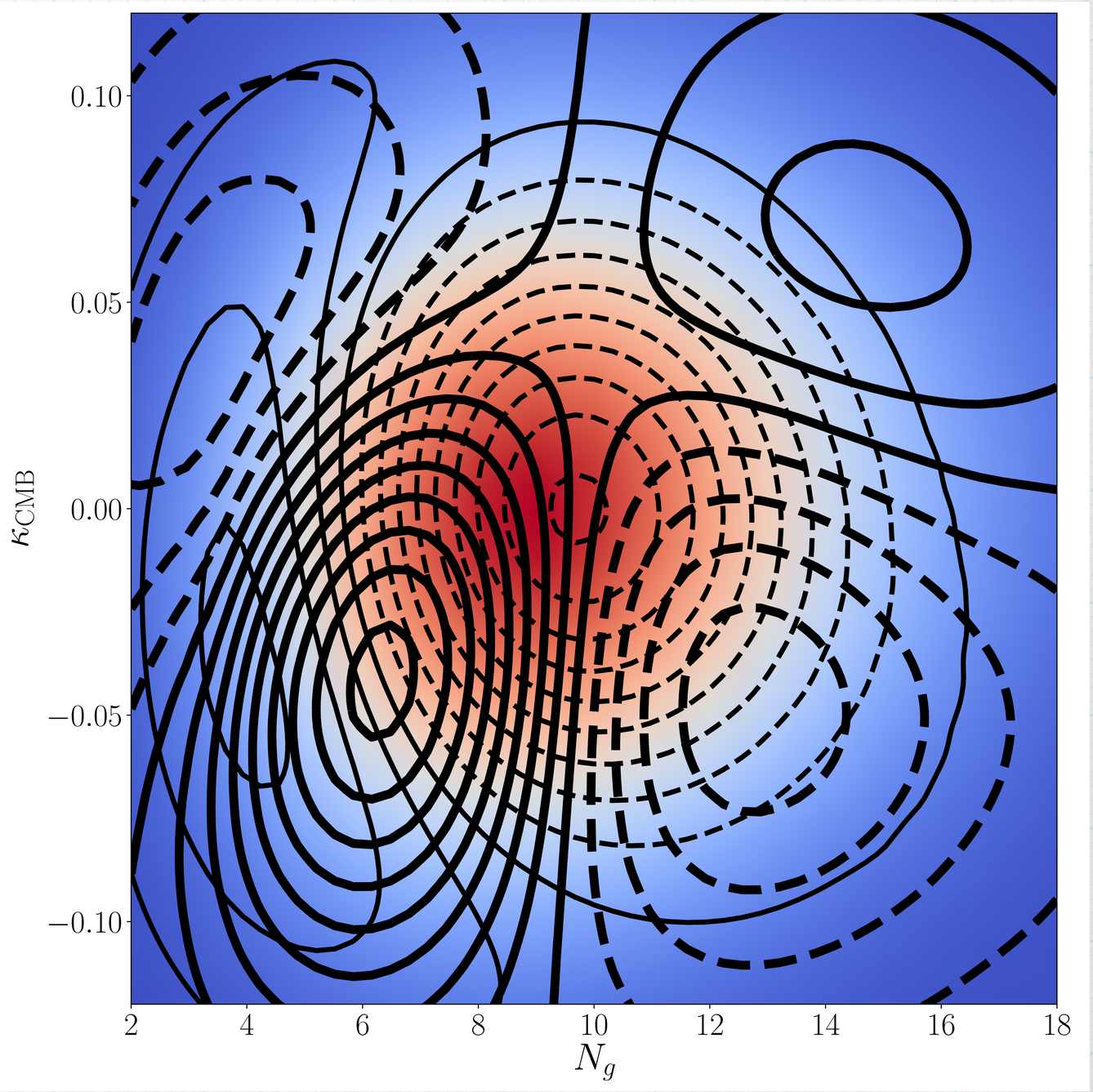
Analysing the PDF of density fluctuations - can it work in real data?

Focus week
“Interpretable higher-order statistics”
Trieste, 30 June 2022

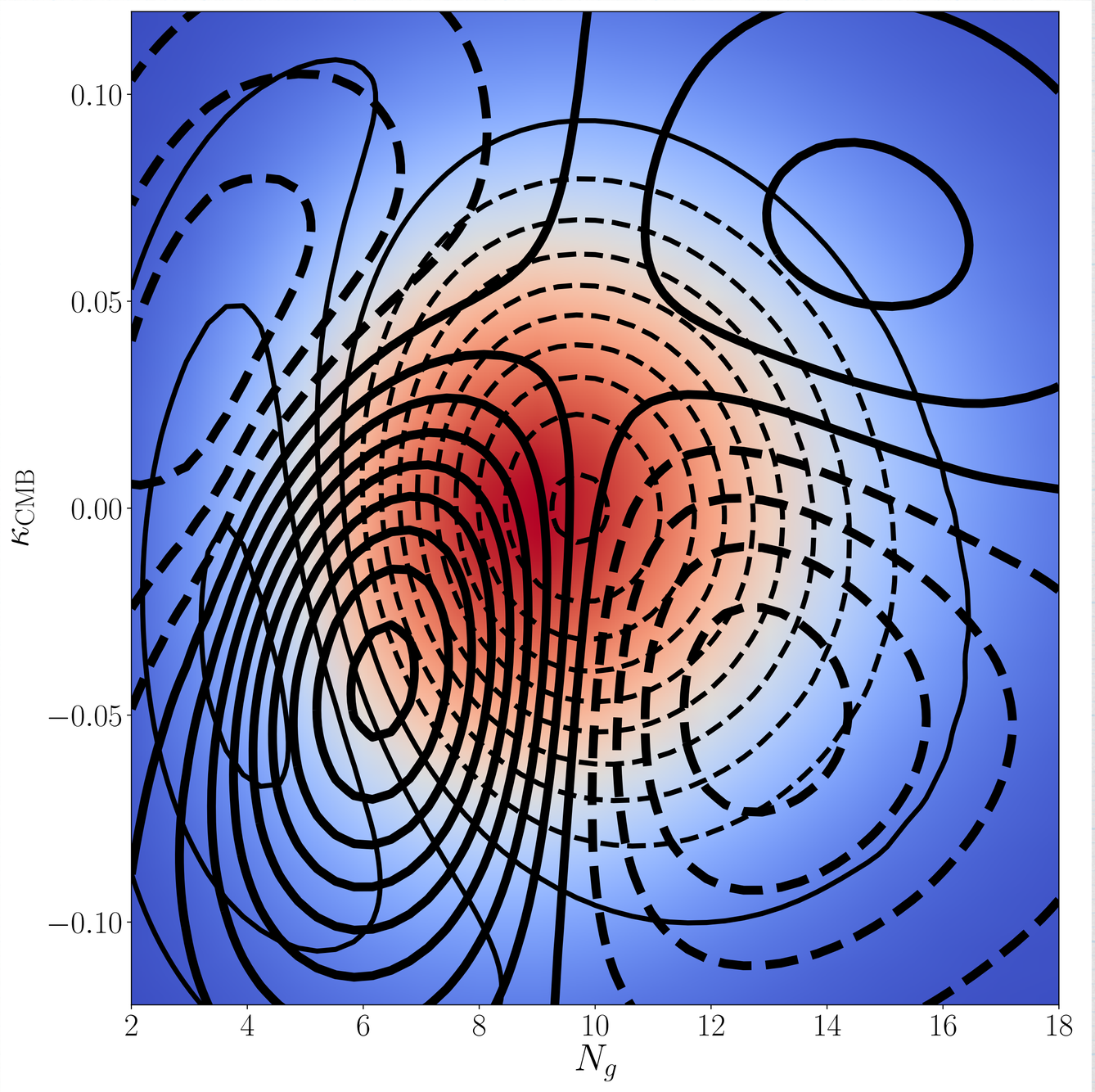
Oliver Friedrich
LMU Munich

work with Anik Halder, Cora Uhlemann, Aoife Boyle, Alexandre Barthelemy, Sandrine Codis, Daniel Gruen, Joe DeRose, Elisabeth Krause, Tobias Baldauf, Francisco Villaescusa-Navarro, Marc Manera, Takahiro Nishimichi and more!



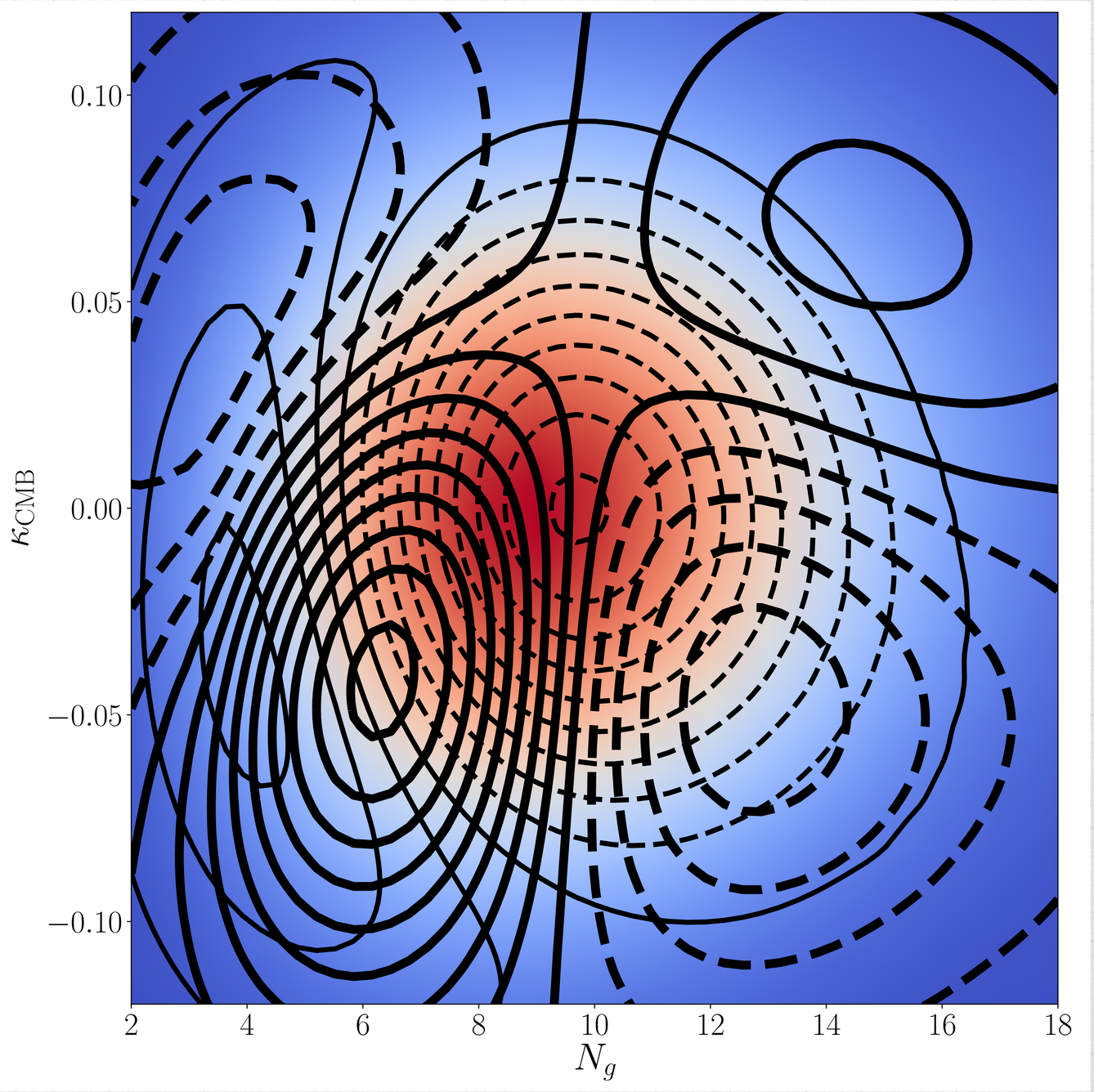
Analysing the PDF of density fluctuations - can it work in real data?

- Why the PDF?
- PDF vs. P(k) — the details
- Let's **interpret** the PDF
- My view of PDF theory

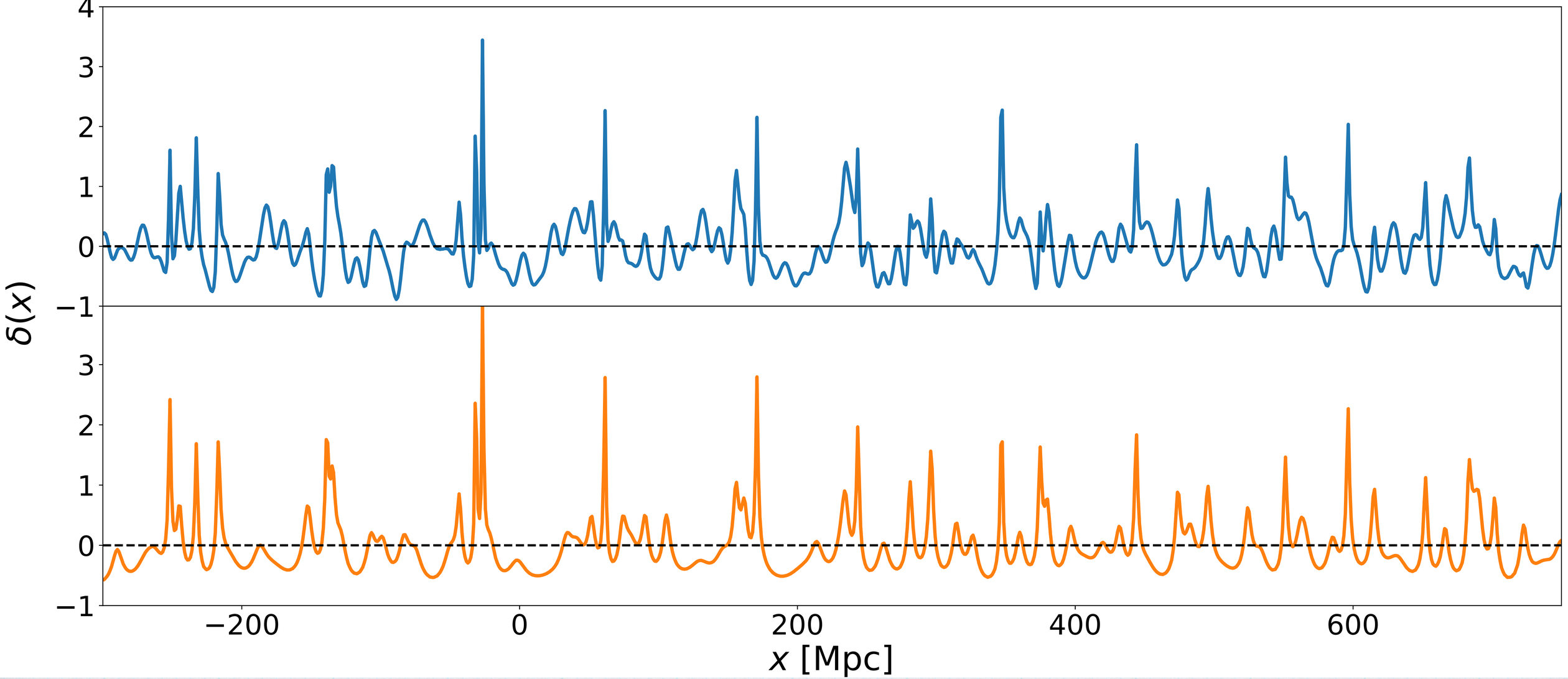


Analysing the PDF of density fluctuations - can it work in real data?

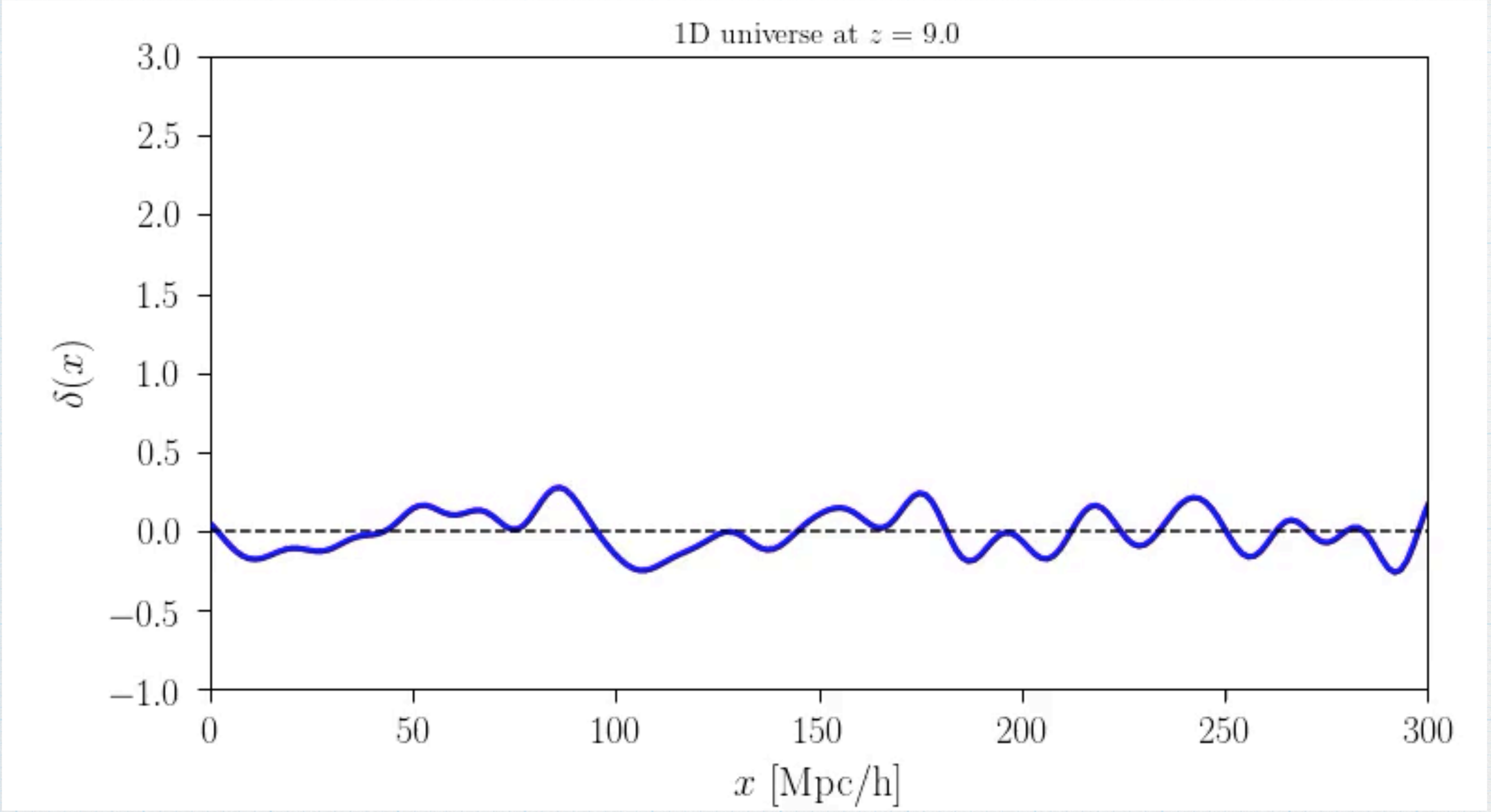
- **Why the PDF?**
- PDF vs. P(k) — the details
- Let's **interpret** the PDF
- My view of PDF theory

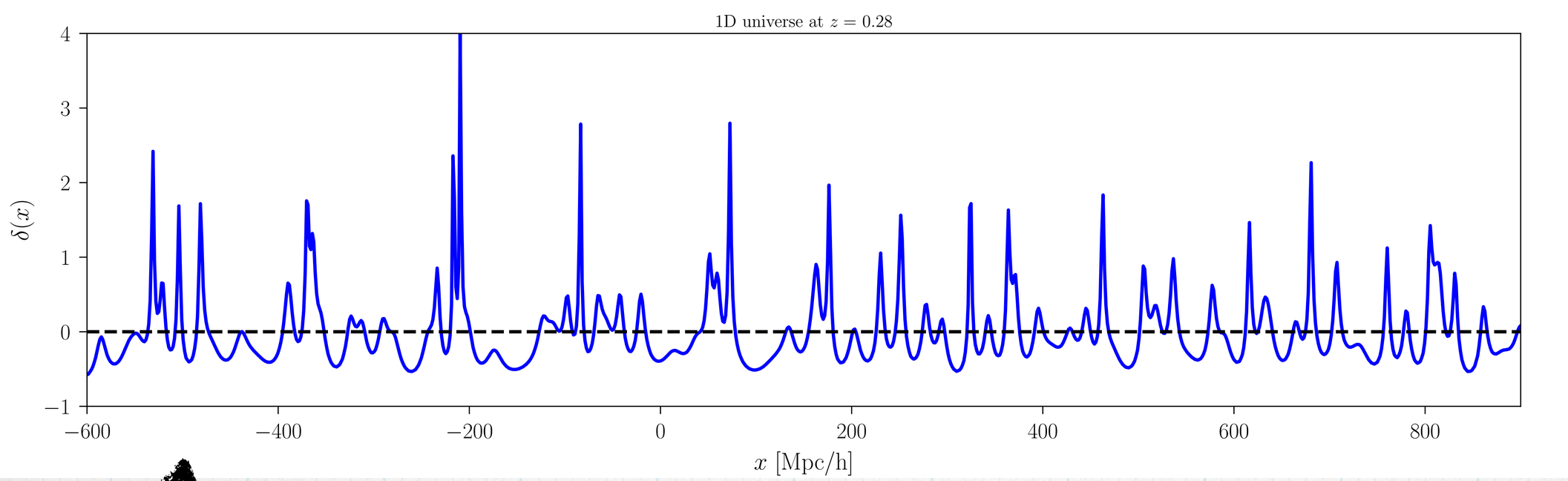


two “density fields” - which one is physical?

