Likelihood-Free Inference Workshop

University of Edinburgh

Simulate a universe on your laptop

Chieh-An Lin (Linc)

March 19th, 2019

Flatiron Institute, New York

Machine learning

The algorithms are alright. The problem is in the data.

Machine learning

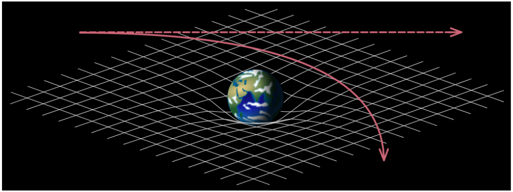
The algorithms are alright. The problem is in the data.

Likelihood-free inference

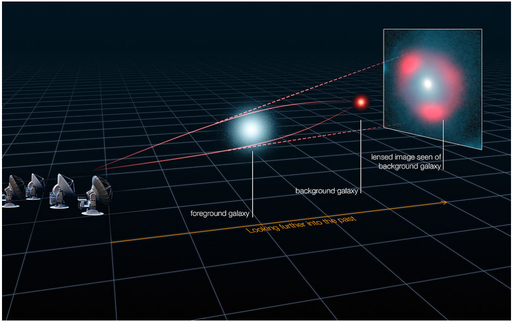
The algorithms are alright. The problem is in the simulator.

