



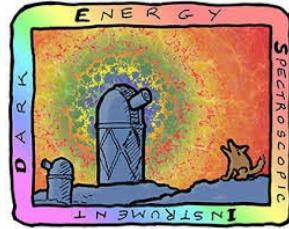
# CARPooling against the cosmological simulation bottleneck

*Nicolas Chartier*

# Motivations

***Data sets of next generation galaxy surveys...***

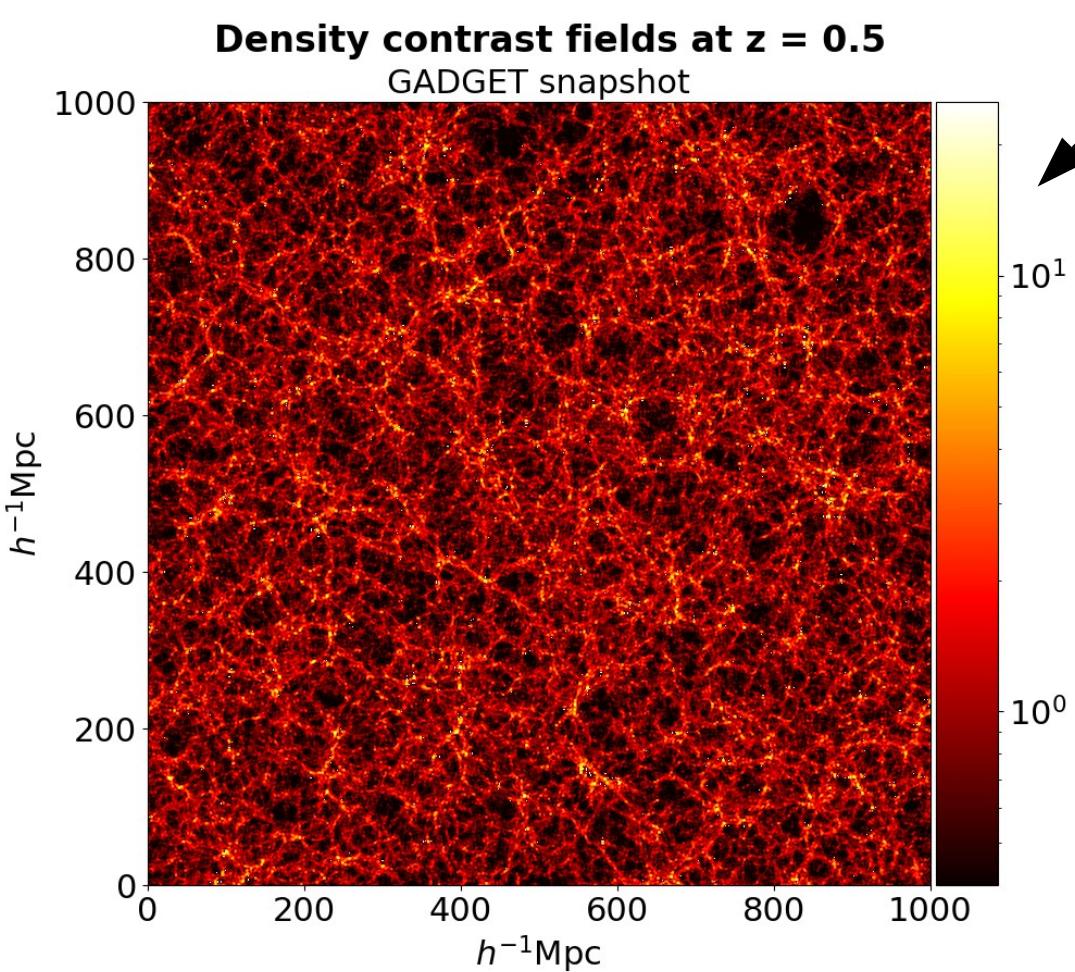
...will provide unmatched statistical power to constrain expansion history



Prime Focus  
Spectrograph

***Goals : constraining cosmological models, estimating cosmological parameters, comparing mocks and observations...***

# Generating mock data samples



$$\delta(x) = \frac{\rho(x)}{\bar{\rho}} - 1$$

**Real-space statistics**  
(matter PDF, matter correlation function)

**Stochasticity!**

→  $y$   
Dimension p

**Fourier Space statistics**  
(power spectrum, bispectrum...)

The most competitive solution is to run **parallel N-body simulations** (ex: CDM particles + baryons). Very expensive!

