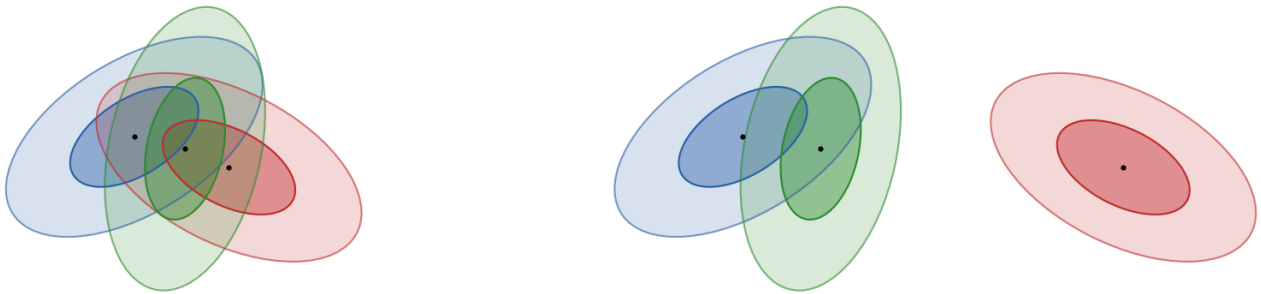
Tension metrics in cosmology

Susana J. Landau[1,2] Matias Leizerovich [1,2] Claudia Scóccola [3]

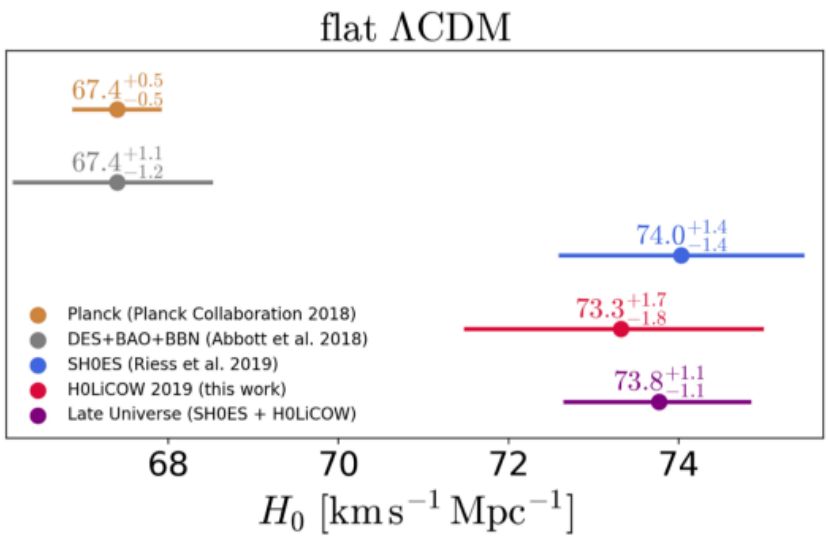
[1]Departamento de Física - Facultad de Ciencias Exactas y Naturales - UBA
[2]Instituto de Física de Buenos Aires - CONICET - Universidad de Buenos Aires
[3]Departamento de Física, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile

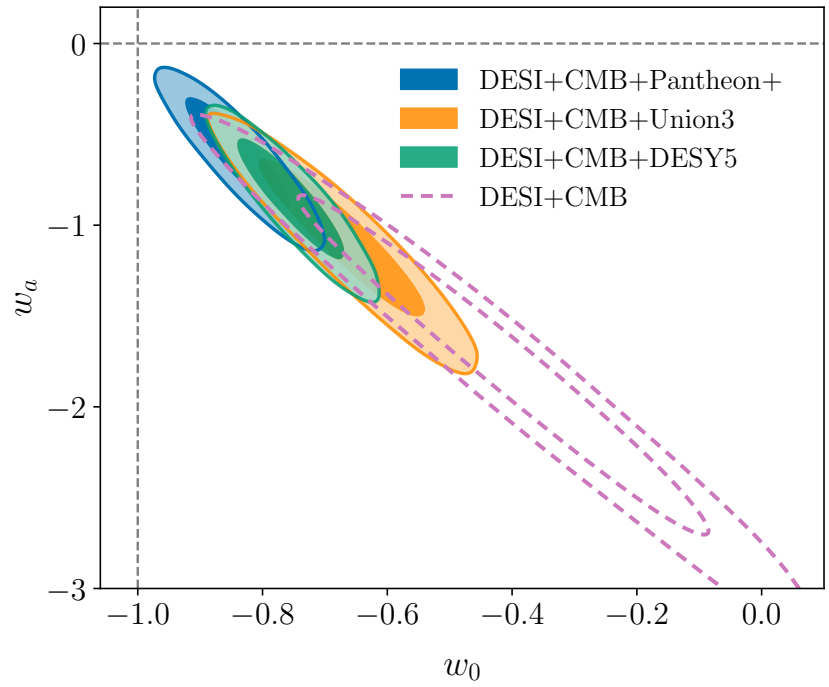
III International Meeting on Cosmology and Gravitation
Universidad de Tarapacá, Arica, Chile
20-23 October 2026



Tensions in Cosmology

- Hubble tension: H_0 from CMB vs local measurements (SH0ES,H0LiCOW,etc.)
- Recent release of the DESI collaboration points to a variable equation of state $w(z)$ for dark energy.





Outline of the talk

- Motivation: Tensions in cosmology and the shared parameter space
- Tension metrics between two datasets
- Generalized tension metrics for multiple datasets
- Geometrical interpretation of the tension
- Application to real data sets
- Conclusions

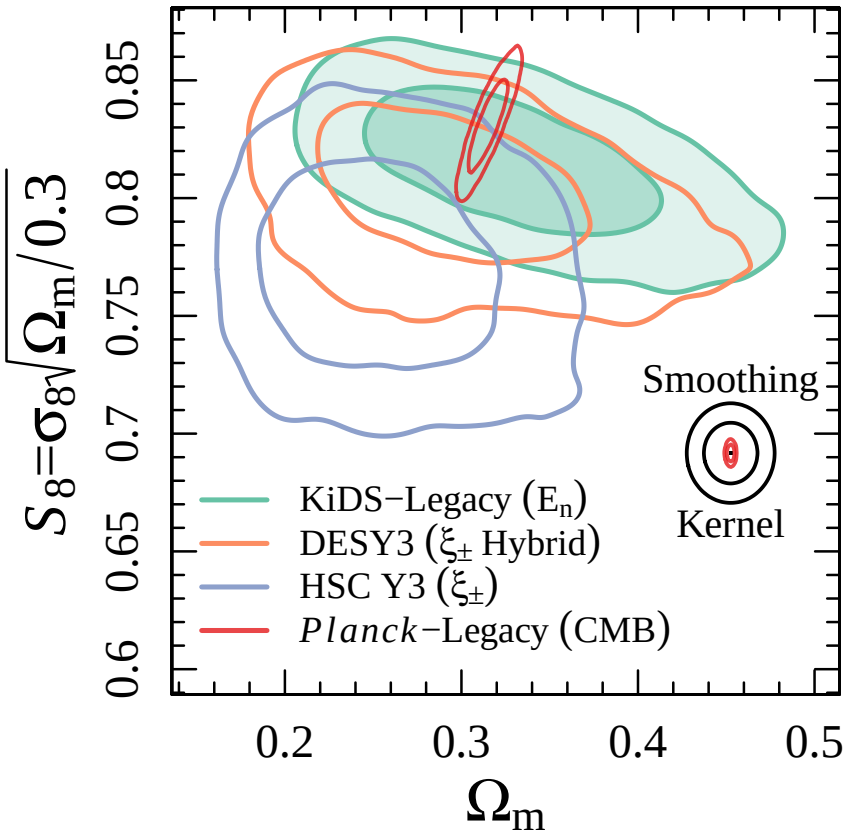
Rule of Thumb estimator (N_σ)

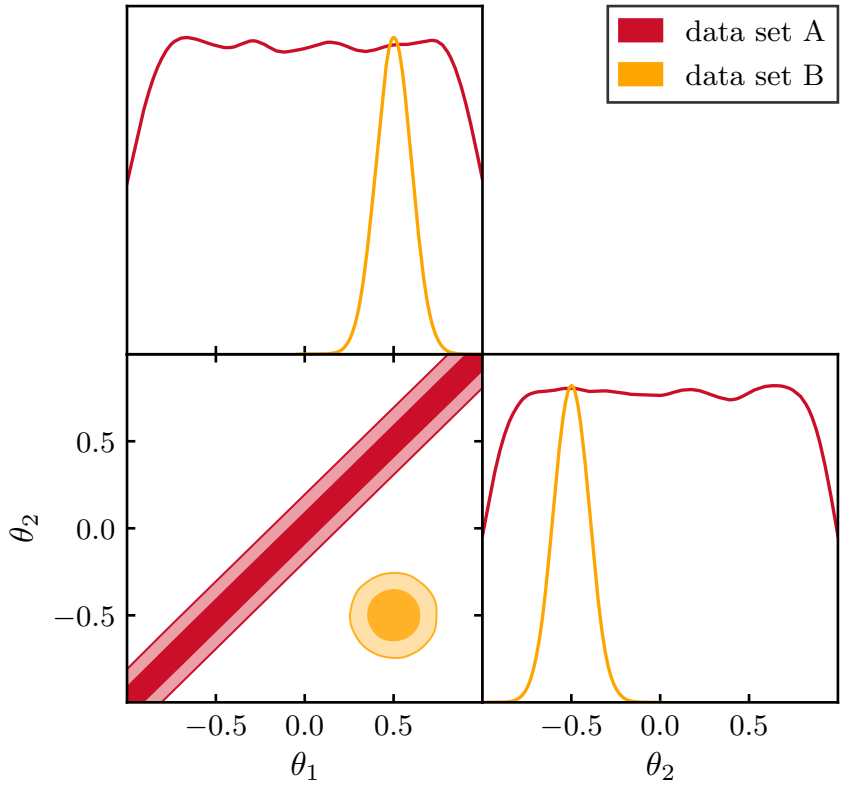
A common way to quantify discrepancies is to compare one-dimensional marginalized posteriors. For a parameter θ estimated from datasets A and B, the rule-of-thumb estimator is defined as:

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

where $\mu_{A,B}$ and $\sigma_{A,B}$ denote the means and standard deviations.

