PHYSICS LETTERS B

Letter

Tensions in cosmology: A discussion of statistical tools to determine inconsistencies

Matías Leizerovich [a,b, ID,*], Susana J. Landau [a,b], Claudia G. Scóccola [c,d]

[a] *Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Departamento de Física, Ciudad Universitaria, 1428, Capital Federal, 1460, Buenos Aires, Argentina*
[b] *CONICET - Universidad de Buenos Aires, Instituto de Física de Buenos Aires (IFIBA), Ciudad Universitaria, 1428, Capital Federal, 1460, Buenos Aires, Argentina*
[c] *Facultad de Ciencias Astronómicas y Geofísicas, Universidad Nacional de La Plata, Observatorio Astronómico, Paseo del Bosque, B1900FWA, La Plata, Buenos Aires, Argentina*
[d] *Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Godoy Cruz 2290, Capital Federal, 1425, Buenos Aires, Argentina*

ARTICLE INFO

ABSTRACT

We present a comprehensive analysis of statistical tools for evaluating tensions in cosmological parameter estimates arising from distinct datasets. Focusing on the unresolved Hubble constant (H_0) tension, we explore the Pantheon Plus + SH0ES (PPS) compilation, which includes low-redshift Cepheid data from the SH0ES collaboration, along with the latest release of CMB data from the Planck collaboration, Cosmic Chronometers (CC) dataset and the most recent Baryonic Acoustic Oscillation (BAO) datasets. Employing various tension metrics, we quantitatively assess the inconsistencies in parameter estimates, emphasizing the importance of capturing multidimensional tensions. Our results reveal substantial tension between PPS and Planck 2018 datasets and moderate tension between the BAO data sets and all other datasets. We highlight the importance of adopting these metrics to enhance the precision of future cosmological analyses and facilitate the resolution of existing tensions.

1. Introduction

In the past twenty years, the amount and precision of cosmological data has increased significantly. As a result, stringent constraints were established for the cosmological parameters such as the baryon and dark matter density parameters, the Hubble constant, and others. On the other hand, it is well known that the predictions of the standard cosmological model are in agreement with the majority of the observations. However, there are tensions between the values of some of the cosmological parameters obtained with different datasets. These discrepancies remain an open question for cosmologists and are usually used as a motivation for studying alternative cosmological models. One of the most important unsolved issues is the so-called H_0 tension: the value of the current Hubble parameter H_0 that has been obtained using data from the Cosmic Microwave Background (CMB) assuming a standard cosmological model [22] is not in agreement with the one inferred in a model-independent way, with data from type Ia supernovae and Cepheids [27]. This issue has been extensively discussed in the litera-ture but there is no agreement within the community about the source of the discrepancy [10,26,28,32]. Another issue is the one called S_8 tension, namely, the difference between the estimation of this parameter (defined as $S_8 = \sigma_8 \left(\frac{\Omega_m}{3} \right)^{0.5}$) from CMB and BAO data with the one obtained from weak lensing and galaxy clustering data [13,1]. Hence, the ability to quantify differences in the estimation of cosmological parameters when different datasets are considered is essential for the success of present and future research in Cosmology.

Given two posterior distributions obtained from two different datasets A and B, the most common method to study statistical tension is to evaluate the one-dimensional marginalization of the posterior distributions. A *rule of thumb* formula is introduced in [16] to measure the discrepancy, expressed in terms of the number of standard deviations σ, between parameter estimations θ derived from distinct datasets. The formulation is as follows:

$$N_\theta = \frac{|\mu_A - \mu_B|}{\sqrt{\sigma_A^2 + \sigma_B^2}}, \tag{1}$$

* Corresponding author.
E-mail address: mleize@df.uba.ar (M. Leizerovich).

where the means $\mu_{A/B}$ and the variances $\sigma_{A/B}$ correspond to the ones of the posteriors obtained from each dataset. However, this method carries some problems. For instance, marginalization can hide tensions that can only be seen in higher dimensions. This is caused by the fact that marginalization over some of the parameters necessarily implies a loss of information. Moreover, the number of dimensions of the problem also affects the inferred tension. The significance of the discrepancy in the parameter estimations from the two experiments depends on the number of shared parameters effectively constrained by both experiments. Therefore, a variety of methods have been developed to determine the consistency between the posterior distributions that are obtained from different datasets. For example, different tension metrics that quantify this problem in the whole parameter space at once have been studied in [24] and [23]. In this work, we focus on a particular type of metrics, namely the ones that are based on the posterior distributions, to study tensions between different data sets.[1] We analyze the tension in the most recently released Pantheon Plus + SH0ES (PPS) [29] with three different datasets: the CMB Planck 2018 release (Planck18), Baryonic Acoustic Oscillation (BAO) BOSS and eBOSS most recent datasets, and Cosmic Chronometers data compilation (CC). Although the tension between PPS and Planck18 datasets may be familiar to the reader, the tension between CC and PPS has not been studied in detail, despite being mentioned in [18]. Since CC data only contains information about the Hubble parameter, a non-trivial tension in the comparison between PPS and CC may point to a tension in the Hubble constant at the background level, providing new insights into the study of the H_0 tension.

This work is organized as follows: in Section 2 we describe the tension metrics that we apply to the different datasets, while in Section 3 we present the results of the statistical analyses that show inconsistencies in the obtained confidence intervals of some parameters. In Section 4 we show the amount of inconsistency that we obtain applying the method proposed in this article, that is, from the metrics described in Section 2 and compare it with the one inferred from the rule of thumb. We also discuss these results in light of the interpretation of what each metric is quantifying. Finally, in Section 5 we present the conclusions of our work.

2. Tension metrics

In this section we describe the different metrics that have been introduced in [24,23]. We recall the Bayes formula that defines the posterior distribution

$$P(\theta|D) = \frac{\mathcal{L}(D|\theta) \cdot \Pi(\theta)}{\varepsilon(D)}, \tag{2}$$

where $\mathcal{L}(D|\theta)$ is the likelihood of the data D given the parameters θ, $\Pi(\theta)$ is the prior distribution and $\varepsilon(D) = \int \mathcal{L}(D|\theta) \cdot \Pi(\theta) \, d\theta$ is the Bayesian Evidence.[2] We can divide the metrics into two types [16]: the ones based on Bayesian Evidence $\varepsilon(D)$ and the ones based on the Posterior distributions $P(\theta|D)$. In this work we focus on the last group, which was implemented in the *Tensiometer* repository.[3] All the results are presented in terms of an effective number of standard deviation, N_σ, which is defined by [24]

