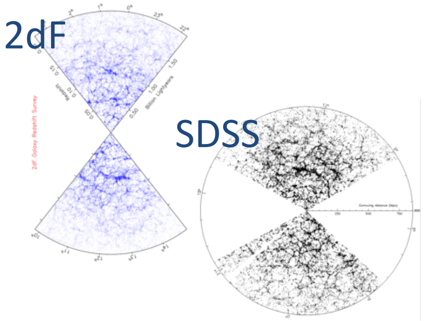
N-body Matter Power Spectrum and Covariance Benchmarks for Future Galaxy Survey Data Analyses

Pier Stefano Corasaniti

CNRS & Observatoire de Paris

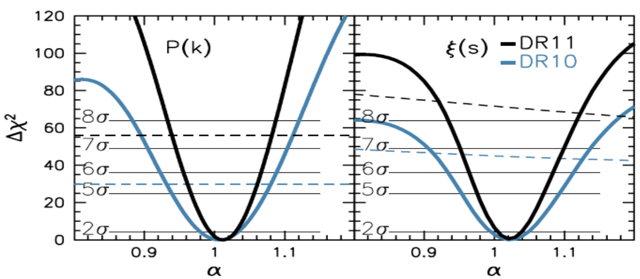
From Past to Present

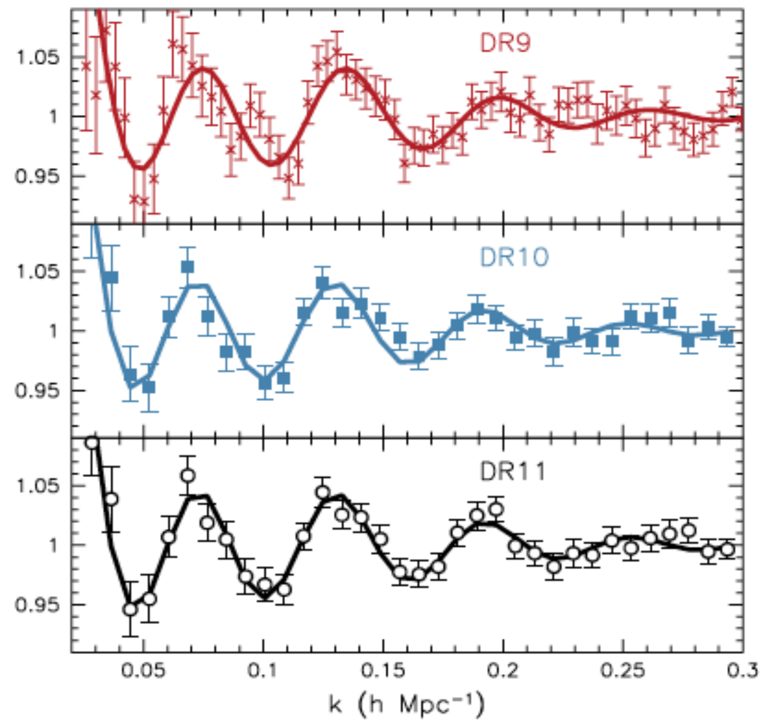
Galaxy Surveys



- 2dF: spectroscopic redshifts of ~200.000 galaxies in 1500 deg^2 at z < 0.3
- SDSS-I: spectroscopic redshifts of ~300.000 galaxies in ~4000 deg^2 at z < 0.7

- SDSS/BOSS: spectroscopic redshifts of ~1.000.000 galaxies from 8500 deg^2 in 0.2 < z < 0.7
- α(z=0.57) = 1.0144 ± 0.0098 (stat+sys)





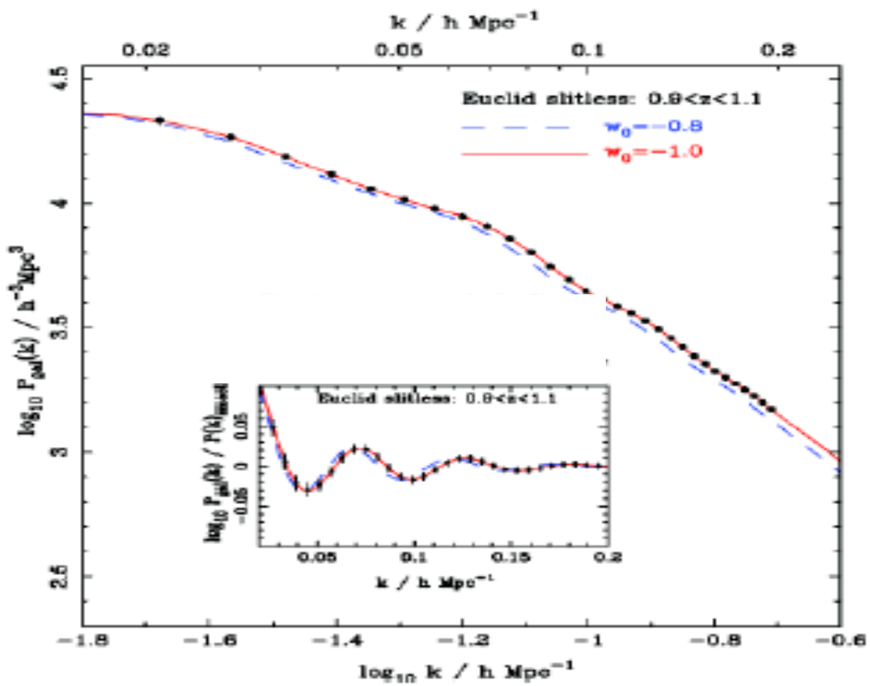
Anderson et al. (2014)

In a not too far future

DES

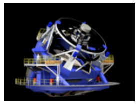
DARK ENERGY SURVEY

Photometric redshifts of ~3 10^8 galaxies in 5000 deg^2 at $z < 1.2$

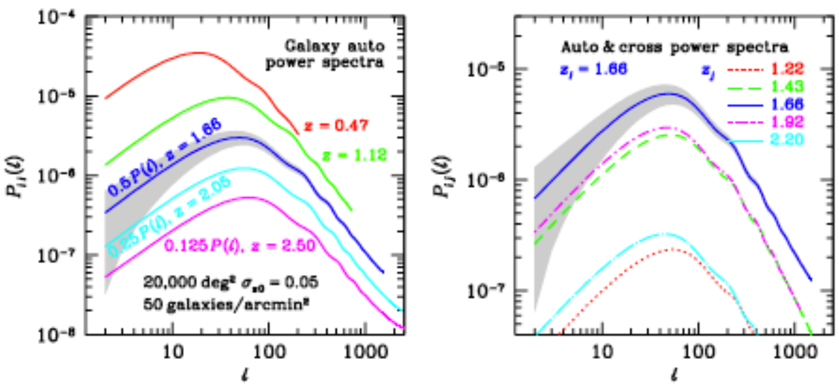


from Euclid red book

LSST



Photometric redshifts of ~10^9 galaxies full sky at $z < 1.5$ and beyond



from LSST red book

Euclid

Spectroscopic redshifts of ~10^9 galaxies full sky at $z < 1.5$ and beyond

Are we ready?

Expectations

- Large Volumes & Wide Redshift Range
- High Galaxy Number Density
- Reach ~ few % Statistical Errors

Requirements

- Cosmological model predictions to few % accuracy
- Estimate Power Spectrum Covariance

Theoretical Challenges

- At % level non-negligible even at $k \sim 0.1$ and $z = 1$
- Non-linear mode correlations
- Deviations from Gaussian statistics

Example: Resolving the BAO Scale

Large Dynamical Range

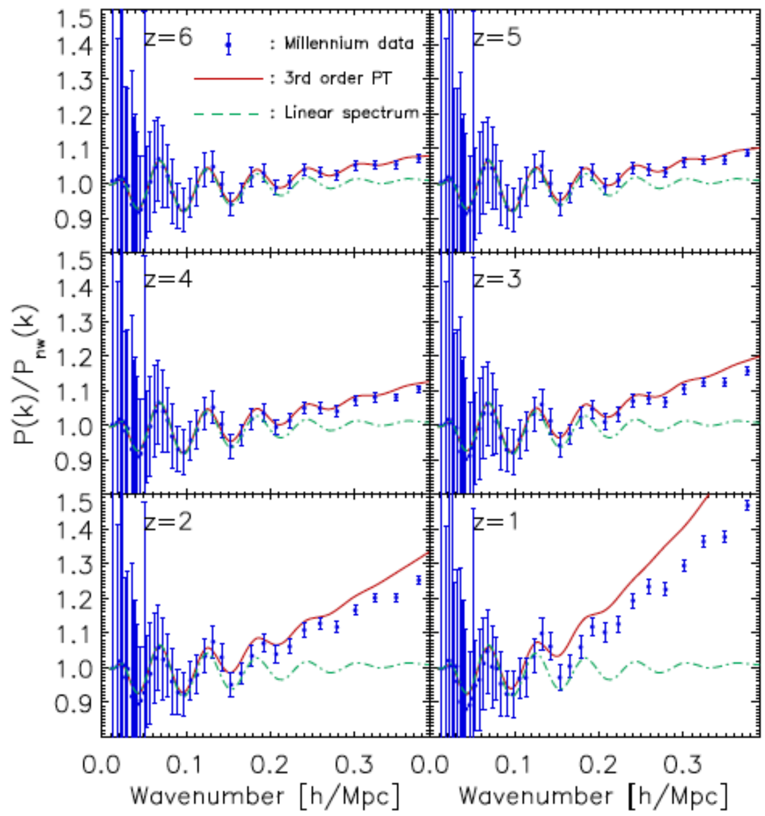
- ~100 Mpc/h
- ~1-10 Mpc/h width

N-body Requirements

- Large Volume (> 1 Gpc)^3
- High Resolution ($\leq 10^{12}$ M_{sun})

Sample Variance Errors

$$\sigma(k) = \sqrt{\frac{2}{N_k}}\left[P(k) + \frac{1}{N_p}\right] \quad \text{with} \quad N_k = \frac{k^2 \Delta k}{2\pi^2} V \qquad \Delta k = \frac{2\pi}{L_{box}}$$