The shared parameter space

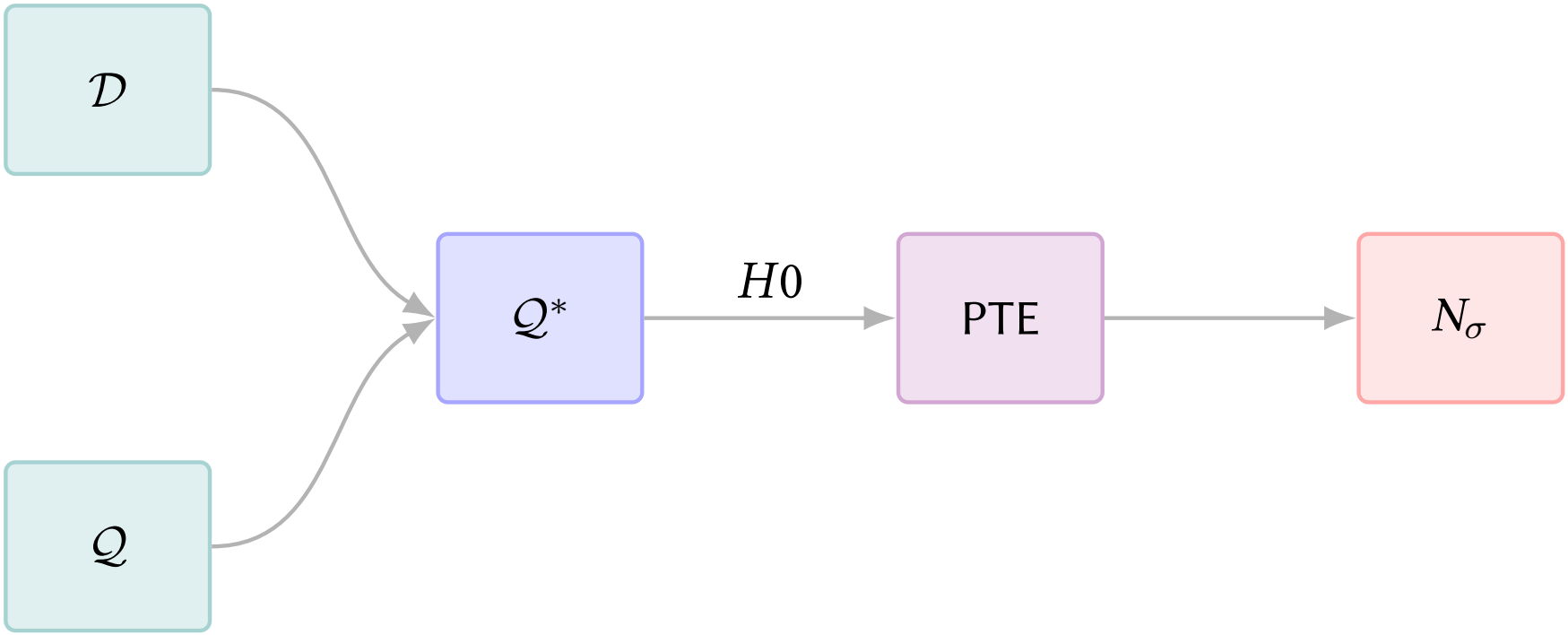




The full-dimensional tension can be larger than that inferred by looking at 1D distributions of the parameters.

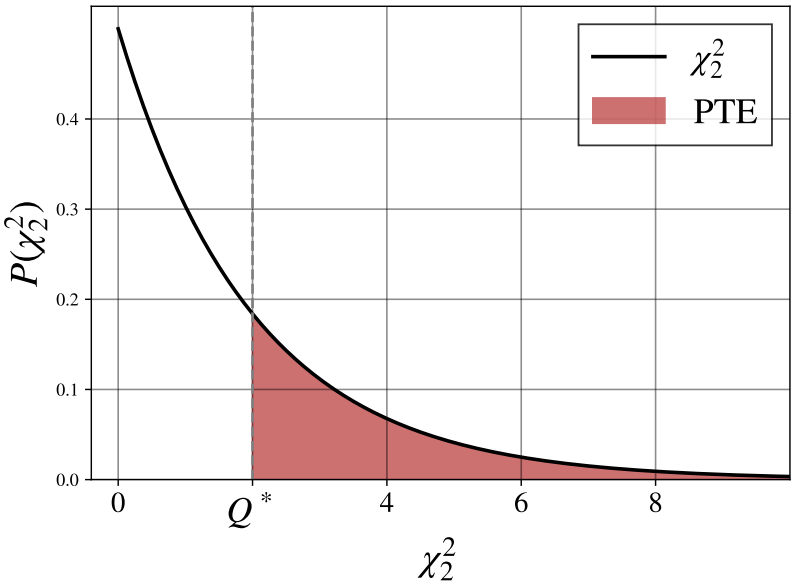
Tension metrics

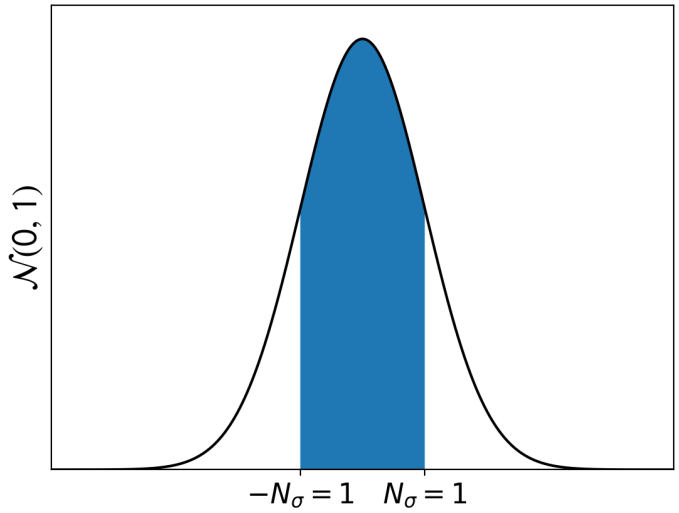
- Given an tension estimator $\mathcal{Q}$ and a data set $\mathcal{D}$, we can calculate its measured value $\mathcal{Q}^*$, and the Probability to Exceed (PTE).
- For this, we assume a null hypothesis to obtain the probability distribution that the estimator $\mathcal{Q}$ follows.



Probability to Exceed (PTE) and equivalent N_σ

The level of agreement or tension between inferred parameter values from different datasets quantified in terms of an equivalent significance N_σ,



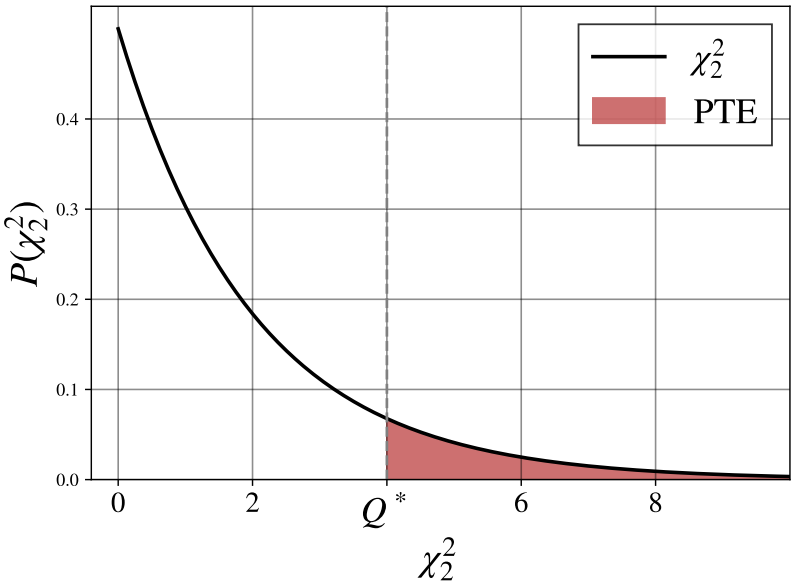


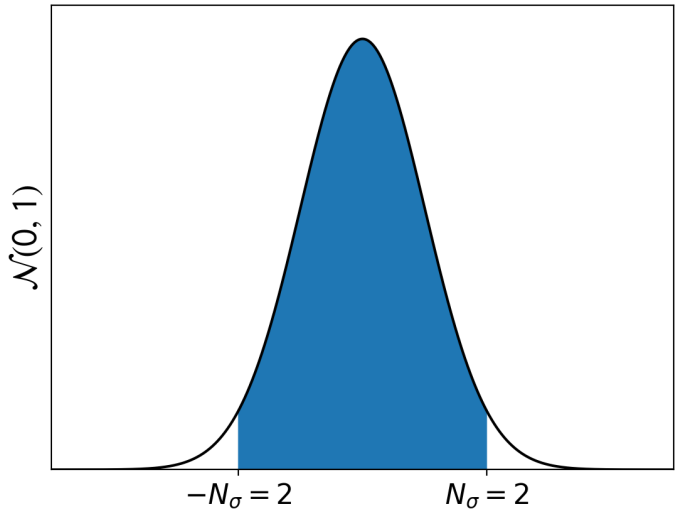
$$\text{PTE} = \int_{\mathcal{Q}^*}^{\infty} \mathcal{Q}|_{H0}(x)\,dx\,.$$

$$1 - \text{PTE} = \frac{2}{\sqrt{2\pi}} \int_0^{N_\sigma} e^{-\frac{u^2}{2}}\,du$$
$$= \text{Erf}\left(\frac{N_\sigma}{\sqrt{2}}\right)$$

Probability to Exceed (PTE) and equivalent N_σ

The level of agreement or tension between inferred parameter values from different datasets quantified in terms of an equivalent significance N_σ,



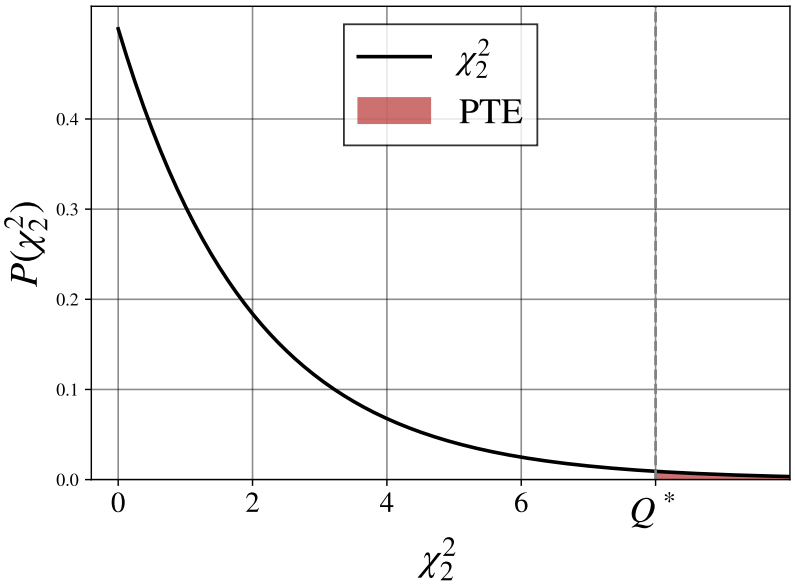


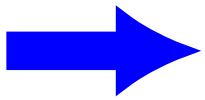
$$\text{PTE} = \int_{\mathcal{Q}^*}^{\infty} \mathcal{Q}|_{H0}(x)\,dx\,.$$

