Testing likelihood accuracy for cluster count cosmology

Constantin Payerne
3^{rd} year PhD student at LPSC, Grenoble
Under the supervision of Dr. Céline Combet

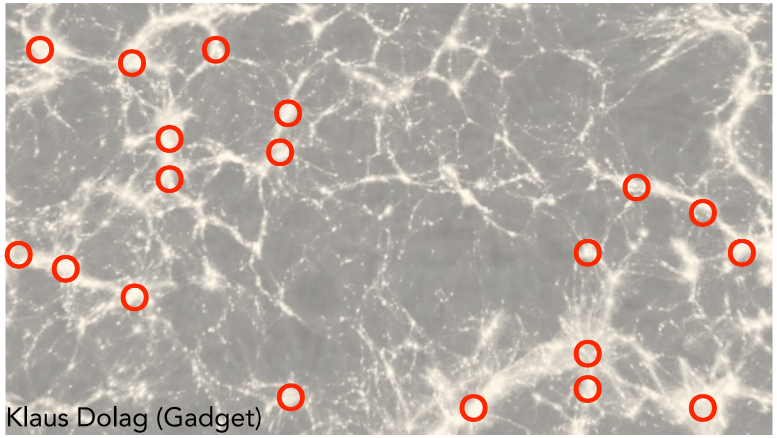
Talk based on:

Testing likelihood accuracy for cluster abundance cosmology
C. Payerne, C. Murray, C. Combet, C. Doux, A. Fumagalli, M. Penna-Lima, 2210.11093

Galaxy Clusters and Cosmology

Galaxy clusters: a brief introduction

- Most massive bound systems with $M \in 10^{13} - 10^{15}\ M_\odot$
- $z < 2$, last step of hierarchical structure formation process
- Densest regions in the cosmic web filaments



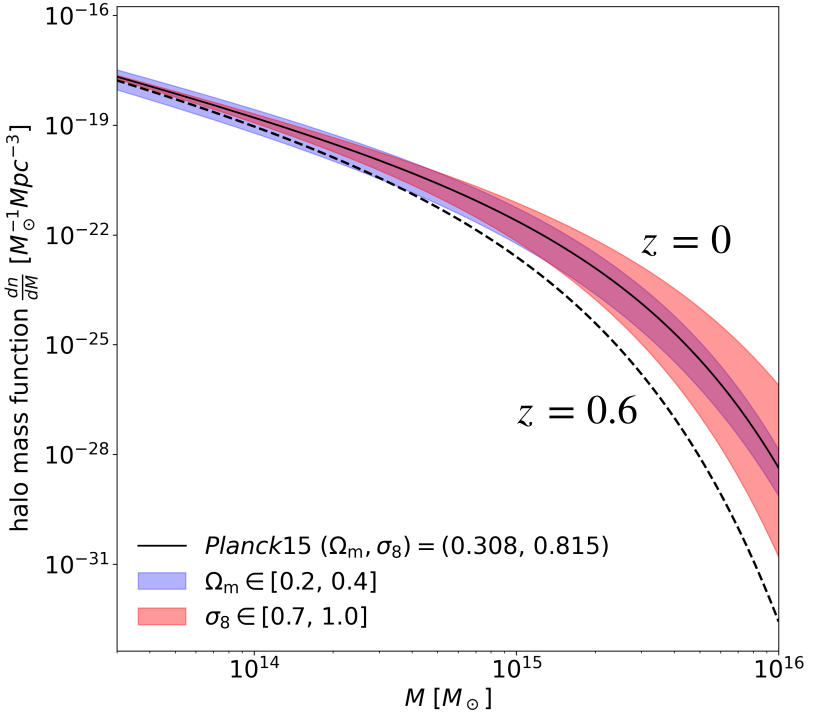
Probing cosmology with galaxy cluster abundance

- Count clusters a function of redshift and mass

Number density $$\frac{\partial^2 N_{\rm th}}{\partial z \partial m} \propto \frac{dn(m,z)}{dm}\frac{dV(z)}{dz}$$

- Depends on:
  - **Halo Mass Function** (matter content $\Omega_{\rm m}$, growth rate of structure $\sigma_8(z)$)
  - **Volume** (background cosmology)