Gravitational lensing

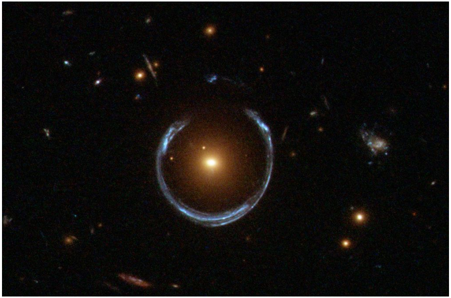
General relativity



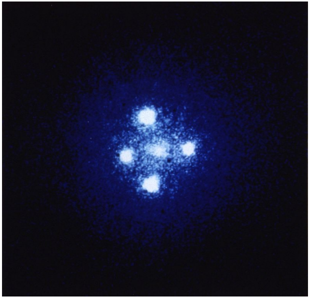
Gravitational lensing

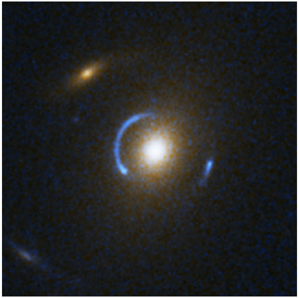


Source: ALMA

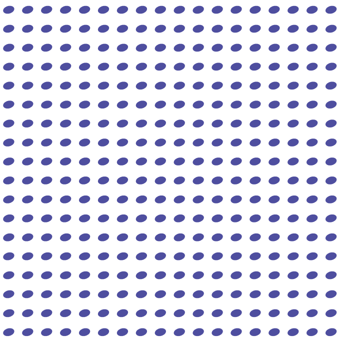


Strong lenses

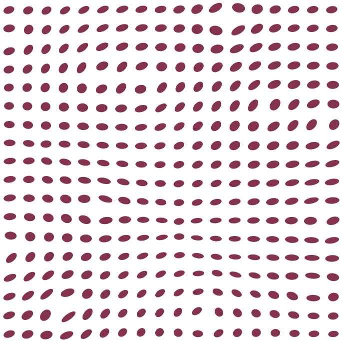




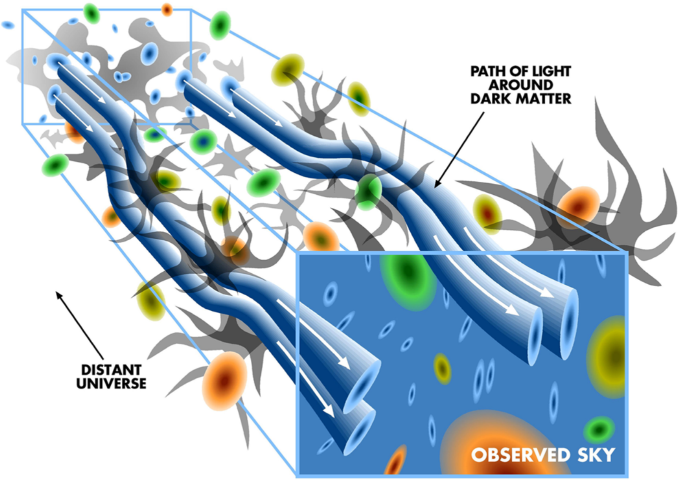
Source: SDSS, HST



Unlensed sources

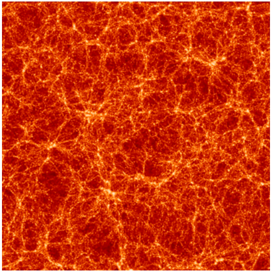


Weak lensing

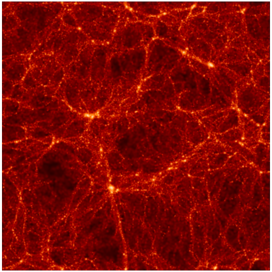

PATH OF LIGHT
AROUND
DARK MATTER
DISTANT
UNIVERSE
OBSERVED SKY

Source: LSST

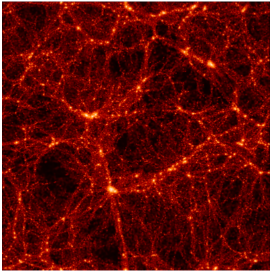
Massive structures from different models



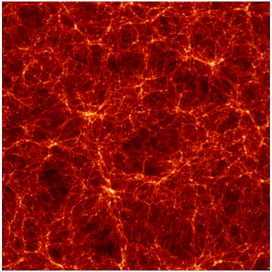
SCDM



OCDM



ΛCDM

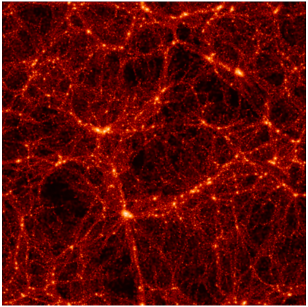


τCDM

Abundance $\Omega_{\rm m}$
Fluctuation σ_8

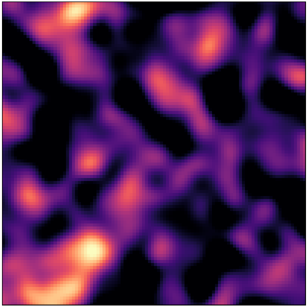
(Credit: J. Colberg, Virgo)

