A Monte Carlo galaxy catalogue generator to predict covariance matrices ?

Atelier Outils de l'action Dark Energy

Philippe Baratta (CPPM/CPT)
PhD student at Aix-Marseille Univerisité

supervisor:
Julien Bel (CPT)
Anne Ealet (IP2I)

collaborator:
Stephane Plaszczynski (LAL)

Motivations

▶ Chalenging the accuracy of galaxy surveys implies to develop strong statistical methods in LSS data analysis to constrain the large variety of cosmological models. The obsevational chain must be controled and unbiased

Need for reliable covariance matrix for a given observable

▶ What kind of observable ? Need for a direct observable that does not suffer from any fiducial bias :

the angular power spectrum C_ℓ

▶ Possibility: galaxy catalogue simulations

- Fast —> **Monte Carlo** sampling
- Few characteristics but well controled

Input

Theoretical Power spectrum P(k)

Probability distribution function

$\longrightarrow$

Output

Cl's

Sampling a field with a given p.d.f. and power spectrum

▶ The simpliest case of a **gaussian p.d.f.** of a matter field

- In a periodic box of lenght L and number of sample per side Ns, we define

$$\left\langle \delta_{\vec{k}} \delta_{\vec{k'}} \right\rangle = \delta^K(\vec{k} + \vec{k'}) k_f^{-3} P(\vec{k})$$

$\delta(\vec{x}) = \Delta\rho(\vec{x})/\rho_0$ density contrast field

$k_f = 2\pi/L$ the fundamental mode

it can be shown that $\delta_{\mathbf{k}} = \sqrt{-\mathcal{P}(\mathbf{k})/k_f^3 \ln(1-\epsilon_1)} e^{2\pi\epsilon_2}$

$\epsilon_1, \epsilon_2 \in [0,1]$ from uniform distributions

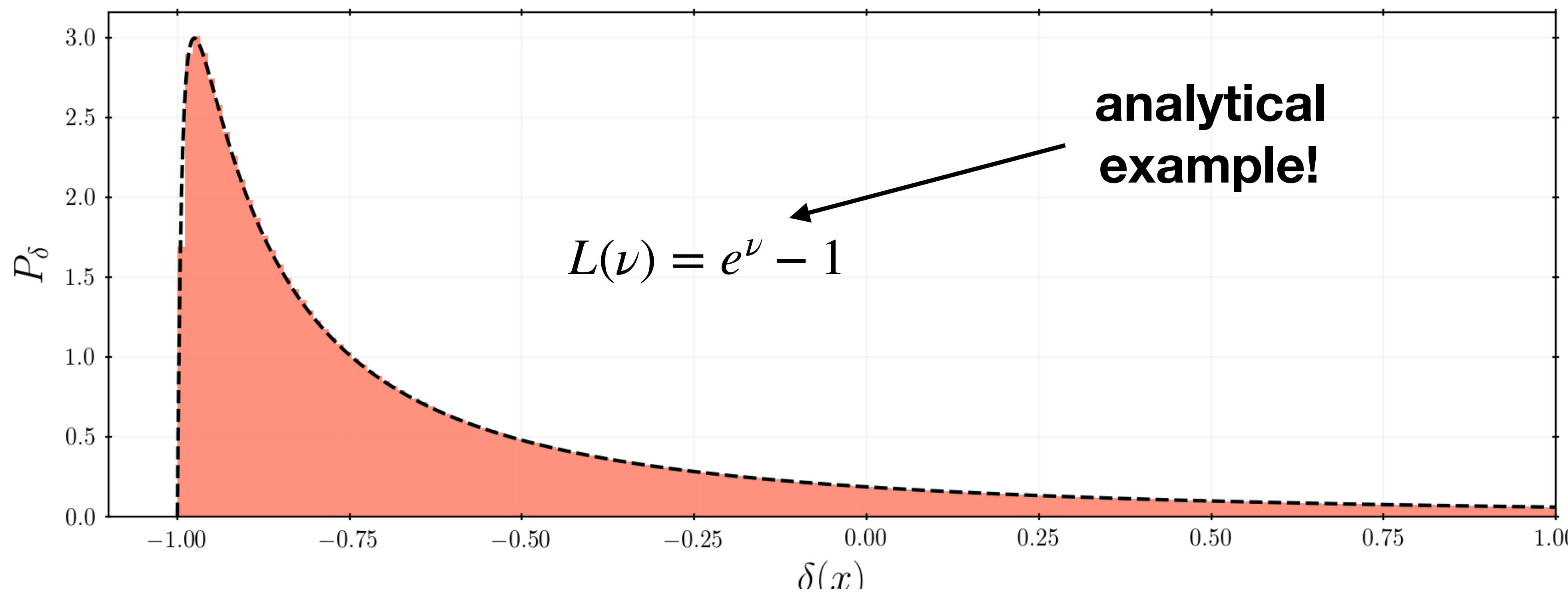
- p.d.f and spectrum as expected with high level of confidence

- Well suited for CMB but dark matter clustering being non linear —> dark matter halo p.d.f non gaussian —> galaxy distribution non gaussian as well that introduce correlations between modes appearing in the covariance matrix of spectra

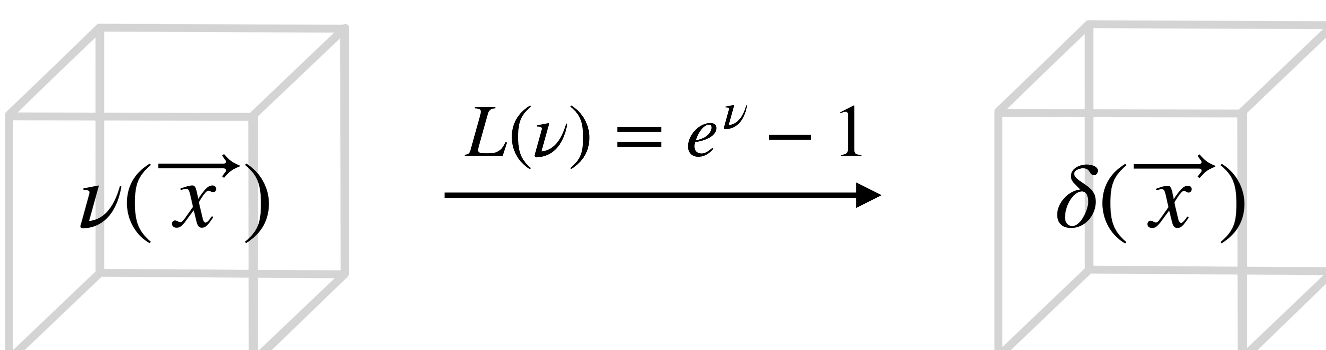
Sampling a field with a given p.d.f. and power spectrum

- The case of a non gaussian field

What choice of pdf ? it can be shown (*Coles & Jones (1991), Clerkin et al. (2017)*) that the log-normal shape is a good approximation to represent the galaxy field



Does it works?



