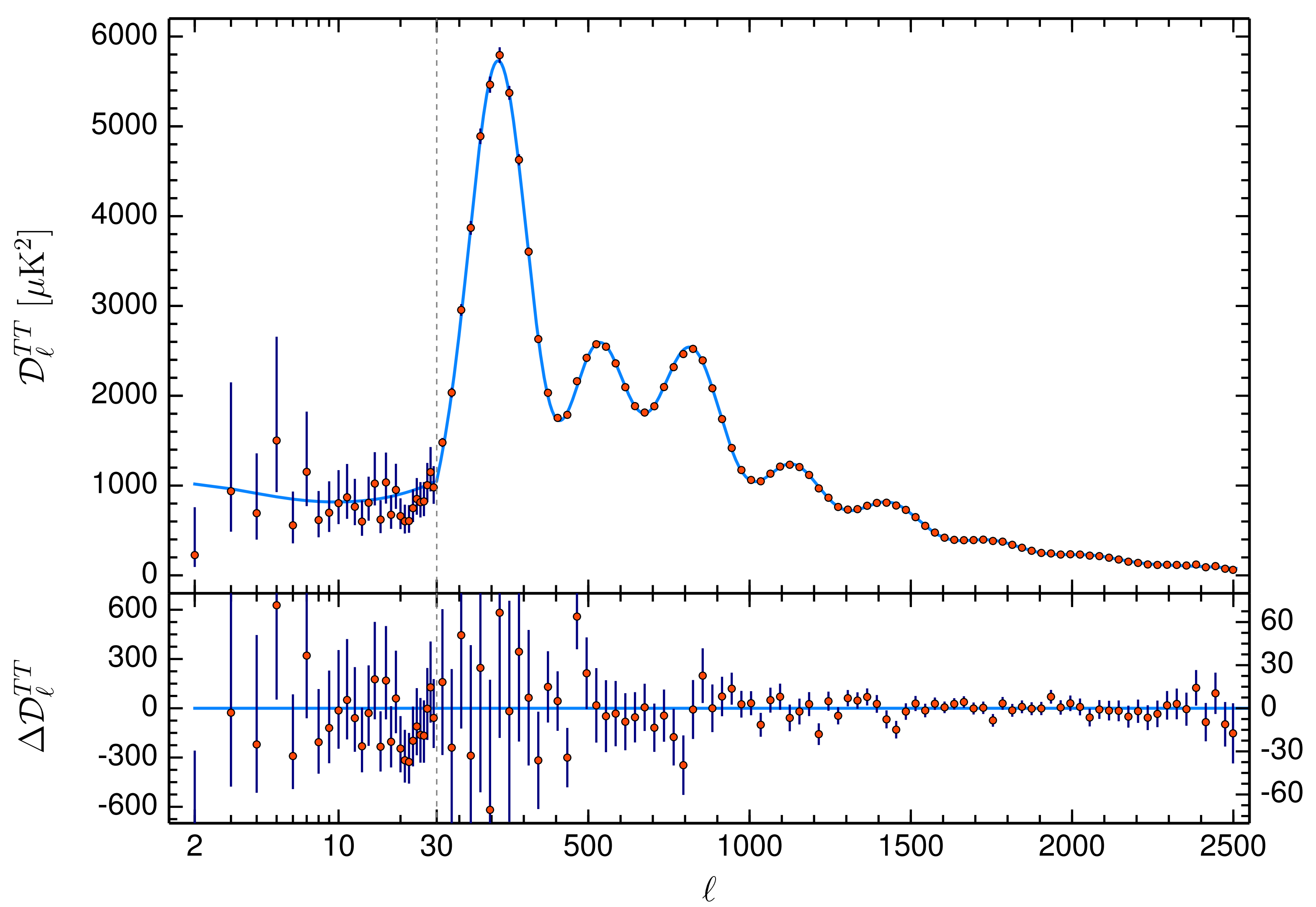
Tensions, DESI, and new cosmological physics

Based on Cortês & Liddle, arXiv:2309.03286 (MNRAS published) and arXiv:2404.08056

instituto de astrofísica e ciências do espaço

The standard six-parameter cosmological model is extraordinarily successful and gives a precision description of our Universe.



Parameter	TT,TE,EE+lowE+lensing 68% limits
$\Omega_b h^2$	0.02237 ± 0.00015
$\Omega_c h^2$	0.1200 ± 0.0012
$100\theta_{MC}$	1.04092 ± 0.00031
τ	0.0544 ± 0.0073
$\ln(10^{10} A_s)$	3.044 ± 0.014
n_s	0.9649 ± 0.0042

Planck 2018 temperature power spectrum ………… and parameter constraints.

Dreams of new physics

LambdaCDM has been the standard cosmological model for two decades. We cosmologists are all desperate to find something new that goes beyond it.

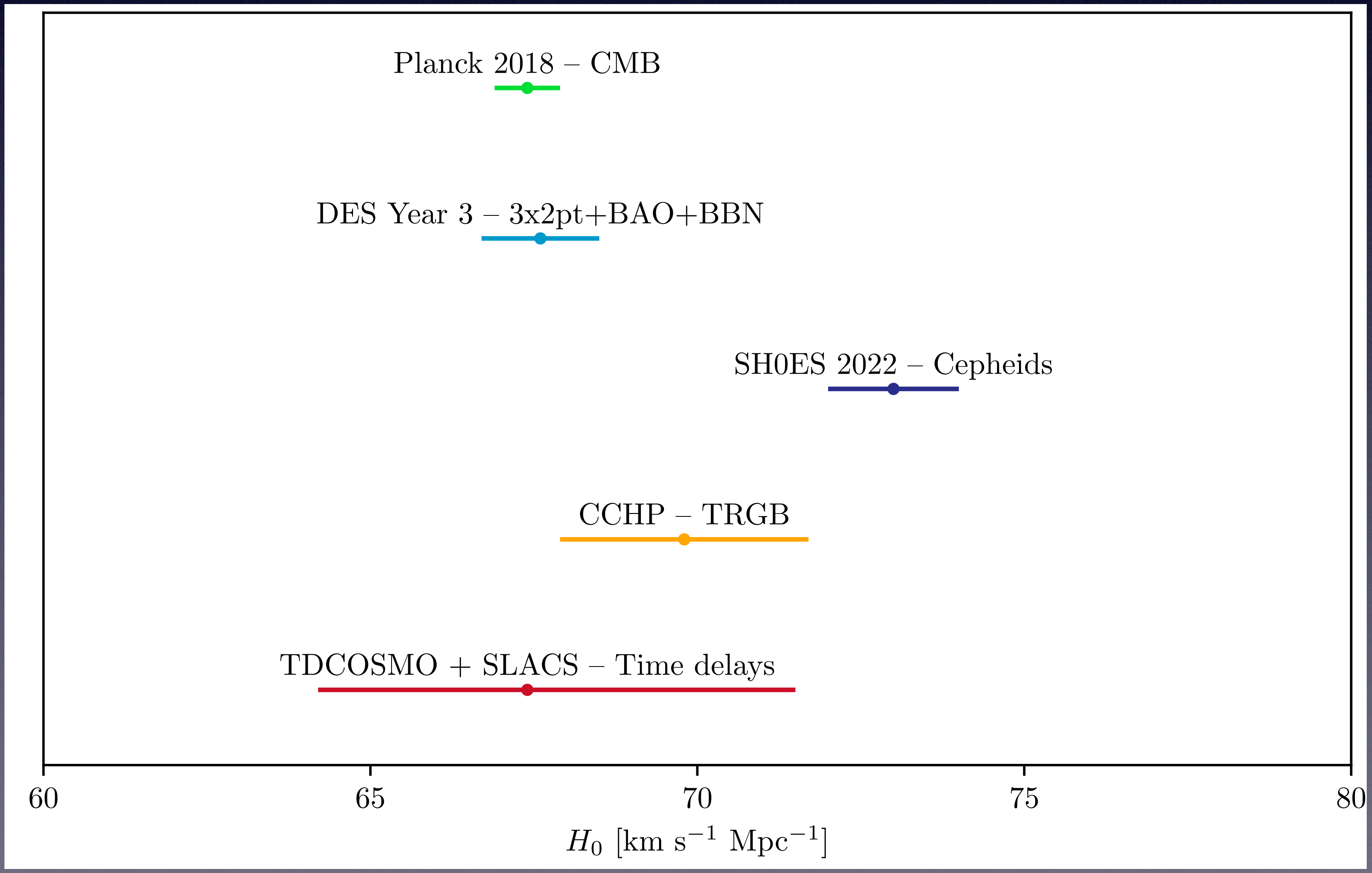
- Tensions between parameters determined by different methods, especially the Hubble tension [`5 sigma′].
- `Tantalizing suggestion′ of evolving dark energy claimed by the DESI collaboration (2024)  [`3 to 4 sigma′].