$$P = \text{Erf}(N_\sigma/\sqrt{2}) \tag{3}$$

where P corresponds to a probability that will identify agreement or disagreement between the considered datasets while N_σ is the number of standard deviations associated to a 1D gaussian distribution with probability P. Moreover, the interpretation of the probability P varies with the chosen metric and will be clarified for each case in what follows.
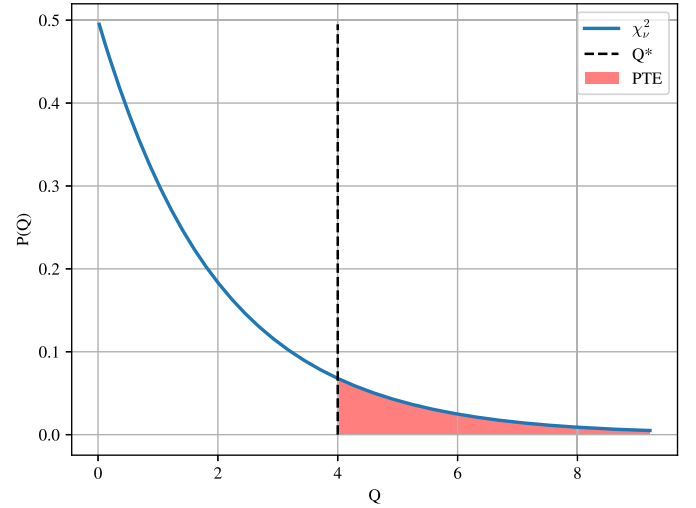
Fig. 1. An example of a quadratic estimator probability density distribution. In red, it is shown the area that represents the probability to exceed, PTE.

2.1. Gaussian metrics

Firstly, we describe tension metrics associated to quadratic estimators, which assumed gaussian posterior distributions. In Fig. 1 we show the area that corresponds to the probability to exceed (PTE) certain observed value of the estimator Q, which is noted as $Q*$. The assigned probability to estimate the tension is $P = 1 - \text{PTE}$, and the equivalent N_σ is obtained using eq. (3).

2.1.1. Parameter differences in standard form

This method consists in calculating the statistical difference between the inferred parameters given by dataset A (θ_A) and the ones given by dataset B (θ_B), with the estimator

$$Q_{DM} = (\hat{\theta}_B - \hat{\theta}_A)^T (\hat{C}_B + \hat{C}_A)^{-1} (\hat{\theta}_B - \hat{\theta}_A), \tag{4}$$

where $(\hat{\theta}_i)$ and $\hat{C}_i$ are the mean and the covariance matrix on the parameter space obtained from the statistical analysis with dataset i (where i correspond to dataset A or B). If both posterior distributions are gaussian, then Q_{DM} follows a χ^2_ν distribution with $\nu = \text{Rank}[\hat{C}_B + \hat{C}_A]$ degrees of freedom.[4] We stress that this metric measures the difference in the obtained confidence intervals of all parameters, while also including the effect of correlations between them. Therefore, it can be regarded as a reasonable generalization of the rule of thumb, with the additional benefit that it quantifies all inconsistencies jointly.

2.1.2. Parameter differences in updated form

Firstly introduced in [24], this method consists in calculating the statistical difference between the inferred parameters given the dataset A (θ_A) and the ones given by the joint datasets (θ_{AB}), with the estimator

$$Q_{UDM} = (\hat{\theta}_{AB} - \hat{\theta}_A)^T (\hat{C}_{AB} - \hat{C}_A)^{-1} (\hat{\theta}_{AB} - \hat{\theta}_A), \tag{5}$$

where $\hat{C}_{AB}$ is the covariance matrix obtained from the statistical analysis with the joint datasets. It can be seen that if the inferred parameters for both datasets A and A+B are gaussian distributed, Q_{UDM} has a χ^2_ν distribution with $\nu = \text{rank}[(\hat{C}_{AB} - \hat{C}_A)]$ degrees of freedom. As it was shown in [24], in order to compute Q_{UDM} it is convenient to break the posterior distribution as a sum over the Karhunen–Loéve (KL) modes of the covariances involved. With this approach, the Q_{UDM} estimator has a χ^2 distribution with ν degrees of freedom, where ν is the number of the KL eigenvalues that are taken into account. We can point out

two interesting features of this metric. Firstly, the directions of the parameter space that shows significant tension can be identified *a priori*, which helps its physical interpretation. Secondly, non-gaussianities are mitigated since we can select the most constraining directions in the parameter space by the two datasets. Finally, it is worth noticing that this metric is asymmetric, since eq. (5) is not invariant under the exchange of A → B. We stress that this metric answers a different question than the other ones, namely, how the results of the statistical analysis using a given dataset are updated if a new dataset is added. For this reason, in our view it is not recommended to quantify the tension with UDM metric if a strong tension has been detected with the other metrics, for example, DM.

2.1.3. Goodness-of-fit loss

This estimator measures the difference in the likelihood function evaluated at the maximum values of the posterior distribution, considering the two datasets jointly and separately in the statistical analysis

$$Q_{\mathrm{DMAP}} = 2ln\mathcal{L}_A(\theta_{pA}) + 2ln\mathcal{L}_B(\theta_{pB}) - 2ln\mathcal{L}_{A+B}(\theta_{pA+B}), \qquad (6)$$

where $\theta_p^{A/B}$ are the Maximum a posteriori (MAP) parameters considering the dataset A/B. Notice that this statistic does not depend explicitly on the covariance of the distribution and only compares the evaluation of the likelihood on certain points. In the case in which the likelihood and the posterior are gaussian distributed, the estimator Q_{DMAP} has a distribution χ^2 with $\Delta\nu$ degrees of freedom, where

$$\Delta\nu = \nu^A + \nu^B - \nu^{A+B}. \qquad (7)$$

In this case, the number of degrees of freedom is defined as

$$\nu = N - tr[C_\Pi^{-1} C_p], \qquad (8)$$

where N is the number of data points and C_Π, C_p are the covariance matrix of the Prior and Posterior distributions, respectively. We emphasize that this metric provides a good quantification of the difference between how well the theoretical predictions can describe both data sets jointly with respect to the same situation but considering the two data sets separately. Therefore, it is a nice tool regarding how the data fit the theoretical prediction but does not give a good estimate of the inconsistency between parameters.

2.2. Non gaussian metrics: Exact Parameter Shift

This method is based on the computation of the parameter difference probability density $P(\Delta\theta)$, where $\Delta\theta = \theta_A - \theta_B$ is the difference between the means of the posterior parameters that correspond to datasets A/B. The general expression for two uncorrelated datasets is given by [24,23]

$$P(\Delta\theta) = \int P_A(\theta)P_B(\theta - \Delta\theta)\, d\theta, \qquad (9)$$

which is analogous to the expression of the cross-correlation function in signal processing. The statistical significance of the shift is calculated by summing over all the values of $P(\Delta\theta)$ over the isocontour corresponding to no shift $\Delta\theta = 0$:

$$\Delta = \int\limits_{P(\Delta\theta)>P(0)} P(\Delta\theta)\, d\Delta\theta. \qquad (10)$$

Here the assigned probability to identify the tension is $P = \Delta$ and the equivalent N_σ is calculated inverting eq. (3). Although it may look like the evaluation of the last equation is straightforward, it is particularly difficult when the parameter space is high dimensional. Taking the difference between the samples of the two MCMC chains as it was described in [25], we can generate the chain of the parameter difference, which is an estimation of the convolution integral in eq. (9). Finally, the integration in eq. (10) can be accomplished using Kernel Density Estimation (KDE) methods. This method does not assume a particular form of the

Table 1
Flat priors on the three cosmological parameters of the standard model used for all analyses in this work. Ω_m and H_0 are derived parameters in Planck18, while the same applies to ω_b in BAO.

