Cosmic Variance

Update your CV with CV

Control Variates

Gabriele Parimbelli
M. Crocce, R. Angulo

ICE INSTITUT DE CIÈNCIES DE L'ESPAI
CSIC
Consejo Superior de Investigaciones Científicas
IEEC

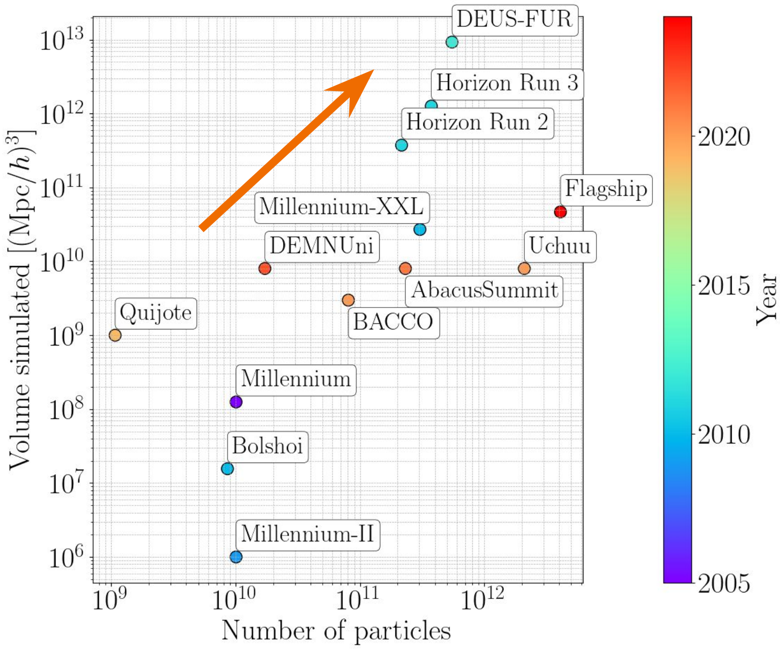
New strategies for cosmology
Sesto, July '25

What the compromise is about

Ideal simulation: $N = \infty \quad V = \infty$

Realistic simulation: $$\frac{\sigma_{P(k)}}{P(k)} \propto \frac{1}{\sqrt{V}}\left(1 + \frac{V}{NP(k)}\right)$$

Largest scales are easier to model but also the ones most affected by statistical uncertainty (cosmic variance)