- Geometry + growth of structures in the Universe



Cosmology with galaxy cluster abundance

Basic recipe for cluster abundance cosmology

- Observations
  - From a galaxy cluster survey with known redshifts, masses
  - Count the number $\overrightarrow{N}_{\rm obs}$ of galaxy clusters within bins of redshift and mass

- Cosmological analysis: define likelihood
  - $\overrightarrow{N}_{\rm th}$ at arbitrary cosmology
  - Statistics:
    - Count of discrete objects in bins
      - Poisson sampling
      - Intrinsic count variance: Shot noise $\mathrm{Var}_{\rm Poiss}(N_k) = N_k$
    - Fluctuation + clustering of the matter density field
      - Gaussian contributions: (Super) Sample Covariance
      - $\mathrm{Corr}_{\rm SSC}(N_k, N_l) \neq 0$
      - $\mathrm{Var}_{\rm SSC}(N_k) \sim b_k^2 \sim P_{\rm mm}(k) \sim N_k^2$
    - Non-linear physics of halo formation → More complications
    - Observational systematics, …

Likelihoods for cluster count cosmology

Likelihoods

- Ideally should describe completely abundance statistics
- There exist approximations

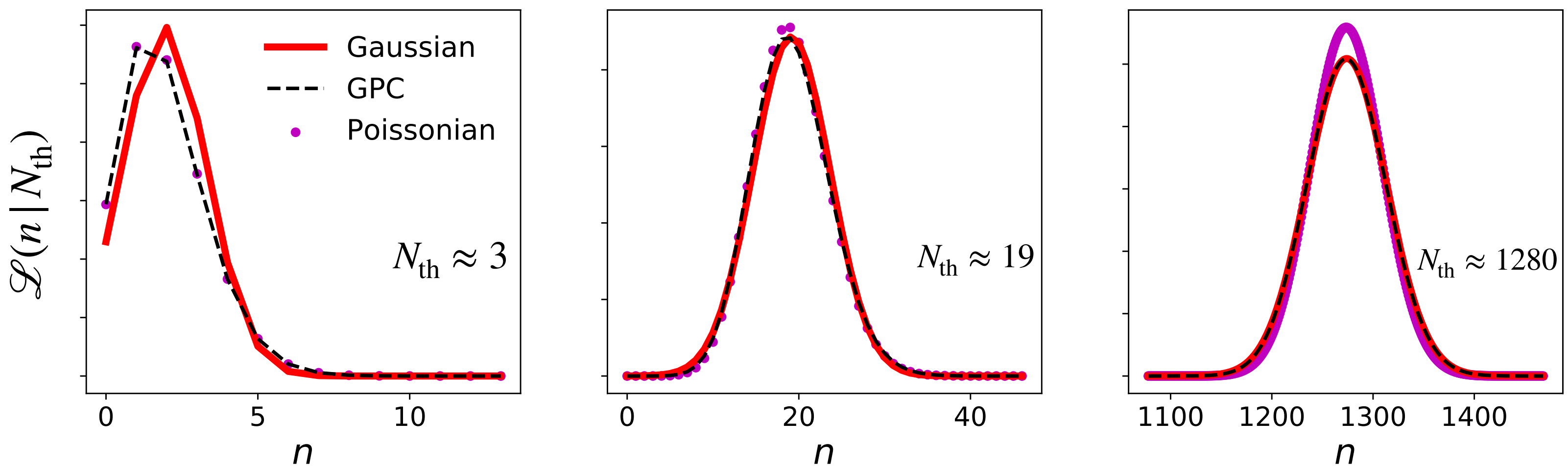
  - **Poisson likelihood** (Planck, 2015 ~ 500 clusters)
    - Accounts for Poisson sampling
    - Does not account for sample covariance
    - Valid for low number of clusters, Shot Noise >> Sample variance

  - **Gaussian likelihood** (DES, 2021 ~ 7000 clusters)
    - Sample covariance
    - Limited to continuous approximation
    - Valid for high number of clusters, Shot Noise ~ Sample variance

  - **Gauss-Poisson Compound** (GPC) (KiDS, 2021 ~ 4000 clusters)
    - Takes into account both Poisson sampling and sample covariance (Hu & Kravtsov, 2003)
    - Computationally expansive to compute
    - Multidimensional integral $\mathscr{L}(\widehat{N} \mid \vec{\theta}) \propto \int d\vec{x}\ \mathscr{N}[\vec{x} \mid \vec{N}(\theta)] \times \prod_{k=1}^{n} \mathscr{P}[\widehat{N}_k \mid x_k]$
    - More precise, can we gain cosmological information?