# Moments of clustering statistics

$$\begin{matrix} & \boldsymbol{y}_1 \\ \boldsymbol{y}_2 & \boldsymbol{y}_N \end{matrix} \xrightarrow{\text{Independent and identically distributed}} \begin{array}{l} \text{Estimator } \hat{\boldsymbol{\mu}} \text{ of } \mathbb{E} [\boldsymbol{y}] \\ \text{Estimator } \hat{\boldsymbol{\Sigma}} \text{ of } \mathbb{E} \left[ (\boldsymbol{y} - \mathbb{E} [\boldsymbol{y}]) (\boldsymbol{y} - \mathbb{E} [\boldsymbol{y}])^T \right] \end{array}$$

---

$$\bar{\mathbf{y}} = \frac{1}{N} \sum_n \mathbf{y}_n \longrightarrow \text{Standard deviation decreases as } \mathcal{O}(N^{-\frac{1}{2}})$$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{y}_i - \bar{\mathbf{y}}) (\mathbf{y}_i - \bar{\mathbf{y}})^T \equiv \frac{N}{N-1} \times \frac{1}{N} \sum_{i=1}^N \mathbf{Y}_i$$

→ rank-deficiency, ill-conditioning (eigenvalues blowing up)...  
4

# Using pairs to reduce variance (« CARPool »)

$y$   
 $c$

→ Intensive N-body simulations

→ Cheap *surrogate* (analytical computation, coarse Particle-Mesh code...)

*Construct new random variables  
from « simulation + surrogate «  
pairs with **mimimal variance***



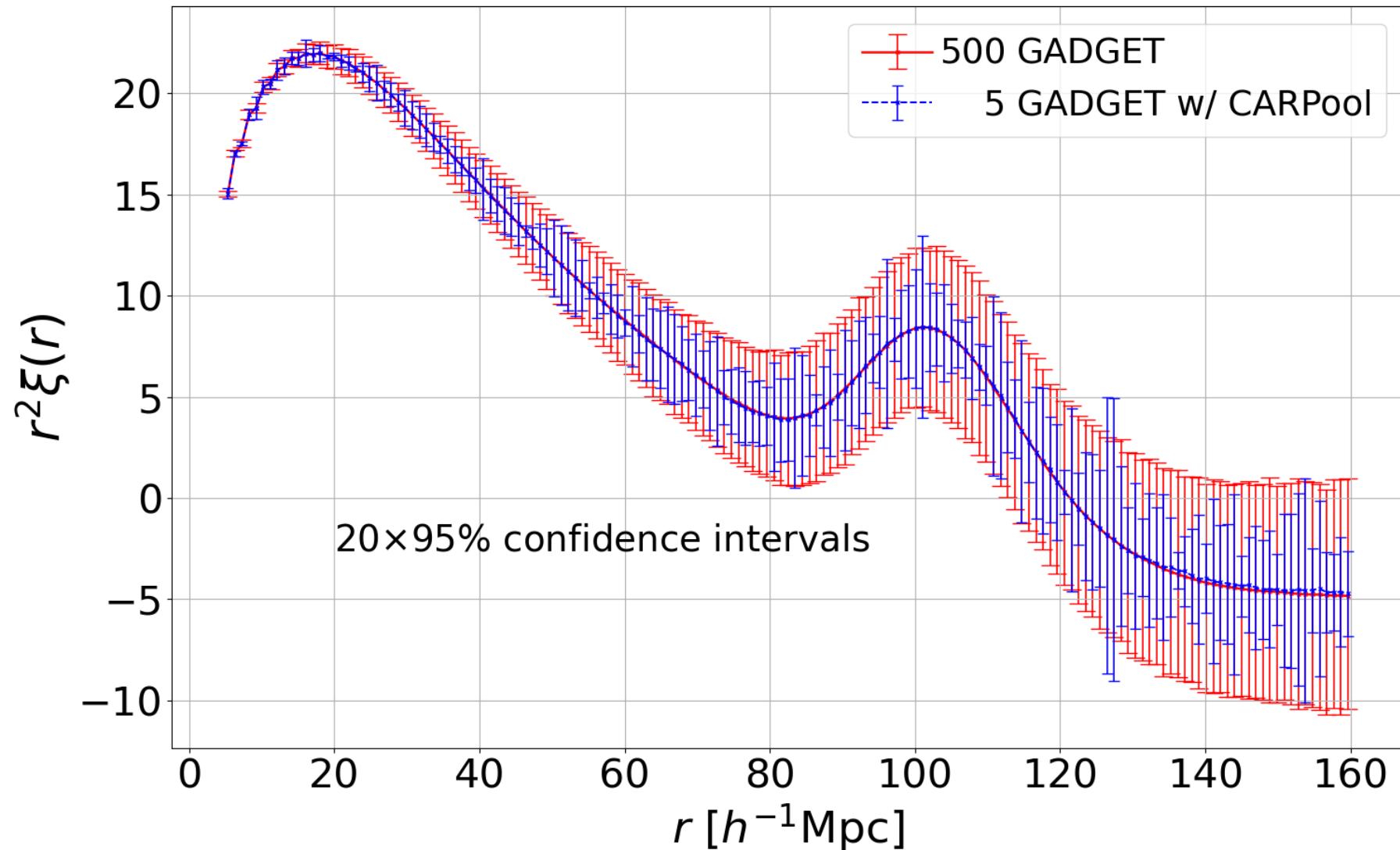
$$x_i(\beta) = y_i - \beta^T (c_i - \mu_c)$$
$$1 \leq i \leq p$$

