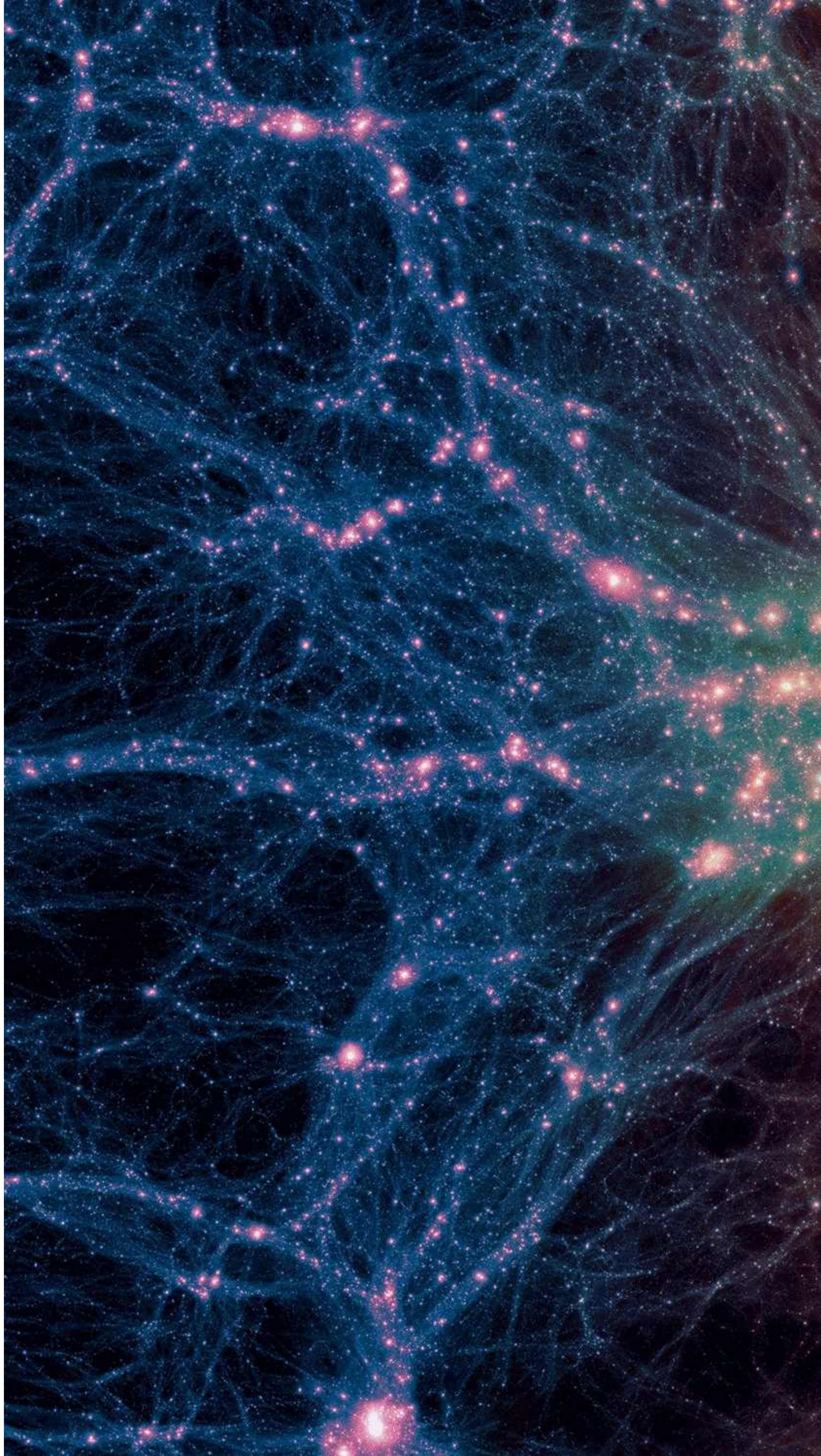
COSMOLOGICAL SIMULATIONS FOR LARGE GALAXY SURVEYS

Linda Blot

INSTITUT DE CIÈNCIES DE L'ESPAI
CSIC IEEC



LARGE SCALE STRUCTURE SURVEYS: WHY WE NEED SIMULATIONS

- Non-linear regime of structure collapse
- Bayesian data analysis:

$$P(\boldsymbol{\theta}|\mathbf{D}, M) = \frac{P(\mathbf{D}|\boldsymbol{\theta}, M)\, P(\boldsymbol{\theta}|M)}{P(\mathbf{D}|M)}$$

