Analytical model for density-split galaxy clustering

New Strategies for Extracting Cosmology from Galaxy Surveys (2nd edition)

Mathilde Pinon - *PhD student at CEA Saclay*

with Arnaud de Mattia, Étienne Burtin, Vanina Ruhlmann-Kleider

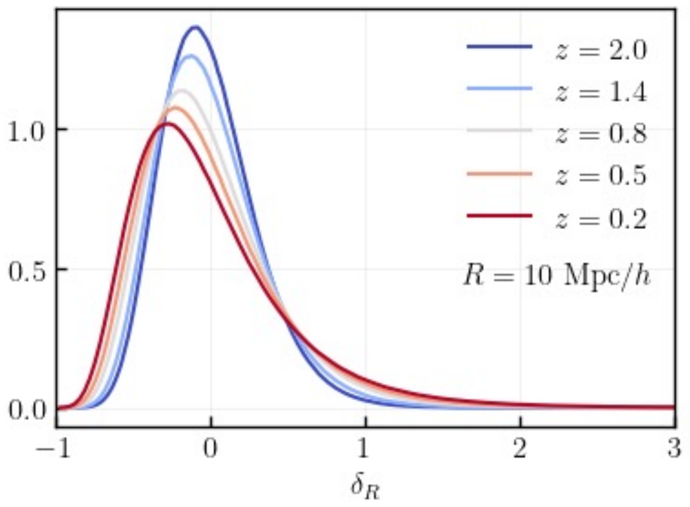
Dark Energy Spectroscopic Instrument
U.S. Department of Energy Office of Science

IRFU – CEA Saclay
Institut de recherche sur les lois fondamentales de l'univers

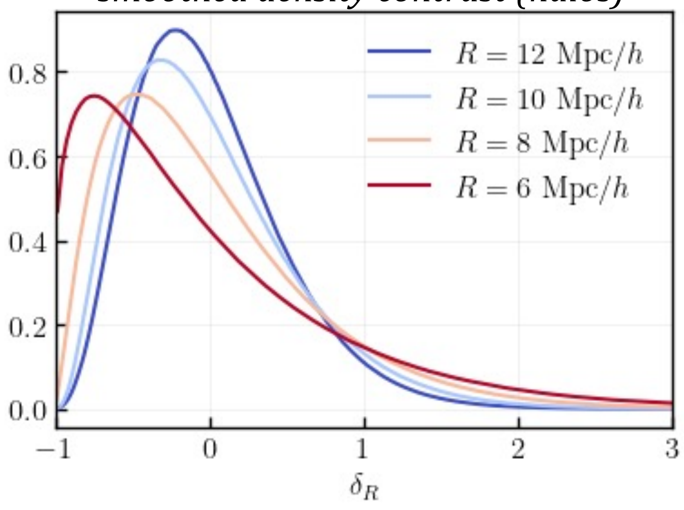
The matter density field has evolved to become non-Gaussian

- The matter density field is **non-Gaussian** at **late time** and **small scales**
- 2-point statistics (correlation function, power spectrum) are not sufficient to describe it entirely
- We need **alternative statistics** to **capture non-Gaussian information** from the density field

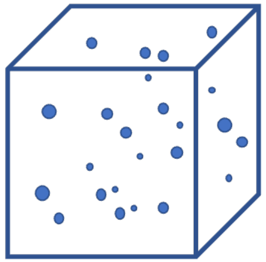
smoothed density contrast (dark matter)



smoothed density contrast (halos)