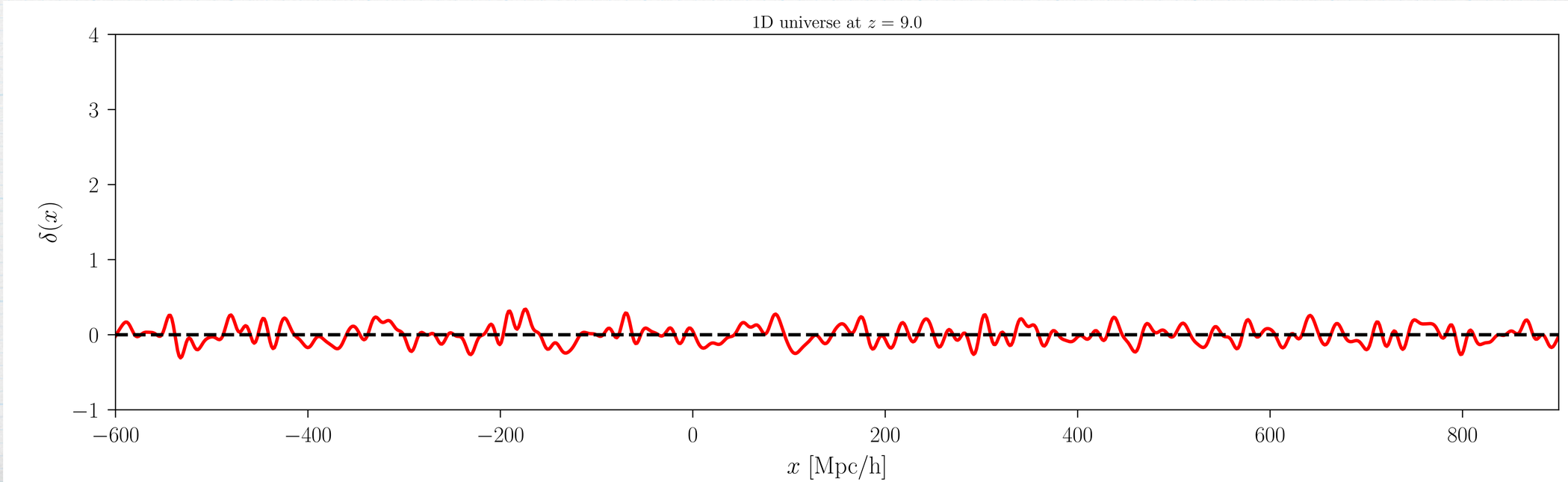


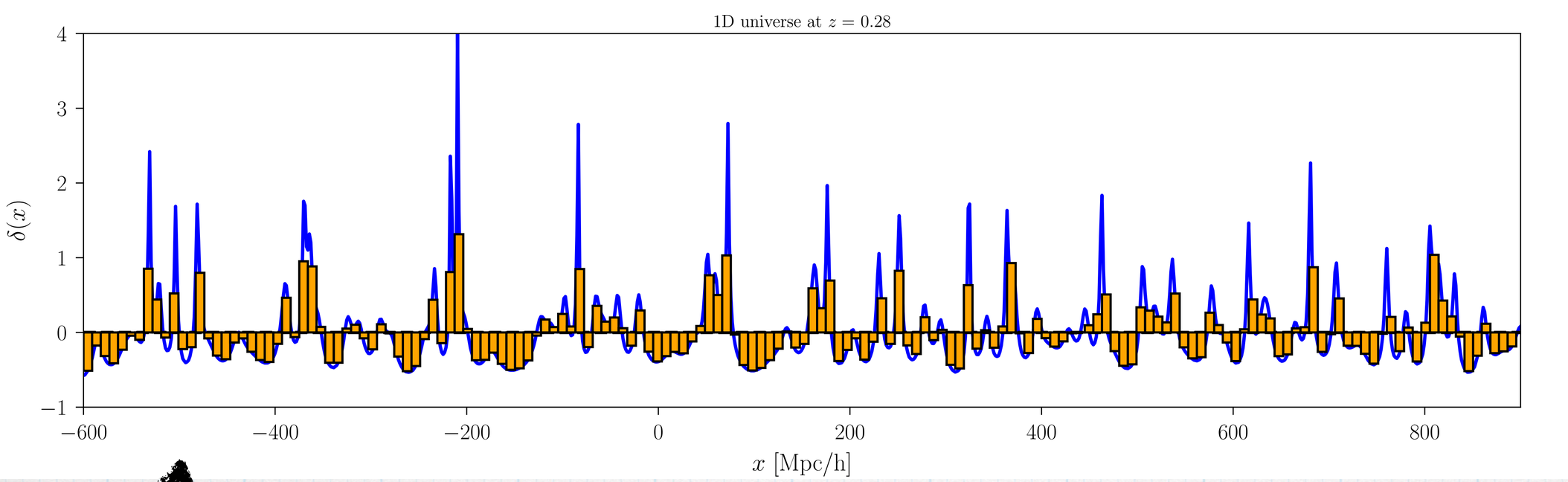
$$\delta_m = (\rho_m - \bar{\rho}_m)/\bar{\rho}_m$$


1D universe at $z = 9.0$
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
$\delta(x)$
0
50
100
150
200
250
300
x [Mpc/h]

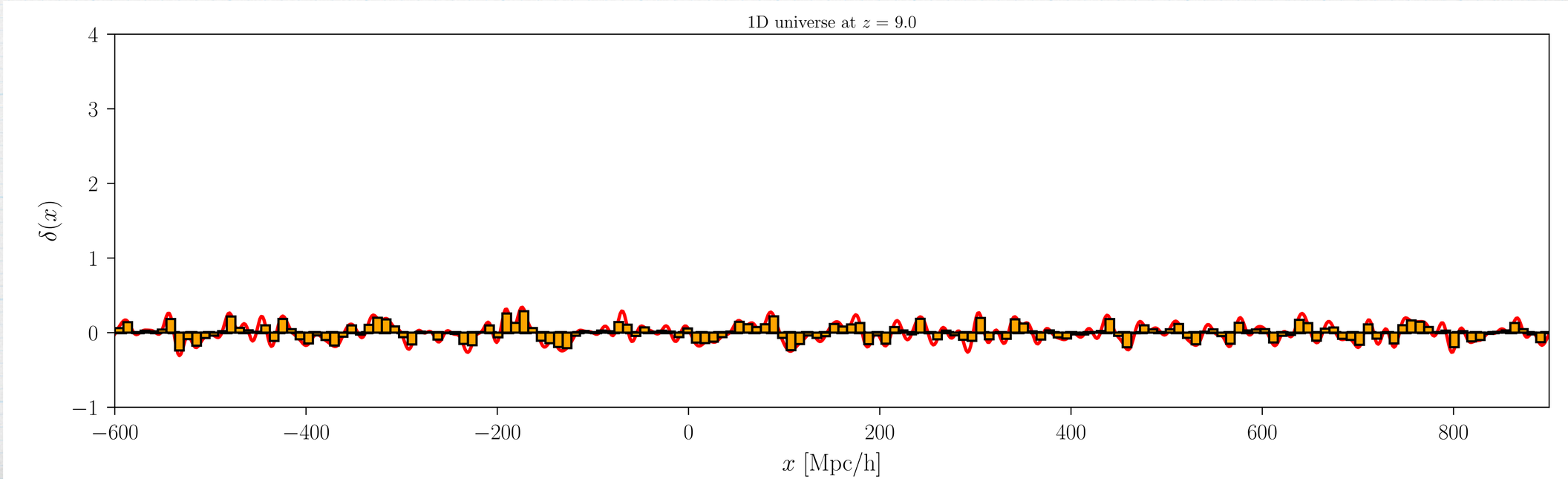


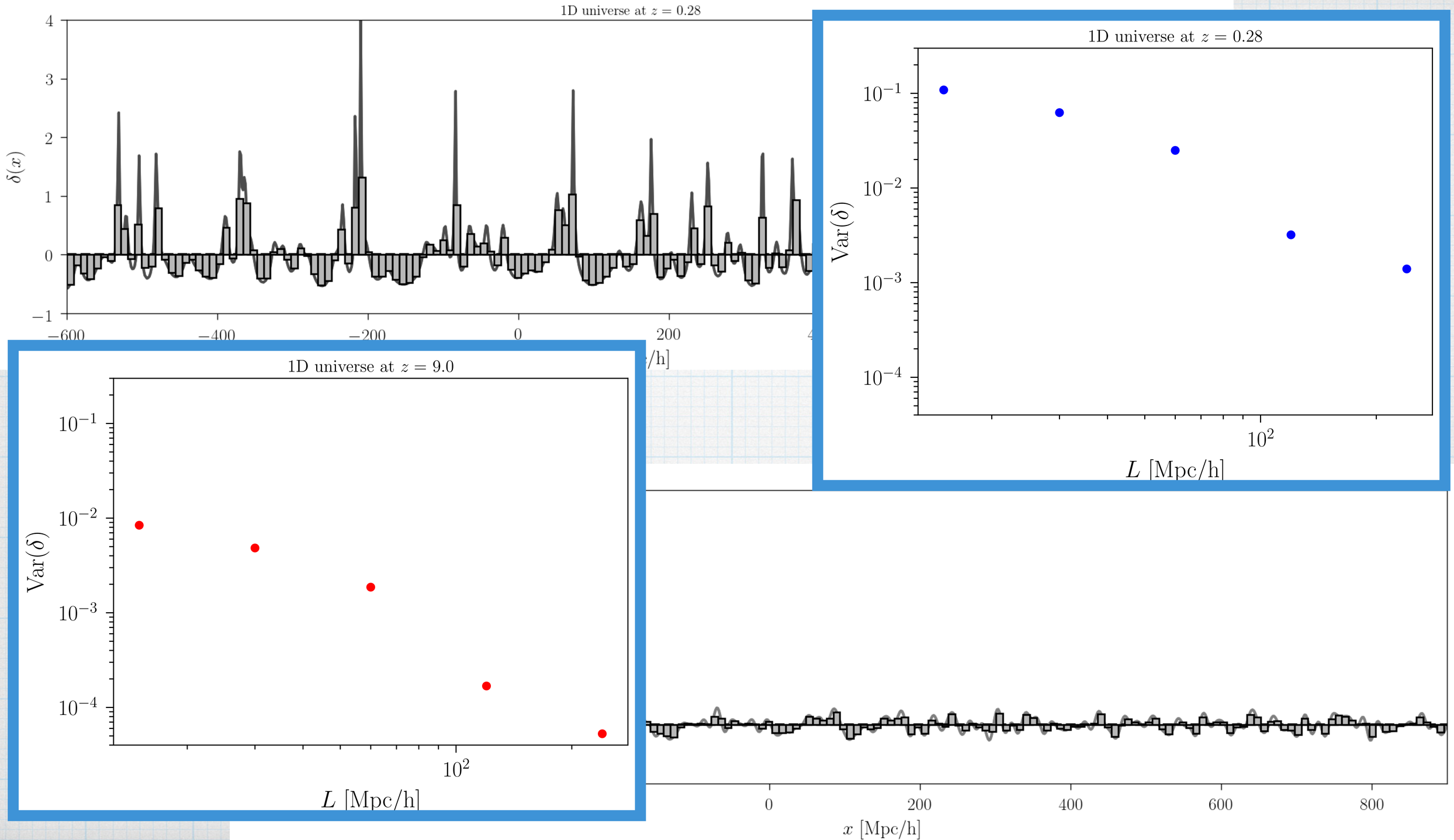
gravitational collapse

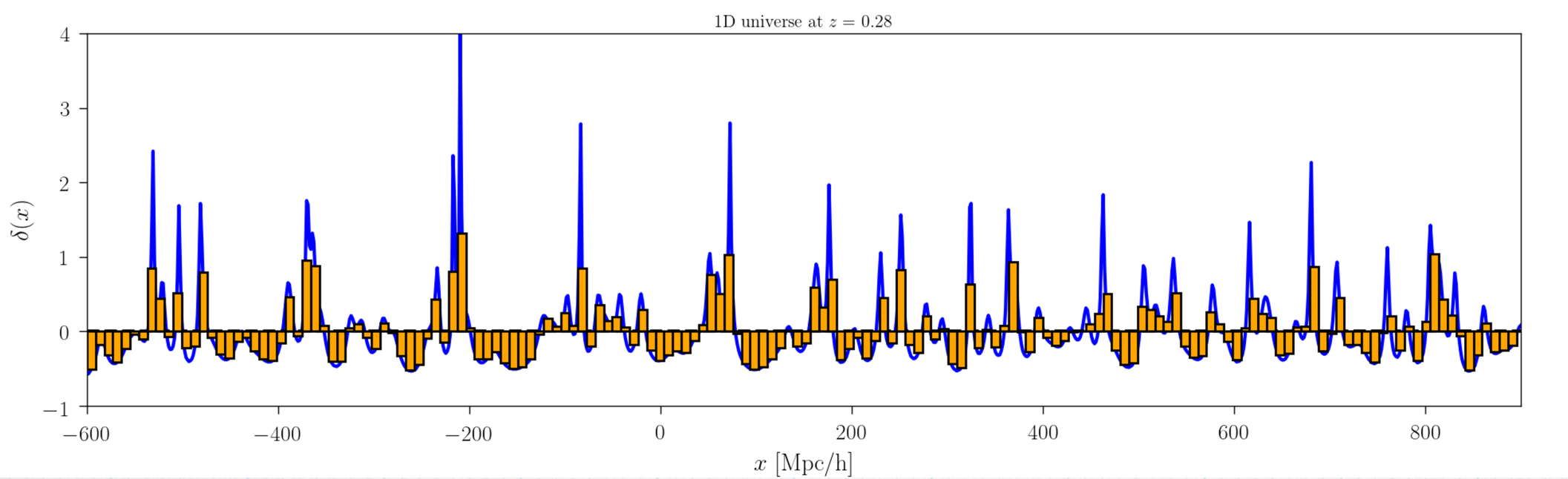


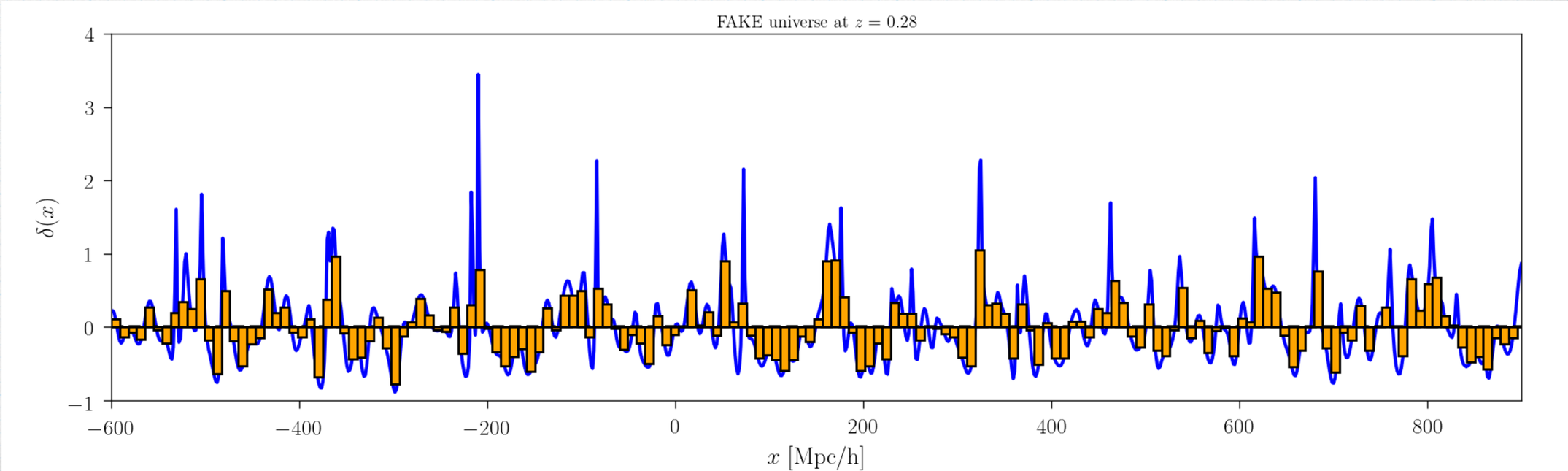


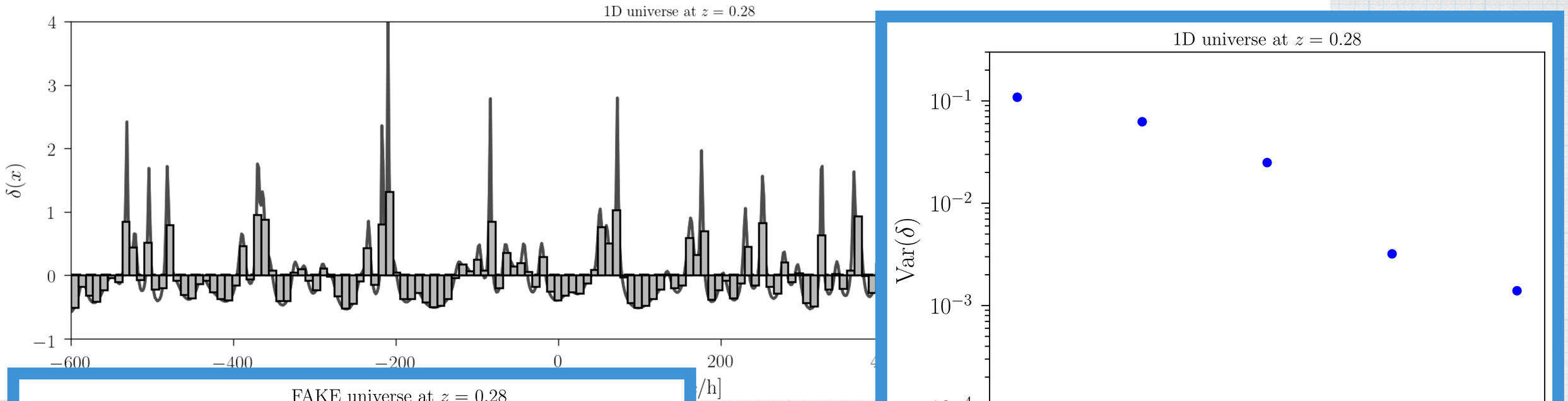
gravitational collapse

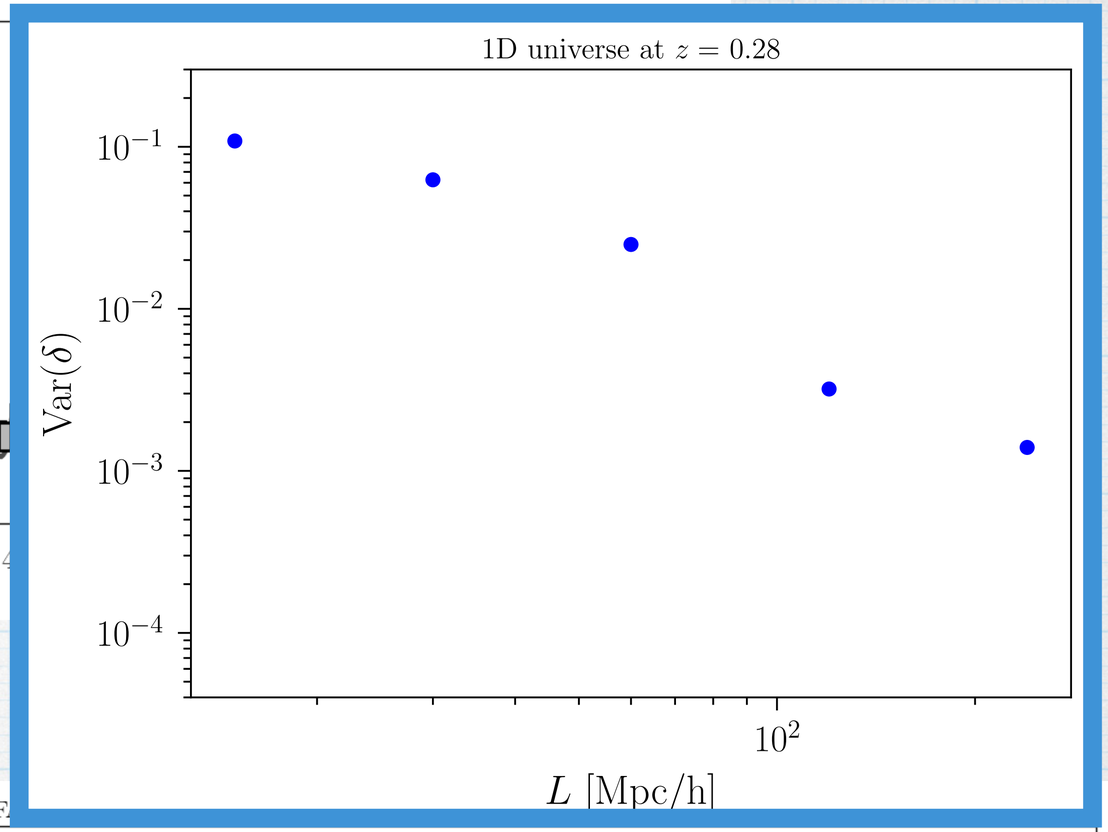


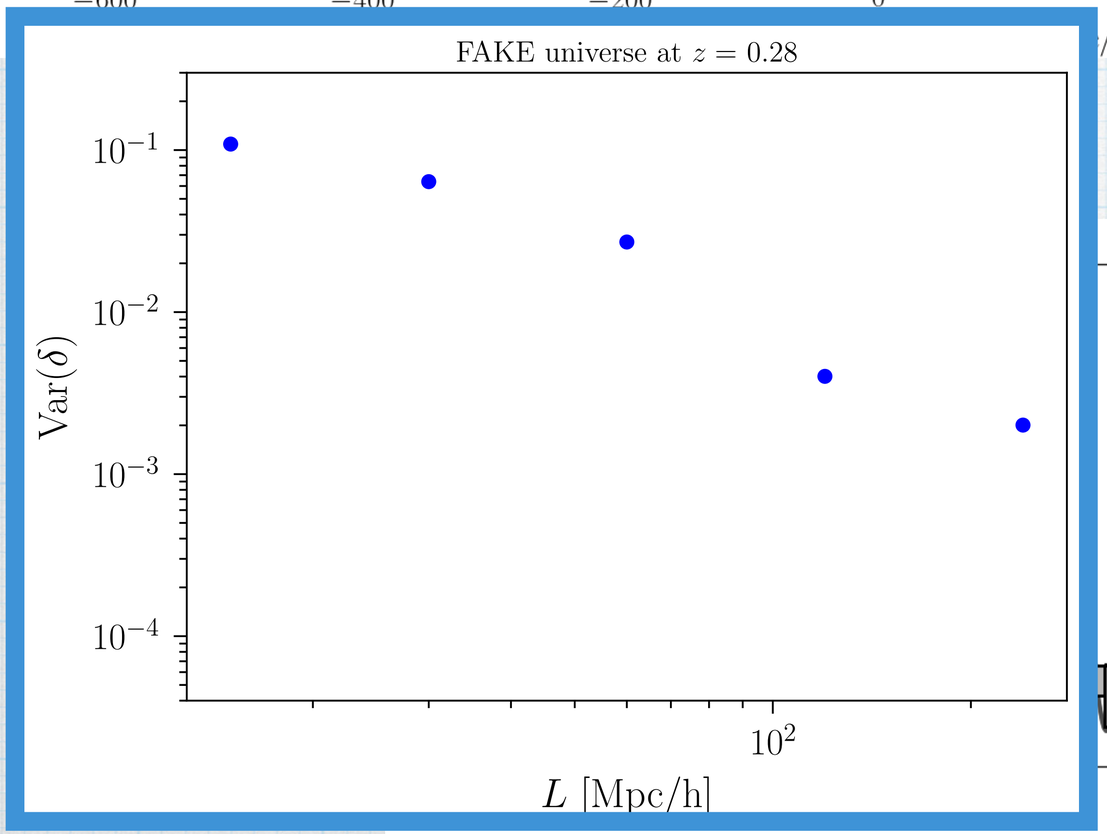

1D universe at $z = 0.28$
$\delta(x)$
4
3
2
1
0
−1
−600
−400
−200
0
200
/h]
1D universe at $z = 0.28$
$\mathrm{Var}(\delta)$
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{2}
L [Mpc/h]
1D universe at $z = 9.0$
$\mathrm{Var}(\delta)$
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{2}
L [Mpc/h]
0
200
400
600
800
x [Mpc/h]