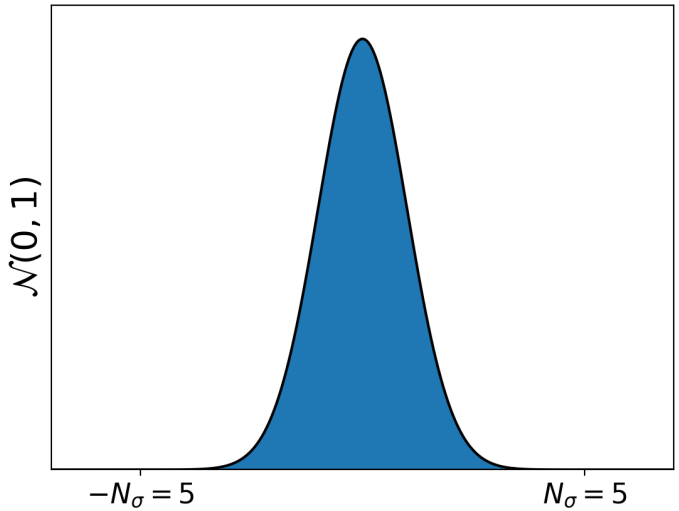
$$1 - \text{PTE} = \frac{2}{\sqrt{2\pi}} \int_0^{N_\sigma} e^{-\frac{u^2}{2}}\,du$$
$$= \text{Erf}\left(\frac{N_\sigma}{\sqrt{2}}\right)$$

Probability to Exceed (PTE) and equivalent N_σ

The level of agreement or tension between inferred parameter values from different datasets quantified in terms of an equivalent significance N_σ,







$$\text{PTE} = \int_{\mathcal{Q}^*}^{\infty} \mathcal{Q}|_{H0}(x)\,dx\,.$$

$$1 - \text{PTE} = \frac{2}{\sqrt{2\pi}} \int_0^{N_\sigma} e^{-\frac{u^2}{2}}\,du$$
$$= \text{Erf}\left(\frac{N_\sigma}{\sqrt{2}}\right)$$

Tension metrics betweem two datasets

- Metrics based on the bayesian evidence:
  - Bayesian ratio statistics (R)
  - Suspiciousness (S)
  - see Handley & Lemos Phys. Rev. D100, 043504, (2019).
- Metrics based on posterior distributions:
  - Parameter differences in standard form ($Q_{\rm DM}$)
  - Parameter differences in updated form ($Q_{\rm UDM}$)
  - Goodness-of-fit loss ($Q_{\rm DMAP}$)
  - Non gaussian metrics: Exact Parameter Shift
  - see M.Leizerovich, Landau & Scóccola Phys.Lett B855, 138844 (2024)

Tension metrics betweem two datasets

- Metrics based on the bayesian evidence:
  - Bayesian ratio statistics (R)
  - Suspiciousness (S)
  - see Handley & Lemos Phys. Rev. D100, 043504, (2019).
- Metrics based on posterior distributions:
  - Parameter differences in standard form ($Q_{\rm DM}$)
  - Parameter differences in updated form ($Q_{\rm UDM}$)
  - Goodness-of-fit loss ($Q_{\rm DMAP}$)
  - Non gaussian metrics: Exact Parameter Shift
  - see M.Leizerovich, Landau & Scóccola Phys.Lett B855, 138844 (2024)

- see The CosmoVerse White Paper: Addressing observational tensions in cosmology with systematics and fundamental physics. Phys. Dark Univ. 49 (2025) 101965 (Section 3.1.8)

Parameter differences in standard form ($Q_{\rm DM}$)

This estimator calculates the statistical difference between the inferred parameters given by dataset A (θ_A) and the ones given by dataset B (θ_B), taking into account the covariance of both datasets. It is defined as

$$\boxed{Q_{\rm DM} = (\hat{\theta}_B - \hat{\theta}_A)^T (\hat{C}_B + \hat{C}_A)^{-1} (\hat{\theta}_B - \hat{\theta}_A)}, \tag{2}$$

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. If both posterior distributions are gaussian, $Q_{\rm DM}$ follows a χ^2_ν distribution with $\nu = {\rm Rank}[\hat{\rm C}_{\rm B} + \hat{\rm C}_{\rm A}]$ degrees of freedom.

This estimator quantifies the distance between the means of the independent posteriors weighted by their covariance matrix. It can be regarded as a generalization of the rule-of-thumb estimator N_σ to the full parameter space.

Parameter differences in updated form ($Q_{\rm UDM}$)

This estimator computes 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

$$\boxed{Q_{\rm UDM} = (\hat{\theta}_{AB} - \hat{\theta}_A)^T (\hat{C}_{AB} - \hat{C}_A)^{-1} (\hat{\theta}_{AB} - \hat{\theta}_A)} \tag{3}$$

where $\hat{C}_{AB}$ is the covariance matrix obtained from the statistical analysis with the joint datasets. If the inferred parameters for both datasets A and A+B are gaussian distributed, $Q_{\rm UDM}$ has a χ^2_ν distribution with $\nu = {\rm rank}[(\hat{\rm C}_{\rm AB} - \hat{\rm C}_{\rm A})]$ degrees of freedom

This estimator metric quantifies how a new dataset updates another one.

Goodness-of-fit loss ($Q_{\rm DMAP}$)

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.

$$Q_{\rm DMAP} = 2ln\mathcal{L}_A(\theta_{pA}) + 2ln\mathcal{L}_B(\theta_{pB}) - 2ln\mathcal{L}_{A+B}(\theta_{pA+B}) \tag{4}$$

where $\theta_p^{A/B}$ are the Maximum a posteriori (MAP) parameters considering the dataset A/B.

In the case in which the likelihood and the posterior are gaussian distributed, the estimator $Q_{\rm DMAP}$ has a distribution χ^2 with $\Delta\nu$ degrees of freedom $\Delta\nu = \nu^A + \nu^B - \nu^{A+B}$.The number of degrees of freedom is defined as $\nu = N - tr[C_\Pi^{-1} C_p]$ where N is the number of data points and C_Π, C_p are the covariance matrix of the Prior and Posterior distributions

This estimator quantifies the distance between the MAP evaluating the likelihood of the joint posterior and the one using the datasets separately

Non gaussian metrics: Exact Parameter Shift

This estimator computes 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.

$$P(\Delta\theta) = \int P_A(\theta) P_B(\theta - \Delta\theta)\, d\theta\,, \tag{5}$$

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 $\Delta\theta = 0$:

$$\boxed{\Delta = \int_{P(\Delta\theta)>P(0)} P(\Delta\theta)\, d\Delta\theta.} \tag{6}$$

Here the assigned probability to identify the tension is $1 - PTE = \Delta$.

This estimator calculates the distance between the two posteriors in the full parameter space (like $Q_{\rm DM}$), without assuming any particular distribution.

- $P(\Delta\theta)$ for three examples with real data, and $\theta = (H_0, \Omega_m)$

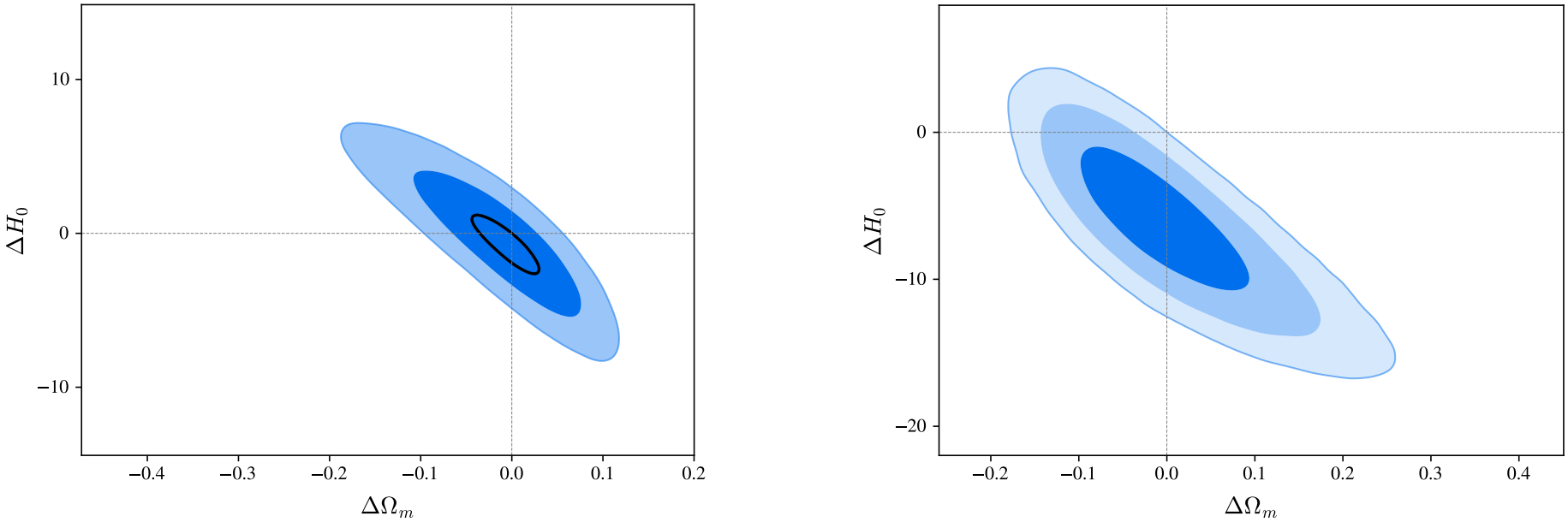
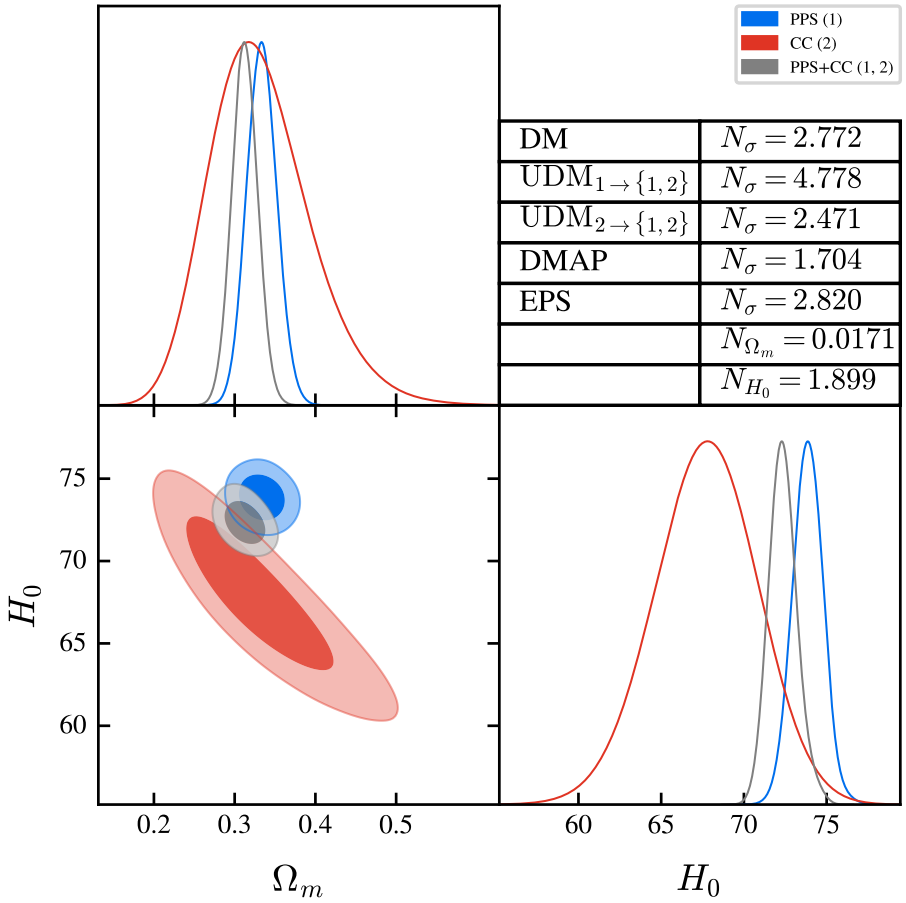