Parameter	Ω_m	H_0	ω_b
Prior range	[0.1 , 0.9]	[20 , 100]	[4×10^{-4}, 0.3]

posterior distribution. This metric measures the discrepancy directly from the chains of the parameter difference defined above. Therefore, like Q_{DM}, it quantifies the tension directly from the outputs of the MCMC process. However, in the case that the isocontour of $P(\Delta\theta = 0)$ is far from the maximum of $P(\Delta\theta)$, the output of the metric is difficult to compute.

Next, we point out some final observations about the metrics described before. As it was shown in eq. (4), DM metric quantifies the difference between the parameters obtained from datasets A and B separately, weighted by the sum of their covariance matrices as it was shown in eq. (6). Furthermore, DMAP metric quantifies the difference between the fit with the joint datasets (A+B) with respect to the ones using the datasets separately. Consequently, both metrics quantify different aspects of the tension in the parameter space and it is not trivial to compare between them. On the other hand, Q_{UDM} evaluates how the results using one dataset are updated when including another one. In short, in the case of gaussian posteriors, Q_{DM} and Parameter Shift quantify the discrepancy between inferred parameters directly from the outputs of the inference process and therefore should be considered the best metrics to quantify discrepancies between parameters.

3. Parameter inferences with different datasets

In this paper, we compute the estimators described in Section 2 for different combinations of four cosmological datasets. Firstly, we consider the recently released Pantheon Plus compilation of type Ia supernovae (SNIa). We note that this release includes the option of using low redshift Cepheid data obtained by the SHOES collaboration, which are crucial for the calibration of SNIa and therefore for the Hubble tension. Therefore, we name this data set as Pantheon Plus + SHOES (PPS). The Pantheon Plus compilation consists of 1,701 SNIa at redshifts between $0.0012 < z < 2.26$, which are available in [29].[5] Secondly, we consider data obtained with the Cosmic Chronometers technique detailed in [30,31,19,33,17,20]. Thirdly, we consider the most recent BOSS and eBOSS datasets [15,2,3,5,11,7,14,21]. Finally, we use CMB data from the latest release of Planck [22] (Planck18).

To estimate the posterior distribution we use the latest available version of *CLASS* [6] and perform a Markov Chain Monte Carlo (MCMC) analysis with the latest release of the *MontePython* software [8,4]. Table 1 shows the priors on the shared parameter space[6] that are considered in our analyses. Note that the prior in the physical baryon density $\omega_b = \Omega_b h^2$ is only used in the comparison between Planck18 and BAO.

For the CC dataset, the matter density Ω_m and the Hubble parameter H_0 are the free parameters of the model. For PPS, the Supernovae absolute magnitude M_{abs} has to be taken into account as a nuisance parameter. For BAO, we consider (Ω_m, H_0, Ω_b). Finally, when Planck18 data is taken into account, all parameters of the standard cosmological model, ΛCDM, (ω_b, ω_{cdm}, θ_s, A_s, n_s, τ_{reio}) are considered, apart from the nuisance parameters of the Planck likelihood. In this case, Ω_m and H_0 can be obtained as derived parameters of this analysis. Our analysis is focused on the comparison of different pairs of the mentioned datasets. Fig. 2 shows the results of the statistical analysis performed considering

---

[5] https://github.com/PantheonPlusSH0ES/DataRelease

[6] In Planck18, Ω_m and H_0 are derived parameters, while in BAO the derived parameter is ω_b.
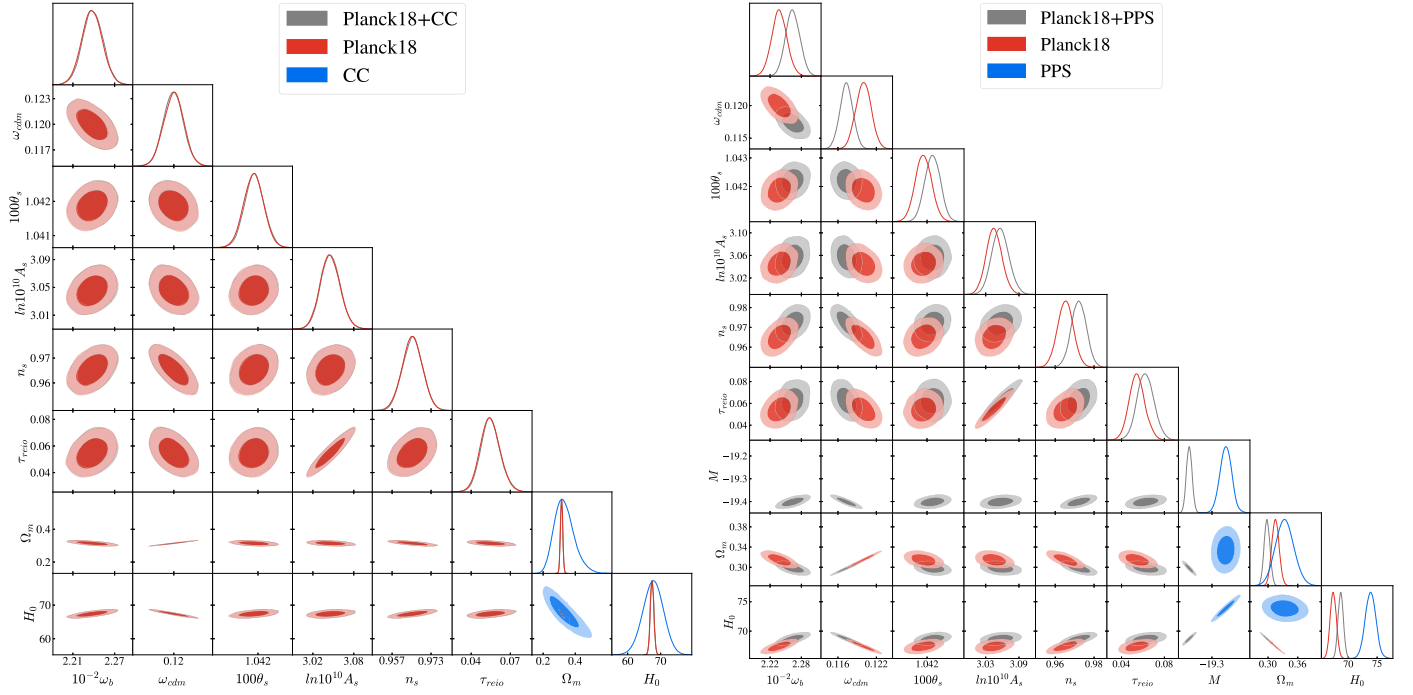
Fig. 2. *Results of the statistical analysis for the ΛCDM model using Planck likelihood 2018 with Cosmic Chronometers (left) and with Pantheon Plus + SH0ES (right),* respectively. The darker and brighter regions correspond to 65% and 95% confidence regions, respectively. The plots in the diagonal show the posterior probability density for each of the free parameters of the model.