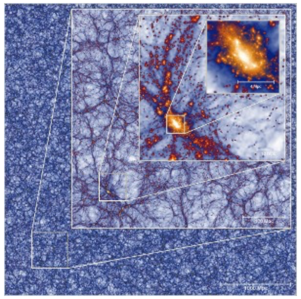
Jeong and Komatsu (2009)

Large Volume Runs

- **Millenium XXL**

  $3.0\ h^{-1}$ Gpc

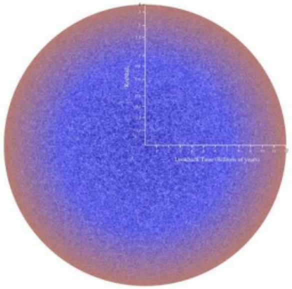
  303 billion particles



- **Horizon Run**

  $10.8\ h^{-1}$ Gpc

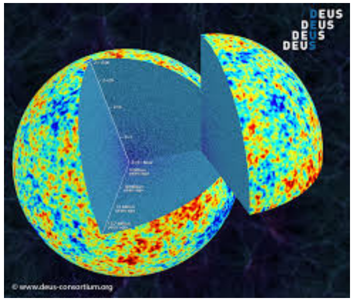
  375 billion particles



- **DEUS FUR**

  $21\ h^{-1}$ Gpc

  550 billion particles



DEUS - Full Universe Runs

- L_{box} = 21 Gpc/h
- $N_p = 8196^3$
- Δx_{coarse} = 40 kpc/h
- $m_p = 10^{12}\ M_{sun}/h$
- RAMSES code
- Models: **LCDM-W7**, RPCDM-W7, WCDM-W7

Curie Thin (80000 cores)

- 1.2 Petabytes of data
- 10 Mhours

Alimi et al. (2012), Proceedings of SC'12, arXiv: 1206.2838

BAO from DEUS-FUR LCDM-7

Statistical & Systematic Errors

- Mass Resolution
- Refinement
- Initial Conditions
- Time Integration

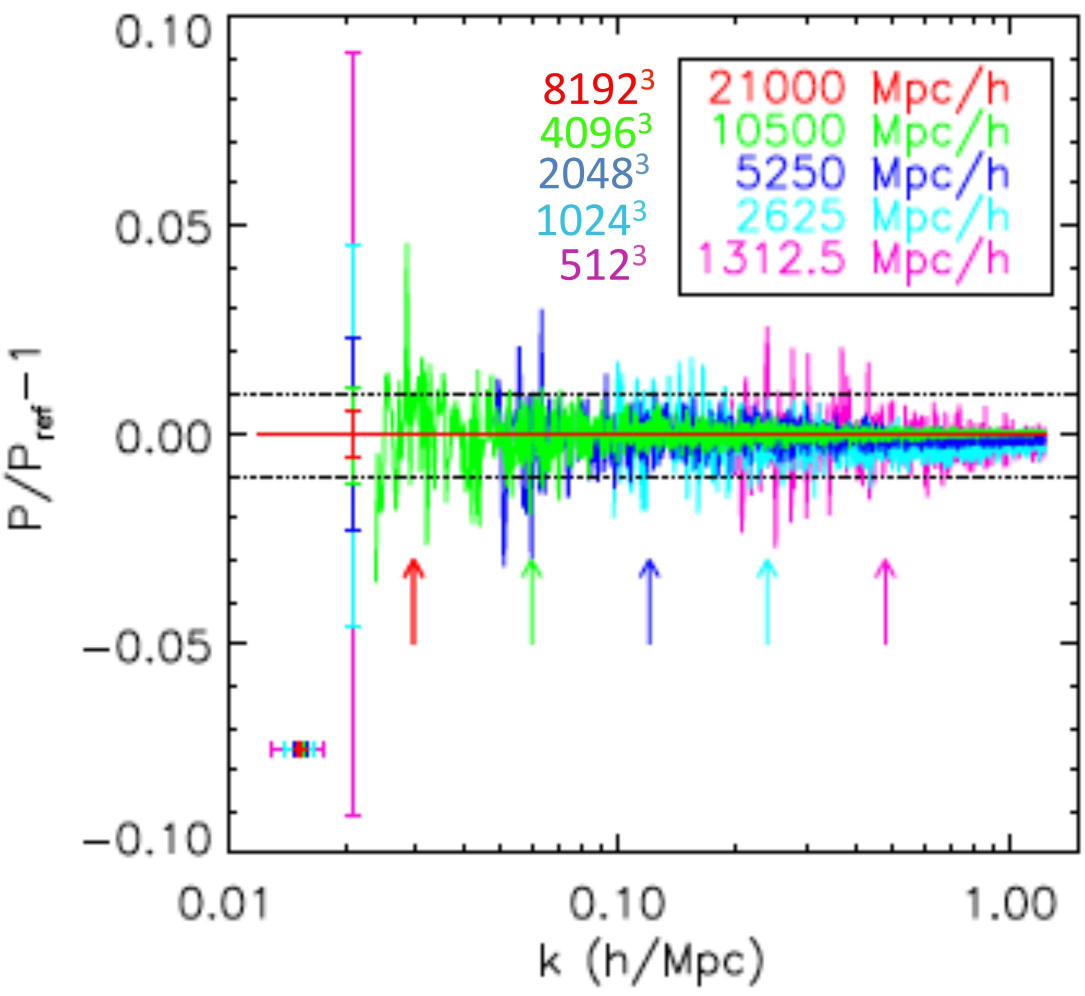
Simulation Suite

L_{box}	n_x	m_{ref}	z_i	C_{dt}
10500	4096	14	106	0.2
5250	2048	14	106	0.2
2625	1024	14	106	0.2
1312	512	14	106	0.2
5250	2048	8	106	0.2
5250	2048	25	106	0.2
5250	2048	14	106	0.08
5250	2048	14	106	0.5
5250	2048	14	272	0.2
5250	2048	14	170	0.2

L_{box}	n_x	m_{ref}	z_i	C_{dt}
5250	2048	14	66	0.2
5250	2048	14	41	0.2
2592	2048	8	56	0.5
2592	1024	8	56	0.5
648	1024	8	93	0.5
648	512	8	93	0.5
648	256	8	93	0.5
2625	1024	14	106*	0.2
5250	2048	14	106**	0.2

Rasera, Corasaniti, Alimi et al. (2014)

Statistical Errors

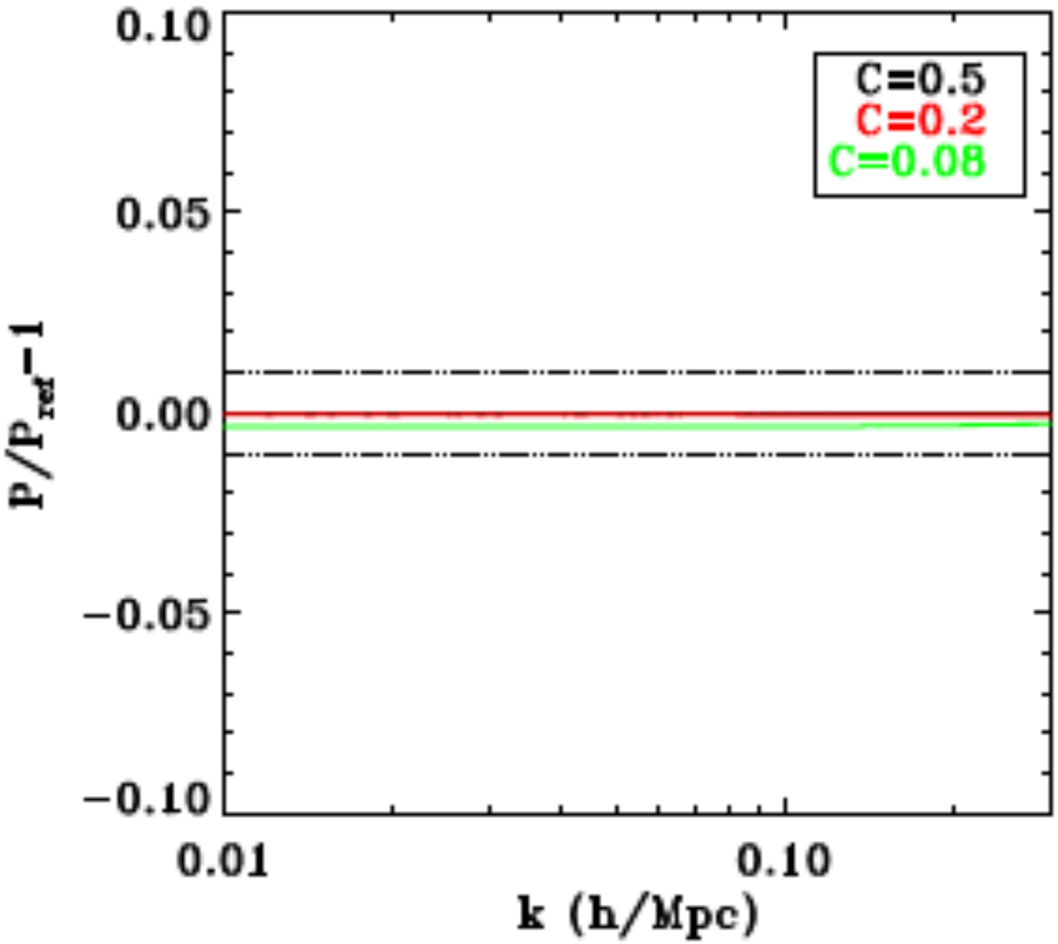


k_{min} setting requirements:

- $E = \sigma_{noise}/P < 1\%$
- $dk/k < 1\%$

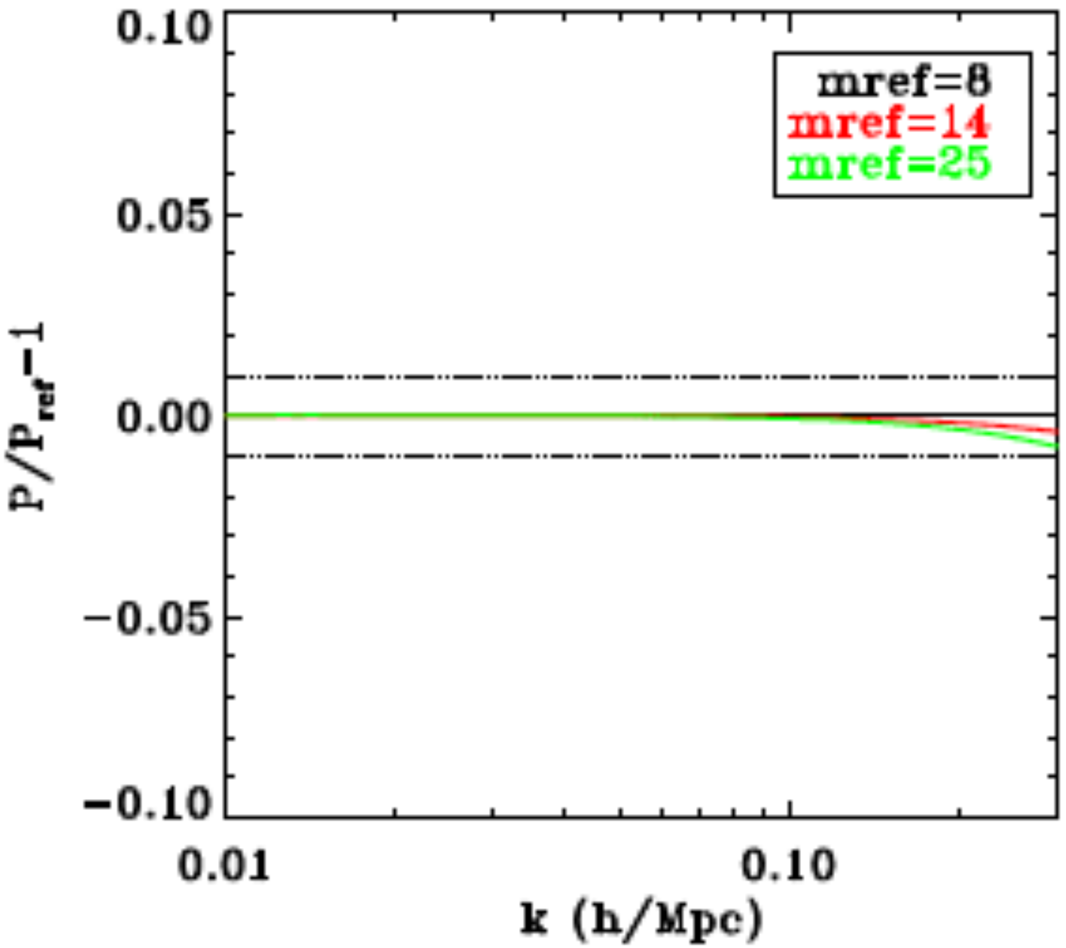
Systematic Errors – Integration dt

L_{box} = 5250 Mpc/h $N_p = 2048^3$



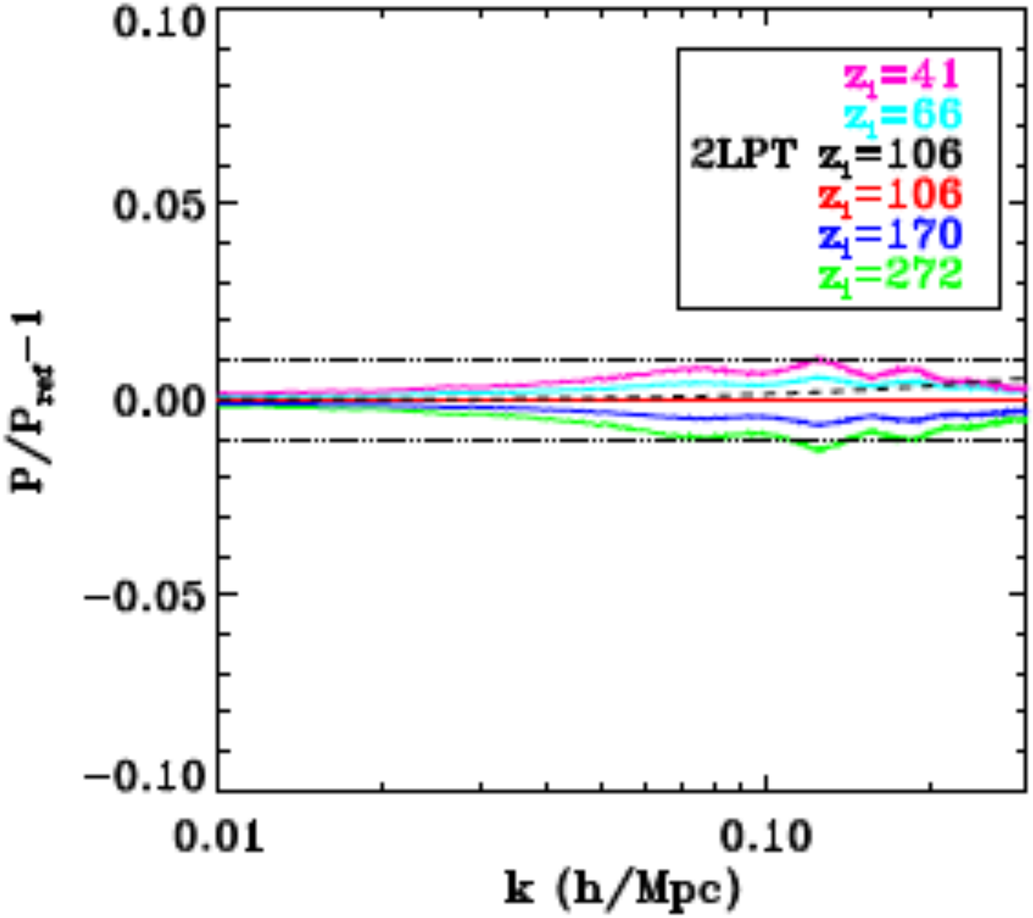
Systematic Errors – Refinement

L_{box} = 5250 Mpc/h N_p = 2048^3

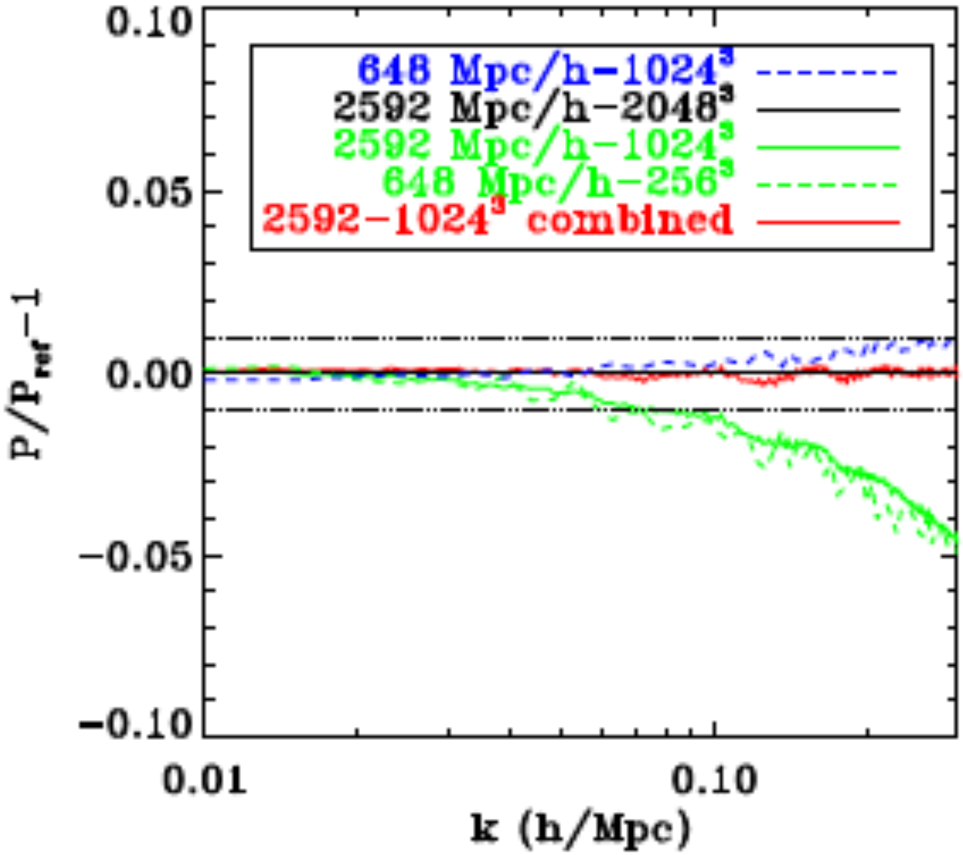


Systematic Errors – IC

L_{box} = 5250 Mpc/h $N_p = 2048^3$



Systematic Errors – Resolution



$m_p = 1.8\ 10^{10}\ h^{-1}\ M_{Sun}$

$m_p = 1.5\ 10^{11}\ h^{-1}\ M_{Sun}$

$m_p = 1.2\ 10^{12}\ h^{-1}\ M_{Sun}$

$$r^{corr}_{poly-fit}(k) = \frac{P_{256^3-648}(k)}{P_{2048^3-2592}(k)}$$

$$P^{corr}_{DEUS-FUR}(k) = \frac{\hat{P}_{DEUS-FUR}(k)}{r^{corr}_{poly-fit}(k)}$$

- 8% drop at k=0.3 for ~3 $10^{12}\ h^{-1}\ M_{sun}$ (Heitmann et al. 2010)
- Accumulated force resolution error during PM computation (linear regime)
- ~8 cells per particle, Zeldovich wave test (Knebe, Green & Binney 2001)