FIGURA: Left: CC vs Planck18, Right: CC vs PPS

$$1 - \text{PTE} = \Delta = \int_{P(\Delta\theta) > P(0)} P(\Delta\theta)\, d\Delta\theta. \tag{7}$$

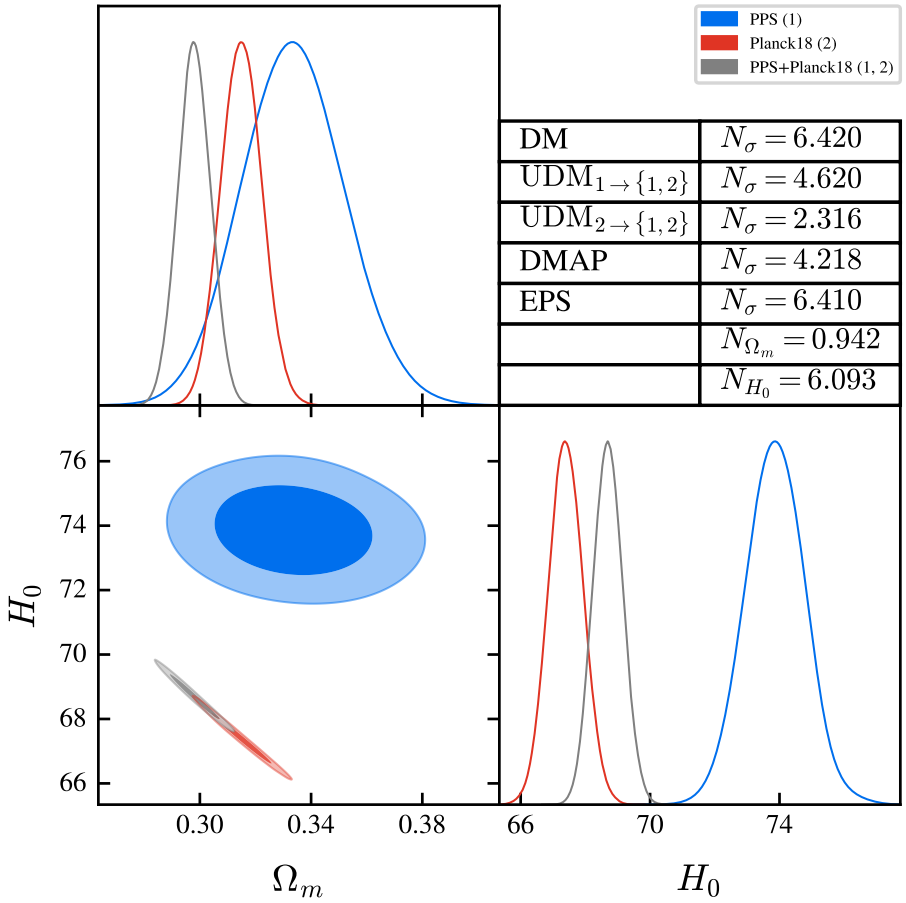
- CC: Cosmic Chronometers
- PPS: type Ia supernovae from the Pantheon Plus Compilation + Cepheid data from SH0ES collaboration
- Planck18: CMB data from Planck

Applying tension metrics between two data sets to real data



PPS (Pantheon Plus + SH0ES) vs Cosmic Chronometers

Applying tension metrics between two data sets to real data



CMB (Planck 2018) vs Pantheon Plus + SH0ES (PPS)

Generalization of tension estimator

Leizerovich, Landau, & Scóccola arXiv:2512.06086

For each pair $k = (i, j)$ of datasets, we define the *tension vector* $\vec{r}_k$ as

$$\vec{r}_k = \frac{1}{\sqrt{\hat{C}_i + \hat{C}_j}} (\vec{\bar{\theta}}_i - \vec{\bar{\theta}}_j) \quad (i \neq j). \tag{8}$$

where $\vec{\bar{\theta}}_i$ and $\hat{C}_i$ are the mean and covariance matrix of the posterior distribution of dataset i in the shared parameter space.

Generalization of tension estimator

Leizerovich, Landau, & Scóccola arXiv:2512.06086

For each pair $k = (i, j)$ of datasets, we define the *tension vector* $\vec{r}_k$ as

$$\vec{r}_k = \frac{1}{\sqrt{\hat{C}_i + \hat{C}_j}} (\vec{\bar{\theta}}_i - \vec{\bar{\theta}}_j) \quad (i \neq j). \tag{8}$$

where $\vec{\bar{\theta}}_i$ and $\hat{C}_i$ are the mean and covariance matrix of the posterior distribution of dataset i in the shared parameter space. The vectors $\vec{r}_k$ are treated as elements of a parameter-difference space and identified with the positions of N_p particles of mass m, which encode the weights of each dataset pair. As uniform weights suffice, we set $m = 1$. The center of mass is

$$\vec{R}_{CM} = \frac{\sum_{k=1}^{N_p} m_k \vec{r}_k}{\sum_{k=1}^{N_p} m_k} = \frac{\sum_{k=1}^{N_p} \vec{r}_k}{N_p} = 0\,, \tag{9}$$

which vanishes by antisymmetry.

We define our new estimator of the tension $\mathcal{Q}$ as the dispersion of the $\vec{r}_k$ vectors w.r.t. $\vec{R}_{CM}$, i.e., the sum of the squared tension vectors

$$\boxed{\mathcal{Q} \equiv \frac{1}{N_p} \sum_{k=1}^{N_p} |\vec{r}_k|^2 .} \tag{10}$$

We define our new estimator of the tension $\mathcal{Q}$ as the dispersion of the $\vec{r}_k$ vectors w.r.t. $\vec{R}_{CM}$, i.e., the sum of the squared tension vectors

$$\boxed{\mathcal{Q} \equiv \frac{1}{N_p} \sum_{k=1}^{N_p} |\vec{r}_k|^2 .} \tag{10}$$

Since the random vectors $\vec{r}_k$ are distributed as $\vec{r}_k \sim \mathcal{N}(\frac{1}{\sqrt{\hat{C}_i+\hat{C}_j}}(\vec{\theta}_i - \vec{\theta}_j), \mathcal{I}_{D\times D})$, we adopt our null hypothesis, $H0$, as

$$\frac{1}{\sqrt{\hat{C}_i + \hat{C}_j}}(\vec{\theta}_i - \vec{\theta}_j) \simeq \vec{0} \;\; \forall i, j, i \neq j \tag{11}$$

In general, an analytical expression for the probability distribution function associated to $H0$ does not exist, since the random vectors $\vec{r}_i|_{H0}$ are not independent of each other. However, when all covariances are equal, the distribution under $H0$ is

$$\mathcal{Q}|_{H0} \sim \Gamma\left(\frac{(N-1)D}{2}, \frac{2}{N-1}\right) \tag{12}$$

where D is the dimension of the shared parameter space.

- From the later, we define the dispersion tensor as

$$\mathcal{C}_{ab} = \frac{1}{N_p} \sum_{k=1}^{N_p} r_k^a \cdot r_k^b \,. \tag{13}$$

where $\mathcal{C}_{ab}$ quantifies the dispersion of $\vec{r}_k$ with respect to the origin. The indices (a, b) run over the directions of the shared parameter space $(1, .., D)$.

- The dispersion tensor is related to $\mathcal{Q}$ through

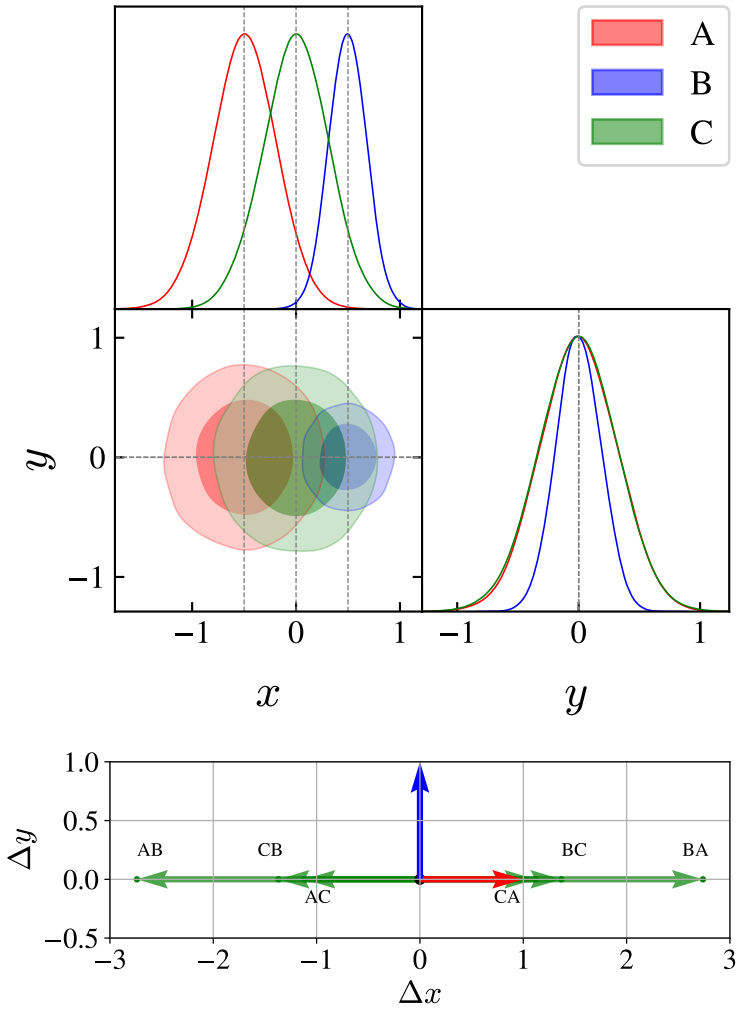
$$\mathcal{Q} = \mathrm{Tr}(\mathcal{C}) = \sum_{\alpha}^{D} \lambda_\alpha \,, \tag{14}$$

where $\{\lambda_\alpha\}$ are the eigenvalues of the tensor $\mathcal{C}$.

