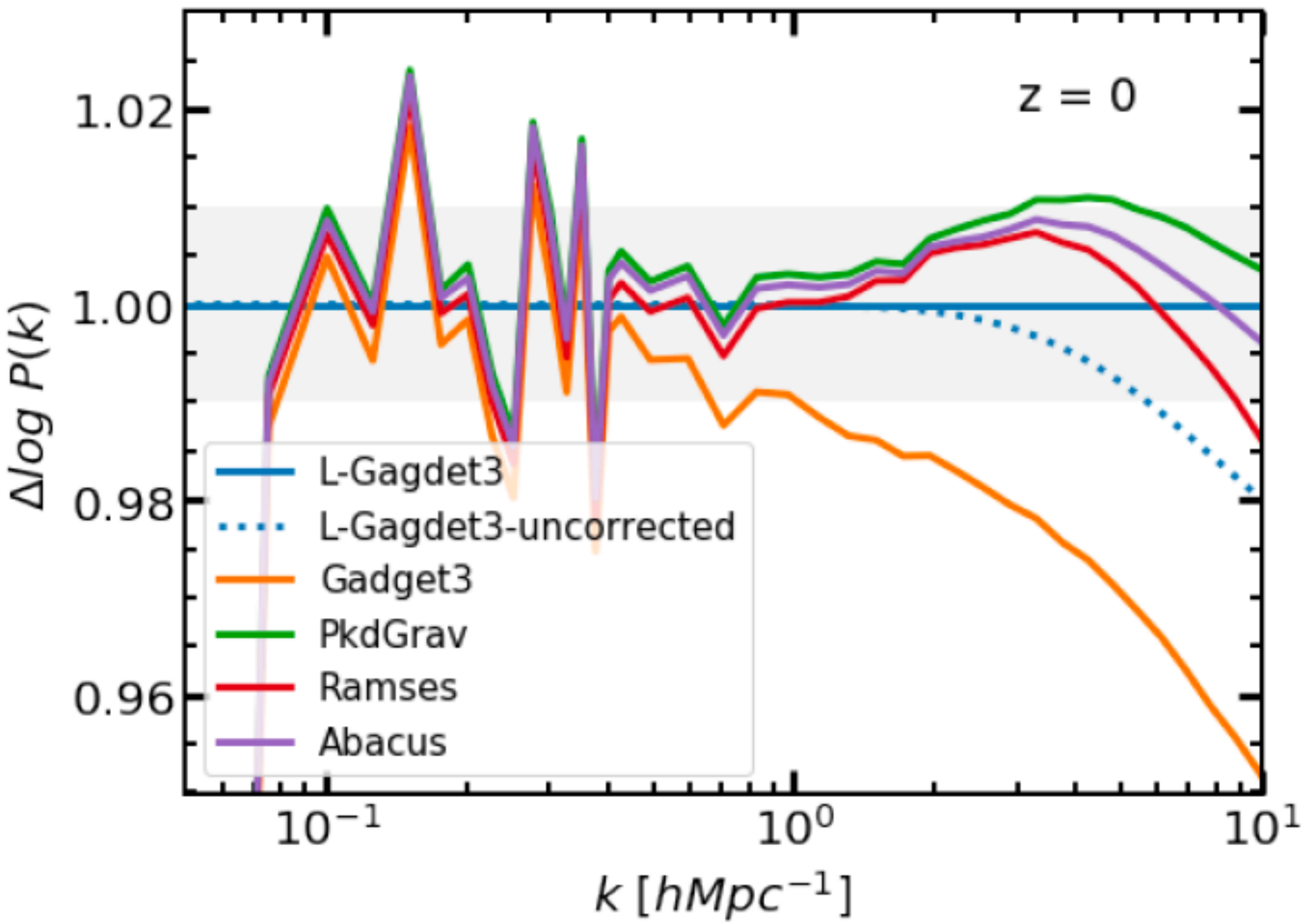
Systematics…





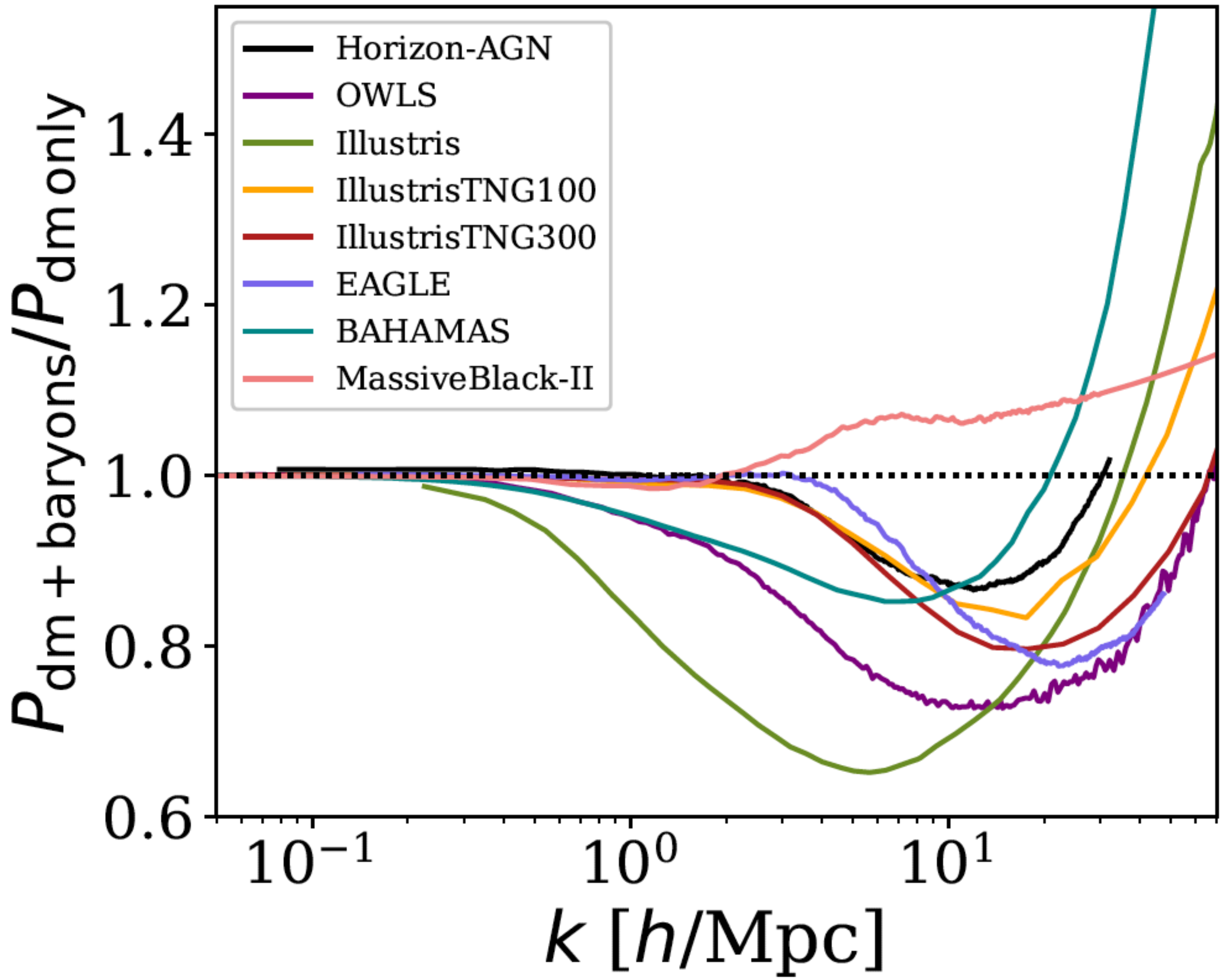
Garcia Garcia et al 2022

Non-linearities…



Angulo et al 2020

Baryons…



Chisari et al 2019

We need models/methods which are

- Comprehensive (include all effects)
- Precise
- Accurate
- Fast
- etc…

