DESC Seminar 12/2024

Likelihood considerations for Stage IV cluster abundance cosmology

A focus on Super-Sample Covariance

Constantin Payerne
Postdoc at Irfu, CEA-Saclay

Collaborators (DESC CL WG)
Calum Murray (DAp, CEA-Saclay)
Céline Combet (LPSC, CNRS-IN2P3)
Mariana Penna-Lima (Dep. de Física, Univ. Estadual de Londrina)
…

Cosmology with cluster counts

The abundance of galaxy clusters

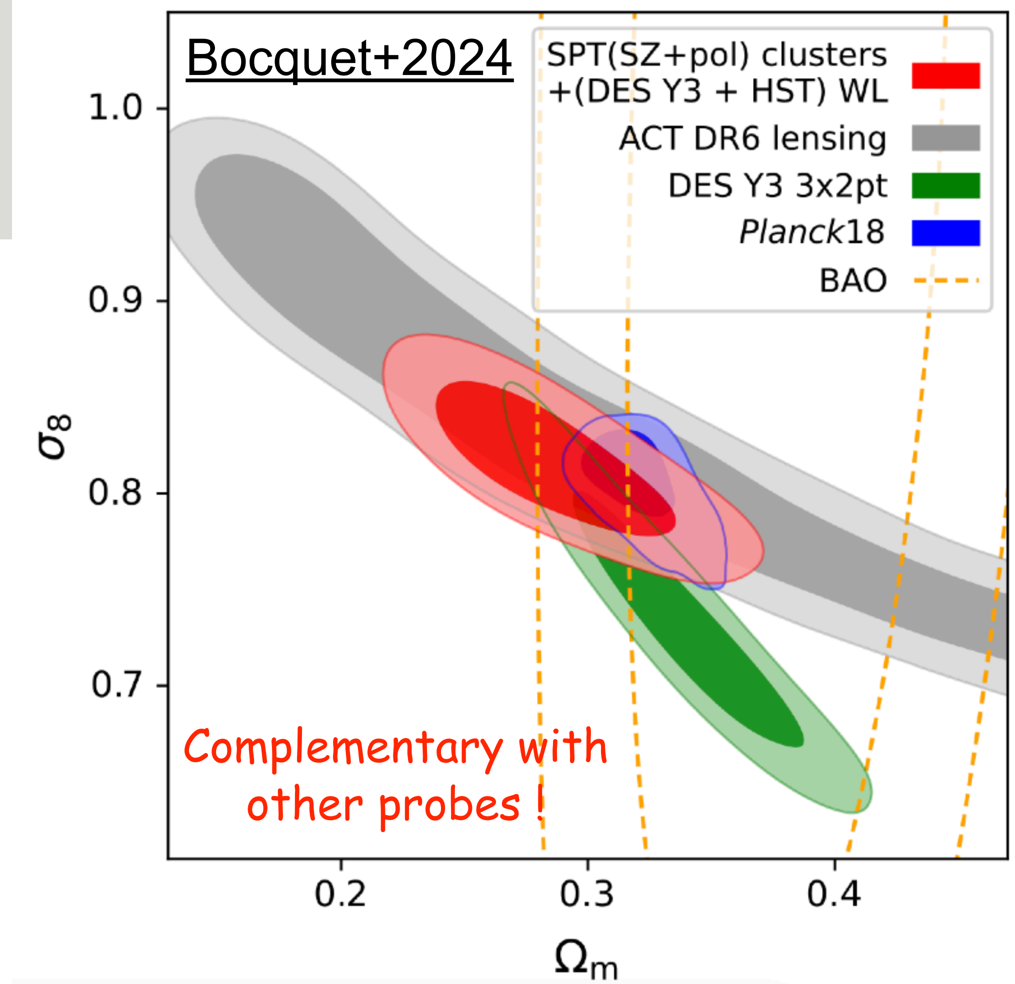
- Massive bound systems $M > 10^{14}\ M_{\odot}$ and detected in λ: X-rays, mm, optical
- Connects proxy-cluster counts to cosmology via a scaling relation:

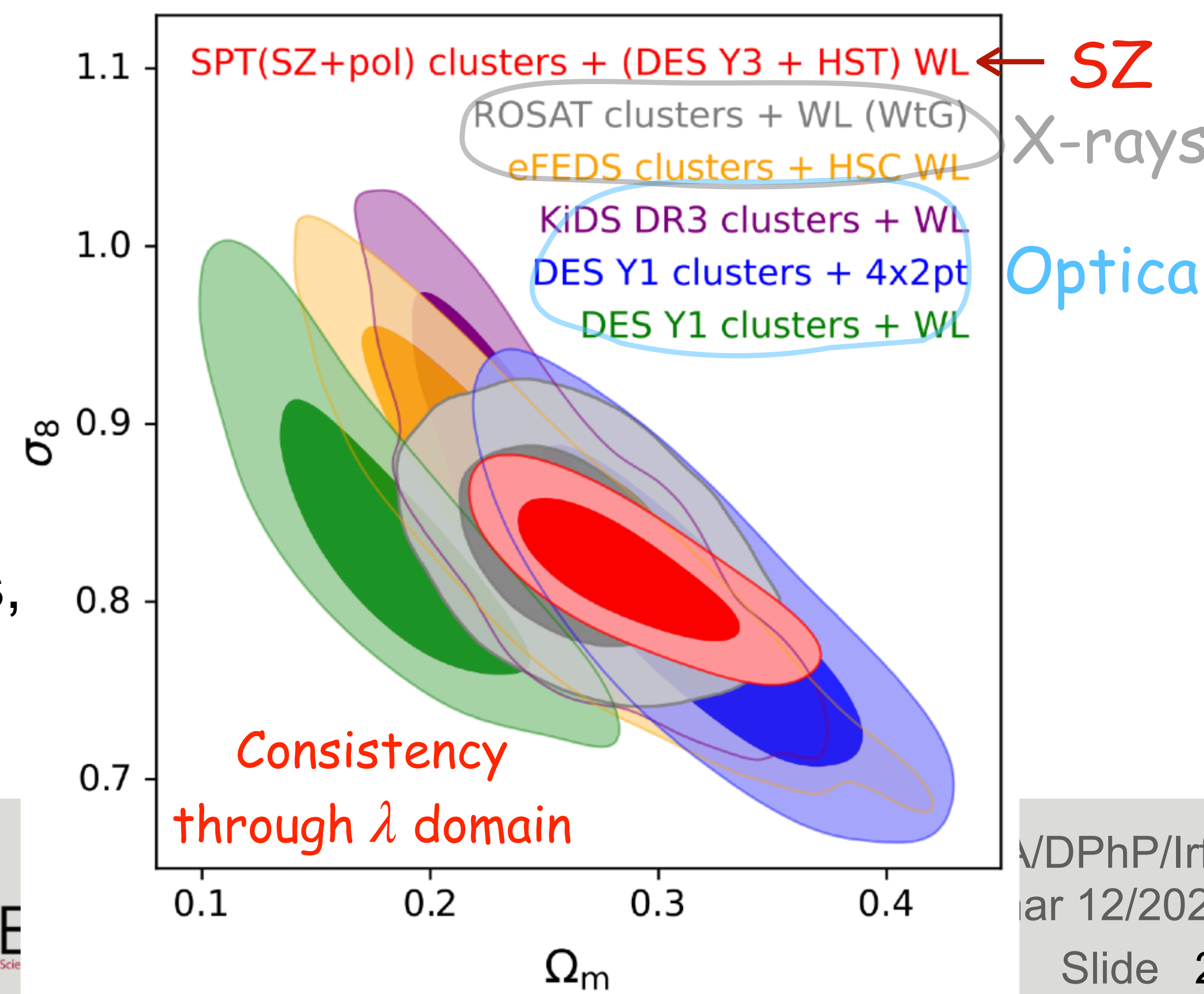
$$\frac{\partial^2 N_{\rm obs}^{\rm clusters}}{\partial \mathcal{O} \partial z} \propto \int dm \frac{\partial^2 N_{\rm th}^{\rm halo}(m,z)}{\partial m \partial z} P(\mathcal{O}\,|\,m,z)\Phi$$

