# Statistics of the epoch of reionization(EoR) 21-cm signal:

# power spectrum error-covariance

**Rajesh Mondal**

**Department of Physics (and CTS), Indian Institute of Technology Kharagpur**

**Workshop on Cosmology with Next Generation Radio Surveys, ICTP, Trieste, Italy**
**22 June 2016**

# Introduction

- It is being anticipated that the EoR 21-cm power spectrum is the main tool to achieve the first detection of the signal.
- Cosmic variance: Any statistical estimation of a cosmological signal comes with intrinsic uncertainty
- Question: How accurately can the power spectrum be estimated from a given EoR 21-cm data set?
- We want to have prior knowledge of the expected error before the detection has been made.

# Motivations

- There have been several works to quantify the sensitivity to the EoR signal for different instruments.
- It is commonly assumed in all sensitivity estimates studies that the EoR 21-cm signal is independent Gaussian random variable

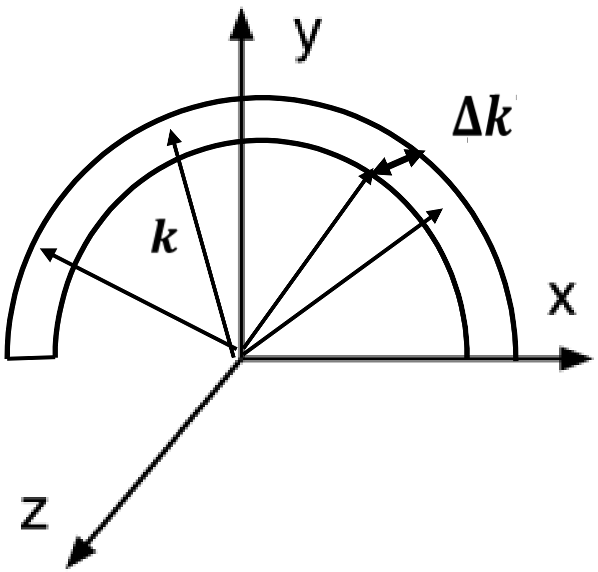
- How good is this assumption?
- How this affects the error predictions for the EoR 21-cm power spectrum.

- Generic: not limited only to the EoR 21-cm signal but can be applied to any non-Gaussian signal (*e.g.* galaxy redshift surveys - Feldman & Peacock, 1994; Neyrinck, 2011; Carron, Wolk & Szapudi, 2014)

# The power spectrum

- Binned power spectrum estimator

$$\hat{P}_{\rm b}(k_i) = \frac{1}{N_{k_i} V} \sum_{\boldsymbol{k}} \tilde{\rm T}_{\rm b}(\boldsymbol{k}) \tilde{\rm T}_{\rm b}(-\boldsymbol{k})$$




averaged over $N_{k_i} \approx \frac{V}{(2\pi)^2} {k_i}^2 \Delta k_i$

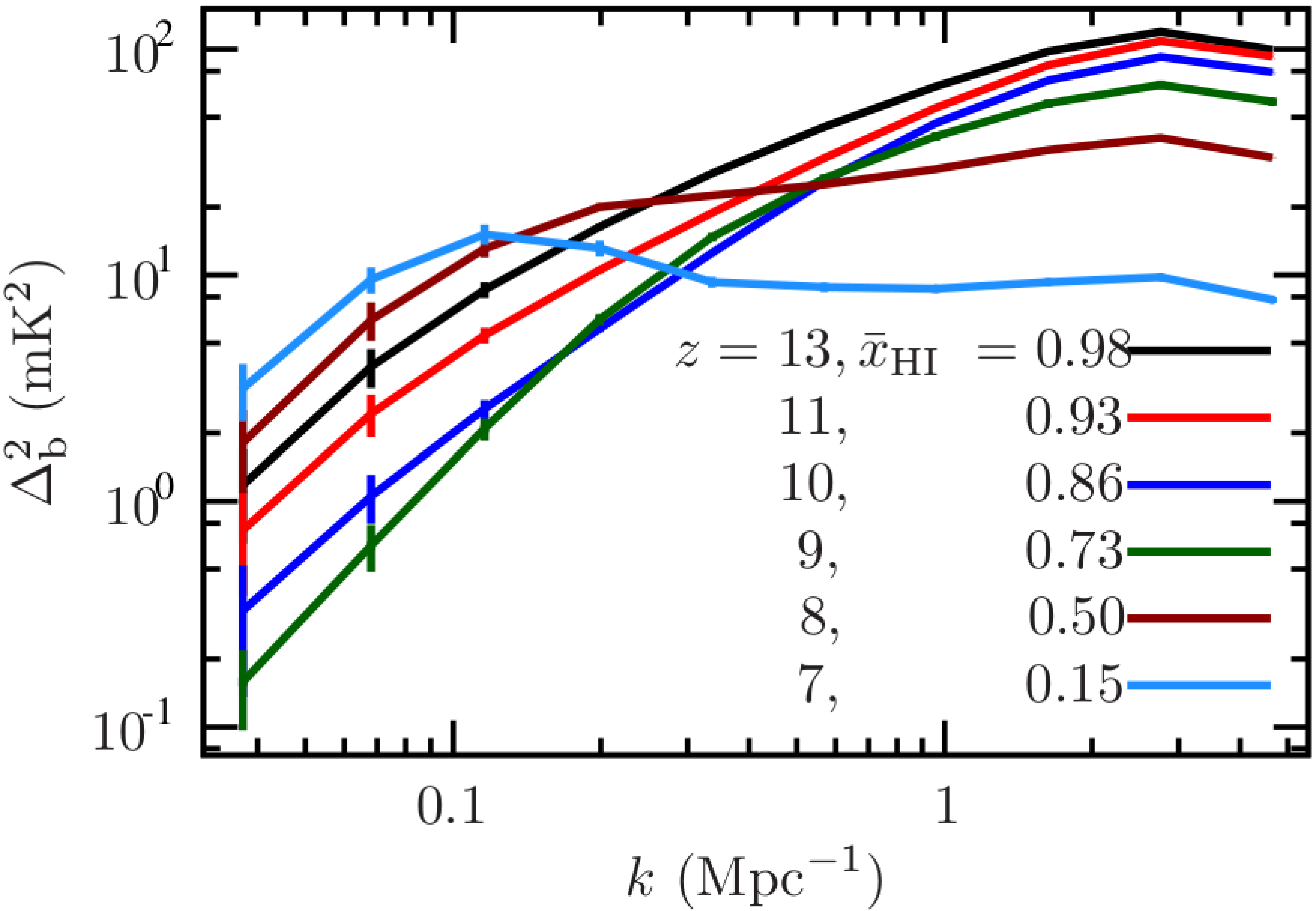
- The power spectrum is defined as

$$P(k) = V^{-1} \langle \tilde{T}_{\rm b}(k)\, \tilde{T}_{\rm b}(-k) \rangle$$

- Bin averaged power spectrum

$$\langle \hat{P}_{\rm b}(k_i) \rangle = \bar{P}_{\rm b}(k_i) = \frac{1}{N_{k_i}} \sum_{\boldsymbol{a}} P_{\rm b}(a)$$

Mondal et al. 2016a, arXiv: 1606.03874



# The error covariance

$$\mathbf{C}_{ij} = \langle [\hat{P}(k_i) - \bar{P}(k_i)]\,[\hat{P}(k_j) - \bar{P}(k_j)] \rangle$$
$$= [\langle \hat{P}(k_i)\,\hat{P}(k_j) \rangle] - \bar{P}(k_i)\,\bar{P}(k_j)$$

We have

$$\langle \hat{P}(k_i)\,\hat{P}(k_j) \rangle = \frac{1}{N_{k_i} N_{k_j} V^2} \sum_{\boldsymbol{k}_a \in i, \boldsymbol{k}_b \in j} \langle \tilde{T}_{\rm b}(\boldsymbol{k}_a)\,\tilde{T}_{\rm b}(-\boldsymbol{k}_a)\,\tilde{T}_{\rm b}(\boldsymbol{k}_b)\,\tilde{T}_{\rm b}(-\boldsymbol{k}_b) \rangle$$

