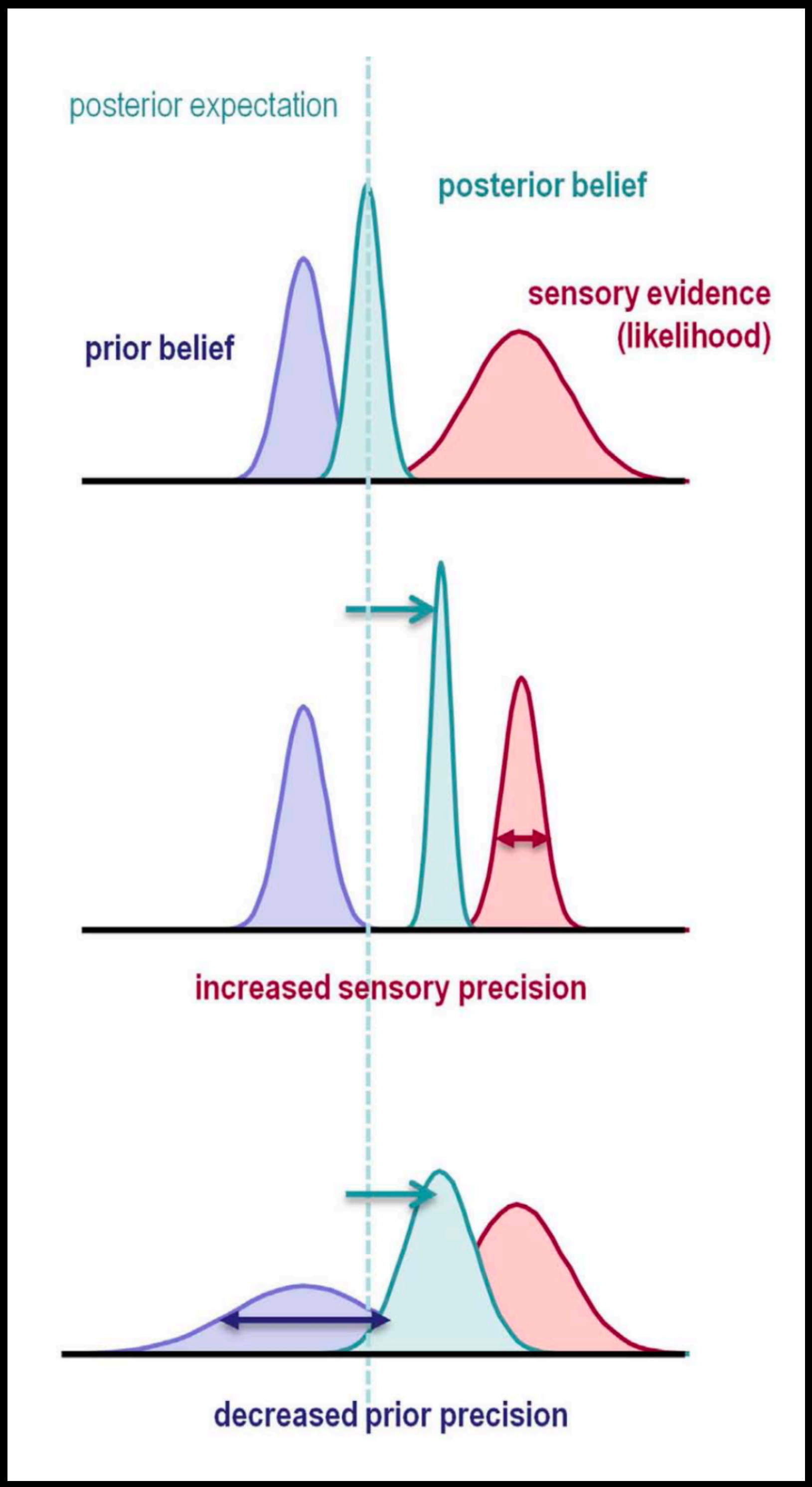
A Monte Carlo galaxy catalogue generator to predict covariance matrices ?

Philippe Baratta (CPPM/CPT)
3rdy PhD student at Aix-Marseille Univerisité

supervised by
Julien Bel (CPT)
Anne Ealet (IP2I)

Motivations ... the Bayesian inference



$$\underbrace{P(\theta|d,M)}_{\text{Posterior}} = \frac{\overbrace{P(d|\theta,M)}^{\text{Likelihood}}\overbrace{P(\theta|M)}^{\text{Prior}}}{\underbrace{P(d|M)}_{\text{Evidence}}}$$

$$\mathscr{L} = \exp(-\chi^2/2)$$

$$\chi^2 = X^T C^{-1} X$$

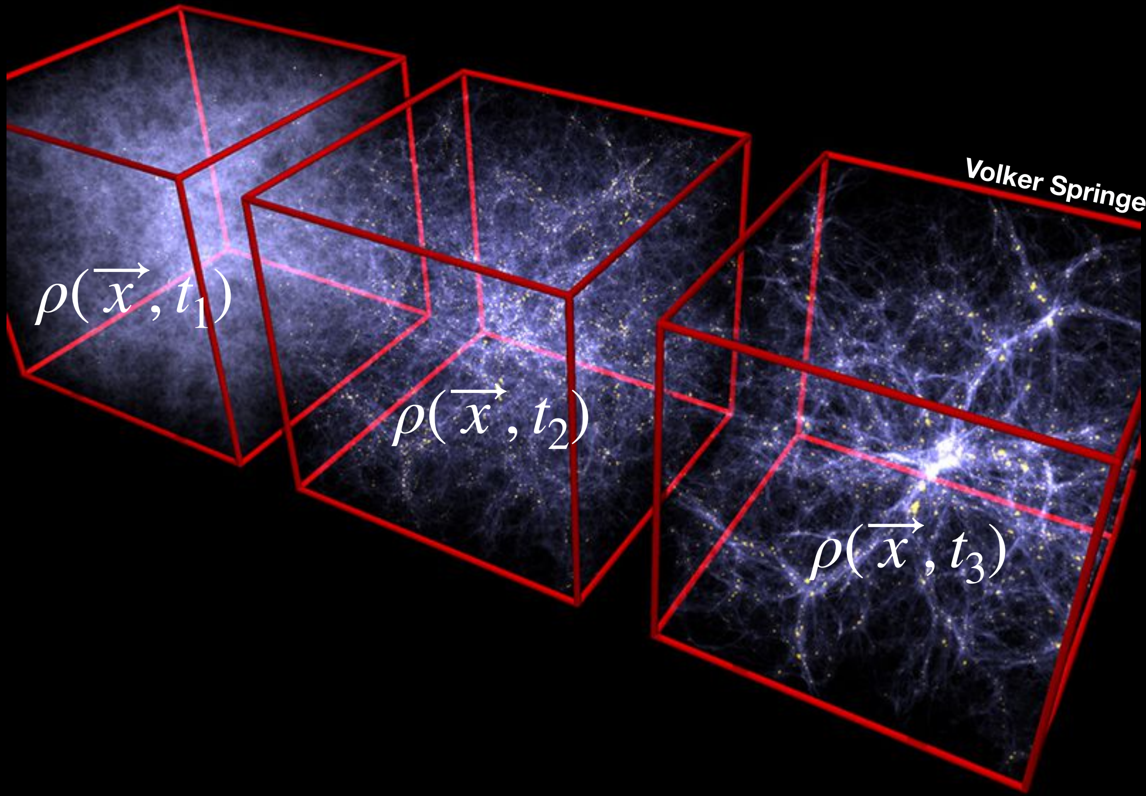
$$X^T = (x_1^{th} - x_1^{meas}, \ldots, x_n^{th} - x_n^{meas})$$

- **C the covariance matrix**

–> diagonal for uncorrelated data
–> non vanishing off-diag otherwise

The spead of the likelihood should not be underestimated

Motivations ... a bit of theory



Fluctuation field

$$\delta(\overrightarrow{x}, t) = \Delta\rho(\overrightarrow{x}, t)/\bar{\rho}$$

$$\delta_{\overrightarrow{k}} = \int \frac{d^3\overrightarrow{x}}{(2\pi)^3} e^{-i\overrightarrow{k}.\overrightarrow{x}} \delta(\overrightarrow{x})$$