Likelihoods for cluster count cosmology

Considering the count in 3 different mass-redshift bins



"low" $\langle N \rangle$ "middle" $\langle N \rangle$ "high" $\langle N \rangle$

Bias on parameter inference

- Deviation of the analysis likelihood from the latent one may bias results
- Most robust constraints with analysis likelihood closest to latent one

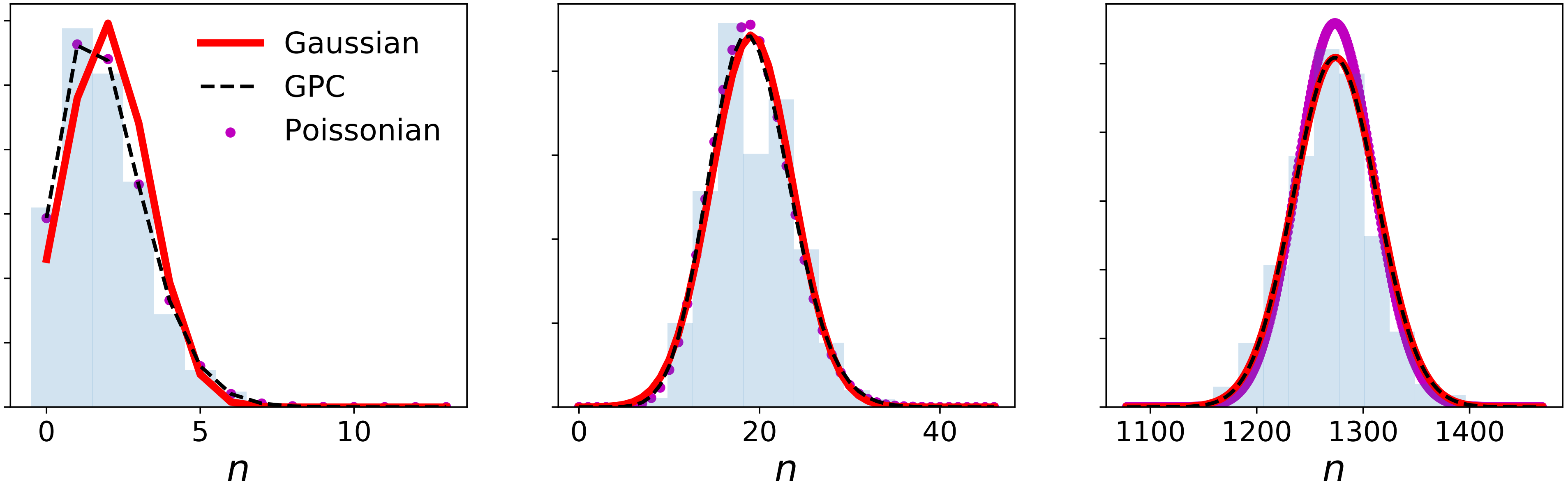
Using simulations to test cluster abundance likelihoods

- Likelihood: statistical properties of the data at input cosmology
- With multiples simulations, can have access to "true" statistics of abundance

We use a set 1000 simulated dark matter halo catalogs

- PINOCCHIO algorithm (Monaco et al., 2013)
- Planck cosmology
- Masses calibrated on known halo mass function (Despali et al., 2015)
- Euclid-like sky area ~ ¼ of full-sky
- $\sim 10^5$ halos per simulation
- $M > 10^{14}\ M_\odot$

Abundance likelihood can be estimated from counts over the 1000 cosmological simulations