Planck18/CC and Planck18/PPS datasets jointly and separately.[7] It follows that for Planck18/CC (left) the datasets are in agreement, while for Planck18/PPS the tension is considerable large. Besides, it is worth noticing that the shared parameter space for all datasets is the $\Omega_m - H_0$ plane, except for the pair Planck18/BAO in which the shared parameter space is (Ω_m, H_0, ω_b). The projections of the posterior distributions on the shared parameter subspace are shown in Figs. 3, 4, and 5. Finally, we point out that the posteriors are gaussian distributions in all cases. This is because priors are not informative.

4. Results

Here we present the results of our analysis. The pair of datasets that have been compared are Planck/CC, CC/PPS, PPS/BAO, Planck18/PPS and Planck18/BAO. Before presenting the results for each pair of datasets, we discuss the results in general for all the metrics.

4.1. Discussion for each metric

4.1.1. DM

Here we discuss DM metric in view of all our results. According to eq. (4), the tension for this metric is proportional to the difference between the means of the posterior distributions, and to the inverse of the sum of the covariance matrices. Following this definition, we encounter two paradigmatic cases: Firstly, the tension between Planck18/CC presents the lowest tension of all the analyses (as it is shown in Fig. 3, the difference between the means of each posterior is negligible). Secondly, the tension between Planck18/PPS is the strongest one (the difference between the means is the most important one, as it is shown in Fig. 4 (right)). Besides, it is remarkable that DM metric indicates a different level of tension with respect to the sum of the 1D rule of thumb for all analysis. This is reasonable since the rule of thumb does not take into account the correlation between parameters.
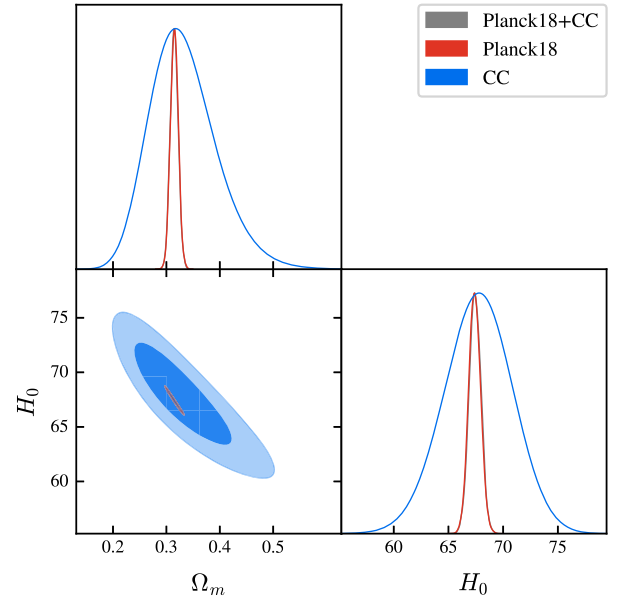

Fig. 3. Results of the statistical analysis for the ΛCDM showing the marginalization on the shared plane on the parameter space $\Omega_m - H_0$ for Planck18+CC. The darker and brighter regions correspond to 65% and 95% confidence regions, respectively. The plots in the diagonal show the posterior probability density for each of the free parameters of the model.

Finally, we encountered three cases with moderate tension: CC/PPS, Planck18/BAO and PPS/BAO.[8] Let us briefly discuss the first two cases. Although in the 1D projection it seems that CC/PPS presents weaker tension than Planck18/BAO, the 2D projection indicates that it is the opposite case. The fact that DM metric takes into account the covari-

---

[7] In the case of Planck18, we do not show the contours corresponding to the nuisance parameters of the Planck likelihood.

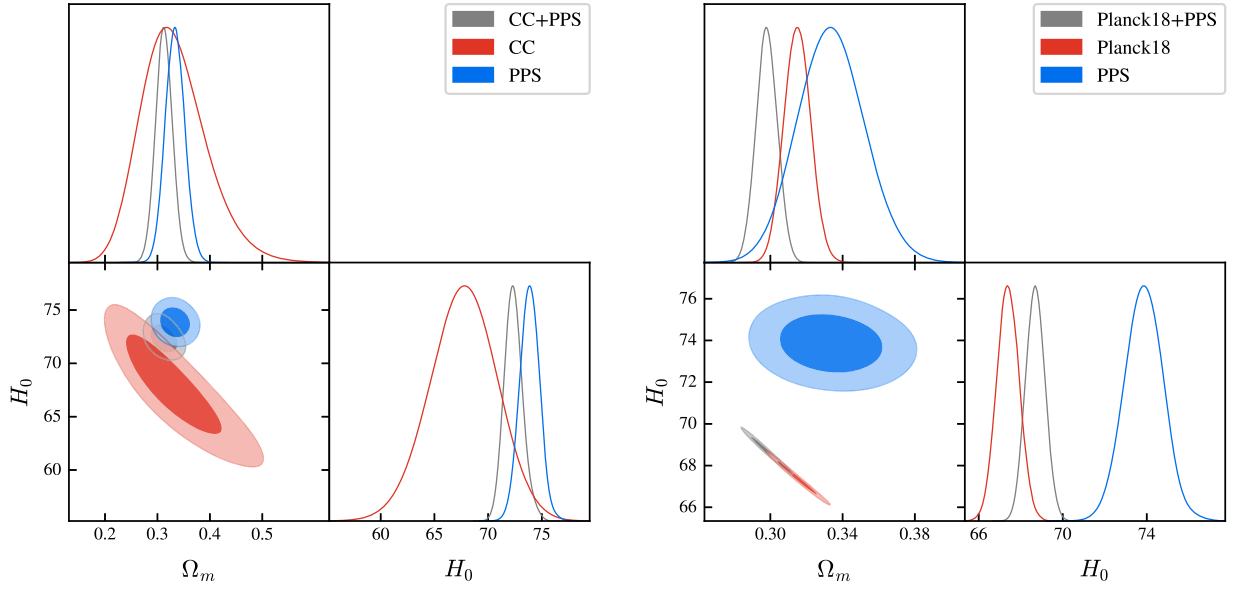[8] This last case shows weak to moderate tension.