Statistical ppties described by their connected moments

$$\left\langle \delta_{\overrightarrow{k}_1} \delta_{\overrightarrow{k}_2} \right\rangle_c = \delta_D(\overrightarrow{k}_{12}) P(\overrightarrow{k}_1)$$ Power spectrum

$$\left\langle \delta_{\overrightarrow{k}_1} \delta_{\overrightarrow{k}_2} \delta_{\overrightarrow{k}_3} \right\rangle_c = \delta_D(\overrightarrow{k}_{123}) B(\overrightarrow{k}_1, \overrightarrow{k}_2, \overrightarrow{k}_3)$$ Bispectrum

$$\left\langle \delta_{\overrightarrow{k}_1} \delta_{\overrightarrow{k}_2} \delta_{\overrightarrow{k}_3} \delta_{\overrightarrow{k}_4} \right\rangle_c = \delta_D(\overrightarrow{k}_{1234}) T(\overrightarrow{k}_1, \overrightarrow{k}_2, \overrightarrow{k}_3, \overrightarrow{k}_4)$$ Trispectrum

$$C_{ij} \equiv \left\langle \hat{P}(k_i)\hat{P}(k_j) \right\rangle - \left\langle \hat{P}(k_i) \right\rangle \left\langle \hat{P}(k_j) \right\rangle = \frac{P(k_i)^2}{M_{k_i}} \delta_{ij}^D + k_f^3 \bar{T}(k_i, k_j)$$

Gaussian contrib (cosmic variance)

non-Gaussianities (Fourier modes correlations)

Motivations ... a bit of theory

$$C_{ij} = \frac{P(k_i)^2}{M_{k_i}}\delta_{ij}^D + k_f^3 \bar{T}(k_i, k_j)$$

bin averaged Trispectrum

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

Using Monte Carlo (later introduced) on a totaly defined set :

- box of volume L x L x L
- analytical P.D.F. (log-normal)
- theoretical Power spectrum

$$\bar{T}(k_i, k_i) \sim 8c_1^2(4c_2^2 + 3c_1c_3)P^3(k_i)$$

$$+24(3c_1^2c_3^2 + 4c_1c_2^2c_3 + 12c_1^2c_2c_4)P^{(2)}(k_i)P^2(k_i)$$

$$+144c_1^2c_3^2P^{(2)}(0)P^2(k_i)$$

c_n some specifics
Hermite polynomials coefficients

Motivations ... a bit of theory

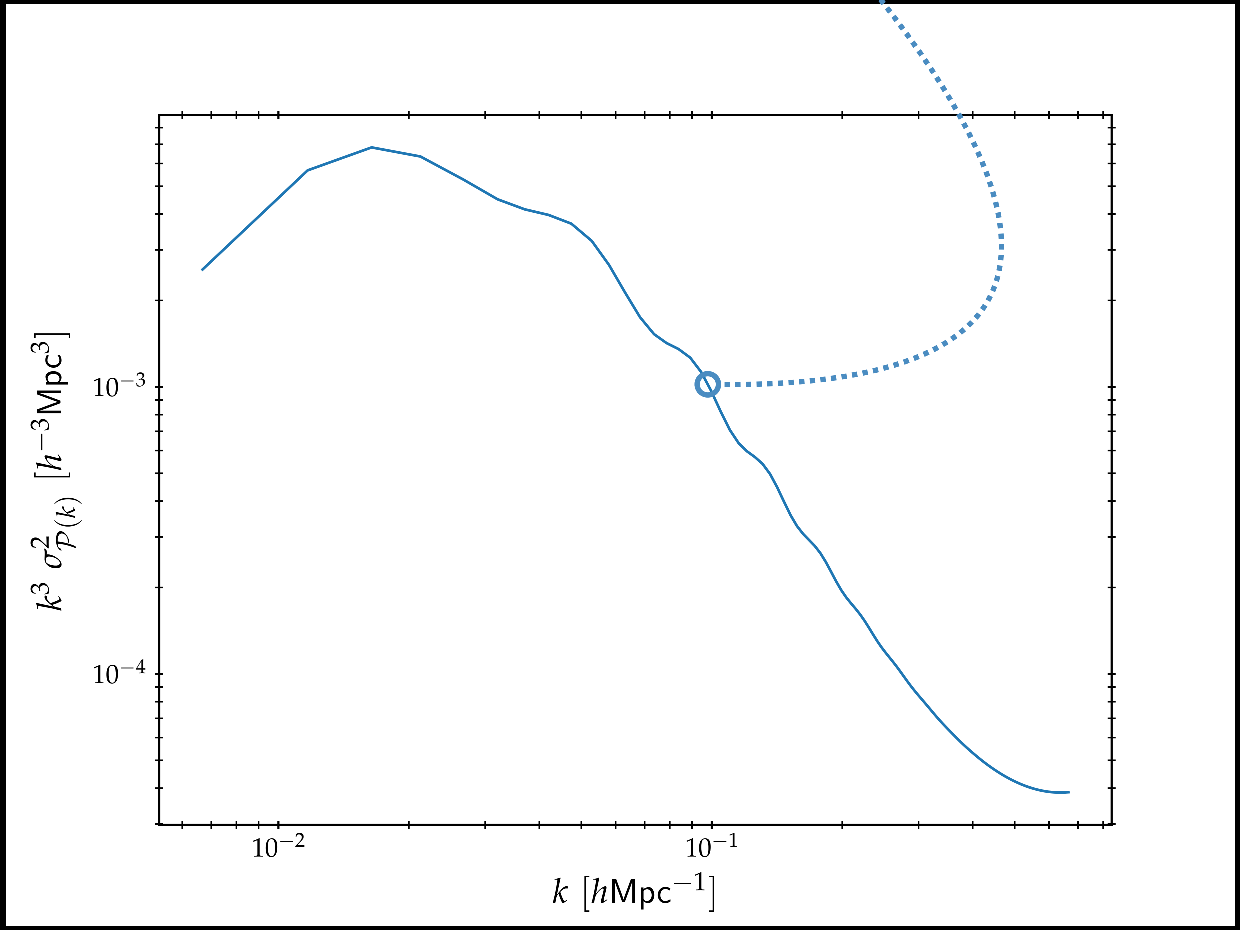
$$C_{ij} = \boxed{\frac{P(k_i)^2}{M_{k_i}}\delta^D_{ij}} + k_f^3 \bar{T}(k_i, k_j)$$

bin averaged Trispectrum

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

Using Monte Carlo (later introduced) on a totaly defined set :

- box of volume L x L x L
- analytical P.D.F. (log-normal)
- theoretical Power spectrum