Frequentist Covariance of Bayesian Estimators

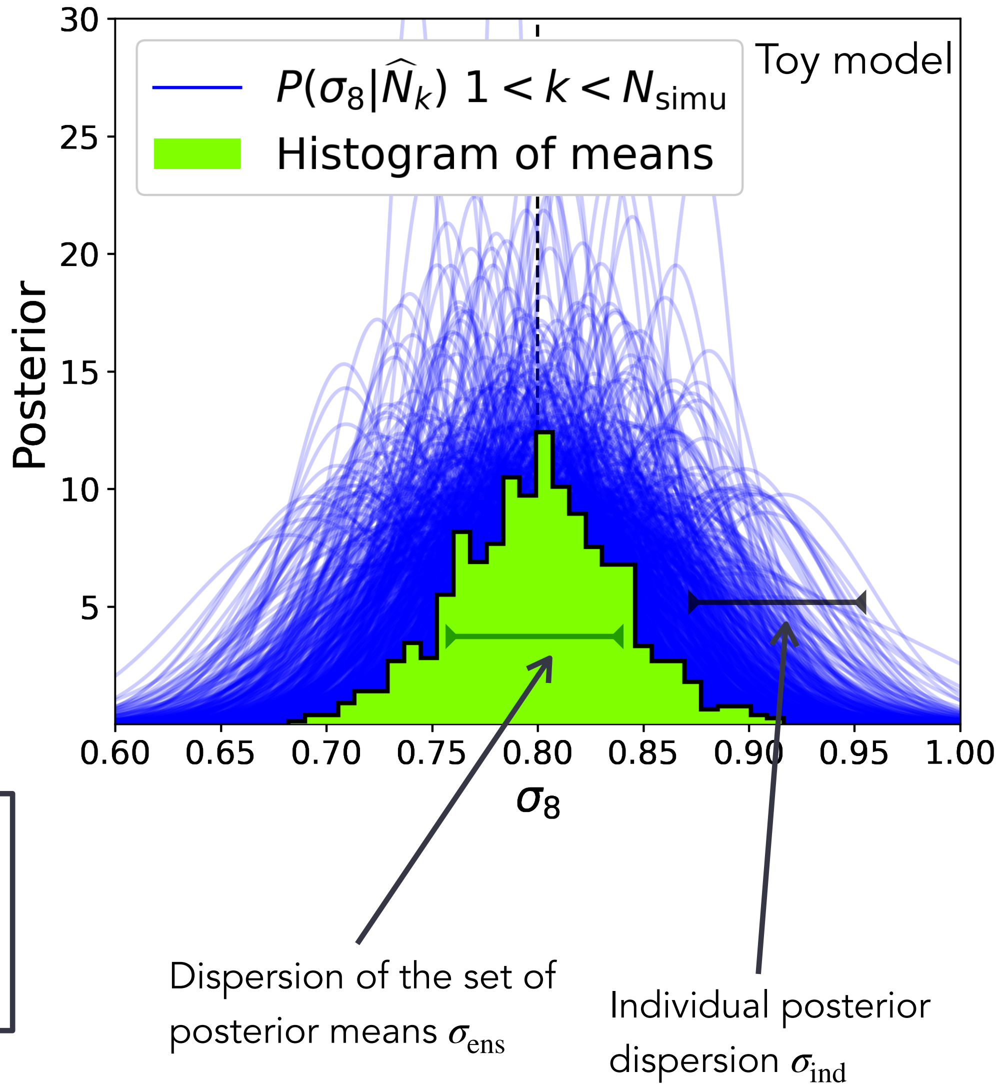
Methodology

- Estimate the posterior for each of the 1000 Pinocchio mocks
- Biases ? Compare the mean of each posterior to input cosmology
- Robustness of errors ? Compare individual posterior dispersion $\sigma_{\rm ind}$ to the spread of posterior means $\sigma_{\rm ens}$ (ensemble dispersion)

More than 1 parameter: compare covariances

$\sigma^2_{\rm ind} \rightarrow C^{\rm ind}$ Individual parameter covariance

$\sigma^2_{\rm ens} \rightarrow C^{\rm ens}$ Ensemble parameter covariance



Why comparing individual errors to the spread of means ?