$\boldsymbol{\theta}$ parameters

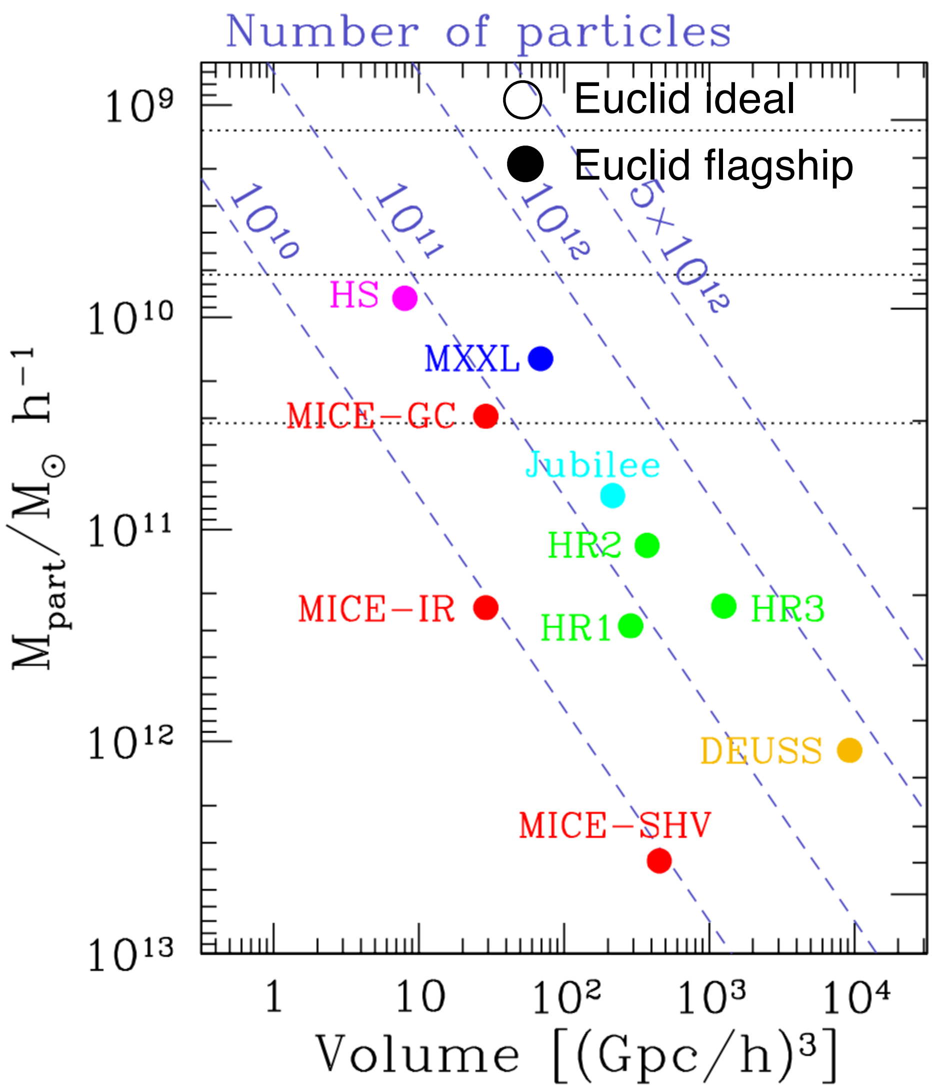
$\mathbf{D}$ data

M model

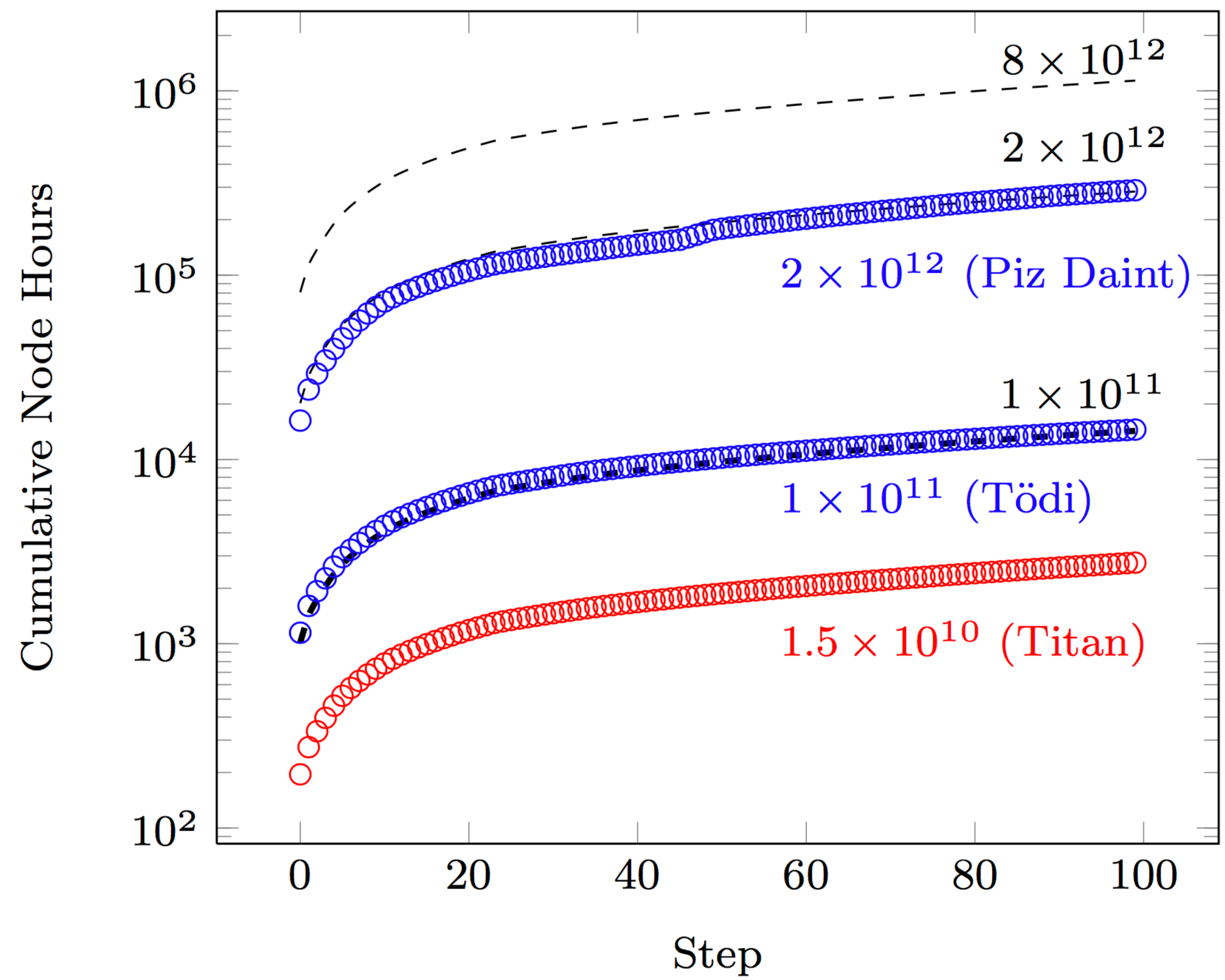
$$P(\mathbf{D}|\theta, M) \propto \exp\left(-\frac{1}{2}\sum_{ij}(D_i - D_M)cov_{ij}^{-1}(D_j - D_M)\right)$$

- Models need simulations to calibrate/validate
- Galaxy mock catalogues to understand observational systematics

COSMOLOGICAL SIMULATIONS: COMPUTATIONAL COST



Fosalba et al. 2015



Potter, Stadel, Teyssier 2016

Flagship used the entire supercomputer, at the time in the top 10 in computing power

COSMOLOGICAL SIMULATIONS: COMPUTATIONAL COST

- 40 Billion halos, 5.5 Tb of [illegible] only for halos in the LC!
- With no observational cuts th[illegible]late into 20 Tb and 60 Billion galaxies, each with more th[illegible]n 200 properties