The four-point statistics

$$\langle \tilde{T}_{\rm b}(\boldsymbol{k}_a)\,\tilde{T}_{\rm b}(\boldsymbol{k}_b)\,\tilde{T}_{\rm b}(\boldsymbol{k}_c)\,\tilde{T}_{\rm b}(\boldsymbol{k}_d) \rangle = V^2 [\,\delta_{a+b,0}\,\delta_{c+d,0}\,P(\boldsymbol{k}_a) P(\boldsymbol{k}_c)$$
$$+ \delta_{a+c,0}\delta_{b+d,0} P(\boldsymbol{k}_a) P(\boldsymbol{k}_b) + \delta_{a+d,0}\delta_{b+c,0} P(\boldsymbol{k}_a) P(\boldsymbol{k}_b)]$$
$$+ V\,\delta_{a+b+c+d,0}\,T(\boldsymbol{k}_a, \boldsymbol{k}_b, \boldsymbol{k}_c, \boldsymbol{k}_d)$$

Mondal et al. 2016b, arXiv: 1508.00896

# The error-covariance

Using the definition of trispectrum (four-point statistics)

$$\mathcal{C}_{ij} = \frac{\overline{P_{\rm b}}^2(k_i)}{N_{k_i}}\delta_{ij} + \frac{\bar{T}_{\rm b}(k_i, k_j)}{V}$$

Where $N_{k_i} \approx \frac{V}{(2\pi)^2} {k_i}^2 \Delta k_i$ and $\overline{P^2}(k_i) = \frac{1}{N_{k_i}} \sum_{\boldsymbol{k}} P^2(\boldsymbol{k})$

the square of the power spectrum averaged over the i-th bin

$$\bar{T}(k_i, k_j) = \frac{1}{N_{k_i} N_{k_j}} \sum_{\boldsymbol{k}_a \in i, \boldsymbol{k}_b \in j} T(\boldsymbol{k}_a, -\boldsymbol{k}_a, \boldsymbol{k}_b, -\boldsymbol{k}_b)$$

the average trispectrum where k_a and k_b are summed over the i-th and the j-th bins respectively

# The dimensionless error-covariance

$$\mathbf{c}_{ij} = \frac{\mathbf{C}_{ij}\, V\, k_i^{3/2} k_j^{3/2}}{(2\pi)^2 \bar{P}(k_i)\, \bar{P}(k_j)} \qquad \mathbf{c}_{ij} = A_i^2 \left( \frac{k_i}{\Delta k_i} \right) \delta_{ij} + t_{ij}$$

where $A_i = \sqrt{\frac{\overline{P^2}(k_i)}{[\bar{P}(k_i)]^2}}$ and $t_{ij} = \frac{\bar{T}(k_i, k_j)\, k_i^{3/2}\, k_j^{3/2}}{(2\pi)^2 \bar{P}(k_i)\, \bar{P}(k_j)}$

Mondal et al. 2016b, arXiv: 1508.00896

---

- The diagonal elements of $\mathbf{c}_{ij}$ quantifies the variance.
- We have $t_{ii} = 0$ if the EoR 21-cm signal is a Gaussian random field then we have $\mathbf{c}_{ii} = A_i^2 (k_i/\Delta k_i)$

# The dimensionless error-covariance

$$\mathbf{c}_{ij} = \frac{\mathbf{C}_{ij}\, V\; k_i^{3/2} k_j^{3/2}}{(2\pi)^2 \bar{P}(k_i)\, \bar{P}(k_j)} \qquad \mathbf{c}_{ij} = A_i^2 \left(\frac{k_i}{\Delta k_i}\right) \delta_{ij} + t_{ij}$$

where $A_i = \sqrt{\frac{\overline{P^2}(k_i)}{[\bar{P}(k_i)]^2}}$ and $t_{ij} = \frac{\bar{T}(k_i, k_j)\; k_i^{3/2}\, k_j^{3/2}}{(2\pi)^2 \bar{P}(k_i)\, \bar{P}(k_j)}$

Mondal et al. 2016b, arXiv: 1508.00896

---

- The diagonal elements of $\mathbf{c}_{ij}$ quantifies the variance.
- We have $t_{ii} = 0$ if the EoR 21-cm signal is a Gaussian random field then we have $\mathbf{c}_{ii} = A_i^2 (k_i/\Delta k_i)$

# The off-diagonal terms of error-covariance

- The off-diagonal terms of $\mathbf{C}_{ij}$ quantify the correlations between the errors in the power spectrum estimated at different bins.

- These terms are all zero if the signal is a Gaussian random field *i.e.* the errors in the different bins are uncorrelated.

- However, the EoR 21-cm signal becomes increasingly non-Gaussian as reionization proceeds, and we expect the off-diagonal terms to develop non-zero values.

- We interpret any statistically significant non-zero off-diagonal component of $\mathbf{C}_{ij}$ as arising from the trispectrum.

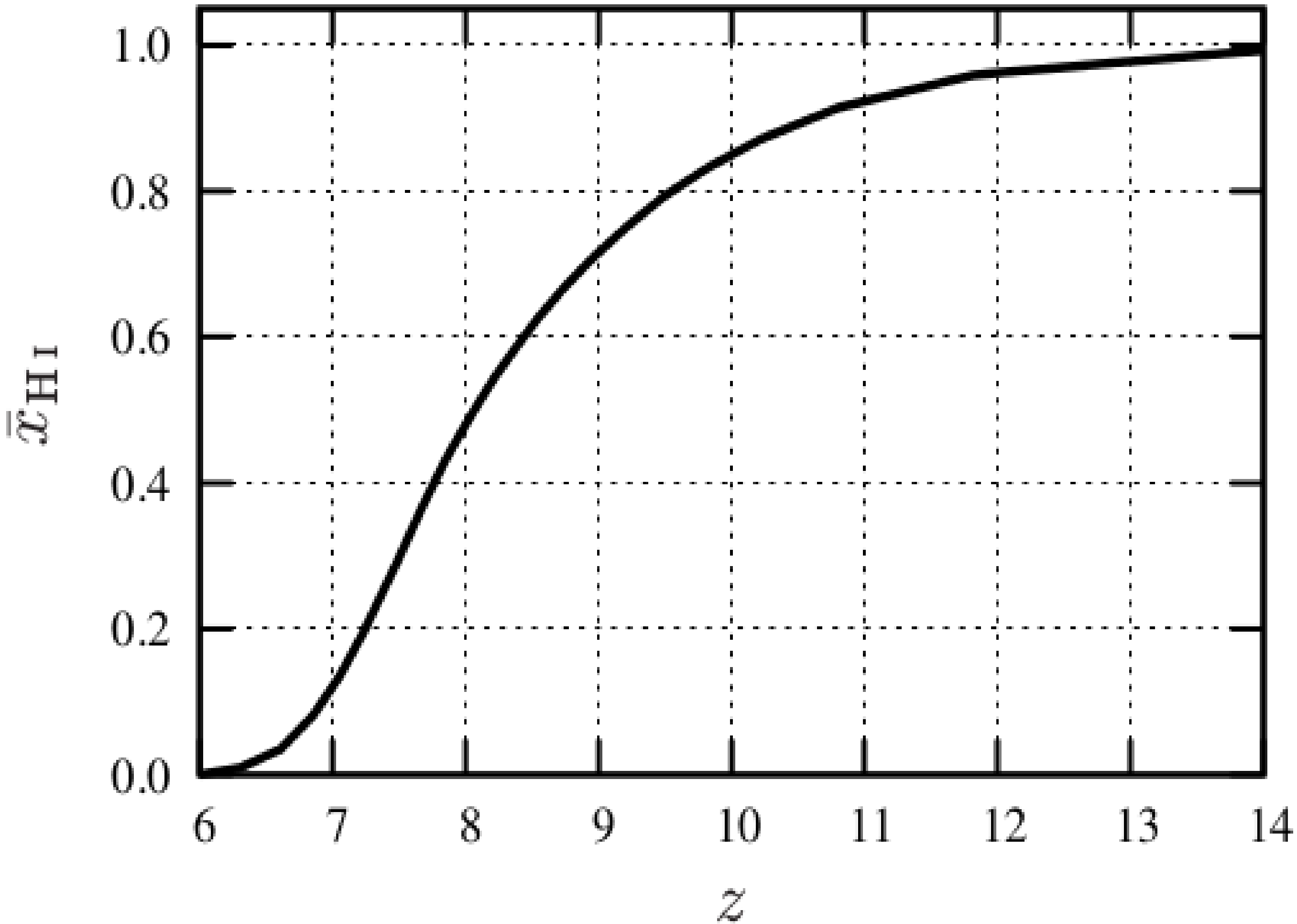
# Signal Ensemble (SE)

We have generated the redshifted EoR 21-cm signal using seminumerical simulations which involve three main steps.