Gaussian field **Non Gaussian field**

pdf controled

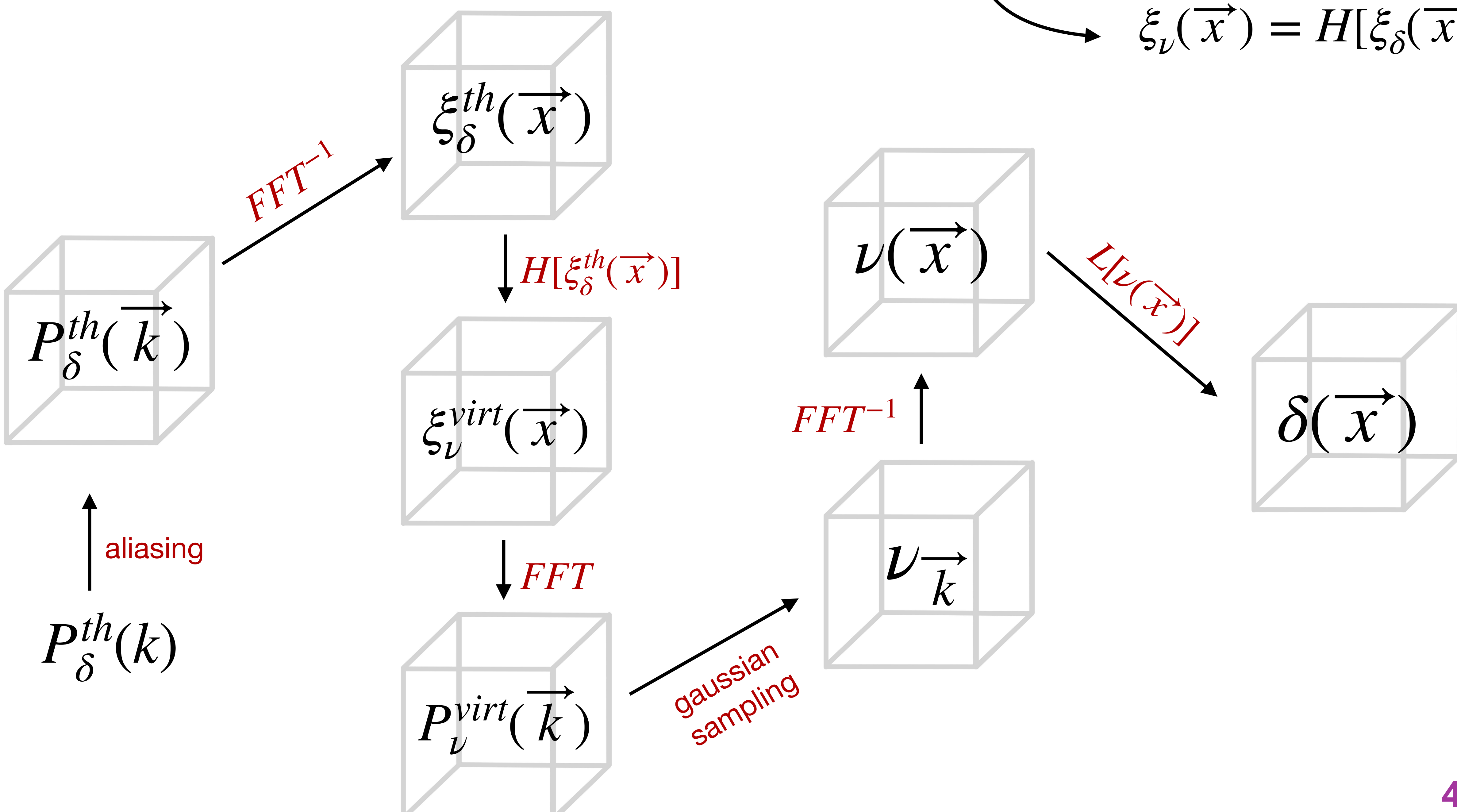
spectrum not controled

Sampling a field with a given p.d.f. and power spectrum

- The case of a non gaussian field
  - Be $\nu(\overrightarrow{x})$ the initial gaussian field and $\delta(\overrightarrow{x})$ the desired non gaussian field
  - $\xi_\nu = FFT^{-1}[P_\nu]$ must be well designed to get a well shaped $\delta(\overrightarrow{x})$
  - Must work on $\xi_\delta \equiv \langle \delta_1 \delta_2 \rangle = \int L(\nu_1) L(\nu_2) \mathscr{B}(\nu_1, \nu_2, \xi_\nu) d\nu_1 d\nu_2$

must find out

$$\xi_\nu(\overrightarrow{x}) = H[\xi_\delta(\overrightarrow{x})]$$



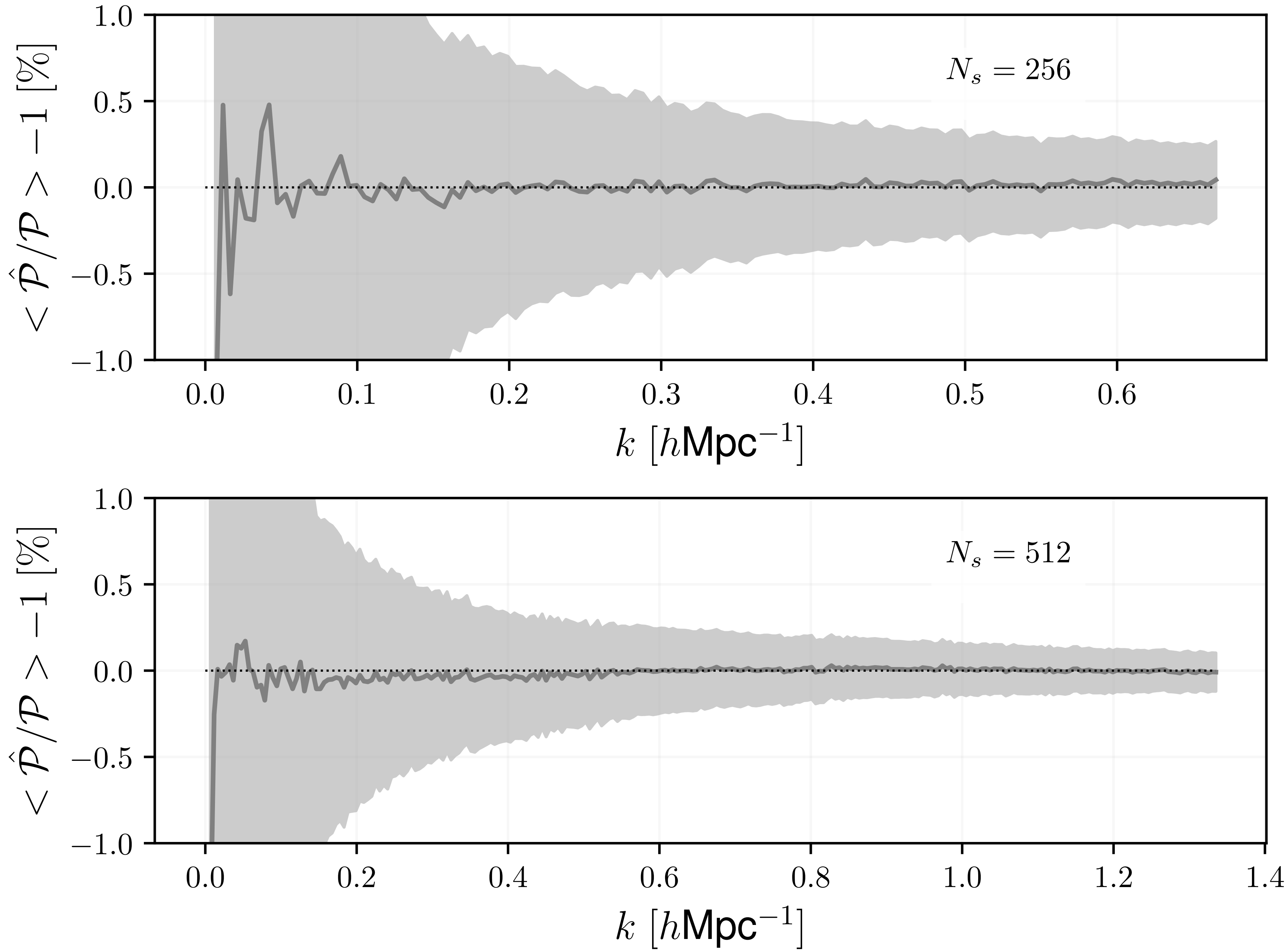


Fig. 2. For 1000 realisations of the density field we compute the averaged 3D power spectrum that we compare relatively to the expected 3D power spectrum. We then compute the shell-averaged monopoles of this residuals in shells of width $|\mathbf{k}| - k_f/2 < |\mathbf{k}| < |\mathbf{k}| + k_f/2$. The result is presented in percent with error bars. The used setting is a sampling number per side of 256 in the top panel and 512 for the other, all in a box of size $L = 1200h^{-1}$Mpc at redshift $z = 0$.