Fig. 4. Results of the statistical analysis for the ΛCDM showing the marginalization on the shared plane on the parameter space $\Omega_m - H_0$. The darker and brighter regions correspond to 65% and 95% confidence regions, respectively. The plots in the diagonal show the posterior probability density for each of the free parameters of the model.
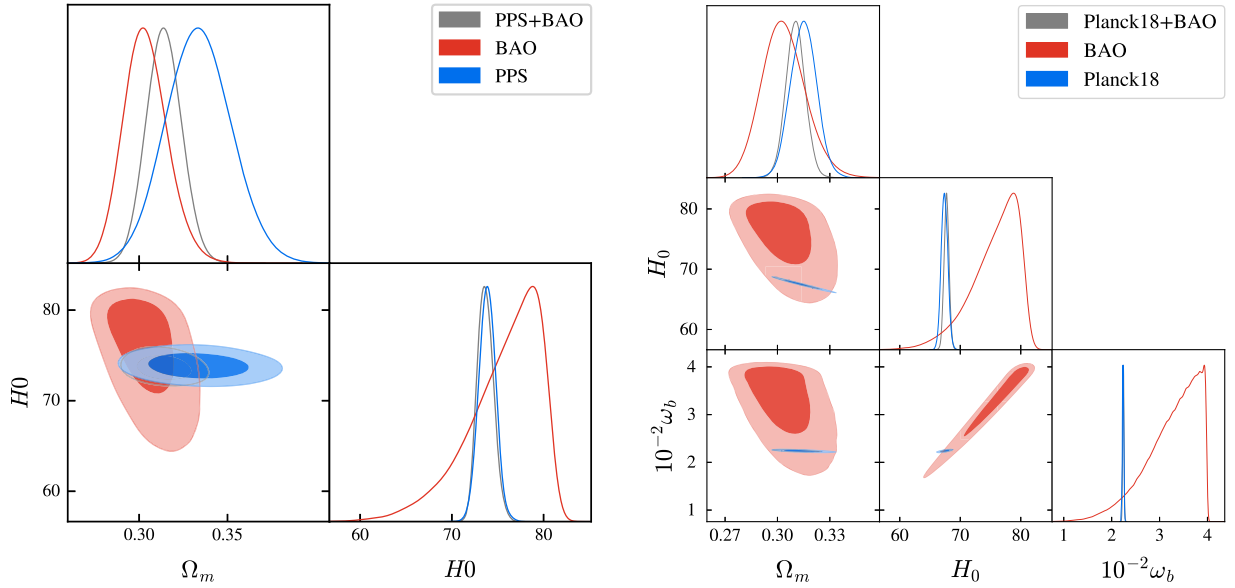

Fig. 5. Results of the statistical analysis for the ΛCDM showing the marginalization on the shared plane on the parameter space $\Omega_m - H_0$ for BAO+PPS and $(\Omega_m, H_0, \omega_b)$ for Planck18+BAO. The darker and brighter regions correspond to 65% and 95% confidence regions, respectively. The plots in the diagonal show the posterior probability density for each of the free parameters of the model.

ance of the posterior distributions allows to quantify better the tension between datasets in moderate cases.

4.1.2. UDM

According to eq. (5), the tension for this metric is proportional to the difference between the means of the posterior distributions of one of the datasets and the joint analysis, and to the inverse of the difference between their covariance matrices. As it was discussed before, this metric is not symmetric and we expect different results when some dataset is updated with another one. In order to gain intuition of UDM results, we analyze two relevant cases.

Firstly, we analyze the case of CC/PPS. When CC results are updated with PPS, the obtained tension is $\sim 2.4\sigma$, while when PPS is updated with CC the tension is $\sim 4.8\sigma$. This can be explained as follows. Accord-

ing to eq. (5), the estimator may indicate higher tension if the difference in the means is high or if the difference $|C_{AB} - C_A|$ is small. In this case, the covariance matrices of the posteriors corresponding to PPS and PPS+CC are quite similar which makes the estimator Q^*_{UDM} and the corresponding N_σ higher.

Secondly, we analyze the case of Planck18/PPS. When Planck18 results are updated with PPS the equivalent N_σ is $\sim 2.3\sigma$, while when PPS is updated with Planck18 we obtain $\sim 4.6\sigma$. Contrary to the case of PPS+CC, here the dominant effect on the metric is the difference between the means of the compared distributions, rather than the difference in the covariance matrices, as it was shown in the right panel of Fig. 4. We note that in this case, the UDM metric has one effective degree of freedom. This is because the KL decomposition points out that the majority of the variance is condensed in the first eigenvalue.

4.1.3. DMAP

According to eq. (6), DMAP compares the likelihoods of the two datasets separately with the one of the joint analysis, all of them evaluated at the maximum of their corresponding posterior distributions. If both datasets are independent, we can write eq. (6) as

$$Q_{\text{DMAP}} = 2\ln\left(\frac{\mathcal{L}_A(\theta_{pA}) \cdot \mathcal{L}_B(\theta_{pB})}{\mathcal{L}_A(\theta_{pA+B}) \cdot \mathcal{L}_B(\theta_{pA+B})}\right). \tag{11}$$

This equation shows that the comparison is between the separate likelihoods evaluated at the maxima of their posteriors and the same likelihoods evaluated at the maxima of the joint posterior distribution. Assuming gaussianity, when θ_{pA+B} is far from θ_{pA} or θ_{pB}, this results in a tension on the DMAP metric.

Unlike DM, this metric does not take into account the covariance matrix of the distributions but has information on the joint analysis. For this reason, both DM and DMAP quantify the tension in different ways and their results cannot be compared. However, like DM, this metric shows that the strongest tension also corresponds to Planck18/PPS and the weakest to Planck18/CC. Finally, since we are using flat uninformative priors, the effective number of degrees of freedom for all analyses is equal to the dimensions of the shared parameter space in all cases.

Also, we discuss two cases of moderate tension: CC/PPS and PPS/BAO. To facilitate the discussion, we assume that the posterior distribution is equal to the corresponding likelihood and that no relevant information is lost in the marginalization process.[9] The left panel of Fig. 4 (CC/PPS) shows that the maximum of the joint posterior is located in the 2σ contour of the corresponding posteriors obtained considering only one dataset, and the left panel of Fig. 5 (PPS/BAO) shows a similar behavior but this time the maximum of the joint posterior is placed in the 1σ contour of the corresponding contours obtained with only one dataset. This is in agreement with the results shown in Tables 3 and 5.

4.1.4. Exact Parameter Shift

Here we discuss our results for Exact Parameter Shift. In Fig. 6, we show the distribution of $P(\Delta\theta)$ for three different levels of tension: the Planck18/CC comparison with negligible tension (left), the CC/PPS comparison with moderate tension (center) and the Planck18/PPS comparison with strong tension (right). Apart from the 1σ and 2σ contours, the last two cases show the isocontour that corresponds to $P(\Delta\theta = 0)$.