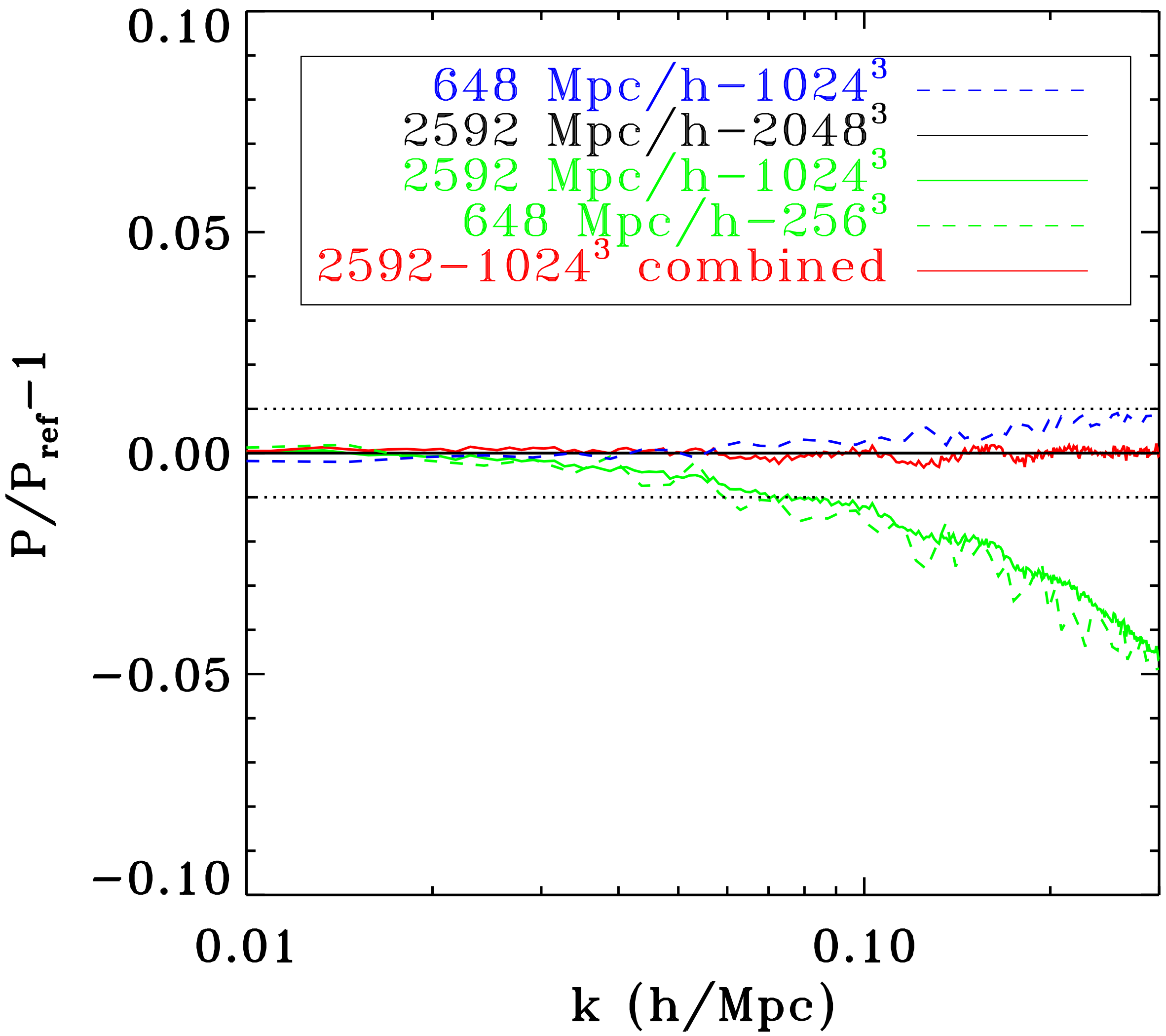
Big data

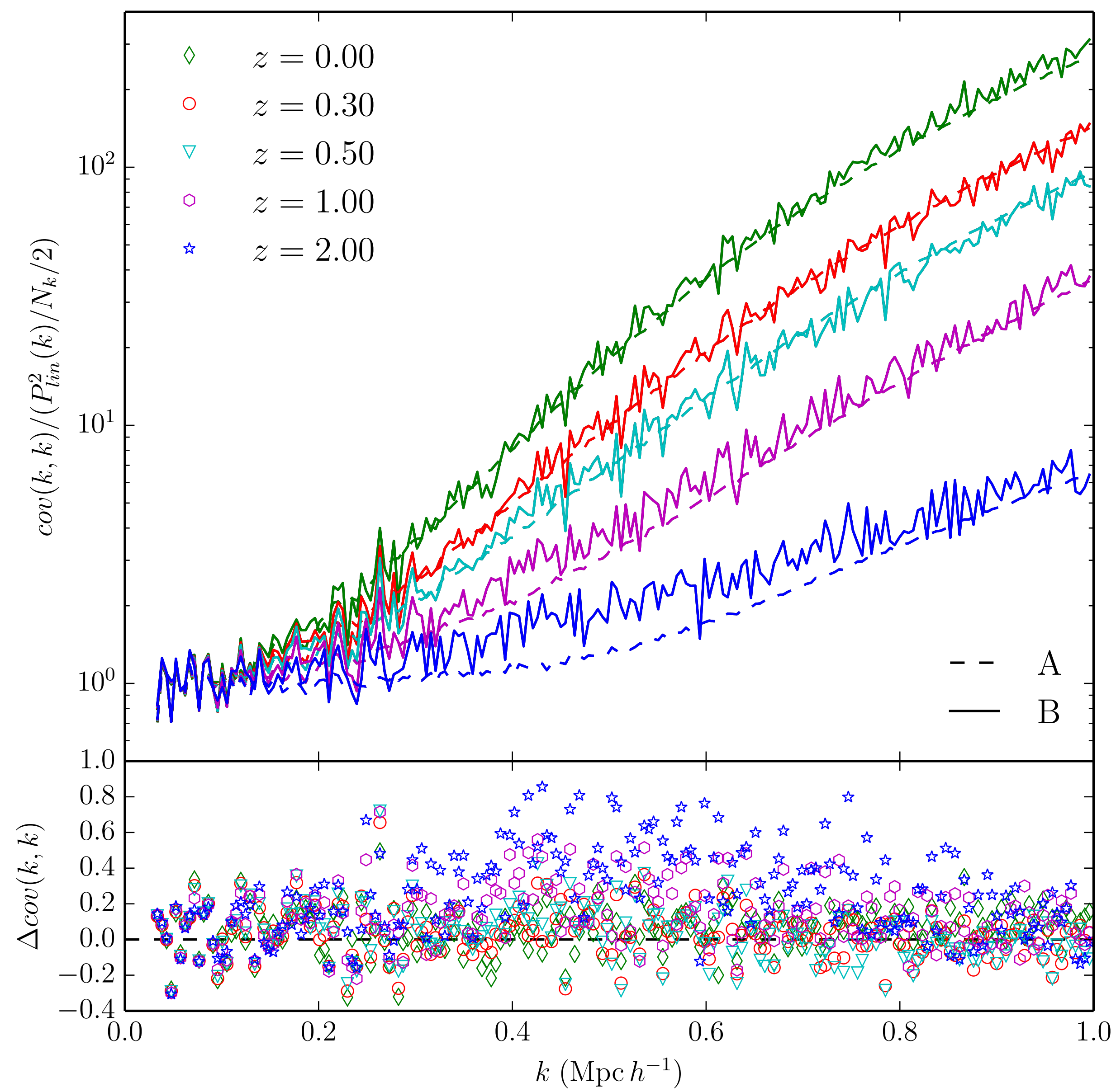
COSMOLOGICAL SIMULATIONS: COMPUTATIONAL COST

- 40 Billion halos, 5.5 Tb of [illegible] only for halos in the LC!
- With no observational cuts [illegible] ate into 20 Tb and 60 Billion galaxies, each with more [illegible] n 200 properties
- Required a paradigm shift:
  - Challenging for standard relational databases ⇒ big data platform
  - Set of python codes optimised to work on this platform

Big data

NUMERICAL SYSTEMATICS





Rasera et al. 2014

Blot et al. 2015

APPROXIMATE METHODS FOR COVARIANCE ESTIMATION

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

- Sample covariance estimator

$$\widehat{\mathrm{cov}}(k_1,k_2) = \frac{1}{N_s-1}\sum_{i=1}^{N_s}[\hat{P}_i(k_1)-\bar{P}(k_1)][\hat{P}_i(k_2)-\bar{P}(k_2)]$$