- Need to caracterise correlations between modes: **the trispectrum**

$$\left\langle \delta_{\overrightarrow{k_1}}\delta_{\overrightarrow{k_2}}\delta_{\overrightarrow{k_3}}\delta_{\overrightarrow{k_4}} \right\rangle_c = \delta^D(\overrightarrow{k_1}+\overrightarrow{k_2}+\overrightarrow{k_3}+\overrightarrow{k_4})\ T(\overrightarrow{k_1},\overrightarrow{k_2},\overrightarrow{k_3})$$

- That have an influence on the **covariance matrix** (*Scoccimarro et al. (1999)*) :

Definition
$C(X) = E[(X - E[X])(X - E(X))^T]$

$$C_{ij} = \frac{\mathcal{P}(k_i)^2}{M_{k_i}}\delta^D_{ij} + k_F^3 \bar{T}(k_i, k_j)$$

$$\bar{T}(k_i,k_j) = \int_{k_i}\int_{k_j} T(\mathbf{k}_1, -\mathbf{k}_1, \mathbf{k}_2, -\mathbf{k}_2)\frac{\mathrm{d}^3\mathbf{k}_1}{V_{k_i}}\frac{\mathrm{d}^3\mathbf{k}_2}{V_{k_j}}$$

- Developping this relation, we get the approximate expression for the diagonal

$$\begin{aligned}\bar{T}(k_i,k_i) \sim\ & 8c_1^2\left\{4c_2^2 + 3c_3c_1\right\}\mathcal{P}^3(k_i) + \\ & + 24\left\{3c_1^2c_3^2 + 4c_1c_2^2c_3 + 12c_1^2c_2c_4\right\}\mathcal{P}^2(k_i)\mathcal{P}^{(2)}(k_i) + \\ & + 144c_1^2c_3^2\mathcal{P}^{(2)}(0)\mathcal{P}^2(k_i)\end{aligned}$$

$$c_n = \frac{1}{n!}\int_{-\infty}^{\infty} L(\nu)H_n(\nu)P_\nu(\nu)d\nu$$ the coefficients of the Hermite polynomials

NL transformation

gaussian p.d.f.

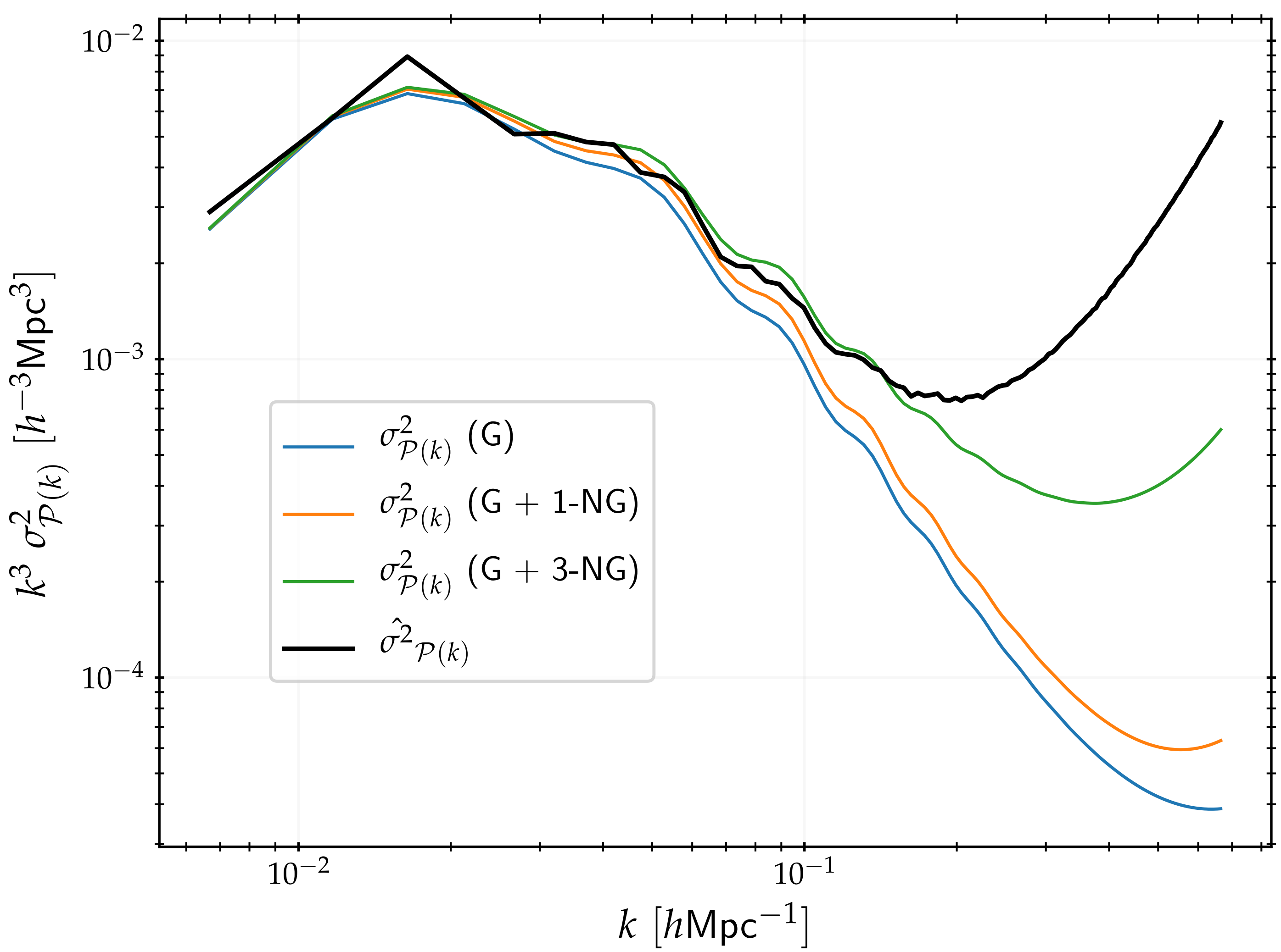


Fig. 5. Measured diagonal of the covariance matrix for 7375 power spectra realizations of the density field using the described method (black curve). The other curves represent their predictions taking into account the gaussian part alone (G) or by adding some non gaussian contributions of equation (18). For example in (1-NG) one keeps only the term in $\mathcal{P}^3(k_i)$ in the trispectrum development presented in equation (20) while in (3-NG) we keep all of them.