“Axis” not shown is mass resolution
Elephant in the room is computational cost

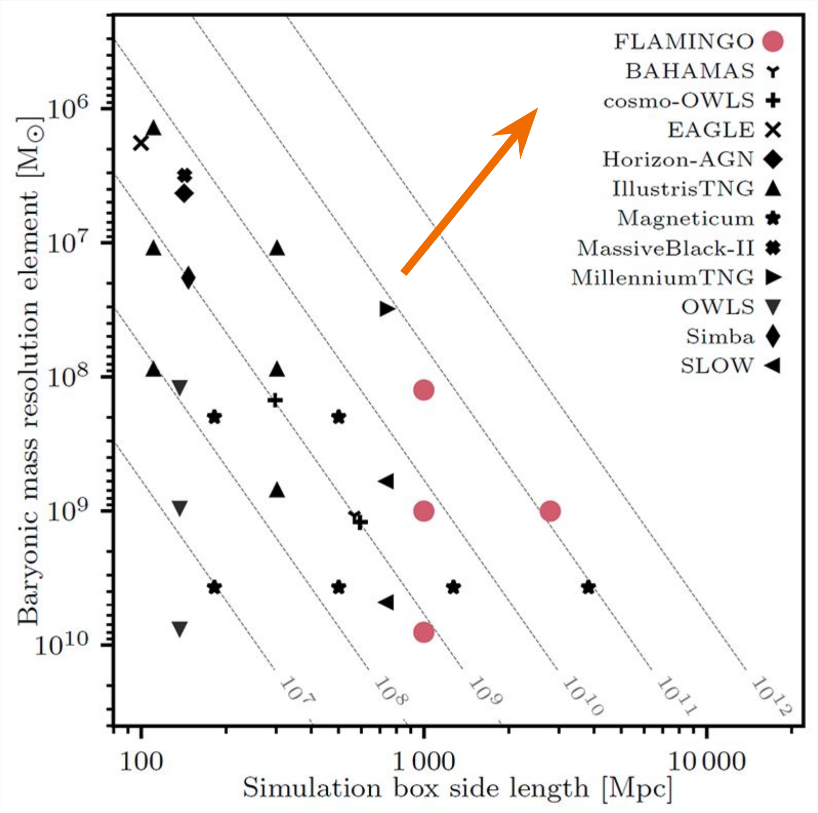
What the compromise is about

Ideal simulation: $N = \infty \quad V = \infty$

Realistic simulation: $\frac{\sigma_{P(k)}}{P(k)} \propto \frac{1}{\sqrt{V}}\left(1 + \frac{V}{NP(k)}\right)$

Largest scales are easier to model but also the ones most affected by statistical uncertainty (cosmic variance)

This is even more evident when dealing with hydro simulations

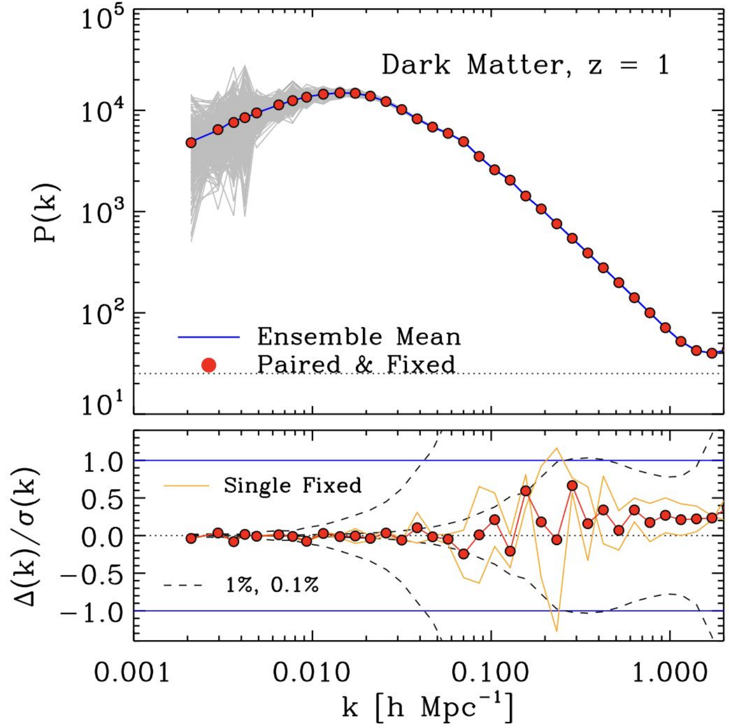


Schaye+23

Variance reduction methods

Several techniques have been used modifying initial conditions

Type	Amplitudes $\|\delta_{ic}(k)\|$	Phases $\phi_{ic}(k)$
Standard	Rayleigh	Uniform $[0,2\pi)$
Paired	Rayleigh	ϕ_k , $\phi_k+\pi$
Fixed	Dirac δ_D	Uniform $[0,2\pi)$
Fixed + paired	Dirac δ_D	ϕ_k , $\phi_k+\pi$



Angulo, Pontzen, 2016

Minimize variance AND noise?