Methodological Revolution?

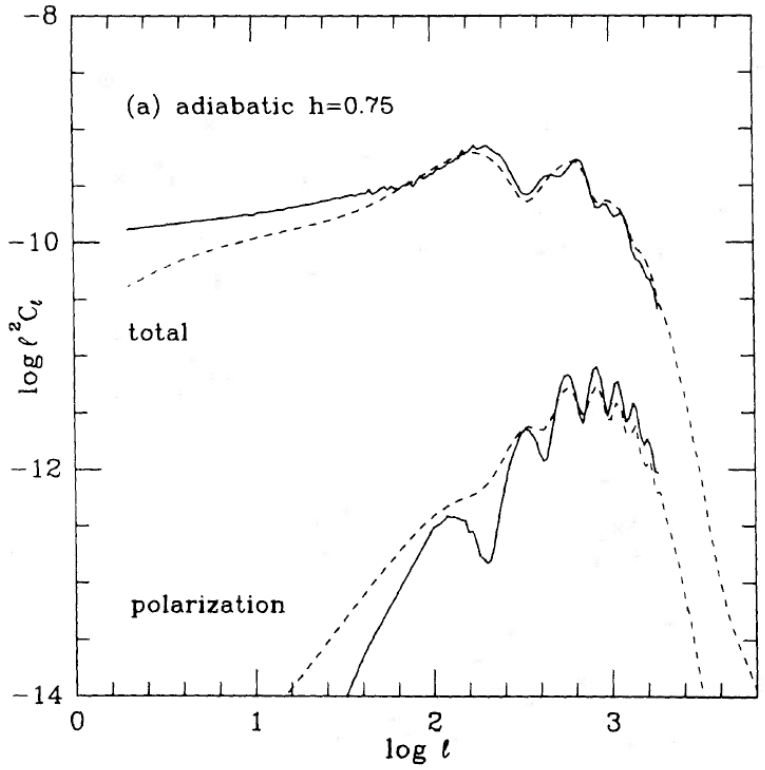
Now: The rise of machine learning …

- Neural Networks
- Gaussian Processes
- Reinforcement Learning
- Generative Adversarial Networks
- etc…

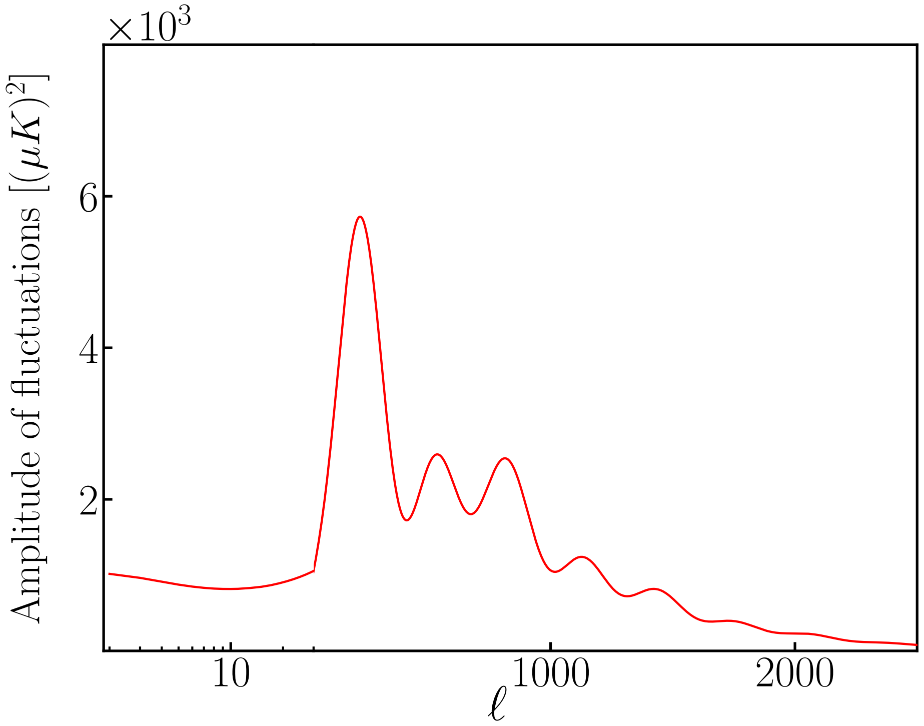
… and advanced numerical/statistical methods.