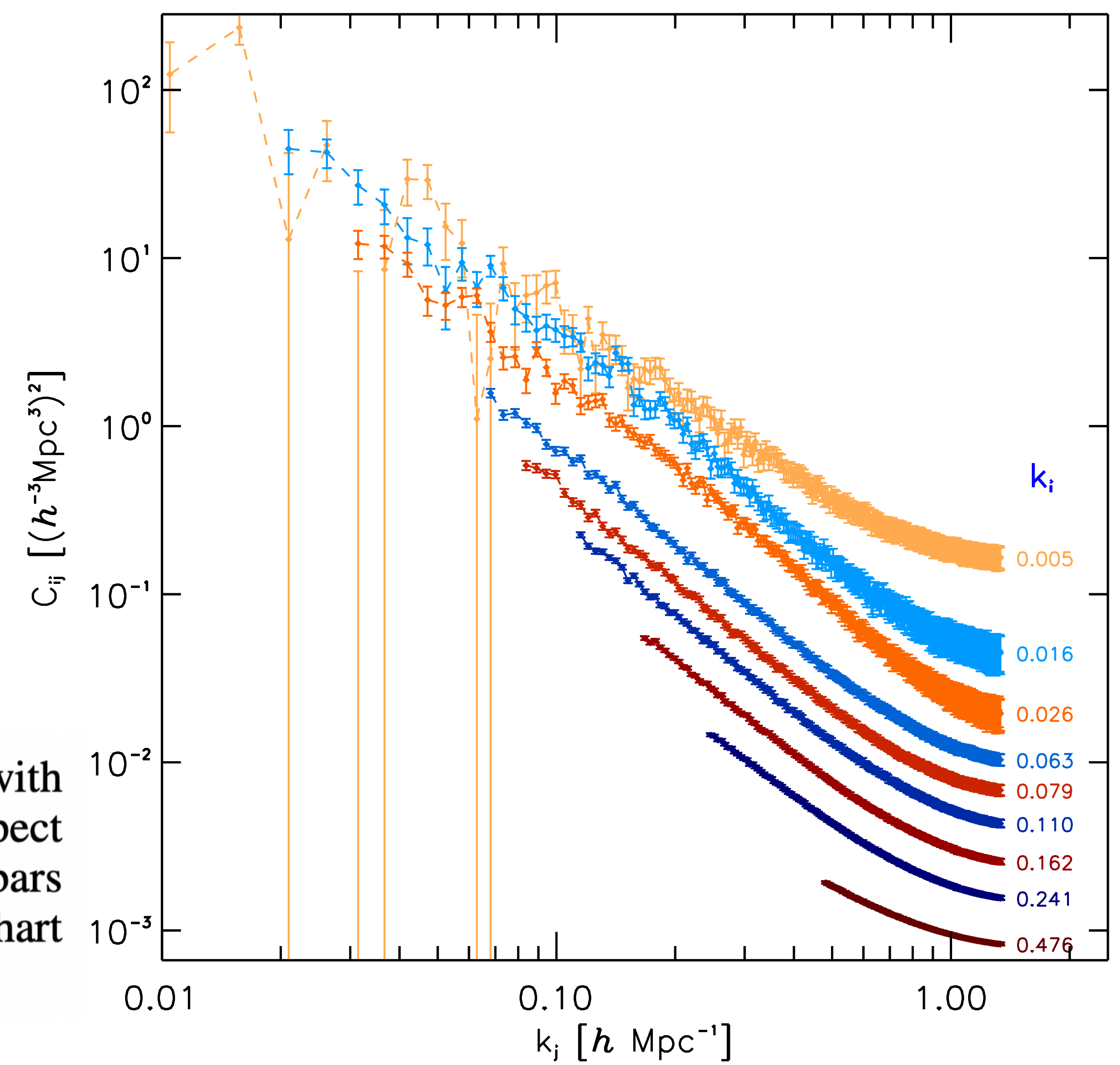
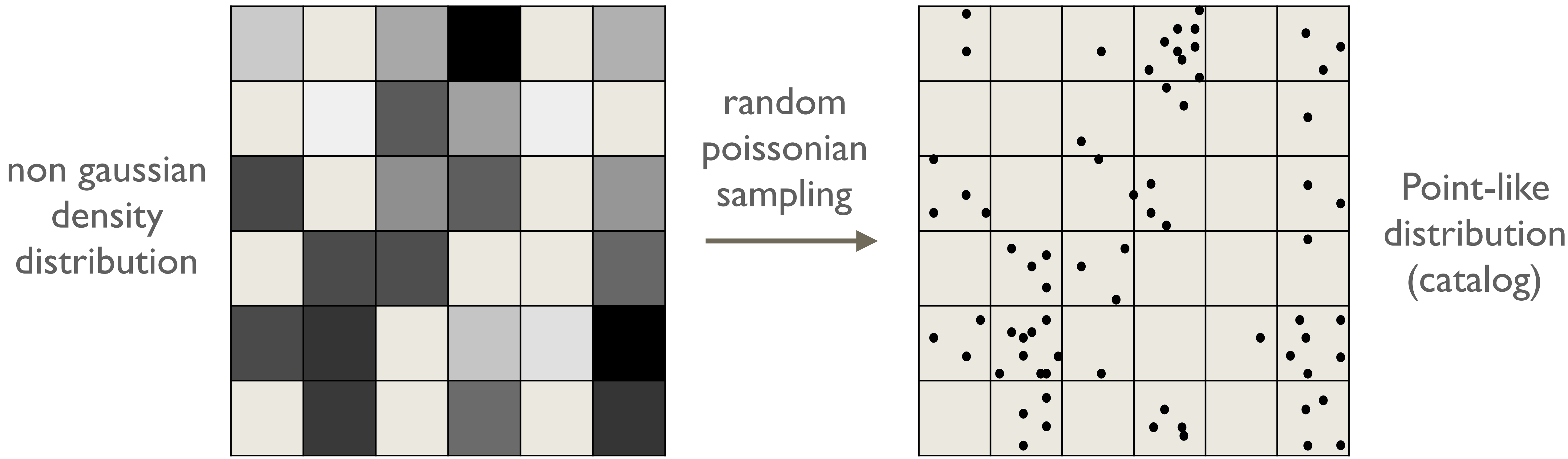


Fig. 6. Off diagonal elements of the covariance matrix estimated with $N = 7375$ realisations, showing the dependance of the C_{ij} with respect to k_j at various fixed k_i labeled on the right of the panel. The error bars are obtained assuming that the covariance coefficients follow a Wishart distribution, i.e. $V[\hat{C}_{ij}] = (C_{ij}^2 + C_{ii}C_{jj})/(N-1)$.

▶Poissonian sampling

- Snapshot : choose an average density and a smoothing scheme



- Predicted spectrum :

$$\hat{P}(\overrightarrow{k}) = |W_{TH}(\overrightarrow{k})|^2 P^{th}(\overrightarrow{k}) + \frac{1}{\rho_0 (2\pi)^3}$$

Convolution function

Shot noise

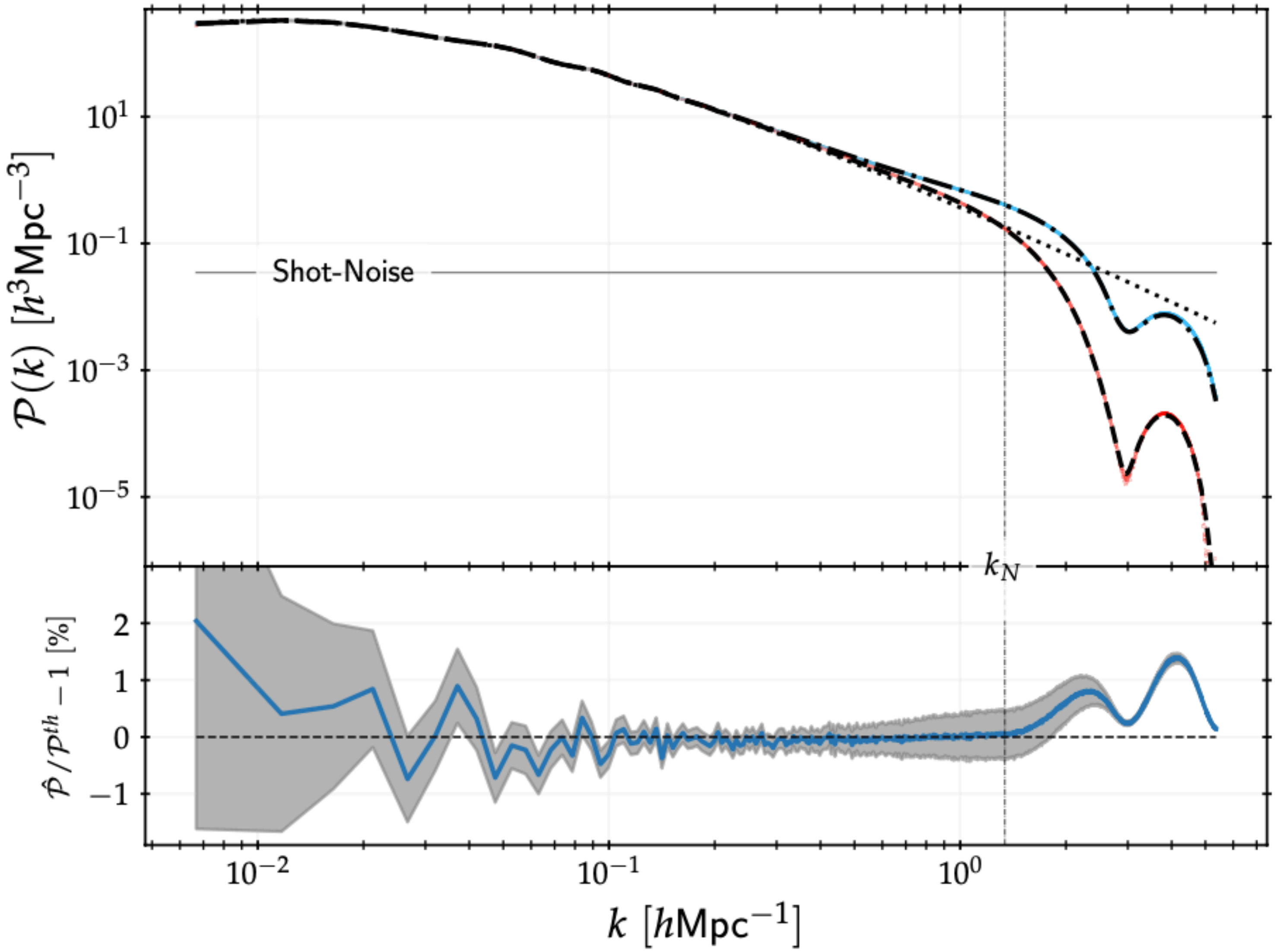
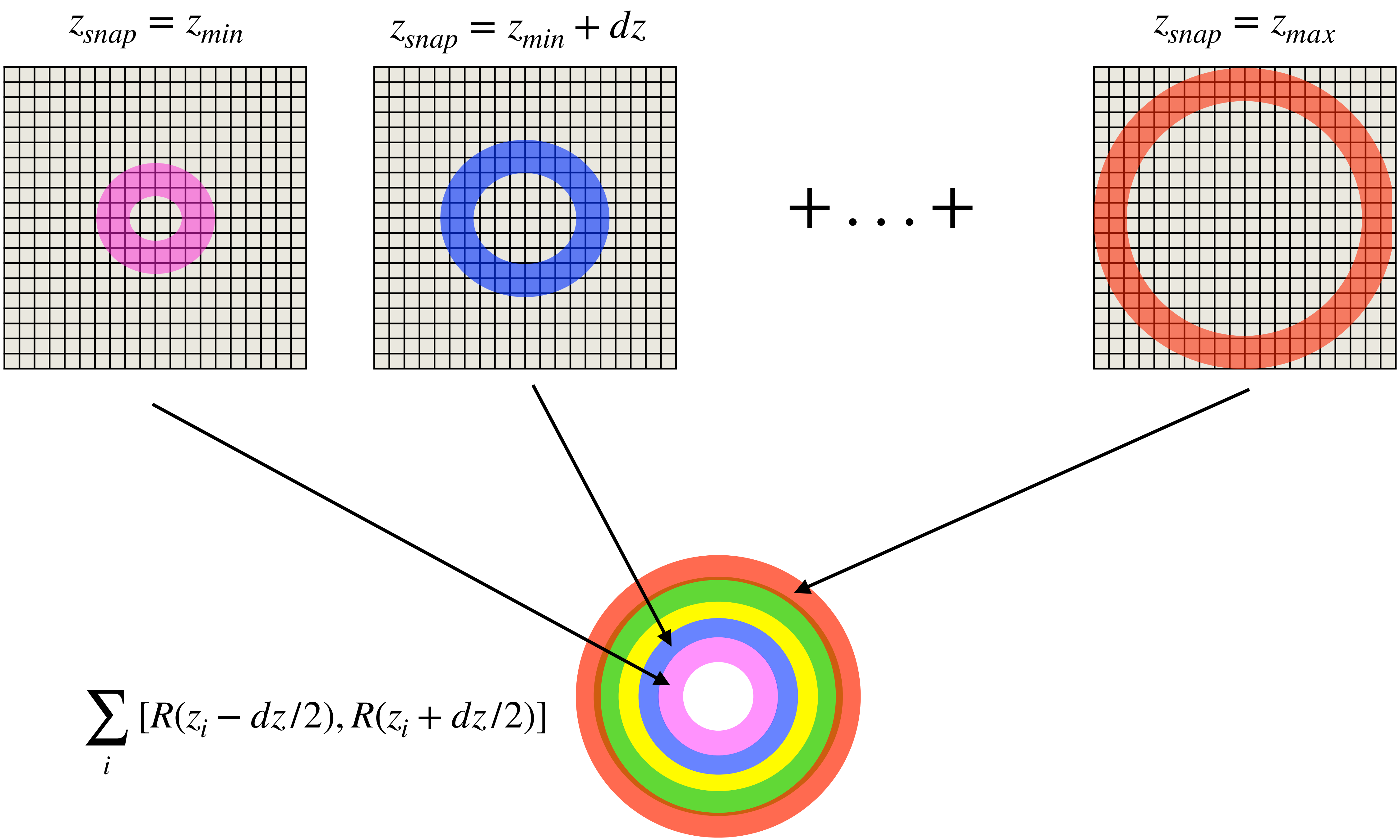