Lensing is a projection

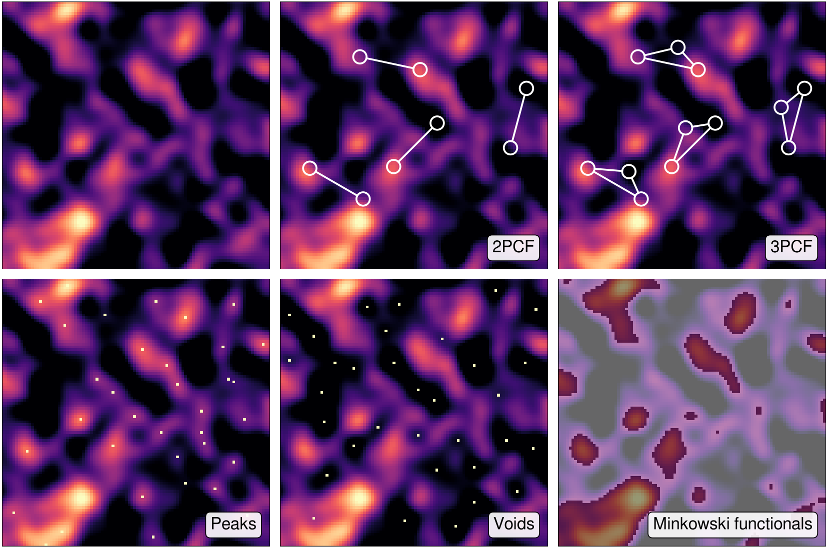


Projection

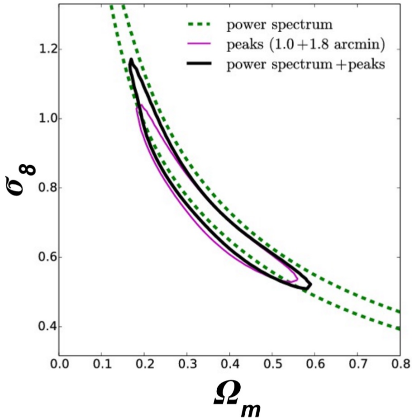
$3D \Rightarrow 2D$



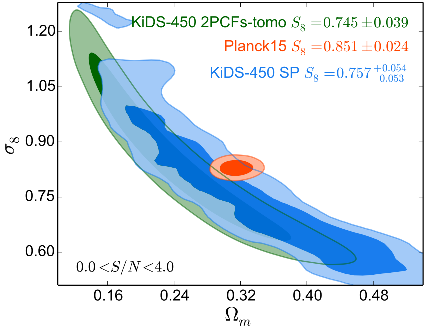
Summary statistics



State of the art



Liu et al. (2015)



Martinet et al. (2018)

I have a dream...

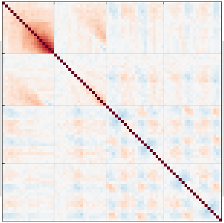
... that all statistics will one day live in the same likelihood
where they will not be judged by their order
but by their degeneracy-breaking character.

(From an anonymous statistician in a sleepy talk?)

Sweet sweet dream

Likelihood

- Is it Gaussian?
- If yes, what is the covariance?
- What about survey effects?

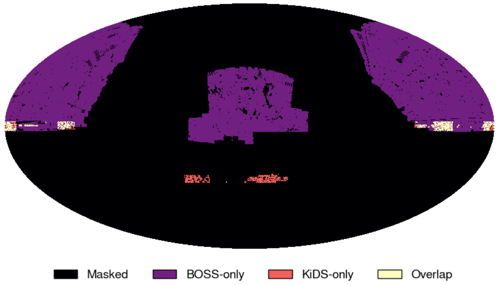


Likelihood-free

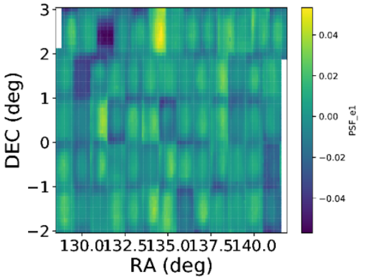
- Is the simulator realistic?
- Can we run it faster?
- Can we infer optimal summary statistics?

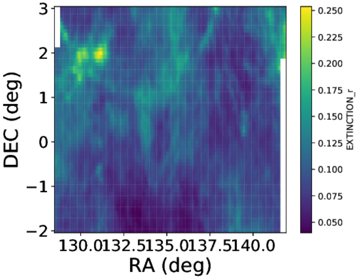
Survey effects: examples

Arbitrary footprints

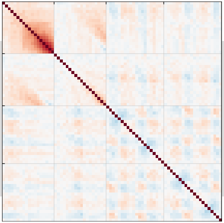


Variable conditions





Sweet sweet dream



Likelihood

- Is it Gaussian?
- If yes, what is the covariance?
- What about survey effects?

Likelihood-free

- Is the simulator realistic?
- Can we run it faster?
- Can we infer optimal summary statistics?

Challenges

How to create realistic simulations for various weak-lensing statistics?

How to do it efficiently?

Fast simulator

Ingredients for fast simulations

It's all about mass distribution

- Large-scales structures : lognormal fields

But the lognormal approximation breaks on small scales

Ingredients for fast simulations

It's all about mass distribution

- Large-scales structures : lognormal fields

But the lognormal approximation breaks on small scales

Let's take advantage of the halo model:

- Small scales: halo profiles & abundance
- Intermediate scales: halo clustering

Ingredients for fast simulations

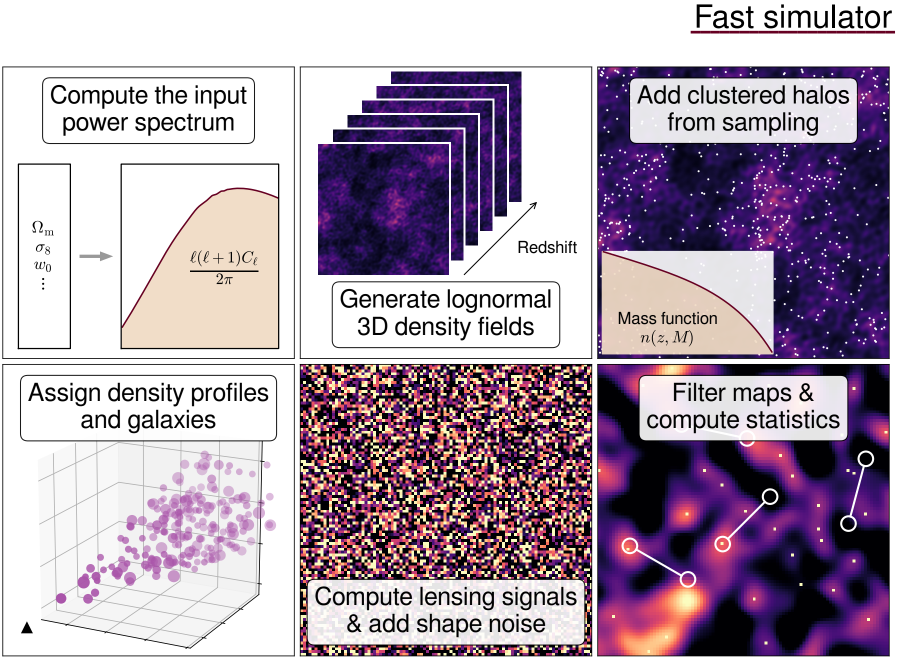
It's all about mass distribution

- Large-scales structures : lognormal fields

But the lognormal approximation breaks on small scales

Let's take advantage of the halo model:

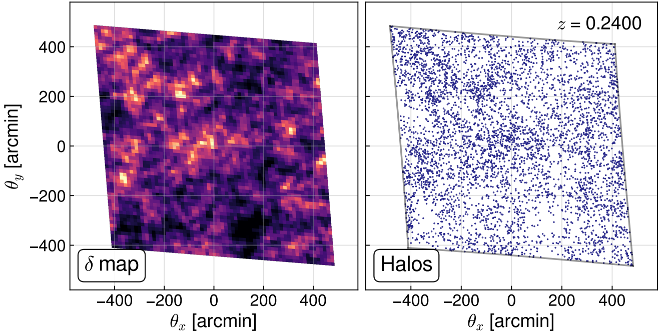
- Small scales: halo profiles & abundance (We know this well)
- Intermediate scales: halo clustering (Measure from N-body sims)


Fast simulator
Compute the input power spectrum
Ω_m
σ_8
w_0
$\vdots$
$\frac{\ell(\ell+1)C_\ell}{2\pi}$
Generate lognormal 3D density fields
Redshift
Add clustered halos from sampling
Mass function
$n(z, M)$
Assign density profiles and galaxies
Compute lensing signals & add shape noise
Filter maps & compute statistics

Public codes

FLASK:
Full-sky lognormal simulations
(Xavier, Abdalla, & Joachimi 2016)

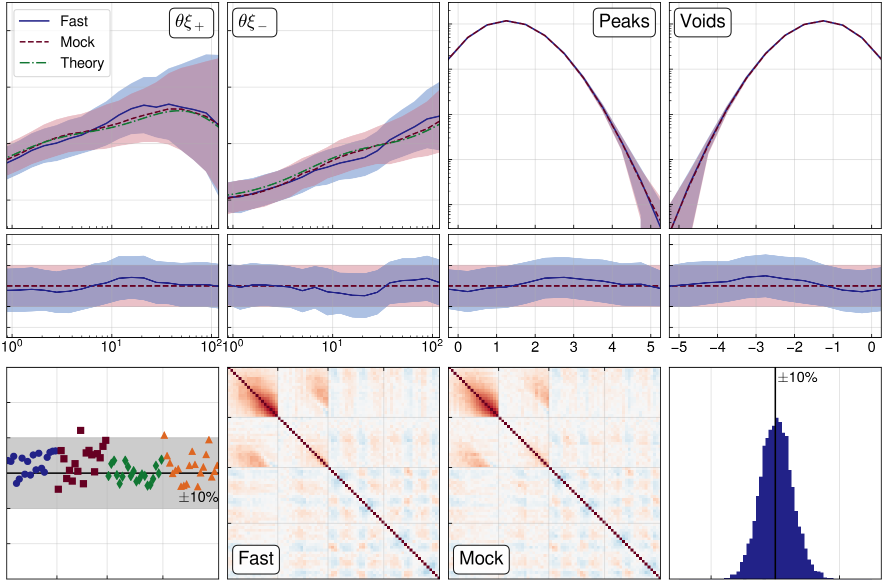
CAMELUS:
Halo simulations & lensing calculator
(Lin & Kilbinger 2015a, b)