Part I: On dataset tensions and signatures of new cosmological physics

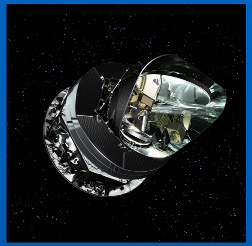
Marina Cortês and Andrew R Liddle,
arXiv:2309.03286, MNRAS 531 (2024) L52

Dataset tensions

- Tensions amongst datasets are one of the defining characteristics and drivers of current cosmology.
- In the last year, `Hubble tension′ has appeared in the title of 66 papers and in the abstract of 250.



Planck (2018)
H_0 = 67.4 ± 0.5



SH0ES (2022)
H_0 = 73.0 ± 1.0

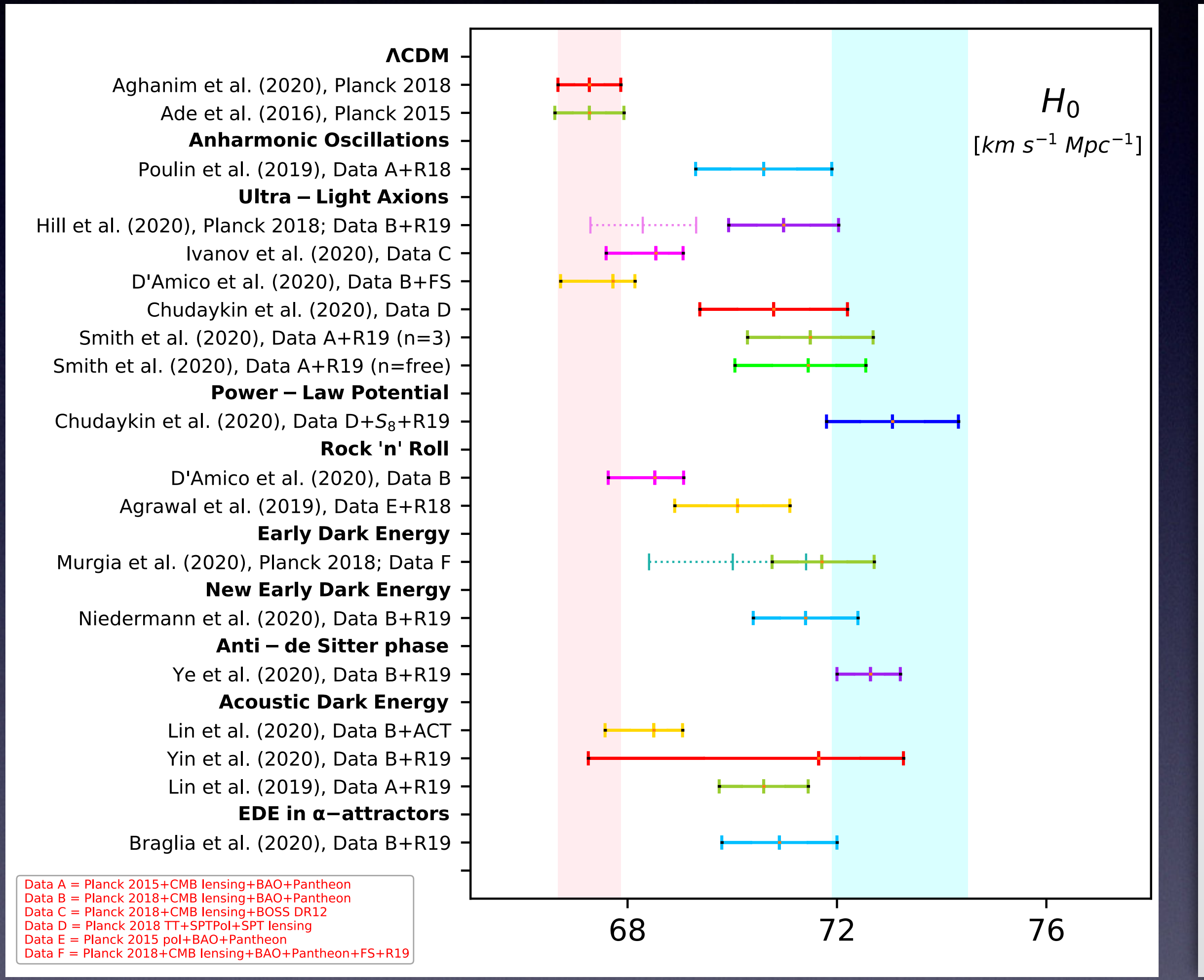
Figure by Pablo Lemos, from Lahav-Liddle, Review of Particle Physics 2023