$$\bar{T}(k_i, k_i) \sim 8c_1^2(4c_2^2 + 3c_1c_3)P^3(k_i)$$

$$+24(3c_1^2c_3^2 + 4c_1c_2^2c_3 + 12c_1^2c_2c_4)P^{(2)}(k_i)P^2(k_i)$$

$$+144c_1^2c_3^2P^{(2)}(0)P^2(k_i)$$

c_n some specifics Hermite polynomials coefficients

Motivations ... a bit of theory

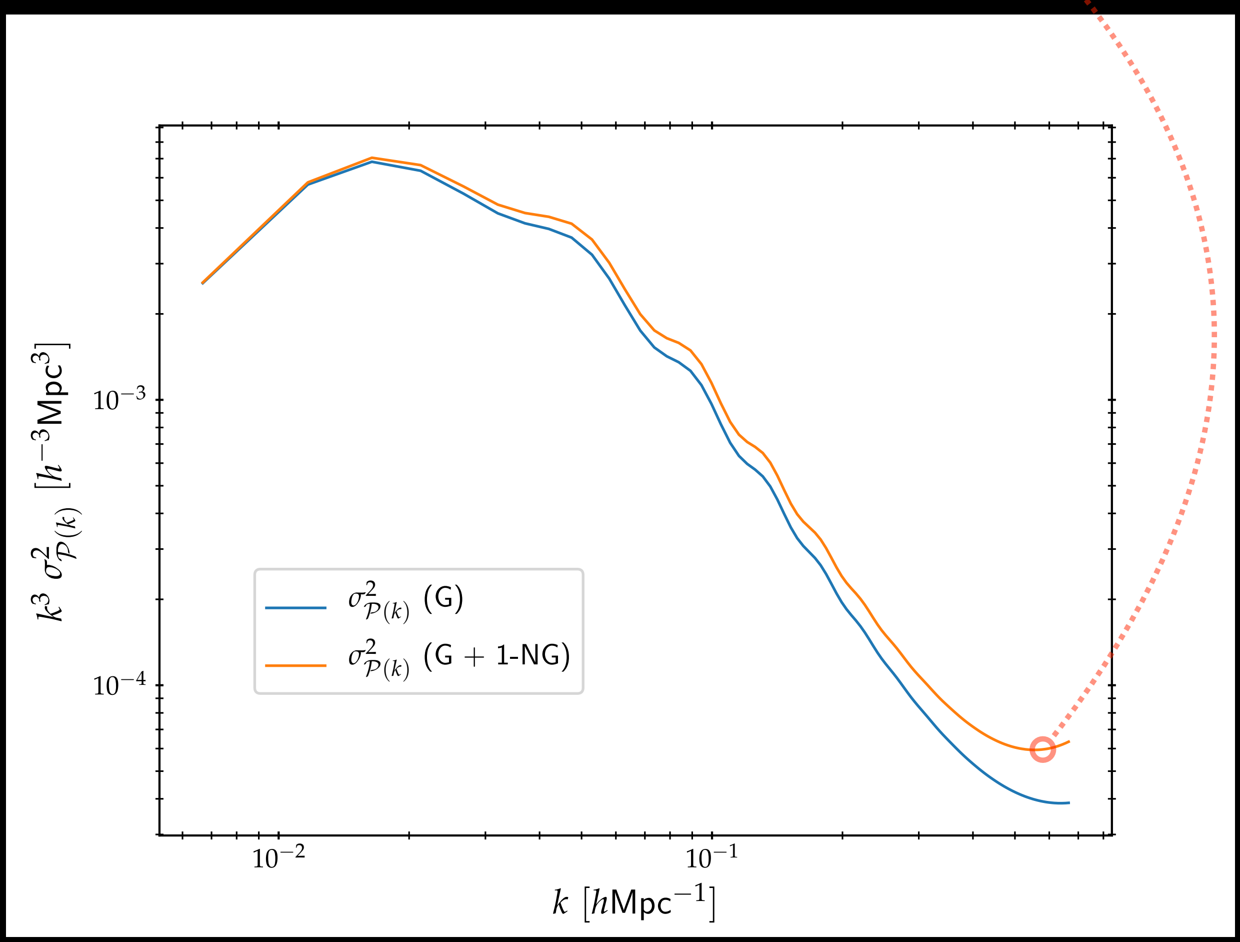
$$C_{ij} = \boxed{\frac{P(k_i)^2}{M_{k_i}}\delta^D_{ij}} + k_f^3\bar{T}(k_i, k_j)$$

bin averaged Trispectrum

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

Using Monte Carlo (later introduced) on a totaly defined set :

- box of volume L x L x L
- analytical P.D.F. (log-normal)
- theoretical Power spectrum



$$\bar{T}(k_i, k_i) \sim \boxed{8c_1^2(4c_2^2 + 3c_1c_3)P^3(k_i)}$$

$$+24(3c_1^2c_3^2 + 4c_1c_2^2c_3 + 12c_1^2c_2c_4)P^{(2)}(k_i)P^2(k_i)$$

$$+144c_1^2c_3^2P^{(2)}(0)P^2(k_i)$$

c_n some specifics
Hermite polynomials coefficients

Motivations ... a bit of theory

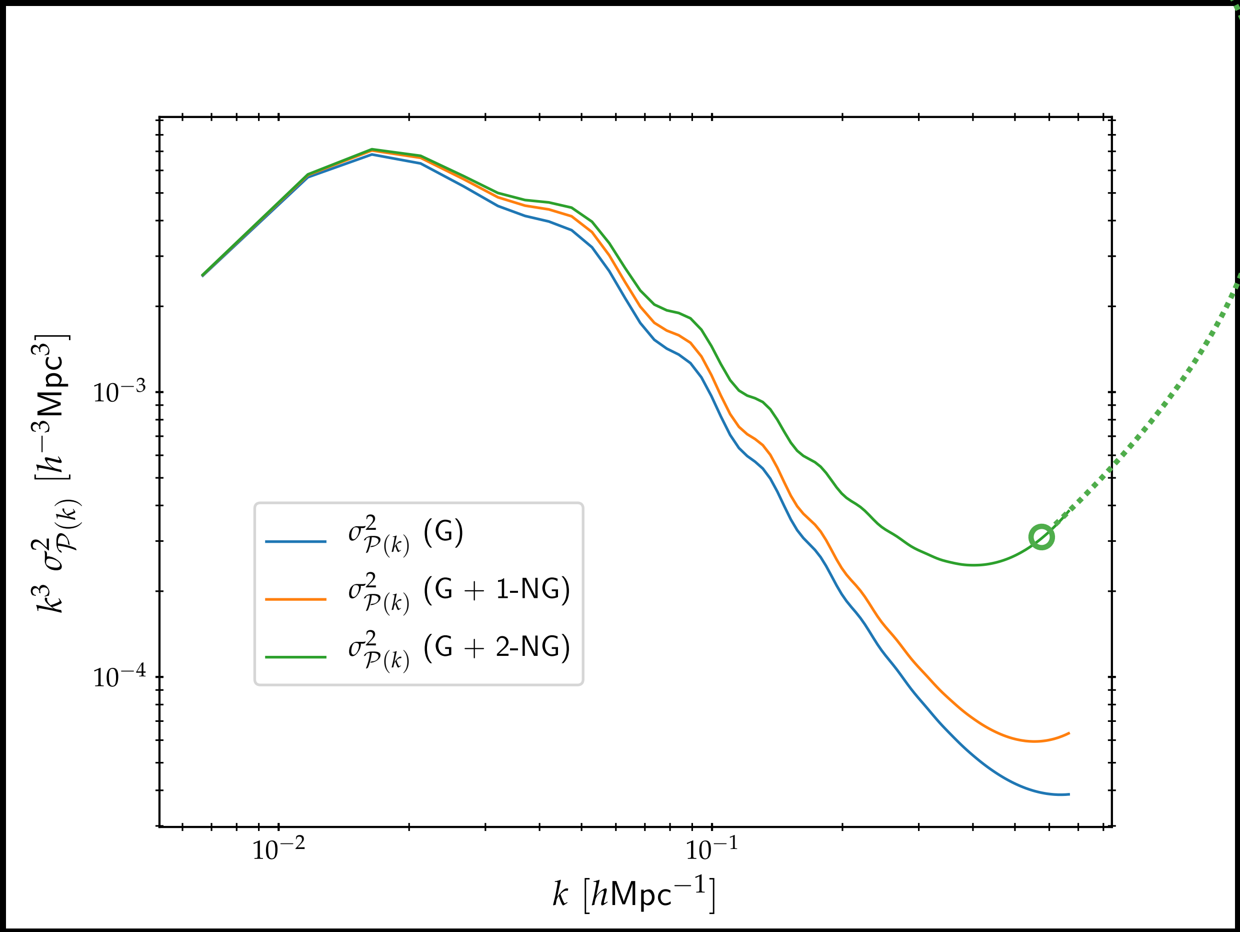
$$C_{ij} = \boxed{\frac{P(k_i)^2}{M_{k_i}}\delta^D_{ij}} + k_f^3 \bar{T}(k_i, k_j)$$

bin averaged Trispectrum

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

Using Monte Carlo (later introduced) on a totaly defined set :