Fig. 10. *Top:* Measured power spectra averaged over 100 realisations of the poissonnian LN field for the TopHat interpolation scheme (blue curve with prediction in semi-dotted black line) and for the linear interpolation scheme (red curve and prediction in dashed line). Note that the shot-noise is subtracted from measures (dotted horizontal line) and is about $3.48 \times 10^{-2}\ h^3\mathrm{Mpc}^3$. The dotted black curve represents the alias-free theoretical power spectrum computed by `CLASS`. *Bottom:* Relative deviation in percent between the averaged realisations (with shot-noise contribution) and prediction (with the same shot-noise added) in blue line with error bar in grey for the TopHat interpolations schemes. Snapshots are computed for a grid of size $L = 1200\ h^{-1}\mathrm{Mpc}$ and parameter $N_s = 512$. Here comparisons are made well beyond the Nyquist (vertical line) frequency at $k_N \sim 1.34\ h\mathrm{Mpc}^{-1}$.

▶ Generate a galaxy catalogue

- choose z_{min},z_{max} for your catalogue and generate N_{shl} snapshots at intermediate redshifts
- place ourself at the center of each box
- select shells in snapshots that correspond to the comoving volume of the redhift interval of the snapshot
- glue all shells to reconstruct the lightcone



$$\sum_i [R(z_i - dz/2), R(z_i + dz/2)]$$

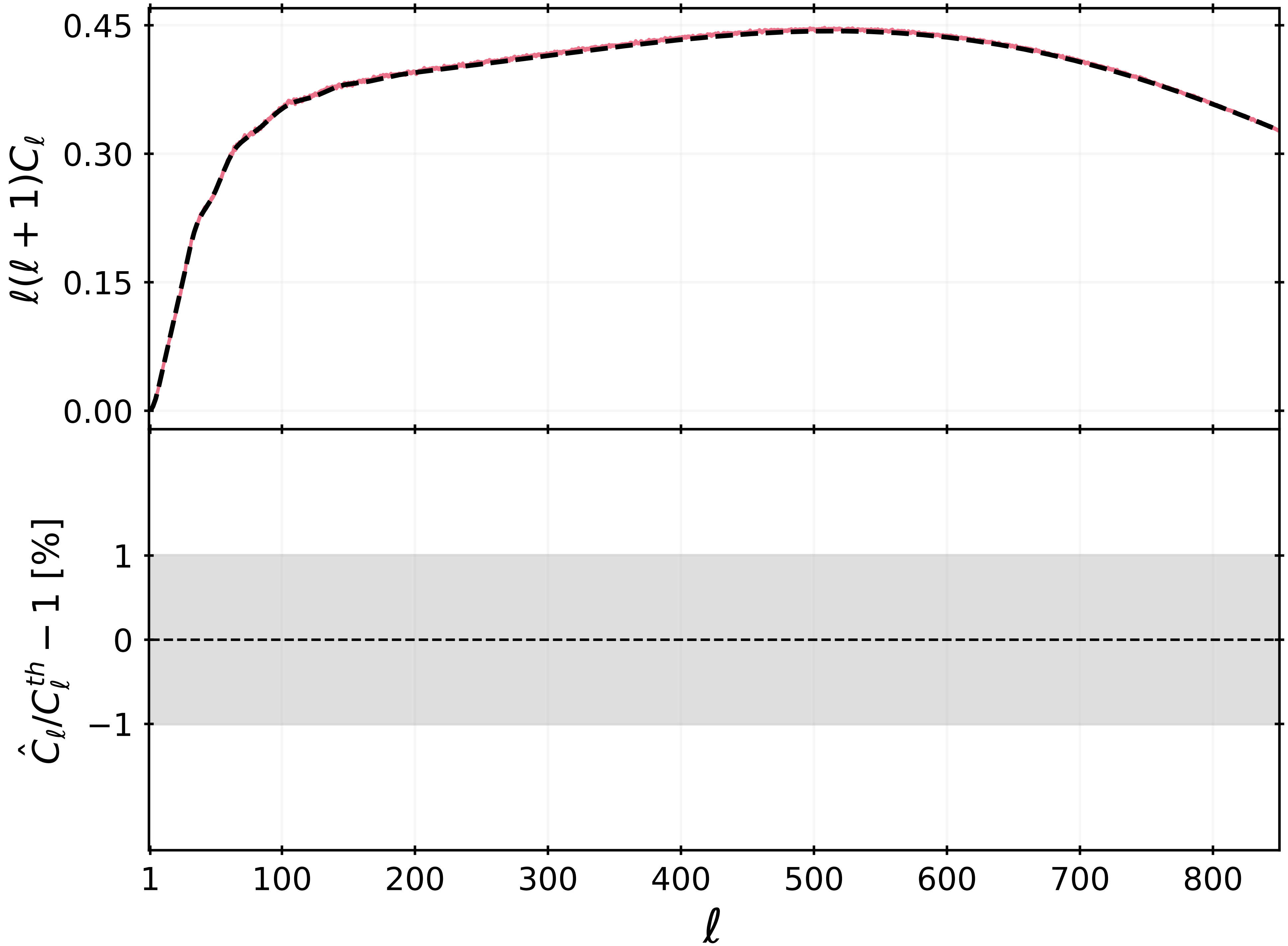


Fig. 8. Top panel : Thousand averaged C_ℓ's for simulated light cones using the shell-method with error bars (red curve) and corresponding prediction (dashed black curve). We simulate here a lightcone between redshifts 0.2 and 0.3 in a sampling $N_s = 512$ and a number of shells $N_{shl} = 250$ to ensure a sufficient level of continuity in the density field. Center panel : relative deviation in percent of the averaged C_ℓ's from prediction with error bars in red.