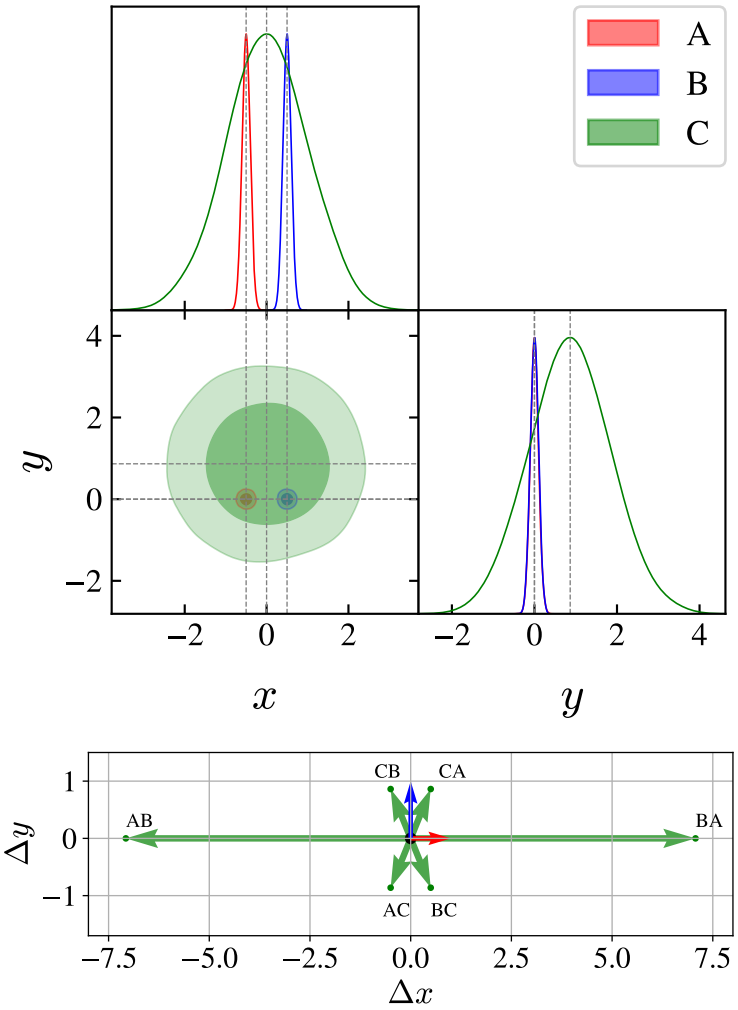
- $\det(\mathcal{C}) = 0$ indicates that at least one eigenvalue vanishes, meaning that the tension effectively lies in a lower-dimensional subspace.
- A second useful descriptor is the eccentricity

$$\mathrm{Ecc} = \sqrt{1 - \frac{\lambda_{\mathrm{mín}}}{\lambda_{\mathrm{máx}}}} \,, \tag{15}$$

which quantifies the anisotropy of $\mathcal{C}$. When $\mathrm{Ecc} \simeq 0$, the tension vectors have comparable magnitudes. In contrast, $\mathrm{Ecc} \simeq 1$ reflects large differences in vector lengths, indicating unequal contributions from the different dataset pairs to the global tension.

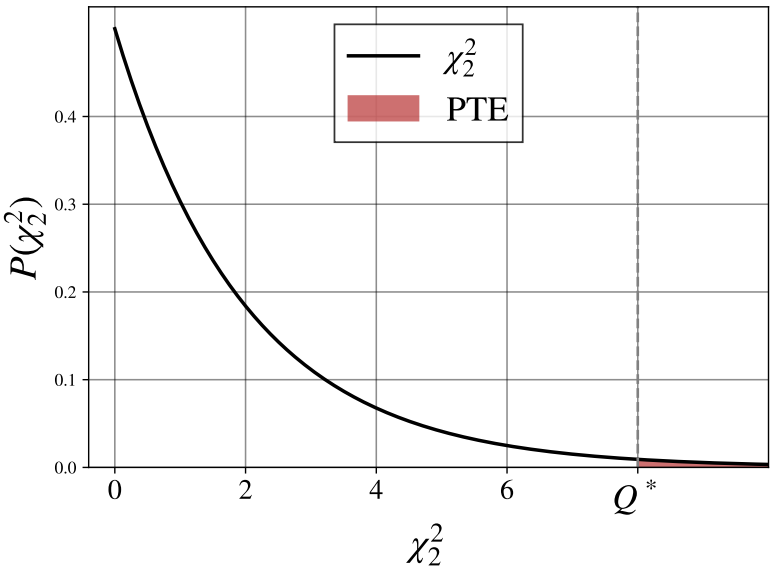


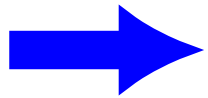
$N_\sigma = 1.5$

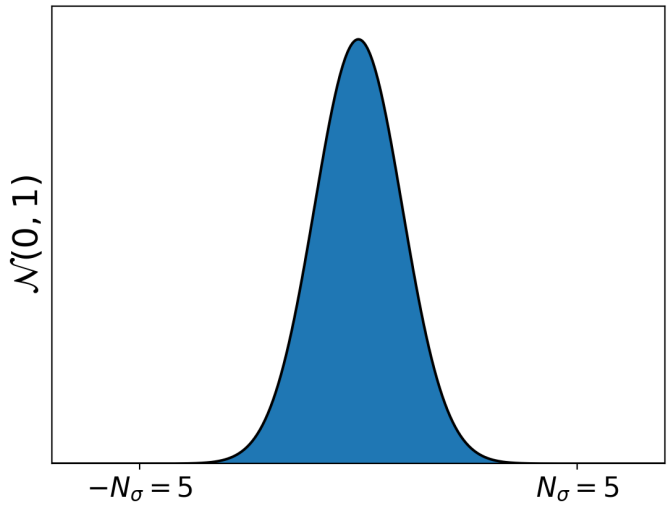


$N_\sigma = 4{,}6$

Recalling the relation between the Probability to Exceed (PTE) and equivalent N_σ





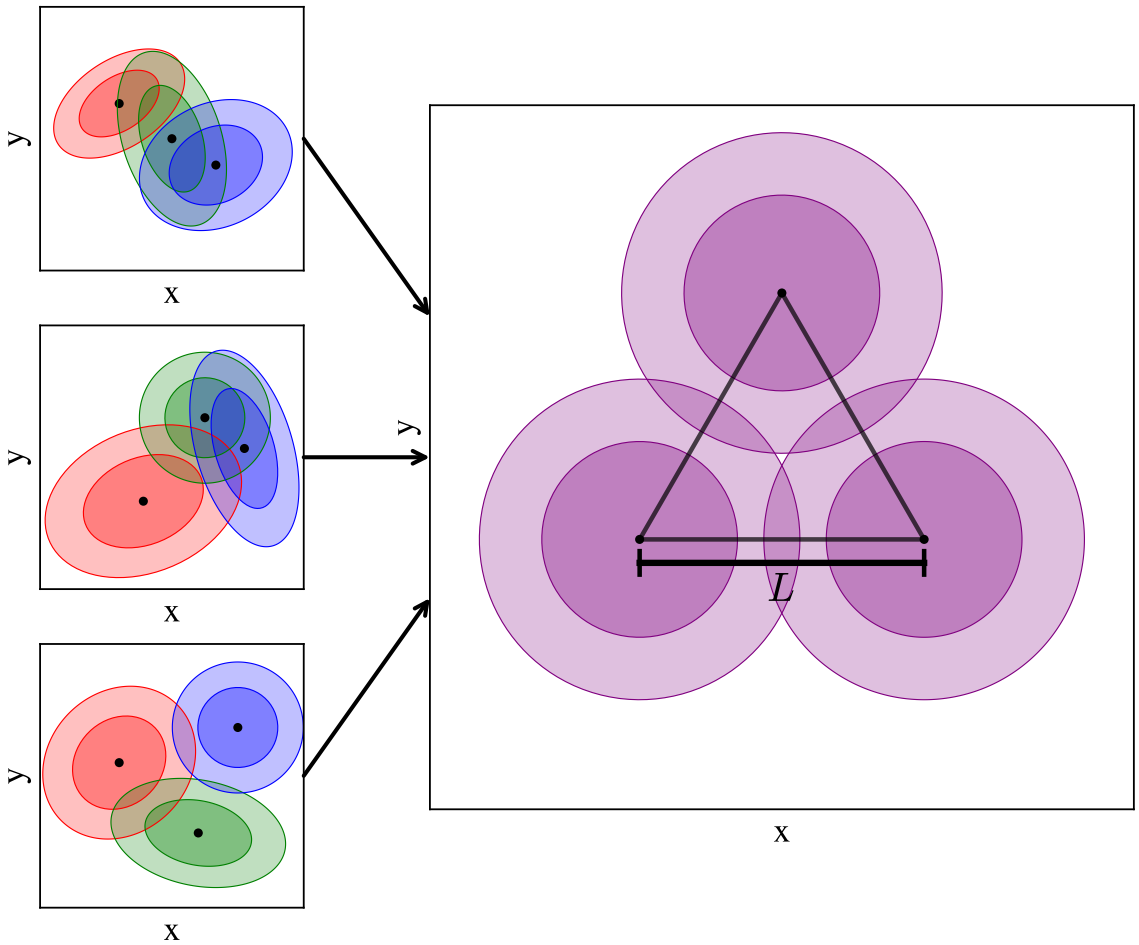


$$\text{PTE} = \int_{\mathcal{Q}^*}^{\infty} \mathcal{Q}|_{H0}(x)\,dx\,.$$

$$1 - \text{PTE} = \frac{2}{\sqrt{2\pi}} \int_0^{N_\sigma} e^{-\frac{u^2}{2}}\,du$$

However, care is required when comparing PTEs (and their associated N_σ values) that arise from physically distinct situations involving different test statistics, or different dimensionalities of the shared parameter space.