- box of volume L x L x L
- analytical P.D.F. (log-normal)
- theoretical Power spectrum



$$\boxed{\begin{aligned}\bar{T}(k_i, k_i) \sim\ & 8c_1^2(4c_2^2 + 3c_1c_3)P^3(k_i) \\ & + 24(3c_1^2c_3^2 + 4c_1c_2^2c_3 + 12c_1^2c_2c_4)P^{(2)}(k_i)P^2(k_i)\end{aligned}}$$

$$+144c_1^2c_3^2P^{(2)}(0)P^2(k_i)$$

c_n some specifics
Hermite polynomials coefficients

Motivations ... a bit of theory

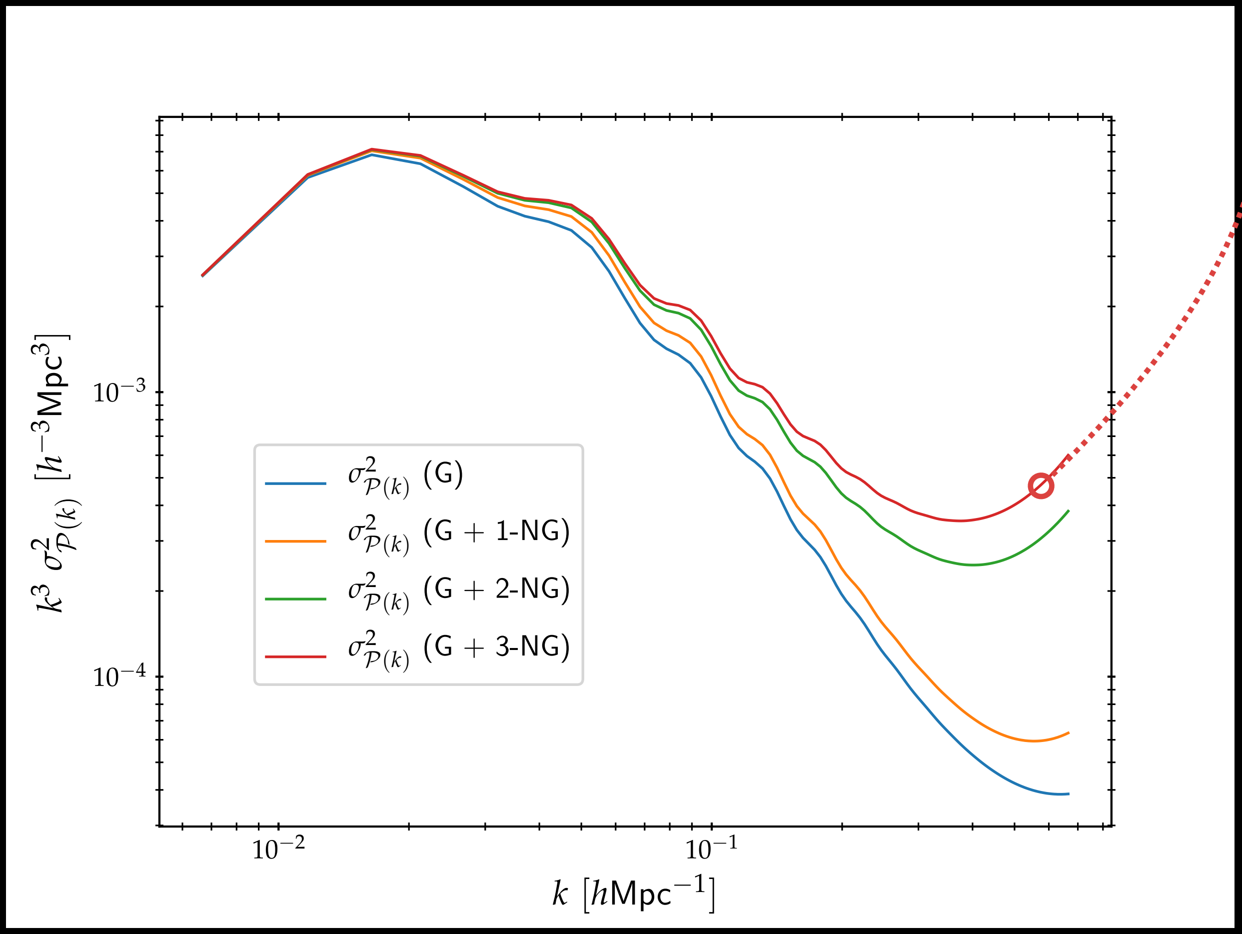
$$C_{ij} = \boxed{\frac{P(k_i)^2}{M_{k_i}}\delta_{ij}^D} + k_f^3 \bar{T}(k_i, k_j)$$

bin averaged Trispectrum

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

Using Monte Carlo (later introduced) on a totaly defined set :

- box of volume L x L x L
- analytical P.D.F. (log-normal)
- theoretical Power spectrum



$$\boxed{\begin{aligned}\bar{T}(k_i, k_i) \sim\ & 8c_1^2(4c_2^2 + 3c_1c_3)P^3(k_i) \\ & + 24(3c_1^2c_3^2 + 4c_1c_2^2c_3 + 12c_1^2c_2c_4)P^{(2)}(k_i)P^2(k_i) \\ & + 144c_1^2c_3^2P^{(2)}(0)P^2(k_i)\end{aligned}}$$

c_n some specifics
Hermite polynomials coefficients

- 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 of a « key » observable

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

Angular power spectrum C_ℓ

$$C_\ell(r, r') = (4\pi)^2 \int_0^\infty dk k^2 P(k) j_\ell(kr) j_\ell(kr')$$

analytically?

How
to predict C_{ij} ?

- 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 of a « key » observable

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

Angular power spectrum C_ℓ

$$C_\ell(r,r') = (4\pi)^2 \int_0^\infty dk k^2 P(k) j_\ell(kr) j_\ell(kr')$$

~~analytically?~~

How to predict C_{ij} ?

numerically? → Nbody sim?

- 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 of a « key » observable

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

Angular power spectrum C_ℓ

$$C_\ell(r, r') = (4\pi)^2 \int_0^\infty dk k^2 P(k) j_\ell(kr) j_\ell(kr')$$

How to predict C_{ij} ?

~~analytically?~~

~~numerically?~~ → ~~Nbody sim?~~

semi-analytically → $\boxed{\text{Monte Carlo}}$

Fast Monte Carlo

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

Input

Theoretical Power spectrum P(k)

Probability distribution function (PDF)

→

Output

Cl's

Generate a galaxy catalogue

- **Step 1 : generate a <u>non gaussian</u> field in a box with a given PDF and P(k,z)**
- **Step 2 : Poisson sample it to get a snapshot**
- **Step 3 : reconstruct the lightcone**

Fast Monte Carlo

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

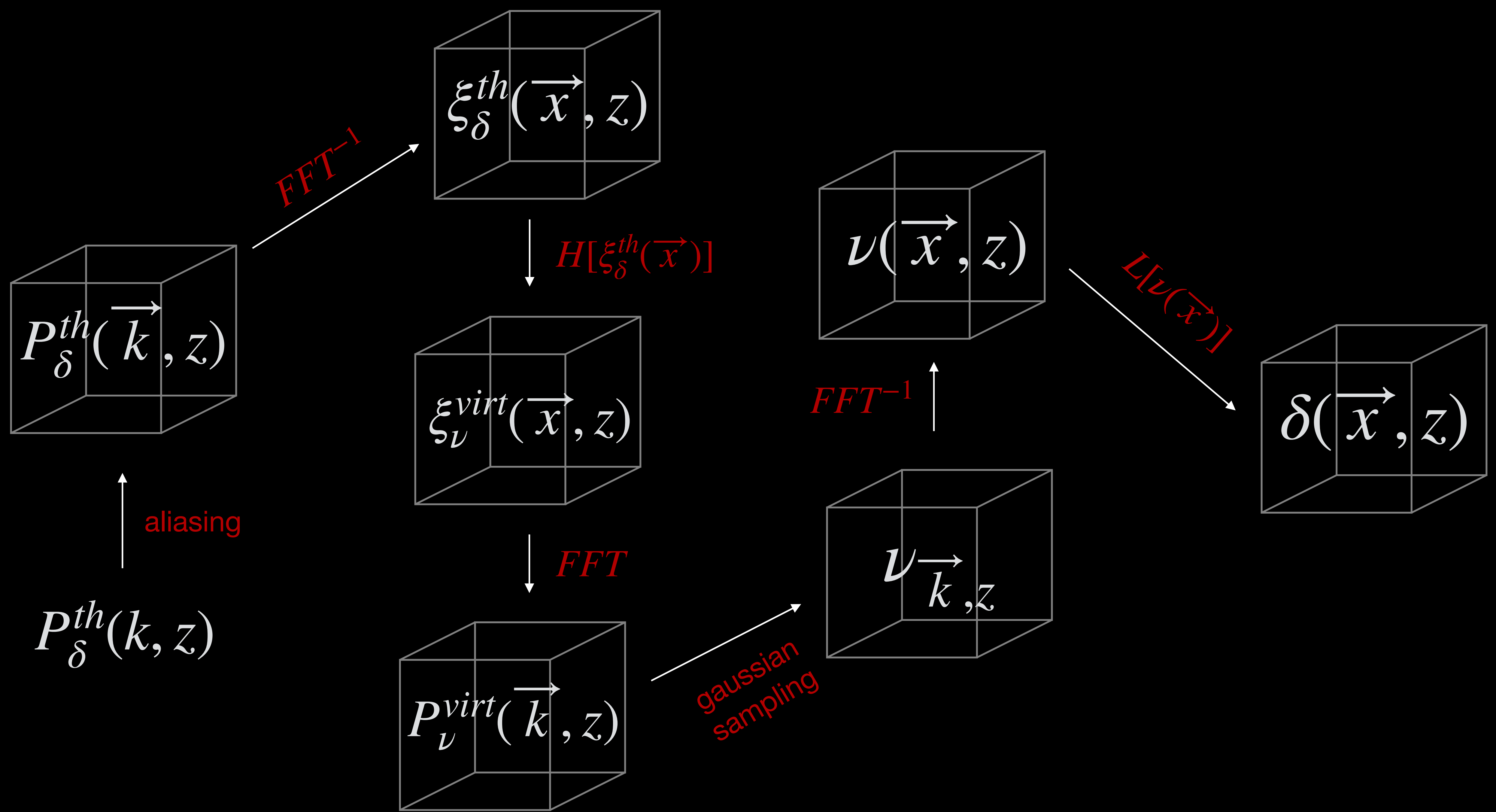
Input		Output
Theoretical Power spectrum P(k) **Probability distribution function (PDF)**	$\longrightarrow$	**Cl’s**