- First, we use a particle mesh *N*-body code to generate the dark matter distribution at the different redshifts.
- In the next step we use the Friends-of-Friends algorithm to identify collapsed halos in the dark matter distribution.
- The third and final step generates the ionization map based on an excursion set formalism (Furlanetto et al. 2004).

- We have run 50 independent realizations of the simulations to generate an ensemble of 50 statistically independent realizations of the EoR 21-cm signal. We refer to this ensembles as the Signal Ensemble (SE).

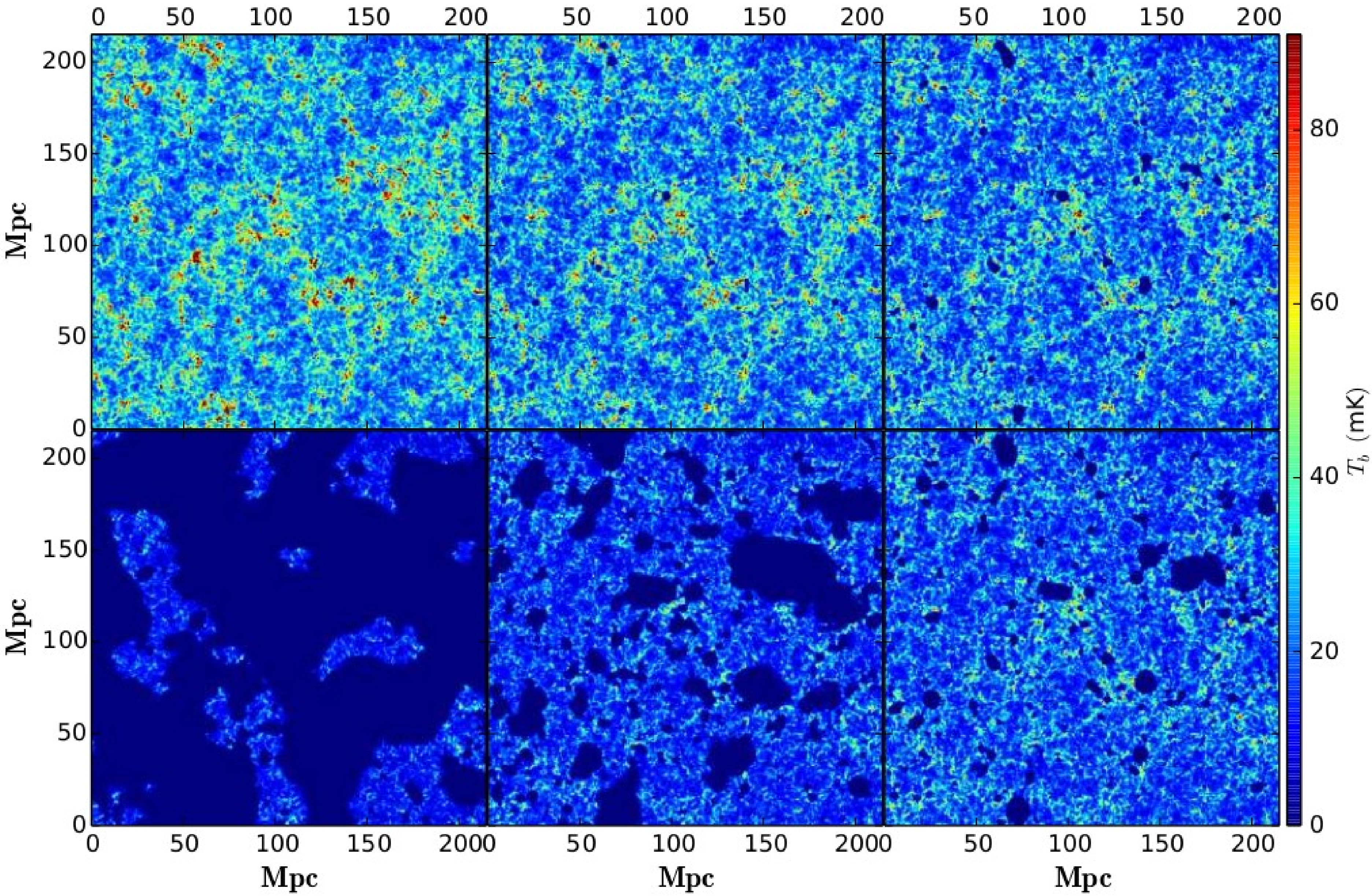
| $z$ | $\bar{x}_{\rm H\,I}$ |
|---|---|
| 13 | 0.98 |
| 11 | 0.93 |
| 10 | 0.86 |
| 9 | 0.73 |
| 8 | 0.50 |
| 7 | 0.15 |




Left: This tabulates the redshifts (z) and corresponding mass averaged neutral fraction

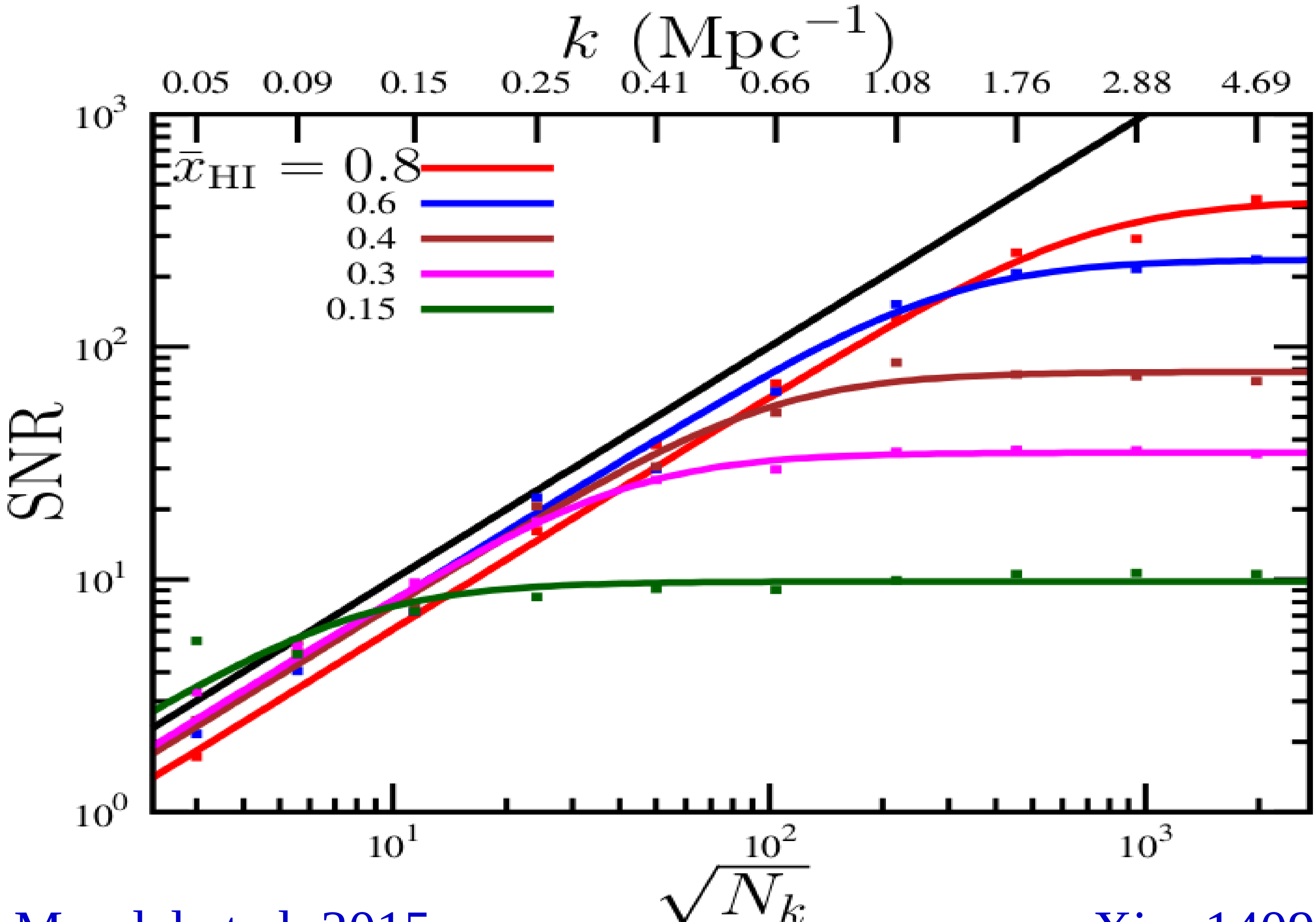
Right: This shows the reionization history