- Additional error on parameters coming from covariance error
- Need a few 1000s simulations to reduce this error to acceptable levels for future surveys
- Not possible with full N-body simulations (too costly)

APPROXIMATE METHODS FOR COVARIANCE ESTIMATION

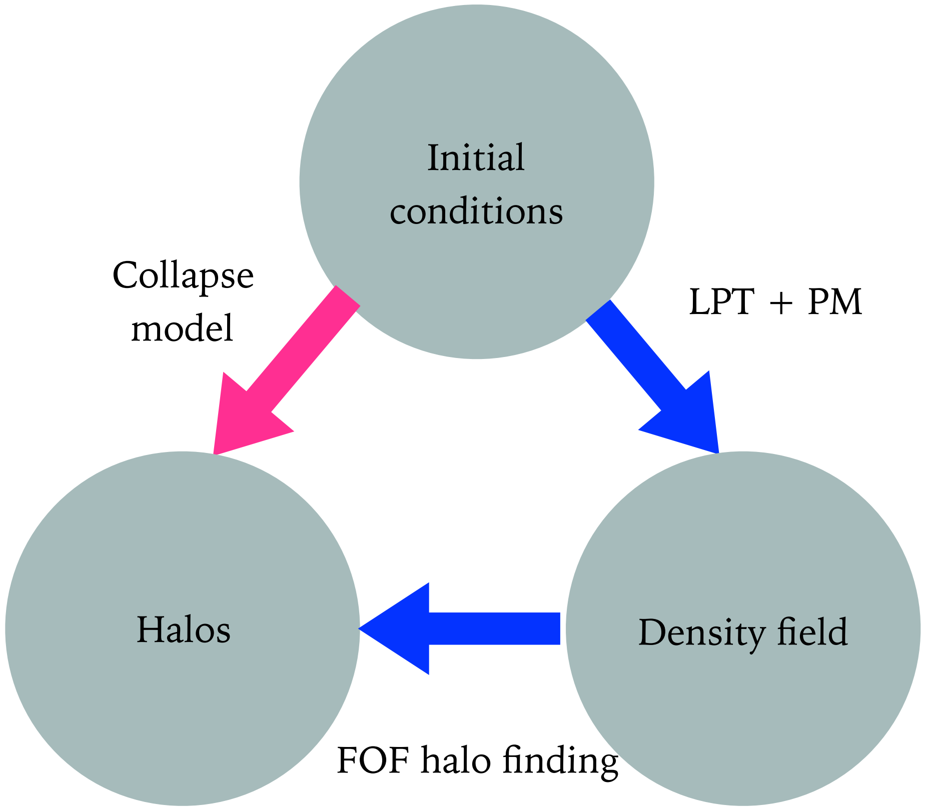
- Benchmark: Minerva simulations
  - 300 realisations
  - L_{box}=1.5 Gpc/h
  - $m_p=2.67 \times 10^{11}\ M_{\odot}/h$
- Approximate methods: matching ICs
- Measurements: 2-pt correlation function multipoles and wedges, power spectrum multipoles, bispectrum real space and monopole
- Propagate errors to cosmological parameters through likelihood analysis

COMPARED METHODS

- One fast PM method: ICE-COLA
- Two predictive methods: Pinocchio, PeakPatch
- Two calibrated methods: Patchy, Halogen
- Two density PDF assumptions: Gaussian, Lognormal