Control variates

Reducing the variance of a random variable (**X**) using a correlated surrogate (**C**)

$$\underbrace{Y}_{\text{Control variate}} = \underbrace{X}_{\text{Fiducial}} - \beta\,(\underbrace{C}_{\text{Surrogate}} - \underbrace{\mu_c}_{\text{Surrogate mean}})$$

Mean

$$\langle Y \rangle \equiv \langle X \rangle$$

Regardless of the value of β

Control variates

Variance

Optimal choice for $\beta = \frac{\mathrm{Cov}[X, C]}{\mathrm{Var}[X]}$

Correlation coefficient $\rho^2_{XC} = \frac{\mathrm{Cov}[X, C]}{\mathrm{Var}[X]\ \mathrm{Var}[C]}$

negligible

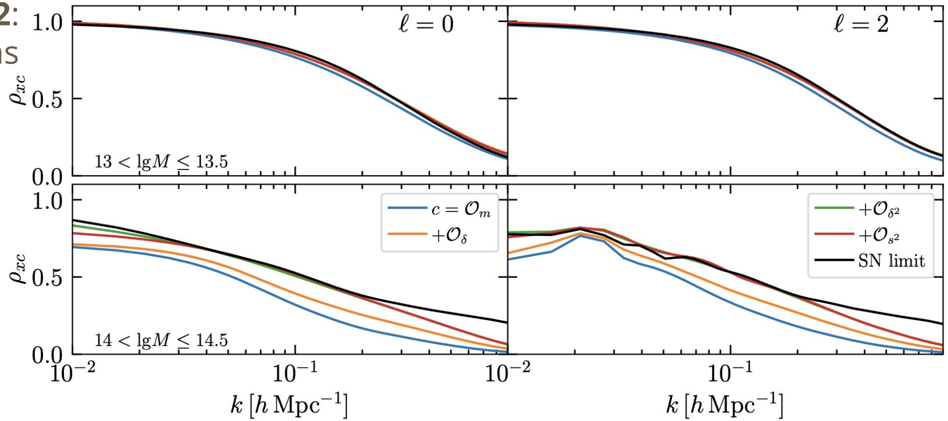
Variance reduction

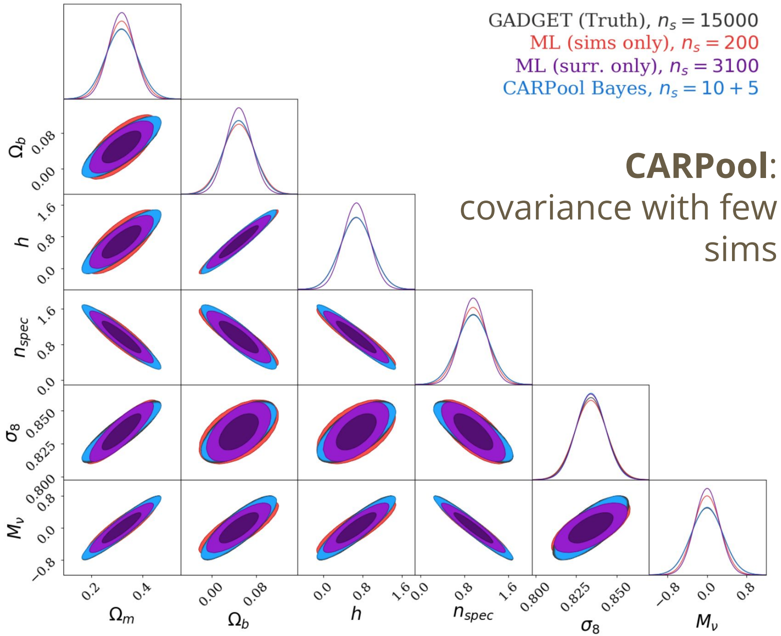
$$\boxed{\mathrm{Var}[Y] = \mathrm{Var}[X]\left(1 - \rho^2_{XC} + \beta^2\ \cancel{\mathrm{Var}[\mu_C]}\right)}$$

- Reduce number of sims for covariance estimate (Chartier+22)
- Mitigating noise of DESI mocks for BAO reconstruction (Hadzhiyska+23)
- Improve clustering predictions of volume-limited hydro sims (Doytcheva+24)