Mondal et al. 2016a, arXiv: 1606.03874




Mondal et al. 2016a, arXiv: 1606.03874

signal to noise ratio $SNR = \bar{P}_b(k)/\delta P_b(k) = \sqrt{N_k}$




Mondal et al. 2015, arXiv: 1409.4420

# Mini Summary

- For Gaussian random field, we expect the SNR to scale as the square-root of the number of independent measurements.
- We find the expected SNR $\propto \sqrt{N_k}$ behaviour at low SNR
- For larger SNR it increases slower than $\sqrt{N_k}$ and finally saturates at a limiting value

$$[\mathrm{SNR}]_l = \sqrt{\frac{[\bar{P}_{\mathrm{b}}(k)]^2 V}{\bar{T}_{\mathrm{b}}(k,k)}}$$

- As the reionization proceeds, the ionized bubbles grow (both in number and size), thus affect power spectrum error estimates more in the later stages of the EoR.

# The diagonal terms of error-covariance

- We expect the diagonal term to have values $\mathbf{c}_{ii} = A_i^2 (k_i/\Delta k_i)$ if the 21-cm signal is a Gaussian random field.
- We interpret any excess relative to this prediction as arising from the trispectrum $t_{ii}$ which arises when the EoR 21-cm signal becomes non-Gaussian.

- The difficulty is that it is not possible to predict the precise value of $\overline{P^2}(k_i)$
- In other words, it is not possible to use the Signal Ensemble (SE) to determine the contribution from the non-Gaussianity

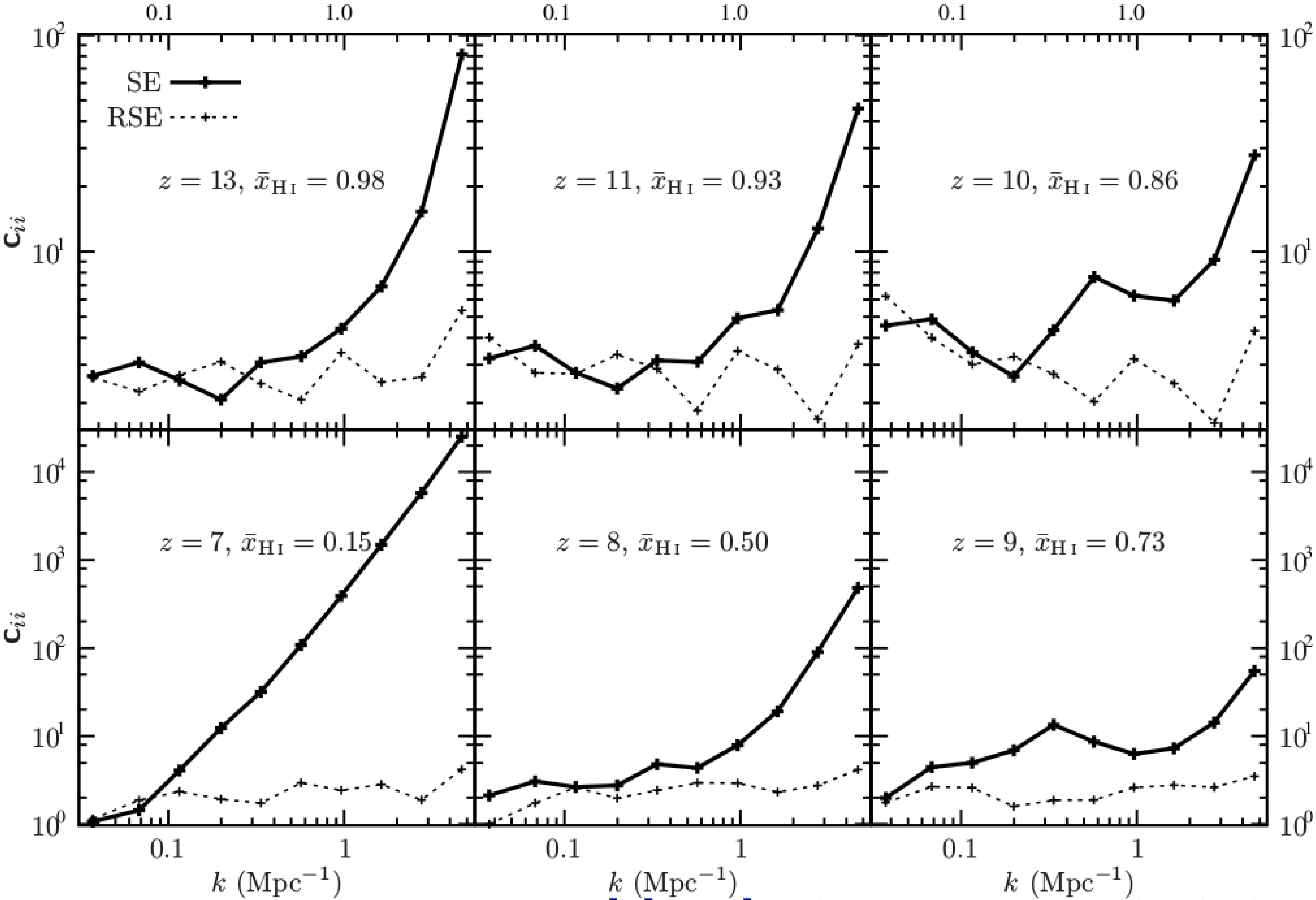
- We use the Randomized Signal Ensemble (RSE) to interpret the diagonal terms of the error-covariance.