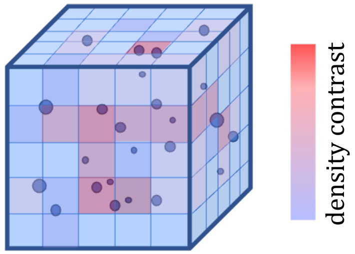


Density-split statistics



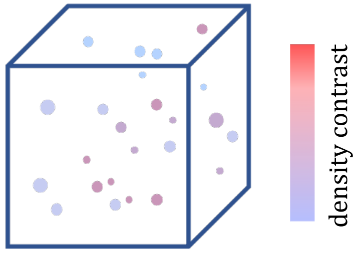
catalog of tracers

1



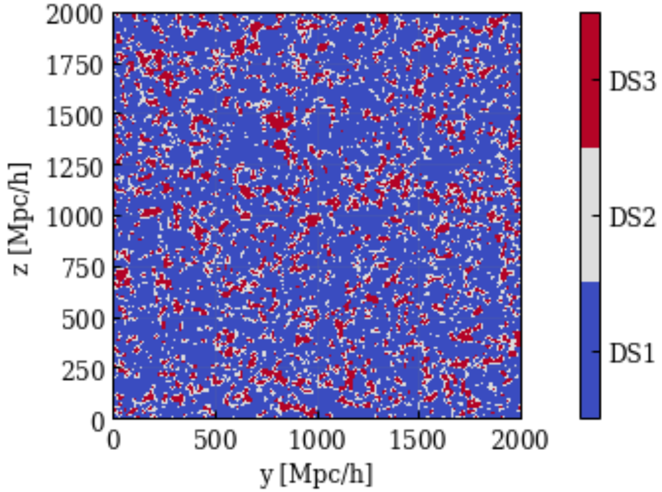
density mesh with cell size R

2



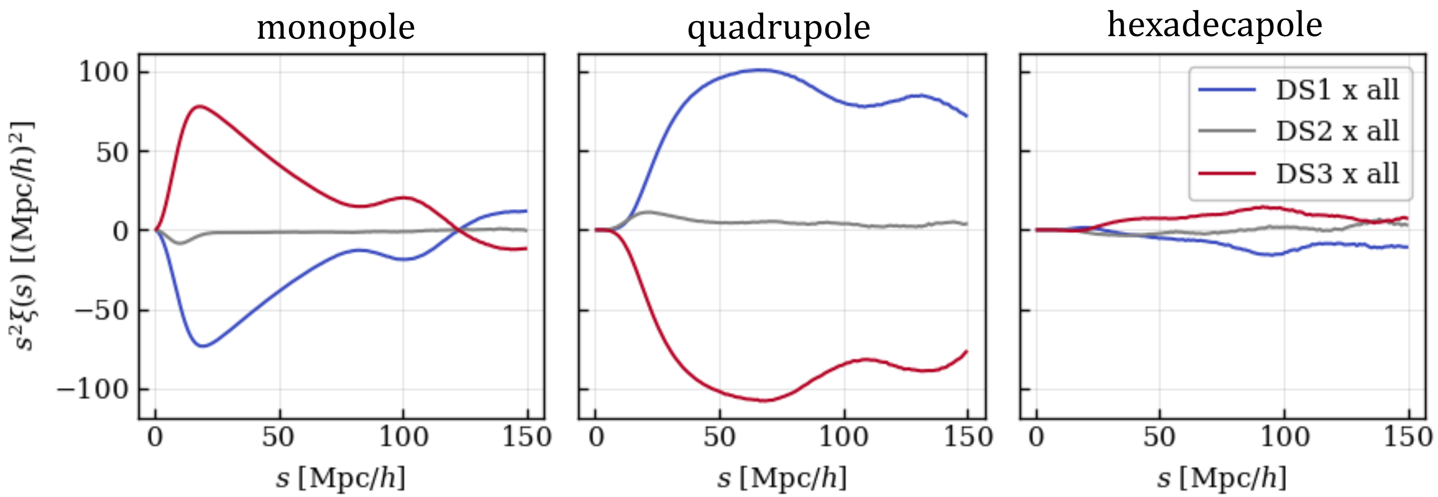
density at each particle location

3



density is split into N quantiles

4



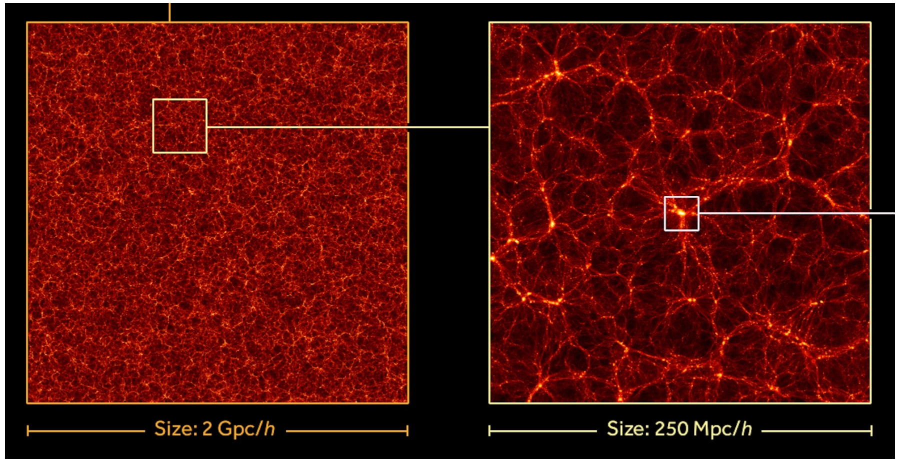
random points in each region are cross-correlated with tracers

5

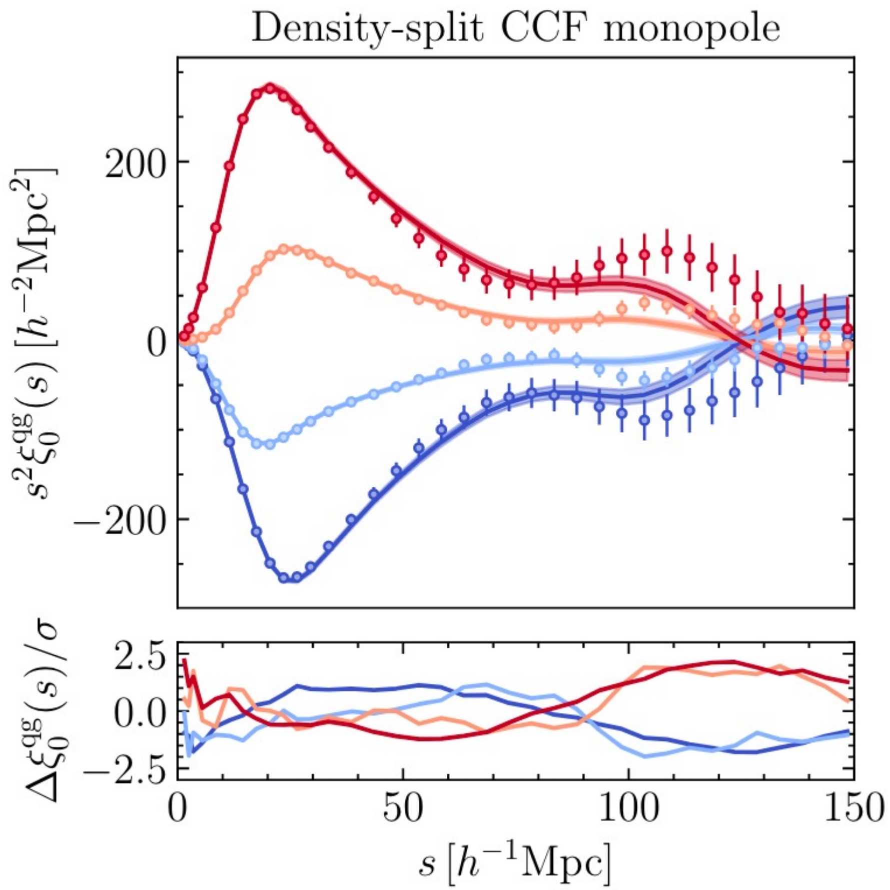
Density splits help tighten constraints on cosmological parameters

Paillas et al. 2023
http://arxiv.org/abs/2309.16541

- BOSS DR12 CMASS sample ($0.45 \leq z \leq 0.6$)
- $R = 10$ Mpc$/h$
- 5 density splits, auto and cross-correlation functions
- emulator trained on `AbacusSummit` N-body simulations



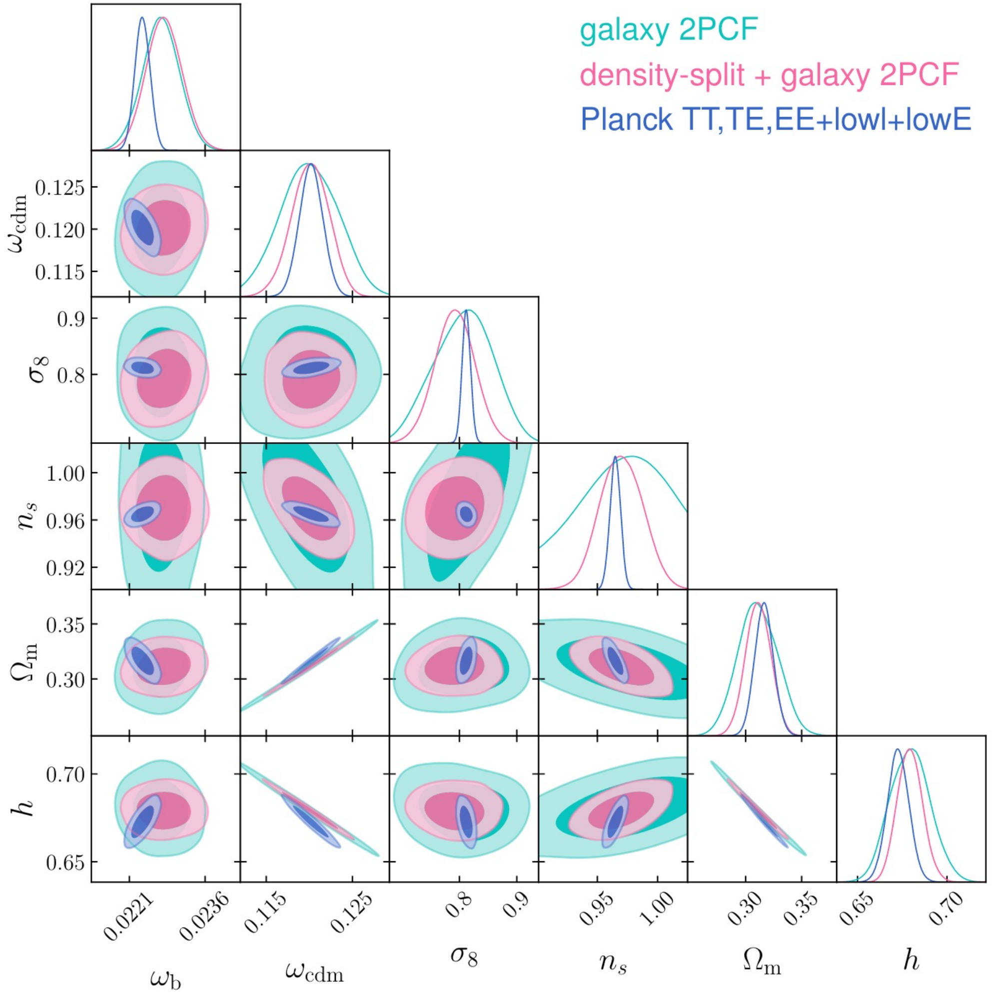
Garrison et al. 2021, Maximova et al. 2021