Geometrical Interpretation of Tension

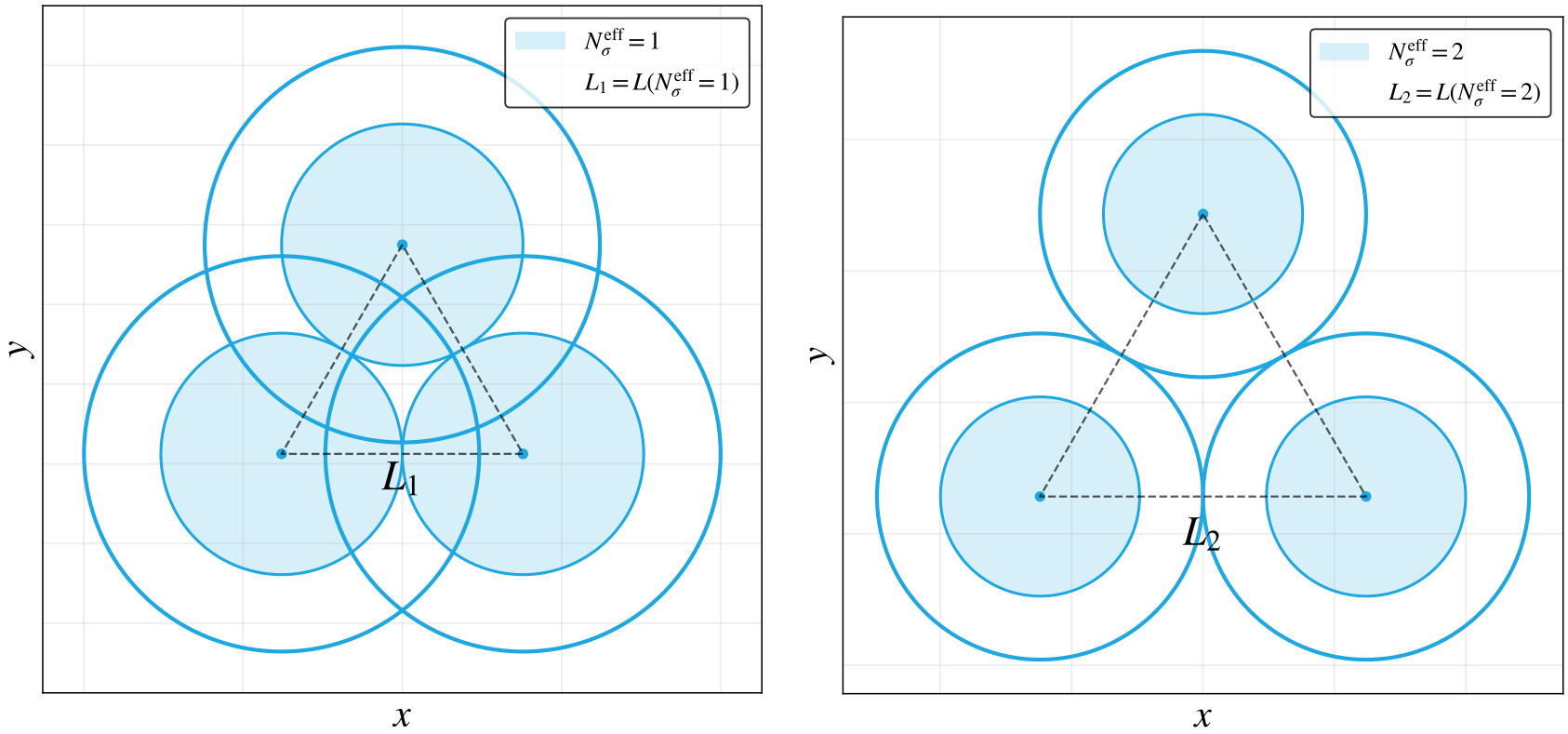


We map any arbitrary configuration of three 2D posteriors onto a reference setup consisting of three posteriors with identical covariance, and means located at the vertices of an equilateral triangle of side lenght L.

The mapping is constructed to preserve the tension level. Therefore, the value L of the reference configuration is chosen such that the corresponding PTE is the same as the one we want to study.

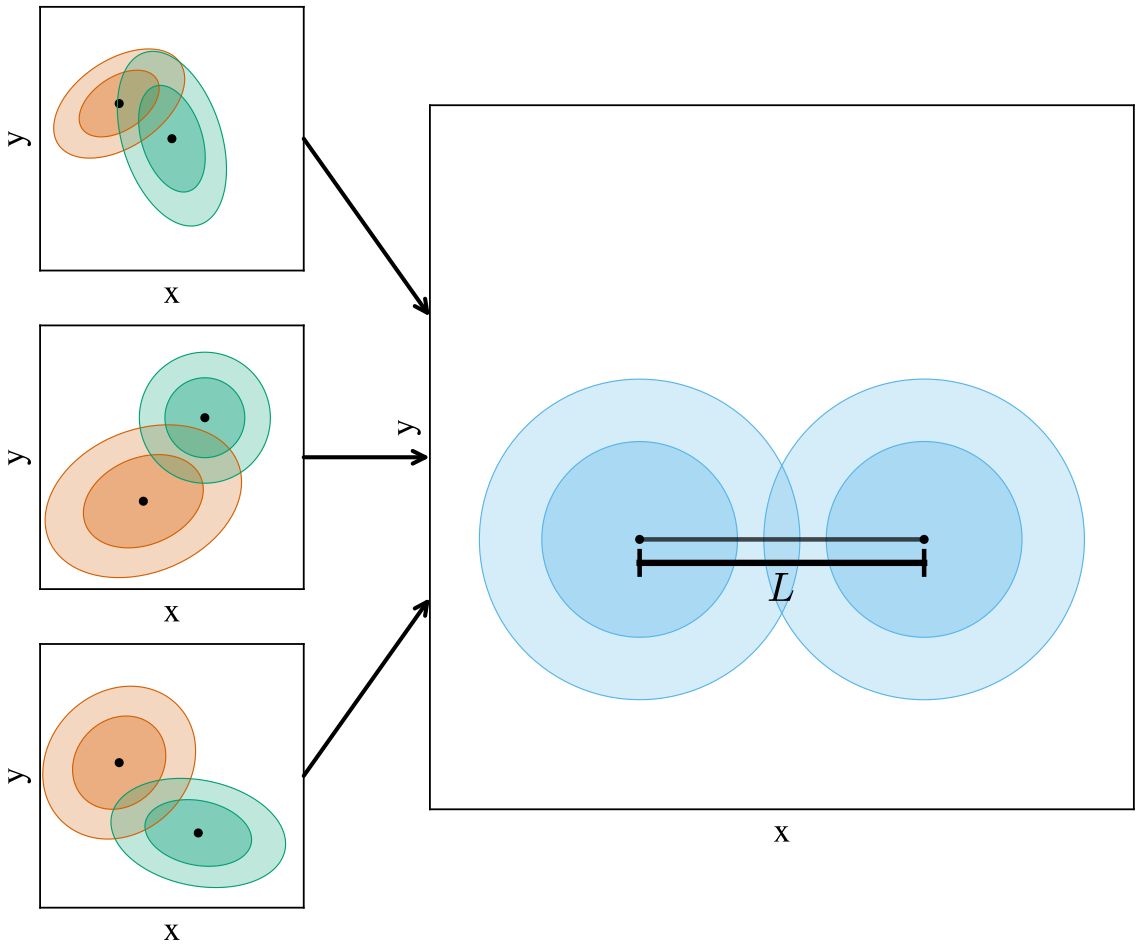
Given a number of datasets N and a shared parameter space of dimension D, there is unique relation between L and N_σ^{eff}.

The equivalent N_σ^{eff}



- For $N = 3$ and $D = 2$, we can define the following scale:
  $L(N_\sigma^{\text{eff}} = 1) = 2{,}143$
  $L(N_\sigma^{\text{eff}} = 2) = 3{,}516$
  $L(N_\sigma^{\text{eff}} = 3) = 4{,}864$

Geometrical Interpretation of Tension

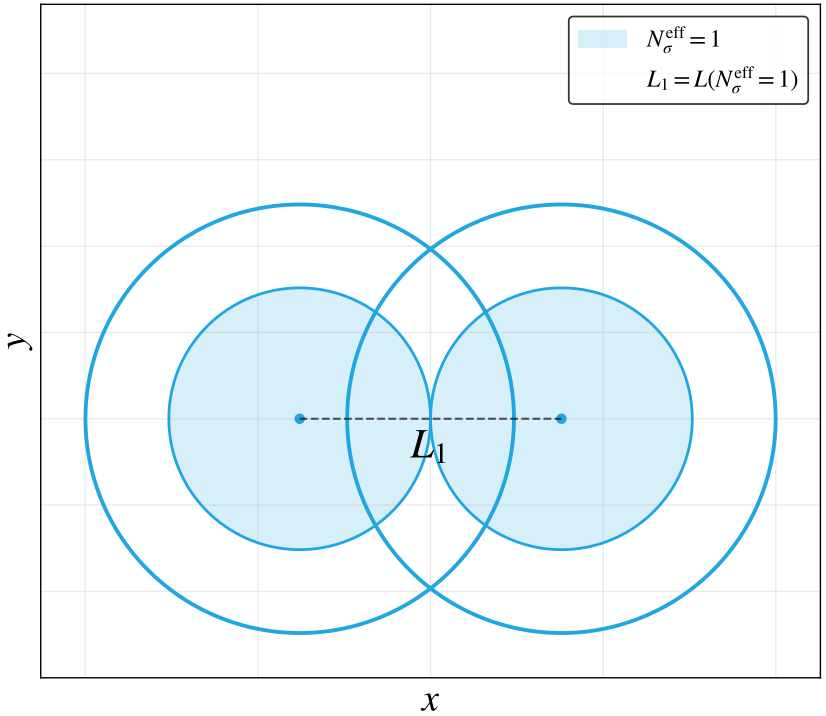


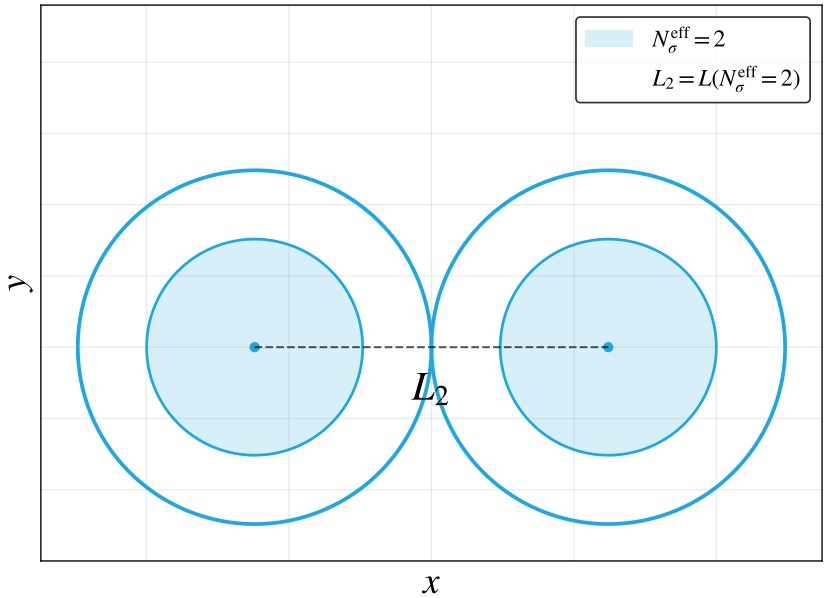
We map any arbitrary configuration of two 2D posteriors onto a reference setup consisting of two posteriors with identical covariance, and means located at the vertices of an equilateral triangle of side lenght L.