- fast PDE solvers
- GPUs
- Differentiable methods
- Samplers
- Higher order stats
- etc…

Models: the linear Universe



Bond & Efstathiou 1987

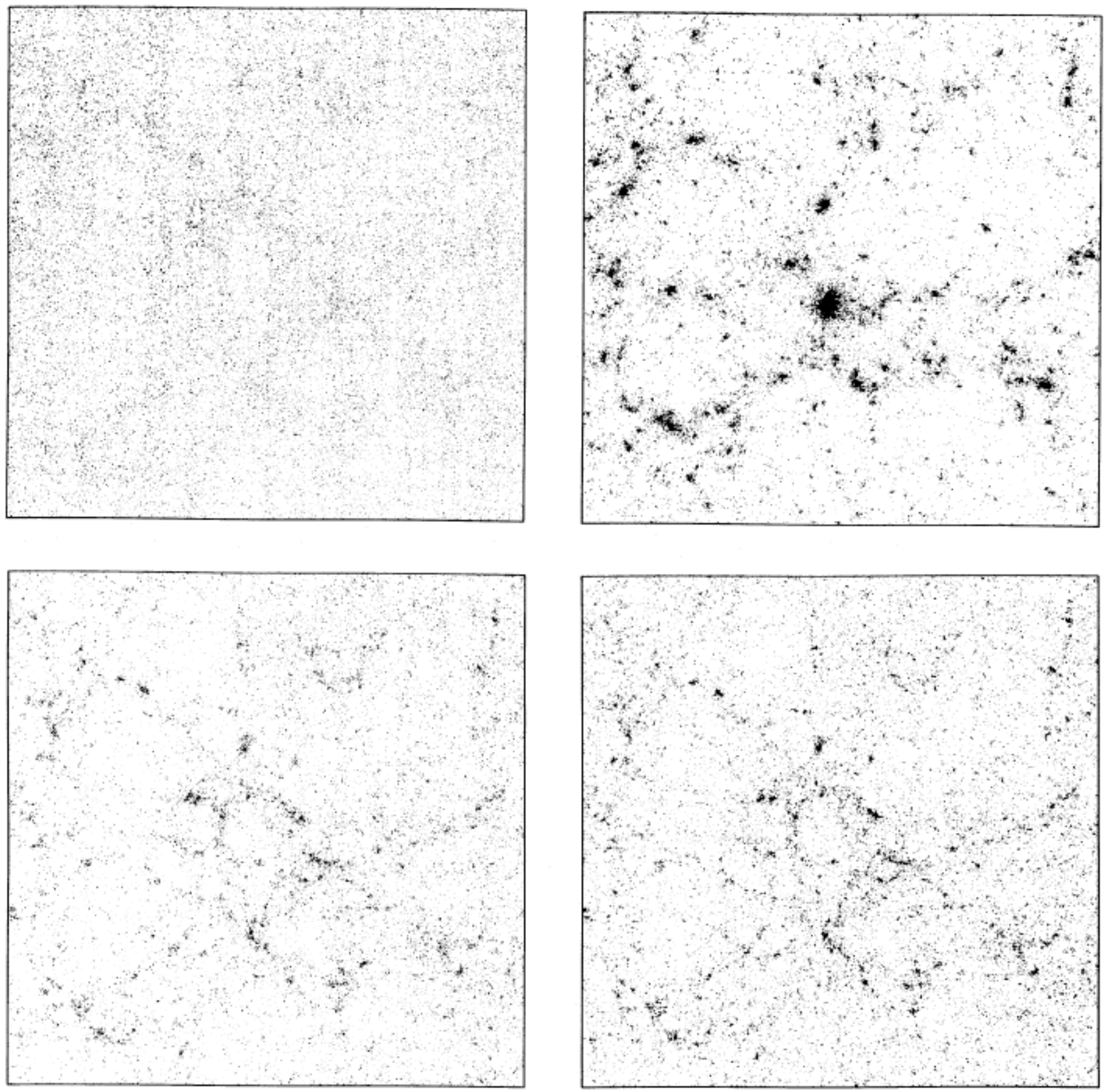


seconds on a laptop

10 hours on a microvax in 1987

- CLASS
- CAMB
- EFTCAMB
- HiCLASS

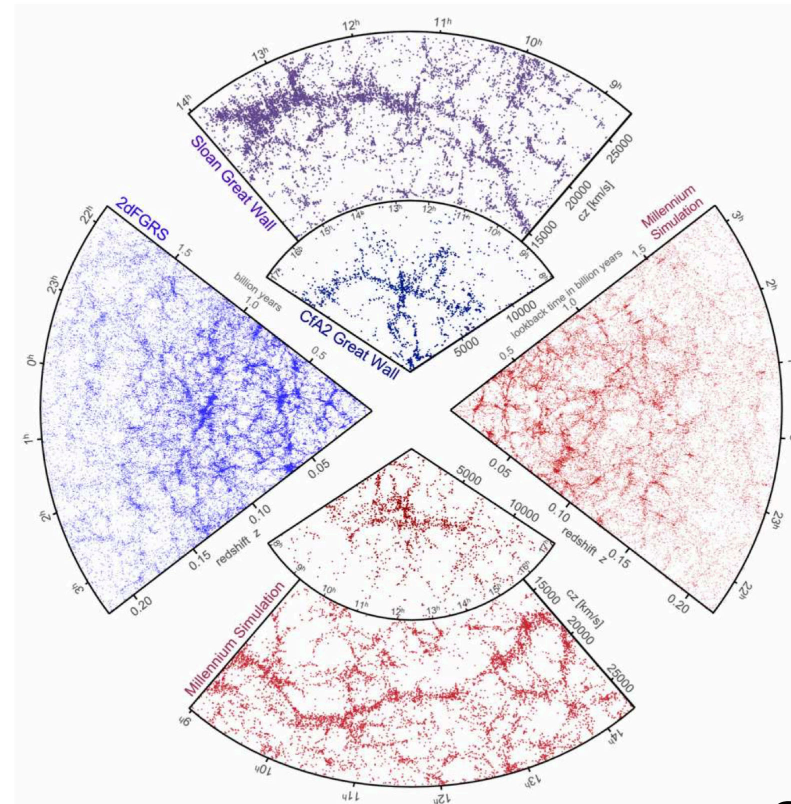
Models: the non-linear Universe



DEFW 1985

Simulating the Universe

- GreeM 0910.0121
- HACC 1411.3396
- 2HOT 1310.4502
- Gadget 1203.3216
- Ramses astro-ph/0111367
- ENZO 1307.2265
- Arepo 1503.00562
- PKDGRAV3 1609.08621