Density splits help tighten constraints on cosmological parameters

Paillas et al. 2023
http://arxiv.org/abs/2309.16541

- BOSS DR12 CMASS sample ($0.45 \leq z \leq 0.6$)
- emulator model fitted in $1 \text{ Mpc}/h < s < 150 \text{ Mpc}/h$
- CMB acoustic scale fixed
- **1.9 to 2.9× improved precision** on $\sigma_8, \omega_{cdm}, n_s$ with respect to 2PCF only
- 4.3% constraint on $f\sigma_8$: ~**2× better** than BOSS main analysis



What are the building blocks of density-split clustering statistics?

$$\xi^{\mathrm{DS}}_{R_1,R_2}(s) = \frac{\langle n^{\mathrm{DS}}_{R_1}(r) n_{R_2}(r+s)\rangle}{\bar{n}^{\mathrm{DS}}_{R_1}\bar{n}} - 1$$

(smoothed) density-split correlation function at separation s

$$\langle n^{\mathrm{DS}}_{R_1}(r) n_{R_2}(r+s)\rangle = \bar{n}^2 \int_{\mathrm{DS}} d\delta_{R_1}(r) \int_{-\infty}^{+\infty} d\delta_{R_2}(r+s)(1+\delta_{R_2}(r+s))\mathcal{P}(\delta_{R_1}(r), \delta_{R_2}(r+s))$$

we need to know the joint PDF of $\delta_{R_1}(r), \delta_{R_2}(r+s)$

here:

- δ_R is the density constrast smoothed by some kernel with smoothing scale R
- DS is a given region of density (« density split »)

A simple case: Gaussian density field

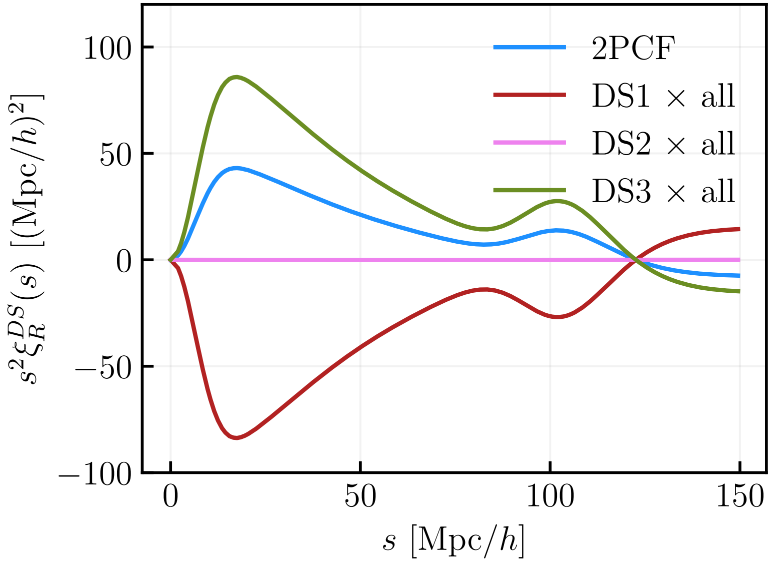
If we assume that $\delta_{R_1}(r), \delta_{R_2}(r+s)$ follows a **bivariate Gaussian distribution**, we find:

average density in density region DS

$$\xi^{DS}_{R_1 R_2}(s) = \frac{\tilde{\delta}_{DS}}{\sigma^2_{R_1 R_2}} \times \xi_{R_1 R_2}(s)$$

smoothed 2PCF

covariance of $\delta_{R_1}, \delta_{R_2}$



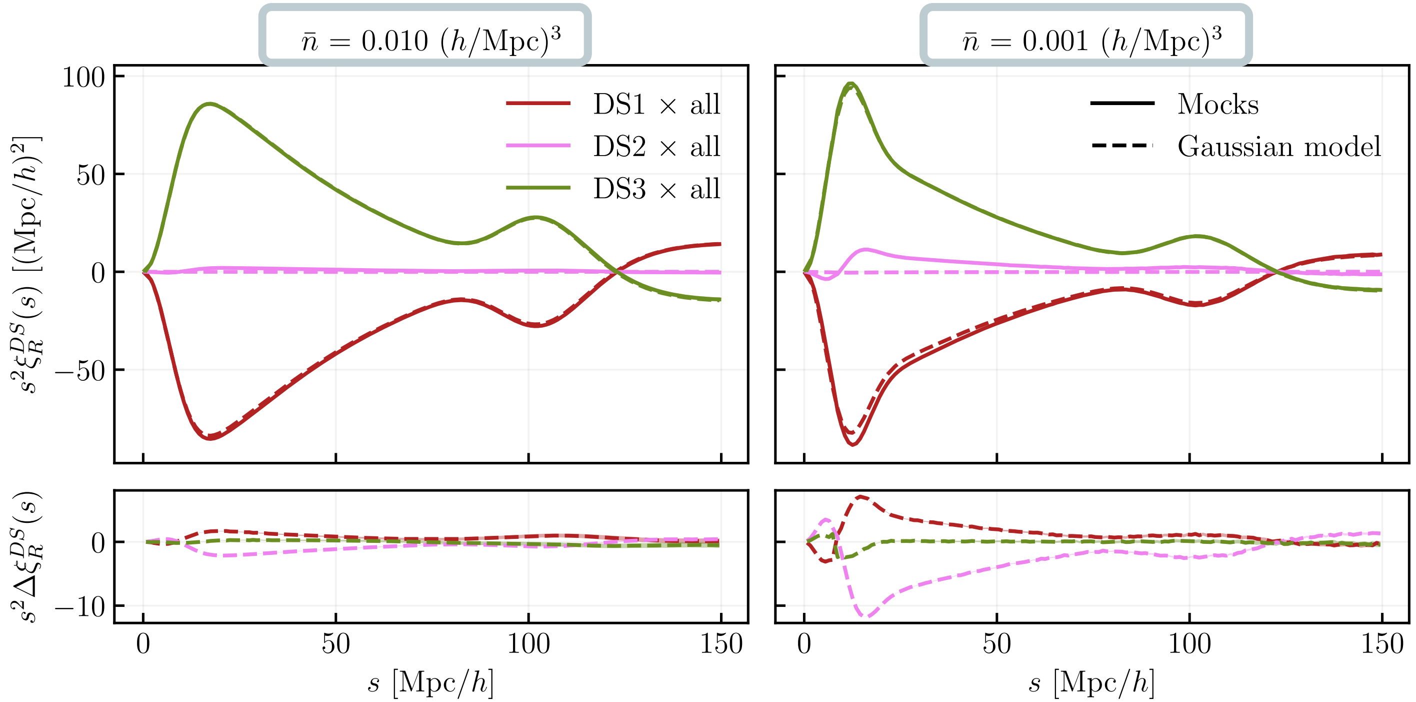
$\xi^{DS}_{R_1 R_2}(s)$ has the same shape as $\xi_{R_1 R_2}(s)$
but rescaled by $\propto$ the average density in DS → *information from the density on small scales is spread out to larger scales*