- Privileged probes for structure formation and geometry in ΛCDM (i.e. Ω_m, S_8)
- Current constraining power: determined by uncertainties on the scaling relation +(usually combined with *e.g.* WL)

Beyond ΛCDM:

- wCDM impacts structure growths !
- Abundance of Massive halos: massive neutrinos, $f(R)$
- Sensitive to initial conditions (inflation), etc.





Experimental landscape

Analysis < June 2024 Payerne+24

Survey	Analysis	Ω_S (deg^2)	$z_{\rm min}$ - $z_{\rm max}$	$N_{\rm tot}$
ROSAT	WtG, Mantz et al. (2015)	400	0–0.5	224
XMM	XXV, Pacaud et al. (2018)	50	0.05–1	178
	XLVI, Garrel et al. (2022)	47.36	0.05–1	178
eROSITA	eFEDS, Chiu et al. (2023)	140	0.1–1.2	455
	eRASS1, Ghirardini et al. (2024)	12 791	0.1–0.8	5259
ACT	S11, Sehgal et al. (2011)	455	0.16–1	9
	H13, Hasselfield et al. (2013)	504	0.15–0.8	15
Planck	XX, Ade et al. (2014)	26 000	0–1	189
	XXIV, Ade et al. (2016)	26 000	0–1	439
	S18, Salvati, Douspis & Aghanim (2018)	26 000	0–1	439
	Z19, Zubeldia & Challinor (2019)	26 000	0–1	439
	×SPT, Salvati et al. (2022)	(26 000)×(2500)	0–1	782
	×*Chandra*, Aymerich et al. (2024)	26 000	0–1	439
	×ACT, Lee, Battye & Bolliet (2024)	(26 000)×(987.5)	0–1.4	439
SPT	B15, Bocquet et al. (2015)	720	$z > 0.3$	100
	H16, de Haan et al. (2016)	2500	0.25–1.7	377
	B19, Bocquet et al. (2019)	2500	0.25–1.75	343
	C22, Chaubal et al. (2022)	2500	0.25–1.75	343
	B24, Bocquet et al. (2024)	5200	0.25–1.75	1005
SDSS	R10, Rozo et al. (2010)	7398	0.1–0.3	10 810
	M13, Mana et al. (2013)	7500	0.1–0.3	13 823
	C19, Costanzi et al. (2019)	11 000	0.1–0.3	7000
	A20, Abdullah, Klypin & Wilson (2020)	11 000	0.045–0.125	756
	P23, Park et al. (2023)	11 000	0.1–0.33	8379
	S23, Sunayama et al. (2023)	11 000	0.1–0.33	8379
	F23, Fumagalli et al. (2023)	11 000	0.1–0.3	6964
DES	Y1, Abbott et al. (2020)	1800	0.2–0.65	7000
	T21, To et al. (2021)	1321	0.2–0.6	4794
	×SPT, Costanzi et al. (2021)	(1800)×(2500)	$z > 0.2$	7000
KiDS	DR3, Lesci et al. (2022)	377	0.1–0.6	3652

X-rays: ~100 to 5,000* cl.
*thanks eROSITA !

SZ: 100 to 1000 cl.

Optical: 1000 to 10,000 cl.

Era of precision cluster cosmology

- Stage IV catalogs ~ 100,000 clusters (LSST, SO, CMB-S4, ongoing: *Euclid, eROSITA*)
- Larger footprint, lower masses, higher redshifts, etc.
- Large statistical power (~ 10xcurrent datasets)
- Requires robust modeling of observables, better control of systematics => likelihood