# The Randomized Signal Ensemble (RSE)

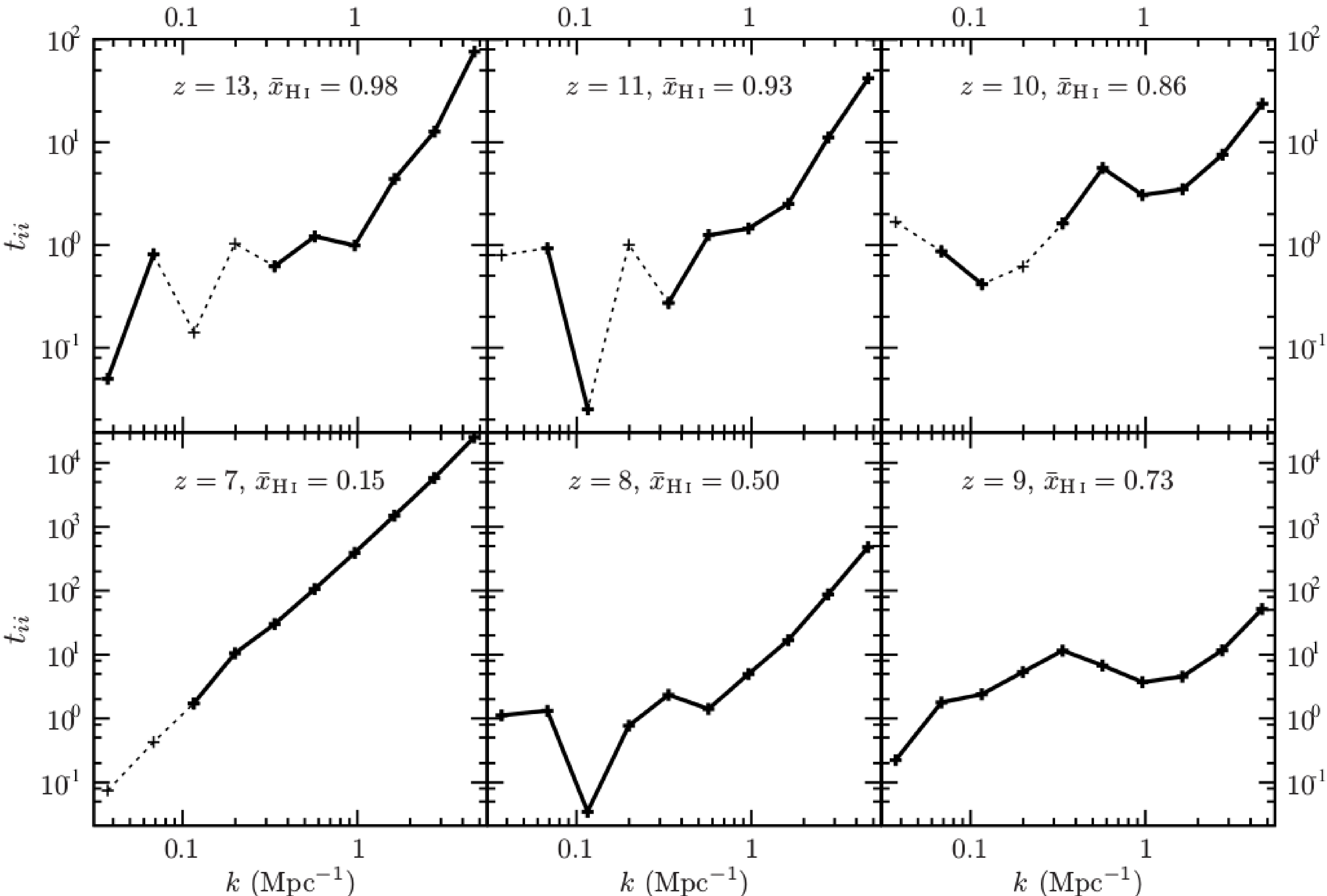
- Each realization in RSE is a mixture of Fourier modes $\tilde{T}_{\rm b}(\boldsymbol{k})$.
- It is expected that modes from one realization in SE is not correlated with those from other realization in SE

- The average trispectrum $\bar{T}_{\rm b}(k_i, k_j)$ is at least 50 times smaller for RSE as compared to SE.
- We expect $\bar{P}_{\rm b}(k_i)$ and $\overline{P_{\rm b}{}^2}(k_i)$ to have exactly the same value in both SE and RSE. Essentially A_i is same
- RSE has been used to estimate the error-covariance that would be expected if the signal were a Gaussian random

- It thus becomes possible to interpret any deviations from this as arising from trispectrum. $[\mathbf{C}_{ii}]_{\rm SE} - [\mathbf{C}_{ii}]_{\rm RSE} = t_{ii}$