Generate a galaxy catalogue

- **Step 1 : generate a <u>non gaussian</u> field in a box with a given PDF and P(k,z)** } **So far not in litterature**
- **Step 2 : Poisson sample it to get a snapshot**
- **Step 3 : reconstruct the lightcone**

Generate a NG field with arbitrary PDF and P(k)

- Chose a theoretical P(k,z)
- chose a target PDF and find the transformation L (ref) to go from a G field nu to the NG field delta
- Given L, how 2pcf trandforms? —> H (ref)



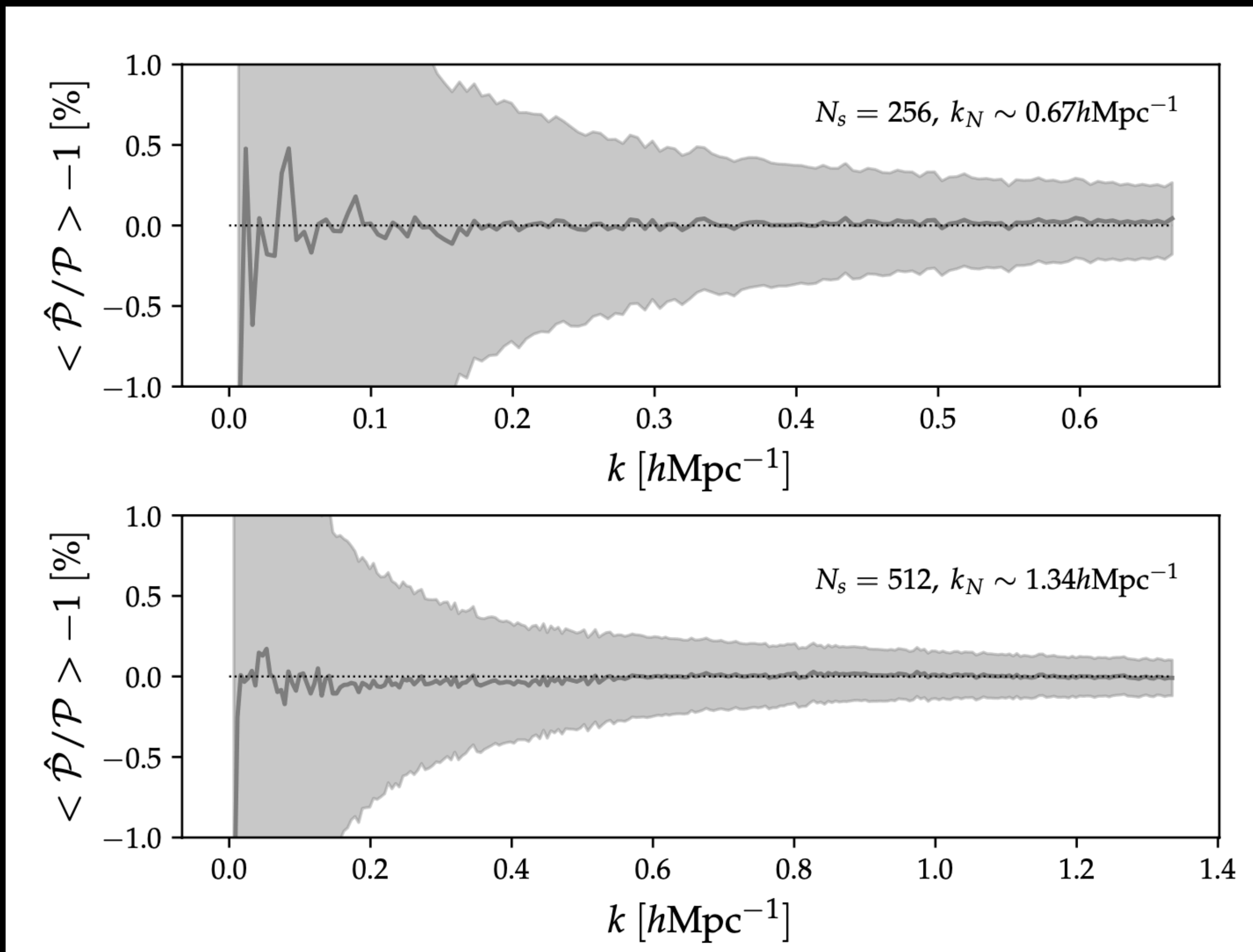
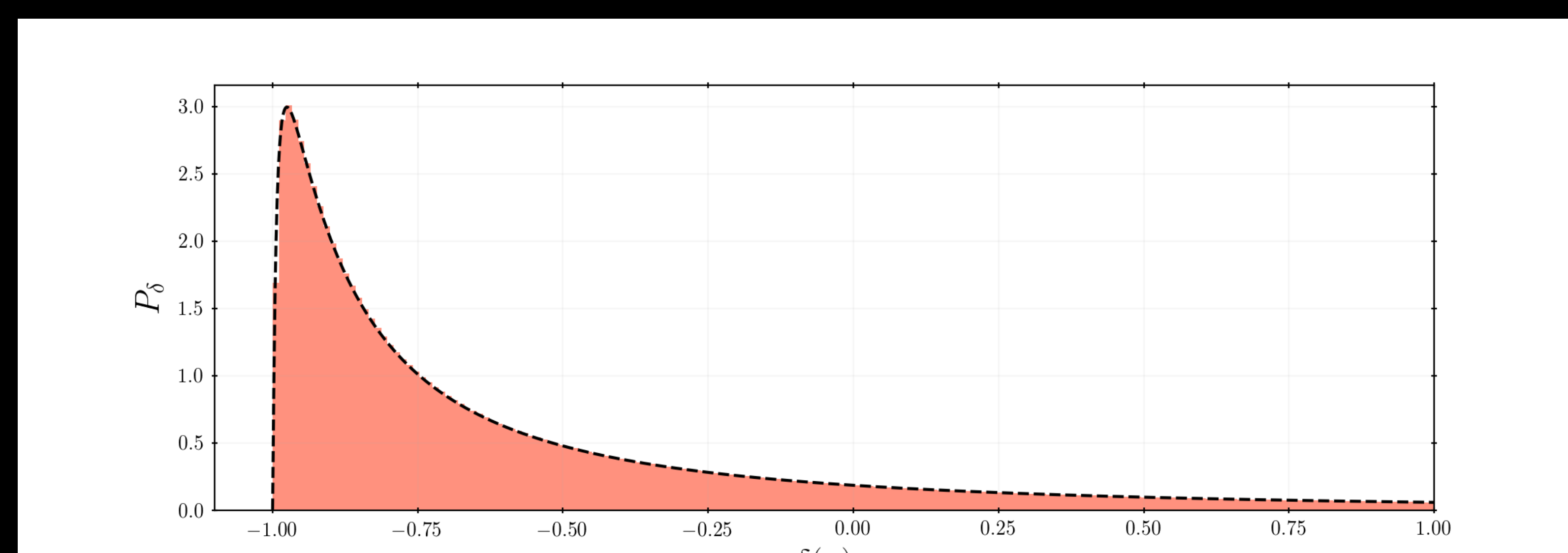