BAO Spectrum at 1%

Interval Range $0.03 < k < 0.3$

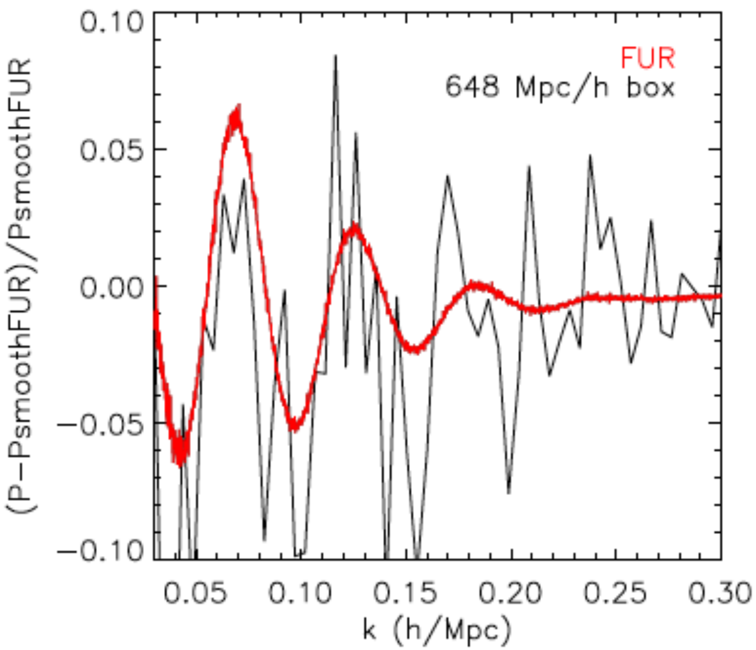
Wiggle-free Spectrum

- Linear wiggle-free ? NL erase BAO at high-end interval
- Polynomial fit ?
- NL evolved wiggle-free initial spectrum

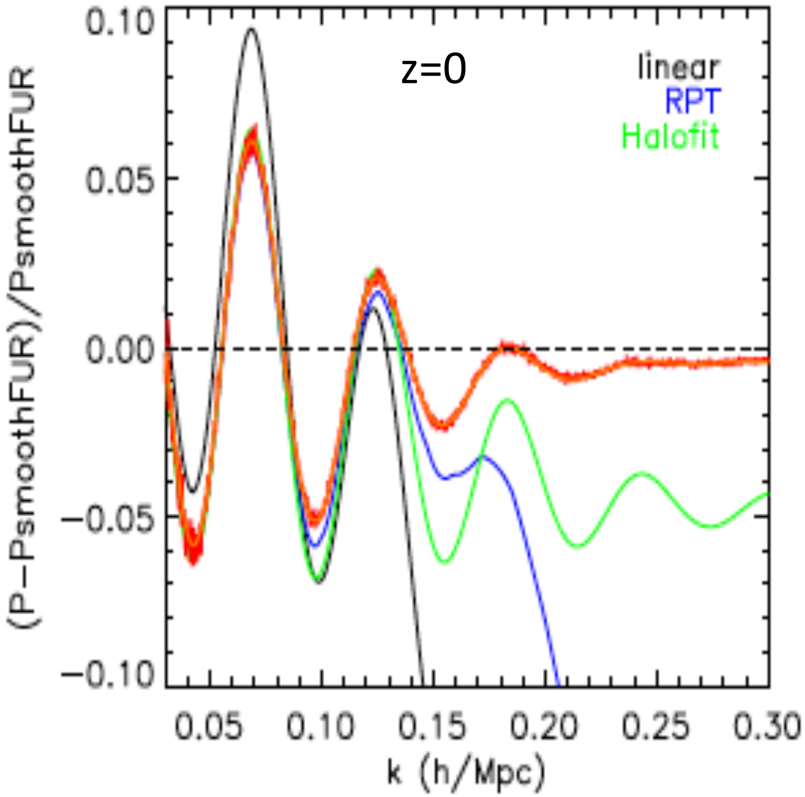
Crocce & Scoccimarro (2008)

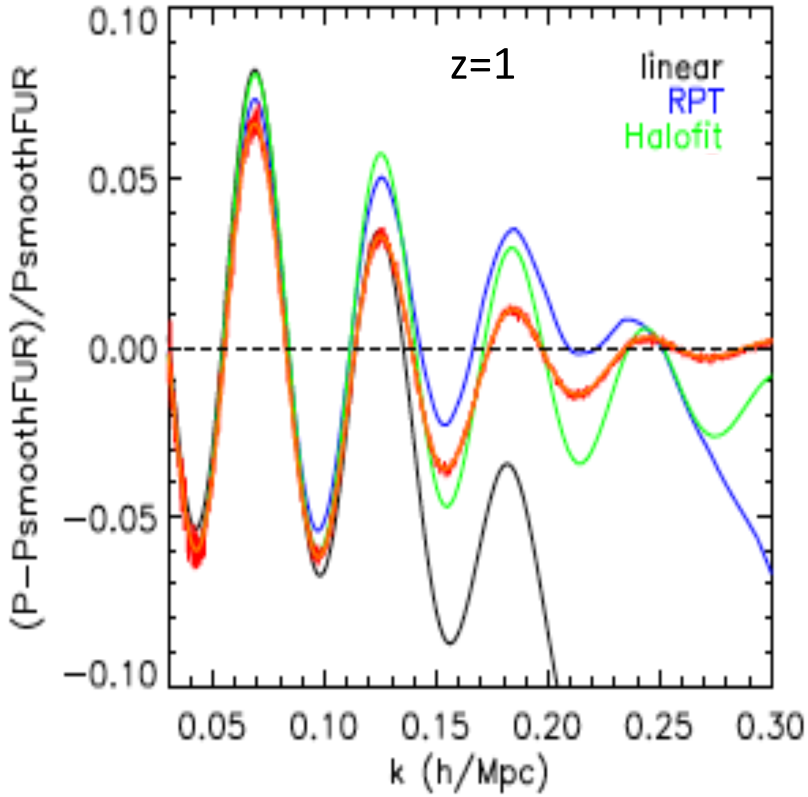
N-body evolved Hu & Eisensten
wiggle-free Initial Conditions

$$P^{BAO}_{DEUS-FUR}(k) = P^{corr}_{DEUS-FUR}(k) - P_{smooth}(k)$$



BAO Spectrum Benchmark



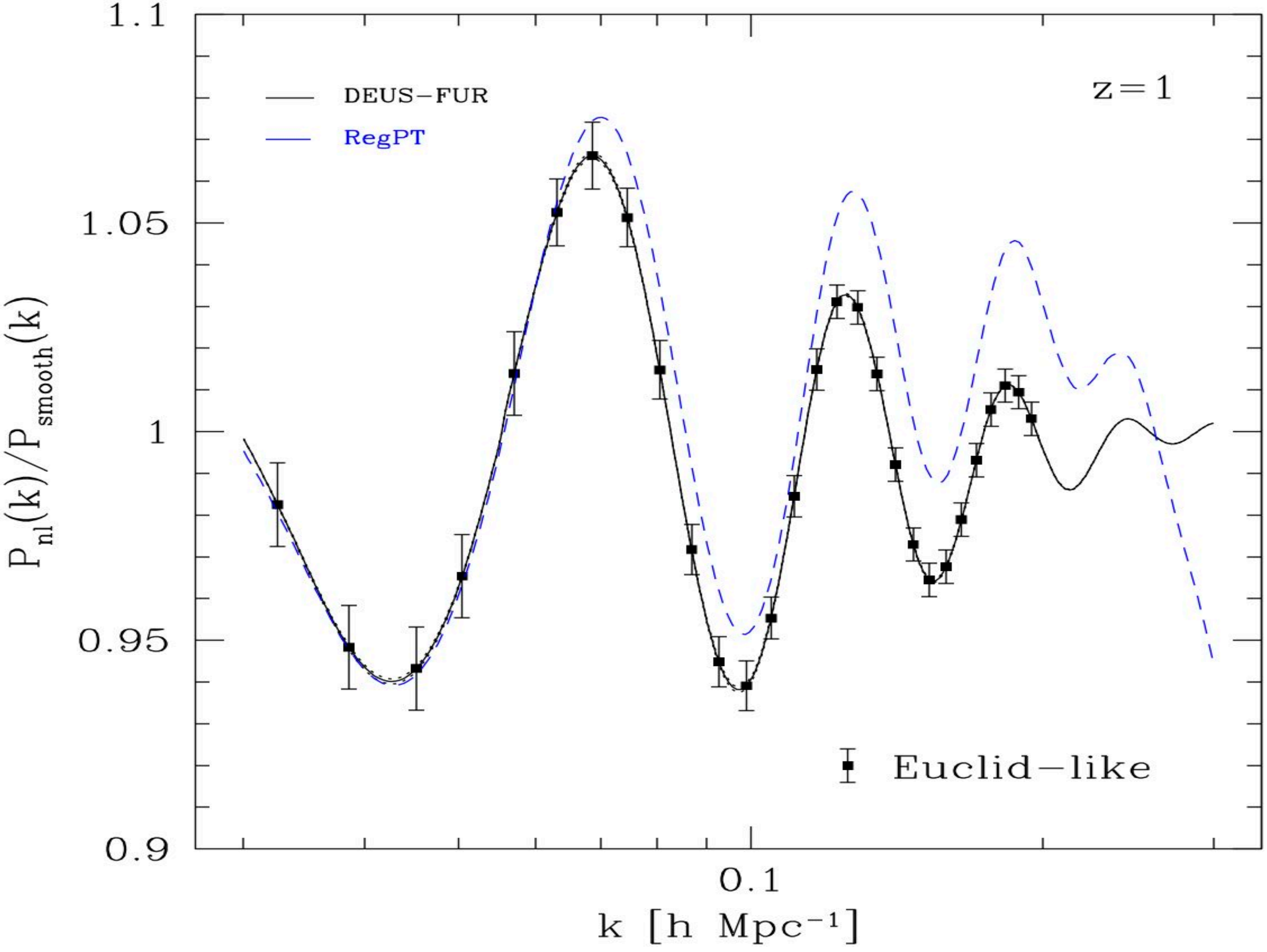


RegPT – Taruya et al. (2012)

Halofit – Smith et al. (2003,2014)

- Location of BAO extrema <1%
- Amplitude difference >1% at k>0.1 and z=1

Implications for Euclid Data Analysis



Power Spectrum Covariance

Non-Linear Contribution

$$\mathrm{cov}(k_1,k_2) = \frac{2}{N_{k_1}} P^2(k_1)\delta_{k_1k_2} + \frac{1}{V}\int_{\Delta_{k_1}}\int_{\Delta_{k_2}} \frac{d^3k'_1}{V_{k_1}}\frac{d^3k'_2}{V_{k_2}}\, T(k'_1,-k'_1,k'_2,-k'_2)$$

- Non-linear regime sources non-Gaussianity ($T \neq 0$)
- Fully analytical trispectrum is not viable (several models on the market still require simulation input)