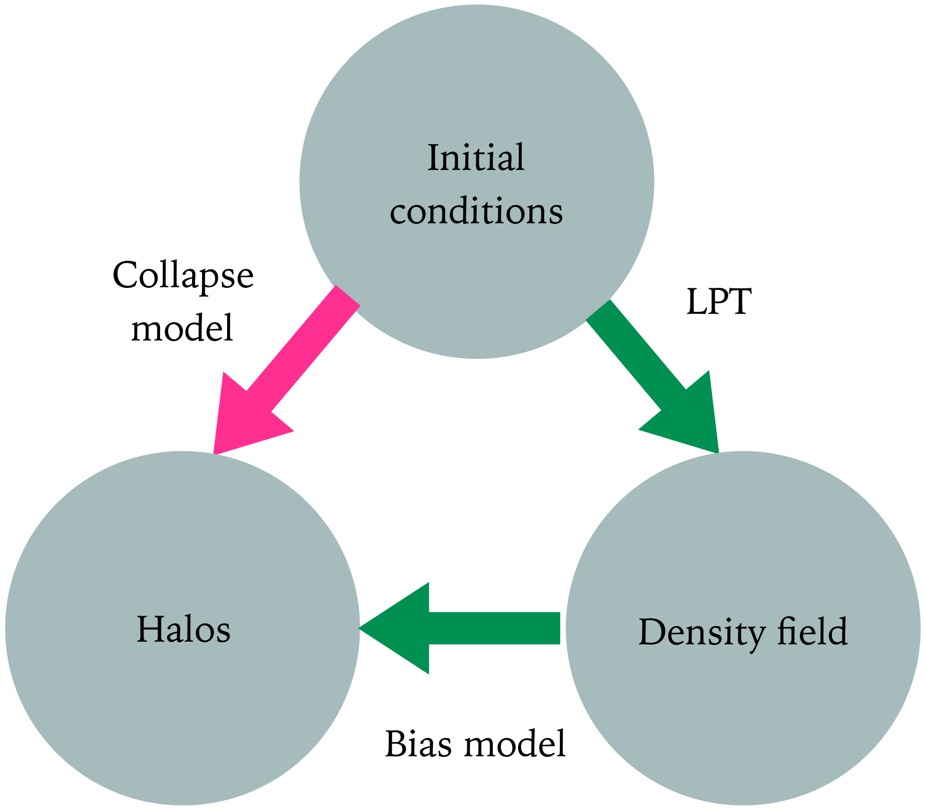
COMPARED METHODS

- One fast PM method: ICE-COLA
- Two predictive methods: Pinocchio, PeakPatch
- Two calibrated methods: Patchy, Halogen
- Two density PDF assumptions: Gaussian, Lognormal



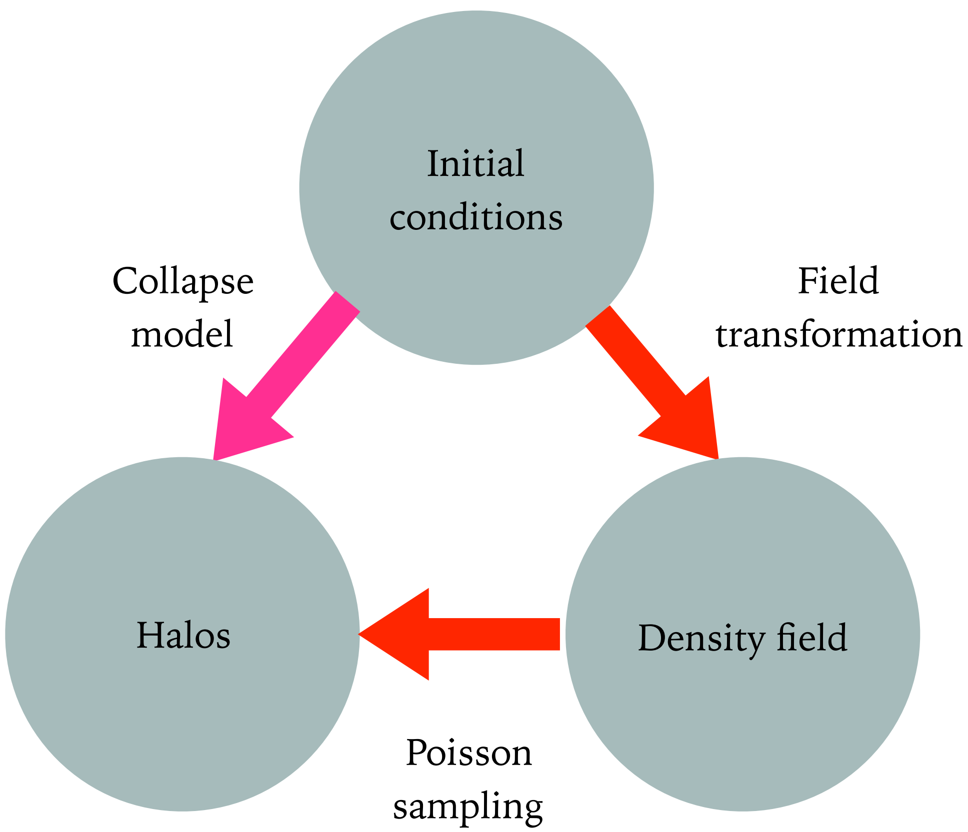
COMPARED METHODS

- One fast PM method: ICE-COLA
- Two predictive methods: Pinocchio, PeakPatch
- Two calibrated methods: Patchy, Halogen
- Two density PDF assumptions: Gaussian, Lognormal



COMPARED METHODS

- One fast PM method: ICE-COLA
- Two predictive methods: Pinocchio, PeakPatch
- Two calibrated methods: Patchy, Halogen
- Two density PDF assumptions: Gaussian, Lognormal



Method	Computational requirements	Mock characteristics
Minerva (N-body)	4500 hours(/2) 660 Gb	Gadget (treePM) 1000^3 particles
ICE-COLA	66 hours(/2) 340 Gb	1000^3 particles $(3x1000)^3$ PM grid
Pinocchio	6.4 hours 256 Gb	1000^3 particles
PeakPatch	1.72 hours* 75 Gb*	1000^3 particles
Halogen	0.6 hours 44 Gb	768^3 particles 300^3 grid
Patchy	0.2 hours 15 Gb	500^3 particles 500^3 grid
Lognormal	0.1 hours 5.6 Gb	256^3 grid