(similar result to Kaiser 1984 10.1086/184341)

Analytical Gaussian model vs Gaussian simulations

- Gaussian density-split model successfully describes density from Gaussian simulations provided that shot noise is low (so that the density is really Gaussian)
- but real matter density field is not Gaussian anyway

$R = 10\ \mathrm{Mpc}/h$



Pinon et al., *in prep.*

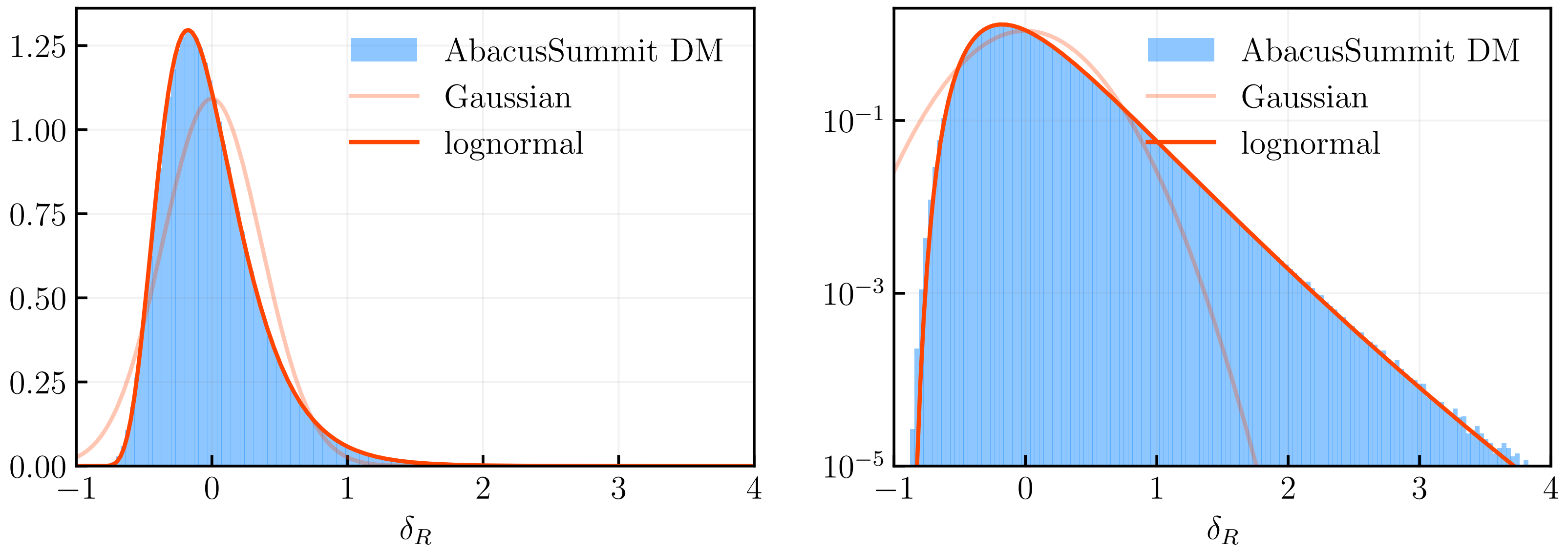
In practice the matter density field is not Gaussian, but is close to lognormal

- Dark matter (DM) density field computed from AbacusSummit simulation
(Garrison et al. 2021 10.1093/mnras/stab2482, Maximova et al. 2021 10.1093/mnras/stab2484)

$R = 10\ \mathrm{Mpc}/h$

simulation characteristics

- N-body simulation
- 2 Gpc/h cubic box
- $\bar{n} = 0.003\ (h/\mathrm{Mpc})^3$
- $z = 0.8$



A more realistic case: lognormal density field

- assumption: $Y_{R_1}, Y_{R_2} = ln\left(1+\frac{\delta_{R_1}}{\delta_{0,R_1}}\right), ln\left(1+\frac{\delta_{R_2}}{\delta_{0,R_2}}\right)$ follows a **bivariate Gaussian distribution**

Gaussian model

$$\xi^{DS}_{R_1R_2}(s) = \frac{\int_{\delta_1}^{\delta_2} d\delta_{R_1}\delta_{R_1}\mathcal{P}(\delta_{R_1})}{\int_{\delta_1}^{\delta_2} d\delta_{R_1}\mathcal{P}(\delta_{R_1})}\frac{\xi_{R_1r_2}(s)}{\sigma^2_{R_1R_2}}$$