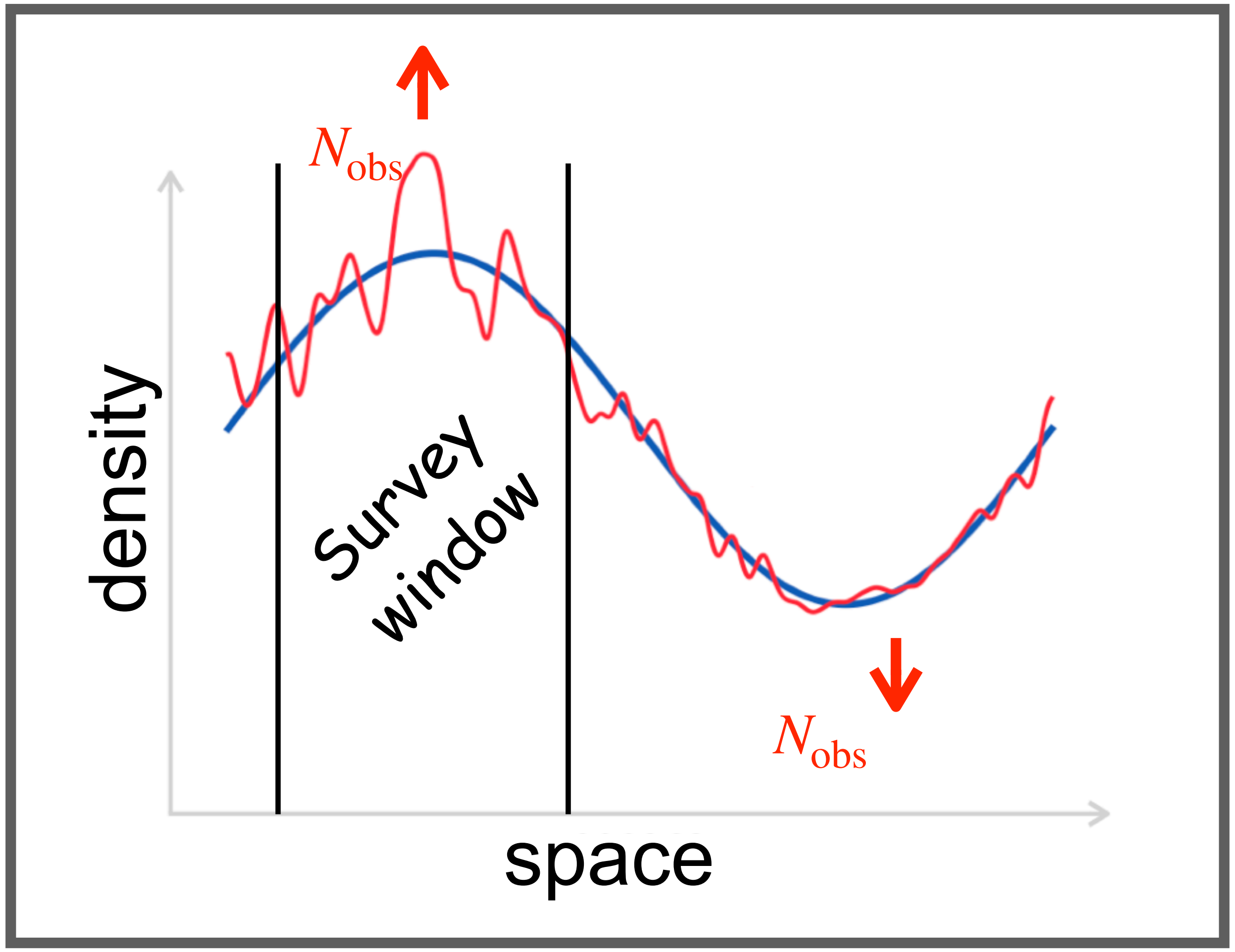
Bayesian inference

Estimation of cosmological params.

- We rely on the Bayes theorem

$$\mathscr{P}(\theta\,|\,\mathrm{obs}) = \frac{\mathscr{L}(\mathrm{obs}\,|\,\theta)\ \pi(\theta)}{\mathscr{L}(\mathrm{obs})}$$

- We need to derive the likelihood => understand count statistics



Adapted from Doux et al. (2017)

Poisson statistics

- Count of discrete objects in bins
- Uncorrelated counts
- Intrinsic Poisson variance $\sigma^2 = N$

Super-Sample Covariance (Hu&Kravtsov2003)

- We observe a finite region in the sky
- Target an over(-under) dense region
- Non-linear couplings between modes
- => Increase(-decrease) the number of halos
- => Correlates counts + additive variance
- Important for large $N_{\rm tot}$

$$\mathrm{Var}_{\rm SSC}(\mathcal{O}) = \int\!\!\int dV_1 dV_2 \frac{\partial \mathcal{O}(z_1)}{\partial \delta}\frac{\partial \mathcal{O}(z_2)}{\partial \delta}\sigma^2_{\mathscr{W}}(z_1, z_2)$$

$$\boxed{\sigma^2_{\rm tot}(N) = N + b_h^2 N^2 \sigma^2_{\mathscr{W}}}$$

Impact of SSC on cosmology

Impact on parameter inference

- Strong impact on cosmological constraints for wide surveys
- -50% on FoM for 3x2pt (Gouyou-Beauchamps+2021)
- Cluster abundance:
    - Impact on binned analyses (count in redshift-proxy bins, Fumagalli+2021)
    - $\sim 10^5$ objects: 25% change in FoM
    - 10% on parameter error!

Standard binned likelihood choice

- Depends on the relative importance between SSC and SN => total number N
- Poisson or SSC dominated regime (or both !)
- Overview of used likelihoods in the literature: Table 2 in Payerne+2024

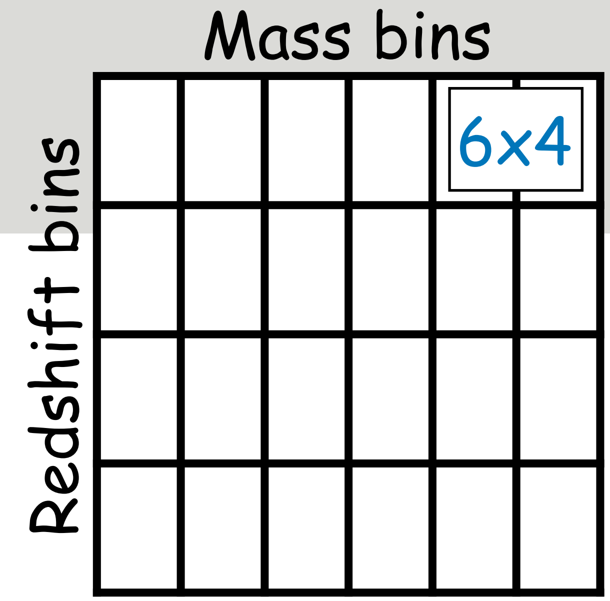
Total number of clusters ?
Survey characteristics ?
…

$N_{\rm tot}$ « small » $\text{Poisson} = \prod_k^n \mathscr{P}(\widehat{N}_k \,|\, N_{\mathrm{th},k})$

$N_{\rm tot}$ « large » $\text{Gaussian} = \mathscr{N}(N - \widehat{N} \,|\, \Sigma_{\rm SN} + \Sigma_{\rm SSC})$

Whatever $\text{GPC} = \left\langle \prod_k^n \mathscr{P}[\widehat{N}_k \,|\, N_{\mathrm{th},k}(1 + \delta_{h,k}^{\rm SSC})] \right\rangle_{\rm SSC}$

This talk: A focus on SSC



Impact of SSC with a Gaussian likelihood (Fumagalli+2021)
- What is the optimal inference setup?
- How does SSC impact cosmological parameters?

I. Validity of « literature » binned likelihoods (Payerne+2023)
- Does likelihoods provide « correct » errors?
- Are « errors » equal to the « latent » dispersion?
- Is there a framework to test the likelihoods?