Fig. 6: Averaged 3D power spectrum compared to the expected 3D power spectrum, for 1000 realisations of the density field. The shell-averaged monopoles of this residuals in shells of width $|\mathbf{k}| - k_f/2 < |\mathbf{k}| < |\mathbf{k}| + k_f/2$ were then computed. The result is presented as percentage with error bars. The setting used 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 $z = 0$. Both results are computed up to the Nyquist frequency.

measured 1p P.D.F. in the case of a target log-normal distribution (analytical)



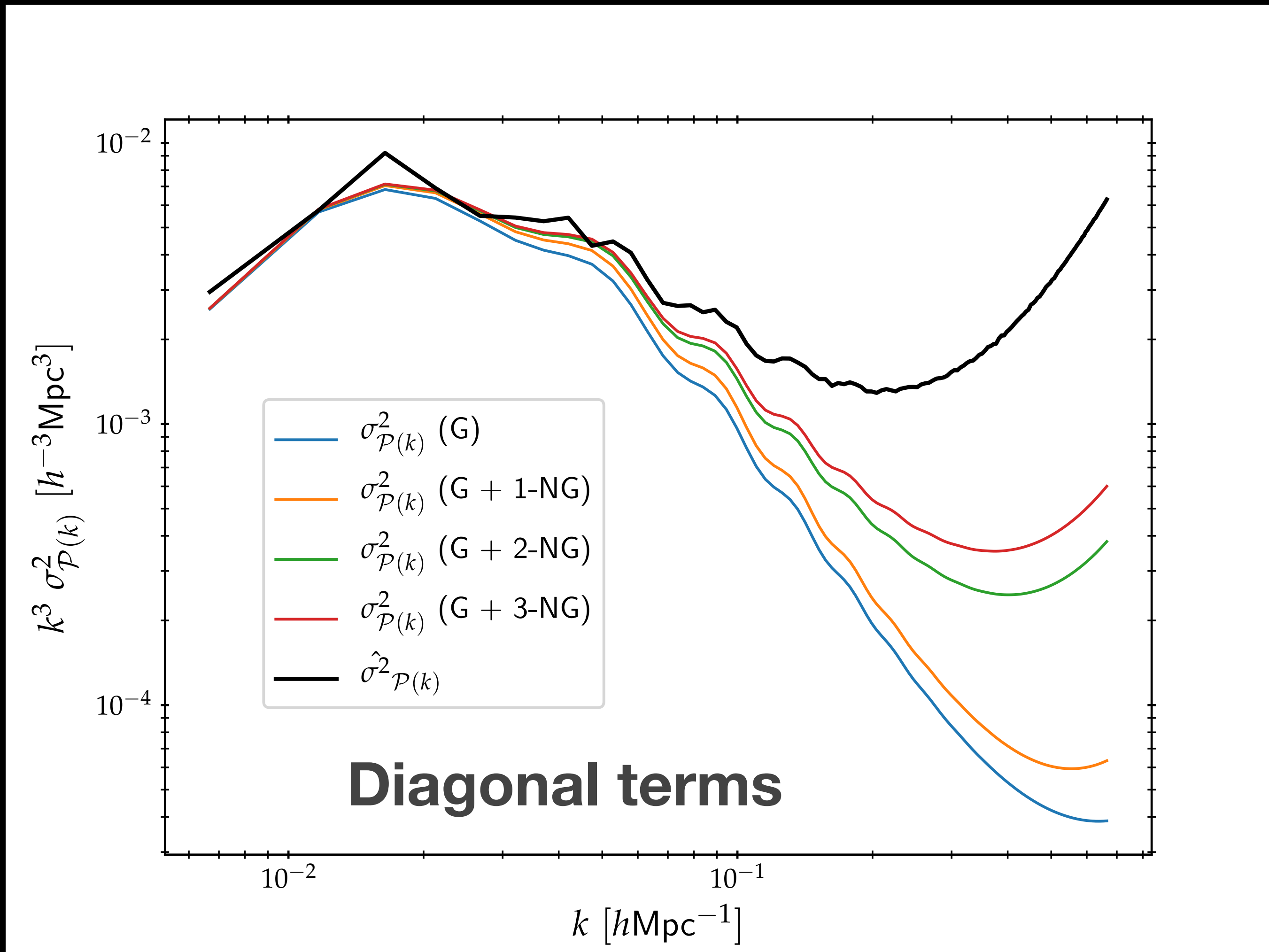


Fig. 5. Measured diagonal of the covariance matrix for 7375 power spectra realizations of the density field using the described method (black curve). The other curves represent their predictions taking into account the gaussian part alone (G) or by adding some non gaussian contributions of equation (18). 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.

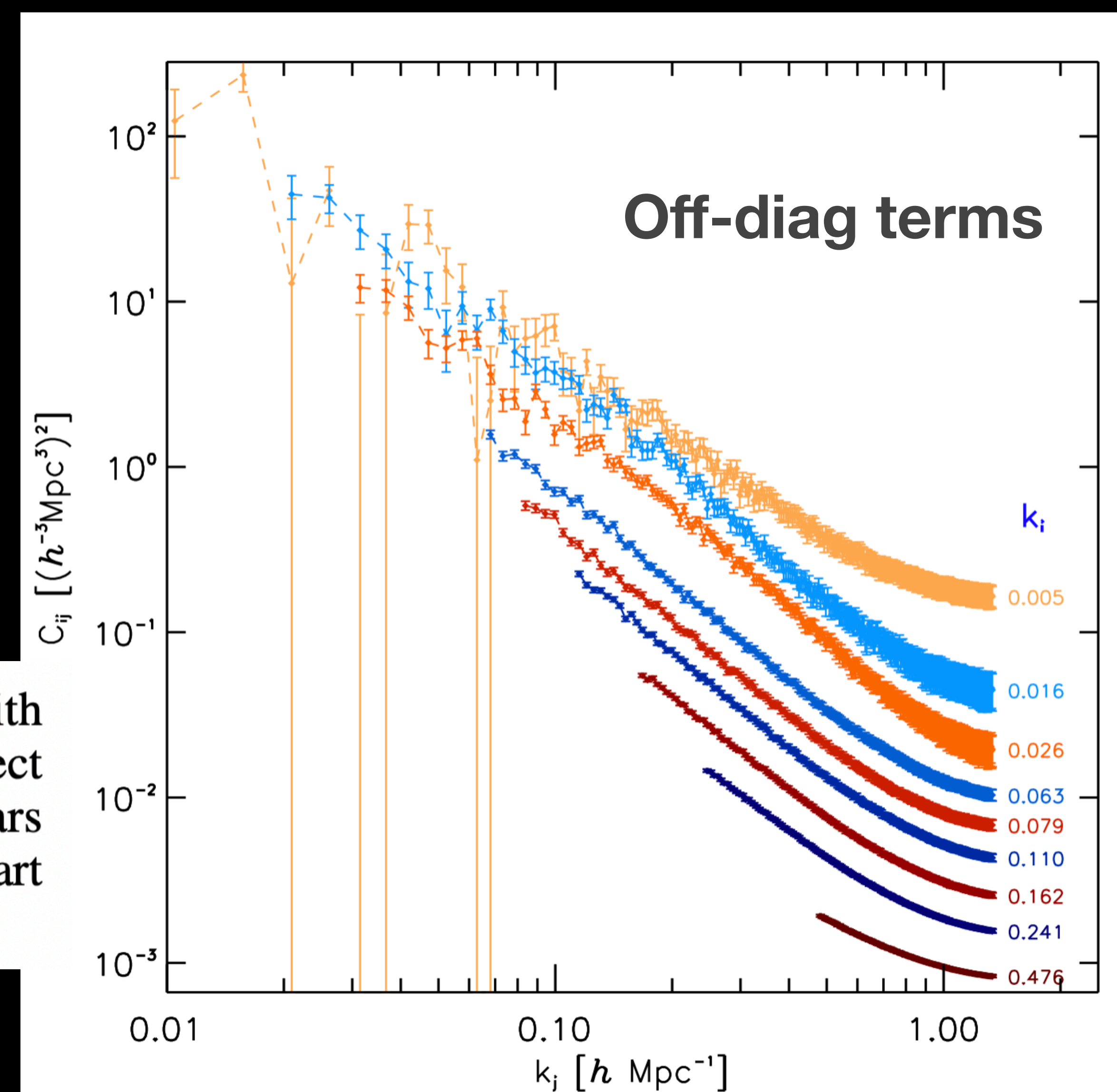
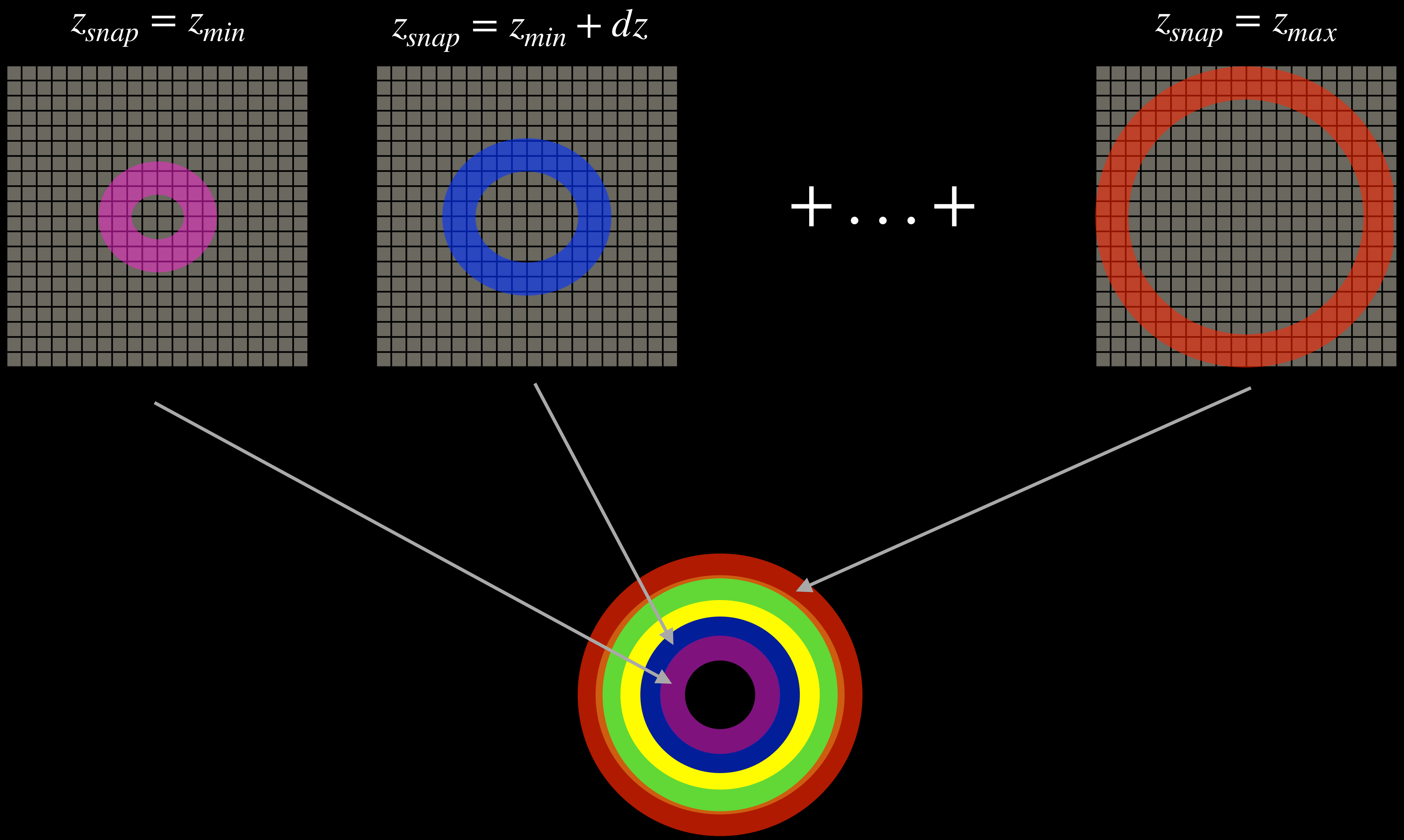