→

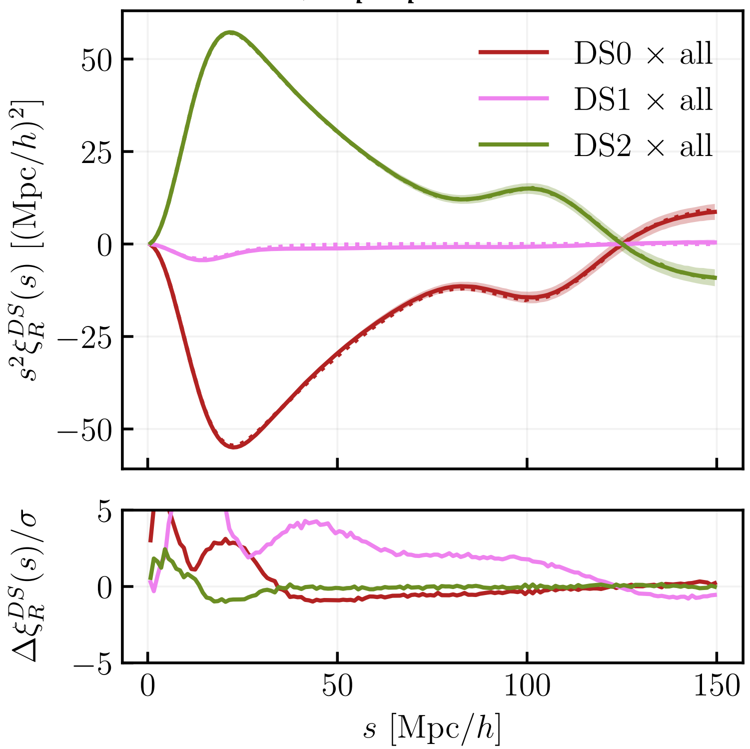
lognormal model

$$\xi^{DS}_{R_1R_2}(s) = \delta_{0,R_2}\frac{\left[\int_{\delta}^{\delta+\left(1+\frac{\delta}{\delta_{0,R_1}}\right)\frac{\xi_{R_1R_2}(s)}{\delta_{0,R_1}}} d\delta_{R_1}\mathcal{P}(\delta_{R_1})\right]_{\delta_1}^{\delta_2}}{\int_{\delta_1}^{\delta_2} d\delta_{R_1}\mathcal{P}(\delta_{R_1})}$$

Lognormal model vs. Abacus dark matter simulations

- assumption: $Y_{R_1}, Y_{R_2} = ln\left(1 + \frac{\delta_{R_1}}{\delta_{0,R_1}}\right), ln\left(1 + \frac{\delta_{R_2}}{\delta_{0,R_2}}\right)$ follows a **bivariate Gaussian distribution**
- good qualitative agreement but not at the level of the mocks' precision for DS1/at small scales

Pinon et al., *in prep.*

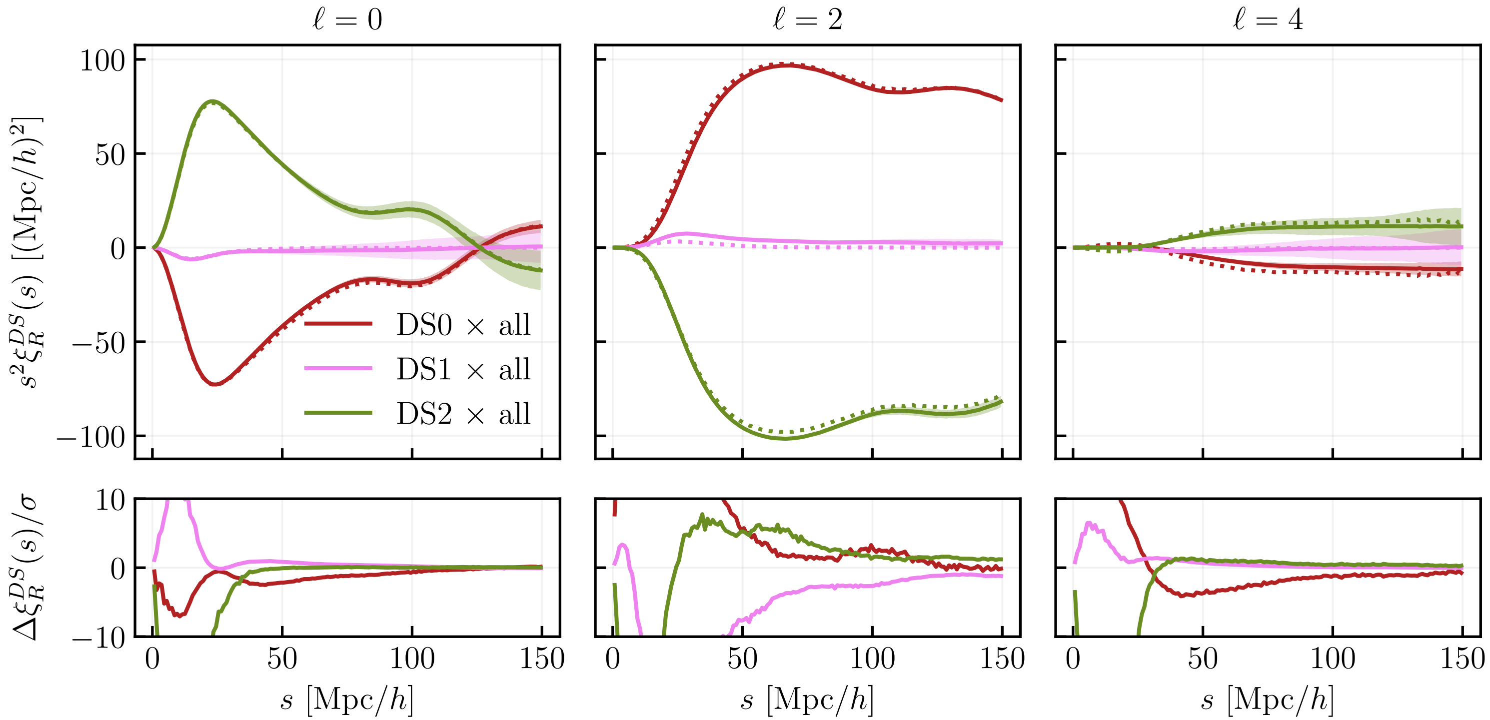


- 25 AbacusSummit boxes
- $\bar{n} = 0.0034\ (h/\text{Mpc})^3$
- $z = 0.8$
- $R_1 = R_2 = 10$ Mpc/h

residuals between model and mocks over mocks' standard deviation

What about redshift space?

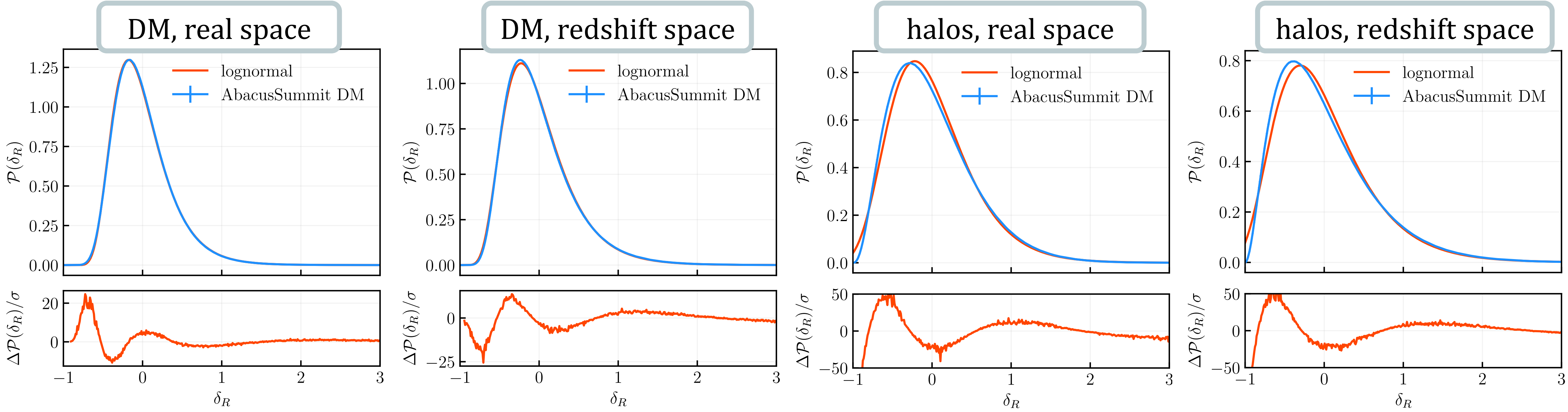
- in practice, we observe galaxies with redshift space distortions (RSD)
- good qualitative agreement but not at the level of the mocks' precision for the quadrupole/small scales