II. SSC in the unbinned regime (Payerne+2024)
- Infinitesimal mass-redshift bins, until the count is at most 1 cluster

$$N_k \sim \frac{d^2 n(m_i, z_i)}{dmdV}\Delta m \Delta V \ll 1$$

- Lot of good reasons to use unbinned ! (multi-λ sample, etc.)
- Historically, generally assumed that SN is the dominant source of uncertainties (small cluster samples) - use only Poisson statistics
- SSC $\sim N^2$, important for Stage IV surveys !
- Not done before, try to include SSC in the unbinned regime!

I. Testing the accuracy of likelihoods

Simulated dark matter halo catalogs

- PINOCCHIO algorithm (Monaco+2013)
- Planck cosmology
- 1000 simulations Euclid/LSST-like

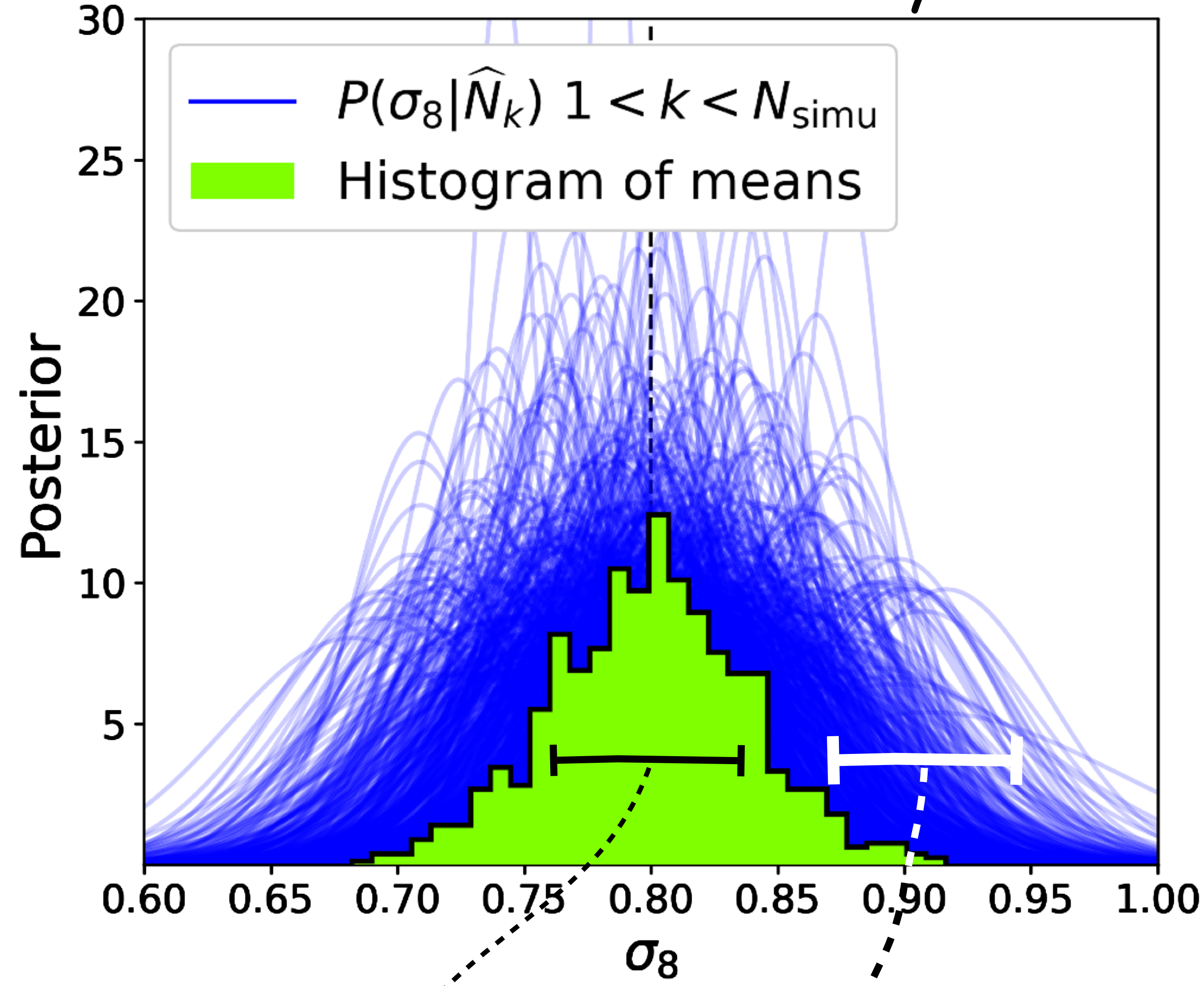
Methodology

- Estimate the posterior for each simulation
- Robustness of errors? Compare individual posterior dispersion $\sigma_{\rm ind}$ to the spread of posterior means $\sigma_{\rm ens}$ (ensemble dispersion)
- Repeat for each analysis likelihood!

$\sigma_{\rm ind}$ Depends on the "analysis" likelihood

$\sigma_{\rm ens}$ Depends on both "analysis" and "latent" likelihood

Toy model



Dispersion of the set of posterior means $\sigma_{\rm ens}$

Individual posterior dispersion $\sigma_{\rm ind}$

I. Testing the accuracy of likelihoods

Why we do that ?