Gaussian:

Latent likelihood $\mathscr{L}_X(\Sigma_X)$

Analysis likelihood $\mathscr{L}_Y(\Sigma_Y)$

$$C^{\rm ens} - C^{\rm ind} = ?$$

$= 0$ if $\Sigma_X = \Sigma_Y$

- Parameter errors are "robust"
- And can be forecasted (Fisher formalism)

$$[(C^{\rm Fisher})^{-1}]_{\alpha\beta} = N_{,\alpha}^T \Sigma_Y^{-1} N_{,\beta}$$

$\neq 0$ if $\Sigma_X \neq \Sigma_Y$

- $C^{\rm Fisher}$ is not sufficient
- $C^{\rm ens}$ can be forecasted

$$C^{\rm ens}_{\alpha\beta} = (C^{\rm Fisher} N)_{,\alpha}^T \Sigma_Y^{-1} \Sigma_X \Sigma_Y^{-1} (C^{\rm Fisher} N)_{,\beta} \neq C^{\rm Fisher}_{\alpha\beta}$$

- Example:

$\Sigma_{Y_{ii}} < \Sigma_{X_{ii}}$ then we have ${C^{\rm ind}}_{\alpha\alpha} < {C^{\rm correct}}_{\alpha\alpha} < {C^{\rm ens}}_{\alpha\alpha}$

- Metric: Using correct likelihood gives $C^{\rm ens} = C^{\rm ind}$
- Likelihood and posterior are not always gaussians
- Rather closeness between individual errors and ensemble error
- Used as a metric to test likelihood accuracy $C^{\rm ind}$ + robustness $C^{\rm ens} - C^{\rm ind} = ?$

Cosmological inference setup

- The Poisson, Gaussian and GPC likelihood are approximations
- Valid at linear scales (clusters are biased tracers of the density field)
- For given count magnitude N_k + for SSV/SN ratio $\sim N_k$
- These quantities can vary by changing the binning the mass-redshift plane

Methodology: Test accuracy of likelihoods for various regimes

For each likelihood

1. Compare C^{ens}, C^{ind} for the overall 1000 PINOCCHIO mocks
2. For 3 binning schemes

Binning setup

	Redshift bins	Mass bins	# of bins	Average # N/bin
#1	4	4	16	5000
#2	20	30	600	150
#3	100	100	10 000	10