Model performance

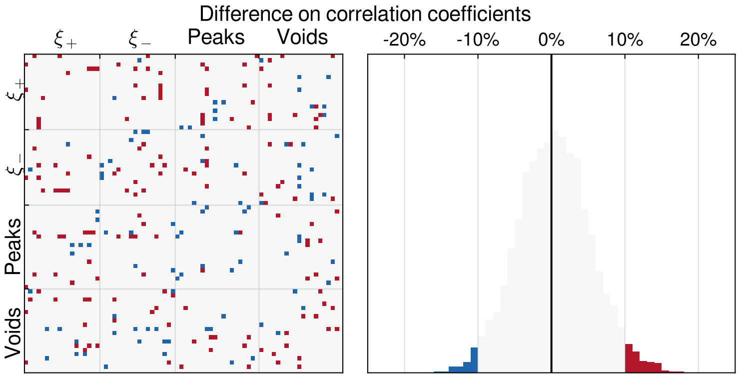
Lin et al. in prep.

Result summary



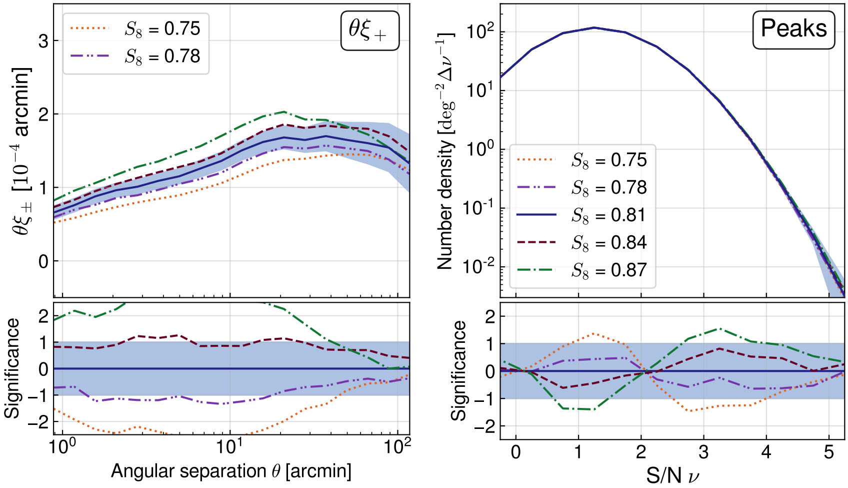
Zoom on covariance

Lin et al. in prep.



Lin et al. in prep.