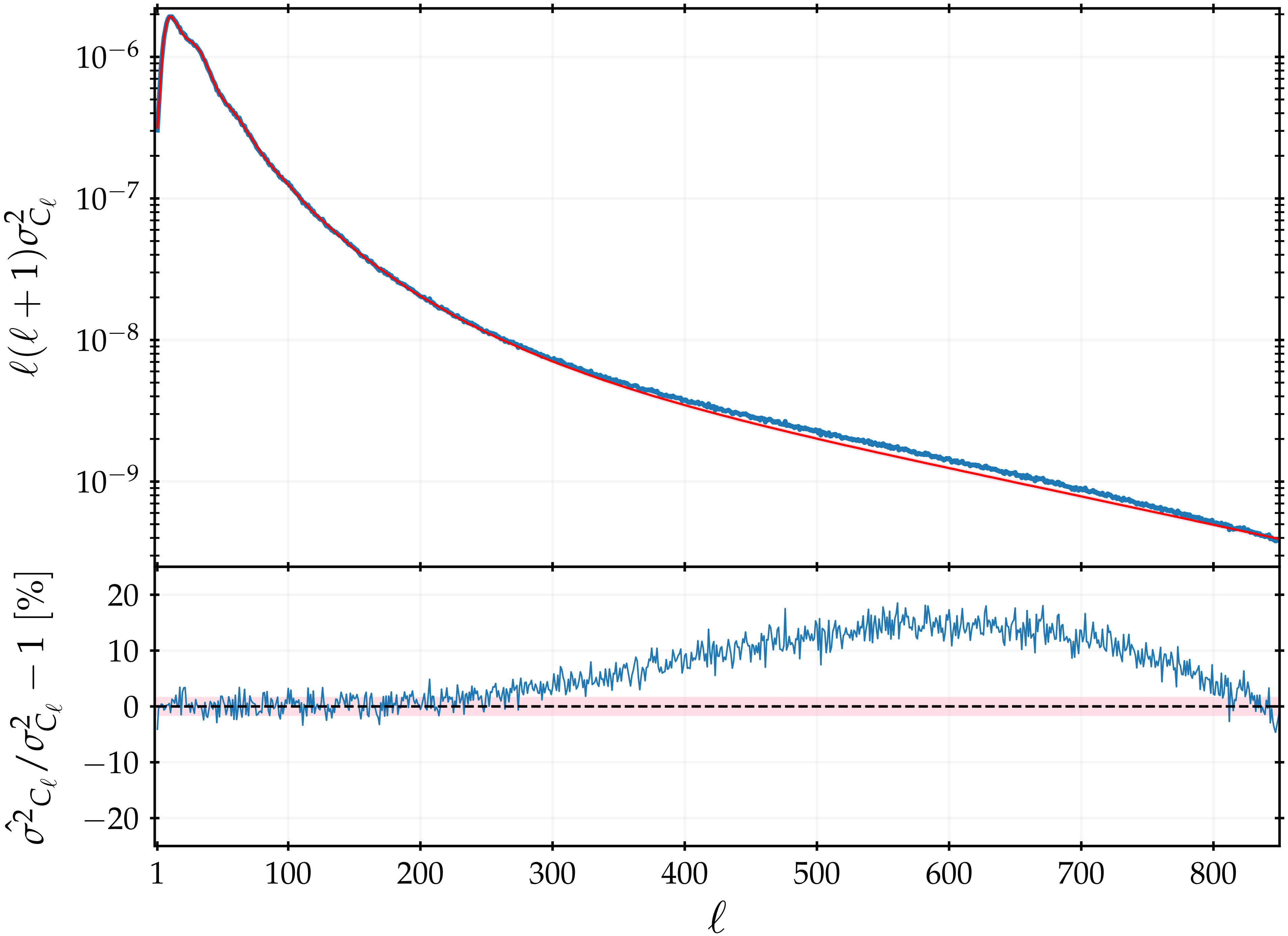


Fig. 13. *Top :* Measured diagonal of the covariance matrix (blue curve) over $N = 10000$ realisations of different light cones. The red curve represent the associated prediction in the case of a gaussian field with errors computed using equation 17. Here we keep the SN effect in the measures and include it in the prediction. *Bottom :* Relative difference in percent following the same color code.

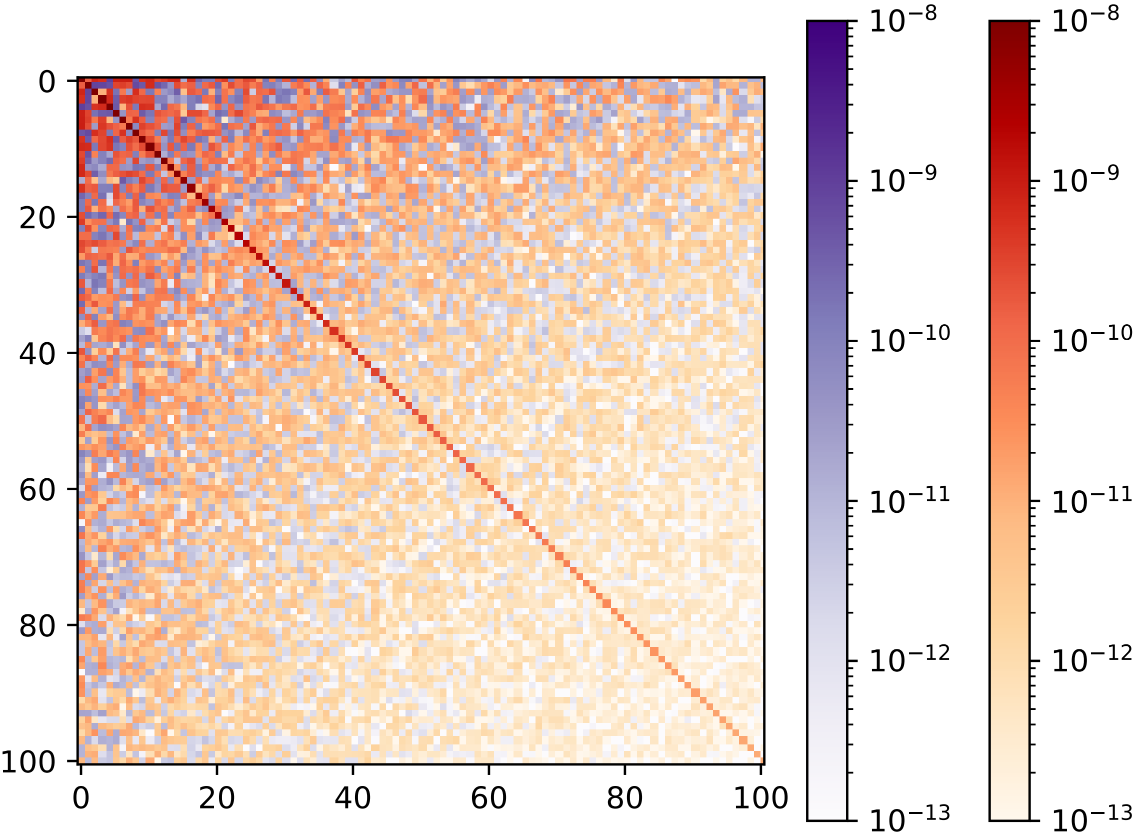


Fig. 11. Covariance matrix for 10000 realizations of C_ℓ's in a simulated universe between redshifts 0.2 and 0.3 and a sampling $N_s = 512$. Only $(\ell \times \ell') = (100 \times 100)$ first elements of the matrix are represented here. Color maps are here logarithmic scale: the red ones are the positive correlations while the blue ones are anticorrelations.