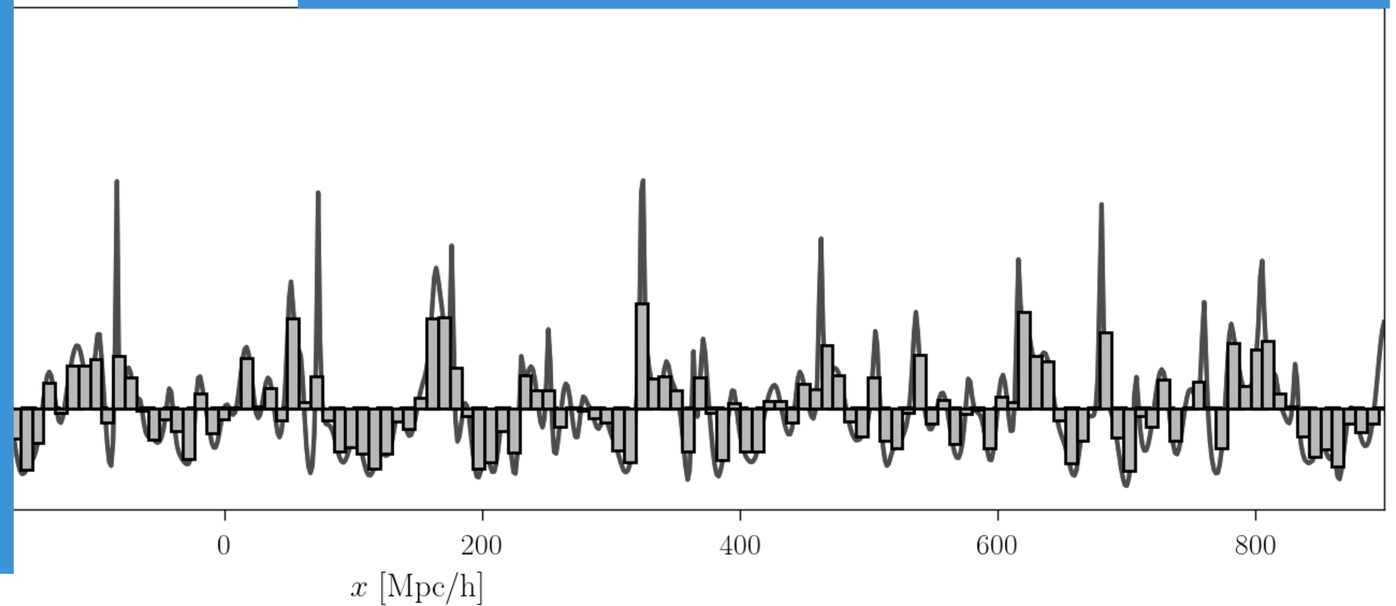

1D universe at $z = 0.28$
$\delta(x)$
x [Mpc/h]
4
3
2
1
0
−1
−600
−400
−200
0
200
400
600
800

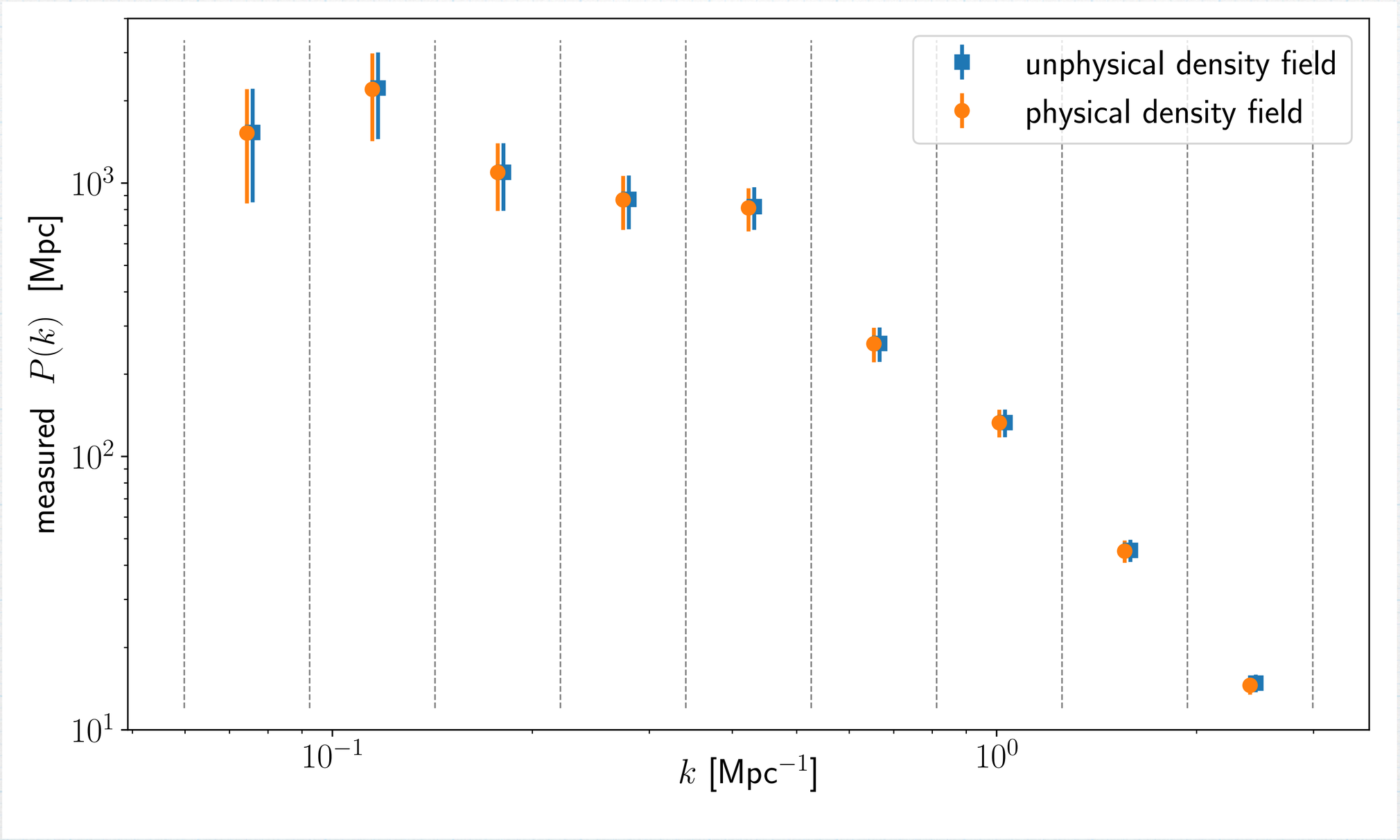

FAKE universe at $z = 0.28$
$\delta(x)$
x [Mpc/h]
4
3
2
1
0
−1
−600
−400
−200
0
200
400
600
800

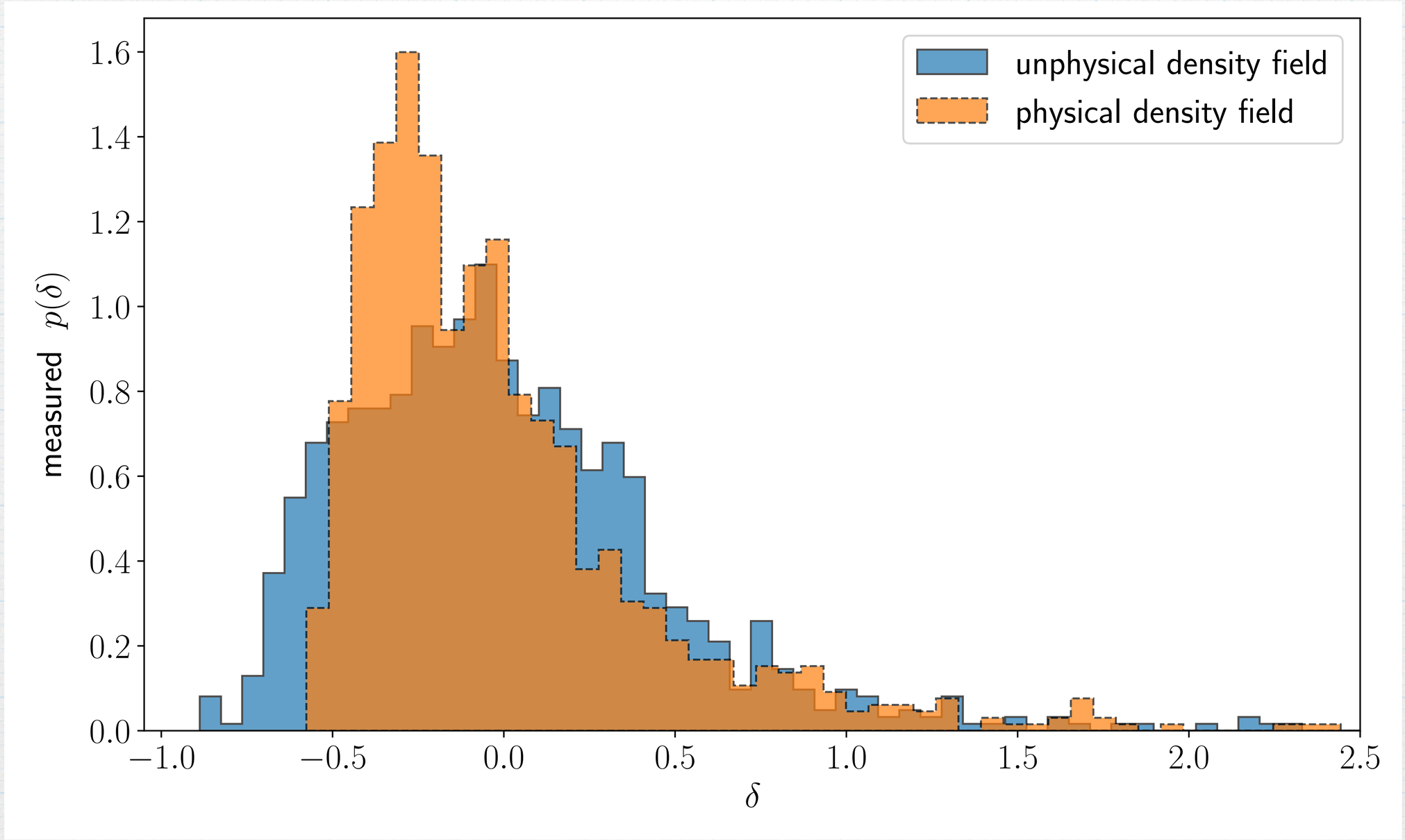

1D universe at $z = 0.28$
$\delta(x)$
4
3
2
1
0
−1
−600
−400
−200
0
200
/h]


1D universe at $z = 0.28$
$\mathrm{Var}(\delta)$
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{2}
L [Mpc/h]

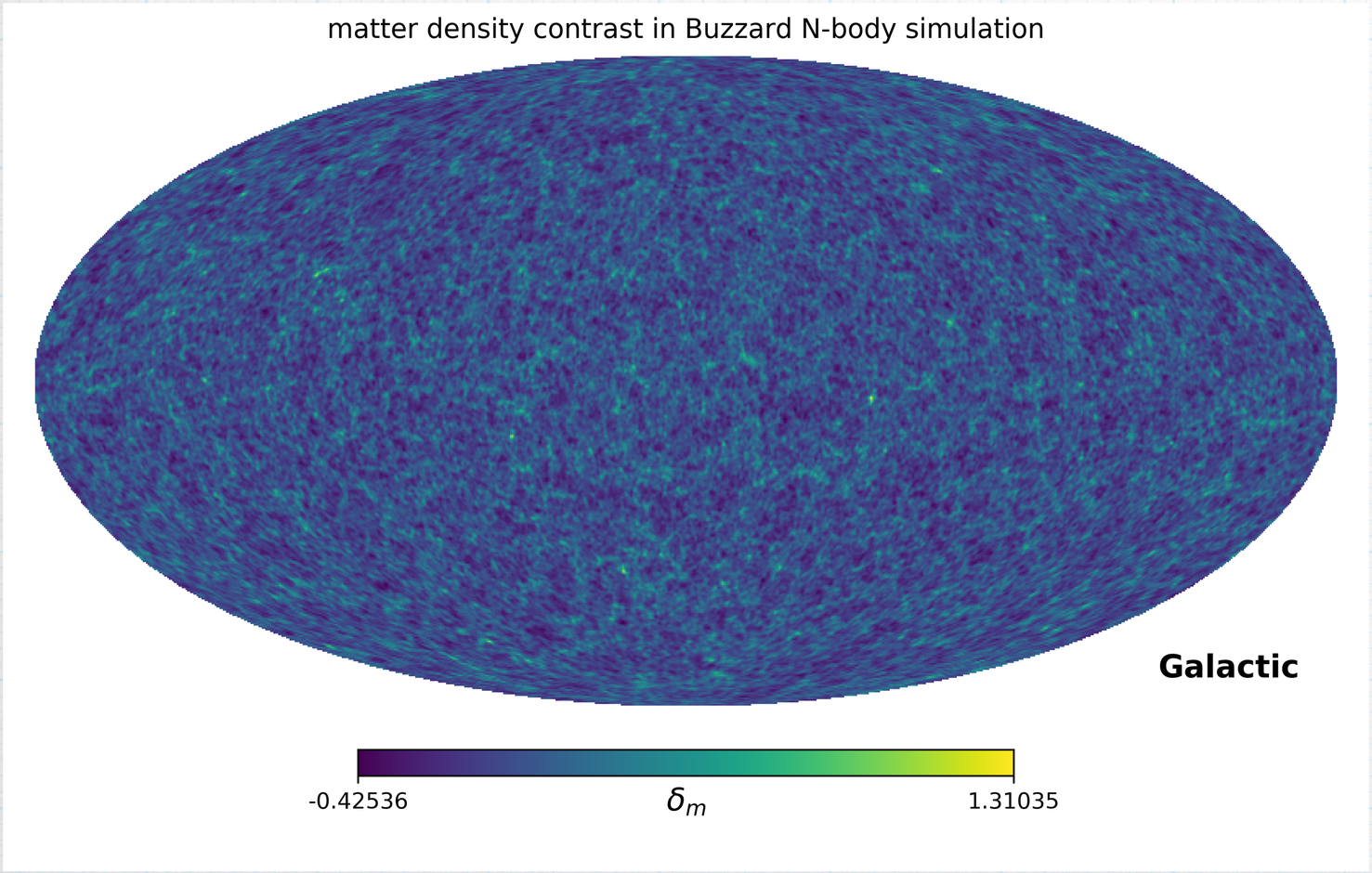

FAKE universe at $z = 0.28$
$\mathrm{Var}(\delta)$
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{2}
L [Mpc/h]

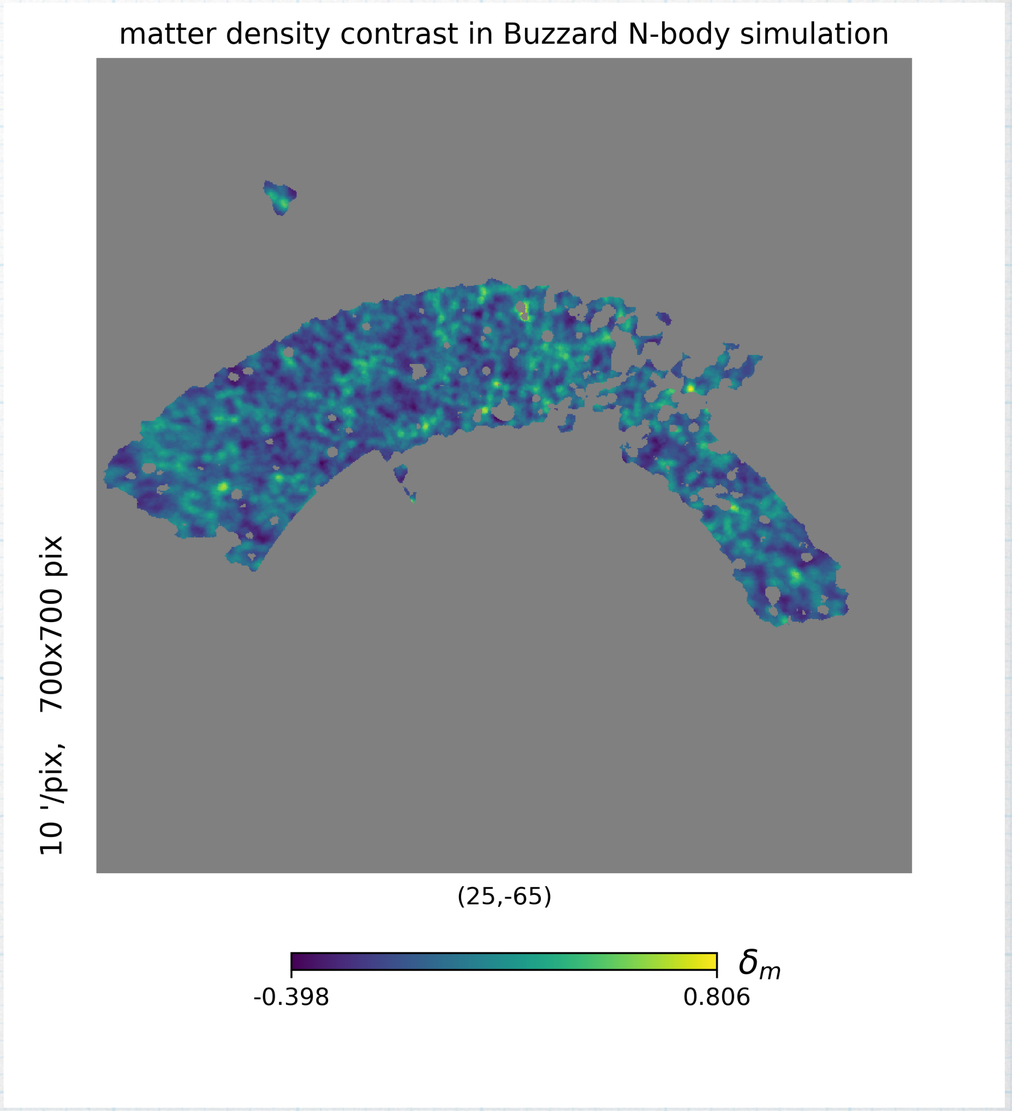

0
200
400
600
800
x [Mpc/h]


unphysical density field
physical density field
measured $P(k)$ [Mpc]
10^3
10^2
10^1
10^{-1}
10^0
k [Mpc^{-1}]


unphysical density field
physical density field
measured $p(\delta)$
δ
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5

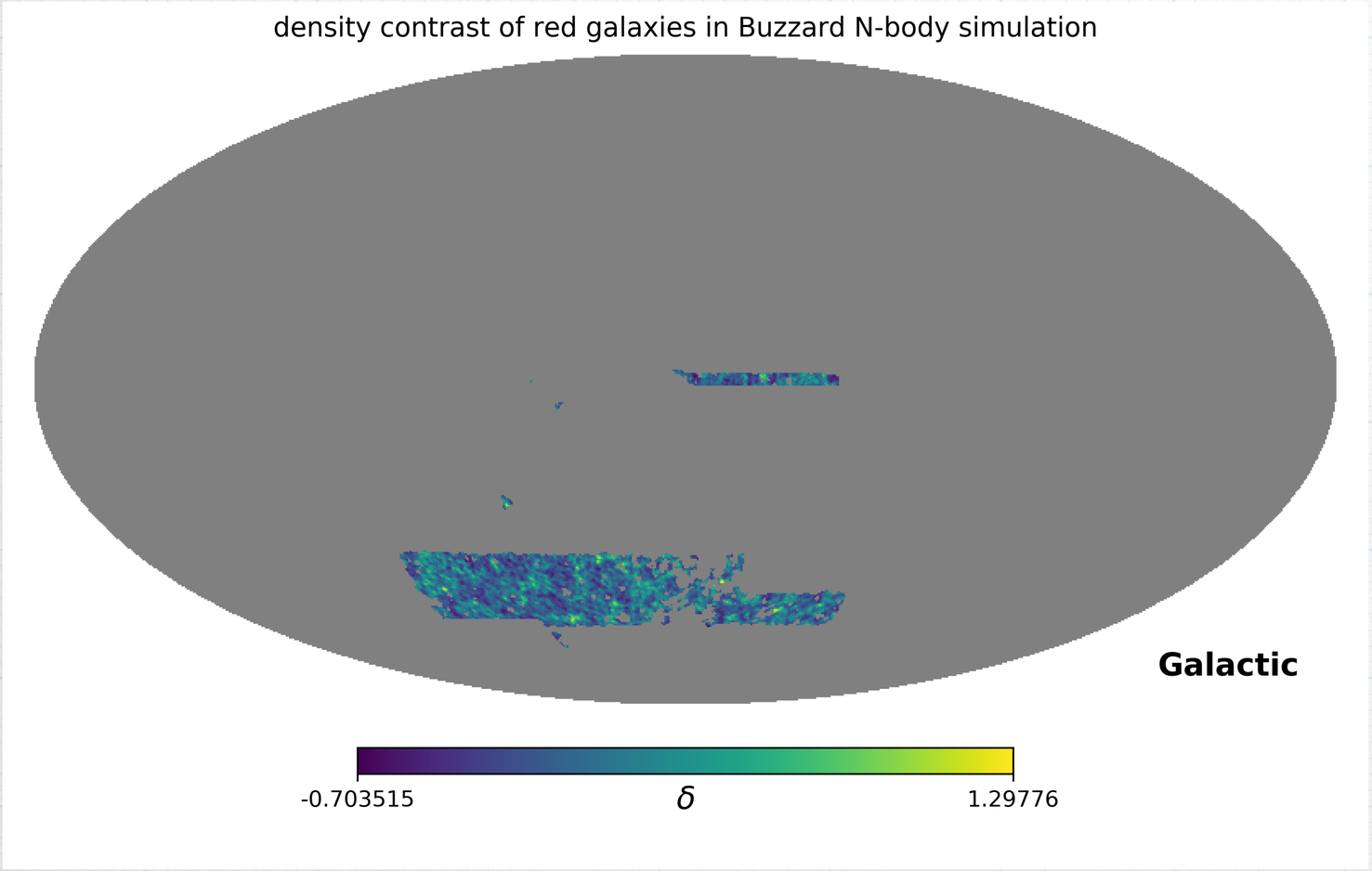
Matter density contrast in Buzzard N-body simulation:

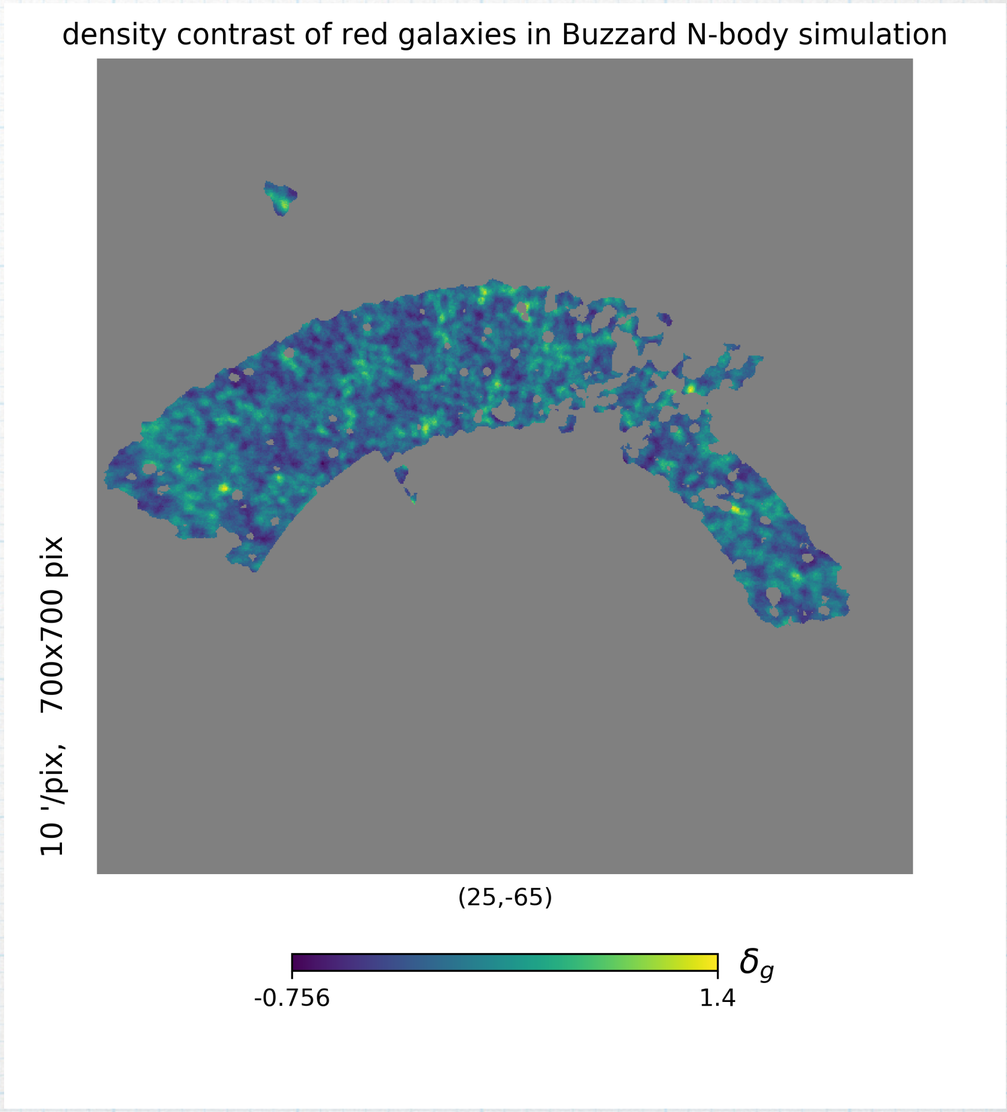




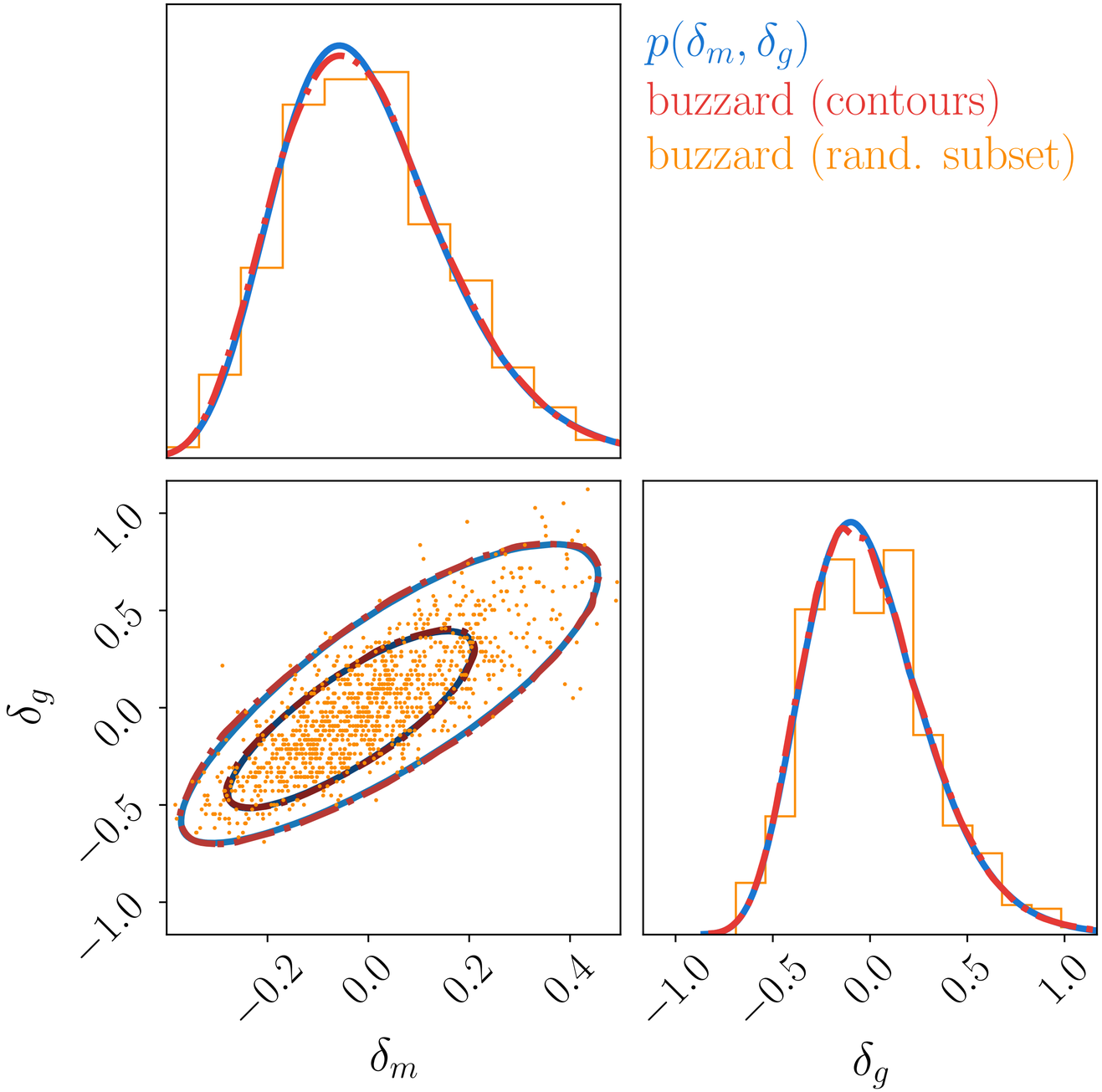
Y1 Buzzard Flock, DeRose++ (2019)

Galaxy density contrast in Buzzard N-body simulation:





Y1 Buzzard Flock, DeRose++ (2019)



galaxies are biased & stochastic tracers of matter density

2-point statistics only extracts three numbers from this!

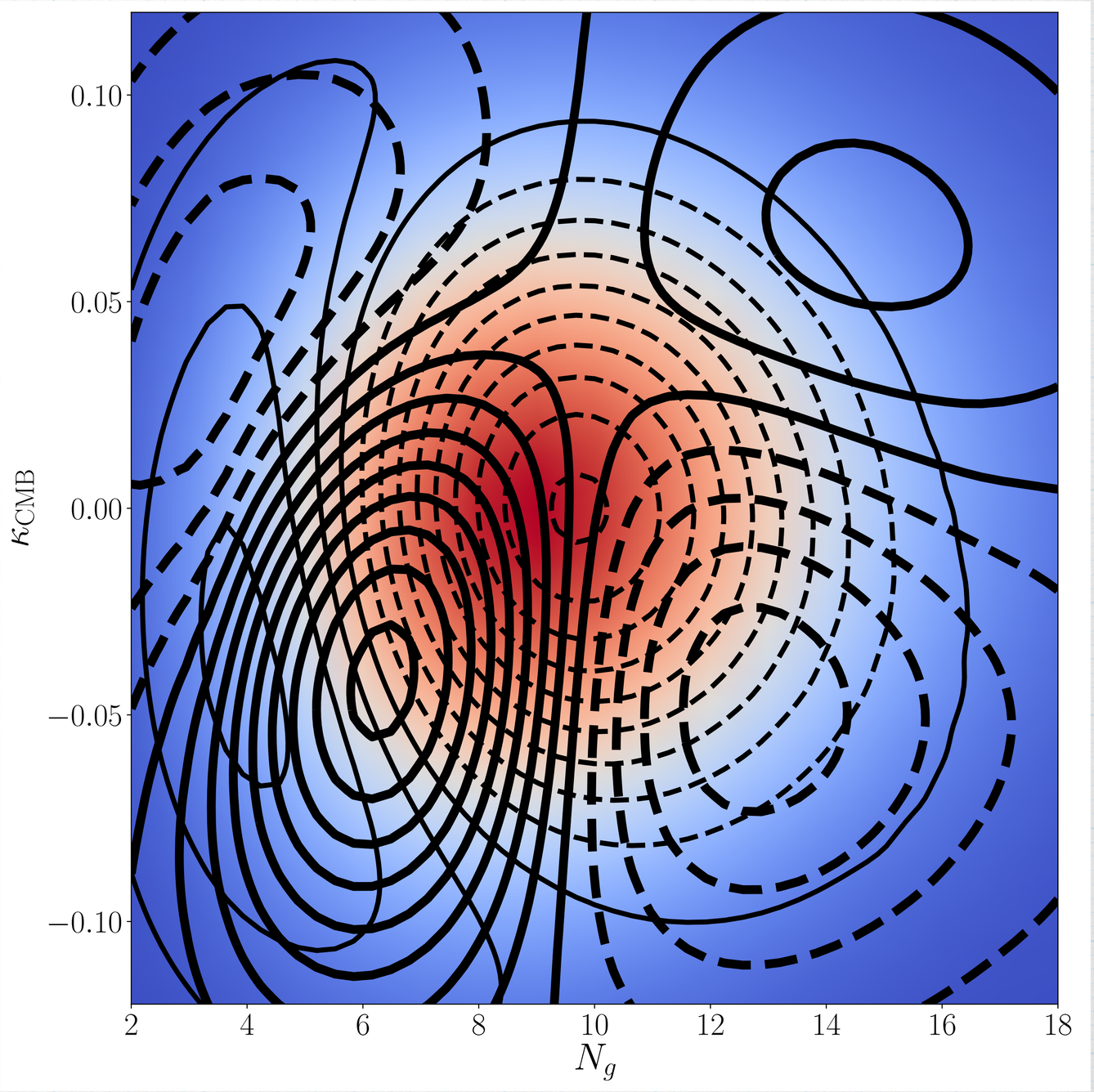
→ way to do better: analyse full shape of the PDF

Blue: model predictions calculated in **Friedrich++(2018)**
Red: data from Buzzard N-body sims (DeRose, Wechsler++2019)

$$\delta_m = (\rho_m - \bar{\rho}_m)/\bar{\rho}_m$$
$$\delta_g = (\rho_g - \bar{\rho}_g)/\bar{\rho}_g$$

The full shape of the density PDF: A new dimension for studies of the large-scale structure

- Why the PDF?
- **PDF vs. P(k) — the details**
- Let's **interpret** the PDF
- My view of PDF theory



PDF / cumulants

- Cumulant generating function (CGF):

$$\varphi_R(\lambda) = \sum_n \langle \delta_R^n \rangle_c \frac{\lambda^n}{n!}$$

$\delta_R =$ density contrast smoothed by radius R

$\langle \delta_R^n \rangle_c =$ Cumulant (part of the nth moment that vanishes for Gaussian field)

2-point statistics

- Power spectrum:

$$\langle \delta_{\mathbf{k}} \delta_{\mathbf{q}} \rangle \sim P(k)\ \delta_D(\mathbf{k} + \mathbf{q})$$

PDF / cumulants

- Actual observable:

PDF $p(\delta_R)$

related to CGF via Laplace transform

$$e^{\varphi_R(\lambda)} = \int \mathrm{d}\delta_R \; p(\delta_R) \; e^{\lambda \delta_R}$$

2-point statistics

- Alternative observable:

2-point function $\xi(r) \equiv \langle \delta(\mathbf{x}) \delta(\mathbf{x}+\mathbf{r}) \rangle$

related to power spectrum via Fourier transform:

$$P(|\mathbf{k}|) \sim \int \mathrm{d}^3 r \; \xi(|\mathbf{r}|) \; e^{-i\mathbf{r}\mathbf{k}}$$

PDF / cumulants

- Leading order perturbation theory
  Friedrich++ 2020, arxiv.org/abs/1912.06621 :

$\varphi_R(\lambda) \approx$ minimum of function

$$s_\lambda(\delta_i, j) = -\lambda F(\delta_i) + j\delta_i + \varphi_{R(1+F(\delta_i))^{1/3},init}(j)$$

Spherical or cylindrical collapse of an initial fluctuation

Initial CGF at initial radius

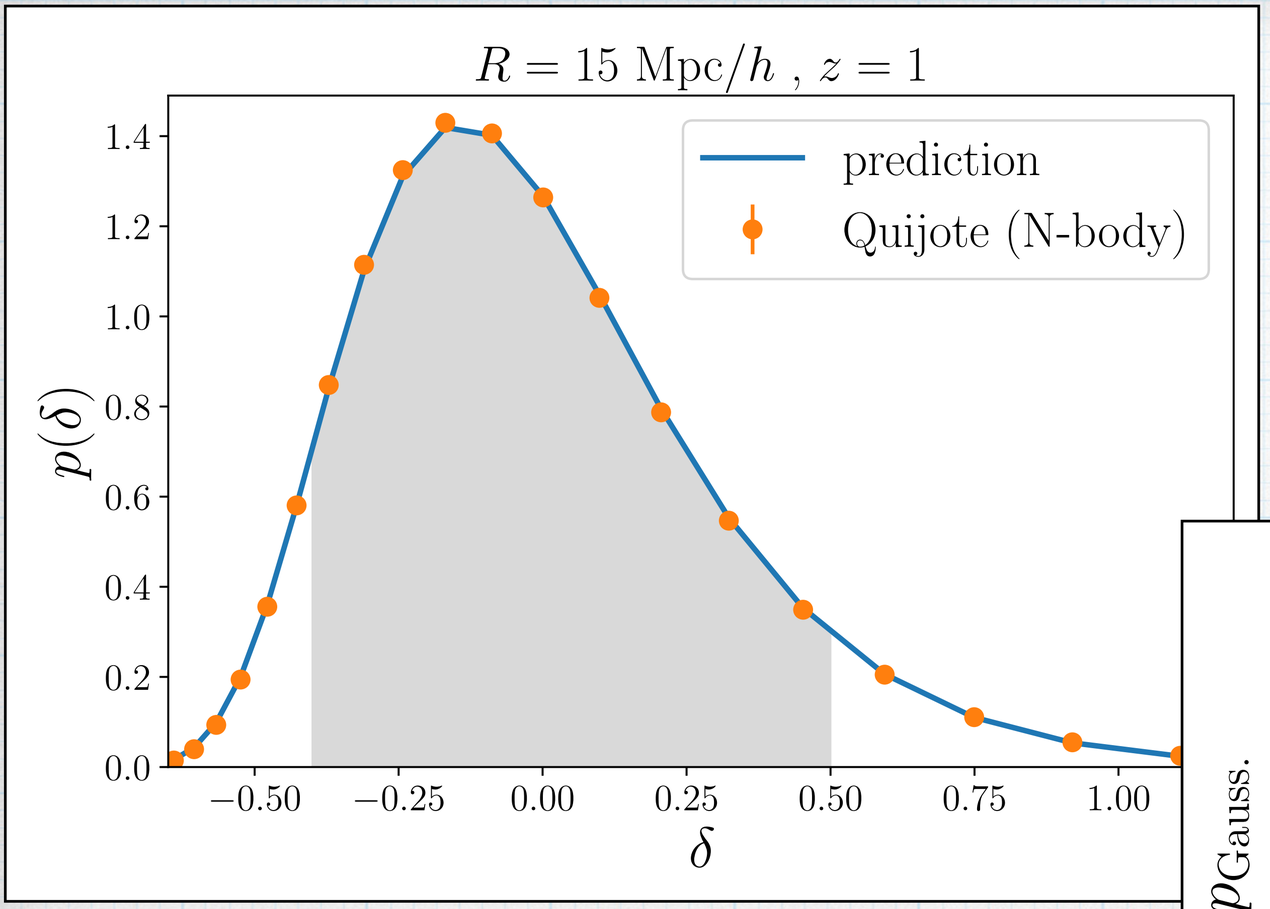
"step 1: Legendre transform initial CGF"
"step 2: Legendre trans. back, but wrt transformed variable

See also:
Valageas (2002), Bernardeau et al. (2015)
***Uhlemann et al. (2017)**, Friedrich et al. (2018)*
Ivanov et al. (2019) … and many more!

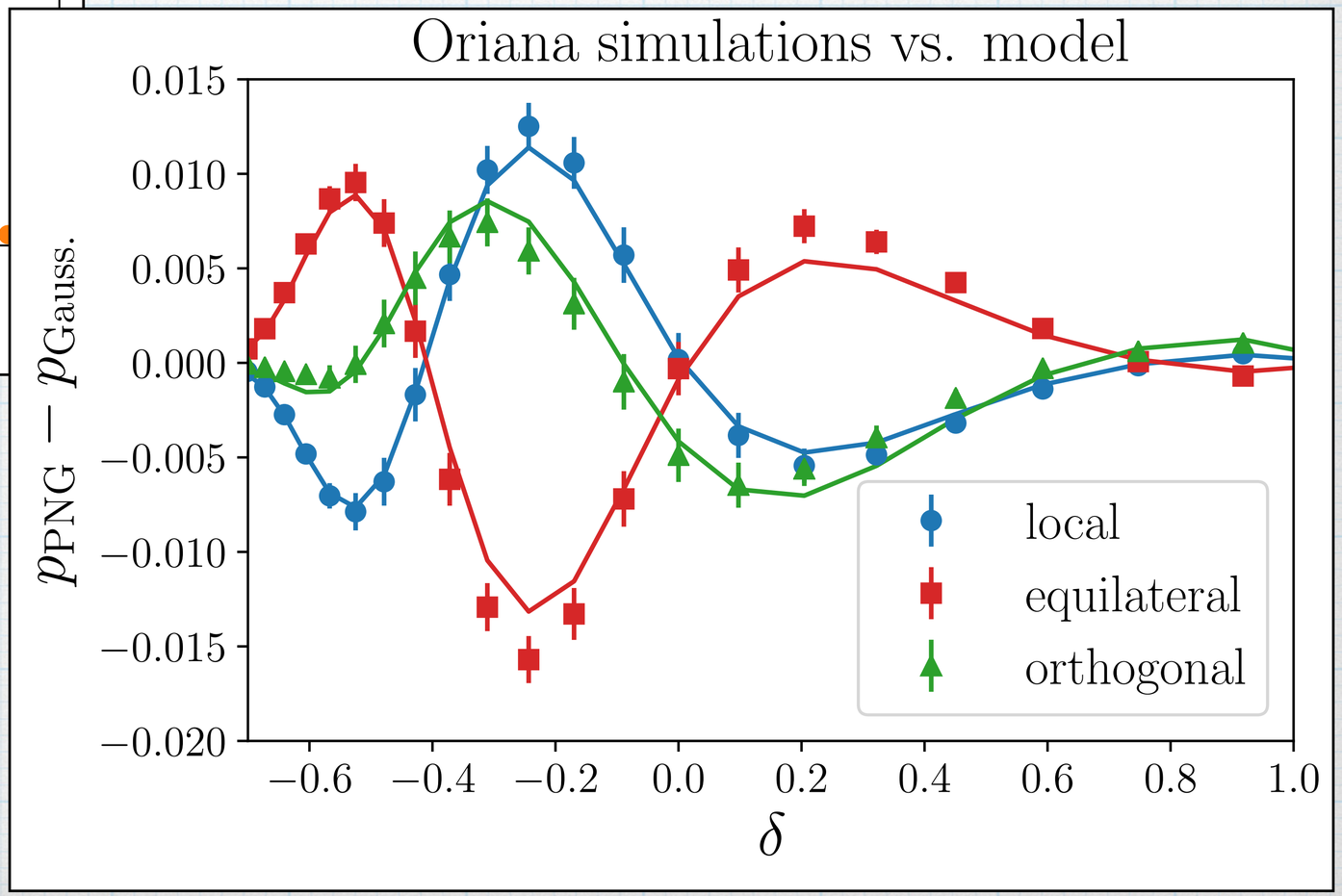
2-point statistics

- Leading order perturbation theory

$$P(k,t) = P(k,t_i)\,\frac{D^2(t)}{D^2(t_i)}$$



Theory prediction
(Friedrich et al. 2020)
compared to the PDF measured in the Quijote N-body simulations
(Villaescusa-Navarro et al. 2020)



Prediction for impact of primordial non-Gaussianity in the PDF
(Friedrich et al. 2020)
compared to Oriana N-body simulations **(Mao et al. 2014)**

PDF / cumulants

- line-of-sight projections:

$$\varphi_{q,\theta}(\lambda) \approx \int \mathrm{d}w \lim_{L\to\infty} \frac{\varphi_{\mathrm{cy},w\theta,L}\left(q(w)L\lambda, w\right)}{L}$$

CGF of cylindrical filter

2-point statistics

- line-of-sight projections:

$$P_q(\ell) \approx \int \mathrm{d}w \left(\frac{q(w)}{w}\right)^2 P_{\mathrm{3D}}\left(\frac{\ell+\frac{1}{2}}{w}, \eta_0 - w\right)$$

PDF / cumulants

- Joint PDF of galaxies and matter:

$$p(\delta_m, \delta_g) = p(\delta_m)\; p(\delta_g \,|\, \delta_m)$$

Sofar mostly:

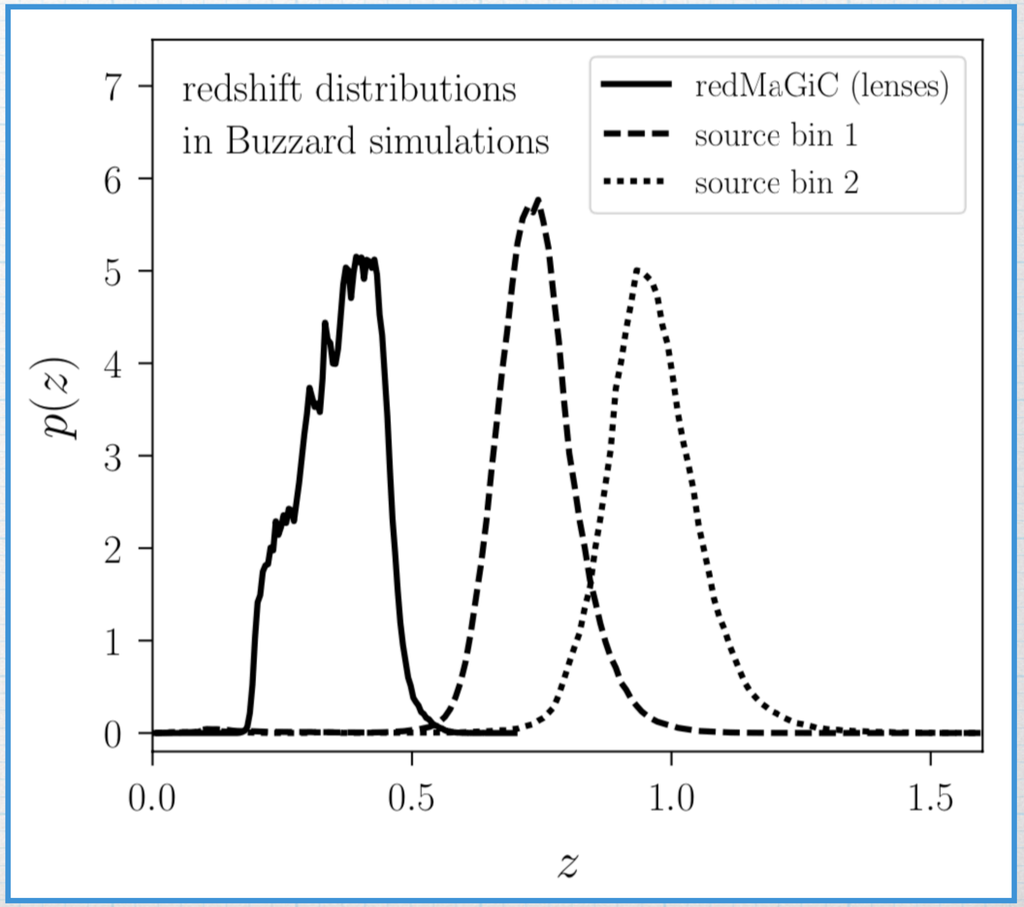
$$p(\delta_g \,|\, \delta_m) = \text{Poisson}(\langle \delta_g \,|\, \delta_m \rangle)$$