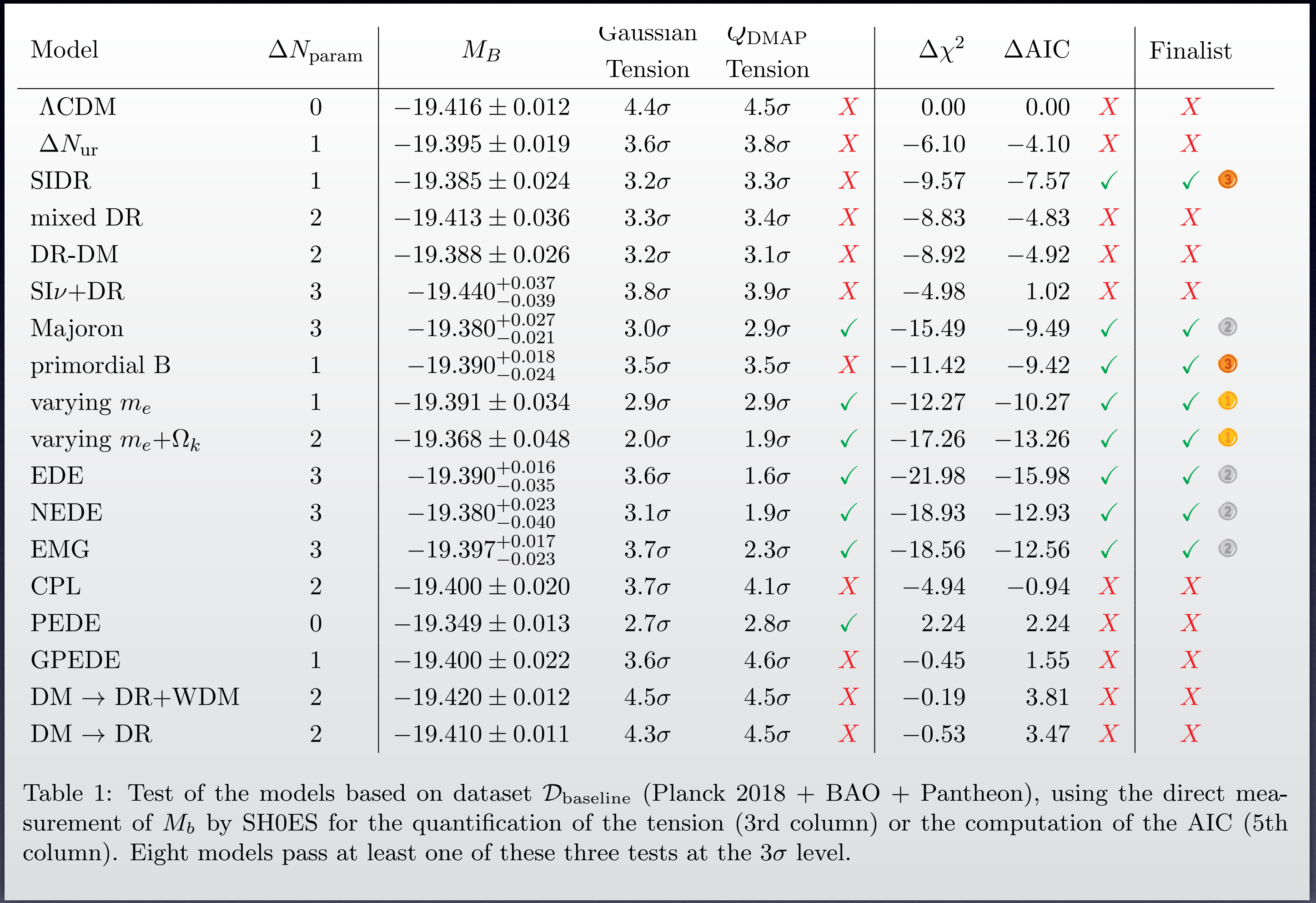
There are numerous reasons why datasets might be in tension.

- Statistical fluke.
- Underreported systematic uncertainties.
- Data analysis pipeline errors.
- Inadequate cosmological model.

Dataset tensions are frequently invoked as signalling new physics. `Alleviating the tension' is a common phrase.

An incredible variety of theoretical models have been examined to see if they can alleviate the Hubble tension. No compelling option has arisen.





Model	$\Delta N_{\rm param}$	M_B	Gaussian Tension	$Q_{\rm DMAP}$ Tension		$\Delta\chi^2$	ΔAIC		Finalist	
ΛCDM	0	-19.416 ± 0.012	4.4σ	4.5σ	X	0.00	0.00	X	X	
$\Delta N_{\rm ur}$	1	-19.395 ± 0.019	3.6σ	3.8σ	X	-6.10	-4.10	X	X	
SIDR	1	-19.385 ± 0.024	3.2σ	3.3σ	X	-9.57	-7.57	✓	✓	③
mixed DR	2	-19.413 ± 0.036	3.3σ	3.4σ	X	-8.83	-4.83	X	X	
DR-DM	2	-19.388 ± 0.026	3.2σ	3.1σ	X	-8.92	-4.92	X	X	
SIν+DR	3	$-19.440^{+0.037}_{-0.039}$	3.8σ	3.9σ	X	-4.98	1.02	X	X	
Majoron	3	$-19.380^{+0.027}_{-0.021}$	3.0σ	2.9σ	✓	-15.49	-9.49	✓	✓	②
primordial B	1	$-19.390^{+0.018}_{-0.024}$	3.5σ	3.5σ	X	-11.42	-9.42	✓	✓	③
varying m_e	1	-19.391 ± 0.034	2.9σ	2.9σ	✓	-12.27	-10.27	✓	✓	①
varying m_e+Ω_k	2	-19.368 ± 0.048	2.0σ	1.9σ	✓	-17.26	-13.26	✓	✓	①
EDE	3	$-19.390^{+0.016}_{-0.035}$	3.6σ	1.6σ	✓	-21.98	-15.98	✓	✓	②
NEDE	3	$-19.380^{+0.023}_{-0.040}$	3.1σ	1.9σ	✓	-18.93	-12.93	✓	✓	②
EMG	3	$-19.397^{+0.017}_{-0.023}$	3.7σ	2.3σ	✓	-18.56	-12.56	✓	✓	②
CPL	2	-19.400 ± 0.020	3.7σ	4.1σ	X	-4.94	-0.94	X	X	
PEDE	0	-19.349 ± 0.013	2.7σ	2.8σ	✓	2.24	2.24	X	X	
GPEDE	1	-19.400 ± 0.022	3.6σ	4.6σ	X	-0.45	1.55	X	X	
DM → DR+WDM	2	-19.420 ± 0.012	4.5σ	4.5σ	X	-0.19	3.81	X	X	
DM → DR	2	-19.410 ± 0.011	4.3σ	4.5σ	X	-0.53	3.47	X	X	

Table 1: Test of the models based on dataset $\mathcal{D}_{\rm baseline}$ (Planck 2018 + BAO + Pantheon), using the direct measurement of M_b by SH0ES for the quantification of the tension (3rd column) or the computation of the AIC (5th column). Eight models pass at least one of these three tests at the 3σ level.

Di Valentino et al., arXiv:2103.01183

Schöneberg et al., arXiv:2107.10291

Bayesian tension metric: definition

We are going to focus on a fully Bayesian analysis, though the same issues apply to the many partly Bayesian or non-Bayesian tension metrics that have been defined.