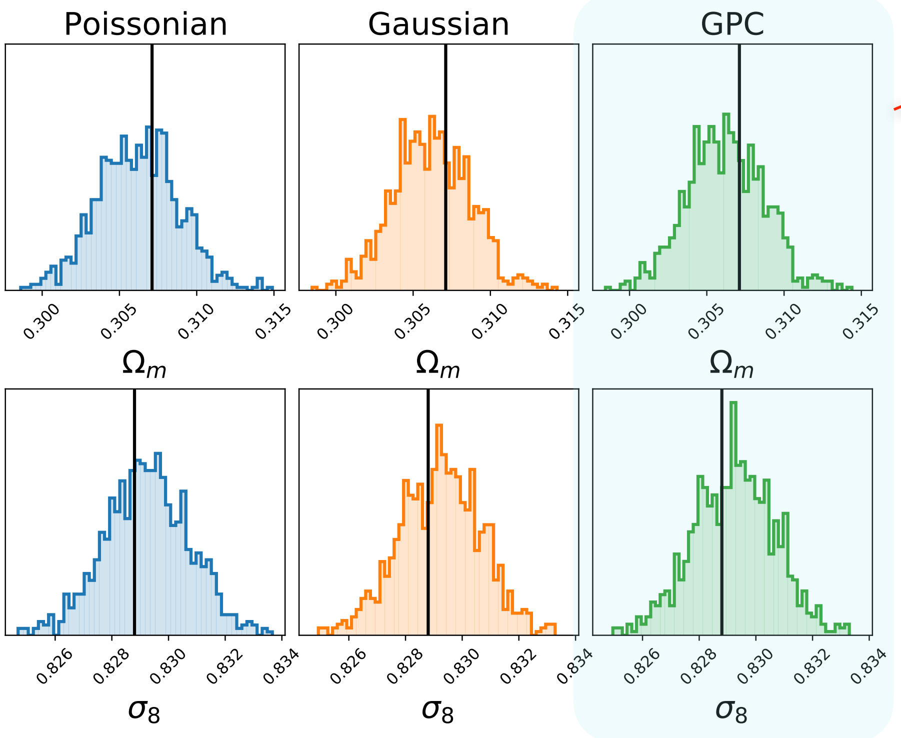
Poisson sampling

Sample Variance

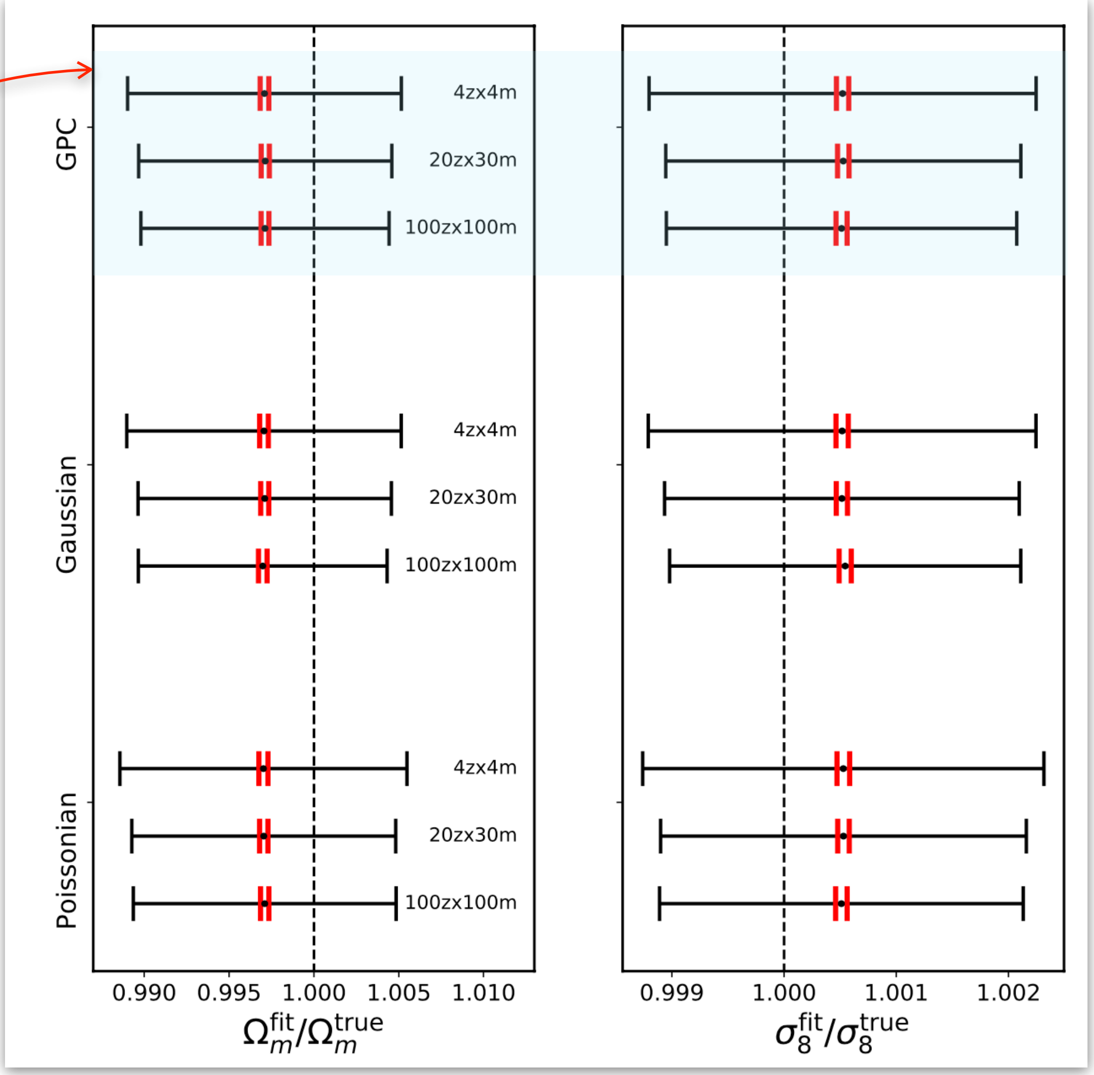
$\sim 10^4$ cosmological constraints ! **Importance sampling (efficient for 2 parameters)**

Results: Bias to input cosmology ?