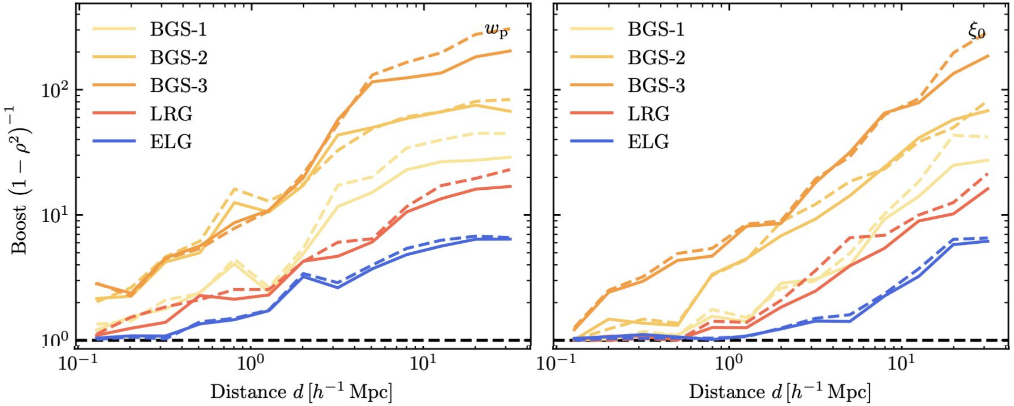
Previous works

DeRose+22: Lagrangian bias





CARPool: covariance with few sims



++ Hadzynska+24: mitigating noise with CV in BAO reconstruction

Doytcheva+24: volume gain in hydro sims

* 50 realizations for small and 8 for big

Our setup

SIMULATIONS

2 box sizes:

- "small" 512 Mpc/h 384^3 particles
- "big" 1440 Mpc/h 1080^3 particles

Realizations (ICs):

- fixed amplitude
- 2 opposite phases

Gravity solver:

- **N-body**
- **m2m** = ZA + "map2map" emulator (Jamieson+22) + Lagrangian bias expansion

Observable: P(k)

- dark matter
- SHAM galaxies* [n_g = 0.001, 0.00054 $(h/Mpc)^3$]

$$Y = X - \beta\,(C - \mu_c)$$

big N-body | small N-body | small m2m | **big m2m**

METHODS:

- direct application
- fit of Lagrangian bias to N-body
- maximization of rho(k)

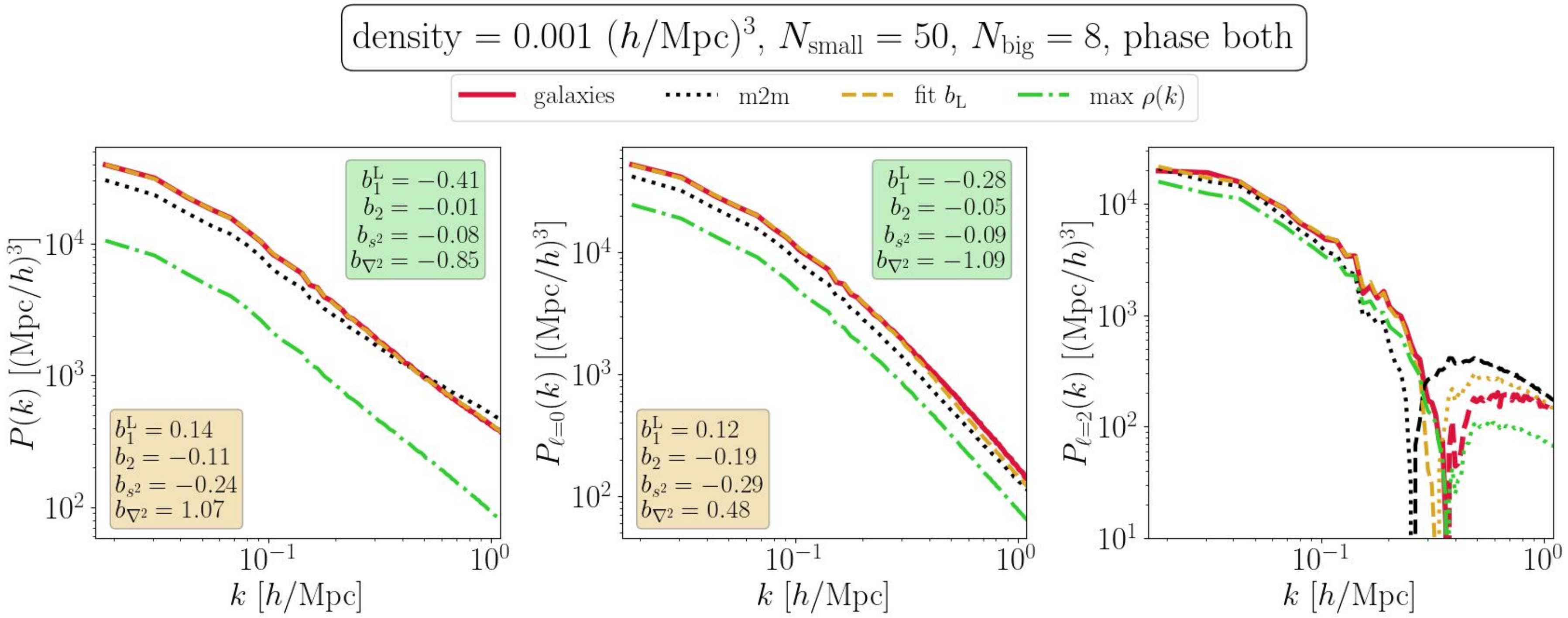
APPROXIMATIONS:

- disconnected approximation

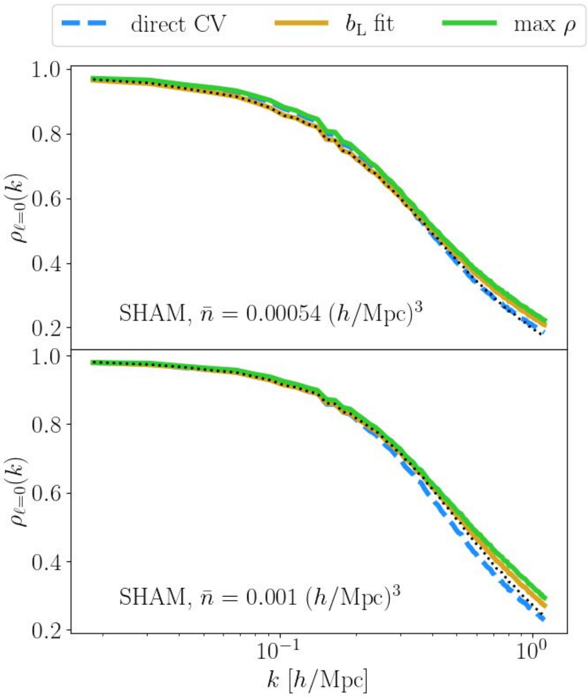
$$\mathrm{Cov}[P_X,P_C] \cong \mathrm{Var}[P_{XC}]$$