Mondal et al. 2016a, arXiv: 1606.03874

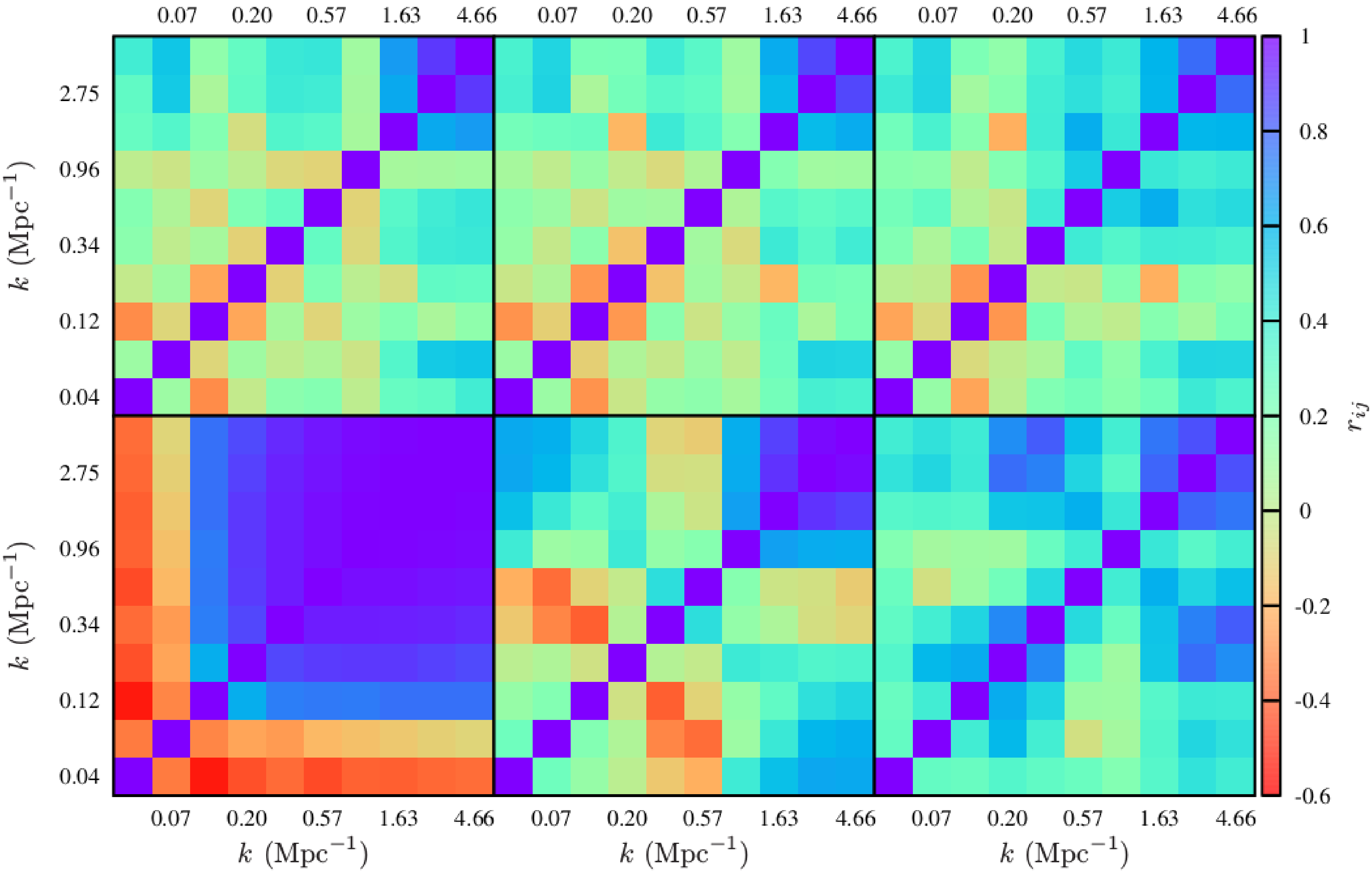



Mondal et al. 2016a, arXiv: 1606.03874

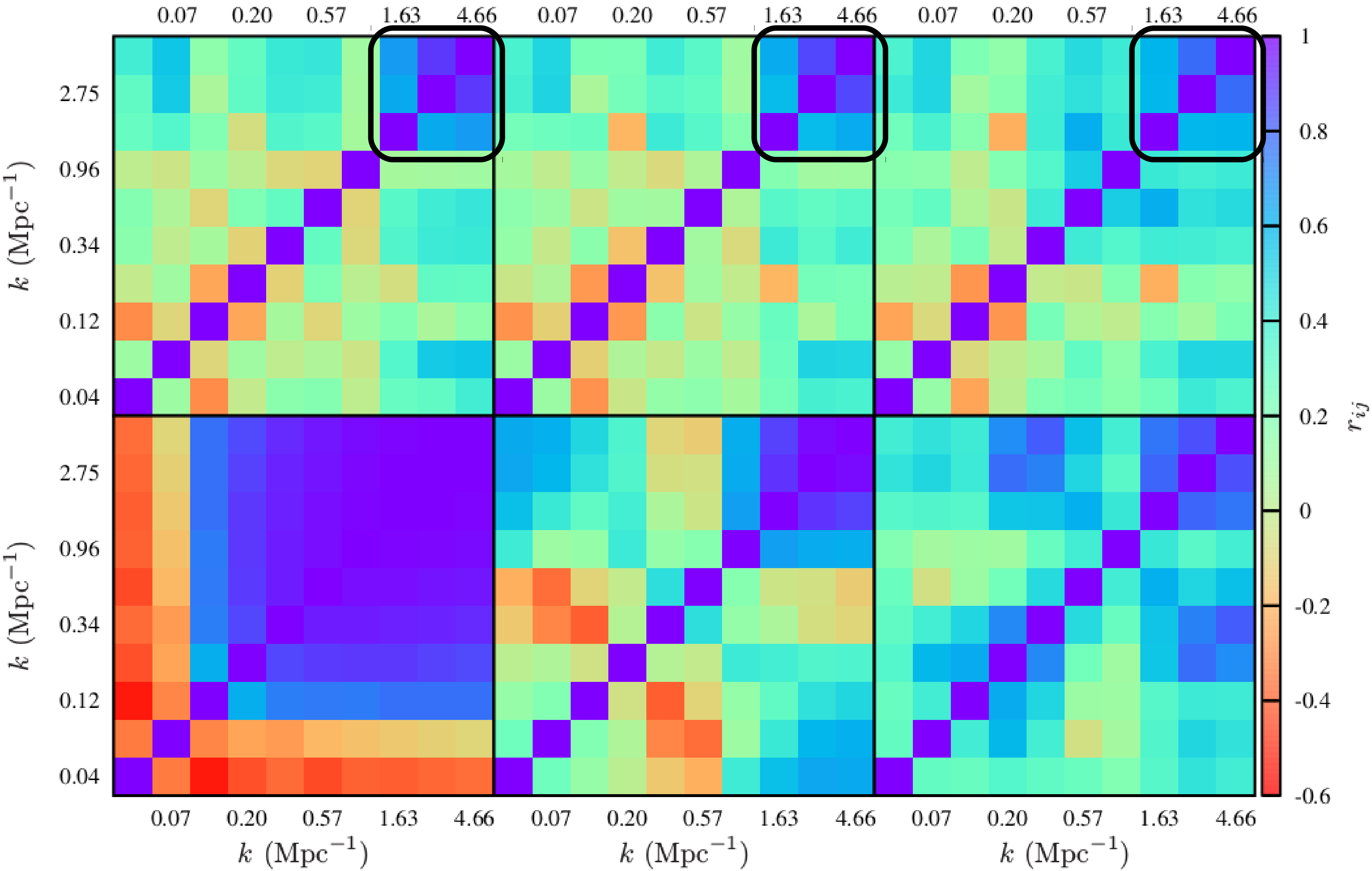
# The off-diagonal terms of error-covariance

- The off-diagonal terms of $\mathbf{C}_{ij}$ quantify the correlations between the errors in the power spectrum estimated at different bins.

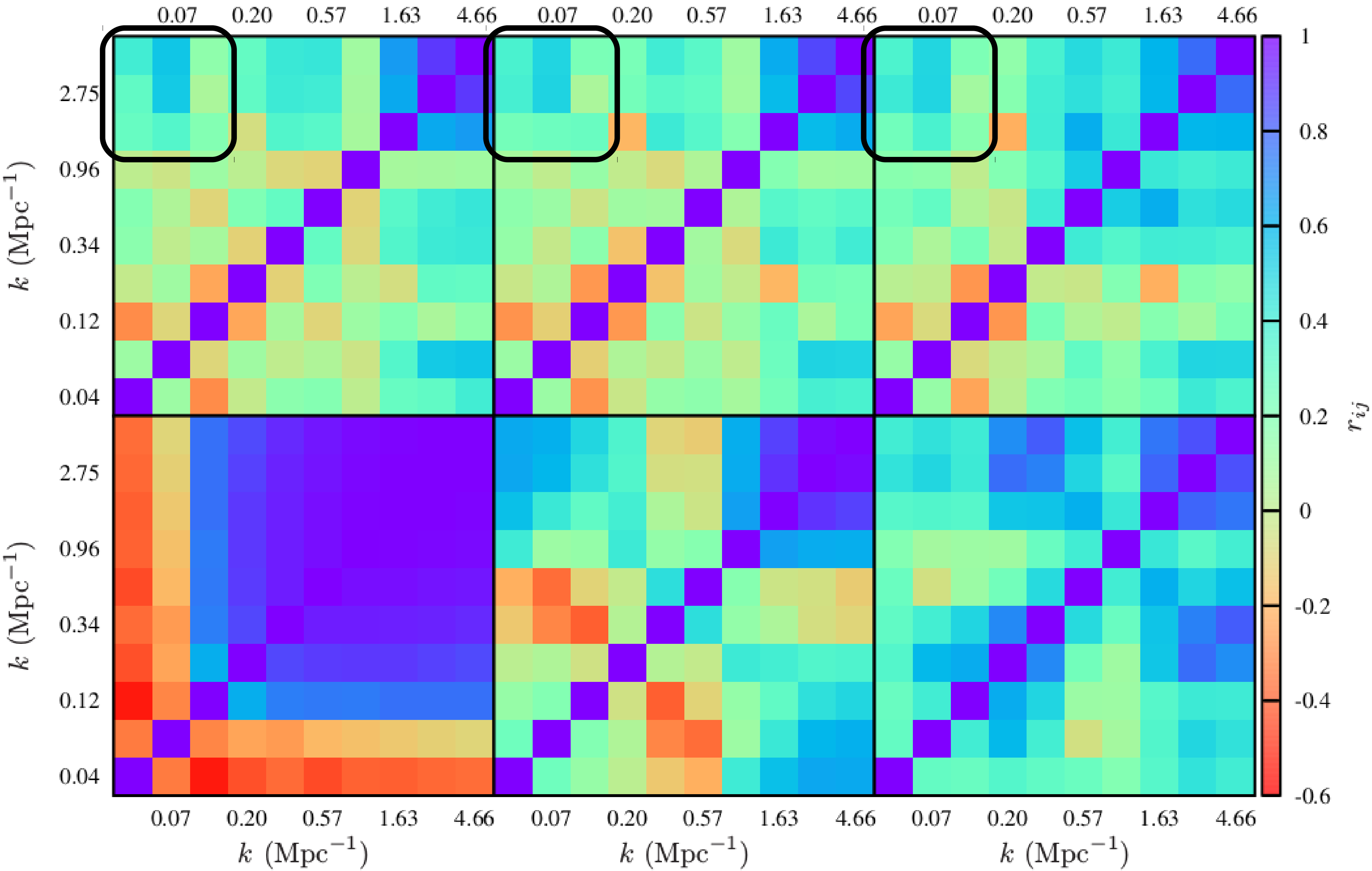
The correlation coefficient $$r_{ij} = \frac{c_{ij}}{\sqrt{c_{ii}\, c_{jj}}}$$