Only binning 4zx4m



All binning scheme



Small constant bias between input and recovered cosmology

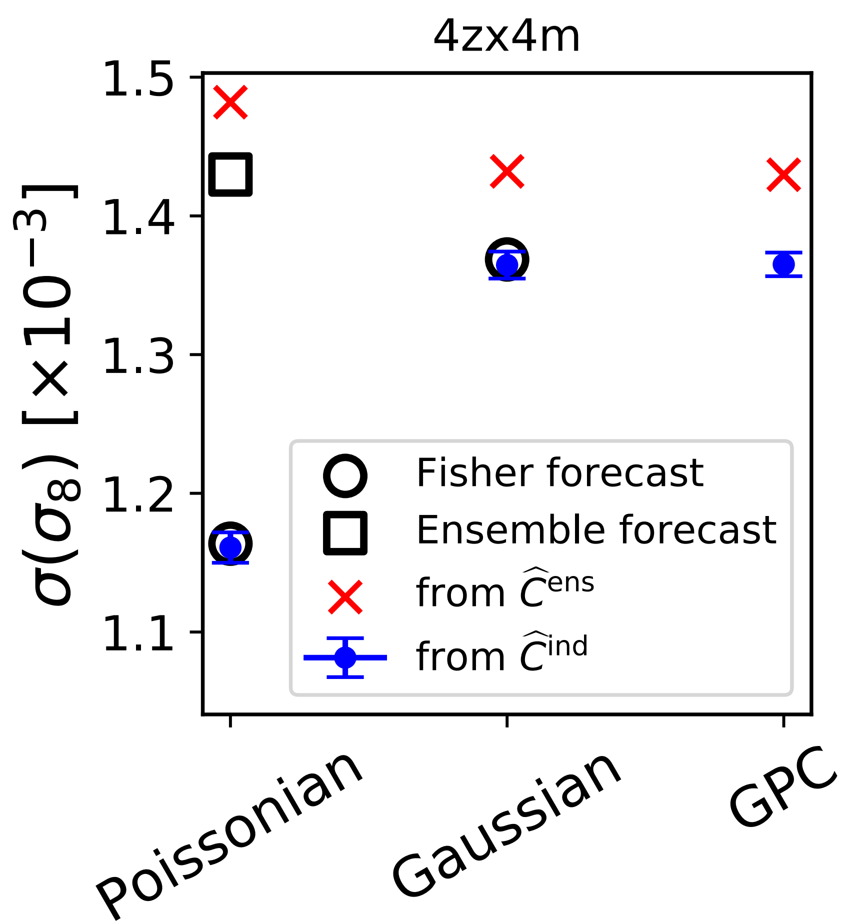
- Accuracy of the underlying halo model
- Numerical error

Results: (4 redshift bins)x(4 mass bins) case

- Individual errors on each simulation (blue)
- Spread of best fits (red)

Parameter error

- Poisson underestimates the errors, since it not takes account of sample variance
- Gaussian = Gauss-Poisson Compound
  - Slightly underestimate errors, likely due to approximations made for the 2-pt statistics
  - The same level of constraints



- Fisher forecasts (circle) in agreement with individual errors
- Ensemble forecast (square) for the spread of posterior means