The level of tension depends both on the datasets, labelled A and B, and on the model being assumed, labelled 1,2, etc.

$$R_1^{AB} \equiv \frac{P(D_A, D_B|M_1)}{P(D_A|M_1)P(D_B|M_1)}$$

Marshall-Rajguru-Slosar, arXiv:astro-ph/0412535

It's much more easily interpreted by rewriting using Bayes' theorem:

$$R_1^{AB} = \frac{P(D_A|D_B, M_1)}{P(D_A|M_1)}$$

Bayesian tension metric: uses

$$R_1^{AB} \equiv \frac{P(D_A, D_B|M_1)}{P(D_A|M_1)P(D_B|M_1)}$$

The original idea of the tension metric was to validate that two datasets are consistent with each other before they are combined into a joint analysis.

But more recently the tension metric has been co-opted to a different purpose. If the tension is less under a different model assumption, M_2, ie. $R_2^{AB} > R_1^{AB}$, this is taken as support for Model 2.

The reduced tension is said to signal the new physics encoded in Model 2.

Bayesian model comparison

The correct Bayesian quantity to judge if one model is preferred over another is the posterior model probability ratio taking into account all the data we have.

$$\frac{P(M_1|D_A,D_B)}{P(M_2|D_A,D_B)}$$

This is determined via the relation

$$\frac{P(M_1|D_A,D_B)}{P(M_2|D_A,D_B)} = B_{12}^{AB}\frac{P(M_1)}{P(M_2)} ,$$

where the Bayes' factor

$$B_{12}^{AB} = \frac{P(D_A,D_B|M_1)}{P(D_A,D_B|M_2)}$$

The model likelihoods, also known as the Bayesian evidence, can be computed for example by nested sampling. The prior model probabilities are to be chosen as you wish.

Tension ratio and Bayes factors

Using the tension ratio to diagnose new physics is saying that it can be taken as a proxy for the model probability ratio. Can it?

$$R_1^{AB} \equiv \frac{P(D_A, D_B|M_1)}{P(D_A|M_1)P(D_B|M_1)}$$

$$B_{12}^{AB} = \frac{P(D_A, D_B|M_1)}{P(D_A, D_B|M_2)}$$

$$\rightarrow \quad B_{12}^{AB} = \frac{R_1^{AB}}{R_2^{AB}} \frac{P(D_A|M_1)P(D_B|M_1)}{P(D_A|M_2)P(D_B|M_2)}$$

This can be neatly written as

$$B_{12}^{AB} = \frac{R_1^{AB}}{R_2^{AB}} B_{12}^{A} B_{12}^{B}$$

Clearly, the tension ratio and the Bayes factor are ***not*** the same.

What gives an improved model fit?

$$B_{12}^{AB} = \frac{R_1^{AB}}{R_2^{AB}} B_{12}^{A} B_{12}^{B}$$

- Model 2 could be favoured either because it better fits dataset A, or better fits dataset B, or because it reduces the tension between these datasets.
- The tension ratio only matches the Bayes factor if $B^{A}{}_{12}$ and $B^{B}{}_{12}$ both equal one. Ie, the models fit datasets A and B equally well, yet one model fails when the datasets are combined.
- Reduction of tension on its own does not justify new physics, because the new model may fit one or both datasets less well. Indeed this is likely because extra parameter freedom reduces model predictiveness.

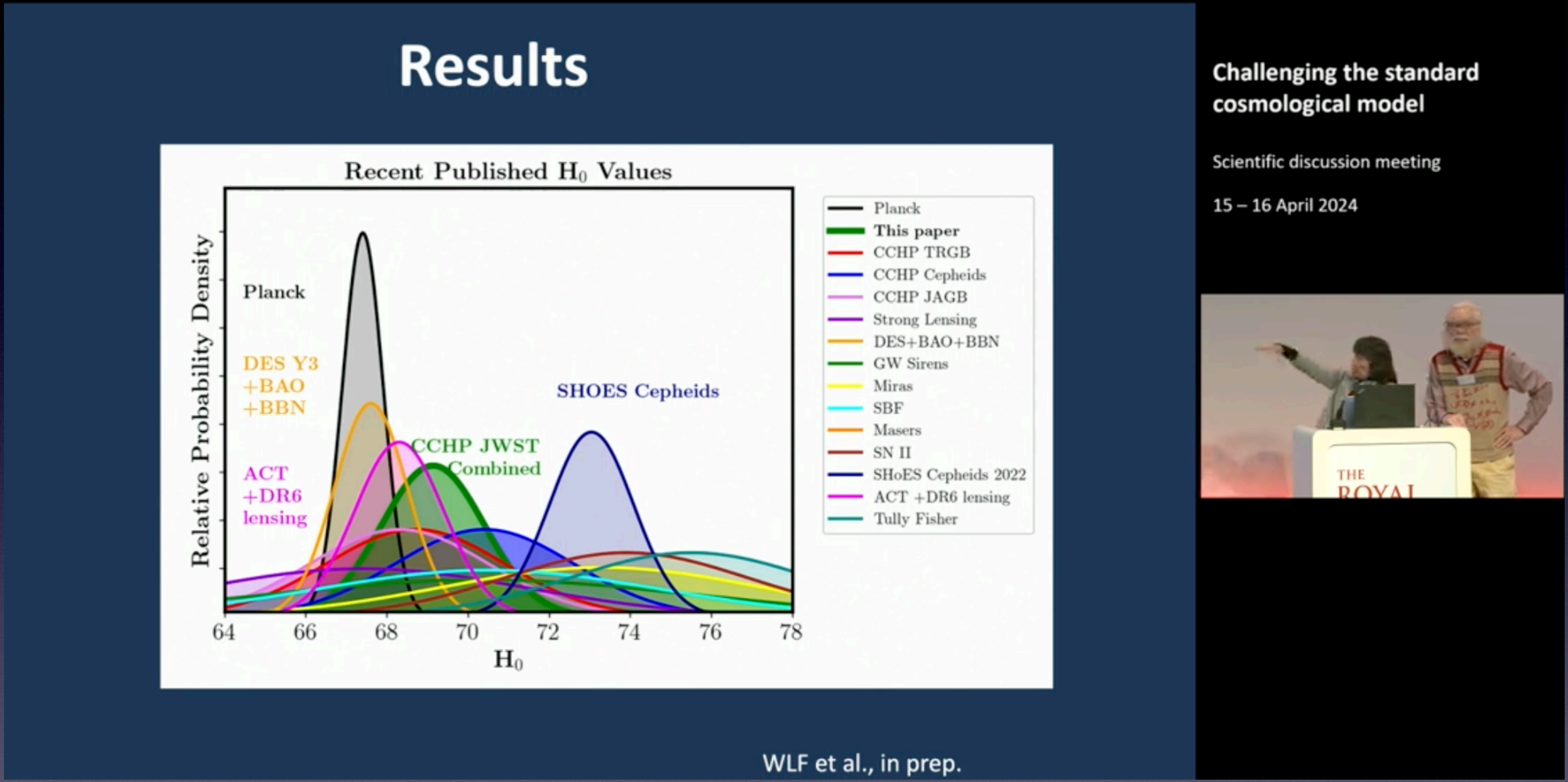
Prior perspective on new physics

- For new physics to show up via tension requires quite a coincidence. The `wrong′ model, here ΛCDM, has to fit the *Planck* data and the SH0ES data as well the new `right′ model, yet fail badly when they are combined.
- This is especially true because of the awesome constraining power of *Planck*: in the 6-dimensional prior parameter space of ΛCDM, the data reduces the allowed volume (posterior versus prior) by a factor of 2×10^{14}!
- Hubble constant probes by contrast compress in a single direction, by a factor of around 20.