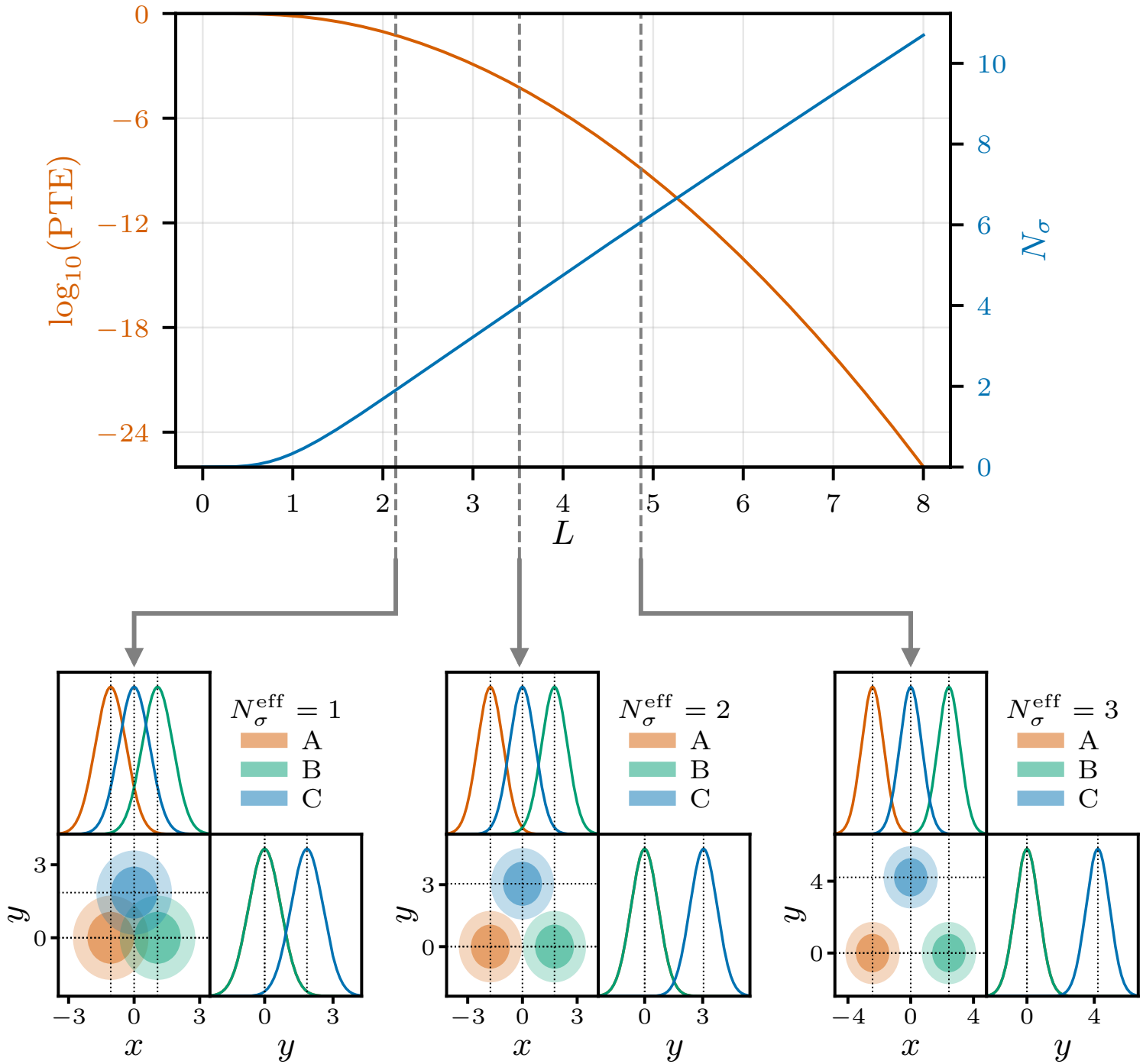
The mapping is constructed to preserve the tension level. Therefore, the value L of the reference configuration is chosen such that the corresponding PTE is the same as the one we want to study.

Given a number of datasets N and a shared parameter space of dimension D, there is unique relation between L and N_σ^{eff}.

The equivalent N_σ^{eff}

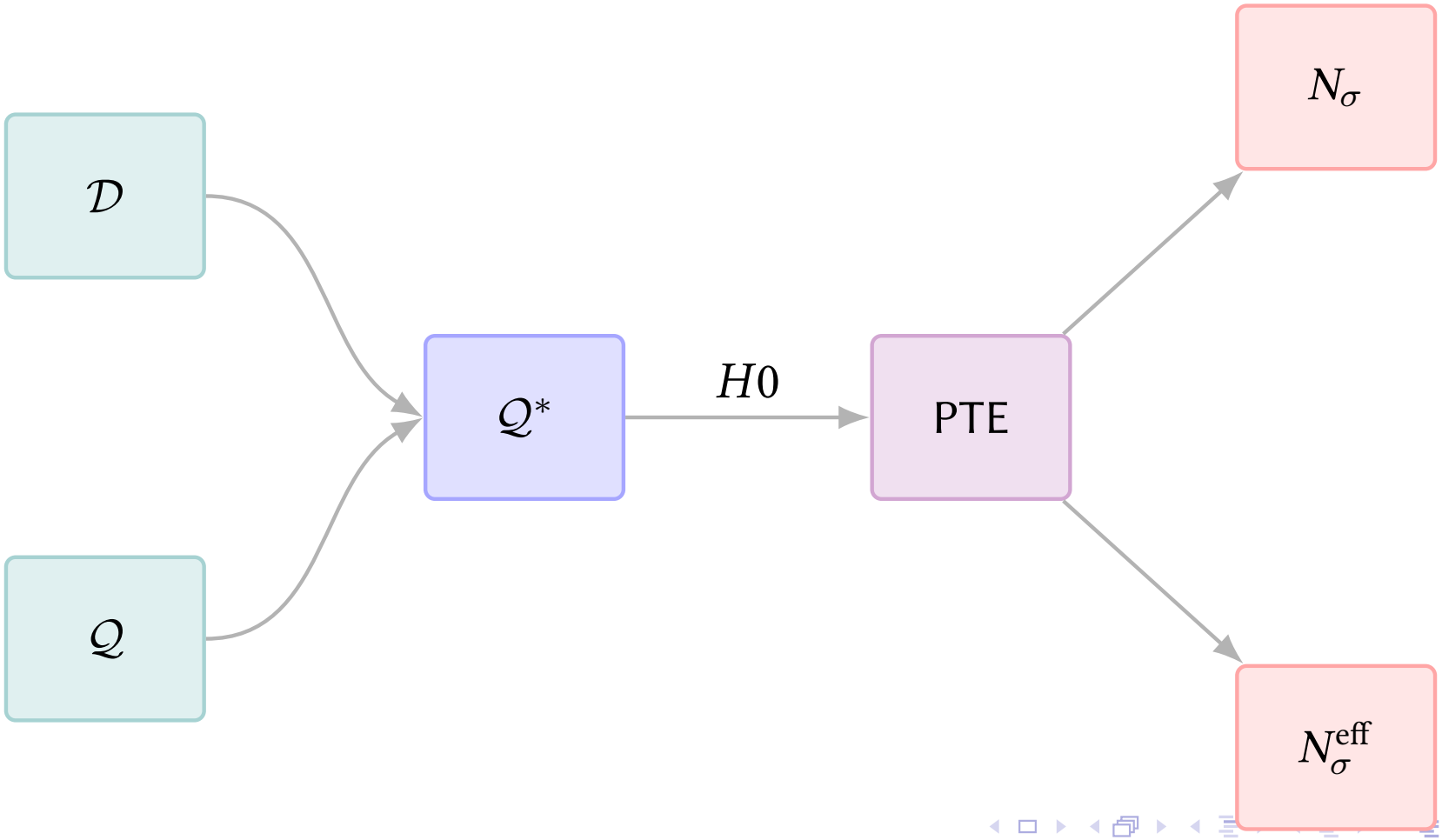




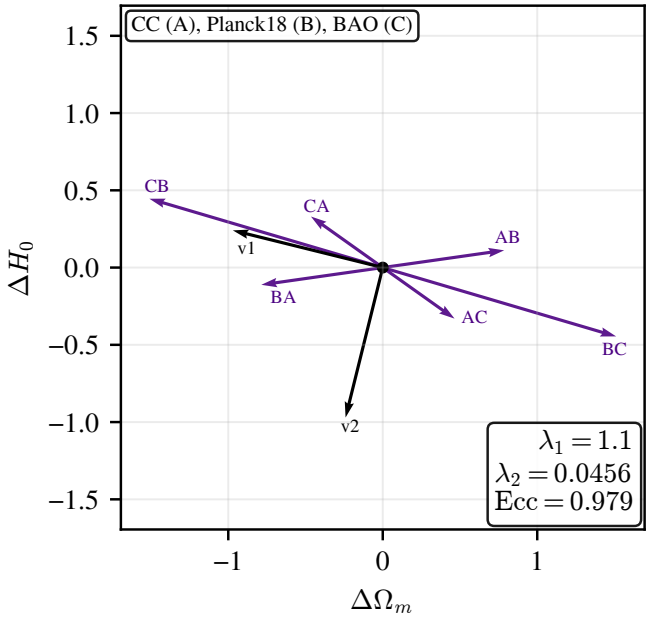

0
−6
−12
−18
−24
$\log_{10}(\mathrm{PTE})$
10
8
6
4
2
0
N_σ
0 1 2 3 4 5 6 7 8
L
$N_\sigma^{\mathrm{eff}} = 1$
$N_\sigma^{\mathrm{eff}} = 2$
$N_\sigma^{\mathrm{eff}} = 3$
A
B
C
x
y