Fig. 6. Off diagonal elements of the covariance matrix estimated with $N = 7375$ realisations, showing the dependance of the C_{ij} with respect to k_j at various fixed k_i labeled on the right of the panel. The error bars are obtained assuming that the covariance coefficients follow a Wishart distribution,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

Reconstructing the light cone

- choose z_{min},z_{max} for your catalogue and generate N_{shl} NG fields
- Poisson sample them to get snapshots at these 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



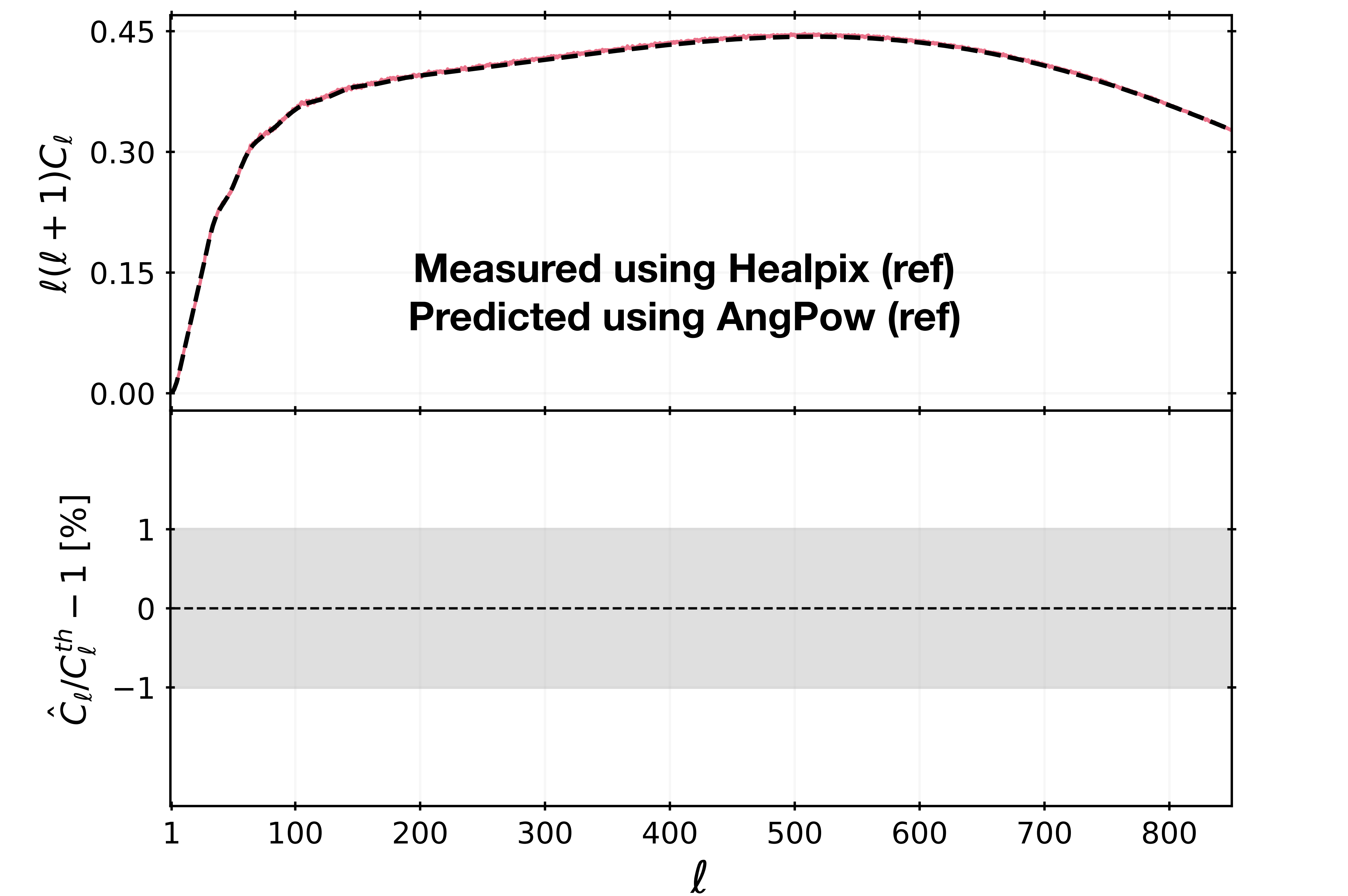


Fig. 8. Top panel : Thousand averaged C_ℓ's for simulated light cones using the shell-method with error bars (red curve) and corresponding prediction (dashed black curve). We simulate here a lightcone between redshifts 0.2 and 0.3 in a sampling $N_s = 512$ and a number of shells $N_{shl} = 250$ to ensure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