Conclusions: Tension as a signature of new physics

1. Any analysis that claims evidence for new physics *solely* on the basis of alleviating dataset tensions should be considered incomplete and suspect.
2. If a new physics model is identified as explaining the tension, it immediately raises a new coincidence that the true parameters of that model were such that the deviations from LambdaCDM were only apparent via tensions.
3. Because of those considerations, we believe unidentified systematic uncertainties in the observations are a more likely source of the tension than new physics.

Stop press: Wendy Freedman, preliminary CCHP results shown at the Royal Society Discussion meeting, April 15th 2024



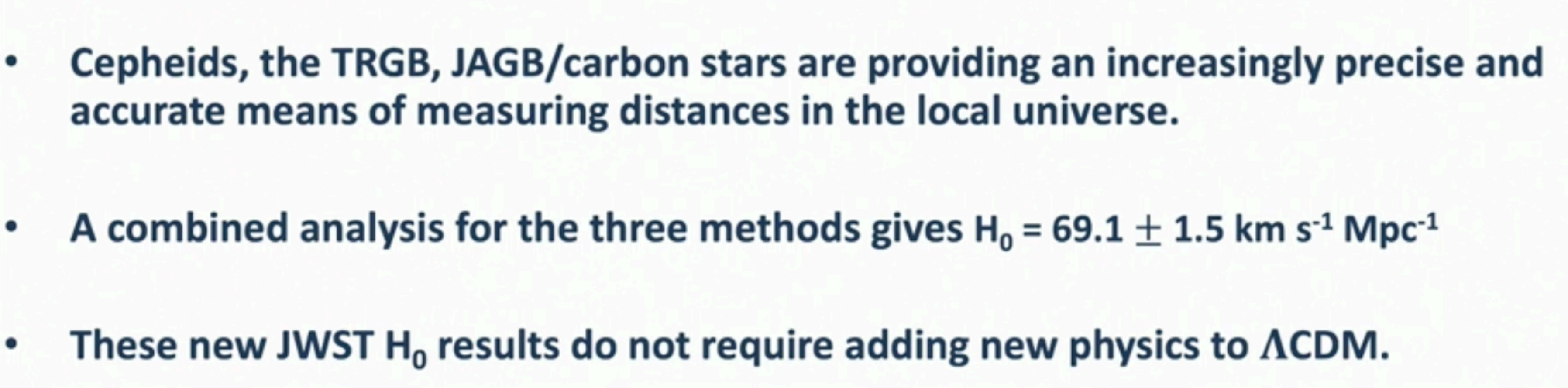
Results

WLF et al., in prep.

Challenging the standard cosmological model

Scientific discussion meeting

15 – 16 April 2024



- **Cepheids, the TRGB, JAGB/carbon stars are providing an increasingly precise and accurate means of measuring distances in the local universe.**
- **A combined analysis for the three methods gives $H_0 = 69.1 \pm 1.5$ km s^{-1} Mpc^{-1}**
- **These new JWST H_0 results do not require adding new physics to ΛCDM.**

Part II: Interpreting DESI's evidence for evolving dark energy

Marina Cortês and Andrew R Liddle,
arXiv:2404.08056

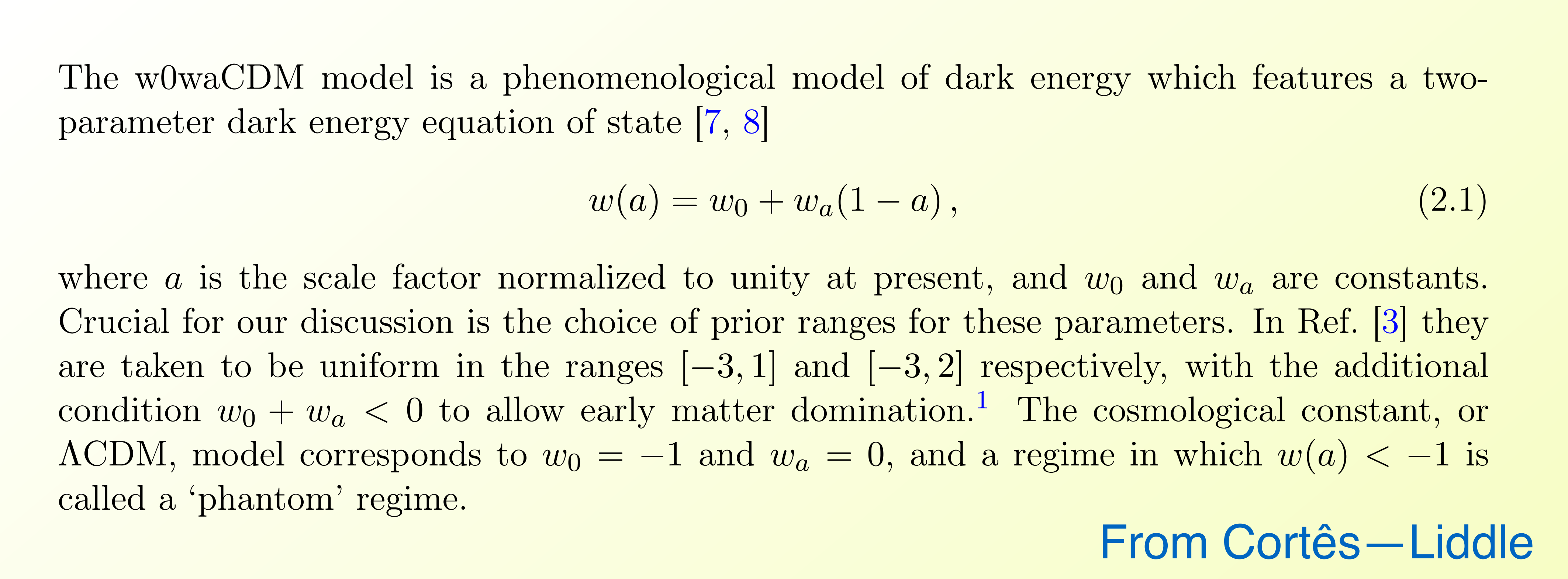
DESI BAO 2024

In April 2024, the DOE-led DESI (Dark Energy Spectroscopic Instrument) survey, the first Stage IV, announced cosmological results utilising their new baryon acoustic oscillation (BAO) observations.