Springel et al 2006

Leading to …

- Aemulus
- Millenium XXL
- Illustrious TNG
- Horizon
- OWLS and Cosmo-OWLS

- CAMELS
- QUIJOTE
- BAHAMAS
- EAGLE
- Tian-ν
- FLAMINGO

Analytical Models

- EFT of LSS 2110.07539
- LPT bias 2012.04636
- HEFT 1910.07097
- HZPT 2104.10676
- EHM 2004.09515
- Baryonification 1510.06034
- SP(k) 2305.09710
- Kinetic Field Theory 1905.01179

Emulators or Surrogates

Case 1: Cold dark matter, adiabatic fluctuations:

$$T_{\mathrm{CDM,\,ad,}\,X}(k) = \frac{\ln(1+2.34q)}{2.34q}\left[1+3.89q+(16.1q)^2+(5.46q)^3+(6.71q)^4\right]^{-1/4}\,,$$

$$q \equiv \frac{k\theta^{1/2}}{(\Omega_X h^2\ \mathrm{Mpc}^{-1})}\,. \qquad \text{(G3)}$$

BBKS 1986

The transfer function is written as a sum of the baryon and cold dark matter contributions at the drag epoch

$$T(k) = \frac{\Omega_b}{\Omega_0}T_b(k) + \frac{\Omega_c}{\Omega_0}T_c(k)\,. \qquad (8)$$

The CDM transfer function can be solved exactly in terms of hypergeometric functions that are more conveniently approximated by the following form:

$$T_c \to \alpha_c \frac{\ln 1.8\beta_c q}{14.2q^2}\,, \qquad (9)$$

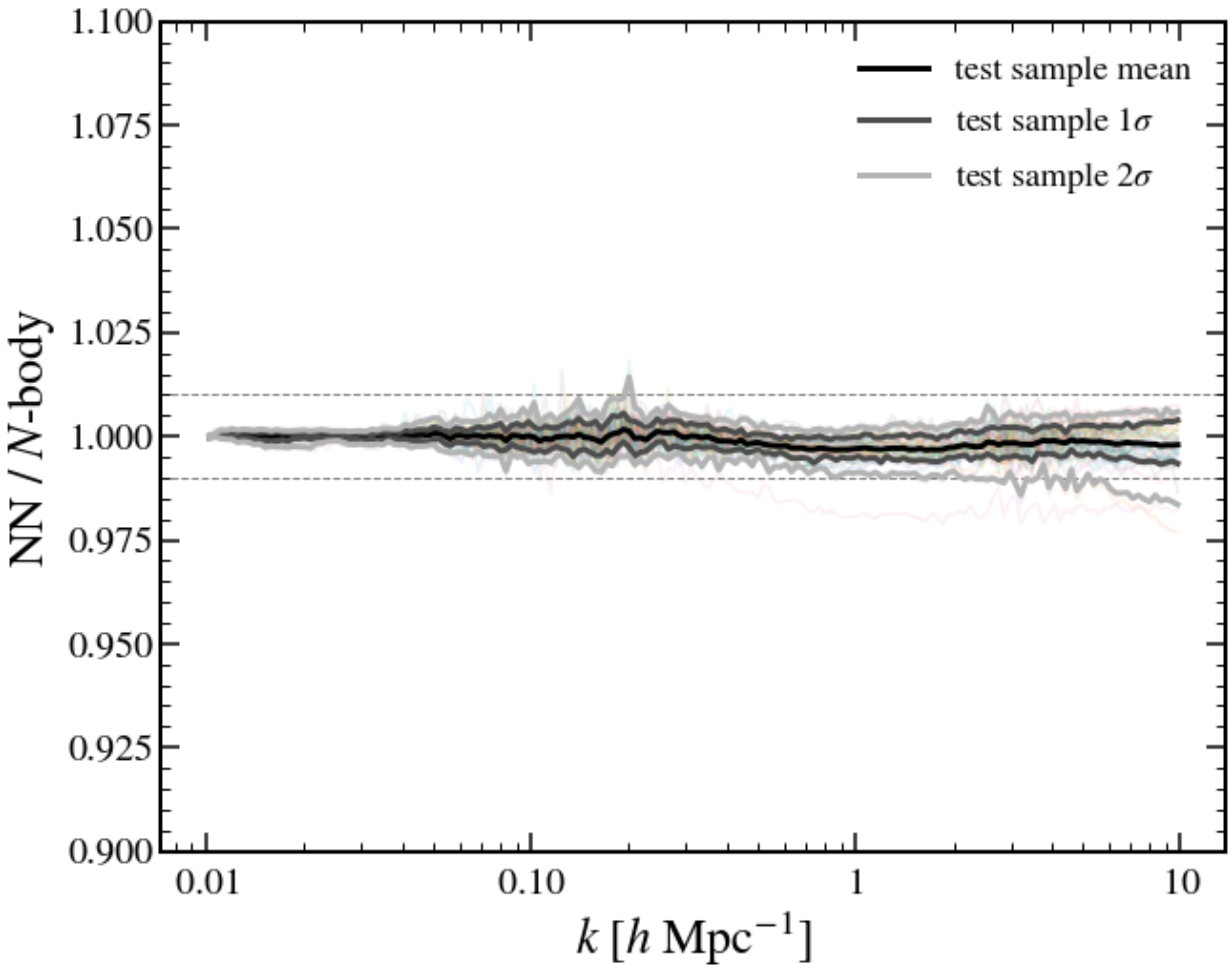
$$q = \left(\frac{k}{\mathrm{Mpc}^{-1}}\right)\Theta_{2.7}^2(\Omega_0 h^2)^{-1} = \frac{k}{13.41k_{\mathrm{eq}}}\,, \qquad (10)$$

where α_c and β_c are fitted by

$$\begin{aligned}
\alpha_c &= a_1^{-\Omega_b/\Omega_0} a_2^{-(\Omega_b/\Omega_0)^3}\,,\\
a_1 &= (46.9\Omega_0 h^2)^{0.670}\left[1+(32.1\Omega_0 h^2)^{-0.532}\right],\\
a_2 &= (12.0\Omega_0 h^2)^{0.424}\left[1+(45.0\Omega_0 h^2)^{-0.582}\right], \qquad (11)\\
\beta_c^{-1} &= 1 + b_1\left[(\Omega_c/\Omega_0)^{b_2} - 1\right],\\
b_1 &= 0.944\left[1+(458\Omega_0 h^2)^{-0.708}\right]^{-1},\\
b_2 &= (0.395\Omega_0 h^2)^{-0.0266}\,. \qquad (12)
\end{aligned}$$