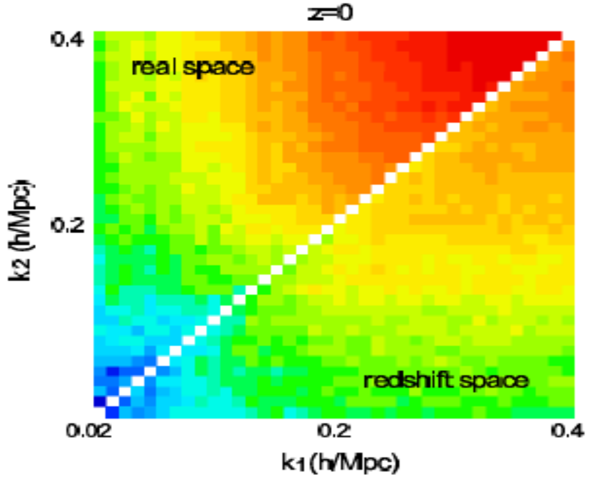
Sampling N-body Ensemble

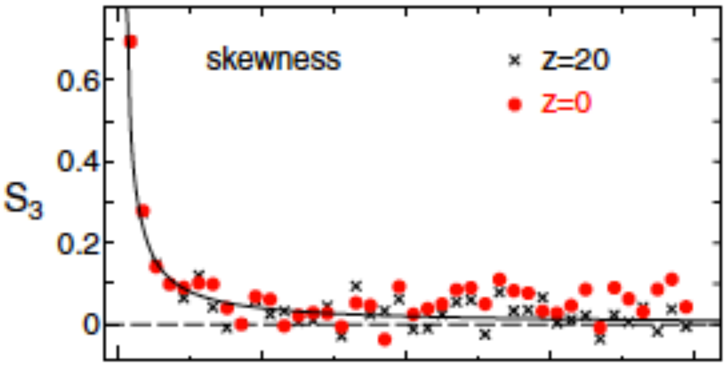
$$\mathrm{cov}(k_1,k_2) = \frac{1}{N_r - 1}\sum_{i=1}^{N_r}\left[\hat{P}_i(k_1) - \overline{P}(k_1)\right]\left[\hat{P}_i(k_2) - \overline{P}(k_2)\right] \qquad \overline{P}(k) = \frac{1}{N_r}\sum_{i=1}^{N_r}\hat{P}_i(k)$$

Previous Studies

Power Spectrum Statistics

- $N_r = 5000$ N-body PM simulations
- $L_{box} = 1$ Gpc h^{-1} & $N_p=256^3$ ($m_p = 4.1 \times 10^{12}\ h^{-1}\ M_\odot$)





Takahashi et al. (2014)

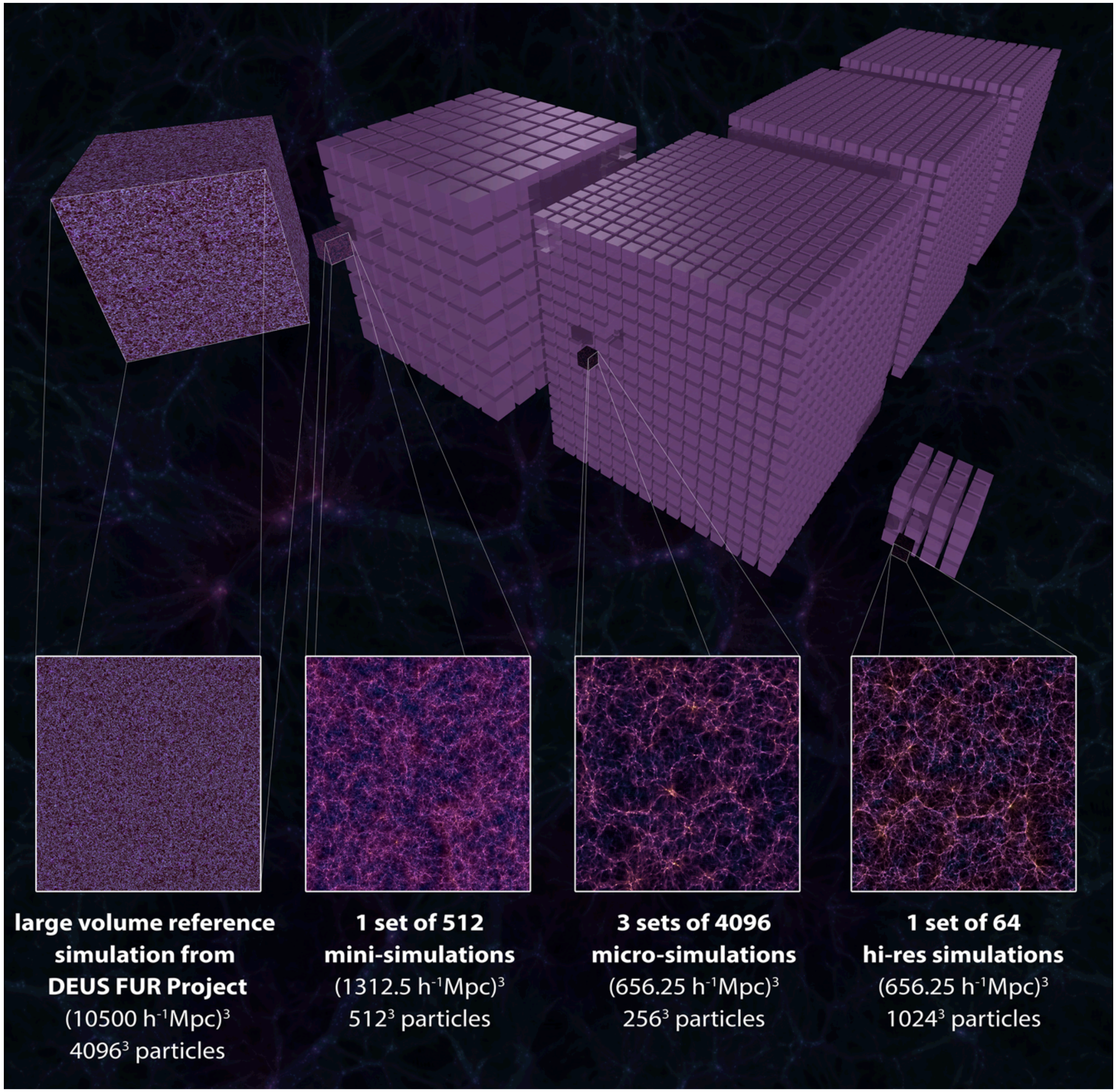
- Not conclusive (still large statistical uncertainties)

DEUS-Parallel Universe Runs

N_r = 12288 Simulations
N_p=256^3 L_{box}=648 Mpc/h
m_p= 1.2×10^{12} M_{sun}

N_r = 512 Simulations
N_p=512^3 L_{box}=1.3 Gpc/h
m_p= 1.2×10^{12} M_{sun}

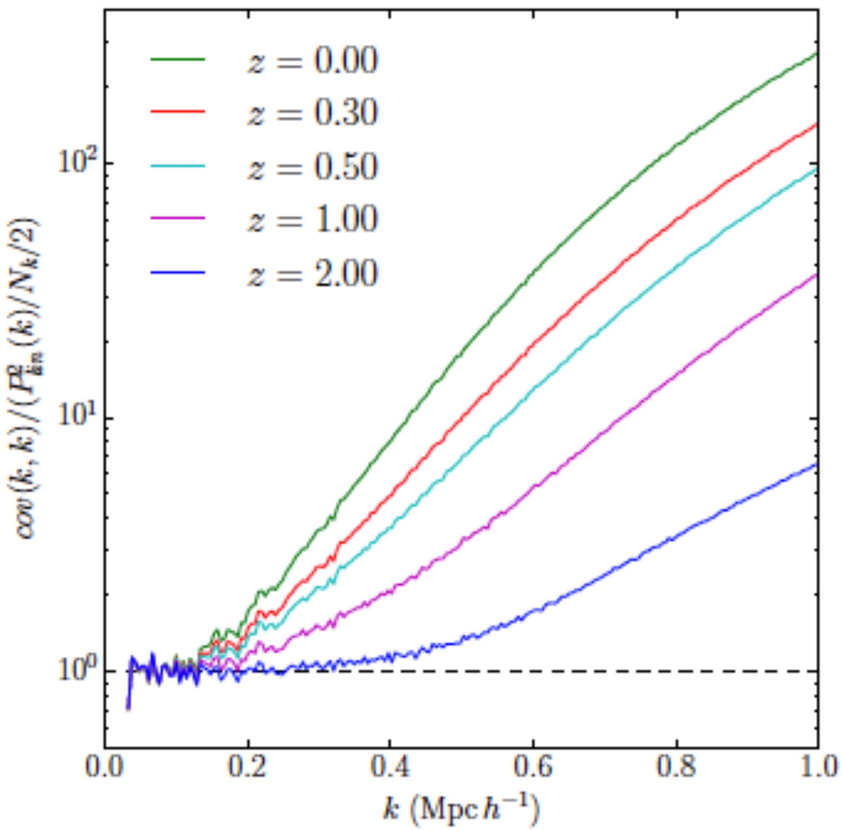
N_r = 96 Simulations
N_p=1024^3 L_{box}=648 Mpc/h
m_p= 1.8×10^{10} M_{sun}



DEUS-PUR Covariance

Blot, Corasaniti, Alimi, Reverdy, Rasera (2015)