The analysis has been carried out by the best researchers, whom we have come to know for their rigorous internal scrutiny for systematics and probe uncertainties. DESI's is an exemplary dataset.

Combining DESI BAO with Planck CMB anisotropy and Supernovae luminosity-redshift data, they found `tantalizing suggestions' of evolving dark energy, in a phenomenological dark energy model.

DESI BAO 2024



The w0waCDM model is a phenomenological model of dark energy which features a two-parameter dark energy equation of state [7, 8]

$$w(a) = w_0 + w_a(1 - a)\,, \tag{2.1}$$

where a is the scale factor normalized to unity at present, and w_0 and w_a are constants. Crucial for our discussion is the choice of prior ranges for these parameters. In Ref. [3] they are taken to be uniform in the ranges $[-3, 1]$ and $[-3, 2]$ respectively, with the additional condition $w_0 + w_a < 0$ to allow early matter domination.[1] The cosmological constant, or ΛCDM, model corresponds to $w_0 = -1$ and $w_a = 0$, and a regime in which $w(a) < -1$ is called a 'phantom' regime.

From Cortês—Liddle

Note: when they fit a constant *w* model to the data, they find (e.g. when combined with the PantheonPlus supernova dataset) *w* = -0.997 ± 0.025, completely consistent with LambdaCDM's *w* = -1.