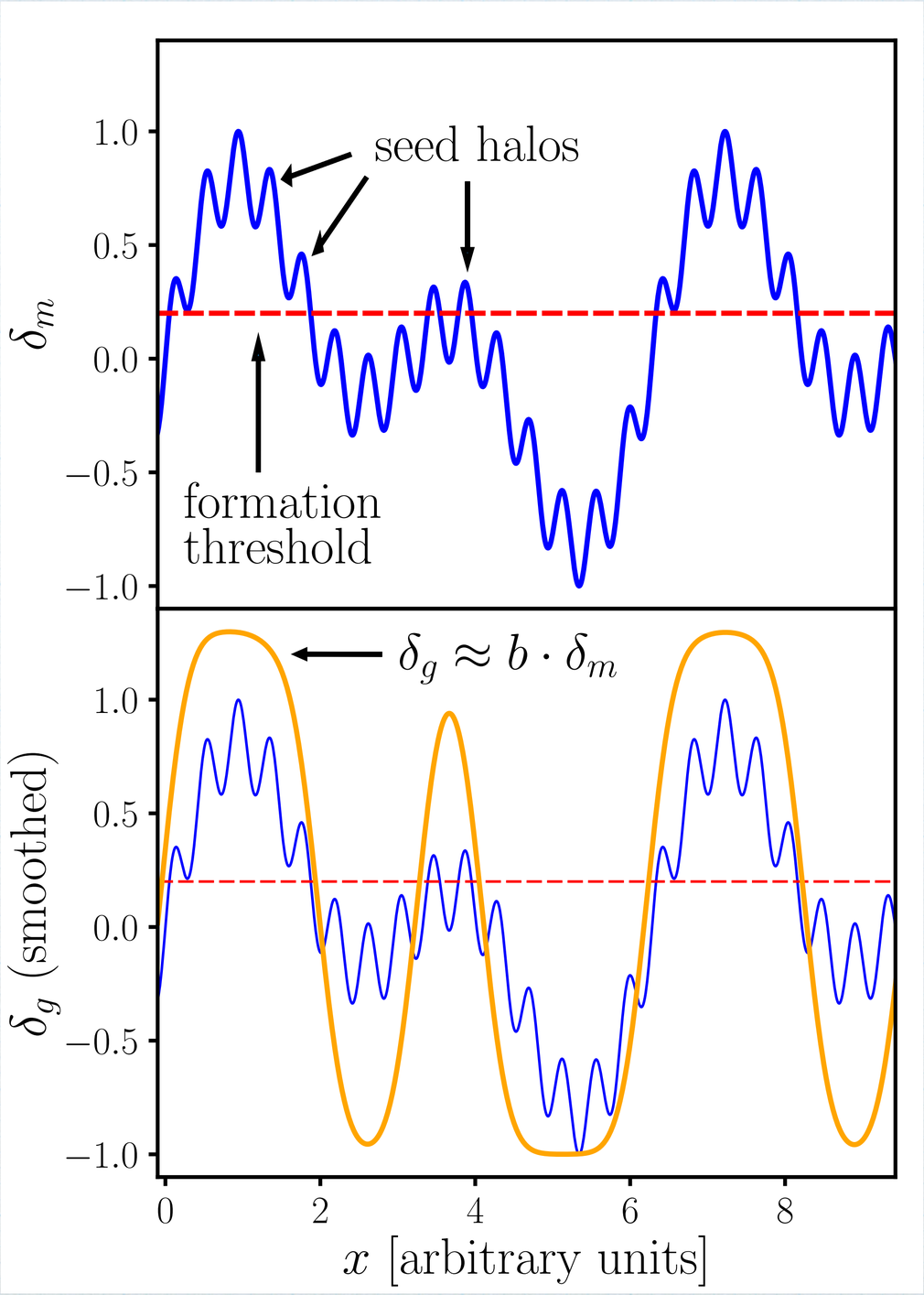
$$\langle \delta_g \,|\, \delta_m \rangle \approx b_1^E \delta_m + \frac{b_2^E}{2}(\delta_m^2 - \langle \delta_m^2 \rangle)$$

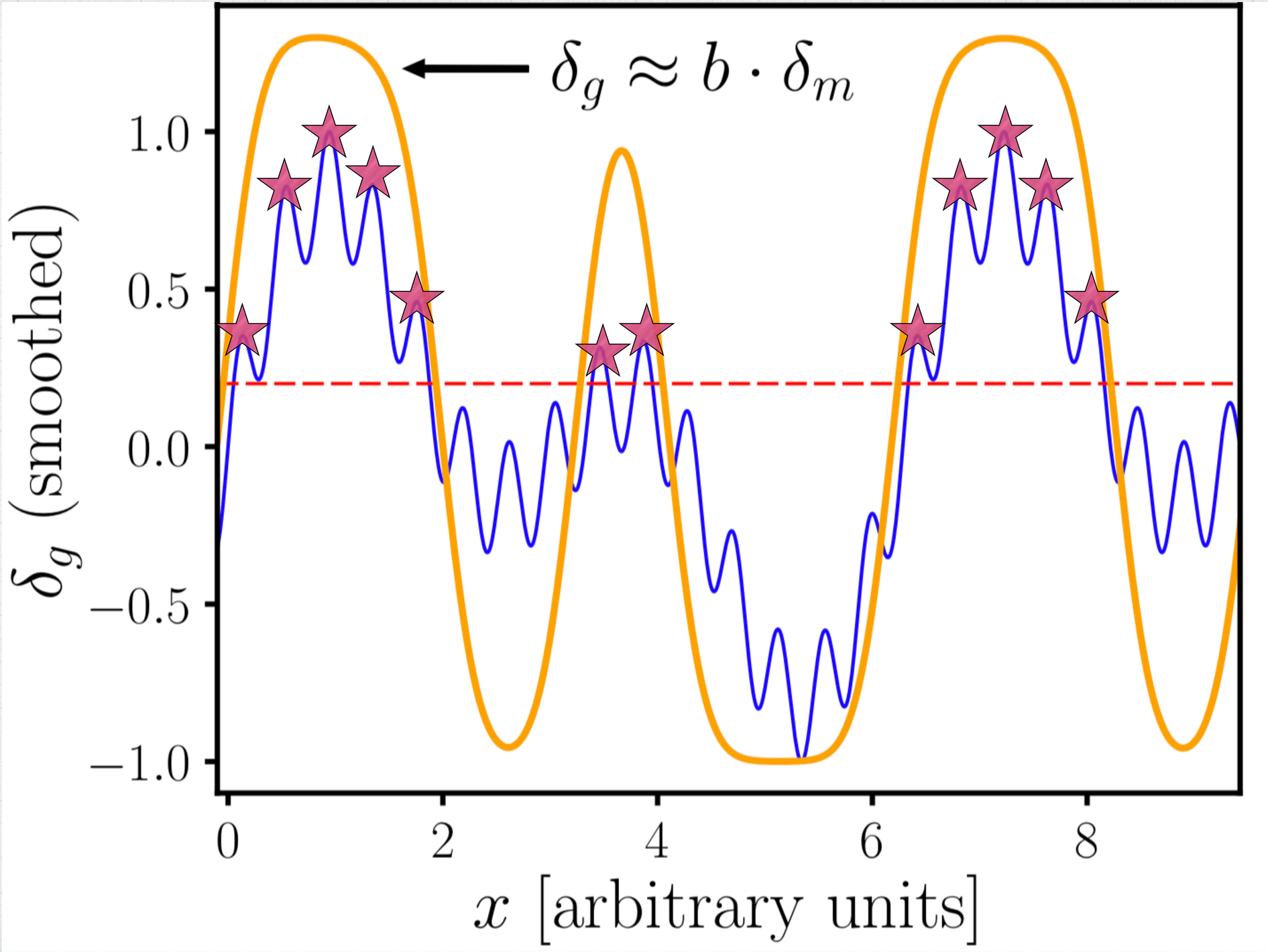
2-point statistics

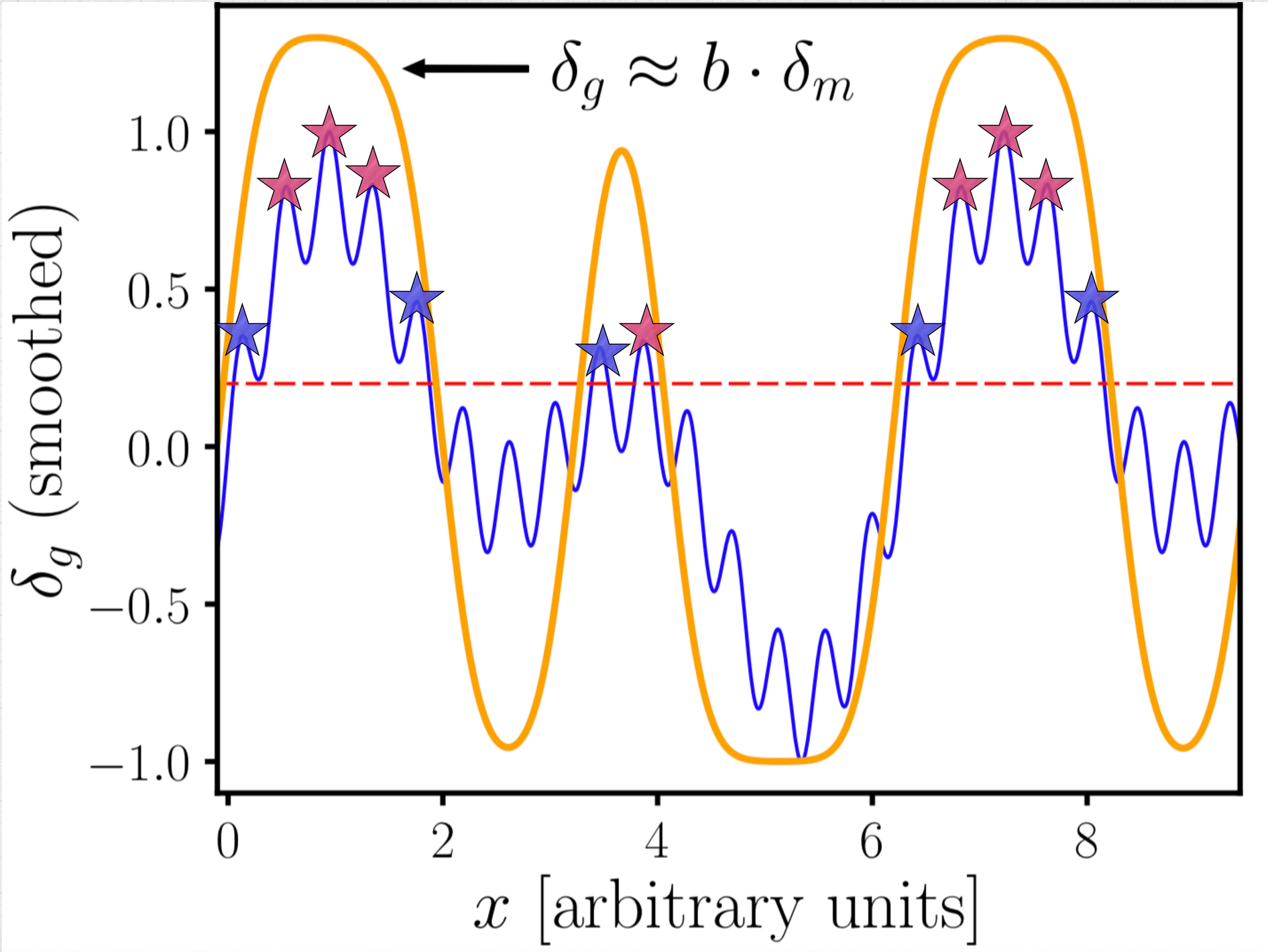
- Galaxy bias & shot-noise:

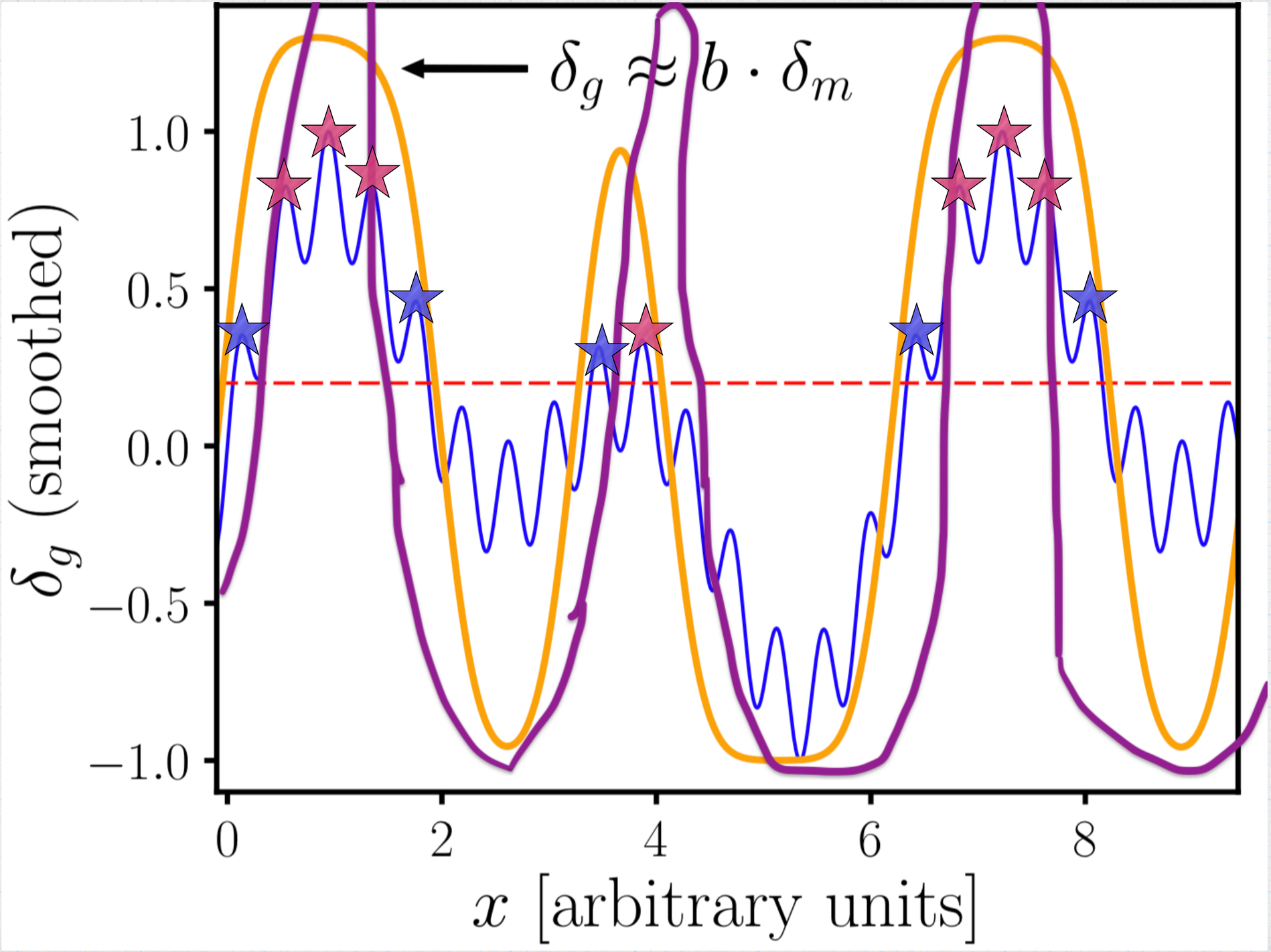
$$P_g(k) \approx b^2 P_m(k) + \frac{1}{n_g}$$




seed halos
formation threshold
δ_m
$\delta_g \approx b \cdot \delta_m$
δ_g (smoothed)
x [arbitrary units]
1.0
0.5
0.0
−0.5
−1.0
0
2
4
6
8


$\delta_g \approx b \cdot \delta_m$
1.0
0.5
0.0
−0.5
−1.0
δ_g (smoothed)
0
2
4
6
8
x [arbitrary units]


$\delta_g \approx b \cdot \delta_m$
1.0
0.5
0.0
−0.5
−1.0
δ_g (smoothed)
0
2
4
6
8
x [arbitrary units]


$\delta_g \approx b \cdot \delta_m$
1.0
0.5
0.0
−0.5
−1.0
δ_g (smoothed)
0
2
4
6
8
x [arbitrary units]

- Sofar mostly:

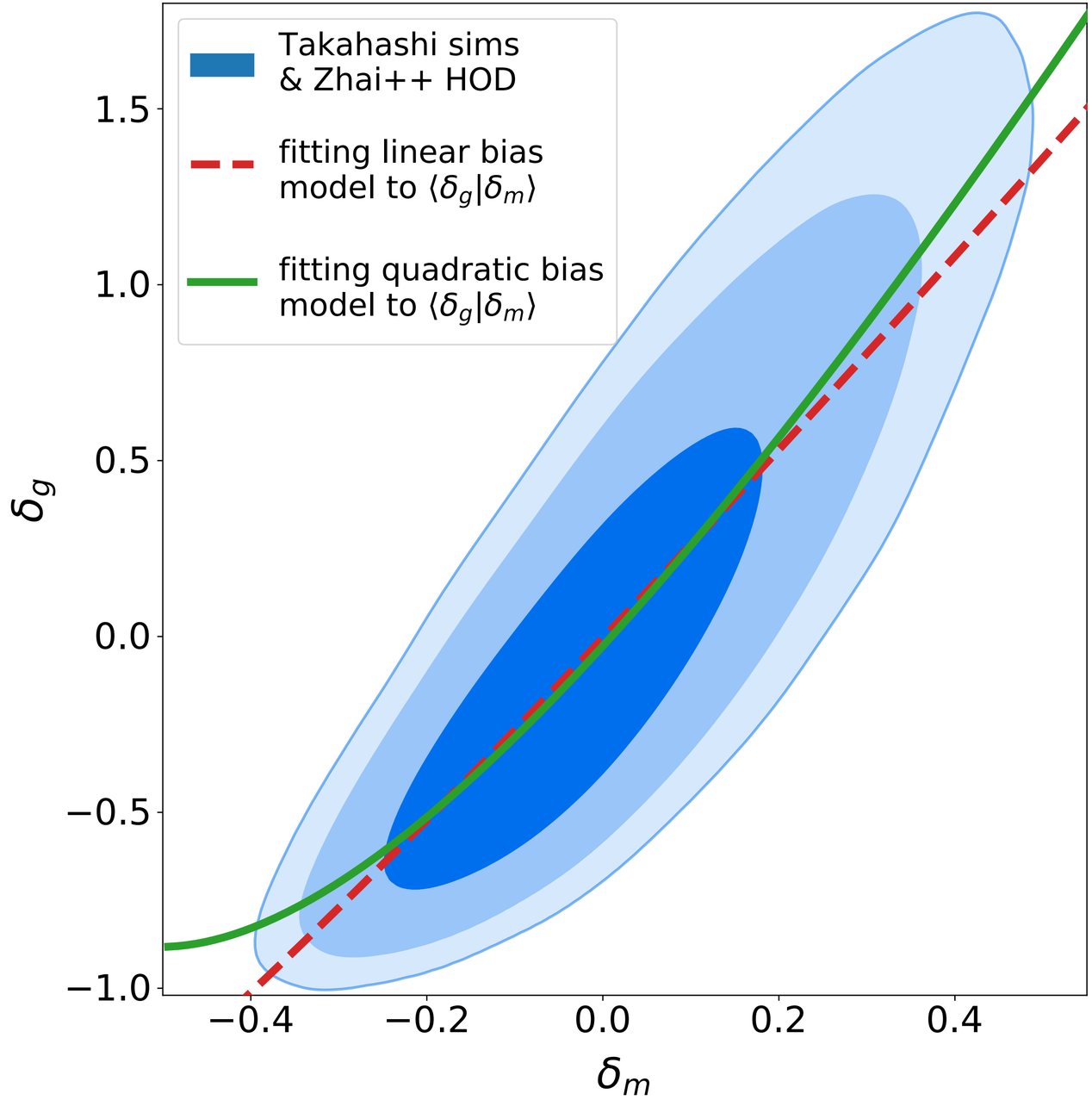
$$p(\delta_g \,|\, \delta_m) = \mathrm{Poisson}(\langle \delta_g \,|\, \delta_m \rangle)$$

$$\langle \delta_g \,|\, \delta_m \rangle \approx b_1^E \delta_m + \frac{b_2^E}{2}(\delta_m^2 - \langle \delta_m^2 \rangle)$$

- **We have implemented:**

  - Lagrangian bias model
    (expansion in terms of initial δ)

  - non-Poisson noise



Friedrich++ 2022b, **arXiv:2107.02300**

- Joint PDF of galaxies and matter:

$$p(\delta_m, \delta_g) = p(\delta_m)\; p(\delta_g \,|\, \delta_m)$$

Sofar mostly:

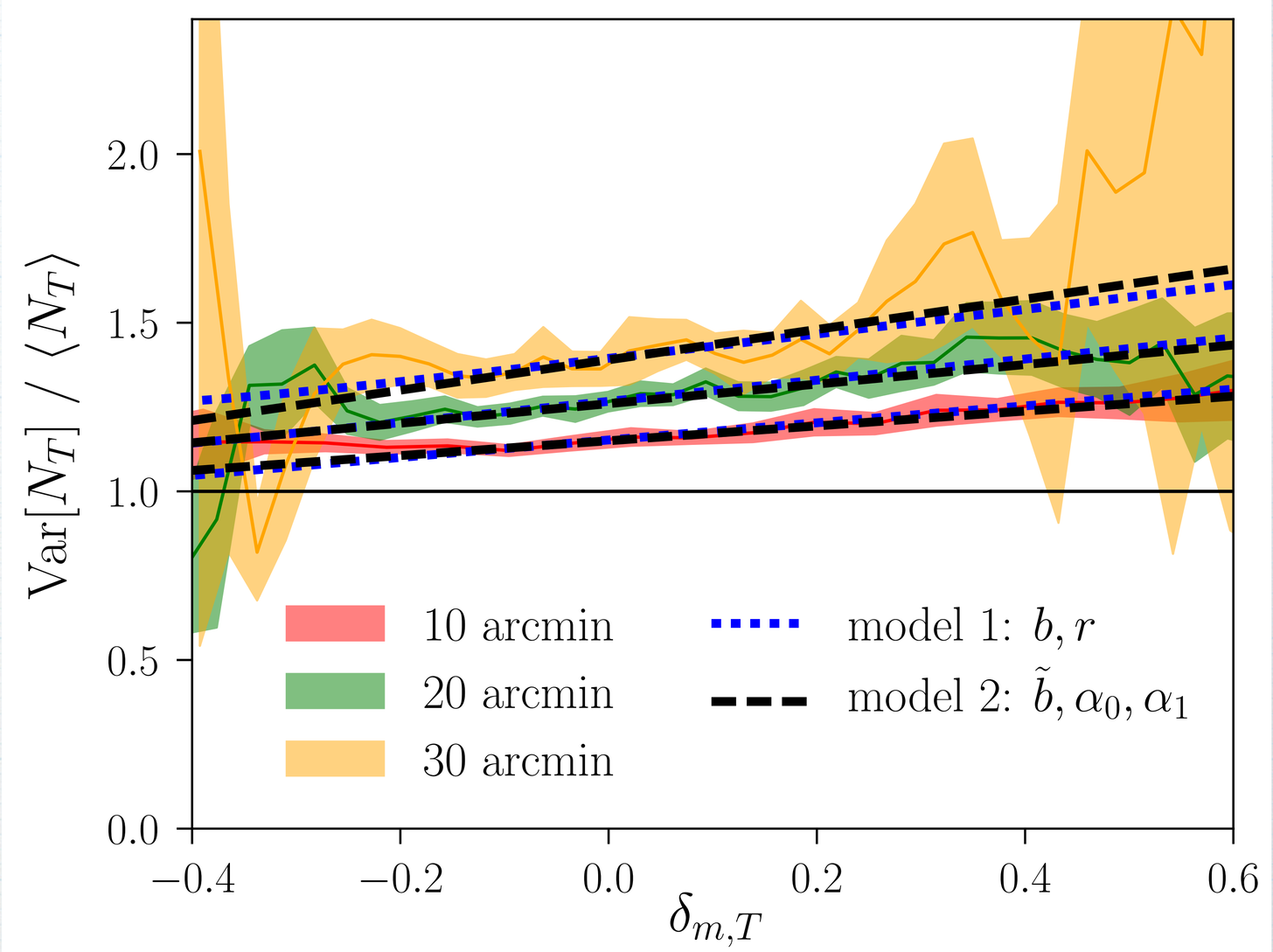
$$p(\delta_g \,|\, \delta_m) = \mathrm{Poisson}(\langle \delta_g \,|\, \delta_m \rangle)$$

$$\langle \delta_g \,|\, \delta_m \rangle \approx b_1^E \delta_m + \frac{b_2^E}{2}(\delta_m^2 - \langle \delta_m^2 \rangle)$$