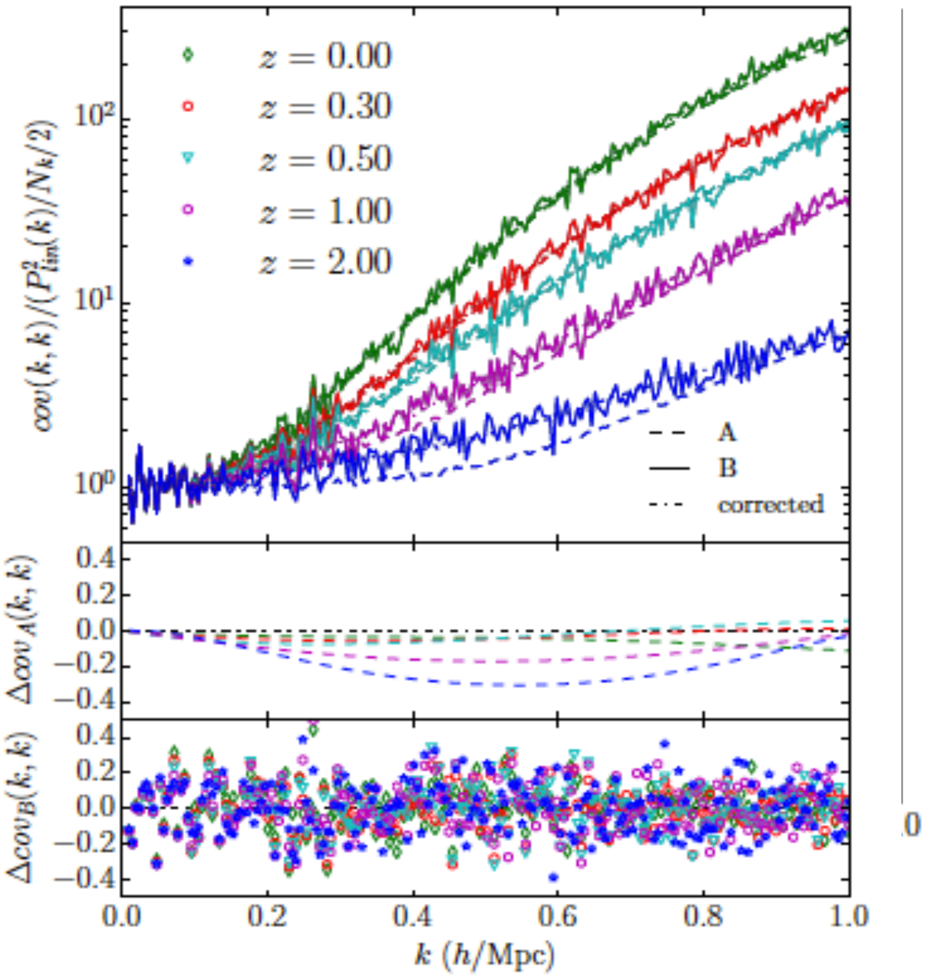
Diagonal Components



- Set A
- Deviations from linear prediction grow towards low redshifts
- The onset of the deviation moves towards lower k at lower redshifts

Numerical Systematics

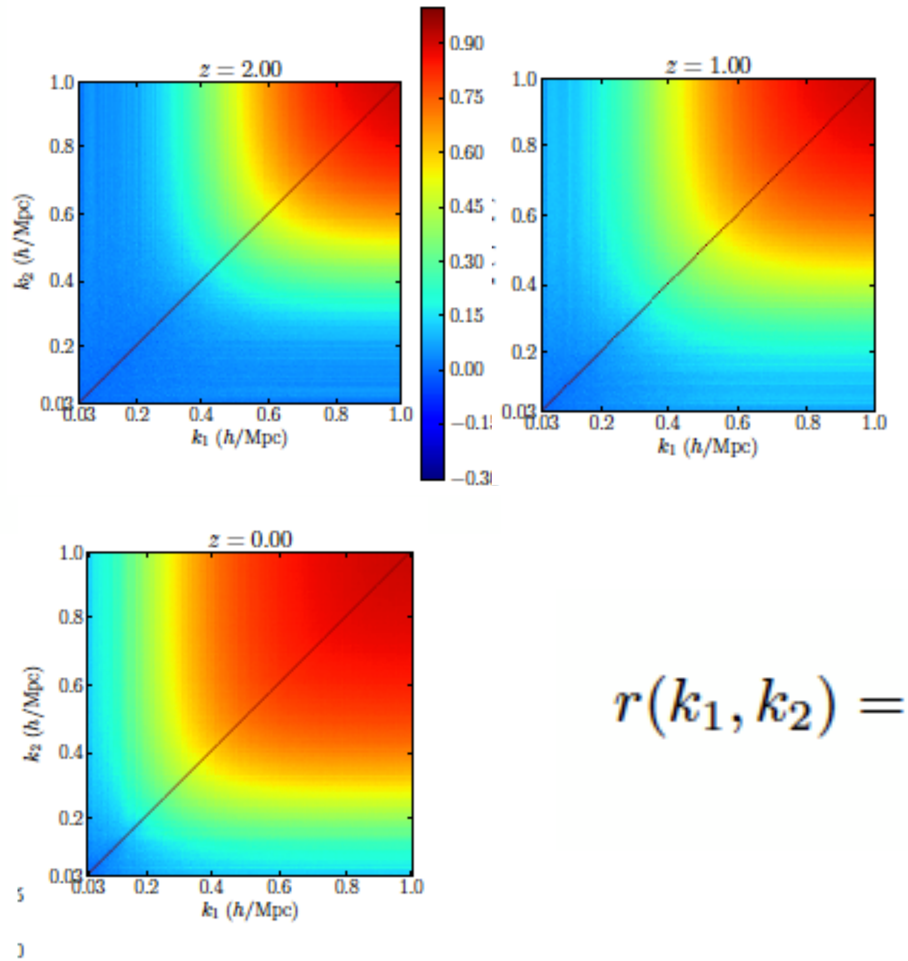
Mass Resolution Errors



- Set A vs Set B
- At intermediate scales lower resolution leads to lower covariance
- Discrepancy decrease with z and within statistical noise for z<0.5
- PM effect on trispectrum, alleviated by refinement
- Corrections

$$\hat{P}_A^{\mathrm{corr}}(k) = \left[\hat{P}_A(k) - \bar{P}_A(k)\right] \frac{\sigma_{\hat{P}_B}(k)}{\sigma_{\hat{P}_A}(k)} + \bar{P}_B(k)$$

Correlation Matrix



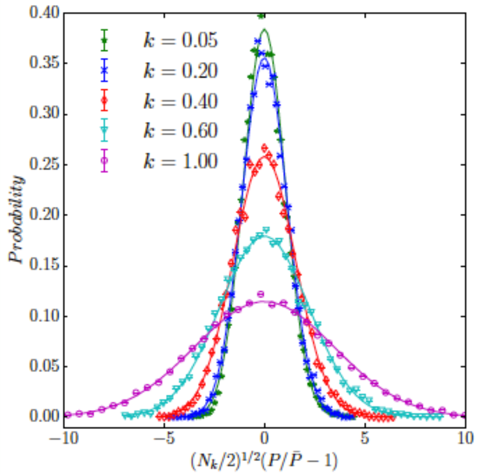
- Non-negligible mode correlations on BAO scales

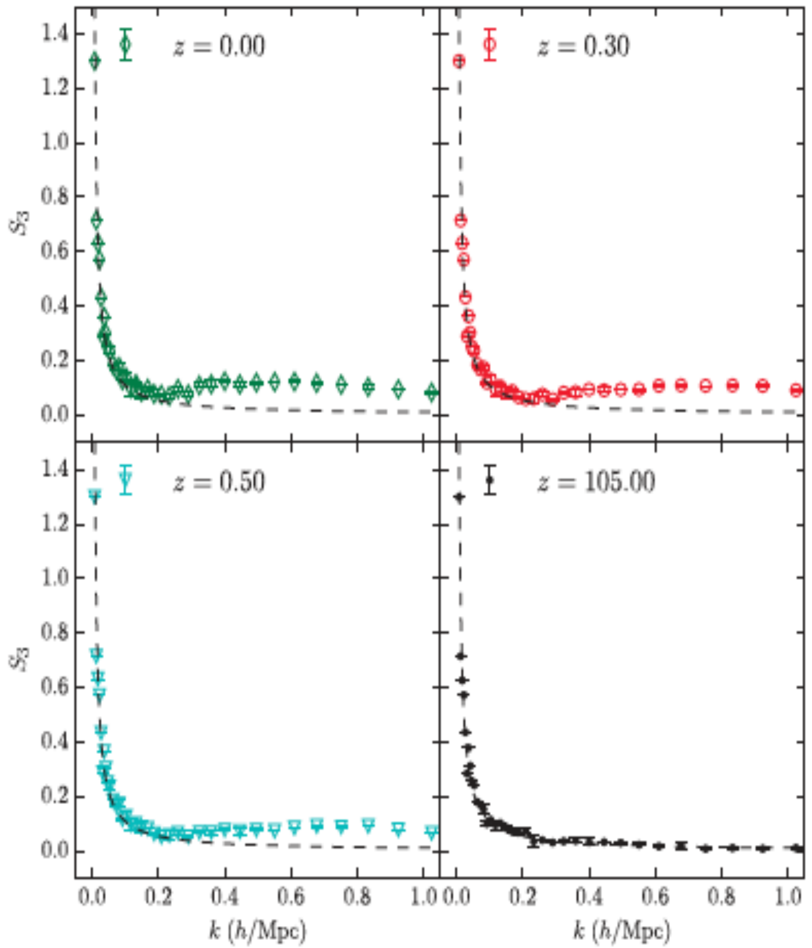
$$r(k_1, k_2) = \frac{\mathrm{cov}(k_1, k_2)}{\sqrt{\mathrm{cov}(k_1, k_1)\,\mathrm{cov}(k_2, k_2)}}$$

Power Spectrum Distribution

Deviations from χ^2 statistics

- P(k) of Gaussian density field is χ^2-distributed





Impact on cosmo params inference of 10% deviation from Gaussian likelihood at k>0.2?