From DESI Paper VI: arXiv:2404.03002

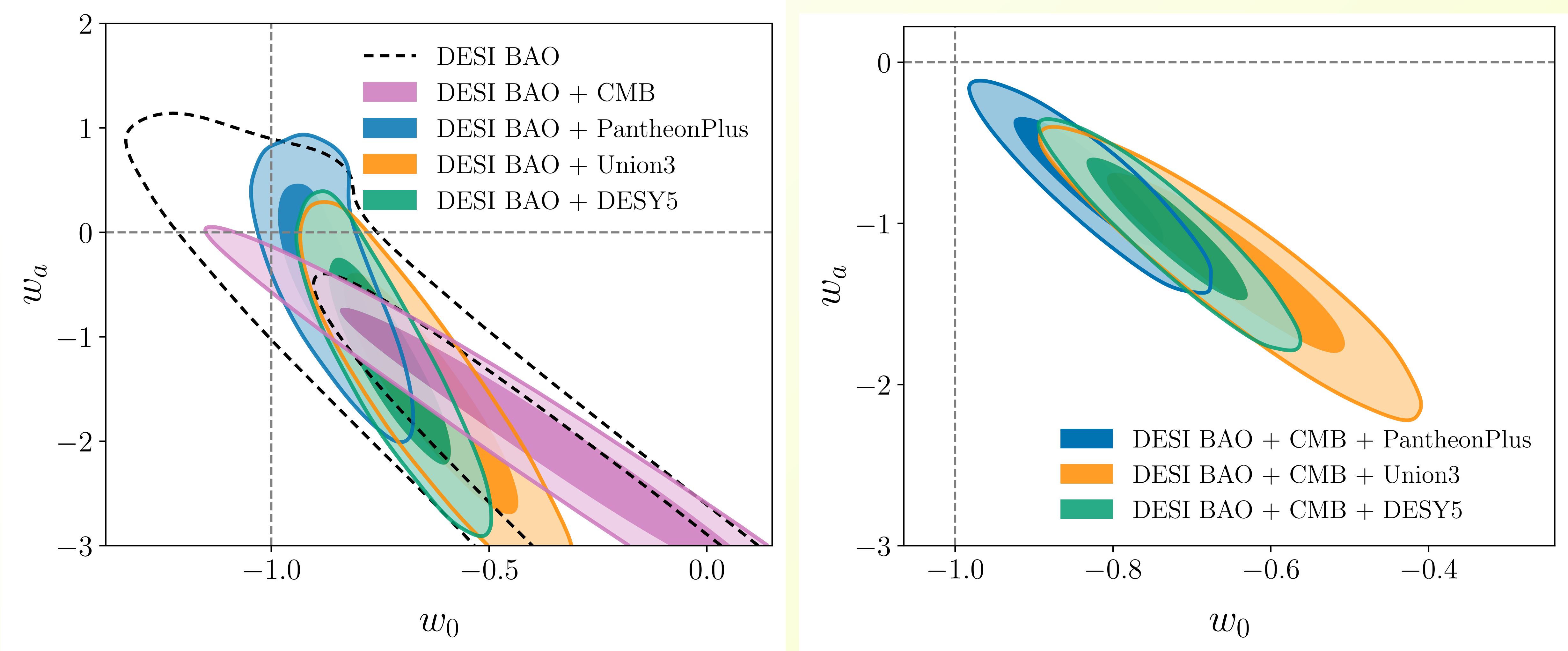


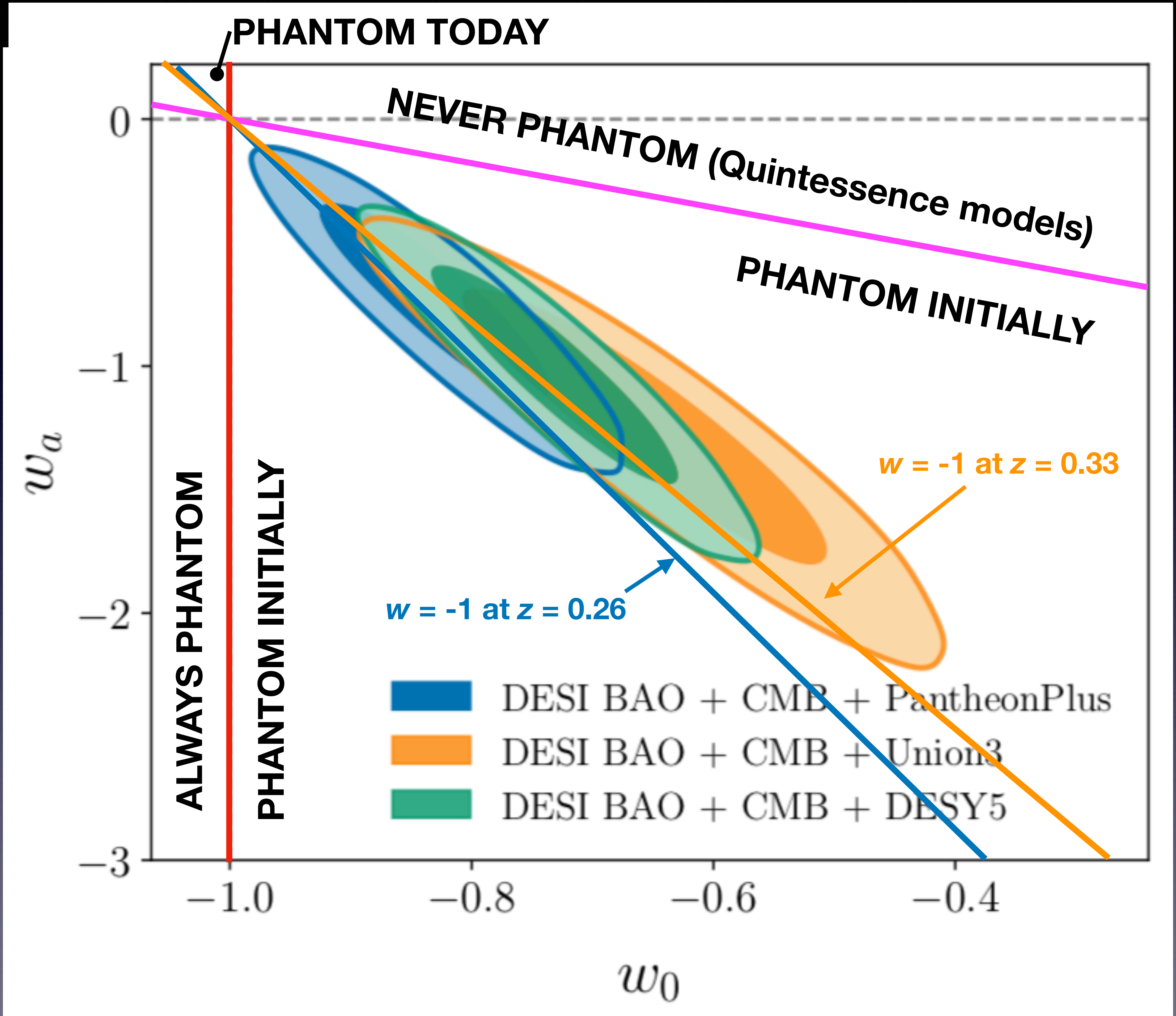
Figure 6. *Left panel*: 68% and 95% marginalized posterior constraints in the w_0–w_a plane for the flat w_0w_aCDM model, from DESI BAO alone (black dashed), DESI + CMB (pink), and DESI + SN Ia, for the PantheonPlus [24], Union3 [25] and DESY5 [26] SNIa datasets in blue, orange and green respectively. Each of these combinations favours $w_0 > -1$, $w_a < 0$, with several of them

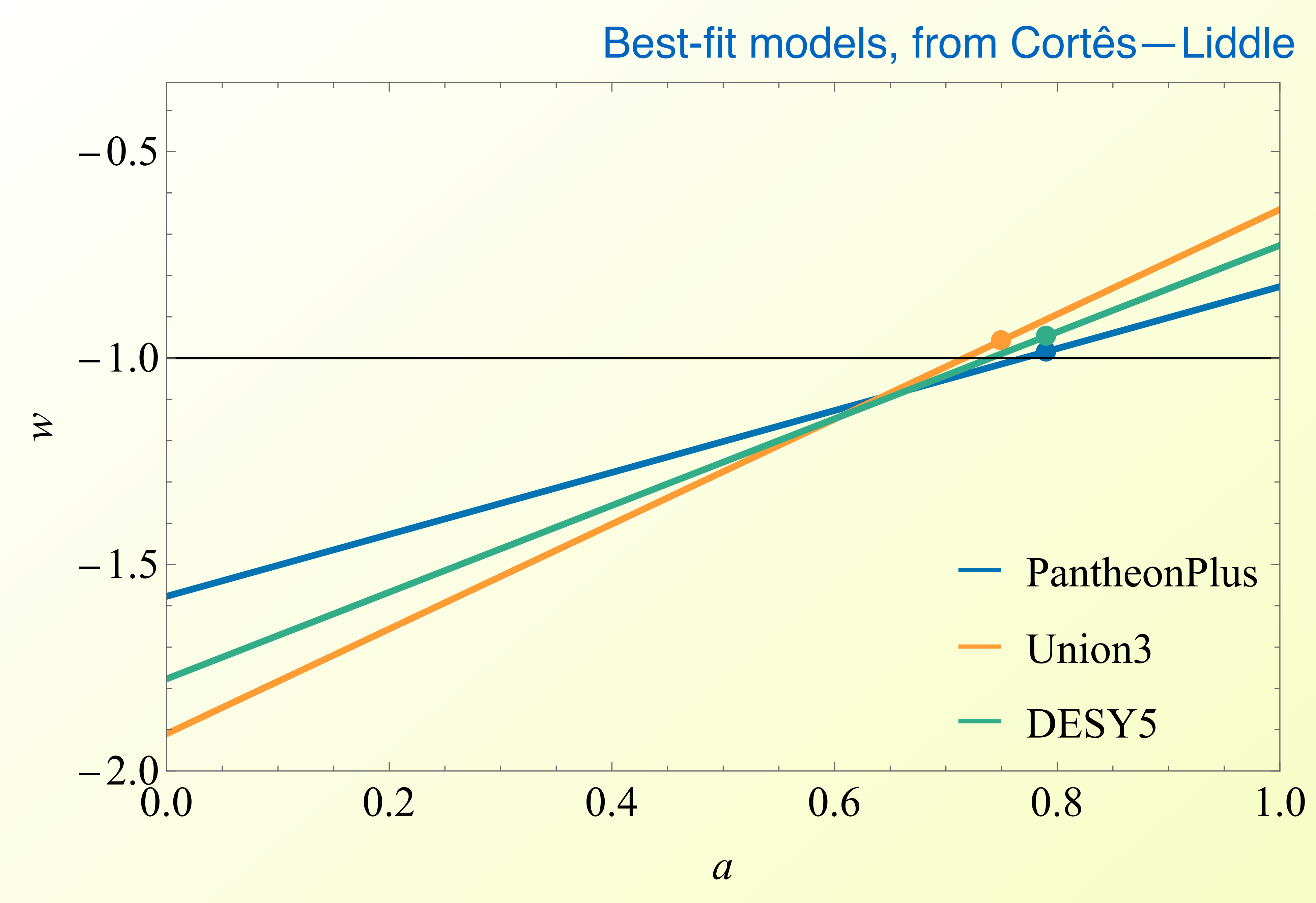
Dark energy pivot

The model $w(a) = w_0 + w_a(1-a)$ has a reparametrization invariance, as we can use any scale a_p to specify the amplitude $w_\mathrm{p} \overset{\mathrm{def}}{=} w(a_\mathrm{p})$. In particular we can choose a_p to decorrelate the estimate of w_p and w_a. This is called the pivot scale, used extensively in the Dark Energy Task Force report (Albrecht et al. 2006) including in their Figure Of Merit.

The pivot scale indicates the scale factor at which the equation of state w is best constrained by the data. The pivot redshifts are $a_\mathrm{p} = 0.79$ for the PantheonPlus and DESY5 supernova choices, and $a_\mathrm{p} = 0.75$ for Union3.