We have implemented:

- Lagrangian bias model
  (expansion in terms of initial δ)

- non-Poisson noise



Friedrich++ 2018, **arXiv:1710.05162**

- Joint PDF of galaxies and matter:

$$p(\delta_m, \delta_g) = p(\delta_m)\; p(\delta_g \,|\, \delta_m)$$

Sofar mostly:

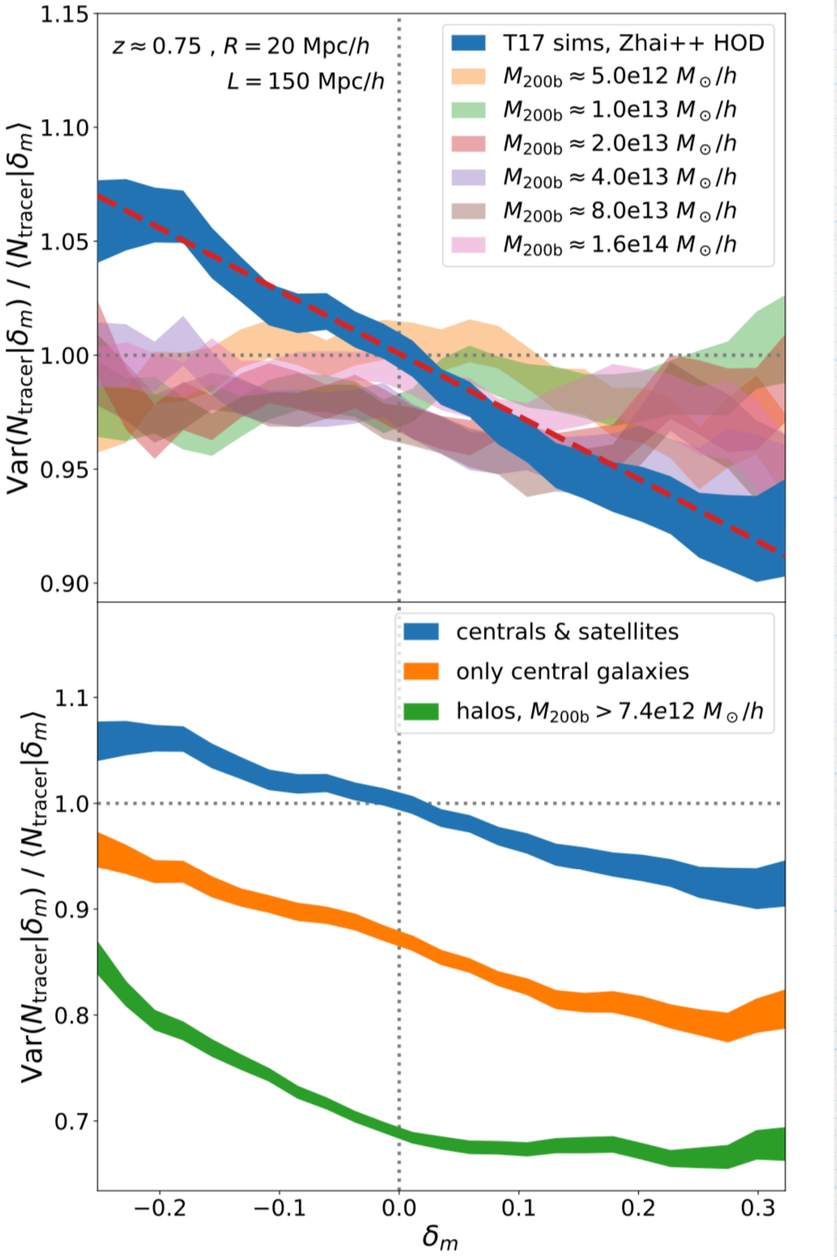
$$p(\delta_g \,|\, \delta_m) = \text{Poisson}(\langle \delta_g \,|\, \delta_m \rangle)$$

$$\langle \delta_g \,|\, \delta_m \rangle \approx b_1^E \delta_m + \frac{b_2^E}{2}(\delta_m^2 - \langle \delta_m^2 \rangle)$$

We have implemented:

- Lagrangian bias model
  (expansion in terms of initial δ)

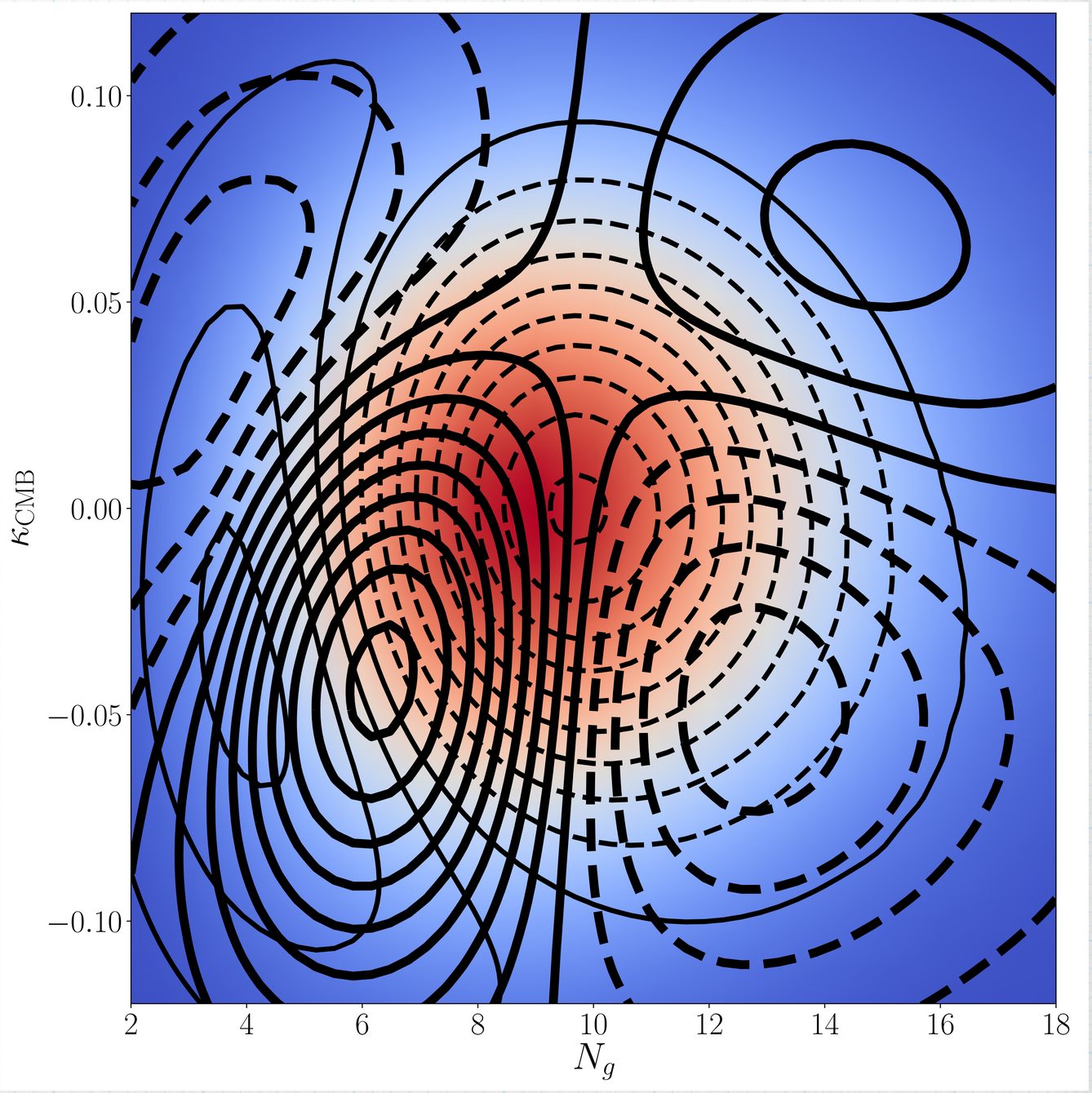
- non-Poisson noise

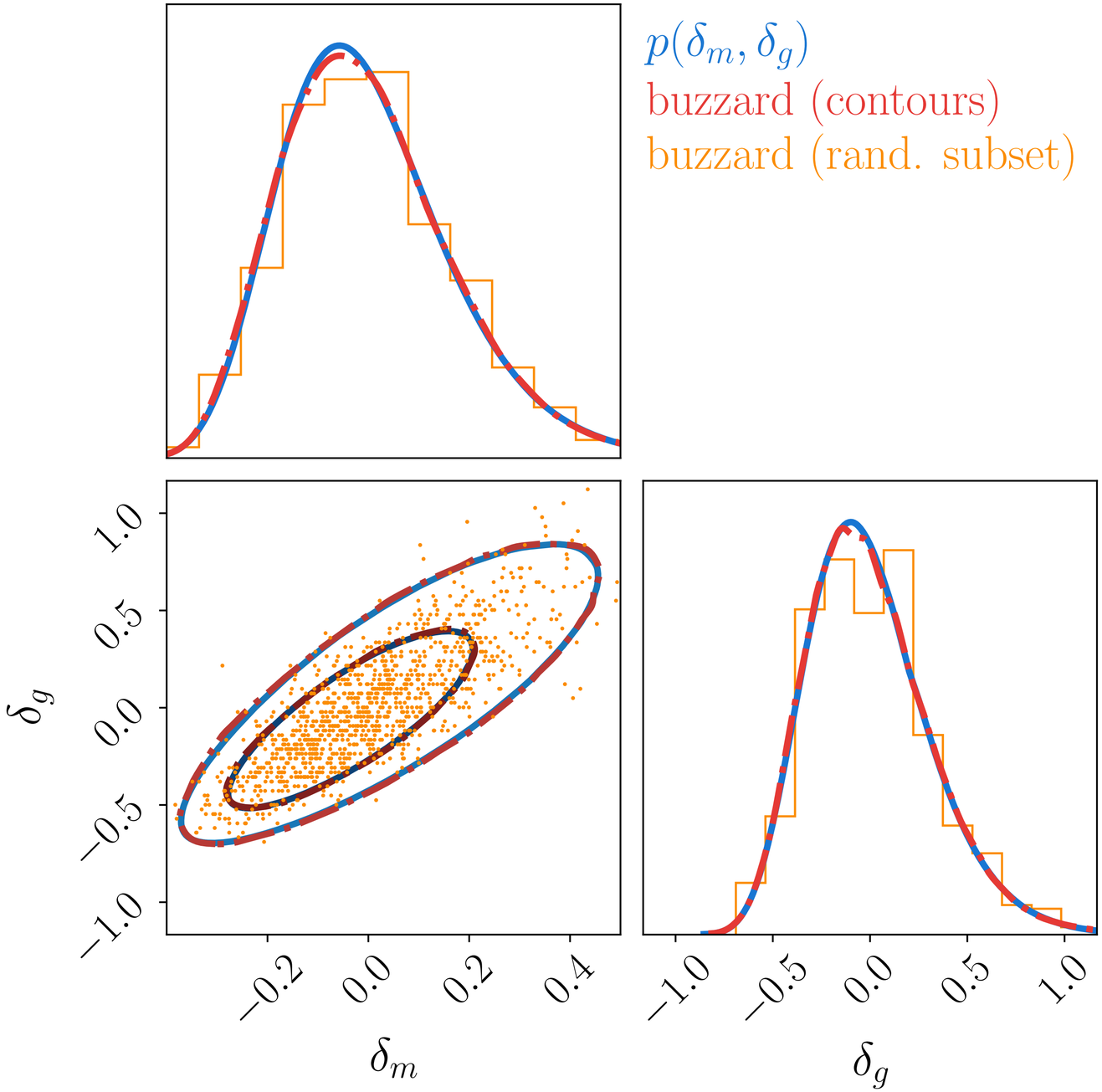


Friedrich++ 2021, **arXiv:2107.02300**

The full shape of the density PDF: A new dimension for studies of the large-scale structure

- Why the PDF?
- PDF vs. P(k) — the details
- **Let’s interpret the PDF**
- My view of PDF theory

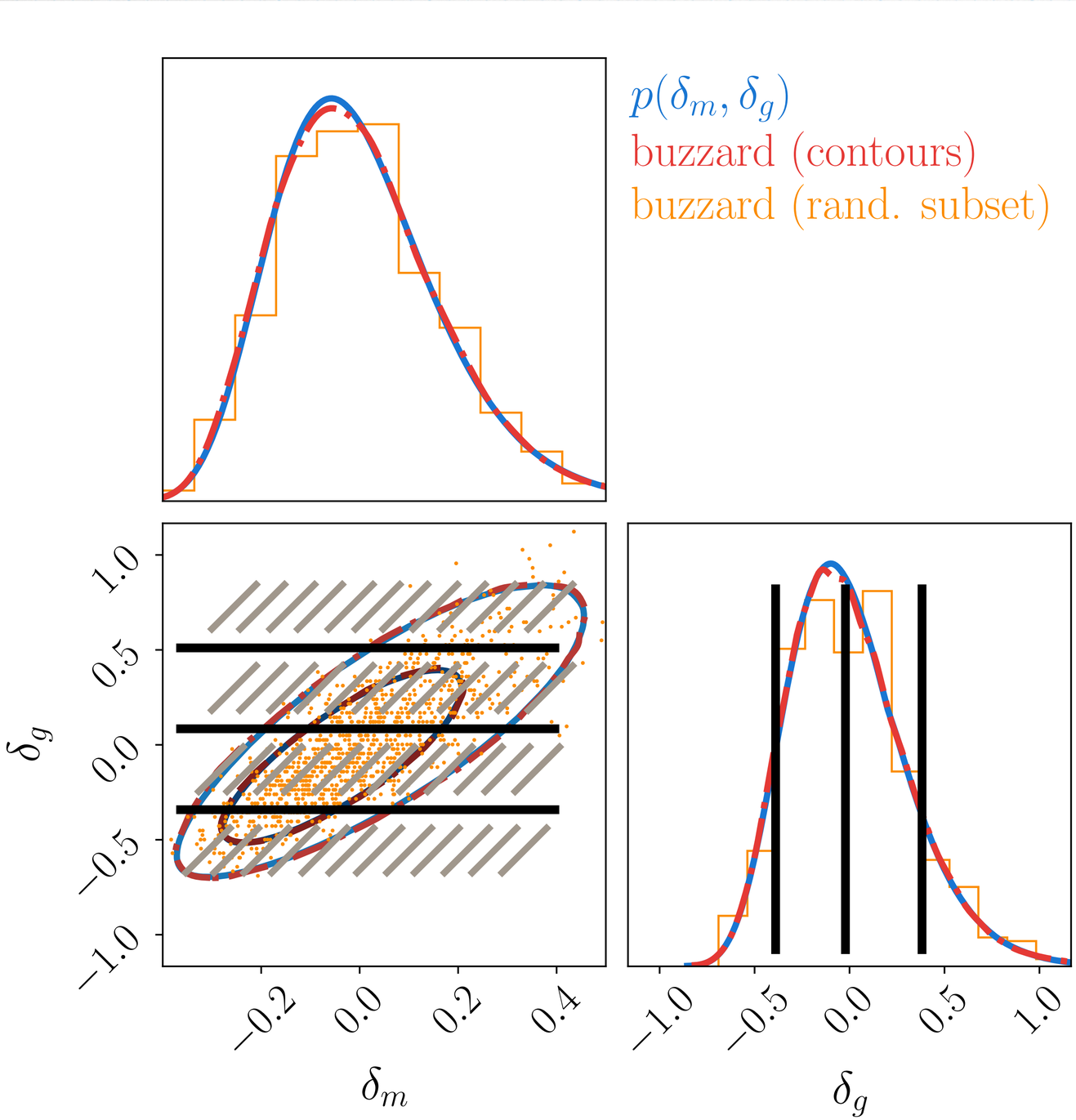




Blue - model predictions calculated in **Friedrich++ (2018)**

Red - data from Buzzard N-body sims (DeRose, Wechsler++2019)

But:
does this work in real data?



Blue - model predictions calculated in **Friedrich++ (2018)**

Red - data from Buzzard N-body sims (DeRose, Wechsler++ 2019)

But:
does this work in real data?

We already (partly) did this with DES year-1 data!

Density split statistics uses gravitational lensing to scan through galaxy density PDF “quantile-by-quantile” to connect it to underlying matter density PDF

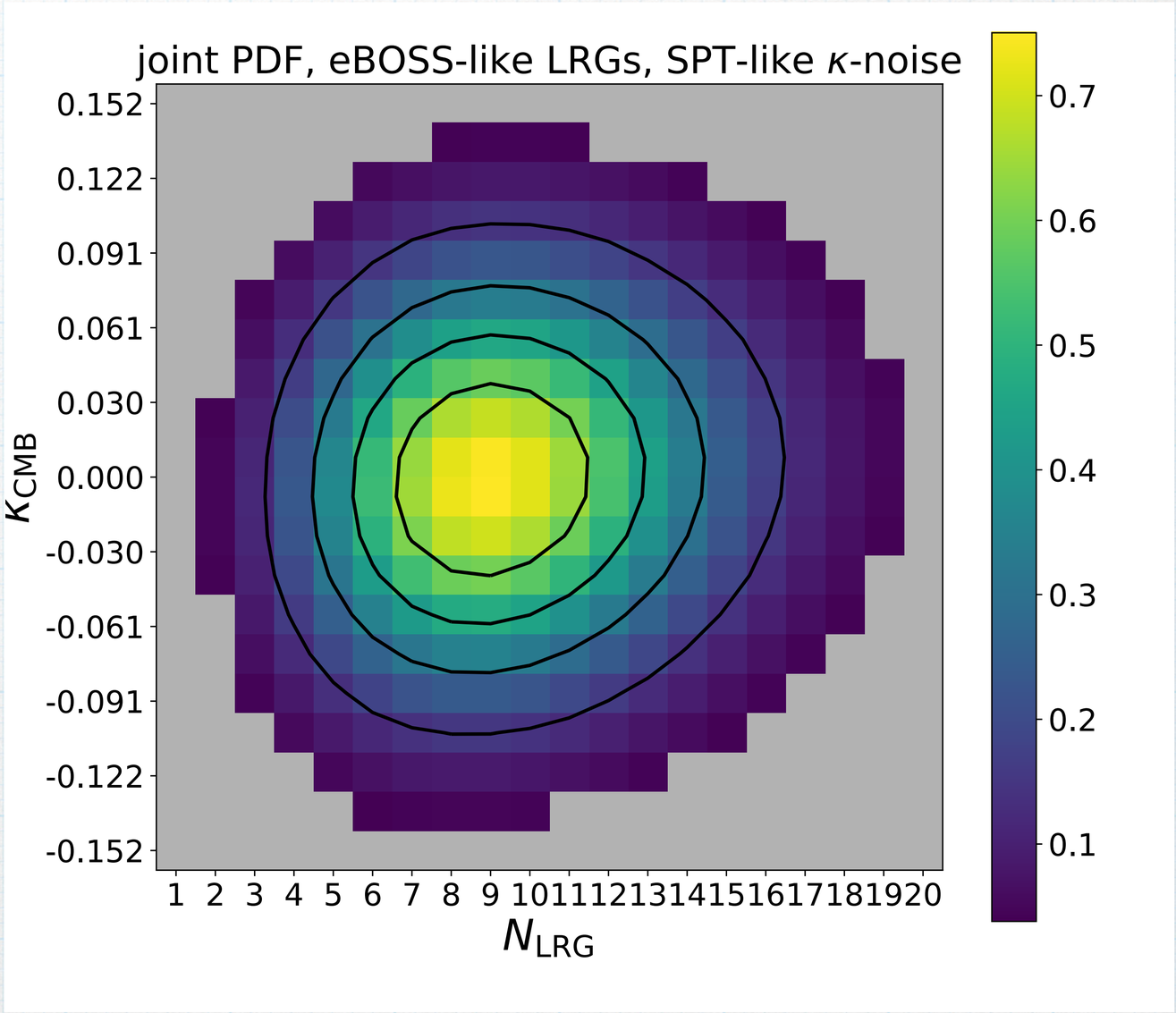
Friedrich et al. (2018)
Gruen et al. (2018)

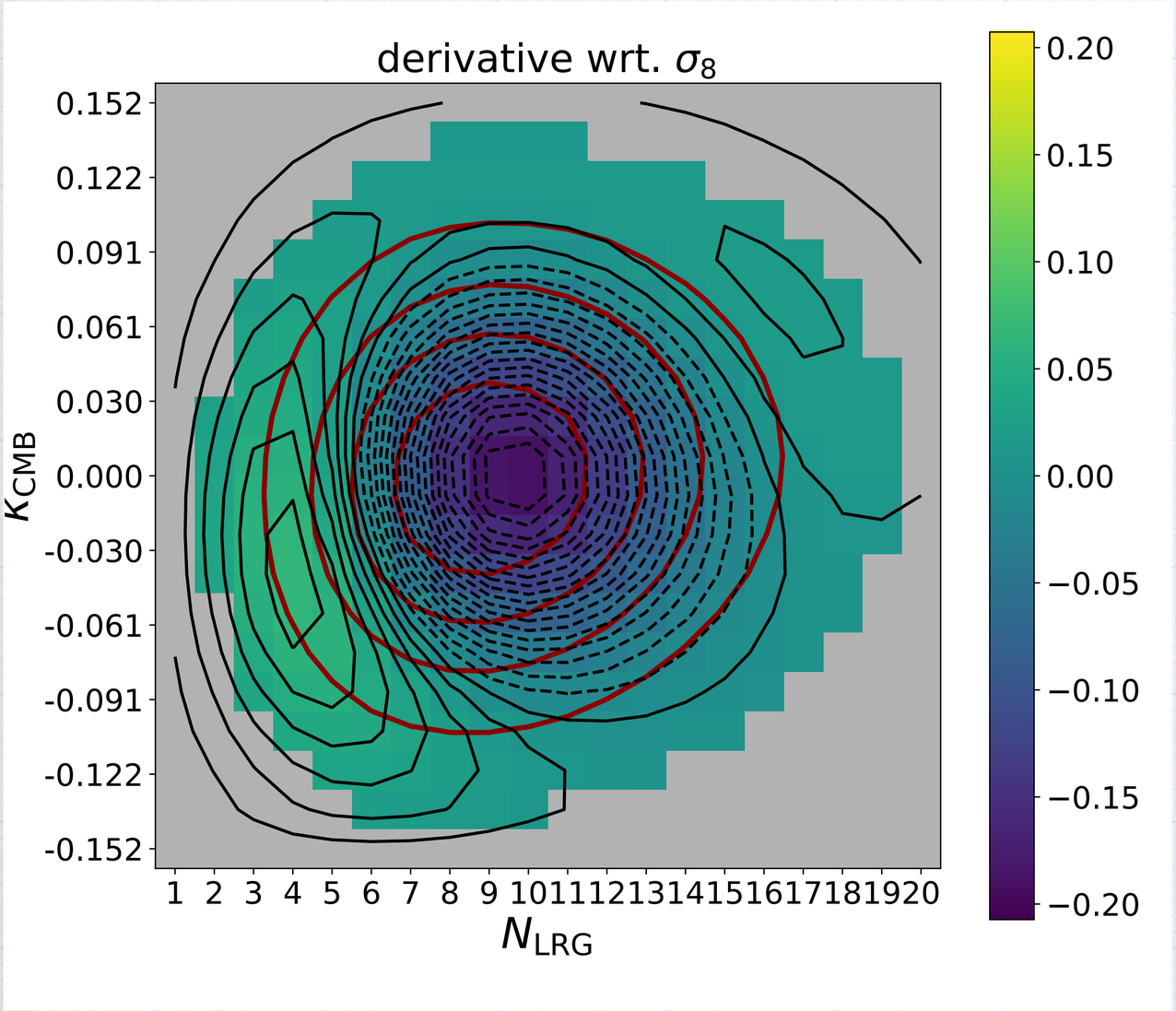
What do cosmological parameters do to the PDF?