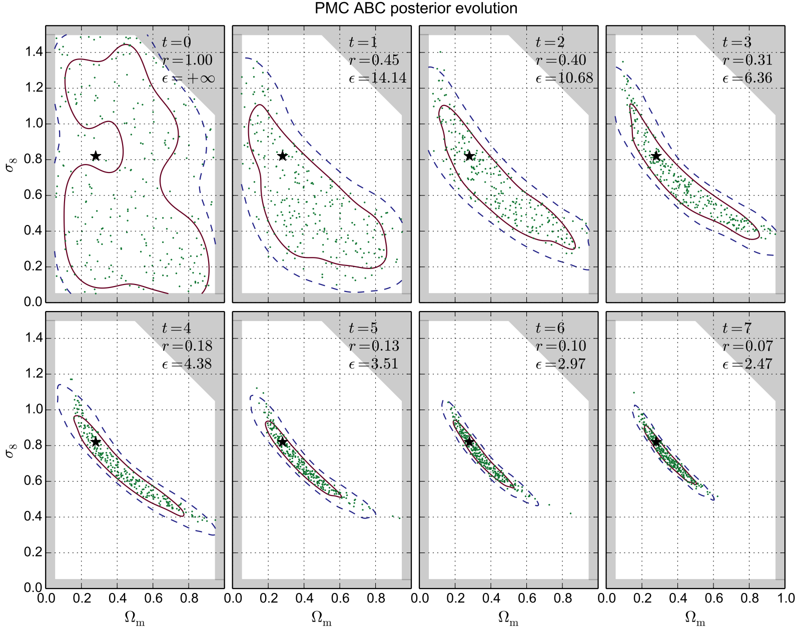
Cosmological sensitivity



Likelihood-free inference

Lin & Kilbinger (2015b)

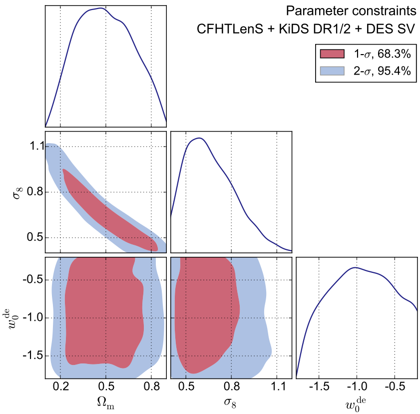
Result from ABC

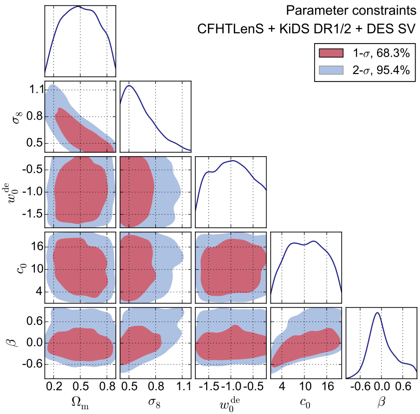


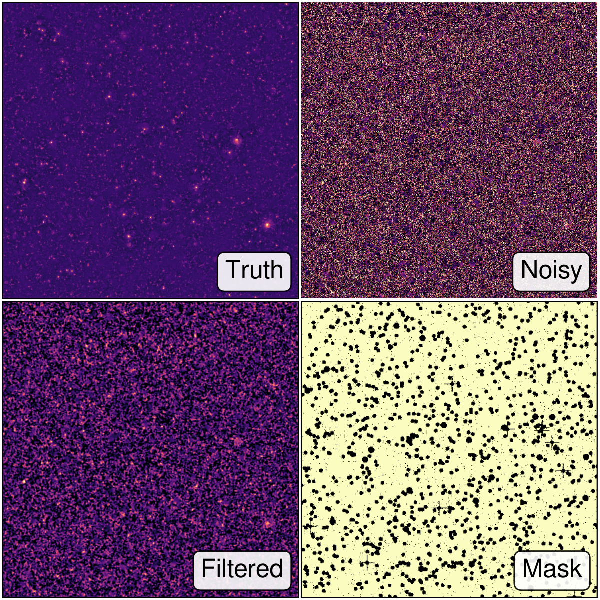
ABC applications to astronomy

- Galaxy morphology (Cameron & Pettitt 2012)
- Type Ia supernovae (Weyant et al. 2013; Jennings et al. 2016)
- Milky Way disk formation (Robin et al. 2014)
- Weak-lensing peak counts (Lin & Kilbinger 2015b)
- Strong lensing (Killedar et al. 2015)
- Halo occupation distribution (Hahn et al. 2016)
- and so on

When the dimensionality increases...