$\sigma_{\rm ens} = \sigma_{\rm ind}$ → "analysis" = "latent"
Correct likelihood

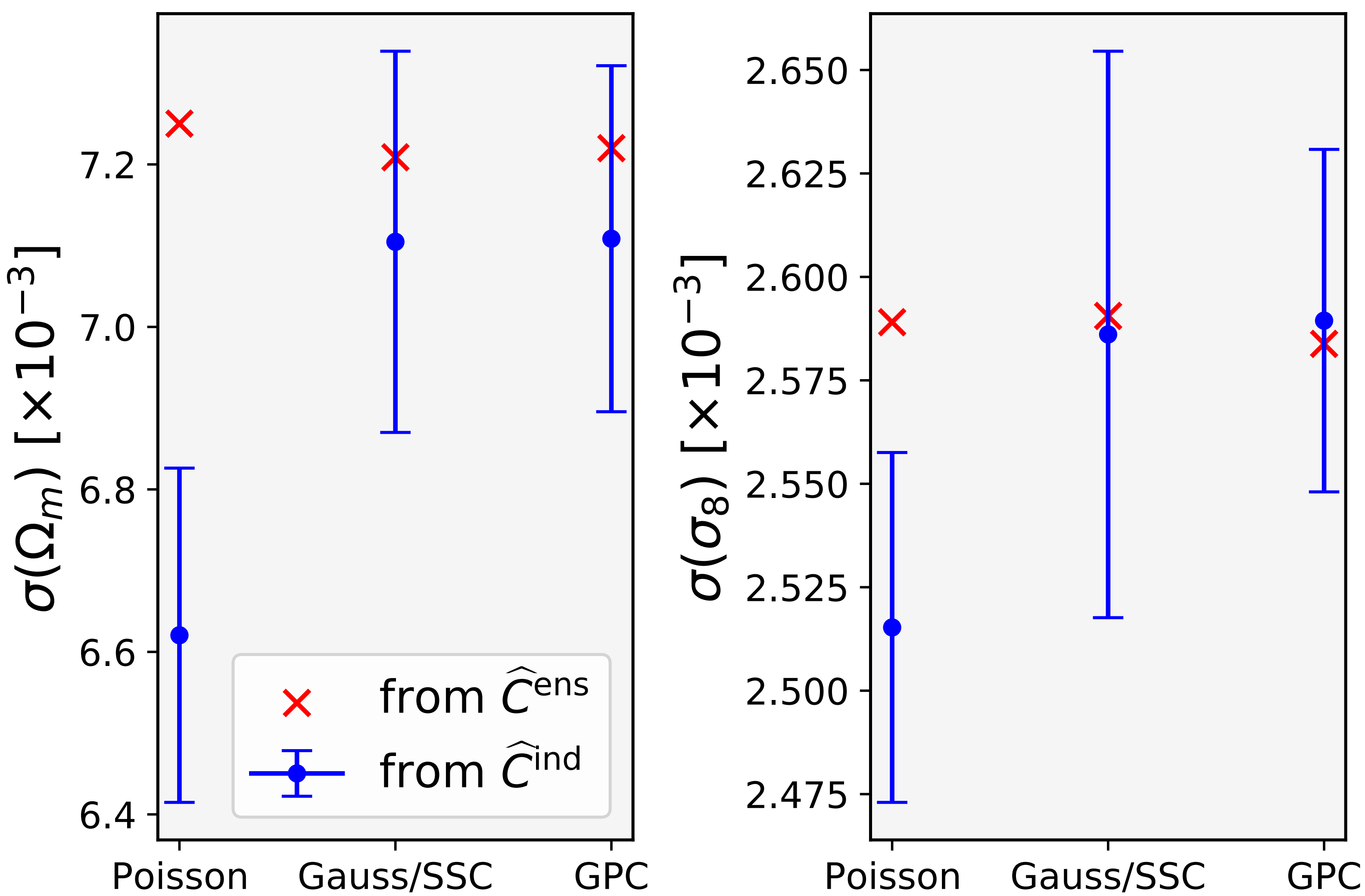
$\sigma_{\rm ens} \neq \sigma_{\rm ind}$ → "analysis" ≠ "latent"
Approximation is not correct
under/over-estimate errors

Results

- Various analysis setups
- Poisson underestimates the errors, since no SSC (even 100x100 bins)
- Gaussian = GPC
- The same level of constraints

Take home message

- New framework to test likelihood « without knowing the latent one »
- **Gaussian likelihood remains an accurate description of the data (for *Euclid*/LSST-like stats)**
- No particular need to use GPC!



This talk: A focus on SSC

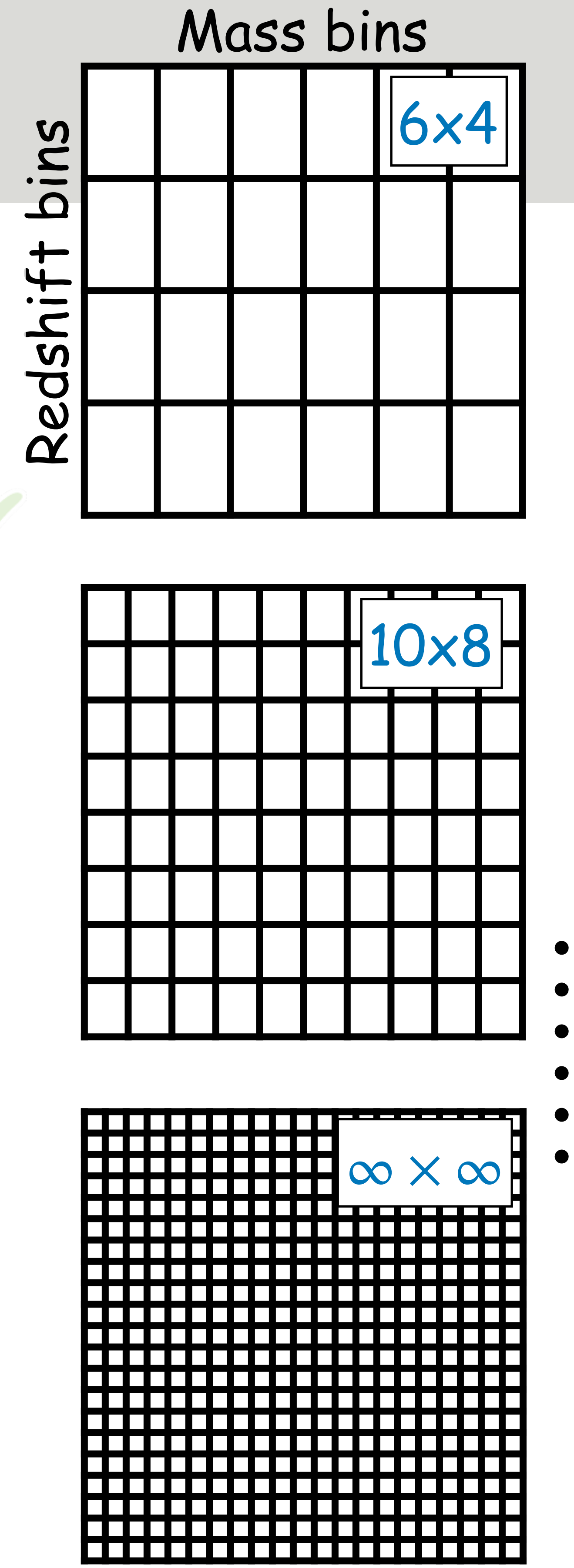
Impact of SSC with a Gaussian likelihood (Fumagalli+2021)
- What is the optimal inference setup?
- How does SSC impact cosmological parameters?

I. Validity of « literature » binned likelihoods (Payerne+2023)
- Does likelihoods provide « correct » errors?
- Are « errors » equal to the « latent » dispersion?
- Is there a framework to test the likelihoods?