The constraint on w_p closely matches that of a model which assumes w is constant (due to the decorrelation property).

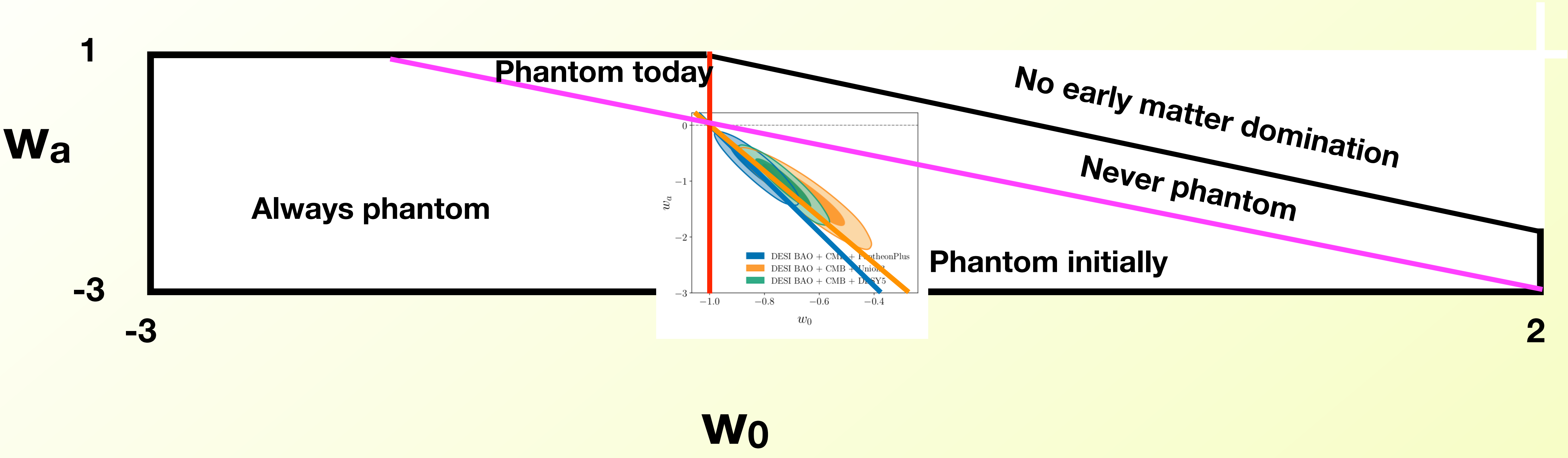

PHANTOM TODAY
NEVER PHANTOM (Quintessence models)
PHANTOM INITIALLY
ALWAYS PHANTOM
PHANTOM INITIALLY
w = -1 at z = 0.33
w = -1 at z = 0.26
DESI BAO + CMB + PantheonPlus
DESI BAO + CMB + Union3
DESI BAO + CMB + DESY5
w_a
0
−1
−2
−3
−1.0
−0.8
−0.6
−0.4
w_0



The best-fit models have the feature that they cross from a phantom regime to a non-phantom regime within the observed window (dots indicate the pivot scale). We call this the PhantomX coincidence.

Equivalently, the coincidence is that the dark energy attains is highest-ever value within the observed window.

FULL DESI PRIOR



However Shlivko and Steinhardt, arXiv:2405.03933, make the interesting observation that quintessence models that are never phantom can nevertheless be well approximated by w0wa models that are in the observationally-preferred sector.

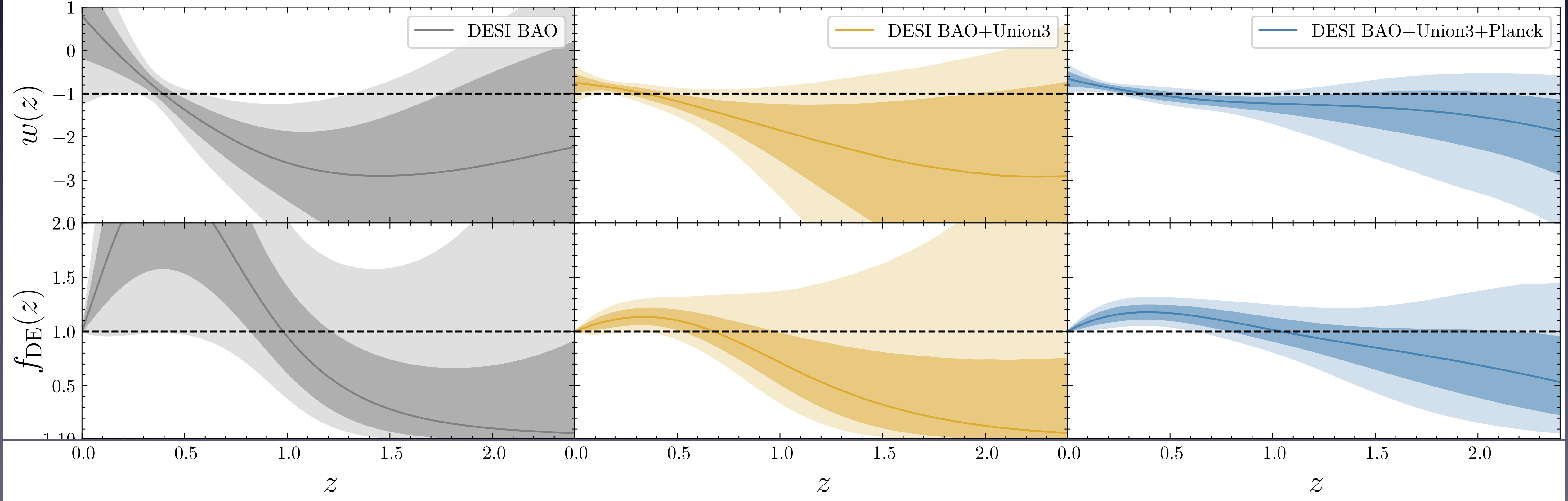
Conclusions: Interpreting DESI's evidence for evolving dark energy

1. DESI's is an exemplary dataset. The analysis in the DOE-led DESI survey, the first Stage IV, has been carried out by the best researchers whom we have come to know for their rigorous internal scrutiny for systematics and probe uncertainties.
2. The DESI result is unexpected in two ways:
   a) The best-fit models are phantom for most of the evolution.
   b) The models cross from phantom to non-phantom within the narrow redshift window probed by observations.
3. The DESI prior appears overly weighted towards phantom models. Altering this could mitigate the phantom crossing coincidence, but would lower the significance of the evolution detection.

Stop press: New DESI analysis paper, May 8th 2024

In this paper (Calderon et al., arXiv:2405.04216), they model *w*(*a*) using a four-term Chebyshev polynomial expansion. They say

> \`Our results hint towards an evolving and emergent dark energy behaviour, with negligible presence of dark energy at $z \gtrsim 1$, at varying significance depending on data sets combined.'



Excerpt from Figure 1 of Calderon et al.