Eisenstein and Hu 1998

Emulators or Surrogates



Zennaro et al 2023

An explosion of Emulators or Surrogates

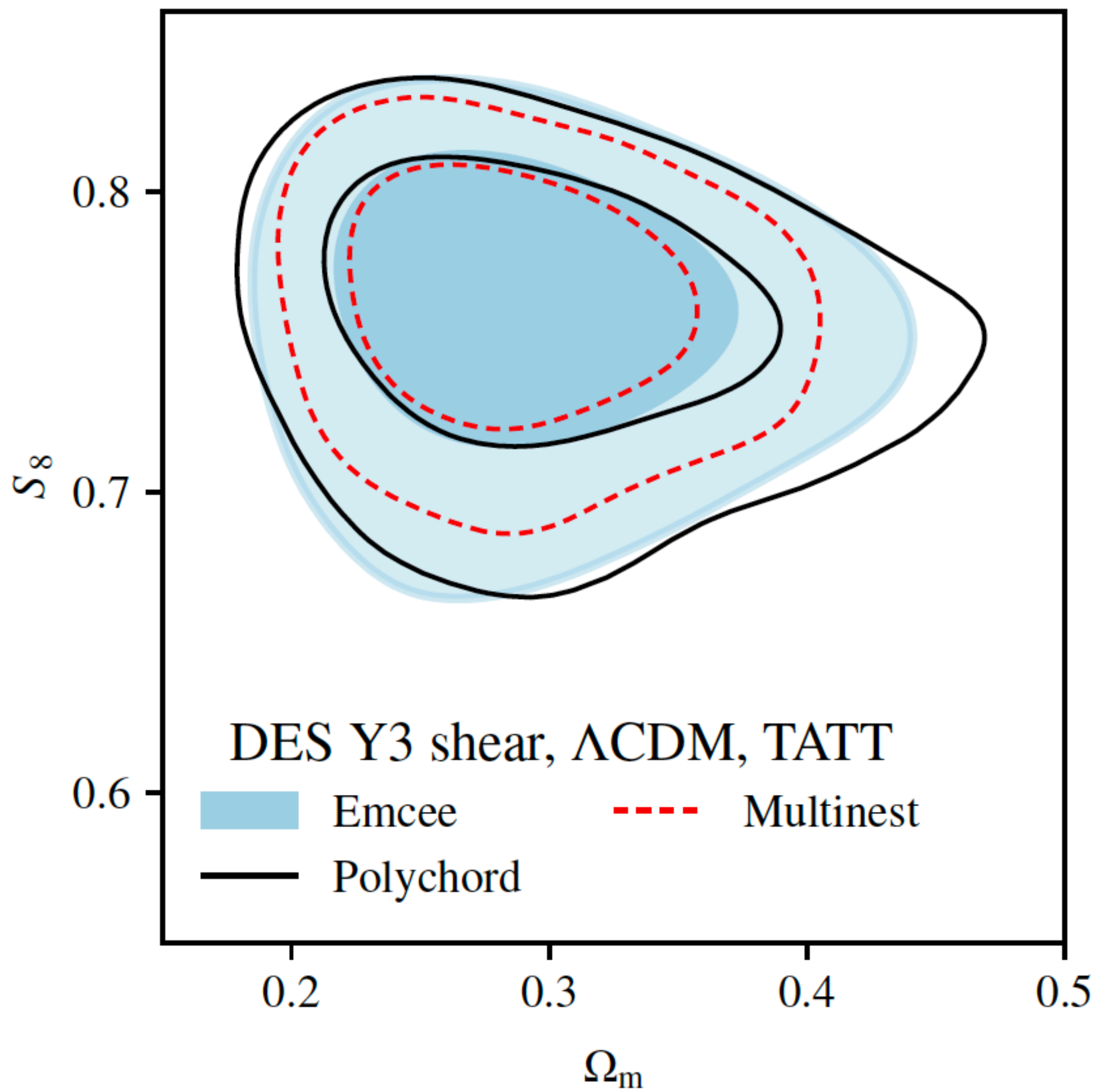
- BACCOEmulator (NN) 2004.06245
- EuclidEmu (PCE) 2010.11288
- PKEmu (GP) 2105.02256
- CosmoPower (NN) 2106.03846
- CAPSE & EFFORT (julia, NN) 2206.14208
- MiraTitan (GP) 1508.02654
- Aemulus (GP and PCE) 1804.05867
- COMET (GP) 2208.01070
- BCEmu (NN) 2108.08863
- MFEmu (GP) 2306.03144

GP - Gaussian Processes
NN - Neural Nets
PCE - Polynomial Chaos Expansion

Samplers

- CosmoMC astro-ph/0205436
- MontePython 1804.07261
- emcee 1202.3665
- Cosmosis 1409.3409
- Cobaya 2005.0529
- Polychord 1506.0017
- CosmoNest astro-ph/0608184
- MultiNest 0809.3437
- CosmoHammer 1206516
- NUTS 1111.4246
- HMC 1701.0243
- MCHMC 2212.08549
- Langevin-HMC 2205.14240
- Black-BOX VI 1601.00670
- VI 1603.00788
- ADVI 2108.03782
- MUSE 2112.09354