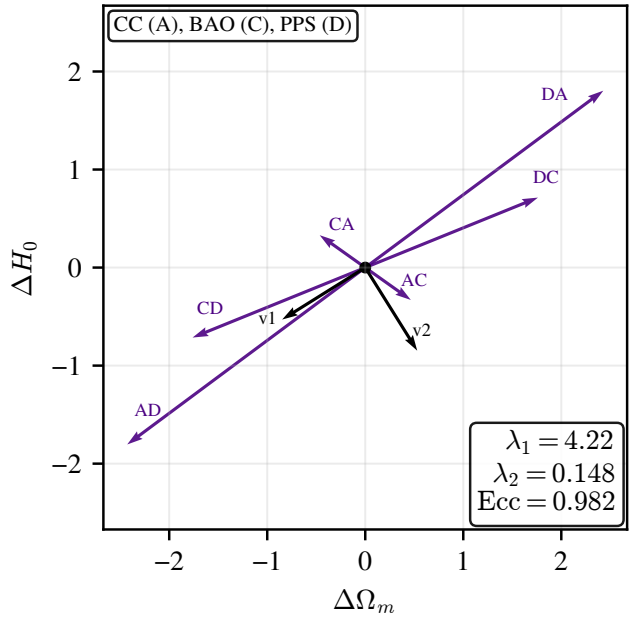
Tension metrics with geometrical interpretation

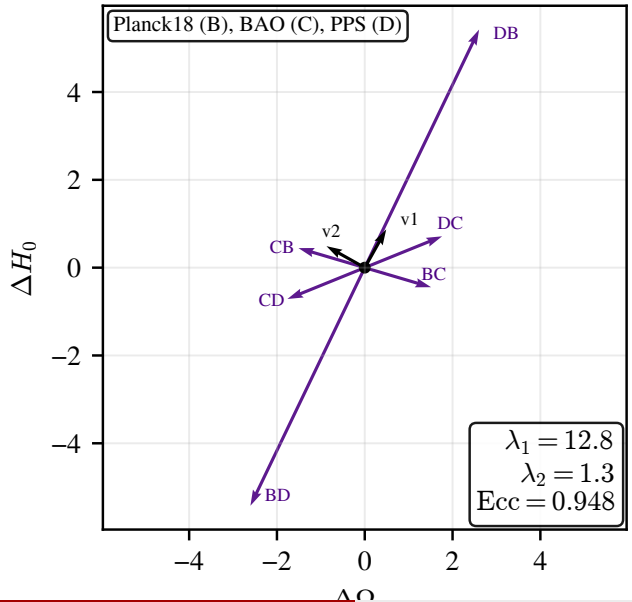
- For a given estimator $\mathcal{Q} \to Q^* \to$ PTE. From this PTE, we can associate the L that corresponds to the equivalent triangle configuration and obtain $N_\sigma^{\rm eff}$

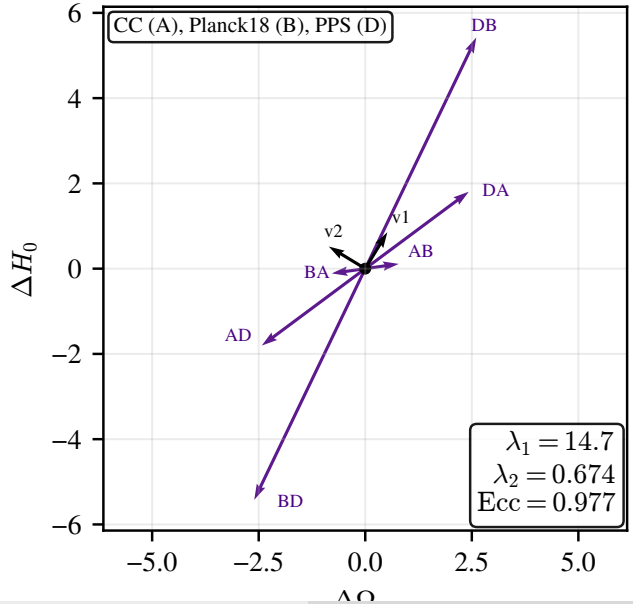


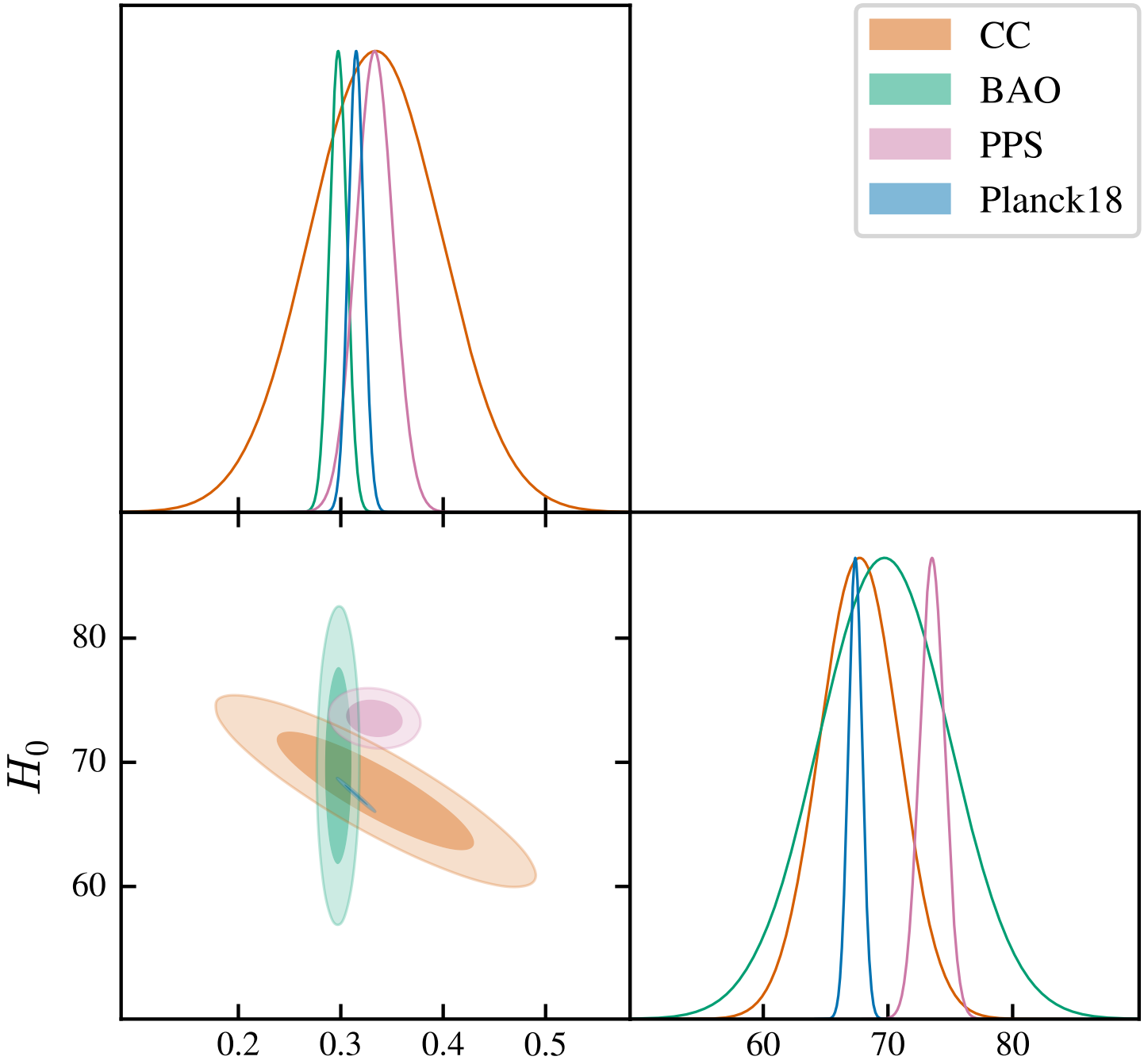
Tension vector in real datasets










CC
BAO
PPS
Planck18
H_0
80
70
60
0.2
0.3
0.4
0.5
60
70
80

Results

Datasets	N_σ	L	N_σ^{eff}
CC + Planck + BAO	0,429	1,09	0,327
CC + BAO + PPS	1,78	2,06	0,944
Planck + BAO + PPS	4,10	3,57	2,04
CC + Planck + PPS	4,26	3,68	2,12
CC + Planck	0,346	0,794	0,185
CC + BAO	0,189	0,57	0,100
CC + PPS	2,57	3,03	1,64
Planck + PPS	5,68	6,03	3,86
BAO + PPS	1,39	1,90	0,833

Cuadro: Tension estimators for real data sets: usual N_σ vs our geometrical indicator N_σ^{eff} derived from the equivalent configuration distance L. PPS: Pantheon Plus + SH0ES , CMB: Planck2018, BAO: DESI DR2 .

Final remarks

- Tension metrics are useful tools to quantify the level of agreement or disagreement between datasets in a given model. However, they do not provide the reason for the tension; they are only tools to quantify and better understand the tension.

Final remarks

- Tension metrics are useful tools to quantify the level of agreement or disagreement between datasets in a given model. However, they do not provide the reason for the tension; they are only tools to quantify and better understand the tension.
- We have presented a new estimator that enables the simultaneous assessment of tensions among multiple datasets.
- We have also provided a geometric interpretation of the tension for two and three dataset cases in a 2D shared parameter space.
- While these tools were developed specifically for evaluating cosmological tensions, they may find broader application in other research fields where discrepancies between datasets emerge.

Final remarks

- Tension metrics are useful tools to quantify the level of agreement or disagreement between datasets in a given model. However, they do not provide the reason for the tension; they are only tools to quantify and better understand the tension.
- We have presented a new estimator that enables the simultaneous assessment of tensions among multiple datasets.
- We have also provided a geometric interpretation of the tension for two and three dataset cases in a 2D shared parameter space.
- While these tools were developed specifically for evaluating cosmological tensions, they may find broader application in other research fields where discrepancies between datasets emerge.

Thank you !!! **Gracias!!!**