Errors on Covariance Estimation

- Sample Covariance in Clustering Analyses
- Propagation of Sample Variance Errors on Covariance
- Gaussian density field covariance: *Wishart Distribution*

$$p(\hat{\mathcal{C}}|\mathcal{C},\nu,\mu) = \left(\frac{\nu^{\nu\mu/2}|\mathcal{C}|^{-\nu/2}|\hat{\mathcal{C}}|^{(\nu-\mu-1)/2}}{2^{\nu\mu/2}\Gamma_\mu[\nu/2]}\right)\exp^{-\frac{\nu}{2}\mathrm{Tr}\hat{\mathcal{C}}\mathcal{C}^{-1}}$$

- Biased Estimator Precision Matrix: $\langle\widehat{\mathcal{P}}_{ij}\rangle = \frac{N_s-1}{N_s-N_d-2}\mathcal{P}_{ij}$ Press (1982)
- Error scaling:

$$\sigma^2(\widehat{\mathcal{C}}_{ij}) = \frac{1}{N_s-1}\left(\mathcal{C}_{ij}^2+\mathcal{C}_{ii}\mathcal{C}_{jj}\right)$$

Taylor, Joachimi & Kitching (2013)

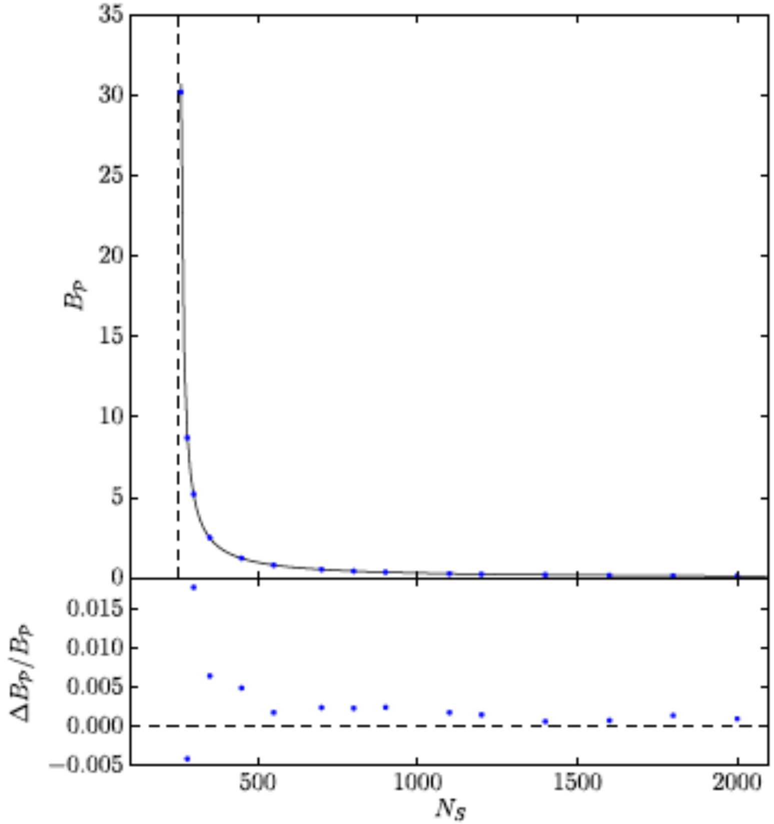
$$\sigma^2(\widehat{\mathcal{P}}_{ij}) = A\left[(N_s-N_d)\mathcal{P}_{ij}^2+(N_s-N_d-2)\mathcal{P}_{ii}\mathcal{P}_{jj}\right]$$

What is the impact of non-linearities on covariance errors?

Testing Sampling Errors

Blot, Corasaniti, Amendola, Kitching (2016)

Unbiased Precision Matrix Estimator



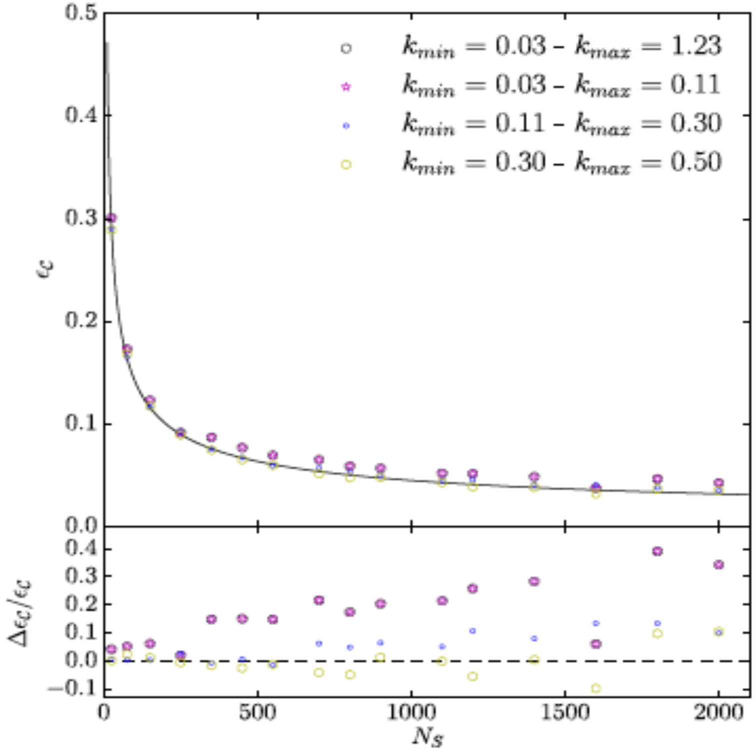
- Divide DEUS-PUR in N_g = int (N_t/N_s)
- Compute <..>=1/N_g Sum

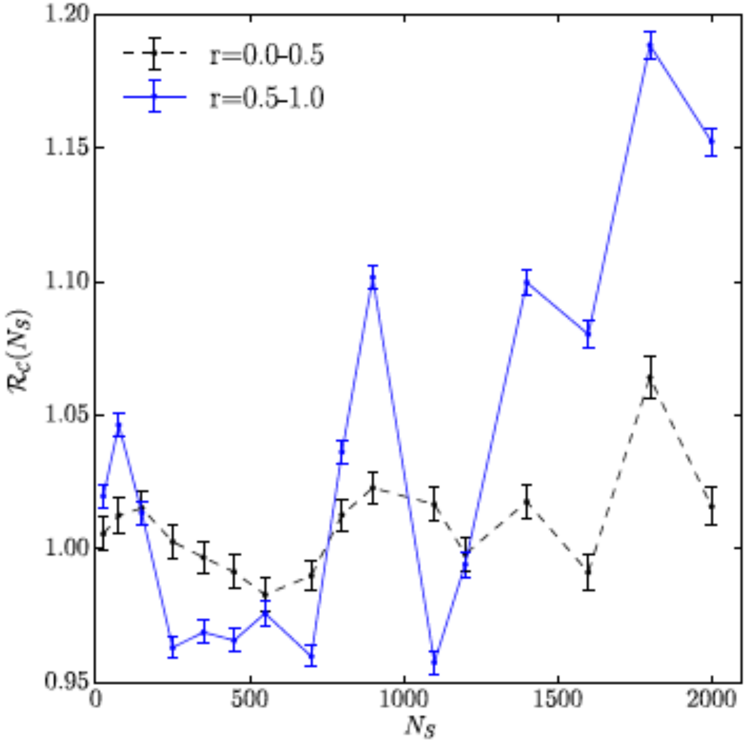
$$B_\mathcal{P} \equiv \frac{\mathrm{Tr}\,\langle\widehat{\mathcal{P}}\rangle - \mathrm{Tr}\,\mathcal{P}}{\mathrm{Tr}\,\mathcal{P}} = \frac{N_s - 1}{N_s - N_d - 2} - 1$$

Unbiased Estimator Precision Matrix not affected by non-linearities

Covariance Sampling Errors

Error Scaling $$\epsilon_{\mathcal{C}} = \sqrt{\frac{\sum_i \sigma^2(\widehat{\mathcal{C}}_{ii})}{\sum_i \langle \widehat{\mathcal{C}}_{ii} \rangle^2}} = \sqrt{\frac{2}{N_s - 1}}, \quad \mathcal{R}_{\mathcal{C}} = \frac{(N_s - 1) \sum_{ij} \sum_{m=1}^{N_g} \left(\widehat{\mathcal{C}}_{ij,m} - \langle \widehat{\mathcal{C}}_{ij} \rangle \right)^2}{(N_g - 1) \sum_{ij} \left(\langle \widehat{\mathcal{C}}_{ij} \rangle^2 + \langle \widehat{\mathcal{C}}_{ii} \rangle \langle \widehat{\mathcal{C}}_{jj} \rangle \right)}$$