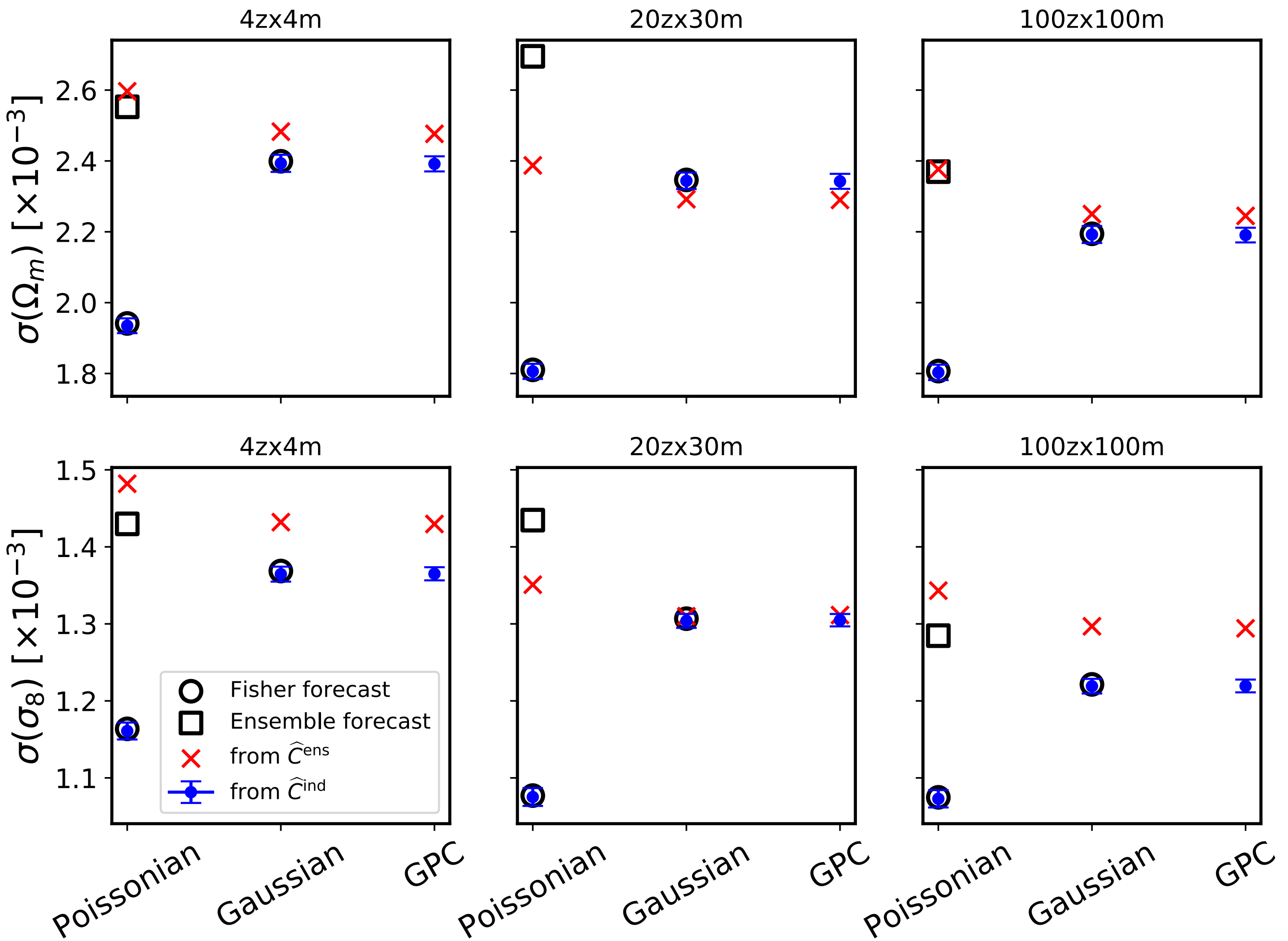
Results: all binning schemes

Parameter error

- Errors decreases with the number of bins (10% improvement from 16 to 10^4 bins)

- **Poisson**
    - Underestimates the error, even for fine binning, does not account for sample variance

- **Gaussian = Gauss-Poisson Comp.**
    - Over/under estimate constraints (approximation for computing the covariance matrix)
    - The same level of constraints



Gaussian likelihood remains an accurate description of the data

Poisson sampling →

Sample variance ←

Conclusions

Recap:

- Tested accuracy of cluster likelihoods with
  - 1000 simulated dark matter halo catalogs
  - By comparing posterior variances to spread of means over the 1000 simulations
  - Sensitive to both analysis and latent likelihood properties

Conclusions: For future Euclid or Rubin-like surveys

- Gaussian gives robust constraints over a wide range of inference setup
- No gain in using Gauss-Poisson Compound (same level of constraints but computationally expansive)
- Gauss-Poisson Compound = Gaussian (under/overestimating errors at most 5%)
- Poisson likelihood always underestimates errors