- small-scale filtering

Results: methodologies



Results: correlation coefficient

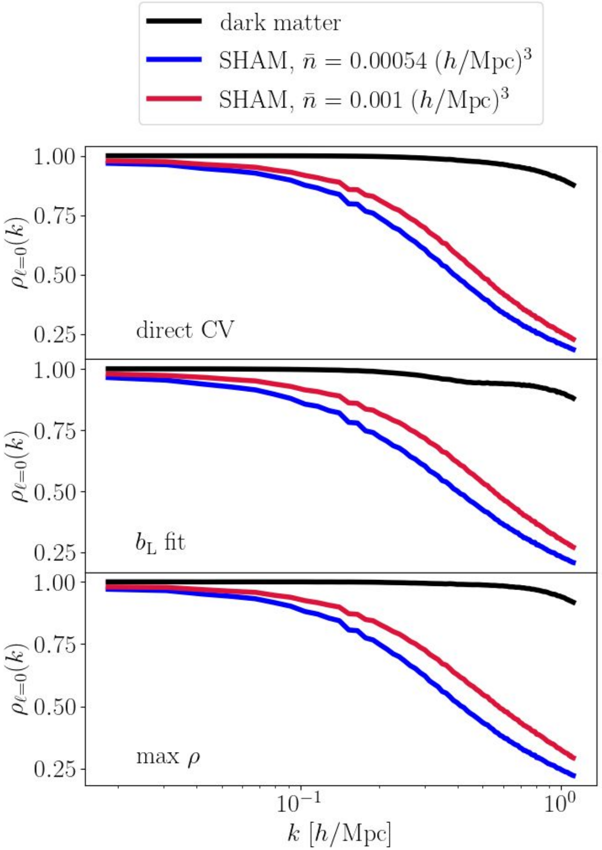


Impact of method

moderate improvement with m2m w.r.t. ZA-only

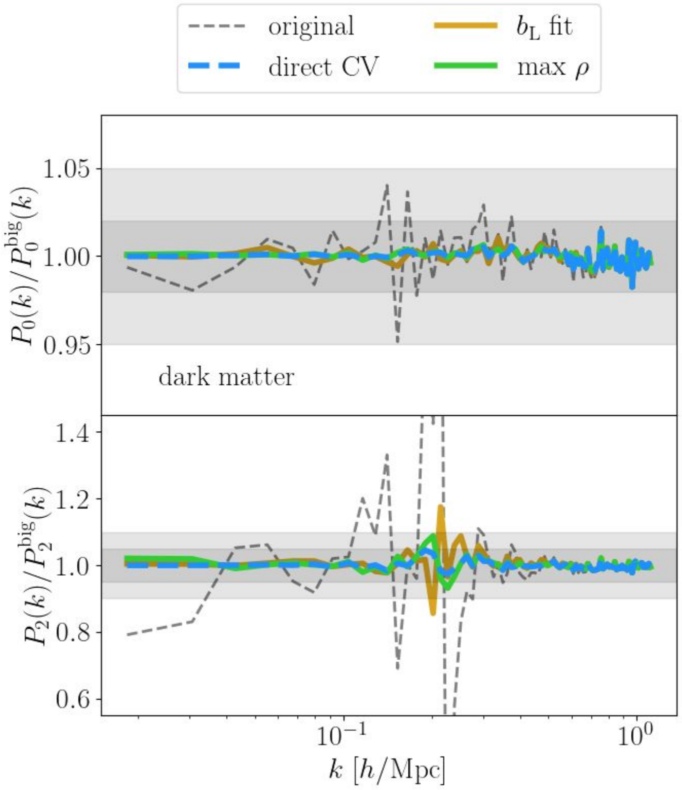
Impact of tracer

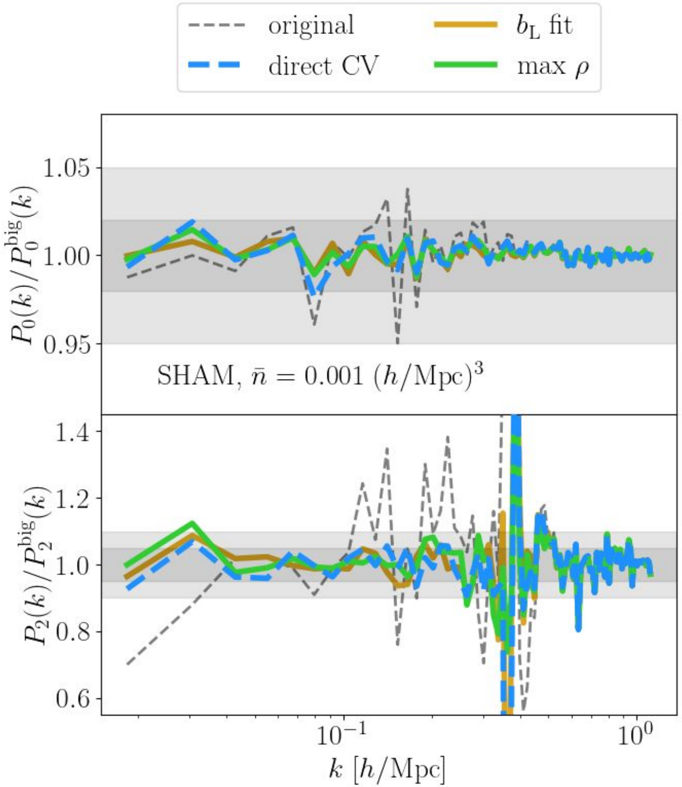
sparser objects and randomness of galaxy-halo connection make it more difficult to keep correlations high