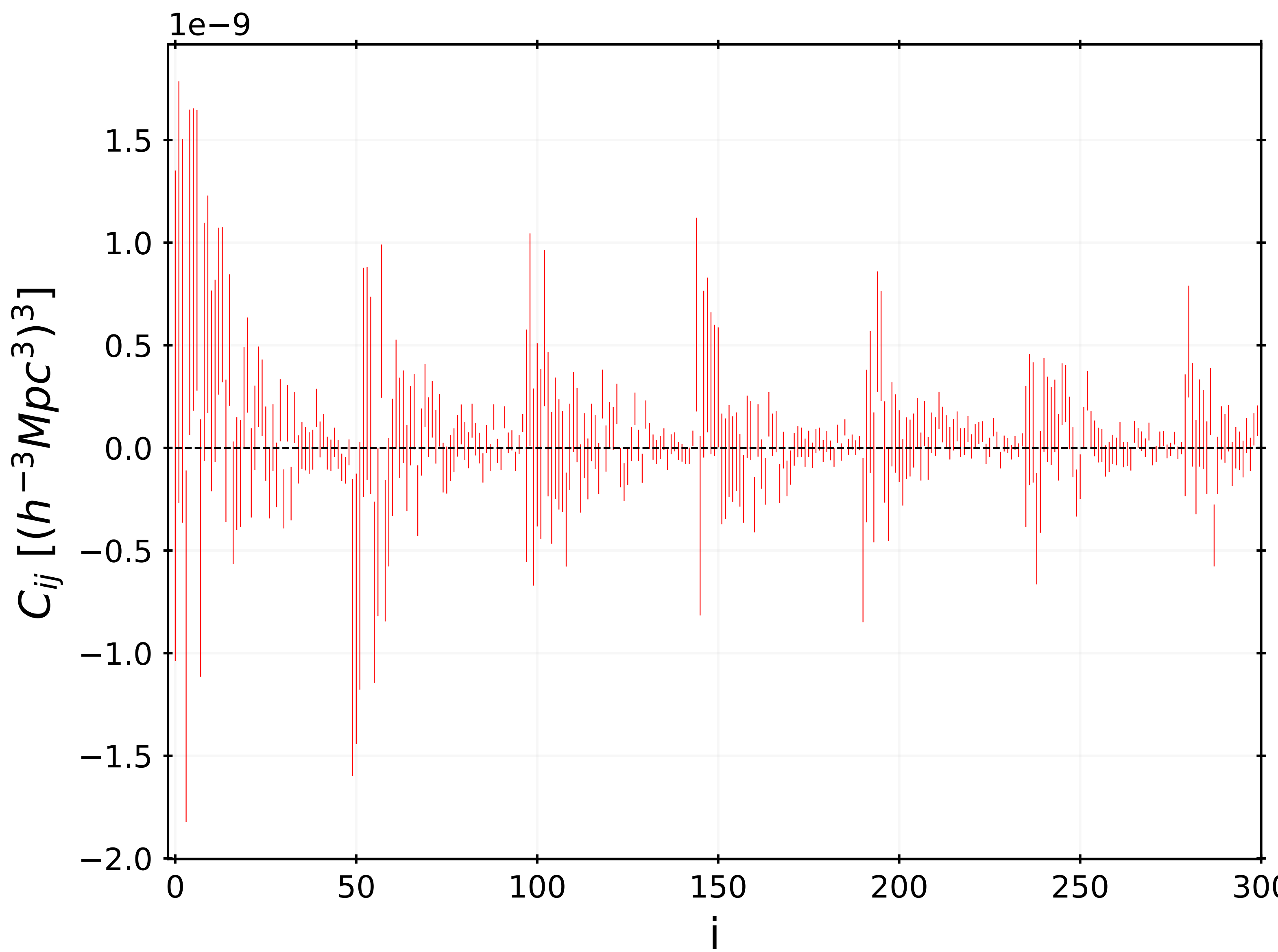


Fig. 13. The 300 first elements measured of the off-diagonal part of the covariance matrix over $n = 10000$ realisations of light cone (black dots) with gaussian errors (in red) computed using $V[C^G_{ij}] = (C^{G\,2}_{ii} + C^G_{ii}C^G_{jj})/(n-1)$. The elements are labeled by the index i and are ordered column by column of the lower half of the matrices without passing by the diagonal.

- Near Gaussian covariance matrix

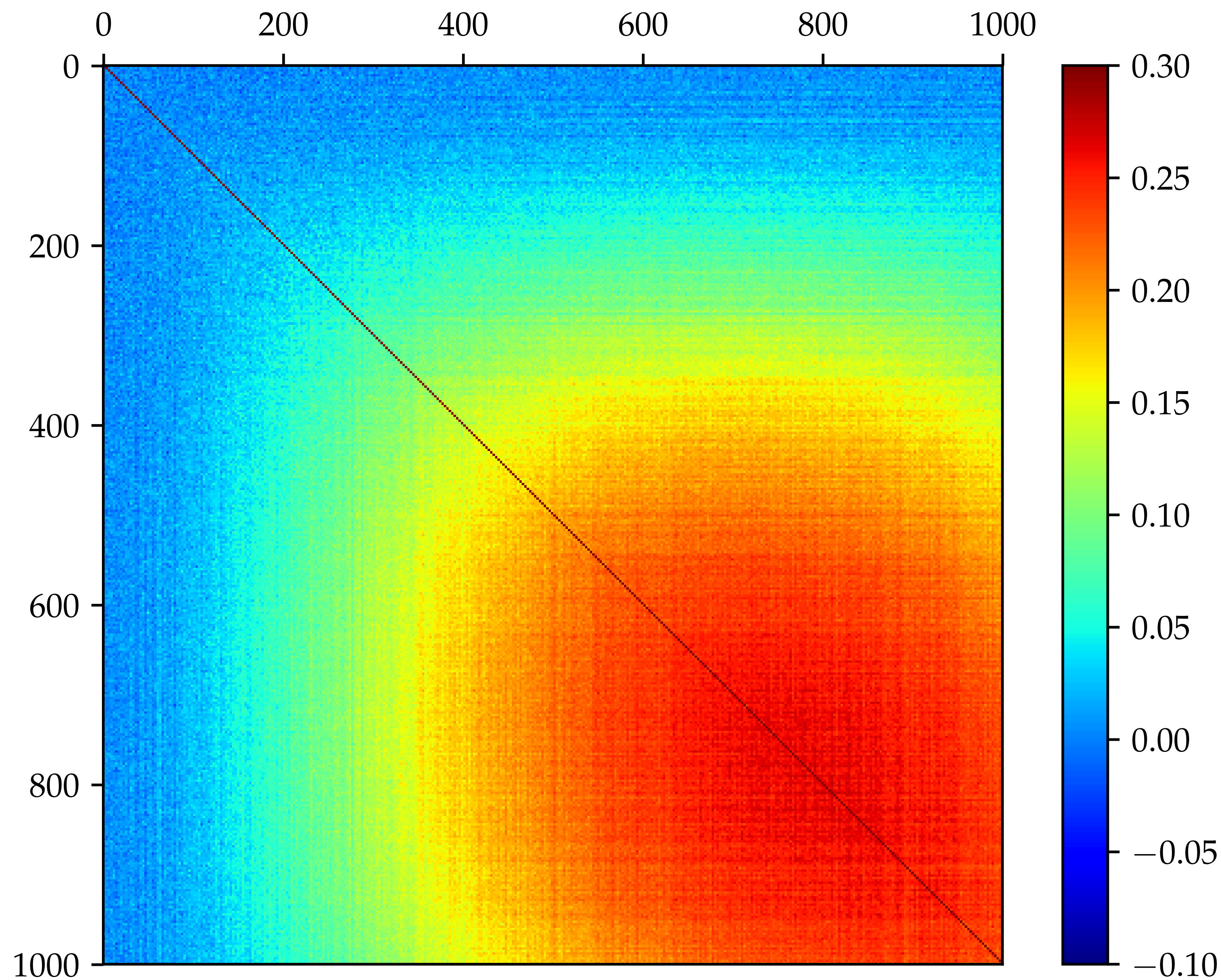


Fig. 15. Logarithm of the correlation matrix $\log(r_{ij} + 1)$ for 10000 realisations of C_ℓ's in a simulated universe between redshifts 0.2 and 0.3 and a sampling $N_s = 512$. The $(\ell \times \ell') = (1000 \times 1000)$ of the matrix are represented here.