Besides, we expect that when the distributions are gaussian, the result of this metric is similar to the one of Q_{DM}. Indeed, this is the case for CC/PPS, Planck18/BAO, Planck18/CC and PPS/BAO. It is particularly relevant the tension between Planck18/PPS, in which Exact Parameter Shift gives an infinite tension. Although this result may be uncomfortable, it can be explained as follows: there is no isocontour of $P(\Delta\theta)$ that corresponds to $P(\Delta\theta = 0)$ as it was shown in Fig. 6 (right). According to eqs. (3) and (10), this implies that $\Delta = 1$ and so the probability is translated to an infinite number of σ. We also checked that this effect is not due to a problem of sampling.

4.2. Comparison between pairs of datasets

In what follows, we describe our results for the tension metrics when taking the different datasets in pairs.

4.2.1. Comparison between Planck18 and CC

The comparison between Planck18 and CC shows weak tension (see Table 2 and Fig. 3). As we discussed in Subsec. 4.1, in this comparison we

---

Table 2
Results of the application of different metrics on the datasets of Planck18/CC. The results are presented in terms of the probability P for a χ^2_ν distribution with ν degrees of freedom when it corresponds (or $1 - \Delta$ in the case of Exact Parameter Shift) and in all cases for the number of standard deviations N_σ.

	ν	Q*	PTE	N_σ
DM	2	0.54	0.76	0.300
UDM$_{\text{Planck18}\to\text{Planck18+CC}}$	2	1.06	0.59	0.541
UDM$_{\text{CC}\to\text{Planck18+CC}}$	1	0.12	0.73	0.348
DMAP	2	0.63	0.73	0.345
Exact Param. Shift	-	-	0.83	0.212
Rule of thumb		$N_{\Omega_m} = 0.280$		
		$N_{H_0} = 0.0995$		

Table 3
Results of the application of different metrics on the datasets of CC/PPS. The results are presented in terms of the probability PTE for a χ^2_ν distribution with ν degrees of freedom when it corresponds (or $1 - \Delta$ in the case of Exact Parameter Shift) and in all cases for the number of standard deviations N_σ.

	ν	Q*	PTE	N_σ
DM	2	10.38	5.57×10^{-3}	2.772
UDM$_{\text{CC}\to\text{CC+PPS}}$	2	8.62	1.35×10^{-2}	2.471
UDM$_{\text{PPS}\to\text{CC+PPS}}$	2	26.58	1.35×10^{-4}	4.778
DMAP	2	4.85	8.84×10^{-2}	1.704
Exact Param. Shift	-	-	4.80×10^{-3}	2.820
Rule of thumb		$N_{\Omega_m} = 0.0171$		
		$N_{H_0} = 1.899$		

expect maximum consistency between datasets. It is worth noting that in all cases, the tension is higher than the one reported by the rule of thumb. This example shows that even in the case of no full consistency, the tension is higher when all shared parameter space is considered.

4.2.2. Comparison between PPS and CC

The comparison between PPS and CC is shown in Table 3 and Fig. 4 (left). As we discussed in Subsec. 4.1. DM and Exact Parameter Shift match their results as expected. Besides, DMAP metric indicates an smaller tension. Finally, note that DM and DMAP indicate moderate tension.

4.2.3. Comparison between Planck18 and BAO

Comparison between Planck18 and BAO is presented in Table 4 and Fig. 5 (right) and shows a weak to moderate tension. Unlike the previous analyses, in this case the shared parameter space has three dimensions: $(\Omega_m, H_0, \omega_b)$. This implies that we cannot visualize the total likelihood but only the 2D projections. As expected, DM metric and Exact Parameter Shift are in agreement, while DMAP shows a weaker tension than the other ones.

4.2.4. Comparison between PPS and BAO

Comparison between PPS and BAO is shown in Table 5 and Fig. 5 (left). In this case, all metrics indicate a moderate tension. The results for the DM metric and Exact Parameter Shift are similar, while DMAP metric indicates a higher tension than the other ones.

4.2.5. Comparison between PPS and Planck18

Finally, the comparison between PPS and Planck18 is shown in Table 6 and Fig. 4 (right). As we discussed in Subsec. 4.1, this is the case that shows the strongest tension. For the DM metric, we obtain a tension of $\sim 6.4\sigma$, which is not far from the estimation obtained from the rule of thumb for H_0. The most relevant result of this analysis is the infinite number of standard deviations using the Exact Parameter Shift metric, which has been already discussed in Subsec. 4.1.
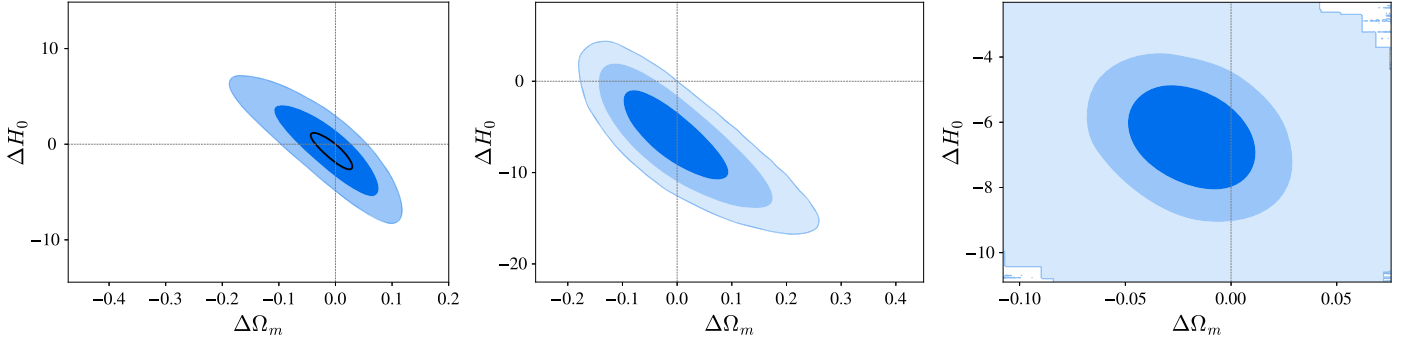
Fig. 6. Distribution of parameter differences defined in eq. (9), for the datasets Planck18/CC (left), CC/PPS (center) and Planck18/PPS (right). The first two delimited contours correspond to 65%, 95% confidence region. On the first figure, the black contour represents the contour that corresponds to $P(\Delta\theta = 0)$. On the last two figures, this contour is outside the 95% confidence region.

Table 4

Results of the application of different metrics on the datasets of Planck18/BAO. The results are presented in terms of the probability P for a χ^2_ν distribution with ν degrees of freedom when it corresponds (or $1 - \Delta$ in the case of Exact Parameter Shift) and in all cases for the number of standard deviations N_σ.