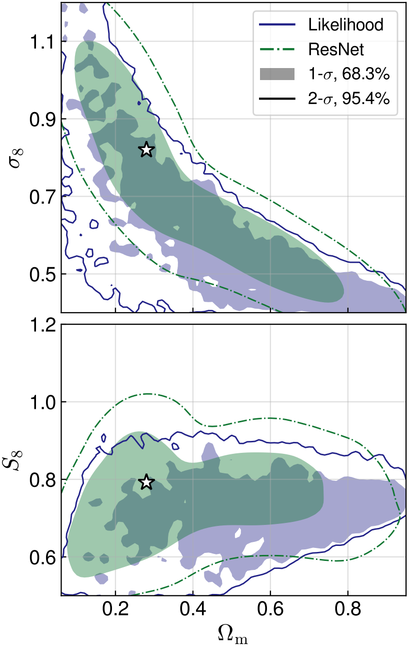






Realistic maps

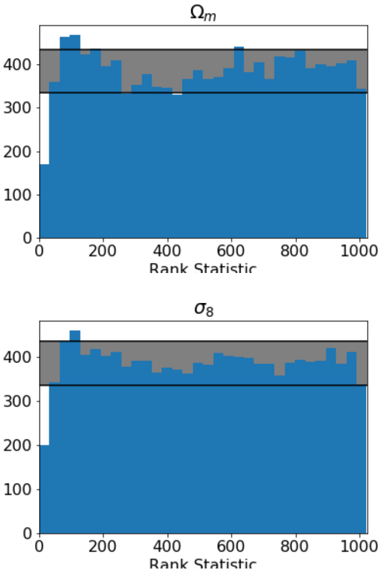
"Inversion leakage" in masks was deliberately kept



LFI robust to systematics

- Deep Residual Network with mixture density output
- Variational mutual information maximization

Lanusse & Lin in prep.



LFI diagnostics

Posterior validation by simulation-based calibration

Lanusse & Lin in prep.

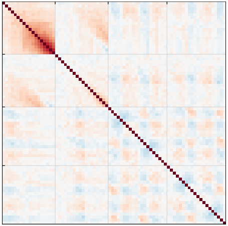
From the discussion session of yesterday

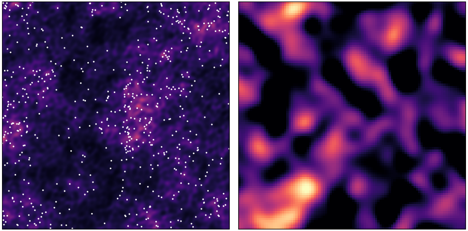
Do we have a list of "realistic" (whatever that means), but tractable toy problems on which we can test algorithms?

Uncertainties on ML predictions: Ensemble methods, bootstrapping, Bayesian networks, variational dropout (is that dead?)...

Summary

- A fast simulator for WL statistics
- Lognormal fields & halo model
- Useful for covariance estimation & likelihood-free inference





Principle collaborators

Shadab Alam

Joachim Harnois-Déraps

Catherine Heymans

Martin Kilbinger

François Lanusse

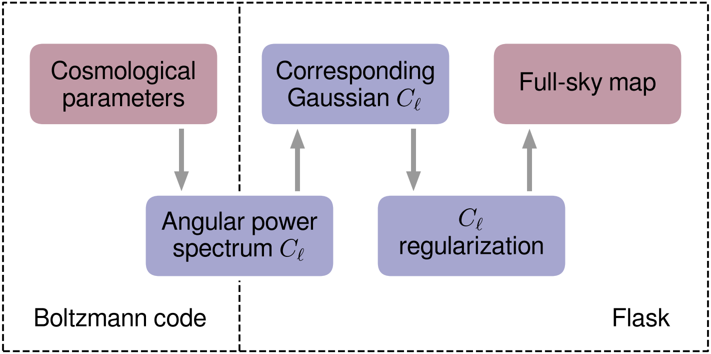
Ann Lee

Backup slides

Xavier, Abdalla, & Joachimi (2016)

Flask

Fast simulations of full-sky lognormal fields

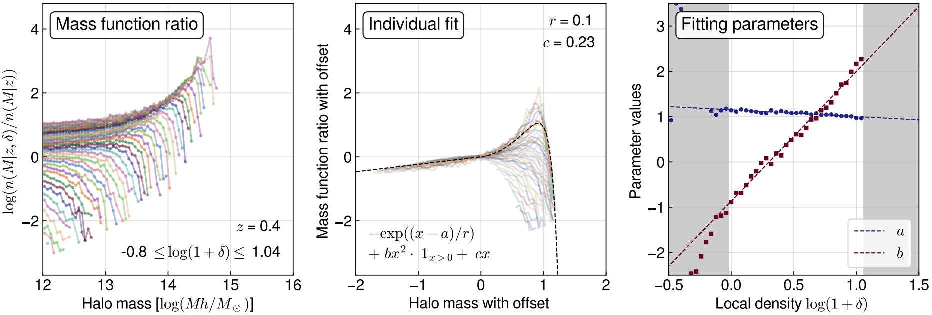


How to determine clustering?