II. SSC in the unbinned regime (Payerne+2024)
- Infinitesimal mass-redshift bins, until the count is at most 1 cluster

$$N_k \sim \frac{d^2 n(m_i, z_i)}{dmdV}\Delta m \Delta V \ll 1$$

- Lot of good reasons to use unbinned ! (multi-λ sample, etc.)
- Historically, generally assumed that SN is the dominant source of uncertainties (small cluster samples) - use only Poisson statistics
- SSC $\sim N^2$, important for Stage IV surveys !
- Not done before, try to include SSC in the unbinned regime!

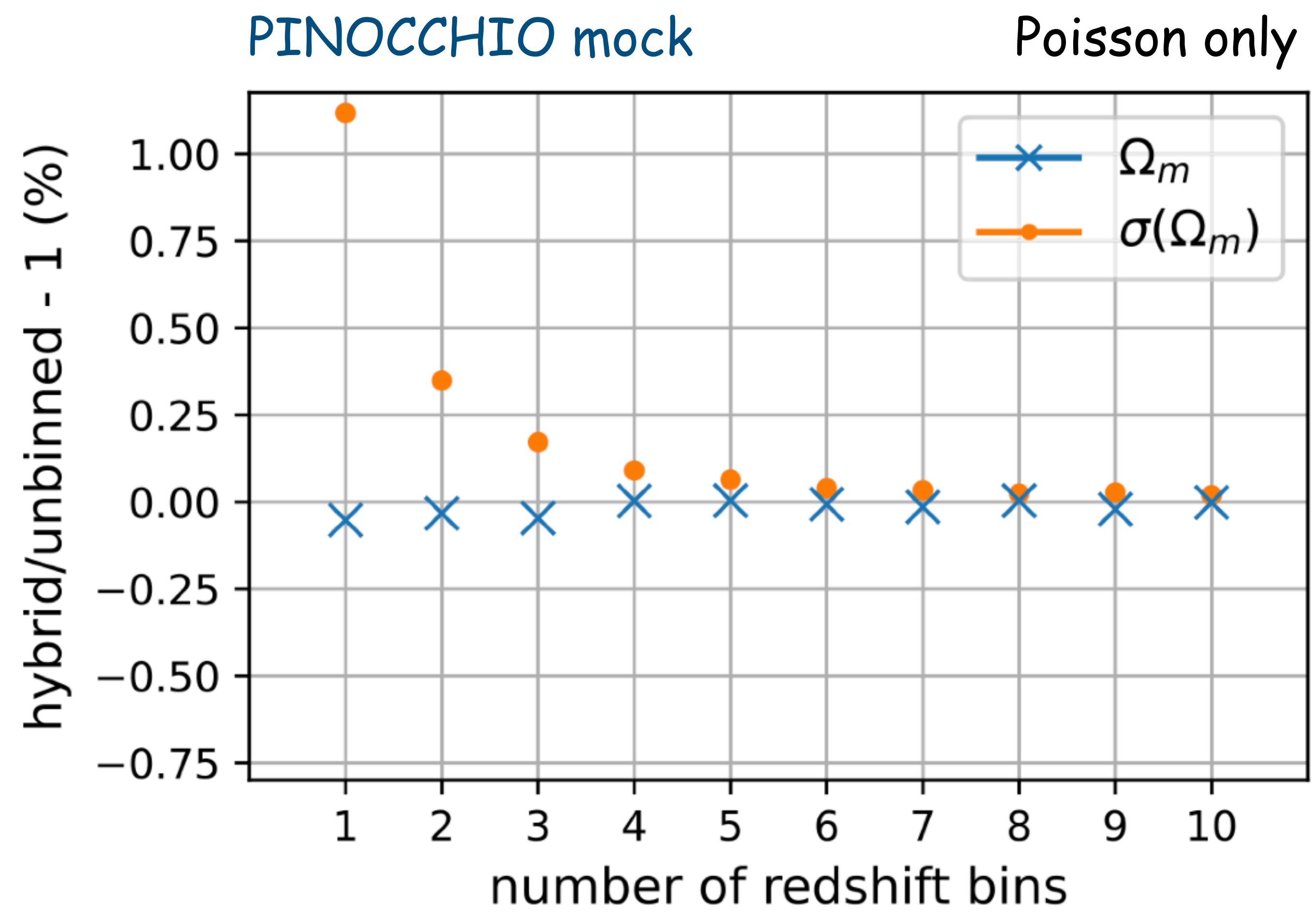


II. SSC in the ~~unbinned~~ Hybrid regime

$$\mathrm{GPC} = \left\langle \prod_{k}^{n} \mathscr{P}[\,\widehat{N}_k \,|\, N_{\mathrm{th},k}(1 + \delta^{\mathrm{SSC}}_{h,k})] \right\rangle_{\mathrm{SSC}}$$