▶ Conclusion

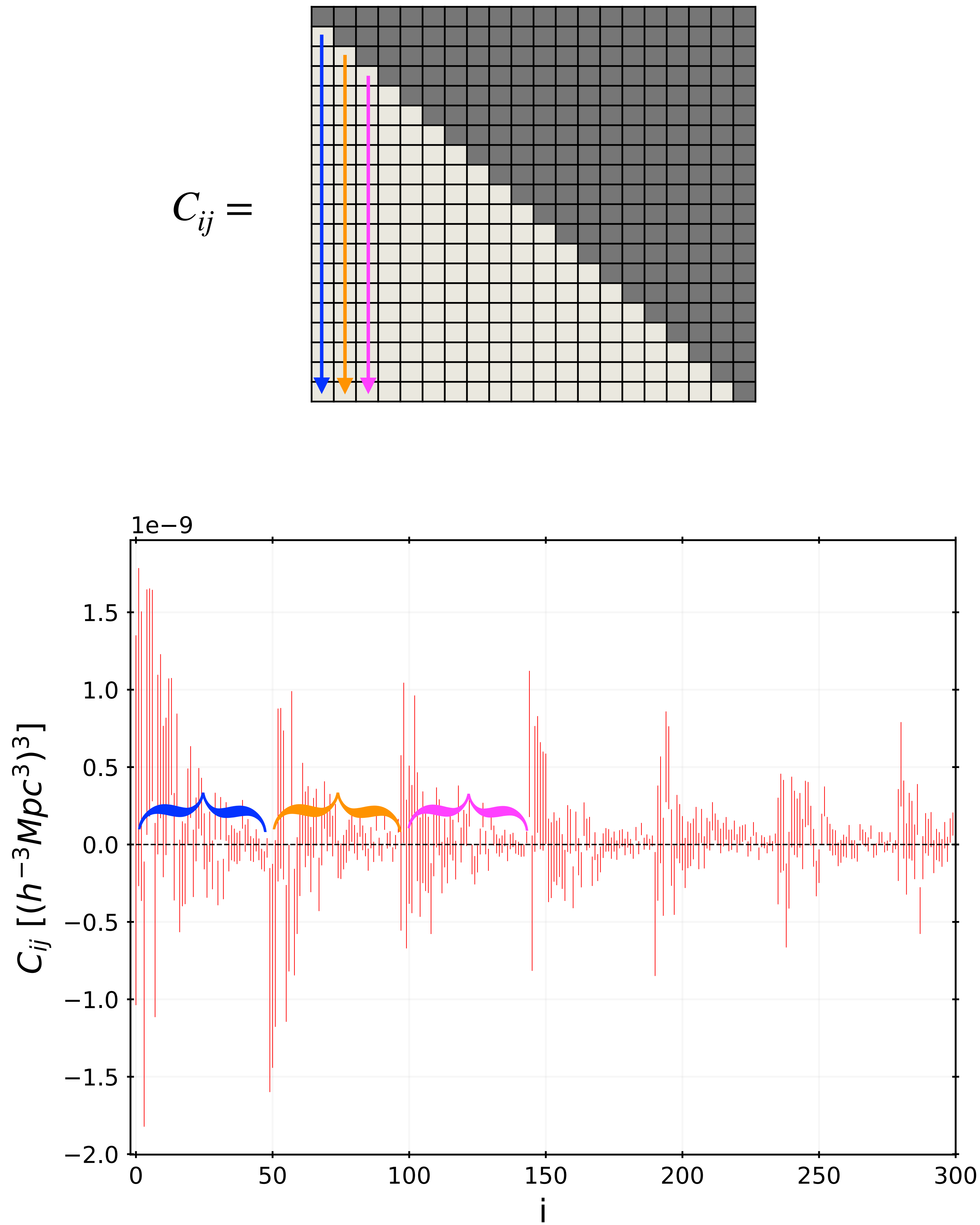
- General code to simulate any universe in a power spectrum oriented analysis
- Fast method for accurate P(k) and C_ℓ's
- Covariance matrix prediction
- Super Sample covariance embedded

—> Baratta, Bel, Plasczcynski, Ealet arXiv:1906.09042
AA/2019/36163

▶ Next developpements

- RSD in next analysis
- Comparison with Nbody codes (DEMNUni)
- Adaptation of FFT's in curved manifold
- Public code
- Surveys forcasts

Extra slides



$C_{ij} =$
1e−9
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0
50
100
150
200
250
300
i
C_{ij} $[(h^{-3}Mpc^3)^3]$

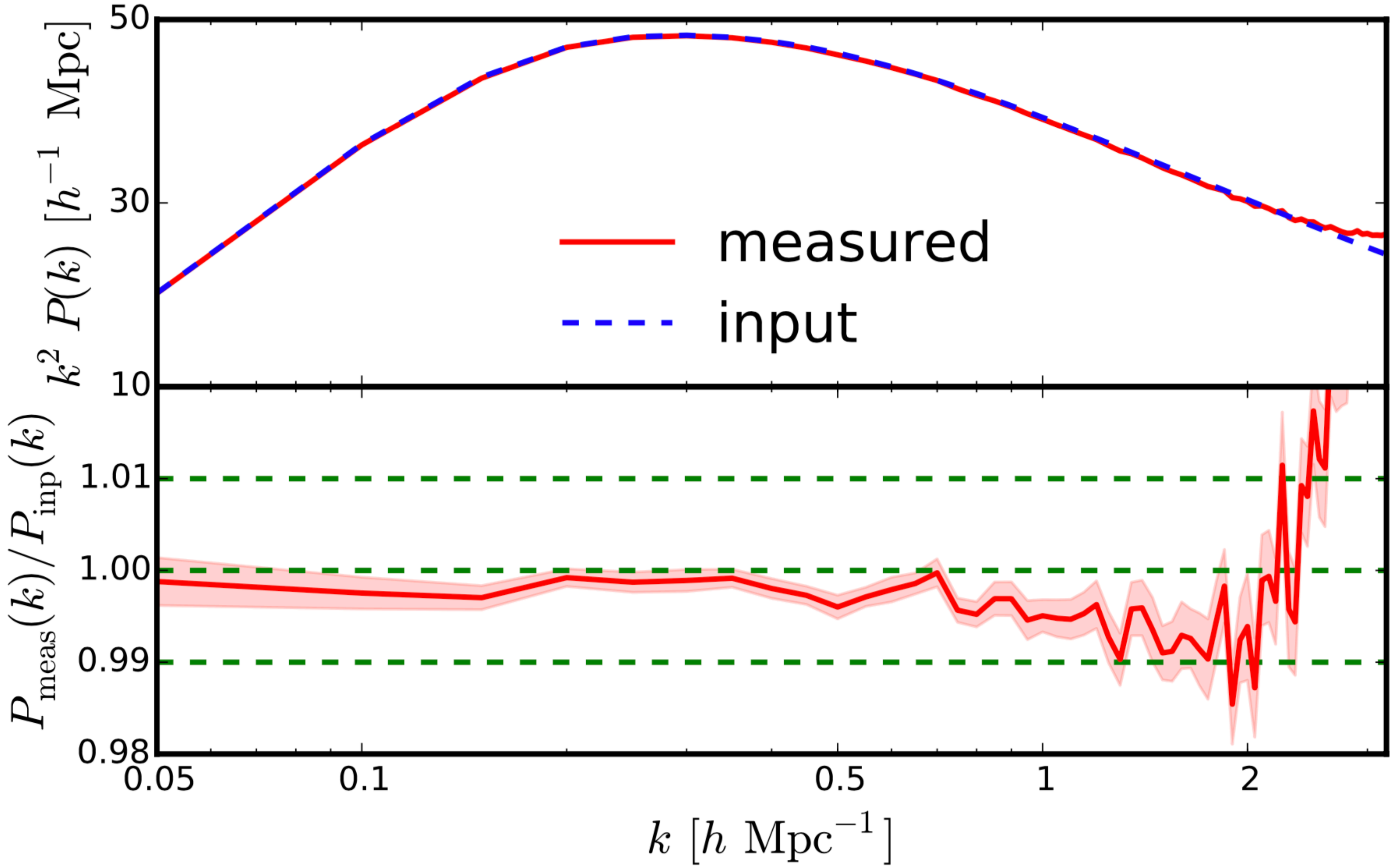


Figure 2: (*Top*) Mean of the real-space galaxy power spectrum measured from 50 log-normal catalogs (solid) and the input power spectrum (dashed). We show $k^2P(k)$. (*Bottom*) Ratio of the two. The band shows the error on the mean estimated from 50 realisations. The Nyquist frequency for these measurements is $k_{\mathrm{Ny}} = 3.22\, h\ \mathrm{Mpc}^{-1}$.

agrawal et al. (2017)