Populate halos with conditional mass functions:

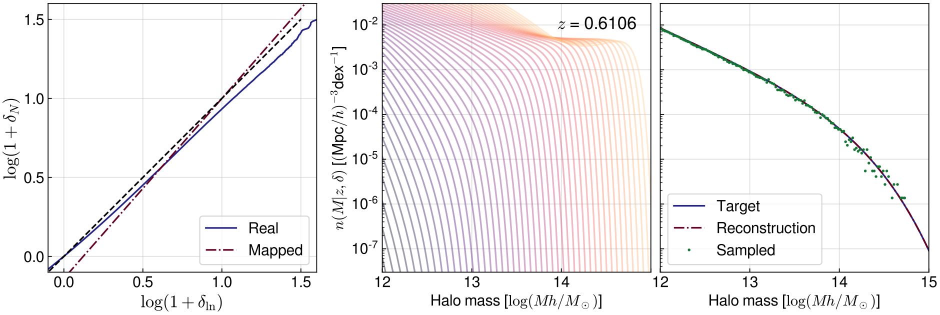
$$n(M|\delta) = \Gamma(M|\delta)\, n(M)$$

where $\Gamma(M|\delta)$ is measured from N-body simulations

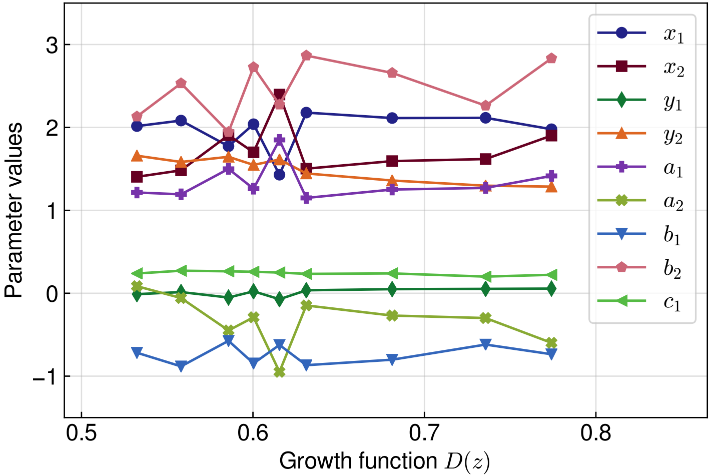


Caveats

- Different cosmologies
- Smoothing volume correspondence
- pdf(δ) from lognormal is not pdf(δ) from N-body



Reshift dependence of fitting parameters

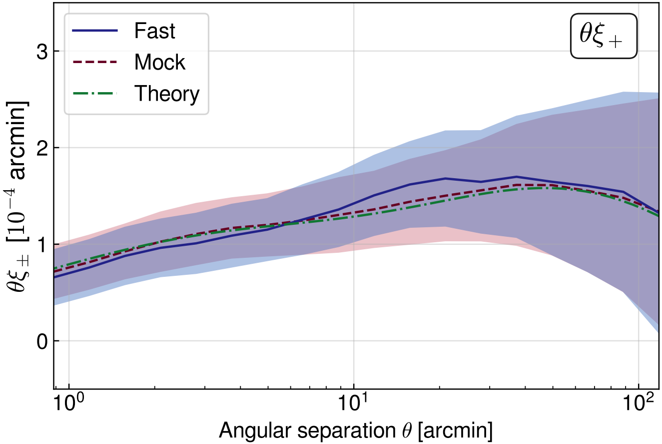


Blue = fast simulator = Flask + Camelus

Red = N-body mocks = the SLICS suite

Green = theory = Nicaea (only for $\xi_{+/-}$)

Results on ξ_+



Lin et al. in prep.

Results on ξ_+