Samplers …

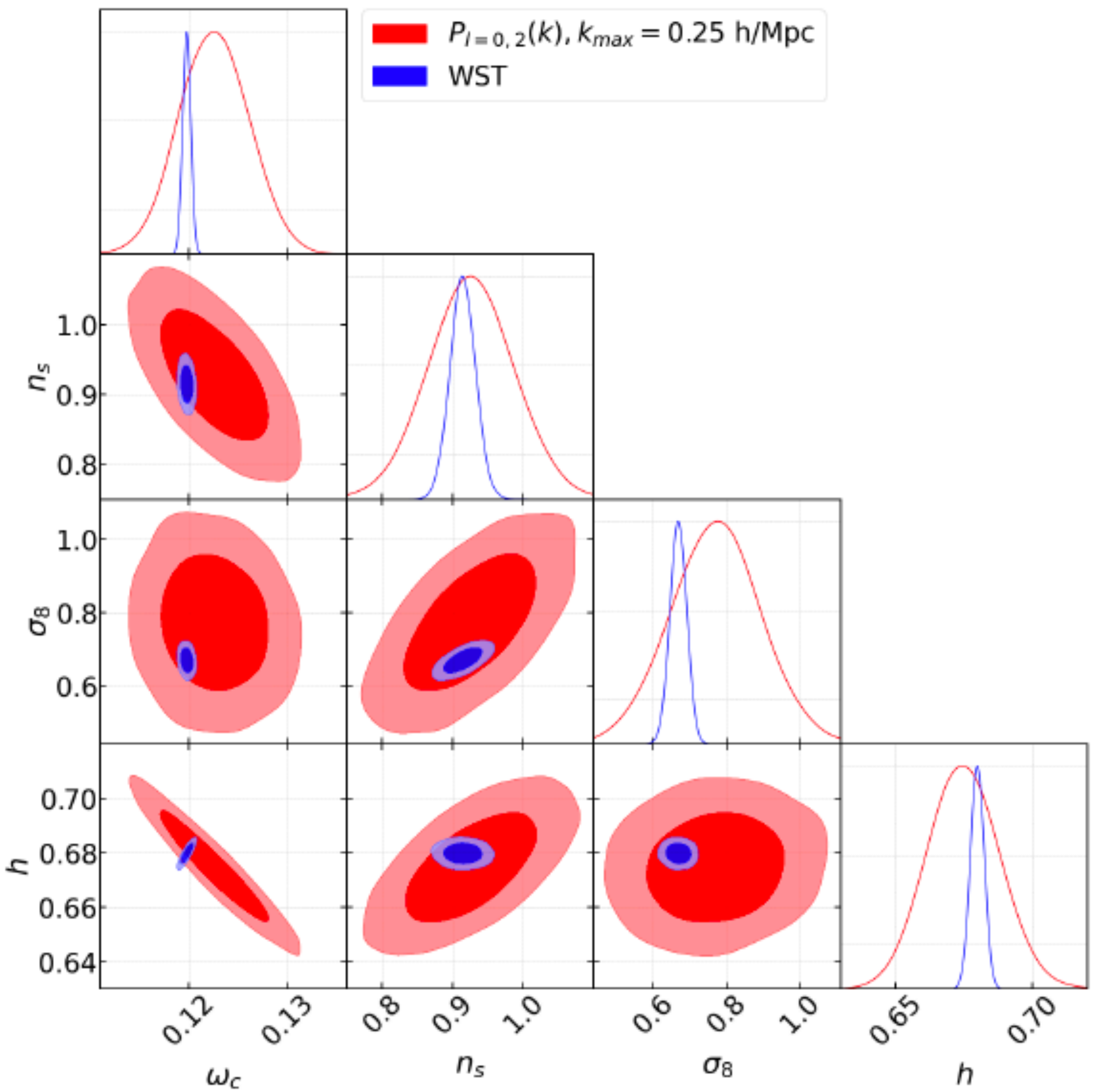


KIDS + DES 2023

Other Statistics

- Bispectrum
- Trispectrum
- Peaks
- Edgeworth Expansion
- Cumulants
- Minkowski functionals
- Wavelet Transform Coefficients
- etc

Other Statistics



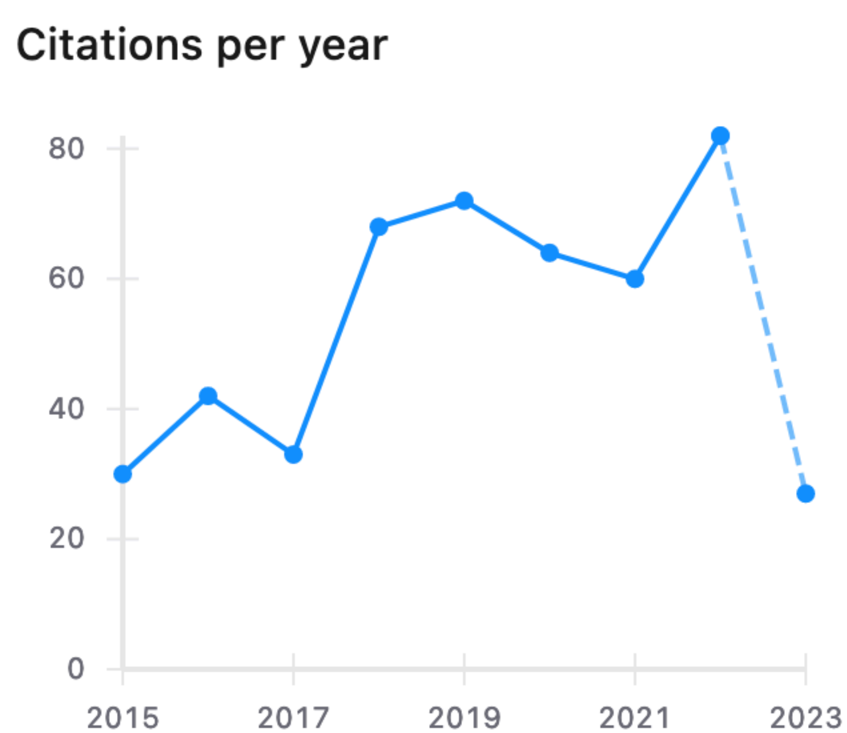
Valogiannis & Dvorkin 2022

Non Gaussian Optimism!