$$\downarrow$$

$$\boxed{N_{\mathrm{th}} \ll 1;\;\; \delta_{\mathrm{SSC}} \ll 1}$$



Recipe

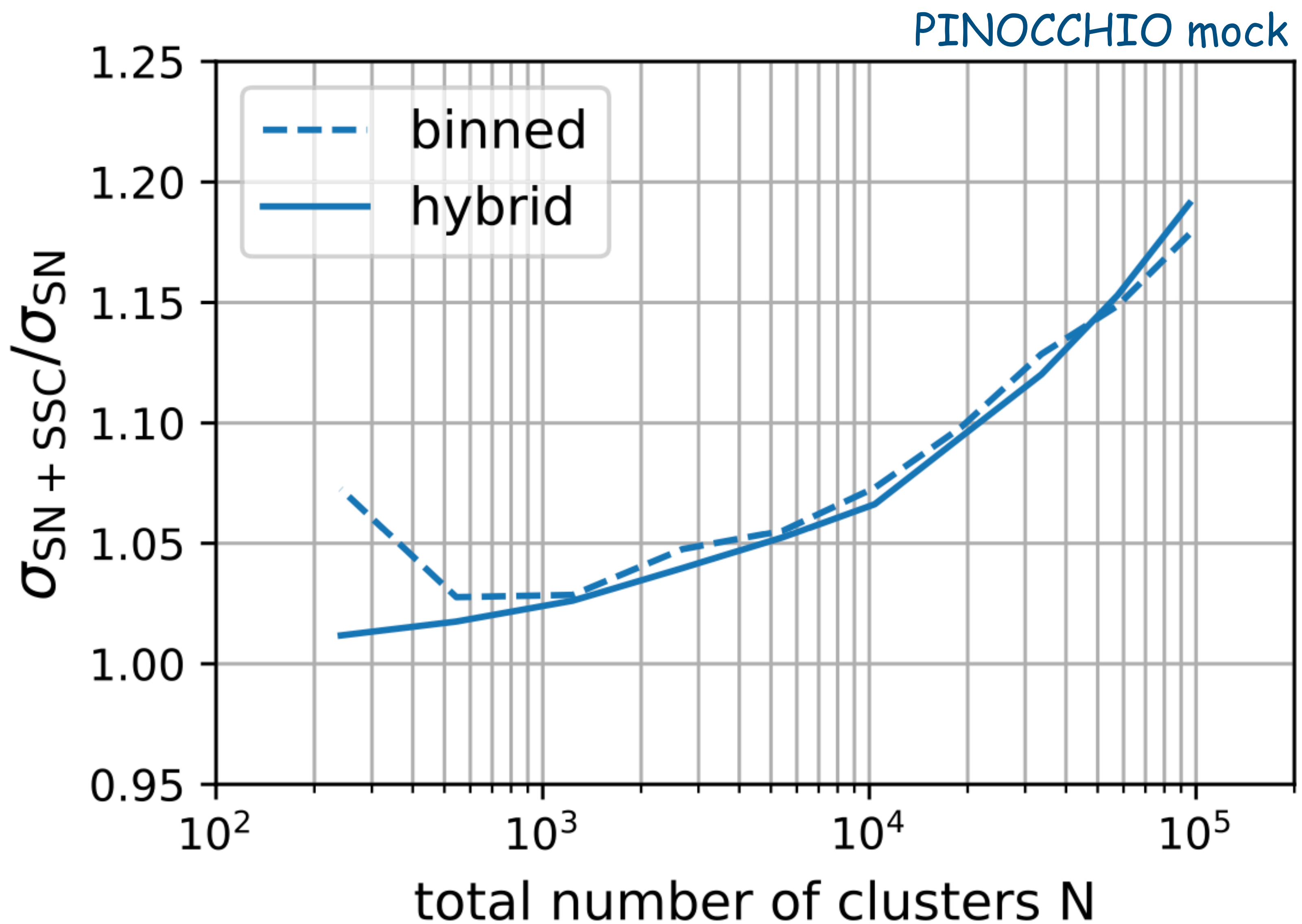
1. We used the GPC formalism (Hu&Kravtsov2003)
2. ~~Linearization to 2nd order in count/over-density (Takada&Spergel2014)~~

→ our approach gives non-physical results ($\delta_{\mathrm{SSC}} \ll 1$ not true)!

Hybrid regime

- « standard » redshift bins, infinitesimal mass bins
- When only Poisson, very close to « standard » unbinned < 0.2% bias

II. SSC in the unbinned regime



SSC now !

- SN+SSC-to-SN ratio compatible with the binned approach
- Implementation of SSC in the hybrid regime behaves as expected
- 15% impact on parameter errors for 10,000 clusters

Summary

Cluster cosmology

- Competitive probe of growth of structure
- Precision cosmology with Stage IV surveys
- Better control of systematics
- Understand count statistics SN+SSC

$$\mathscr{P}(\theta\,|\,\mathrm{obs}) = \frac{\mathscr{L}(\mathrm{obs}\,|\,\theta)\ \pi(\theta)}{\mathscr{L}(\mathrm{obs})}$$

Cluster likelihoods

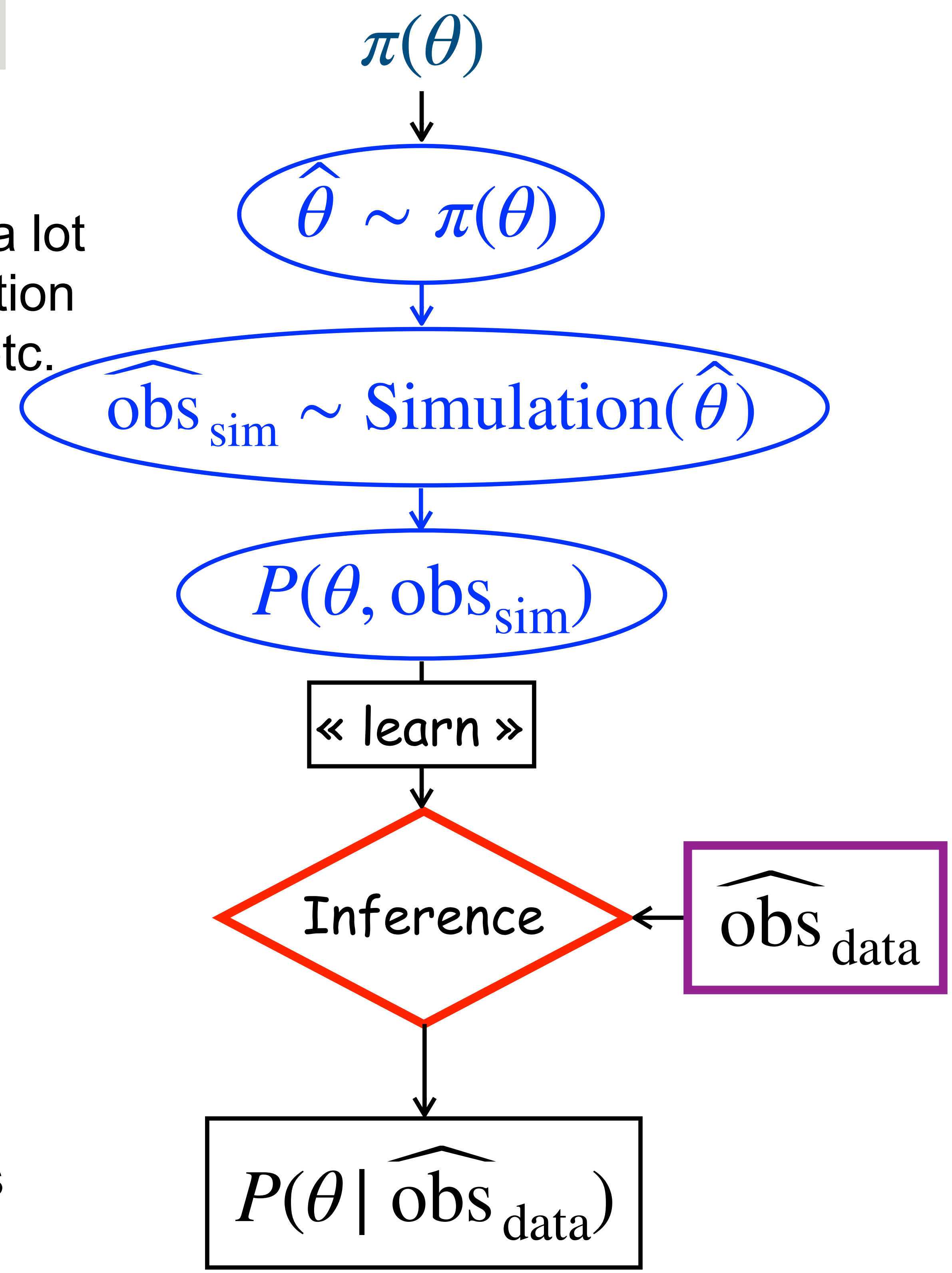
- ***Robustness of binned likelihood Payerne+2023***
  - **Framework to test the robustness of binned likelihoods**
  - Gaussian is good approximation for LSST/Euclid stats.
- ***SSC in the unbinned regime Payerne+2024***
  - From now, SSC ignored (small cluster samples)
  - This work: only possible using a new hybrid regime (standard redshift bins)
  - SSC behaves as expected, **new formalism for the community!**
    - $\mathscr{L} = \mathscr{L}_{\mathrm{Poisson}} \times \mathrm{SSC\ term}$
  - Could be used with *eROSITA*, SPT (multi-λ cluster analysis)