Results: CV prediction

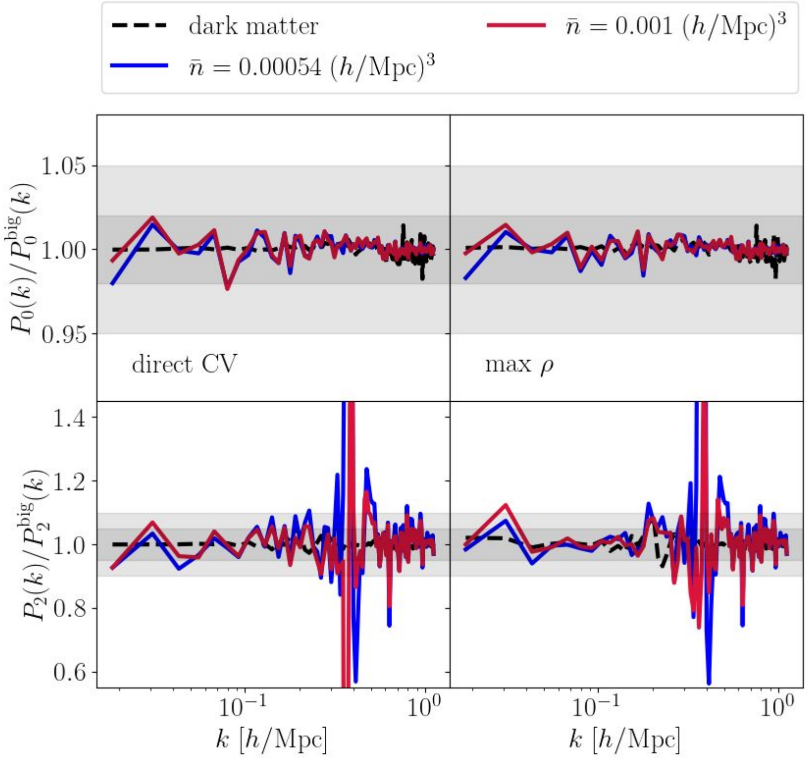
- For DM direct works better than Lagrangian bias fit
- For SHAM we can recover P(k) of big sim at ~2% level