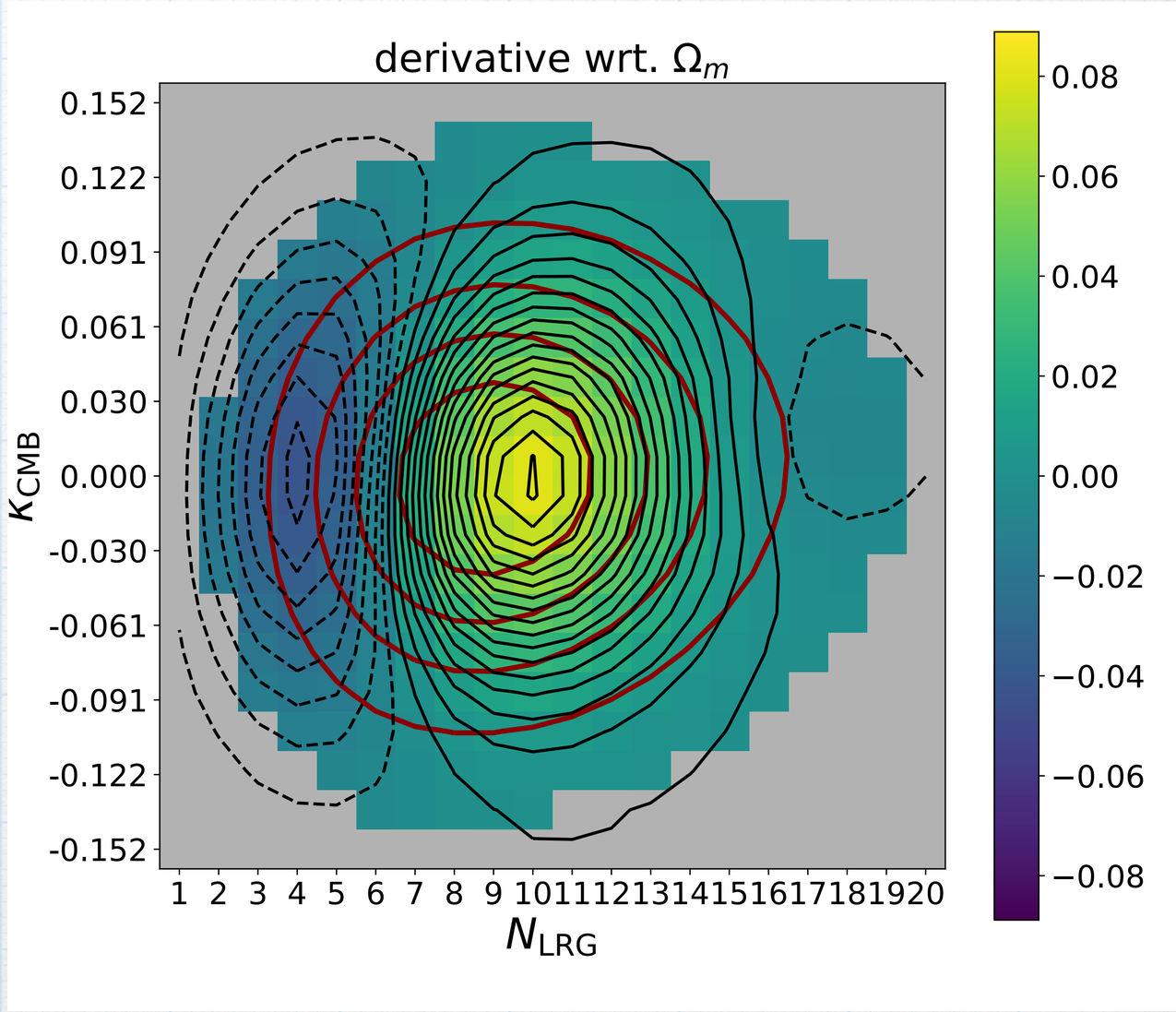
in the following assume:
(from Friedrich++ in prep)

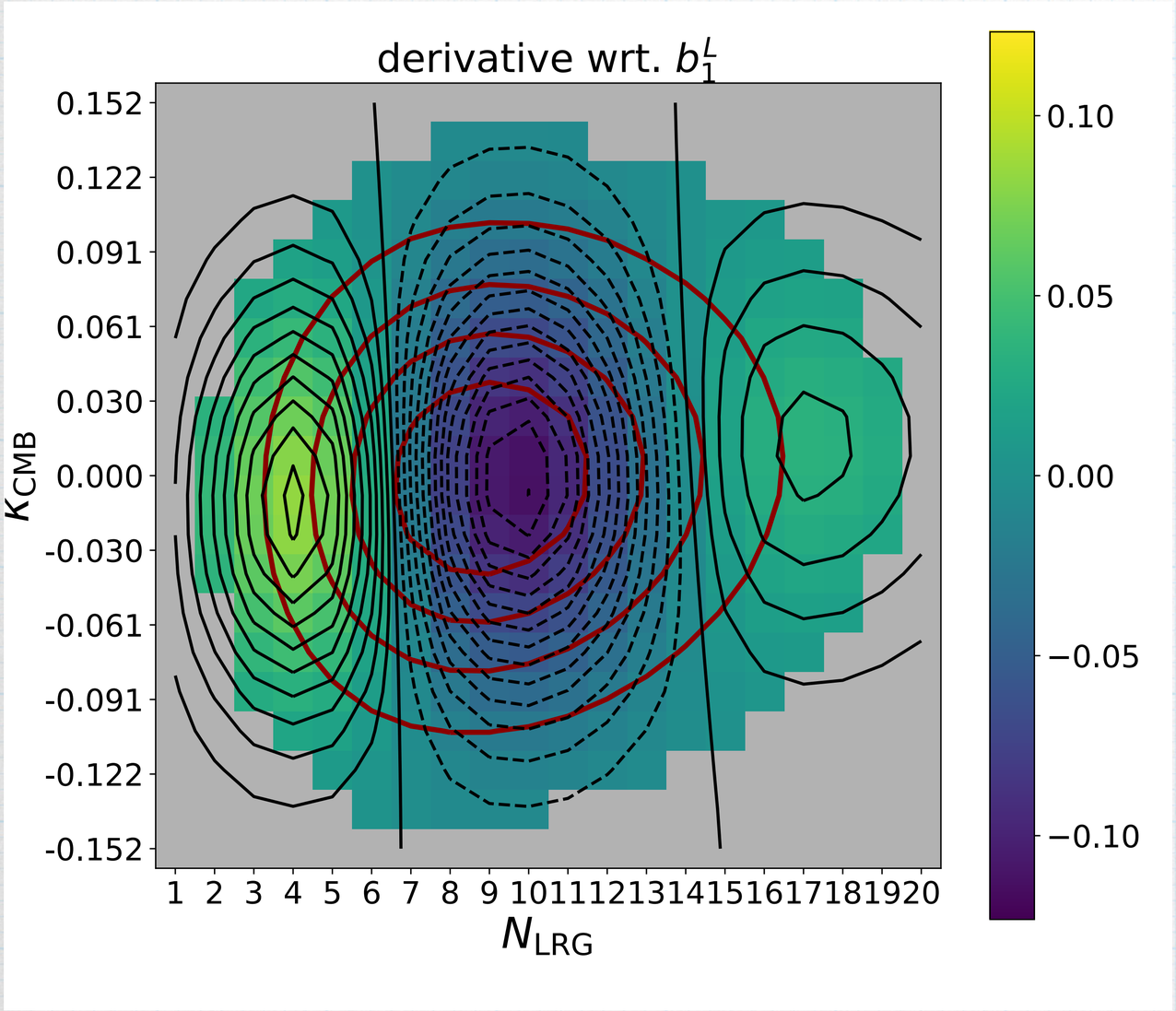
- LRGs (luminous red galaxies) within $0.6 < z < 0.9$
- CMB lensing with SPT-like noise
- 20 arcmin smoothing scale
- 5000 square degrees on the sky

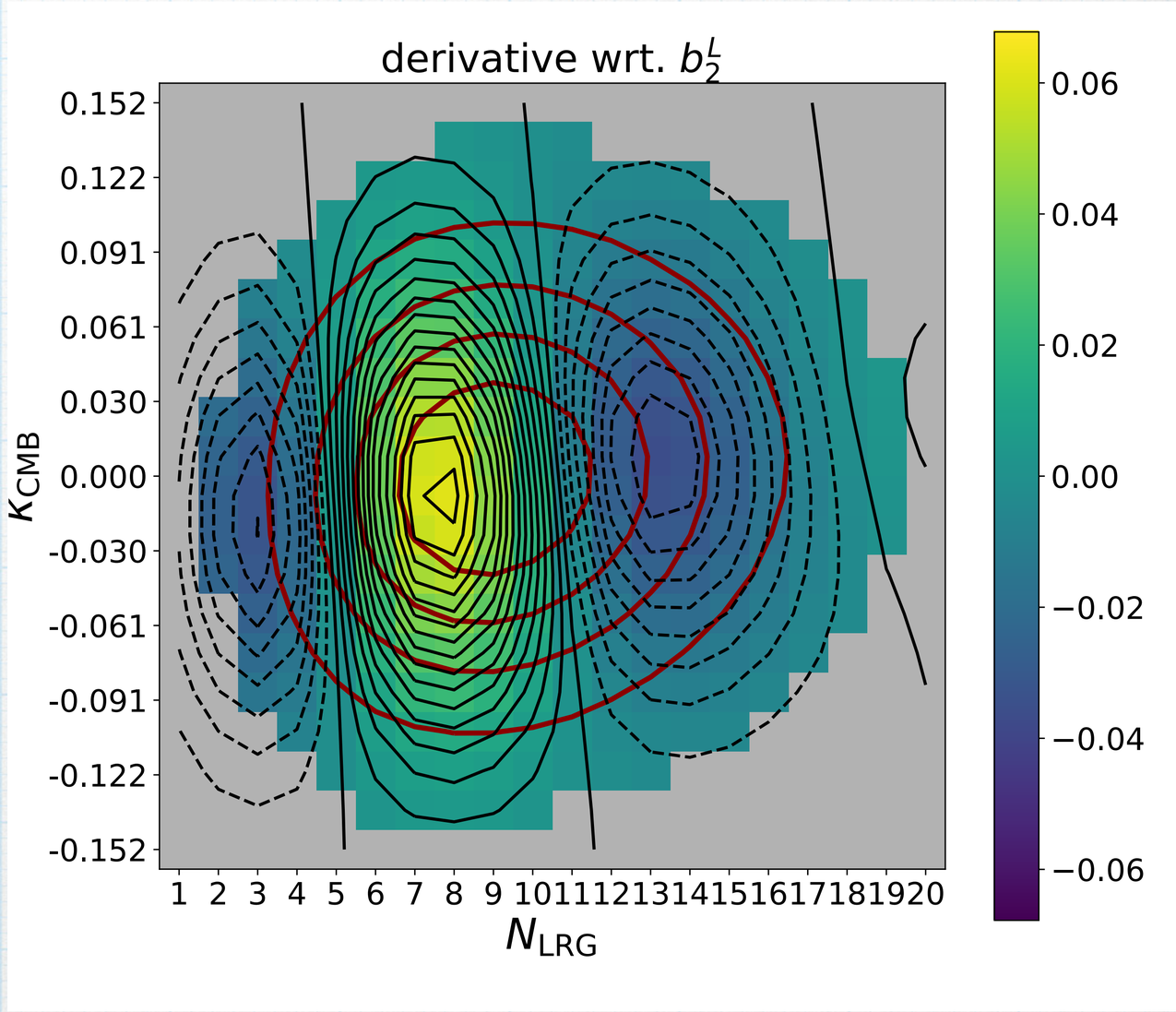
→ **measurable in real data!**

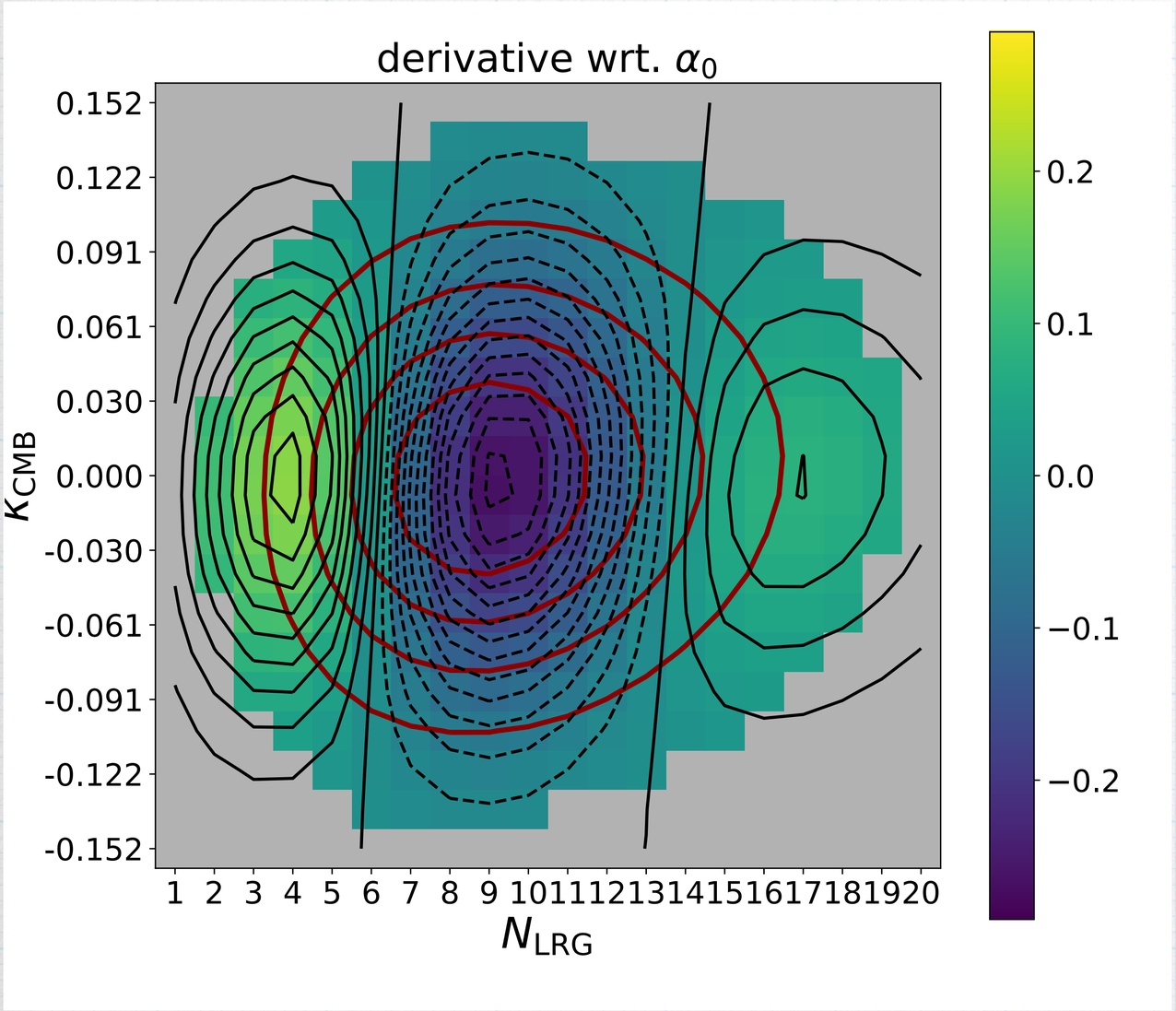


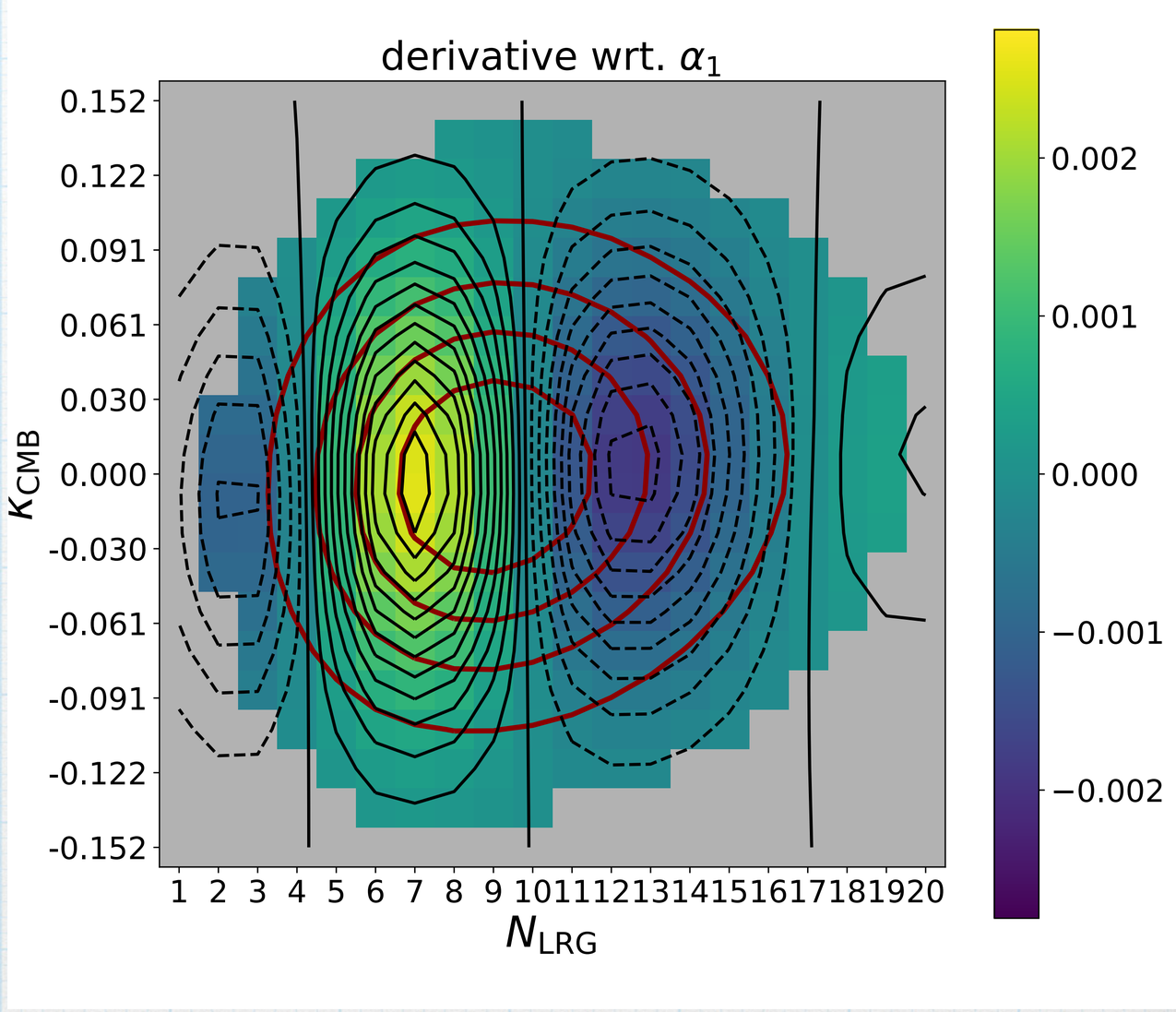

derivative wrt. σ_8
κ_{CMB}
N_{LRG}
0.152
0.122
0.091
0.061
0.030
0.000
-0.030
-0.061
-0.091
-0.122
-0.152
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
0.20
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
−0.20


derivative wrt. Ω_m
κ_{CMB}
N_{LRG}
0.152
0.122
0.091
0.061
0.030
0.000
-0.030
-0.061
-0.091
-0.122
-0.152
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
0.08
0.06
0.04
0.02
0.00
−0.02
−0.04
−0.06
−0.08


derivative wrt. b_1^L
κ_{CMB}
N_{LRG}
0.152
0.122
0.091
0.061
0.030
0.000
-0.030
-0.061
-0.091
-0.122
-0.152
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
0.10
0.05
0.00
−0.05
−0.10


derivative wrt. b_2^L
κ_{CMB}
N_{LRG}
0.152
0.122
0.091
0.061
0.030
0.000
-0.030
-0.061
-0.091
-0.122
-0.152
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
0.06
0.04
0.02
0.00
−0.02
−0.04
−0.06