# Numerical experiments

$\Lambda$ CDM cosmology, Redshift z=0.5

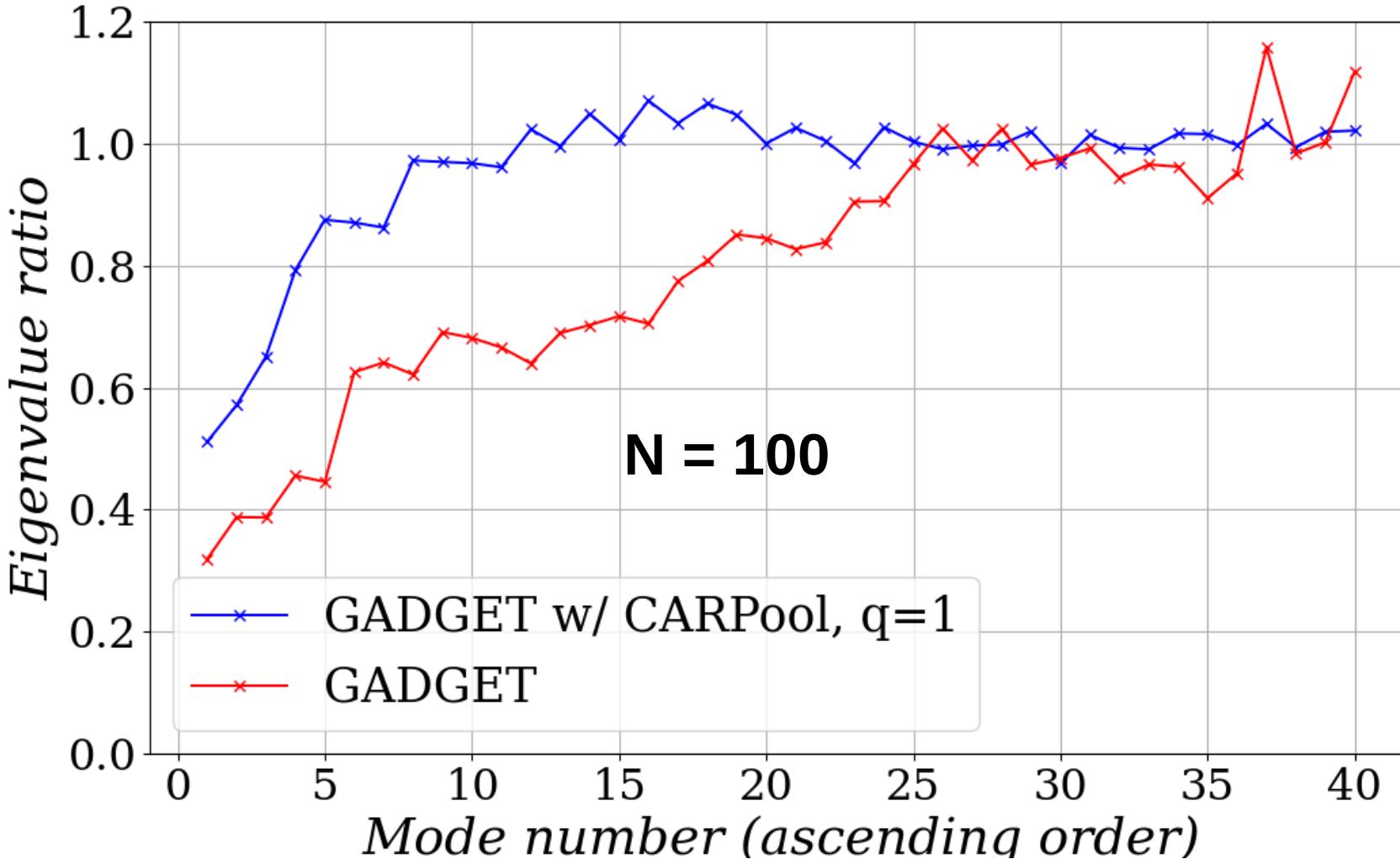
- $y$   N-body simulations – GADGET – are from the *Quijote Simulations* ([Villaescusa-Navarro et al., 2019](#))
- $c$   The cheap surrogate is L-PICOLA ([Howlett, 2015 b](#)),  
an MPI implementation of COLA ([Tassev, 2013](#))

# Mean of correlation function



- Student t-score confidence intervals

# Eigenvalue ratio (covariance of bispectrum)



- See *Bai and Yin (1993)* for beautiful results on the convergence of the smallest (or  $p-N+1$  smallest) eigenvalue(s) of the sample covariance matrix.
- The « true » covariance uses 12,000 sims.

# Take home messages

- If you have access to a **cheap** surrogate solver that is **highly correlated** with the simulations, try the method either for the mean or covariance estimation (non-invasiveness, easy theory...).
- Two papers for details : *Chartier N., Wandelt B., Akrami Y., Villaescusa-Navarro F., 2021, MNRAS, 503, 1897* (arXiv :arXiv:2009.08970) and *Chartier N., Wandelt B. D., 2021* (arXiv:2106.11718)



# Thank you !