* do not resolves smaller mass halos

SAMPLES

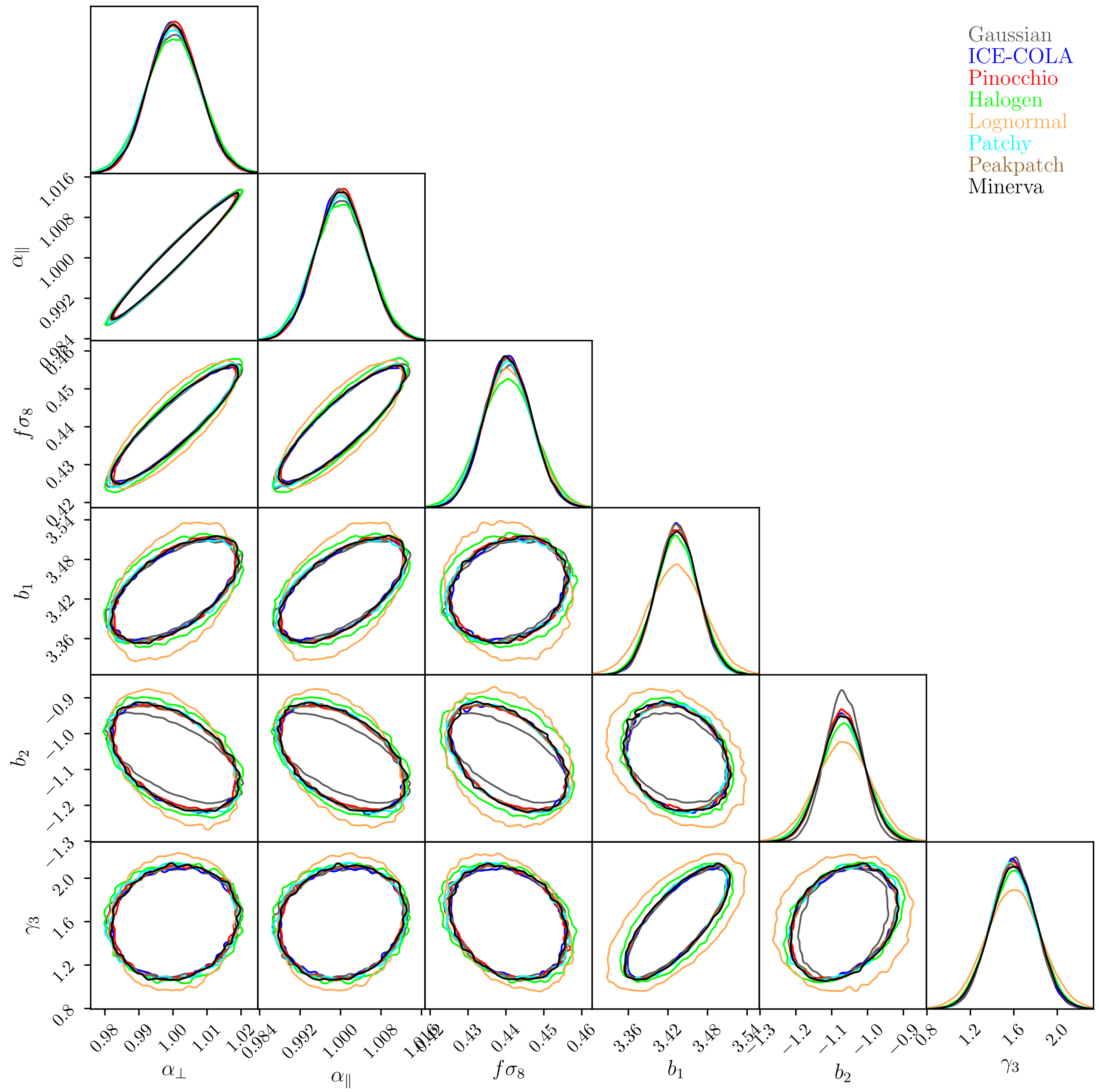
Halo samples at z=1

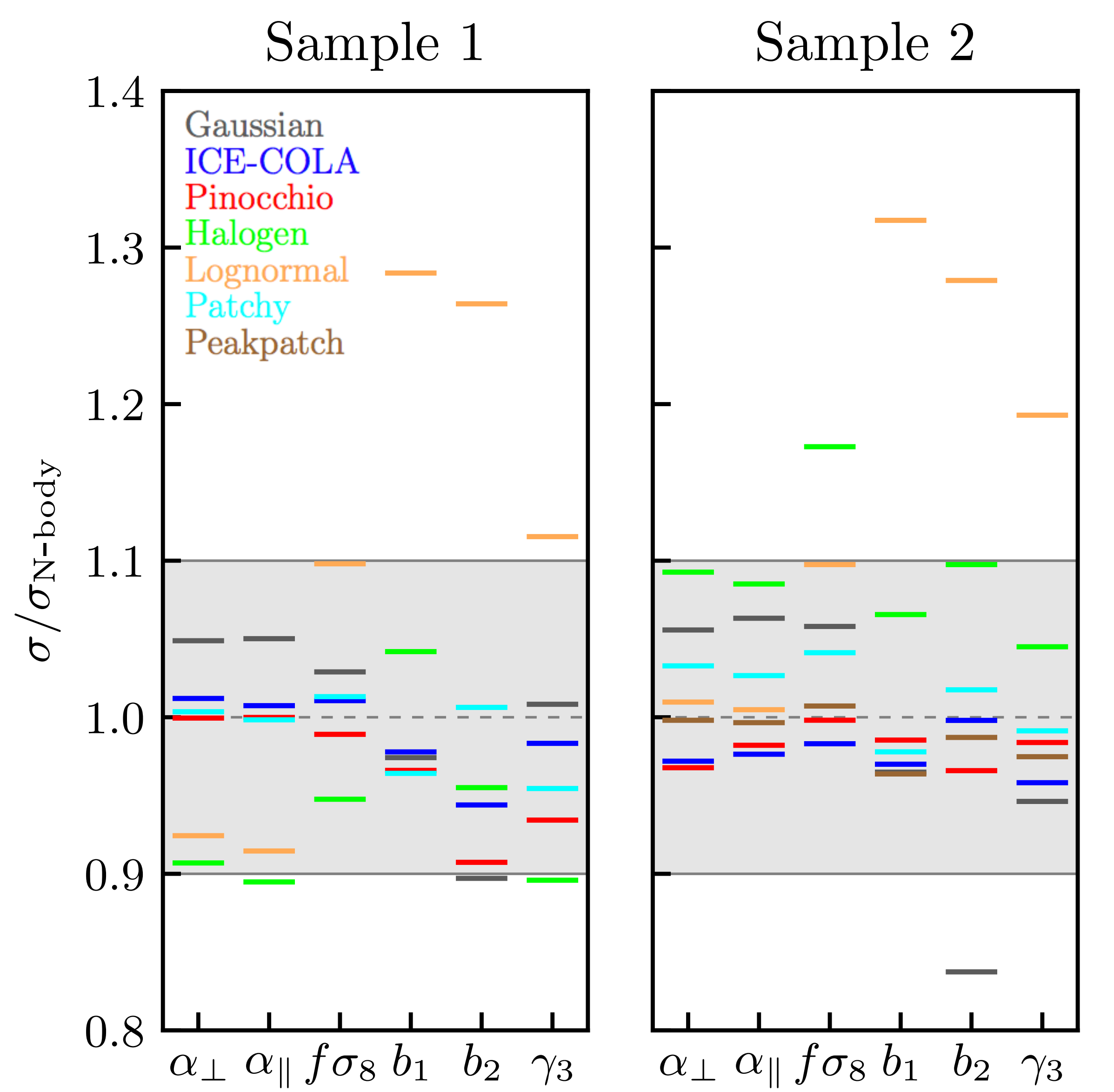
Method	$\bar{n}_{halos}$ (h^3Mpc^{-3})	M_{min} (h^{-1} M$_\odot$)
	Sample 1	
N-body	2.130×10^{-4}	1.121×10^{13}
ICE-COLA	2.123×10^{-4}	1.086×10^{13}
PINOCCHIO	2.148×10^{-4}	1.044×10^{13}
HALOGEN	2.138×10^{-4}	1.121×10^{13}
Lognormal	2.131×10^{-4}	1.121×10^{13}
PATCHY	2.129×10^{-4}	1.121×10^{13}
	Sample 2	
N-body	5.441×10^{-5}	2.670×10^{13}
ICE-COLA	5.455×10^{-5}	2.767×10^{13}
PINOCCHIO	5.478×10^{-5}	2.631×10^{13}
HALOGEN	5.393×10^{-5}	2.670×10^{13}
Lognormal	5.441×10^{-5}	2.670×10^{13}
PATCHY	5.440×10^{-5}	2.670×10^{13}
PEAKPATCH	5.439×10^{-5}	2.355×10^{13}