Pinon et al., *in prep.*

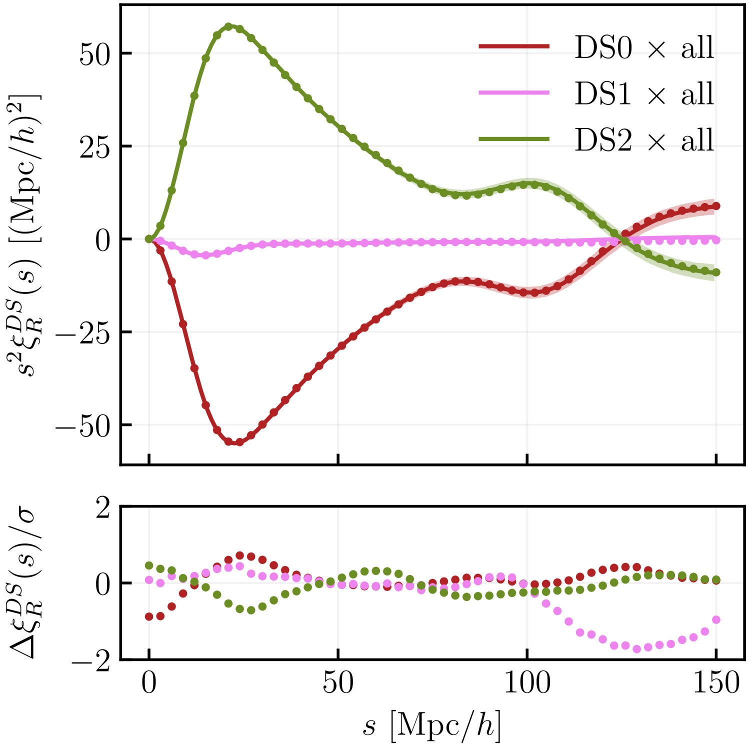
Lognormal assumption is not accurate enough

- **high density** regions (DS2) are the better modelled by the lognormal assumption in real space
- but lognormal model fails for low density regions (DS0, DS1)



Model computed directly from the measured joint PDF

- If we know the true joint PDF of $\delta_{R_1}(r), \delta_{R_2}(r+s)$, we can model density-split correlation very well

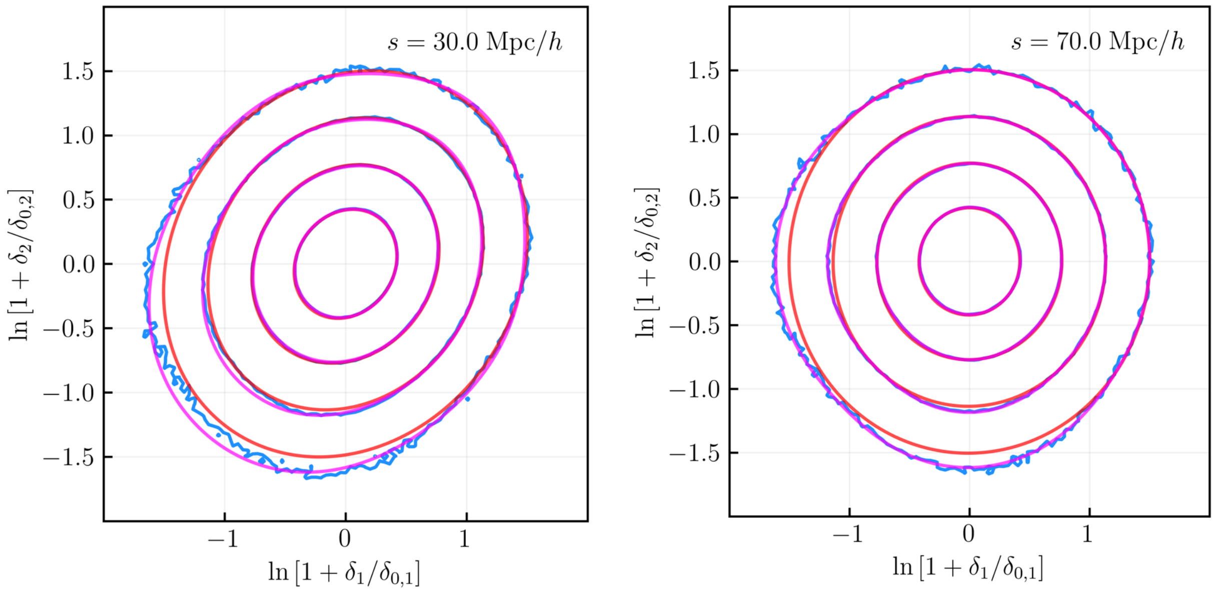


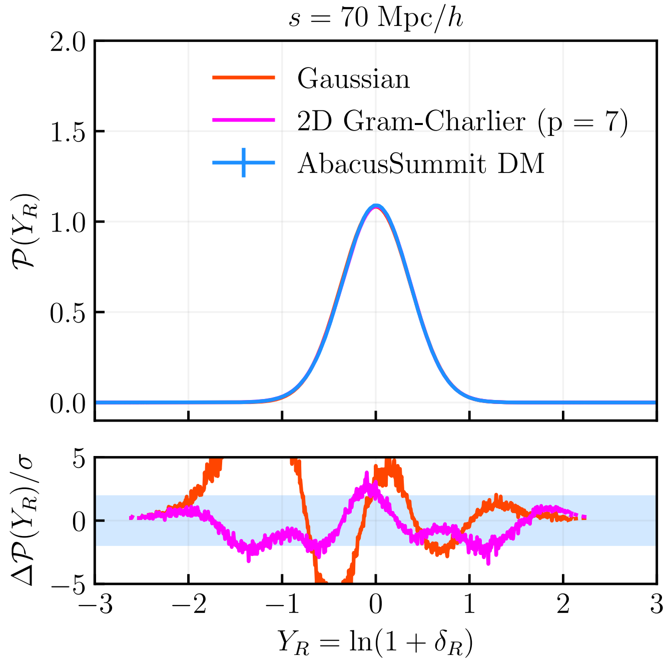
residuals between model and mocks over mocks' standard deviation

Gram-Charlier expansion to improve the model?