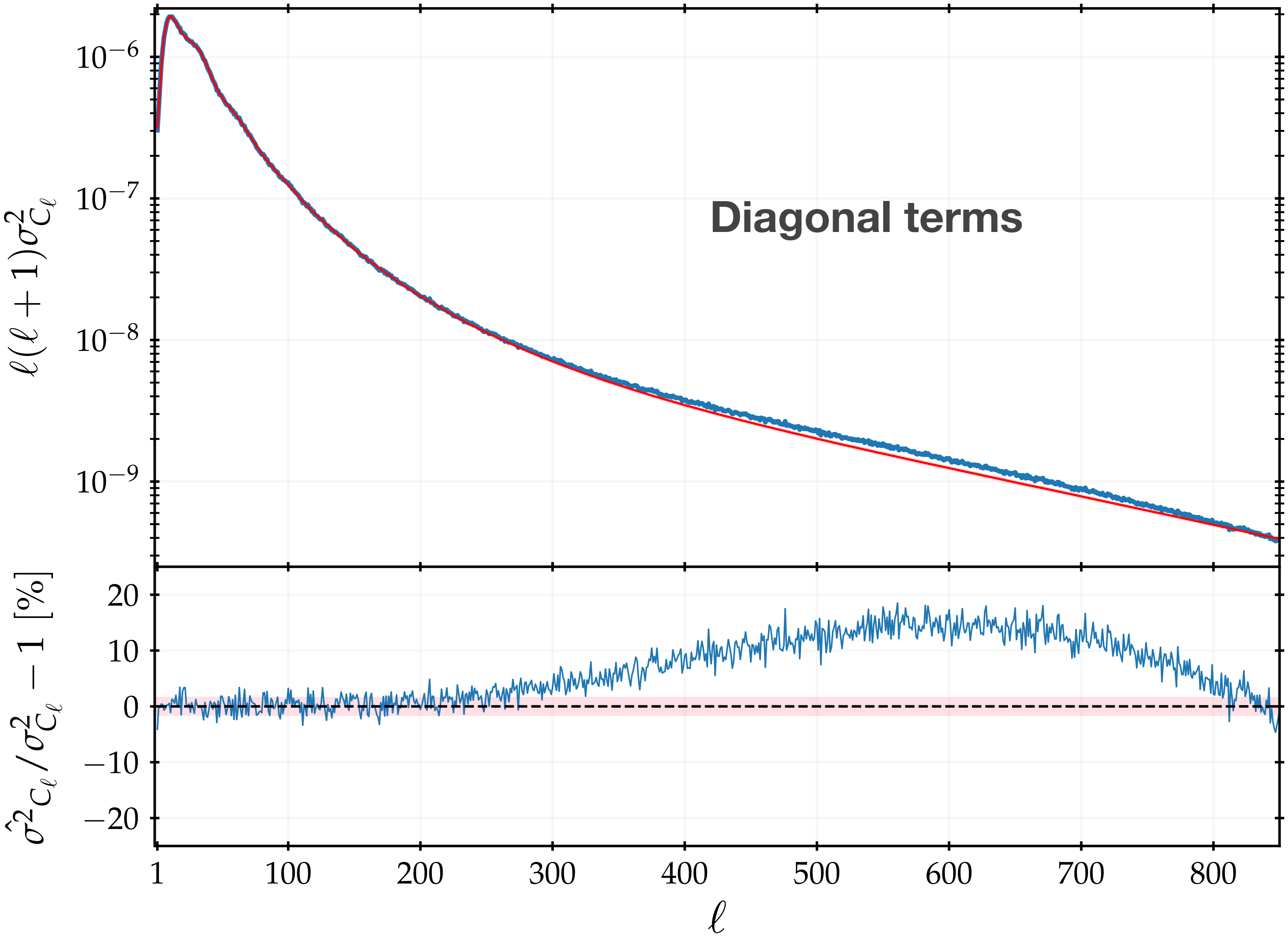


Fig. 13. *Top :* Measured diagonal of the covariance matrix (blue curve) over $N = 10000$ realisations of different light cones. The red curve represent the associated prediction in the case of a gaussian field with errors computed using equation 17. Here we keep the SN effect in the measures and include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

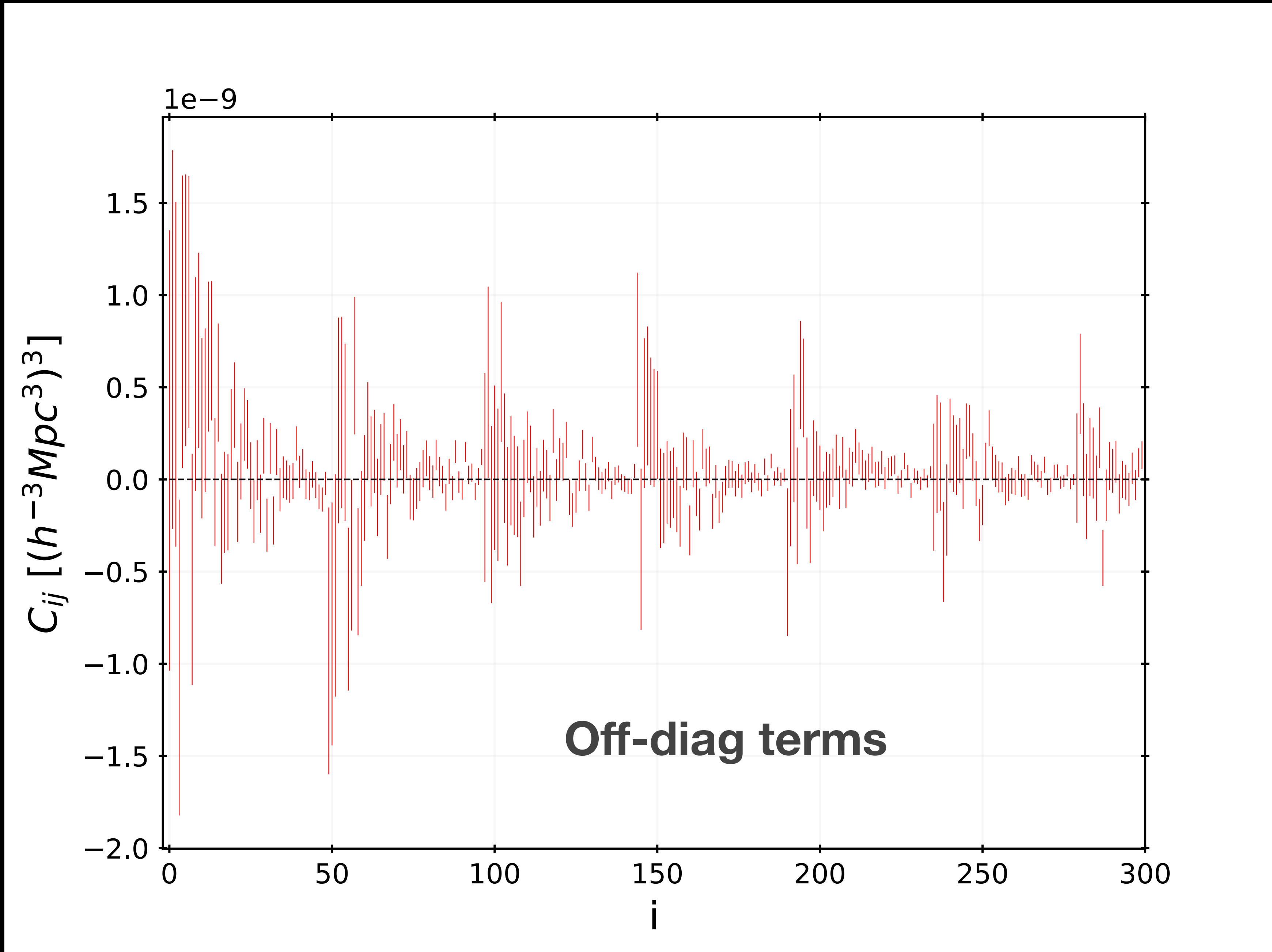


Fig. 13. The 300 first elements measured of the off-diagonal part of the covariance matrix over $n = 10000$ realisations of light cone (black dots) with gaussian errors (in red) computed using $V[C^G_{ij}] = (C^{G\,2}_{ii} + C^G_{ii}C^G_{jj})/(n-1)$. The elements are labeled by the index i and are ordered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

▶ 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

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

▶ Next developpements

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