Cosmological Collider Physics

Nima Arkani-Hamed and Juan Maldacena

Institute for Advanced Study, Princeton, NJ 08540, USA



478 citations

Optimism tempered with pragmatism…

Example: target $f_{NL} \sim \mathcal{O}(1)$

History:

CMB $\begin{cases} \bullet \text{ MAXIMA: } \sigma(f_{NL}) \sim 10^3 & \text{astro-ph/0107588} \\ \bullet \text{ WMAP: } \sigma(f_{NL}) \sim 10^2 & \text{1005.5568} \\ \bullet \text{ Planck: } \sigma(f_{NL}) \sim 10^1 & \text{1905.0569} \end{cases}$

but…

LSS $\begin{cases} \bullet \text{ WMAP: } \sigma(f_{NL}) \sim 10^2 & \text{1005.5568} \\ \bullet \text{ BOSS: } \sigma(f_{NL}) \sim 10^2 & \text{1905.0569} \end{cases}$

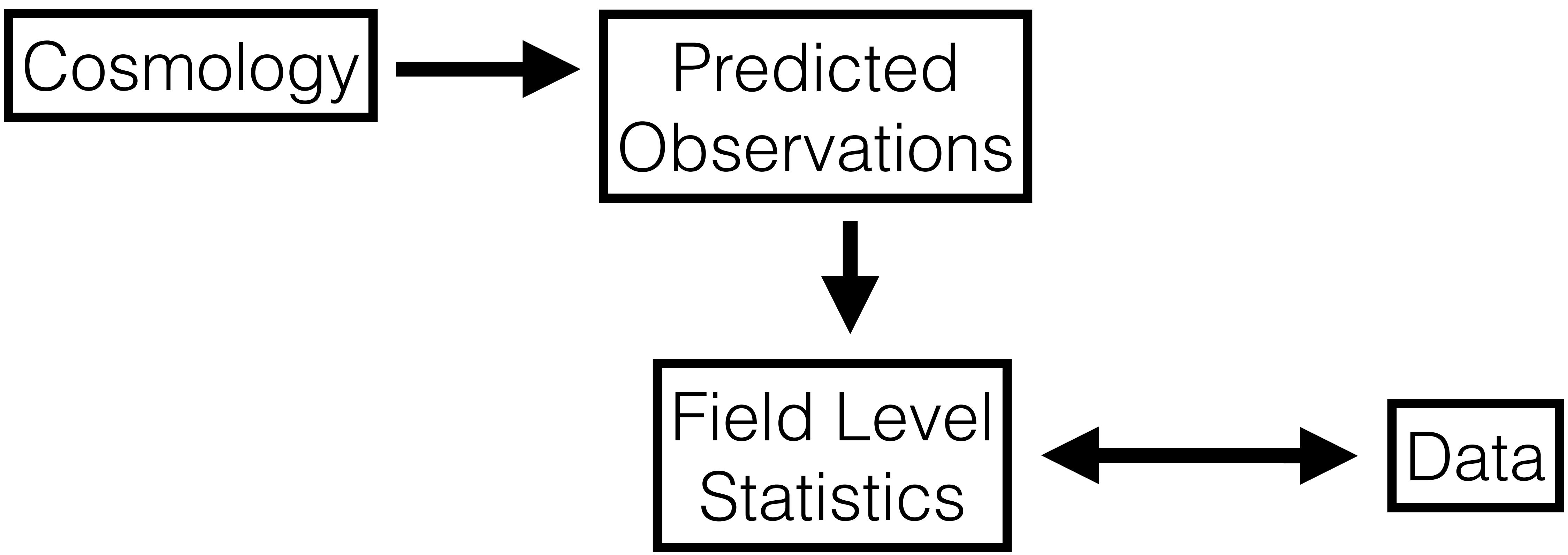
Differentiable Methods

- JAXCOSMO 2302.05163
- CosmoPower-JAX 2305.06347
- Limberjack.jl
- CMBLensing.jl
- Bolt.jl

Using

- Turing (in julia)
- Stan (in C)
- NumPyro (in JAX)

Bayesian forward modelling and Field Level Inference



Statistics pixel by pixel can access all information.

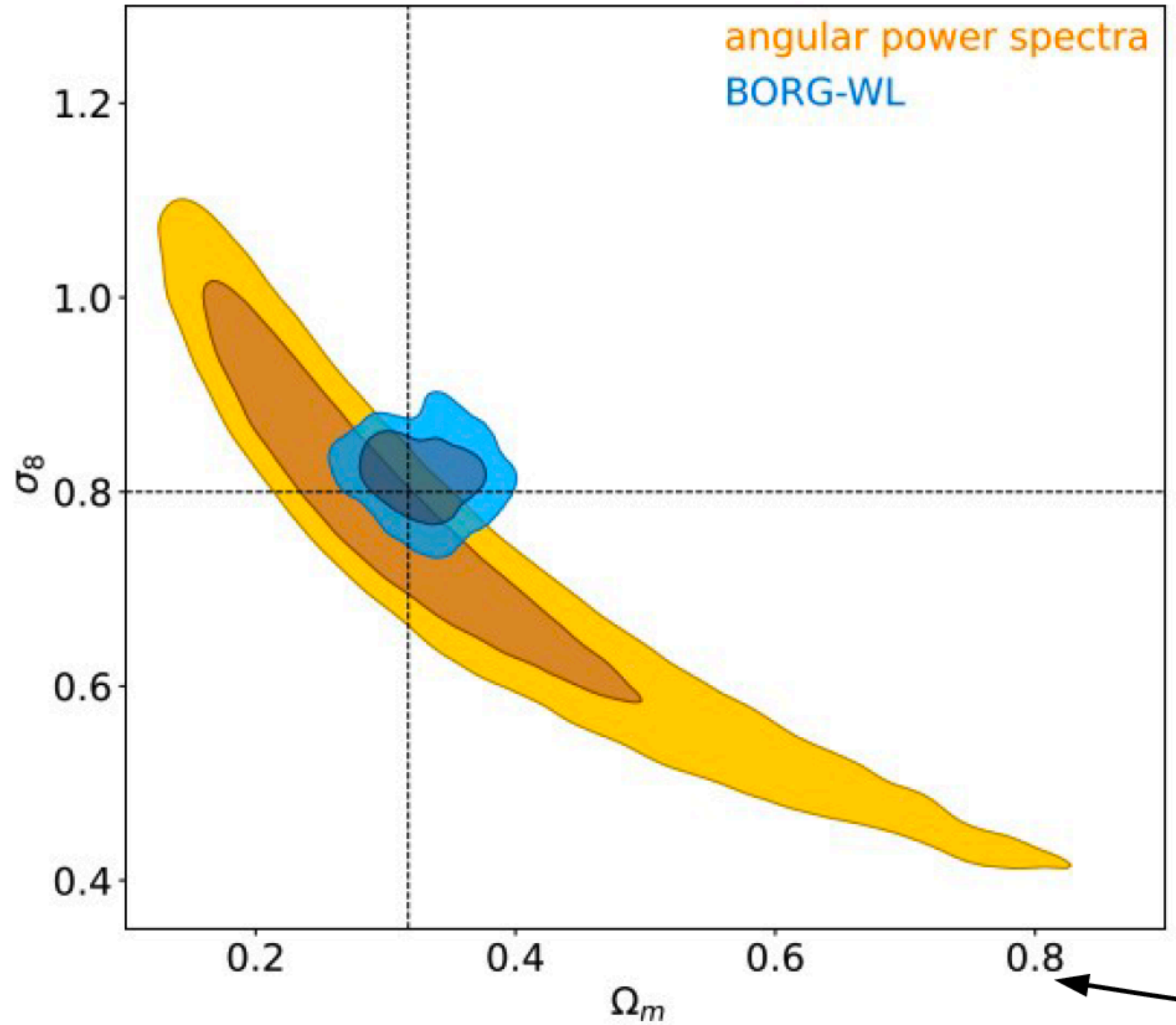
Probabilistic forward model → Simulations constrained by data