From Euclid redbook:

1.81×10^{-3} h^3 Mpc^{-3}

APPROXIMATE METHODS FOR COVARIANCE ESTIMATION

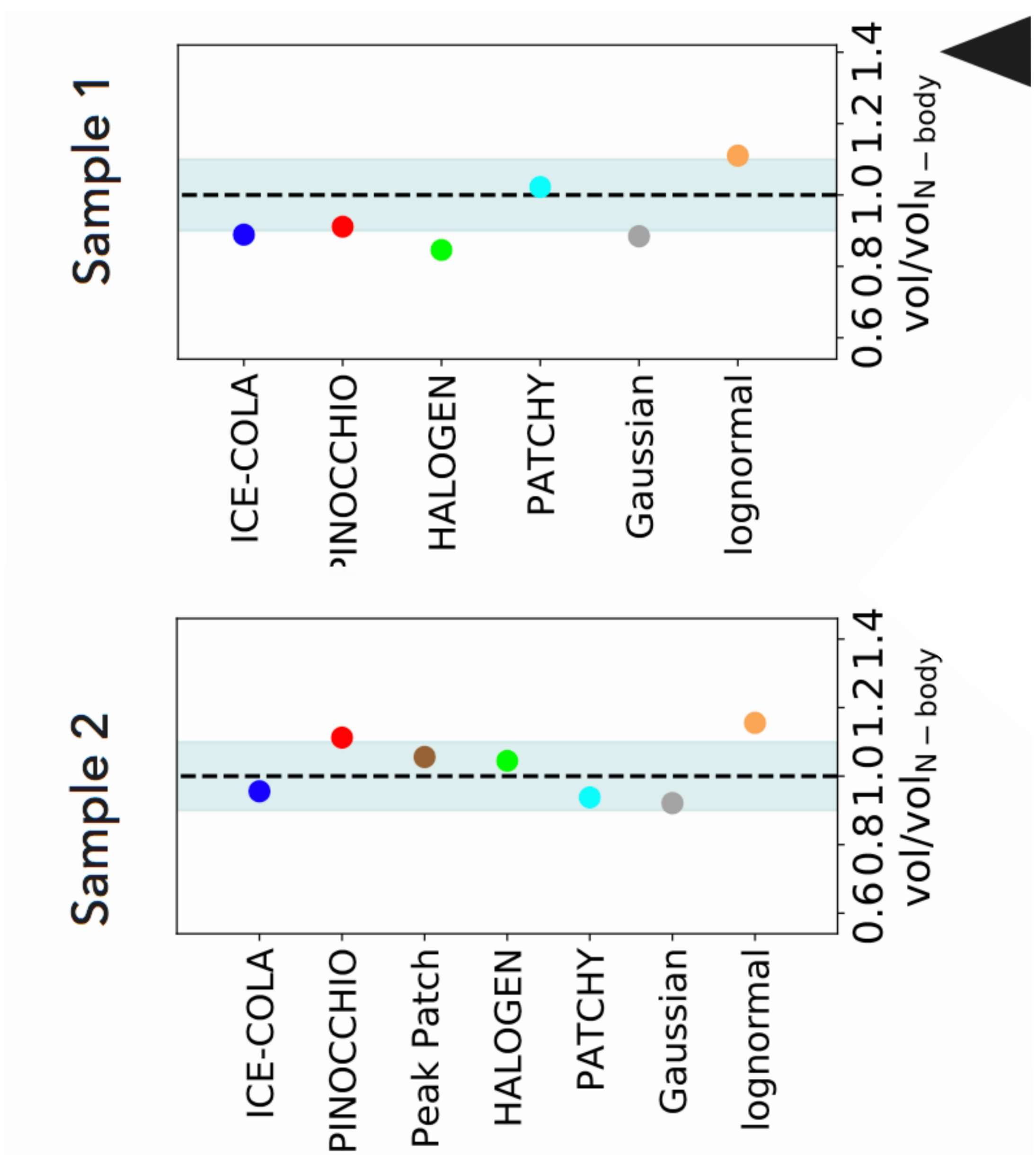


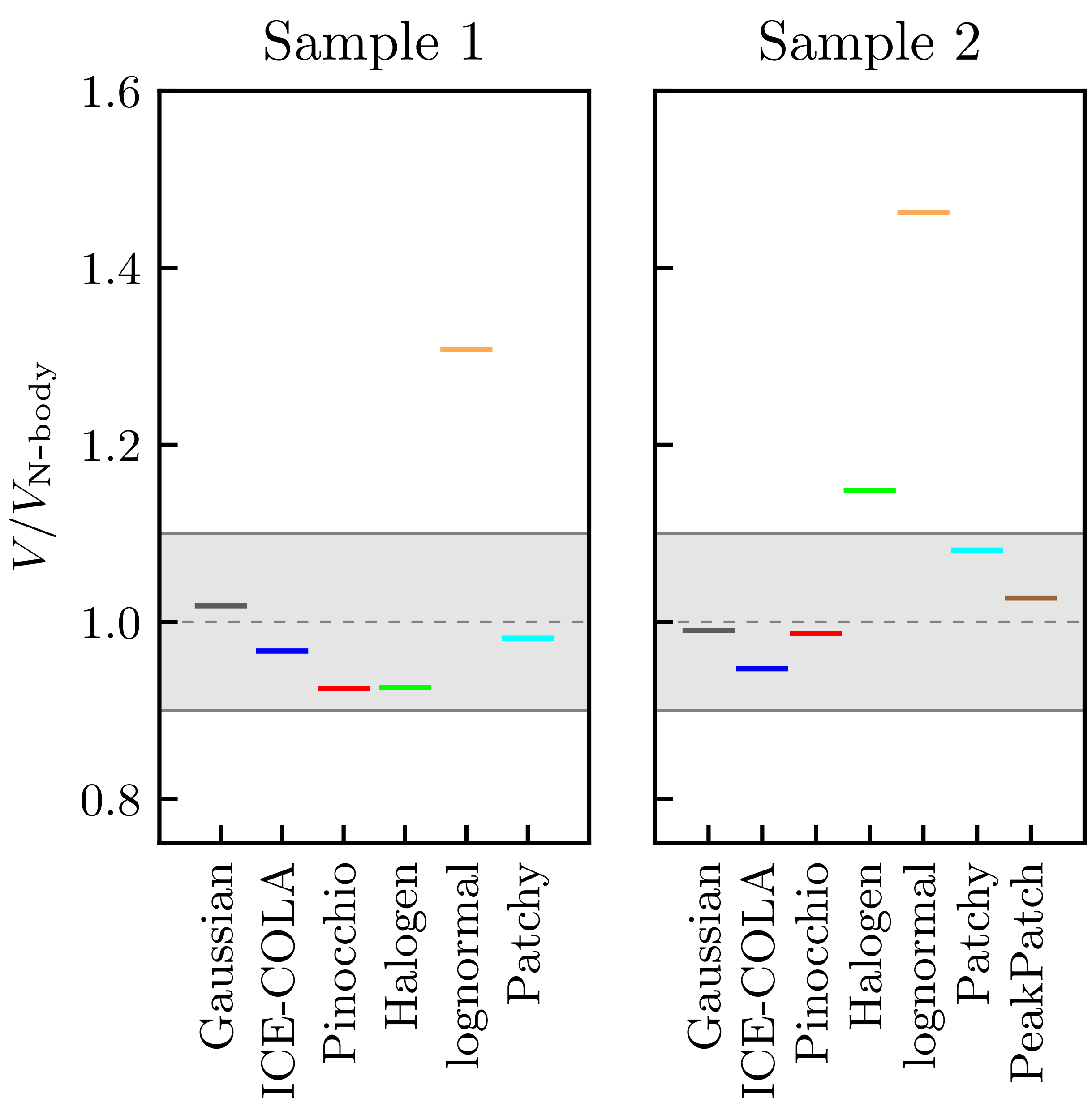


Blot et al. arxiv:1806.09497

APPROXIMATE METHODS FOR COVARIANCE ESTIMATION

$$V = \sqrt{\det cov(\alpha_{\parallel}, \alpha_{\perp}, f\sigma_8)}$$

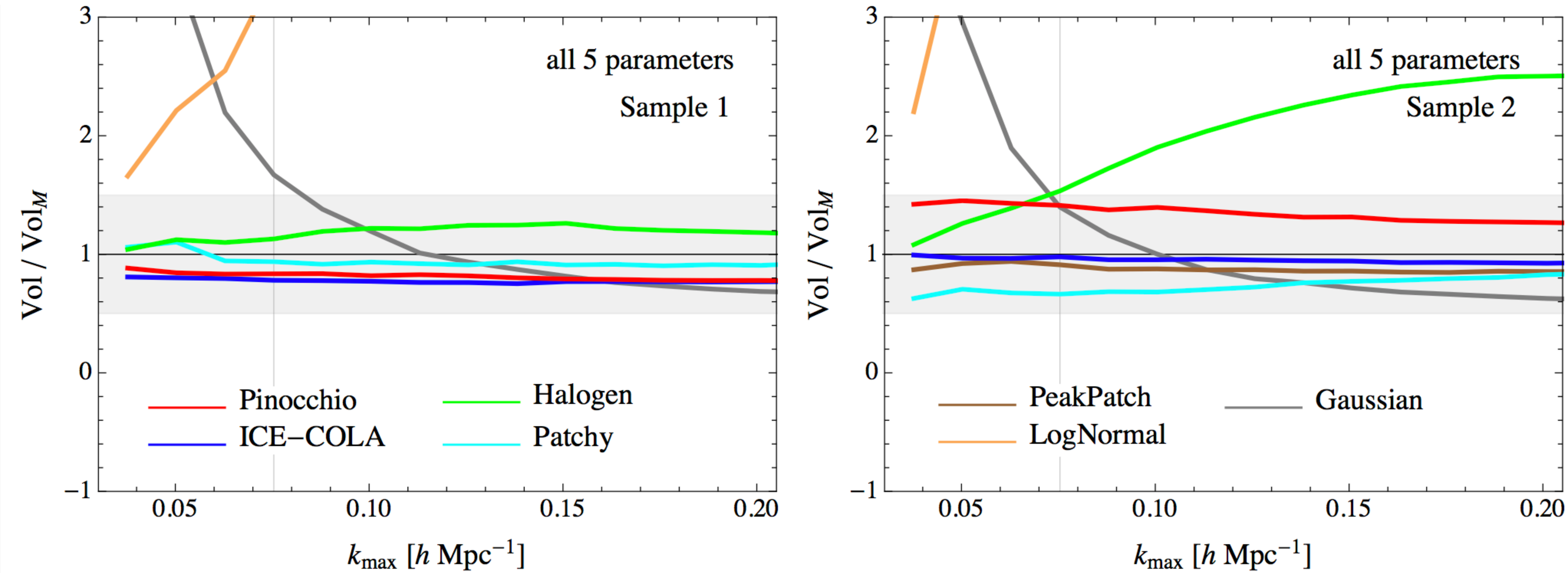




Lippich et al. arxiv:1806.09477

Blot et al. arxiv:1806.09497

APPROXIMATE METHODS FOR COVARIANCE ESTIMATION



Colavincenzo et al. arxiv:1806.09499

CONCLUSIONS

- Simulations are fundamental for LSS analysis
- Computational cost is the main limiting factor
- Simulations have systematics
- Approximate methods look promising to estimate covariances

FOOD FOR THOUGHTS

- How to secure computing time for yet another (thousands) LCDM simulation
- Cosmological simulations should be in the budget of cosmological surveys!
- Many upcoming surveys with similar goals: sharing simulation data?
- Need more interactions with software engineers