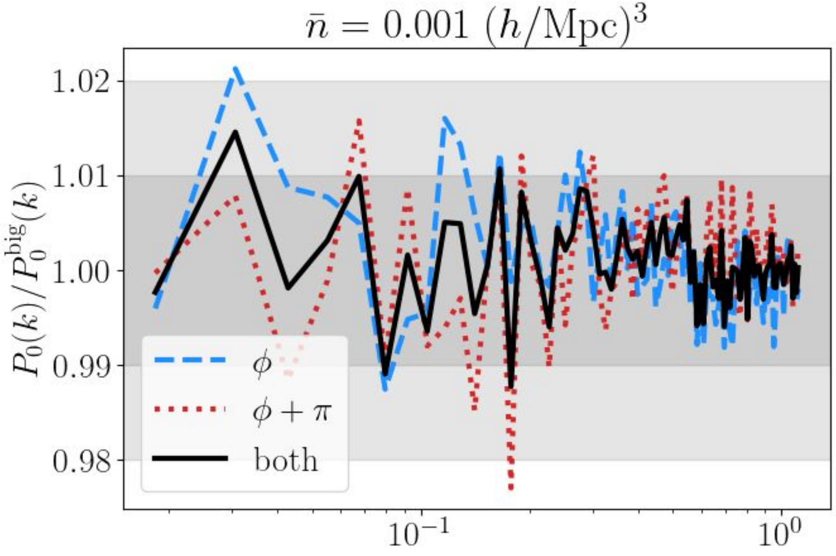


Results: impact of number density and phases

impact of number density on CV



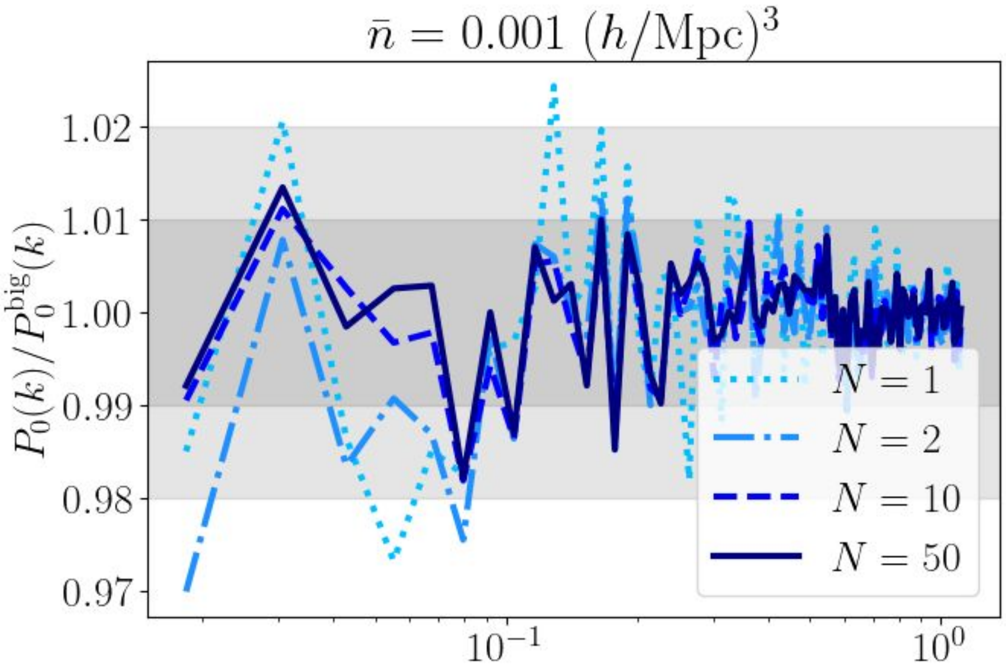
Opposite phases help increasing the accuracy even more



Results: impact of number of SHAM realization

Increasing the number of SHAM realizations increases the accuracy

Valid for each methodology used (direct, Lagrangian bias, maximization of ρ)



Further applications

$$Y = X - \beta(C - \mu_c)$$

N-body/SHAM

LPT/COLA

$$Y = X - \beta(C - \mu_c)$$

Hydro

N-body/SHAM

1. Beat down cosmic variance in hydro
2. Super realistic galaxy-halo connection
3. Library of mocks with varying cosmology
4. **Priors to galaxy bias!**

Conclusions

- Hydro simulations are costly and limited in volume
  - handful of galaxy formation models
- **Control variates** represent a useful tool to beat down both cosmic variance and noise
  - reproduce larger volume with a surrogate observable (ZA for N-body, HOD/SHAM/SAM for hydros?)
  - **realistic** galaxy-halo connections to put priors to galaxy bias
- **Maximize** signal extracted by reproducing summary statistics from a big box using a small box
  - *novelties*: fixed+paired+CV all together, m2m as an improvement to ZA
  - *methods*: direct application of CV, Lagrangian bias fit, maximization of $\rho(k)$
  - *results*: large number of SHAM, opposite phases, max. $\rho \rightarrow$ reproduce P(k) at ~1-2%
  - *caveats*: difficult to quantify gain, disconnected approximation…
  - *next*: bispectrum (cross-covariance?)