	ν	Q*	PTE	N_σ
DM	3	4.62	0.20	1.276
$UDM_{Planck18 \rightarrow Planck18+BAO}$	3	1.45	0.69	0.395
$UDM_{BAO \rightarrow Planck18+BAO}$	1	0.21	0.65	0.460
DMAP	3	3.28	0.35	0.933
Exact Param. Shift	-	-	0.14	1.473
Rule of thumb		$N_{\Omega_m} = 0.767$		
		$N_{H_0} = 2.041$		
		$N_{\omega_b} = 1.847$		

Table 5

Results of the application of different metrics on the datasets of PPS/BAO. The results are presented in terms of the probability P for a χ^2_ν distribution with ν degrees of freedom when it corresponds (or $1 - \Delta$ in the case of Exact Parameter Shift) and in all cases for the number of standard deviations N_σ.

	ν	Q*	PTE	N_σ
DM	2	1.80	0.41	0.829
$UDM_{BAO \rightarrow PPS+BAO}$	2	2.32	0.314	1.007
$UDM_{PPS \rightarrow PPS+BAO}$	2	2.52	0.284	1.072
DMAP	2	1.55	0.460	0.739
Exact Parameter Shift	-	-	0.311	1.014
Rule of thumb		$N_{\Omega_m} = 1.339$		
		$N_{H_0} = 0.434$		

Table 6

Results of the application of different metrics on the datasets of Planck18/PPS. The results are presented in terms of the probability P for a χ^2_ν distribution with ν degrees of freedom when it corresponds (or $1 - \Delta$ in the case of Exact Parameter Shift) and in all cases for the number of standard deviations N_σ.

	ν	Q*	PTE	N_σ
DM	2	45.44	1.36×10^{-10}	6.420
$UDM_{Planck18 \rightarrow Planck18+PPS}$	1	5.36	2.06×10^{-2}	2.316
$UDM_{PPS \rightarrow Planck18+PPS}$	1	21.34	3.84×10^{-6}	4.620
DMAP	2	21.23	2.46×10^{-5}	4.218
Exact Parameter Shift	-	-	0	∞
Rule of thumb		$N_{\Omega_m} = 0.942$		
		$N_{H_0} = 6.093$		

5. Summary and conclusions

In this work, we discuss the importance of using tension metrics to determine tensions between different datasets. We show that the different metrics analyzed here can quantify the tension more precisely than the widely applied Rule of Thumb.

Among the metrics used, three of them (i.e., DM, DMAP, Exact Parameter Shift) quantify the difference between the inferred parameters while the other one (UDM) measures how much one dataset updates the results of another when added to the statistical analysis. Also, we discuss some implementation details; for example, some computational difficulties appeared when computing the Exact Parameter Shift while analyzing the tension between Planck18 and PPS.

Our results show two extreme cases: i) a very good agreement between CC and Planck18, and ii) a strong tension between PPS and Planck18. We also find three intermediate cases with moderate tension (BAO and Planck18; BAO and CC; BAO and PPS). These moderate tensions were not pointed out nor quantified before. Therefore, our analyses show that the tension metrics are excellent tools for quantifying moderate tension between distinct data sets. With the availability of new data in the near future, it is expected that the errors will be reduced, and the posterior contours will be narrower, potentially increasing the tension between datasets.

Applying the tension metrics that we have analyzed here, allows us to perform a detailed and precise analysis of the tension between different datasets, including also the correlation between parameters. We emphasize that the use of tension metrics determines the tension between distinct datasets in the plane of shared parameters rather than the tension in a single parameter as in the case of the rule of thumb. We expect that the use of these metrics will become relevant for the analysis of future datasets. For example, some of these metrics have been used to discuss the recently released DESI results [9]. The increase in the precision of new results may lead to stronger tensions, which can be quantitatively described using tension metrics.

Finally, we discuss which metric is best to quantify the tension between datasets. To give a final answer, several aspects have to be taken into account. Firstly, as it has been pointed out before, UDM metric quantifies how a new dataset updates another one, and its results are not comparable with the ones of DM/Exact Parameter Shift or DMAP. Secondly, although the rest of the metrics indeed quantify the tension between independent datasets, DM/Exact Parameter Shift and DMAP quantify different aspects of the tension: while DMAP quantifies the distance between the MAP evaluating the likelihood of the joint posterior and the one using the datasets separately, DM quantifies the distance between the means of the independent posteriors weighted by their covariance matrix. Therefore, in our opinion, DM/Exact Parameter Shift answers accurately the question about the tension between independent datasets.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Susana J. Landau reports financial support was provided by CONICET. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

References

[1] T.M.C. Abbott, M. Aguena, A. Alarcon, et al., DES Collaboration, Dark Energy Survey Year 3 results: cosmological constraints from galaxy clustering and weak lensing, Phys. Rev. D 105 (2022) 023520, https://doi.org/10.1103/PhysRevD.105.023520, arXiv:2105.13549.

[2] S. Alam, M. Ata, S. Bailey, F. Beutler, et al., The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample, Mon. Not. R. Astron. Soc. 470 (2017) 2617–2652, https://doi.org/10.1093/mnras/stx721.

[3] S. Alam, M. Aubert, S. Avila, C. Balland, et al., Completed sdss-iv extended baryon oscillation spectroscopic survey: cosmological implications from two decades of spectroscopic surveys at the apache point observatory, Phys. Rev. D 103 (2021) 083533, https://doi.org/10.1103/PhysRevD.103.083533, https://link.aps.org/doi/10.1103/PhysRevD.103.083533.

[4] B. Audren, J. Lesgourgues, K. Benabed, S. Prunet, Conservative Constraints on Early Cosmology: an illustration of the Monte Python cosmological parameter inference code, J. Cosmol. Astropart. Phys. 1302 (2013) 001, https://doi.org/10.1088/1475-7516/2013/02/001, arXiv:1210.7183.

[5] J.E. Bautista, R. Paviot, M. Vargas Magaña, S. de la Torre, et al., The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the BAO and growth rate of structure of the luminous red galaxy sample from the anisotropic correlation function between redshifts 0.6 and 1, Mon. Not. R. Astron. Soc. 500 (2020) 736–762, https://doi.org/10.1093/mnras/staa2800.

[6] D. Blas, J. Lesgourgues, T. Tram, The cosmic linear anisotropy solving system (class). Part II: Approximation schemes, J. Cosmol. Astropart. Phys. 2011 (2011) 034, https://doi.org/10.1088/1475-7516/2011/07/034.

[7] H. du Mas des Bourboux, J. Rich, A. Font-Ribera, V. de Sainte Agathe, et al., The completed sdss-iv extended baryon oscillation spectroscopic survey: baryon acoustic oscillations with lyα forests, Astrophys. J. 901 (2020) 153, https://doi.org/10.3847/1538-4357/abb085.

[8] T. Brinckmann, J. Lesgourgues, Montepython 3: boosted mcmc sampler and other features, Phys. Dark Universe 24 (2019) 100260, https://doi.org/10.1016/j.dark.2018.100260.

[9] DESI Collaboration, A.G. Adame, J. Aguilar, S. Ahlen, et al., DESI 2024 VI: cosmological constraints from the measurements of baryon acoustic oscillations, arXiv e-prints, arXiv:2404.03002, 2024.