Field Level Inference

- Aquila 0911.2496
- Elucid 1407.3451
- Almanac 2210.13260
- LEFTfield 2303.09876
- CCA 2302.14101
- KARMMA 2105.14699
- DLPosterior 2201.05561
- FGPA 2305.10428
- FORECAST 2305.19166

Field Level Inference

marginal error
3 times smaller



(1 Gpc/h)^2, 4.5 Gpc/h
(16 deg)^2, 13 arcmin
LPT

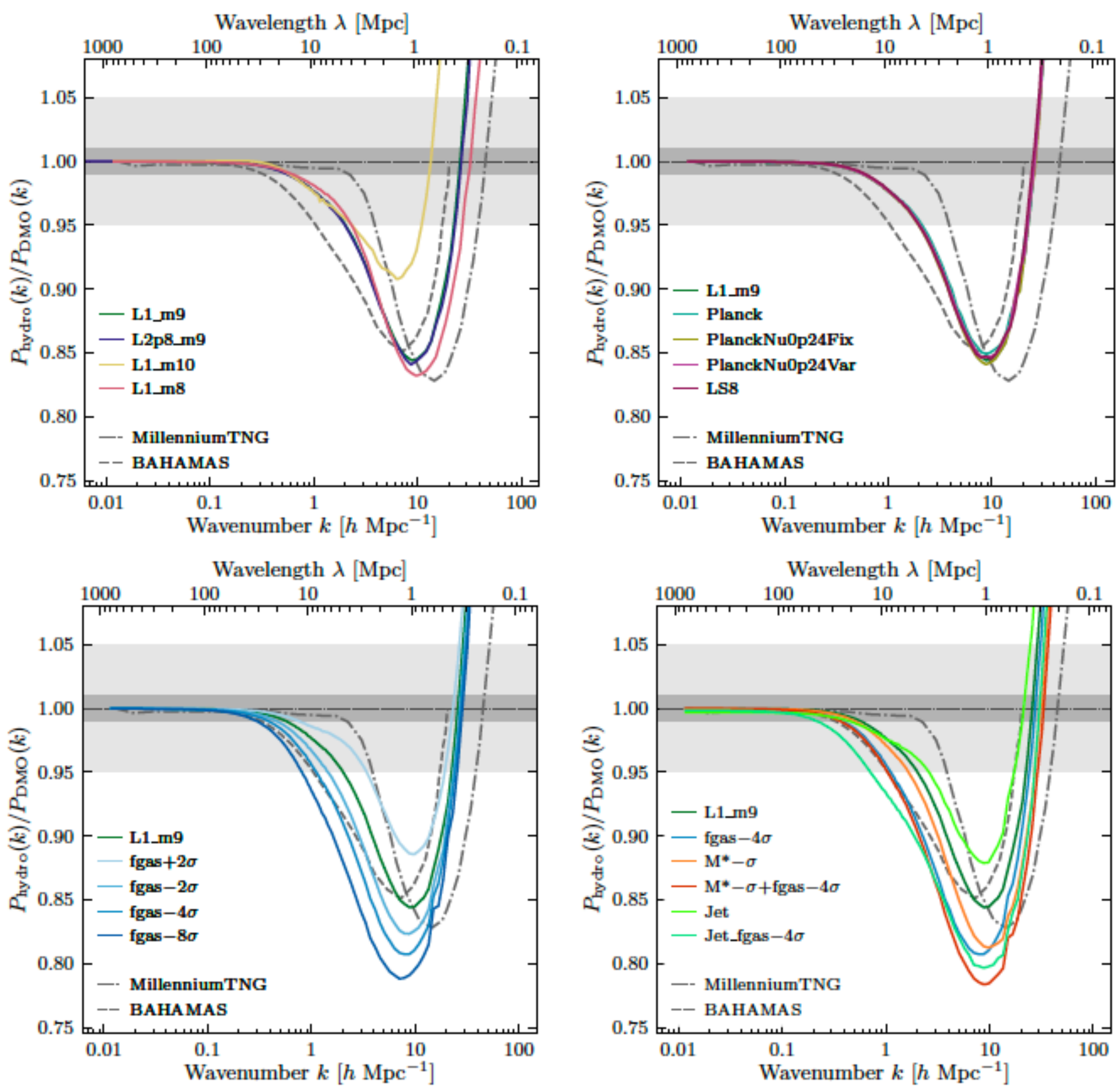
marginal error
5 times smaller

Field-level approach **lifts degeneracy** by extracting more information from the data

Porqueres+2023

Questions for the future

Is there a limit to how deep into the non-linear/baryonic regime we can go? Should we cut our losses?



Schaye et al 2023

Questions for the future

Is there a limit to how deep into the non-linear/baryonic regime we can go? Should we cut our losses?

How important are analytic methods versus emulators?

Should we all aim for differentiable methods?

Can we already see a clear winner for ML methods?

How important are field level methods versus summary statistics?

Should we start using new languages (Julia) or architectures?

Special Section: **HOW TO HAVE FUN OUTDOORS**

MECHANIX ILLUSTRATED

THE HOW-TO-DO MAGAZINE

Best Buy
25c
JUNE