- let's look at the lognormal transform of $\delta_{R_1}, \delta_{R_2}$ which is nearly Gaussian
- Gram-Charlier expansion **breaks down at scales < 40 Mpc/h**

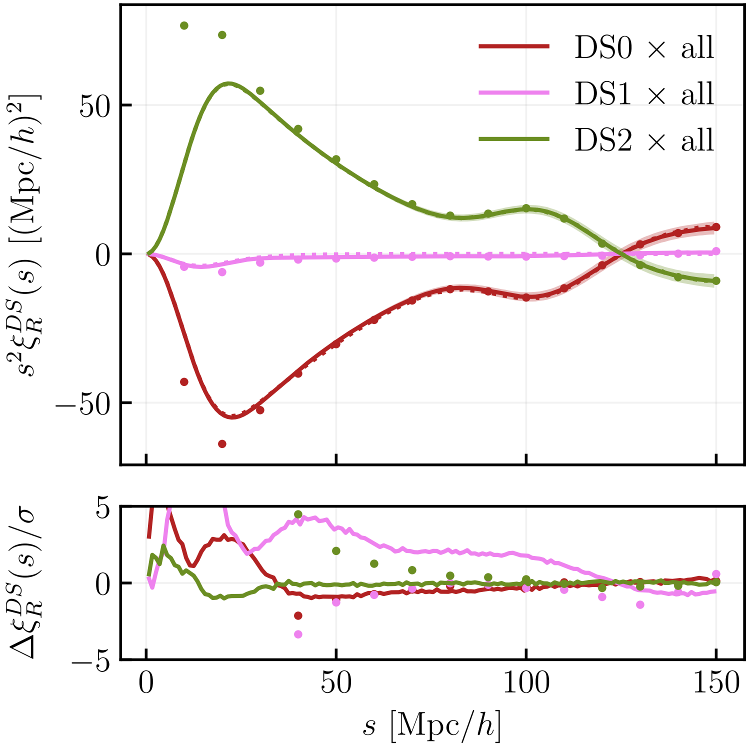
blue: AbacusSummit simulations
red: 2D Gaussian
magenta: Gram-Charlier up to order 7





Gram-Charlier expansion to improve the model?

- Gram-Charlier expansion (dots) does not do better than the lognormal model



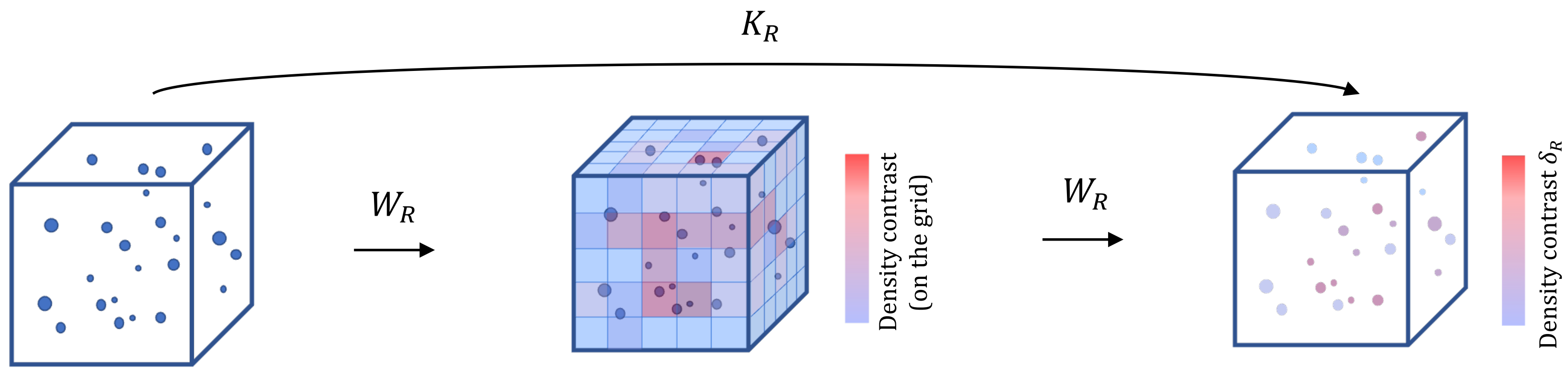
residuals between model and mocks over mocks' standard deviation

Conclusions

- Density-split clustering statistics is a **promising alternative statistics** to **extract information from galaxy surveys** such as DESI
- Previous work obtained cosmological constraints from BOSS using a **simulation-based model**
- An **analytical model** might help us understand what is the **additional physical information** encoded in density-splits statistics compared to standard statistics
- We can predict the density-split correlation from the 2D PDF of $\delta_{R_1}(r), \delta_{R_2}(r+s)$
- Assuming a **lognormal density field** seems reasonable for dark matter in real and redshift space, although not at the level of DESI-like precision
- We can try to expand the density PDF around the lognormal model, e.g. with **Gram-Charlier expansion** (seems not accurate enough) or **normalizing flows**? (work in progress)
- We can try to use results from Large Deviation Theory (e.g. Uhlemann et al. 2016 arXiv:1607.01026, Codis et al. 2016 arXiv:1602.03562)

Back up

Smoothing kernel



Catalog of particles (halos, galaxies…)

Density mesh
Cell size = R

Density is computed for each particle

$$W_R(\mathbf{r}) = W^{\mathrm{TSC}}\left(\frac{r_x}{R}\right) W^{\mathrm{TSC}}\left(\frac{r_y}{R}\right) W^{\mathrm{TSC}}\left(\frac{r_z}{R}\right)$$

$$W^{\mathrm{TSC}}(s) = \begin{cases} \frac{3}{4} - |s|^2 & \text{if} |s| < \frac{1}{2} \\ \frac{1}{2}\left(\frac{3}{2} - |s|^2\right)^2 & \text{if} \frac{1}{2} \leq |s| < \frac{3}{2} \\ 0 & \text{otherwise.} \end{cases}$$

used twice →

Global smoothing kernel:

$$K_R(r, x) = \frac{1}{V} \sum_{i,j,k} W_R(r - r_{i,j,k}) W_R(x - r_{i,j,k})$$

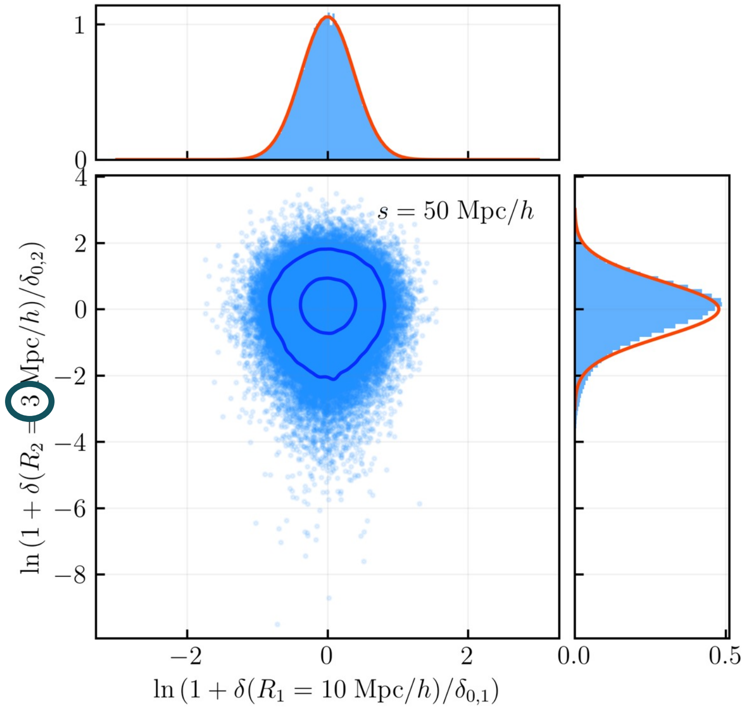
Unsmoothed density splits

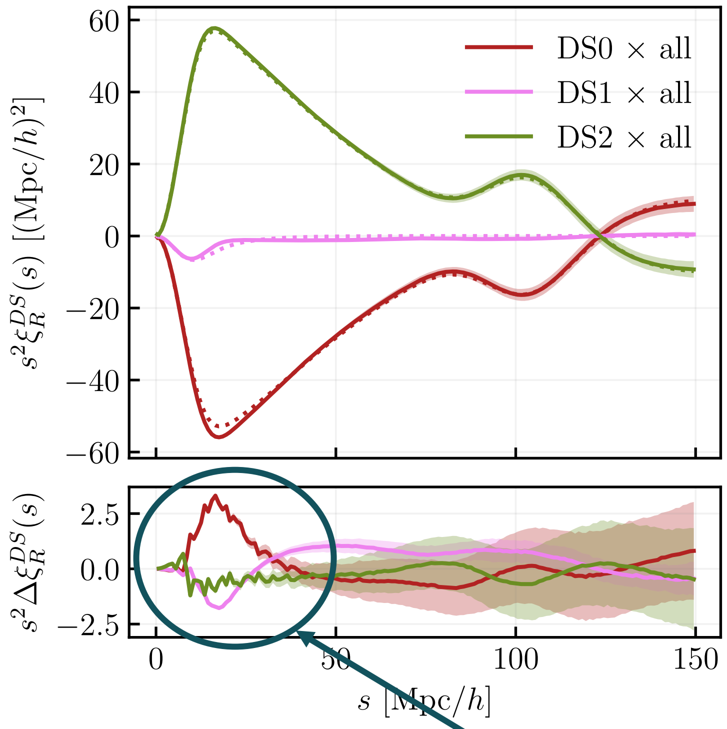
$$\xi^{\mathrm{DS}}_{R_1,R_2}(s) = \frac{\langle n^{\mathrm{DS}}_{R_1}(r) n_{R_2}(r+s)\rangle}{\bar{n}^{\mathrm{DS}}_{R_1}\bar{n}} - 1$$

unsmoothed

$$\xi^{\mathrm{DS}}_{R}(s) = \frac{\langle n^{\mathrm{DS}}_{R}(r) n(r+s)\rangle}{\bar{n}^{\mathrm{DS}}_{R}\bar{n}} - 1$$

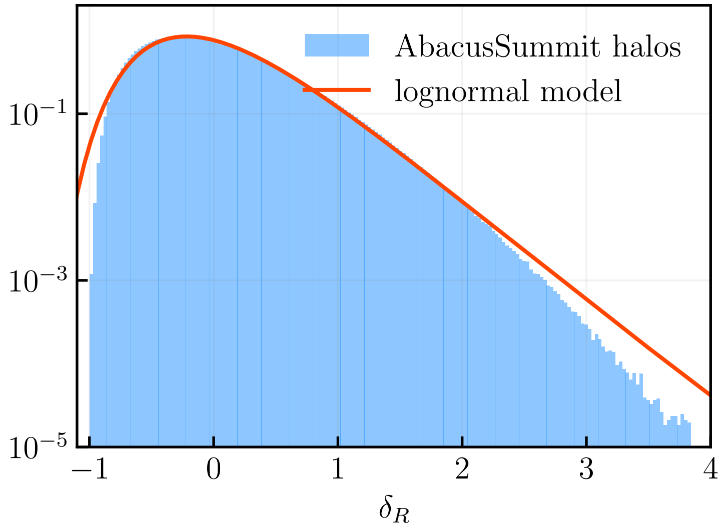
We can't measure the "unsmoothed" density contrast, but as we go to smaller smoothing scales, the density becomes further away to lognormal

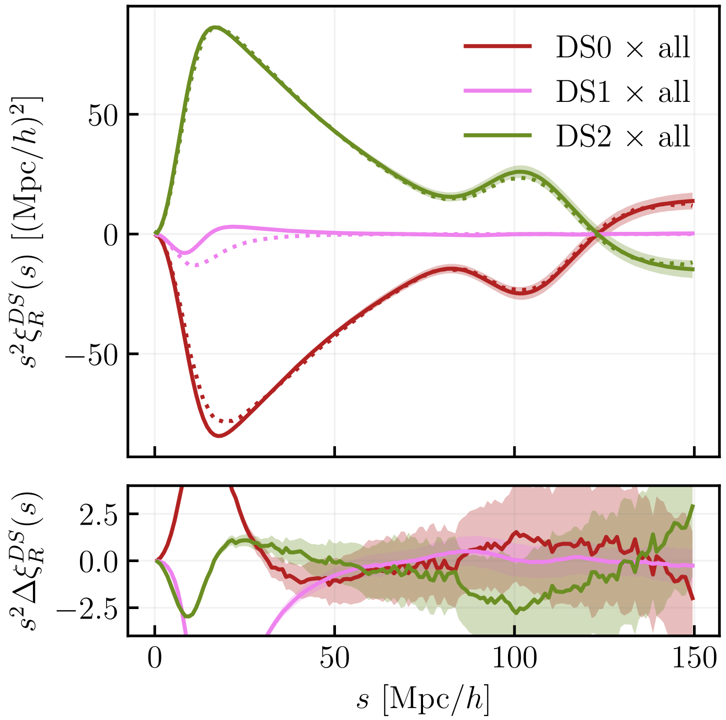




"unsmoothed" lognormal model fails on scales below ~40 Mpc/h

Lognormal model compared to dark matter halos (real space)





breaks down below ~40 h/Mpc
(unsmoothed)