[10] W.L. Freedman, Measurements of the Hubble constant: tensions in perspective, Astrophys. J. 919 (2021) 16, https://doi.org/10.3847/1538-4357/ac0e95, arXiv:2106.15656.

[11] H. Gil-Marín, J.E. Bautista, R. Paviot, M. Vargas-Magaña, et al., The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the BAO and growth rate of structure of the luminous red galaxy sample from the anisotropic power spectrum between redshifts 0.6 and 1.0, Mon. Not. R. Astron. Soc. 498 (2020) 2492–2531, https://doi.org/10.1093/mnras/staa2455.

[12] A. Gómez-Valent, Fast test to assess the impact of marginalization in Monte Carlo analyses and its application to cosmology, Phys. Rev. D 106 (2022) 063506, https://doi.org/10.1103/PhysRevD.106.063506, arXiv:2203.16285.

[13] C. Heymans, T. Tröster, M. Asgari, et al., KiDS-1000 Cosmology: multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints, Astron. Astrophys. 646 (2021) A140, https://doi.org/10.1051/0004-6361/202039063, arXiv:2007.15632.

[14] J. Hou, A.G. Sánchez, A.J. Ross, A. Smith, et al., The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: BAO and RSD measurements from anisotropic clustering analysis of the quasar sample in configuration space between redshift 0.8 and 2.2, Mon. Not. R. Astron. Soc. 500 (2020) 1201–1221, https://doi.org/10.1093/mnras/staa3234.

[15] C. Howlett, A.J. Ross, L. Samushia, W.J. Percival, M. Manera, The clustering of the SDSS main galaxy sample – II. Mock galaxy catalogues and a measurement of the growth of structure from redshift space distortions at z = 0.15, Mon. Not. R. Astron. Soc. 449 (2015) 848–866, https://doi.org/10.1093/mnras/stu2693.

[16] P. Lemos, M. Raveri, A. Campos, et al., DES Collaboration, Assessing tension metrics with dark energy survey and Planck data, Mon. Not. R. Astron. Soc. 505 (2021) 6179–6194, https://doi.org/10.1093/mnras/stab1670, arXiv:2012.09554.

[17] M. Moresco, Raising the bar: new constraints on the Hubble parameter with cosmic chronometers at z 2, Mon. Not. R. Astron. Soc. Lett. 450 (2015) L16–L20, https://doi.org/10.1093/mnrasl/slv037.

[18] M. Moresco, Addressing the Hubble tension with cosmic chronometers, arXiv:2307.09501, 2023.

[19] M. Moresco, A. Cimatti, R. Jimenez, et al., Improved constraints on the expansion rate of the Universe up to z ∼ 1.1 from the spectroscopic evolution of cosmic chronometers, J. Cosmol. Astropart. Phys. 8 (2012) 006, https://doi.org/10.1088/1475-7516/2012/08/006, arXiv:1201.3609.

[20] M. Moresco, L. Pozzetti, A. Cimatti, R. Jimenez, C. Maraston, L. Verde, D. Thomas, A. Citro, R. Tojeiro, D. Wilkinson, A 6% measurement of the Hubble parameter at z∼0.45: direct evidence of the epoch of cosmic re-acceleration, J. Cosmol. Astropart. Phys. 5 (2016) 014, https://doi.org/10.1088/1475-7516/2016/05/014, arXiv:1601.01701.

[21] R. Neveux, E. Burtin, A. de Mattia, A. Smith, et al., The completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: BAO and RSD measurements from the anisotropic power spectrum of the quasar sample between redshift 0.8 and 2.2, Mon. Not. R. Astron. Soc. 499 (2020) 210–229, https://doi.org/10.1093/mnras/staa2780.

[22] Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, et al., Planck 2018 results. VI. Cosmological parameters, Astron. Astrophys. 641 (2020) A6, https://doi.org/10.1051/0004-6361/201833910, arXiv:1807.06209.

[23] M. Raveri, C. Doux, Non-gaussian estimates of tensions in cosmological parameters, Phys. Rev. D 104 (2021) 043504, https://doi.org/10.1103/PhysRevD.104.043504, https://link.aps.org/doi/10.1103/PhysRevD.104.043504.

[24] M. Raveri, W. Hu, Concordance and discordance in cosmology, Phys. Rev. D 99 (2019) 043506, https://doi.org/10.1103/PhysRevD.99.043506, https://link.aps.org/doi/10.1103/PhysRevD.99.043506.

[25] M. Raveri, G. Zacharegkas, W. Hu, Quantifying concordance of correlated cosmological data sets, Phys. Rev. D 101 (2020) 103527, https://doi.org/10.1103/PhysRevD.101.103527, https://link.aps.org/doi/10.1103/PhysRevD.101.103527.

[26] A.G. Riess, L. Breuval, The local value of H_0, arXiv e-prints, arXiv:2308.10954, 2023.

[27] A.G. Riess, W. Yuan, L.M. Macri, et al., A comprehensive measurement of the local value of the Hubble constant with 1 km s^{-1} Mpc^{-1} uncertainty from the Hubble Space Telescope and the SH0ES team, Astrophys. J. Lett. 934 (2022) L7, https://doi.org/10.3847/2041-8213/ac5c5b, arXiv:2112.04510.

[28] N. Schöneberg, G.F. Abellán, A.P. Sánchez, S.J. Witte, V. Poulin, J. Lesgourgues, The H_0 Olympics: a fair ranking of proposed models, Phys. Rep. 984 (2022) 1–55, https://doi.org/10.1016/j.physrep.2022.07.001, arXiv:2107.10291.

[29] D. Scolnic, et al., The Pantheon+ analysis: the full data set and light-curve release, Astrophys. J. 938 (2022) 113, https://doi.org/10.3847/1538-4357/ac8b7a, arXiv:2112.03863.

[30] J. Simon, L. Verde, R. Jimenez, Constraints on the redshift dependence of the dark energy potential, Phys. Rev. D 71 (2005) 123001, https://doi.org/10.1103/PhysRevD.71.123001, arXiv:astro-ph/0412269.

[31] D. Stern, R. Jimenez, L. Verde, M. Kamionkowski, S.A. Stanford, Cosmic chronometers: constraining the equation of state of dark energy. I: H(z) measurements, J. Cosmol. Astropart. Phys. 2 (2010) 008, https://doi.org/10.1088/1475-7516/2010/02/008, arXiv:0907.3149.

[32] S. Vagnozzi, Seven hints that early-time new physics alone is not sufficient to solve the Hubble tension, Universe 9 (2023) 393, https://doi.org/10.3390/universe9090393.

[33] C. Zhang, H. Zhang, S. Yuan, S. Liu, T.J. Zhang, Y.C. Sun, Four new observational H(z) data from luminous red galaxies in the Sloan Digital Sky Survey data release seven, Res. Astron. Astrophys. 14 (2014) 1221–1233, https://doi.org/10.1088/1674-4527/14/10/002, arXiv:1207.4541.