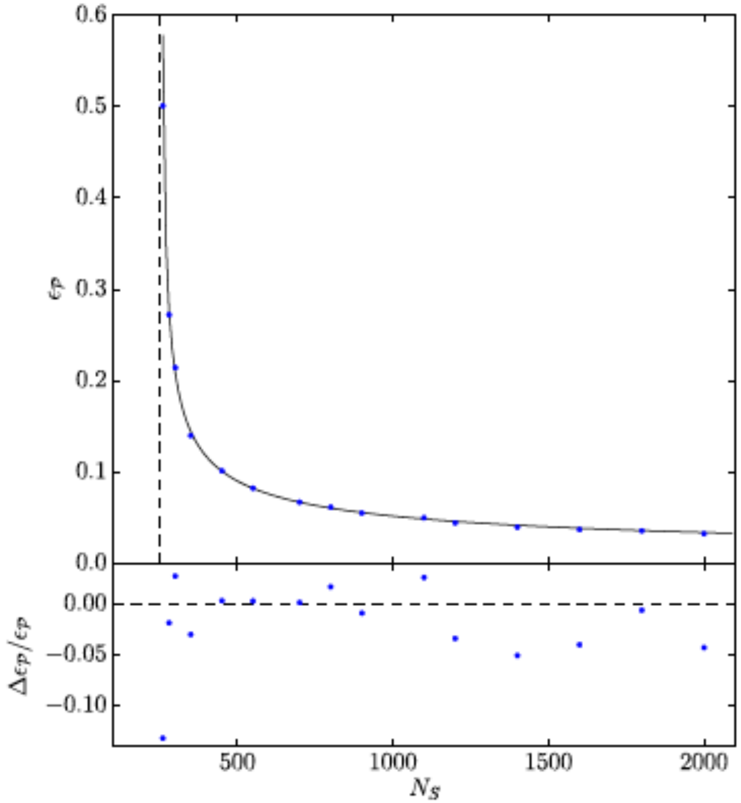


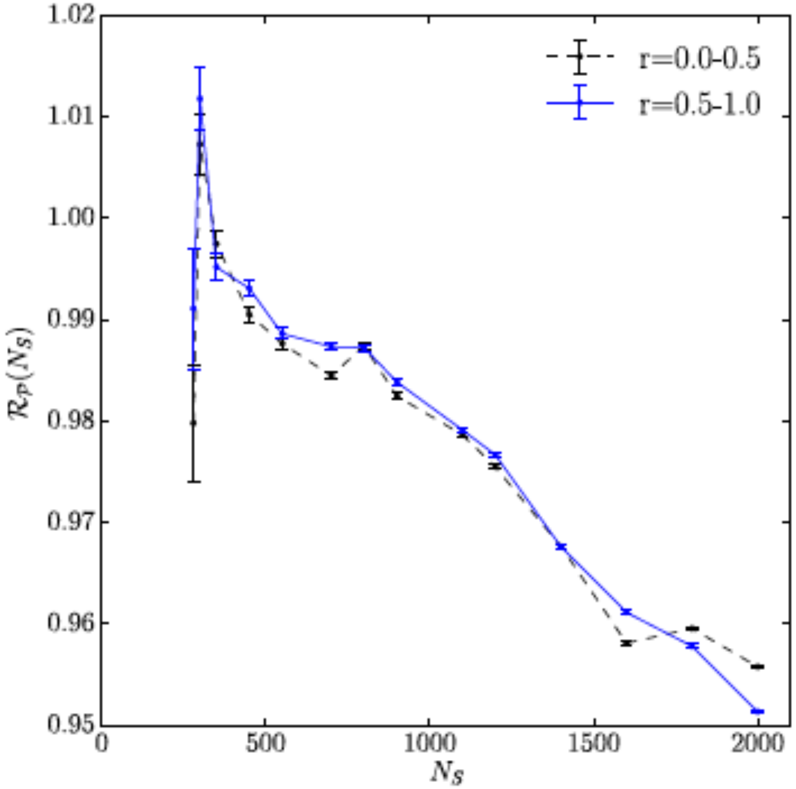
- *On BAO scales Gaussian prediction to within 10%*

Precision Sampling Errors

Error Scaling

$$\epsilon_{\mathcal{P}} = \sqrt{\frac{\sum_i \sigma^2(\widehat{\mathcal{P}}_{ii})}{\sum_i \langle \widehat{\mathcal{P}}_{ii} \rangle^2}} = \sqrt{\frac{2}{N_s - N_d - 4}},$$





Fisher Forecast – Euclid-like Survey

Cosmological Parameters $\theta = \{\Omega_m, w, \sigma_8, n_s, \Omega_b, b_1, .., b_{N_z}\}$

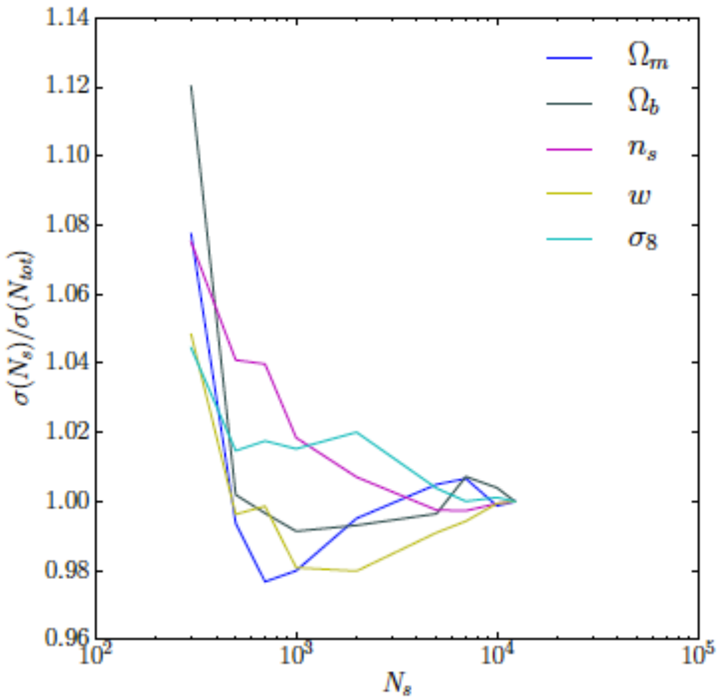
$$P_z^g = b_z^2 P_z + \frac{1}{\bar{n}_g(z)} \qquad \mathrm{cov}_z^g(k_i, k_j) = b_z^4 \hat{\mathcal{C}}_{ij} + 2b_z^2 [P_z(k_i) P_z(k_j)]^{1/2} \bar{n}_g^{-1}(z) + \bar{n}_g^{-2}(z)$$

- 5 bins 0.5 < z < 2
- Marginalized over constant bias b_i

z	$\bar{n}_g(z)$
0.5	4.2×10^{-3}
0.7	2.99×10^{-3}
1.0	1.81×10^{-3}
1.5	0.77×10^{-3}
2.0	0.15×10^{-3}

$$F_{\alpha\beta} = \sum_{l=1}^{N_z} \sum_{i,j=1}^{N_d} \frac{\partial P_{z_l}^g}{\partial \theta_\alpha}(k_i) \frac{\partial P_{z_l}^g}{\partial \theta_\beta}(k_j) \mathrm{cov}_{z_l}^g(k_i, k_j),$$

- Errors convergence for large N_r
- For N_r > 5000 Covariance Estimation Errors <1%



Conclusions

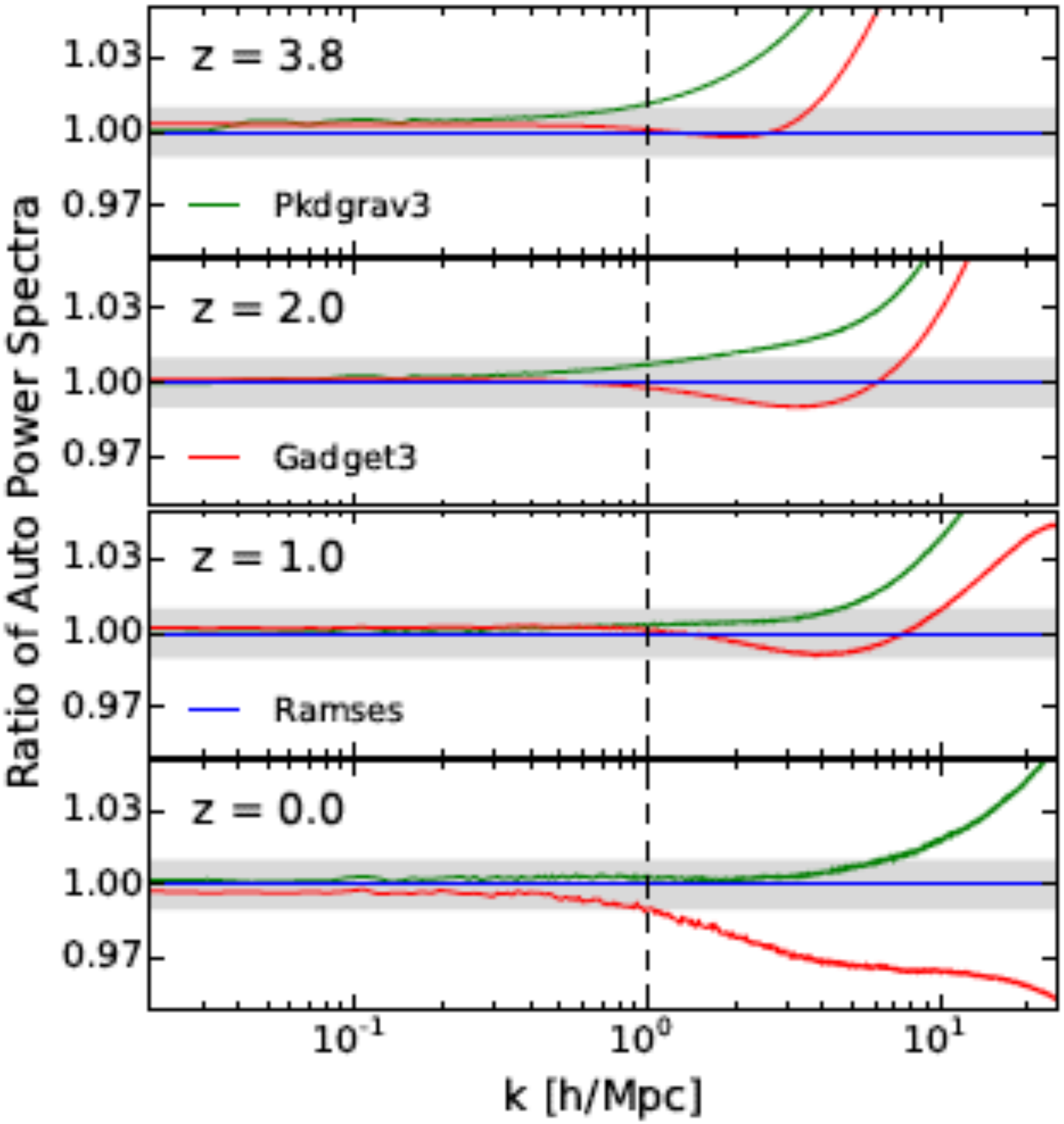
- Accounting for non-linearities at 1% level requires large volume high-resolution simulations for the clustering and large simulation ensembles for convariance
- Realization of benchmarks is challenging, but necessary and needs to be done only once
- More systematic comparison against benchmarks, of approximate numerical methods and semi-analytic approaches

Physicality of Newtonian Simulations

GR Effects (?)

- Newtonian gravity at dozens Gpc/h scales?
- On large scales at leading order in Φ, GR particle-trajectories = ZA which is Newtonian solution at leading order
  Chisari & Zaldarriaga (2011)
- Purely relativistic effects only at 2nd order in Φ > Newtonian terms at 2nd order, but suppressed compared to ZA displacements
- Corrections to $P_{lin}(k)$ are $O(10^{-5})$ at z=49 and $O(10^{-3})$ at z=0
  Rigopoulos & Valkenburg (2015)
- Confirmed by Relativistic N-body Simulations
  Adamek et al. (2016)

Different N-body codes



Schneider et al. (2016)

Fisher Forecast – Euclid-like Survey

Effects on Model Parameter Errors

$$P_z^g = b_z^2 P_z + \frac{1}{\bar{n}_g(z)}$$

$$\theta = \{\Omega_m, w, \sigma_8, n_s, \Omega_b, b_1, .., b_{N_z}\}$$

z	$\bar{n}_g(z)$
0.5	4.2×10^{-3}
0.7	2.99×10^{-3}
1.0	1.81×10^{-3}
1.5	0.77×10^{-3}
2.0	0.15×10^{-3}