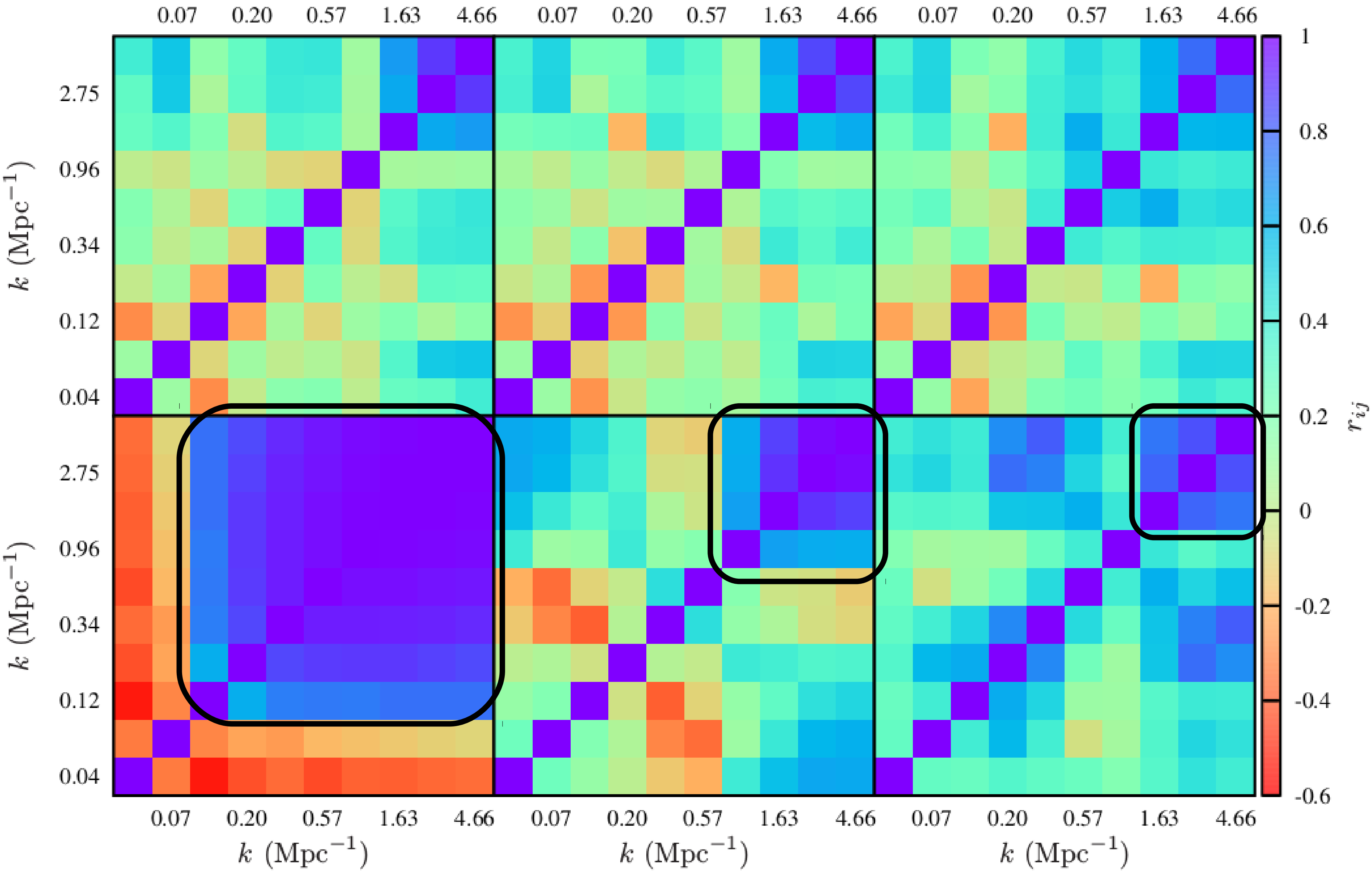

Mondal et al. 2016a, arXiv: 1606.03874



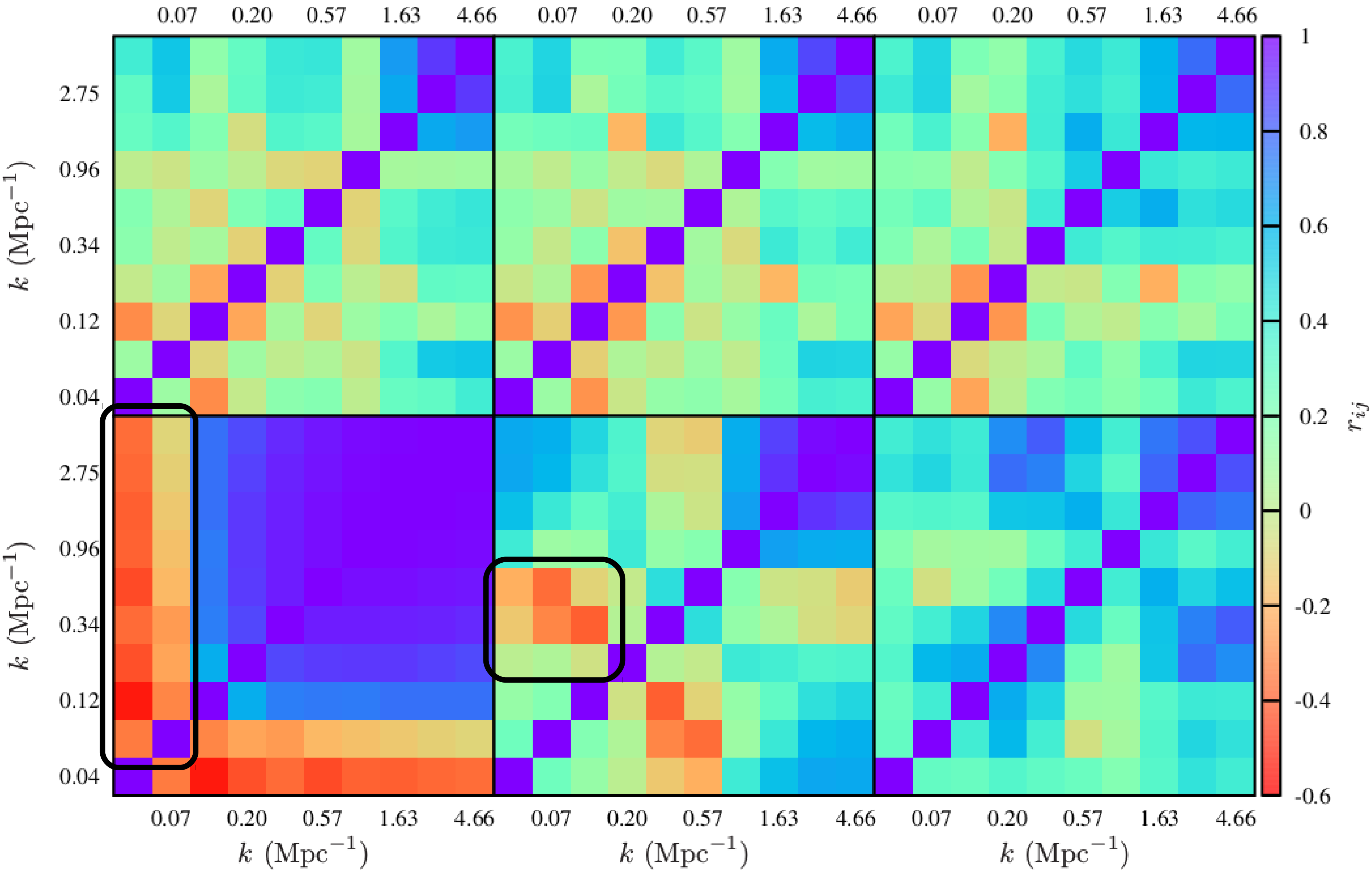

Mondal et al. 2016a, arXiv: 1606.03874




Mondal et al. 2016a, arXiv: 1606.03874



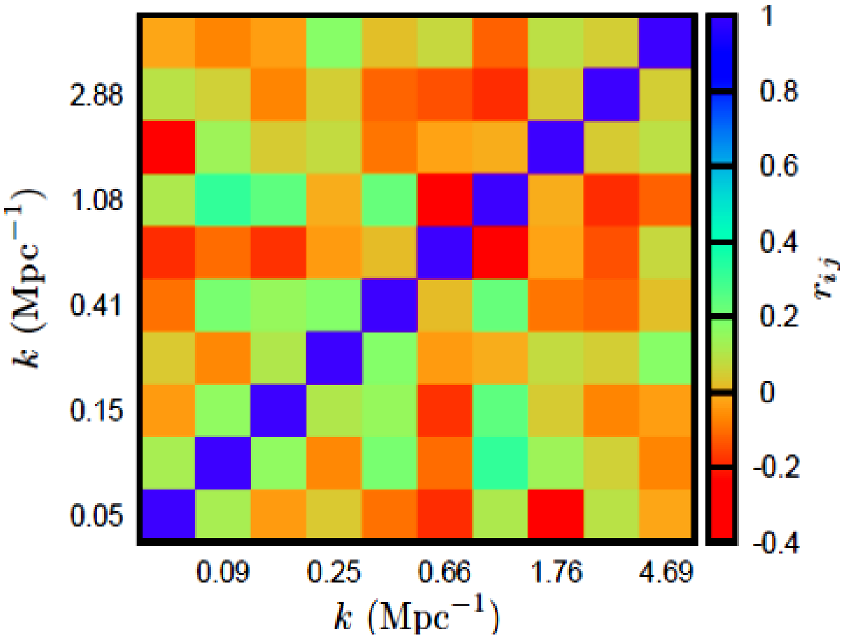

Mondal et al. 2016a, arXiv: 1606.03874

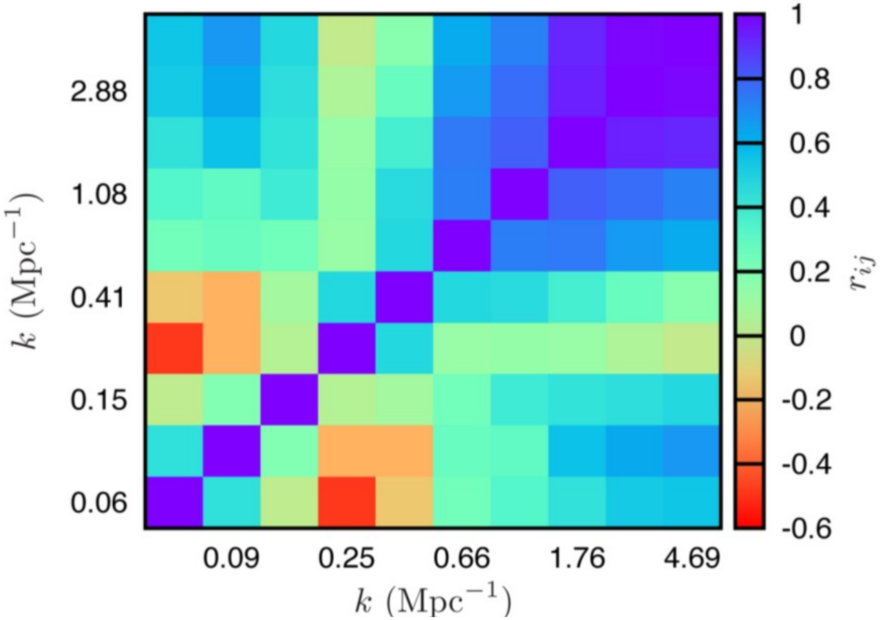



Mondal et al. 2016a, arXiv: 1606.03874

# The Gaussian Random Ensemble (GRE)

- The GRE contains 50 realizations of the 21-cm signal, the signal in each realization is a Gaussian random field.
- Error-covariance for a Gaussian random field is diagonal



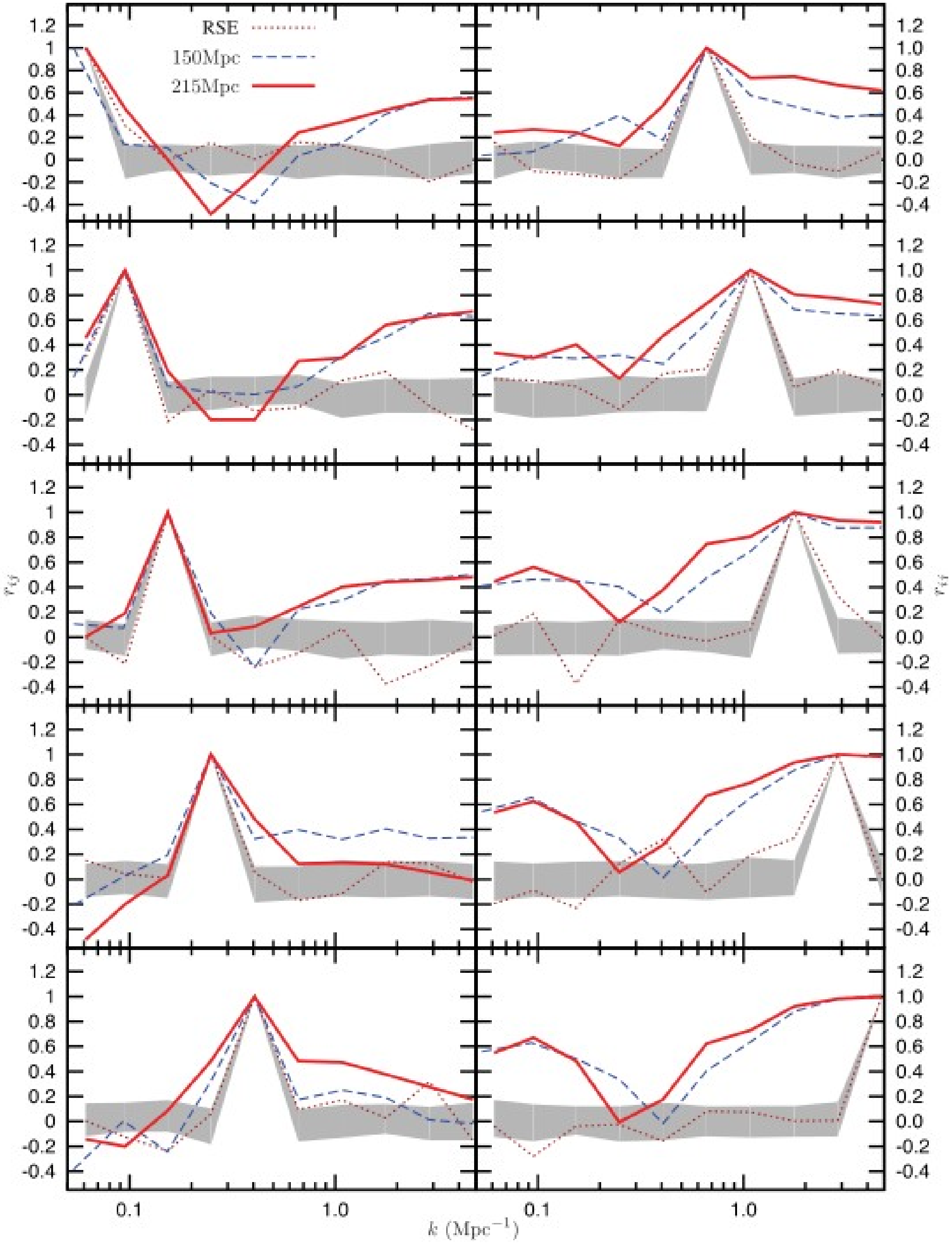

Mondal et al. 2016b, arXiv: 1508.00896

# Ensemble of Gaussian Random Ensembles (EGRE)

- The off-diagonal terms of the error-covariance estimated from GRE will not be zero due to finite number of realizations

- 50 independent GREs are used to construct an EGRE which we have used to estimate the variance of covariance.

- We compare the error-covariance $\mathbf{c}_{ij}$ against the variance of covariance from EGRE to determine statistical significance




Mondal et al. 2016b, arXiv: 1508.00896

# Summary and discussion

- The non-Gaussian components are correlated, which is quantified through trispectrum in $\mathbf{C}_{ij}$.
- The EoR 21-cm signal becomes increasingly non-Gaussian. This manifests itself as a non-zero trispectrum in $\mathbf{C}_{ij}$.
- The diagonal elements quantifie the variance of error in P(k).
- It is not possible to use the SE to independently determine the contributions of non-Gaussianity in $\mathbf{C}_{ij}$
- We have overcome this problem by constructing the RSE
- The difference $[\mathbf{C}_{ii}]_{\rm SE} - [\mathbf{C}_{ii}]_{\rm RSE} = t_{ii}$
- The off-diagonal terms of $\mathbf{C}_{ij}$ quantify the correlations between the errors in the P(k) estimated at different bins.
- We interpret any statistically significant non-zero off-diagonal component as arising from $t_{ij}$ (using EGRE).