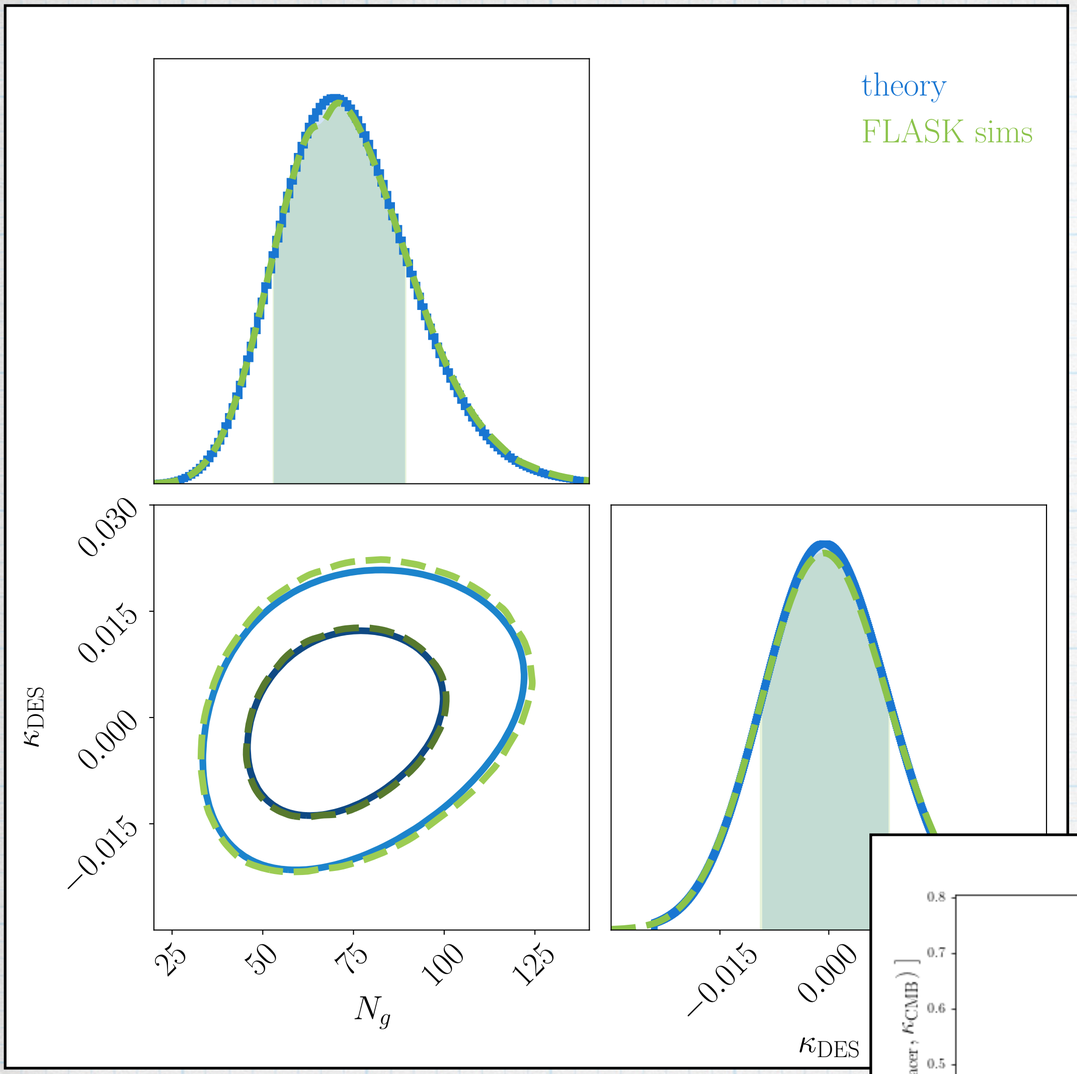

derivative wrt. α_0
κ_{CMB}
0.152
0.122
0.091
0.061
0.030
0.000
-0.030
-0.061
-0.091
-0.122
-0.152
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
N_{LRG}
0.2
0.1
0.0
−0.1
−0.2

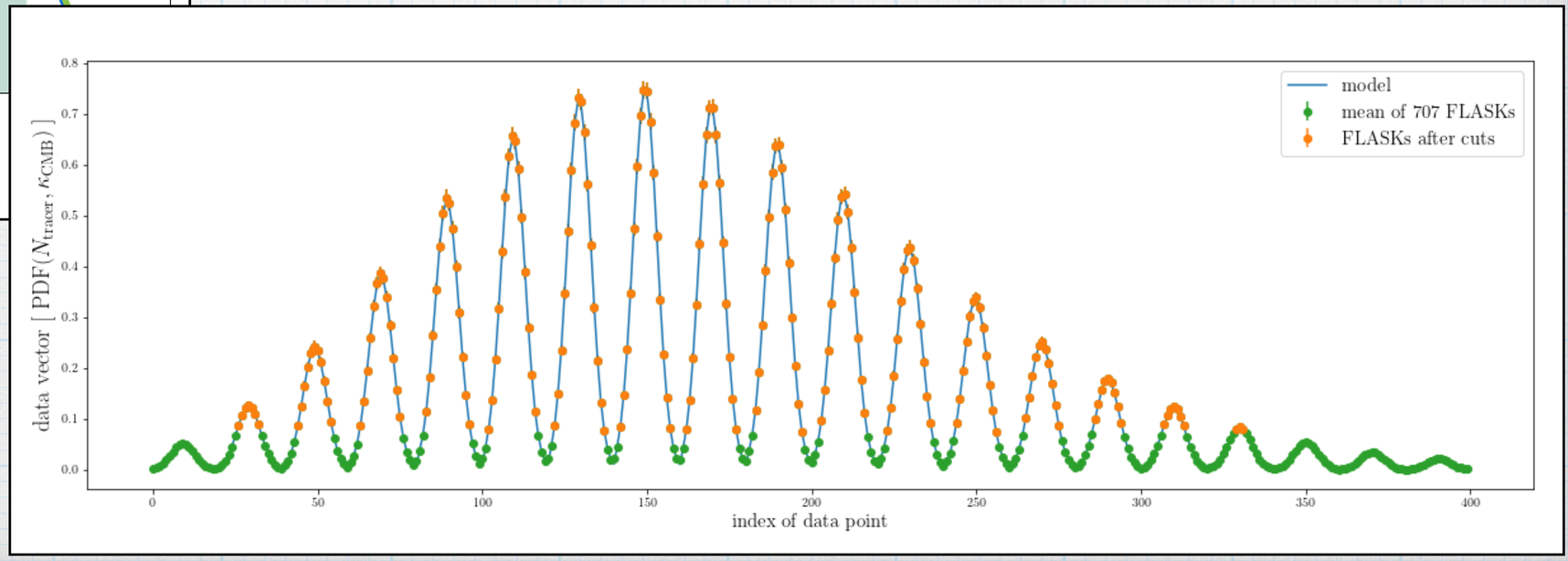

derivative wrt. α_1
κ_{CMB}
0.152
0.122
0.091
0.061
0.030
0.000
-0.030
-0.061
-0.091
-0.122
-0.152
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
N_{LRG}
0.002
0.001
0.000
−0.001
−0.002

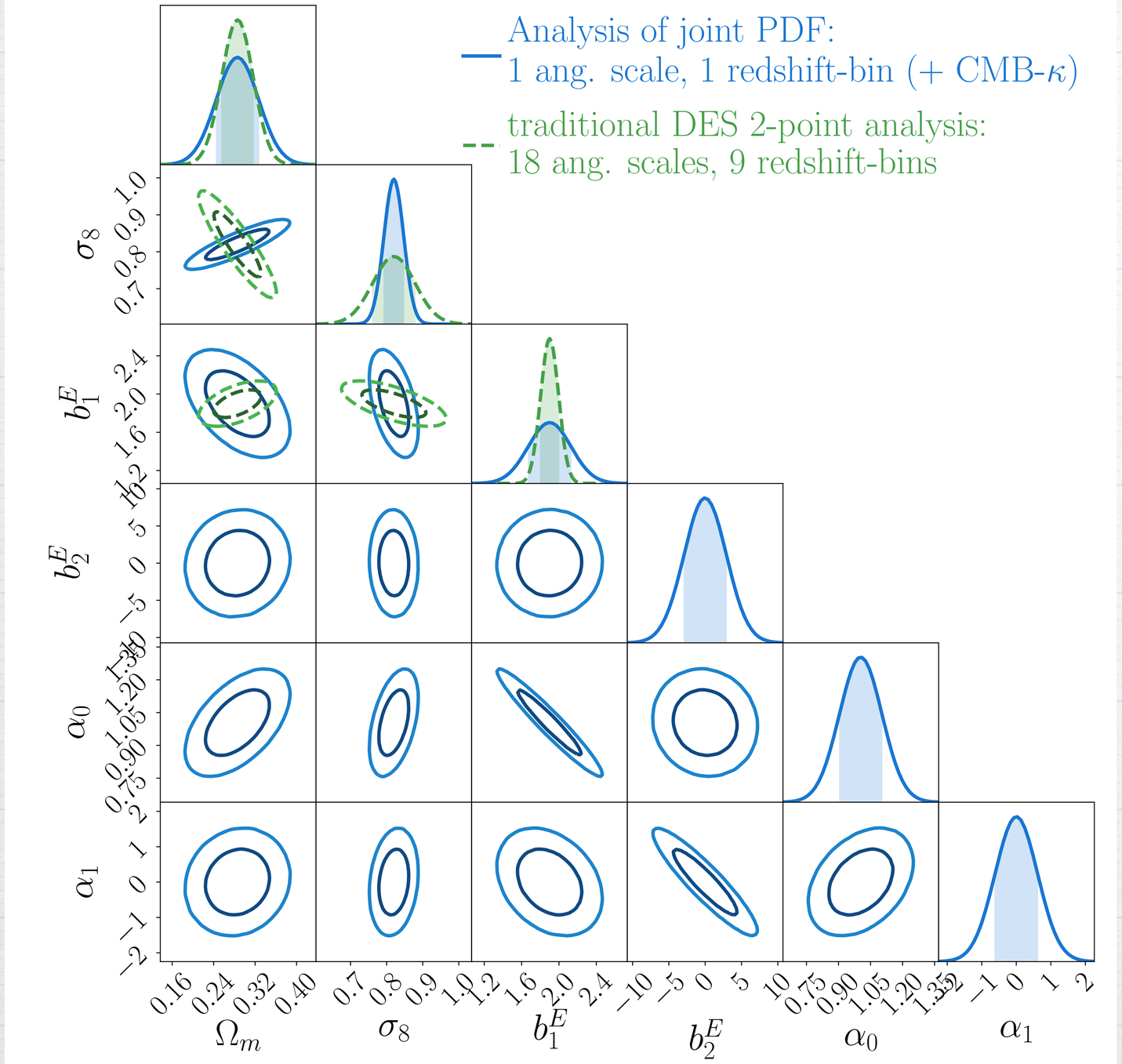
Covariance from log-normal Simulations

- have 700 FLASK simulations (Xavier et al. 2016)
- Combining several log-normal fields to capture non-Gaussian correlation structure

Work by Anik Halder, LMU Munich





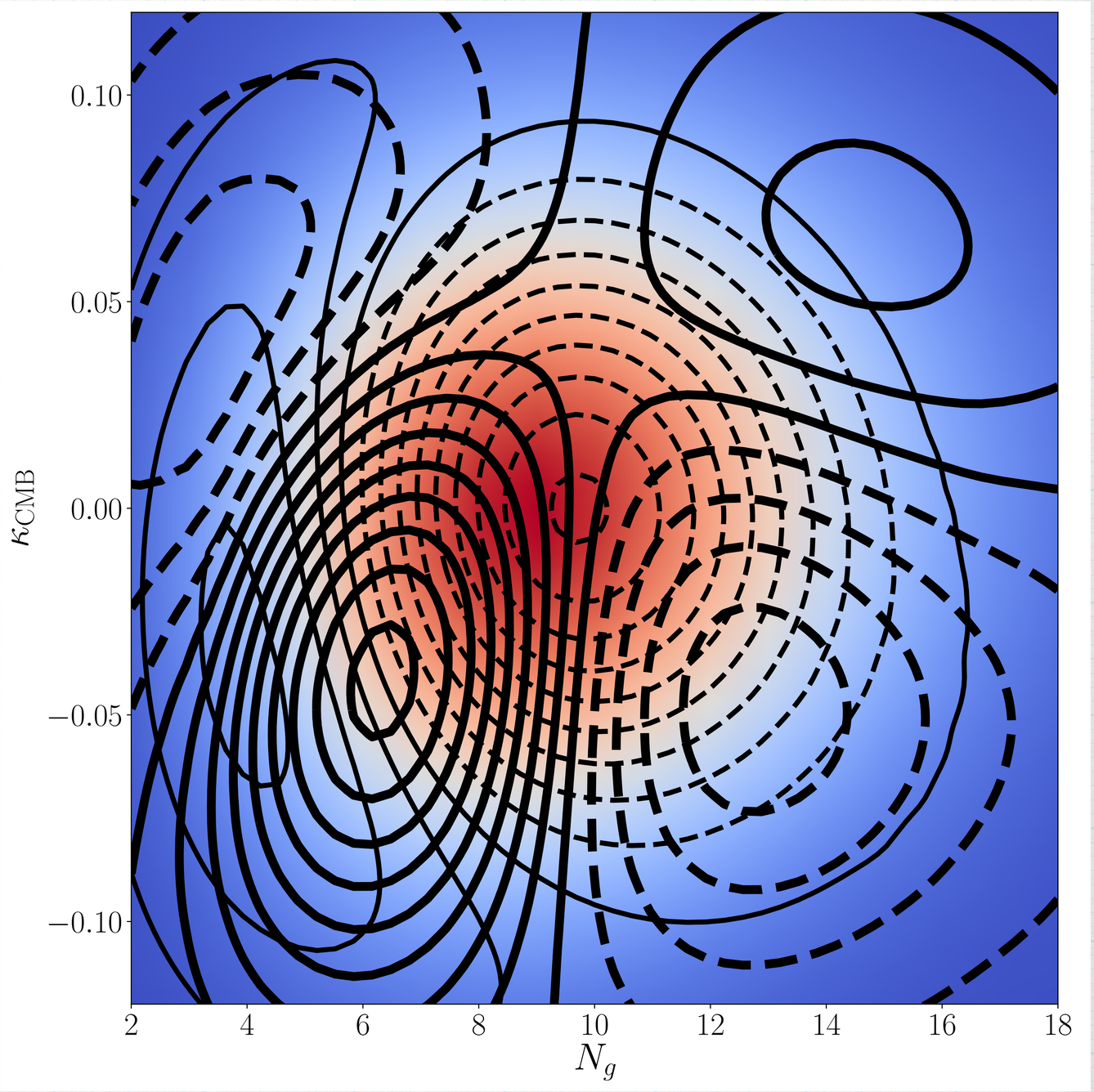


Remember:

at a single scale 2-point statistics only extracts 3 numbers & could never measure 6 properties of the cosmos at once

The full shape of the density PDF: A new dimension for studies of the large-scale structure

- Why the PDF?
- PDF vs. P(k) — the details
- Let's **interpret** the PDF
- **My view of PDF theory**



The core of PDF theory

Remember that

$$e^{\varphi_R(\lambda)} = \int \mathrm{d}\delta_R \; p(\delta_R) \; e^{\lambda\delta_R} \equiv \boxed{\langle e^{\lambda\delta_R} \rangle}$$

$$= \int \mathscr{D}\delta_{\rm lin} \; e^{\lambda\delta_R[\delta_{\rm lin}]} \; \mathscr{P}[\delta_{\rm lin}]$$

The core of PDF theory

Remember that

$$e^{\varphi_R(\lambda)} = \int \mathrm{d}\delta_R \; p(\delta_R) \; e^{\lambda\delta_R} \equiv \boxed{\langle e^{\lambda\delta_R} \rangle}$$

$$= \int \mathscr{D}\delta_{\rm lin} \; e^{\lambda\delta_R[\delta_{\rm lin}]} \; \mathscr{P}[\delta_{\rm lin}]$$

$$\sim \int \mathscr{D}\delta_{\rm lin} \; \mathscr{D}J_{\rm lin} \; e^{-S_\lambda[\delta_{\rm lin}, J_{\rm lin}]}$$

$$\text{with } S_\lambda[\delta_{\rm lin}, J_{\rm lin}] \equiv -\lambda\delta_R[\delta_{\rm lin}] + iJ_{\rm lin} \cdot \delta_{\rm lin} - \Phi[iJ_{\rm lin}] \; .$$

The core of PDF theory

Remember that

$$e^{\Phi[J]} \equiv \sum_n \int \prod_{i=1}^{n} \mathrm{d}^3 x_i \; \xi_{n,\mathrm{lin}}(\mathbf{x}_1 \, , \ldots, \mathbf{x}_n) \frac{J(\mathbf{x}_1) \ldots J(\mathbf{x}_n)}{n\,!}$$

$$\sim \int \mathscr{D}\delta_{\mathrm{lin}} \; \mathscr{D}J_{\mathrm{lin}} \; e^{-S_\lambda[\delta_{\mathrm{lin}}, J_{\mathrm{lin}}]}$$

$$\text{with } S_\lambda[\delta_{\mathrm{lin}}, J_{\mathrm{lin}}] \equiv -\lambda \delta_R[\delta_{\mathrm{lin}}] + i J_{\mathrm{lin}} \cdot \delta_{\mathrm{lin}} - \Phi[i J_{\mathrm{lin}}] \; .$$

The core of PDF theory

Remember that

$$e^{\varphi_R(\lambda)} = \int \mathrm{d}\delta_R \; p(\delta_R) \; e^{\lambda\delta_R} \equiv \langle e^{\lambda\delta_R} \rangle$$

$$= \int \mathscr{D}\delta_{\rm lin} \; e^{\lambda\delta_R[\delta_{\rm lin}]} \; \mathscr{P}[\delta_{\rm lin}]$$

$$\sim \int \mathscr{D}\delta_{\rm lin} \; \mathscr{D}J_{\rm lin} \; e^{-S_\lambda[\delta_{\rm lin},J_{\rm lin}]}$$

$$\text{with } S_\lambda[\delta_{\rm lin}, J_{\rm lin}] \equiv -\lambda\delta_R[\delta_{\rm lin}] + iJ_{\rm lin}\cdot\delta_{\rm lin} - \Phi[iJ_{\rm lin}] \; .$$

The core of PDF theory

Approximate integral at saddle point configuration
- i.e. for **fields that minimise action**!

$$\Rightarrow \varphi_R(\lambda) \approx -S_\lambda[\delta^*_{\rm lin}, J^*_{\rm lin}]$$

The saddle point is spherically symmetric!

$$\Rightarrow \delta_R[\delta_{\rm lin}] \equiv \text{spherical collapse}$$

The core of PDF theory

Approximate integral at saddle point configuration
- i.e. for **fields that minimise action**!

$$\Rightarrow \varphi_R(\lambda) \approx -S_\lambda[\delta^*_{\rm lin}, J^*_{\rm lin}]$$

+ functional determinant of Hessian of action 😱 (cf. Valageas 2005, Ivanov++ 2019)

The saddle point is spherically symmetric!

$$\Rightarrow \delta_R[\delta_{\rm lin}] \equiv \text{spherical collapse}$$

What we are really doing...

$$\tilde{\varphi}_R(\lambda) = \sum_n \frac{\langle \delta_R^n \rangle_c}{\langle \delta_R^2 \rangle_c^{n-1}} \frac{\lambda^n}{n!} = \sum_n S_n \frac{\lambda^n}{n!}$$

$$\tilde{\varphi}_R(\lambda) = \tilde{\varphi}_R(\lambda \cdot \langle \delta_R^2 \rangle_c) / \langle \delta_R^2 \rangle_c$$

You can do these calculations yourself with **CosMomentum**:

https://github.com/OliverFHD/CosMomentum

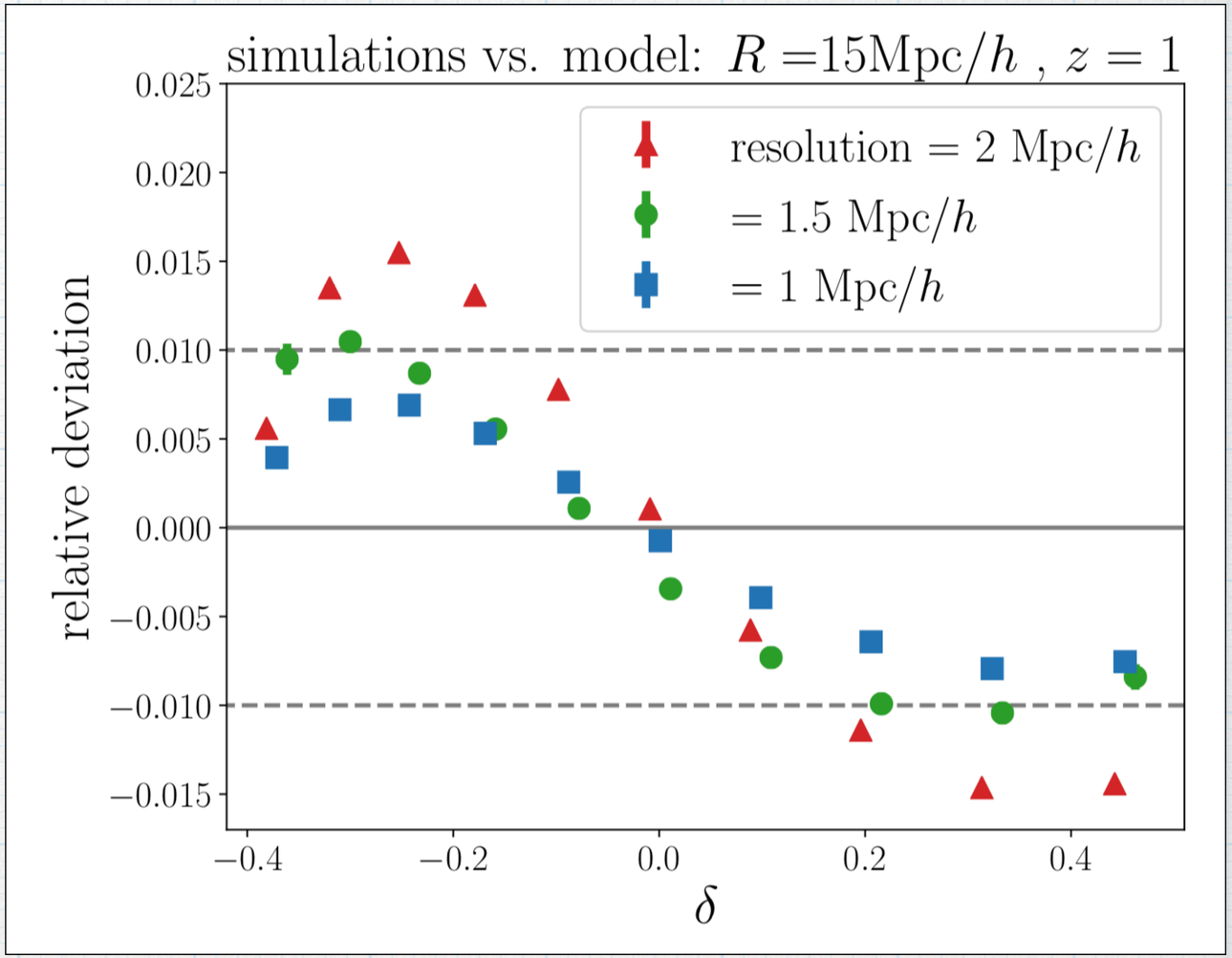
Also, section 4.6 of https://arxiv.org/abs/1912.06621 describes **step-by-step details of numerical implementation**!

Conclusions / Outlook

- Can open up complete “plane of perturbations” for LSS studies

- complete LCMD analysis possible even with complicated bias model and simple analysis setup

- Much more!
  (see e.g. primordial non-Gaussianity in **Friedrich ++ 2020** or neutrino physics from PDF in **Uhlemann, Friedrich et al. 2020 ; Boyle, Uhlemann, Friedrich et al 2021**)

- Most of these tools are already publicly available in the **CosMomentum** package:
  https://github.com/OliverFHD/CosMomentum

  - 3D PDFs and cumulant generating functions
  - PDF of gravitational lensing convergence
  - 2D and 3D halo counts-in-cells PDFs