Further works

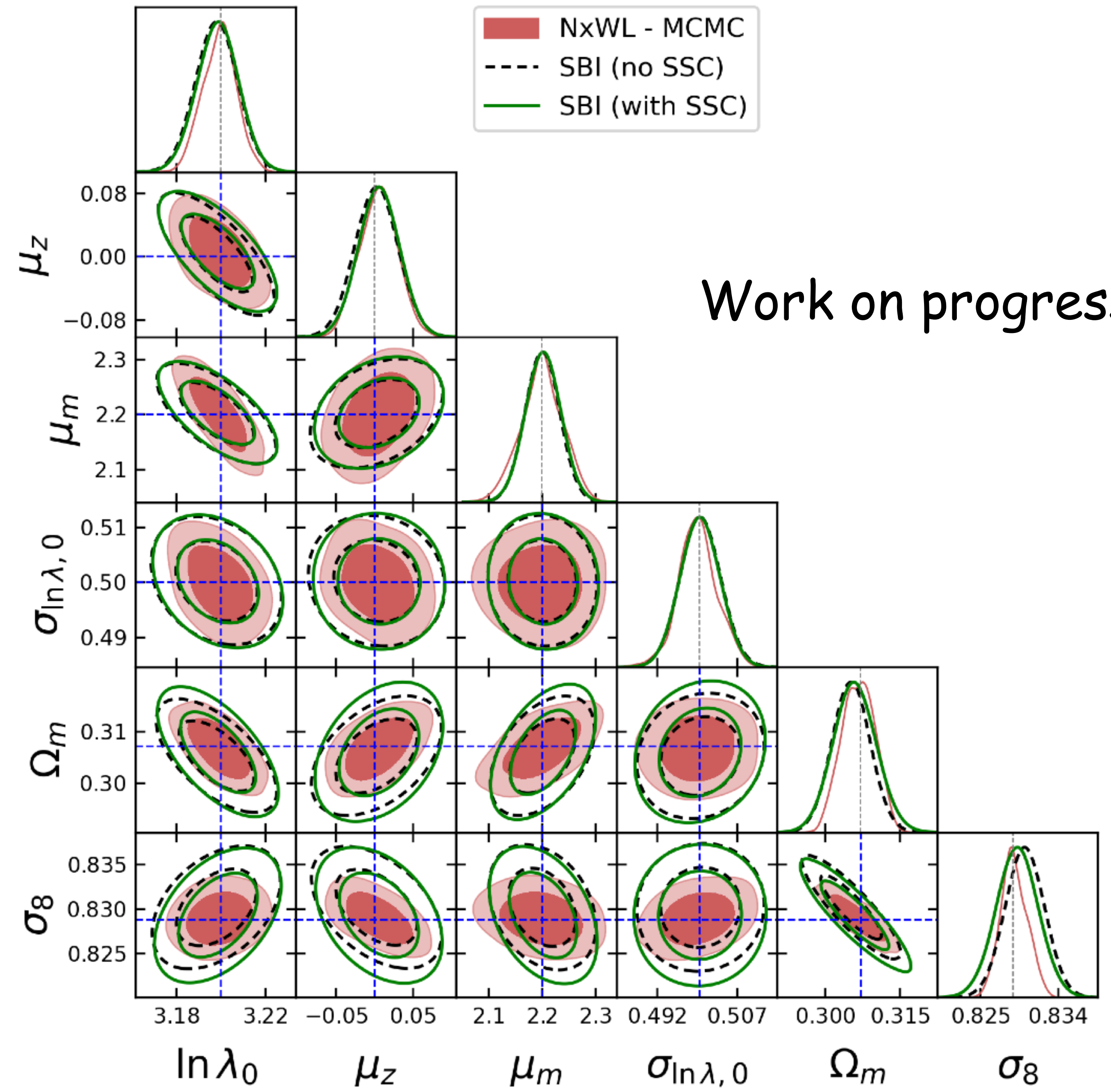
Limitation of « explicit » likelihood
- Explicit = Analytical formula (e.g. for a MCMC)
- May be difficult to obtain, when needing to include a lot of observational systematic effects (photoz's, selection function, selection bias) + sophisticated statistics, etc.

Beyond likelihood
- Simulation-Based Inference (aka *Likelihood-Free-Inference* or *Neural-Posterior-Estimation*)
- Forward simulation only! + ML-based
- GC SDSS/BOSS: Lemos+2023
- CL (simulations): Tam+2021, Reza+2023,2024

With Calum Murray
- Comparing explicit to SBI (sbi package)
- Exploring SSC with more complicated statistics!
- SSC in the unbinned regime! (no needs hybrid regime)
- Revisit current CL+WL constraints with SBI - focus on cluster finder selection effects



Further works - an example



Work on progress!


Thank you!

Appendix: hybrid likelihood with SSC

$$\mathscr{L}(\{m_k, z_k\}) = \left(e^{-N_{\rm th}} \prod_{i=1}^{n_z} \prod_{k=1}^{\widehat{N}_{\rm tot}(\in i)} N_k^{(i)} \right) \times \sqrt{\frac{\det S^{-1}}{\det C}} \exp\{\frac{1}{4} v^T C^{-1} v\} \, .$$

Poisson term

SSC term

$$N_k \approx \Delta m \int_{z_i}^{z_{i+1}} n_h(m_k, z') dV_k(z')$$

$$v_i = \langle Nb \rangle_i - \langle \widehat{Nb} \rangle_i$$

$$C_{ij} = \frac{1}{2}[S_{ij}^{-1} + \langle \widehat{Nb^2} \rangle_i \delta_{ij}^K]$$