Title: Precision Cosmology from the Clustering of Galaxies

Speakers: Marcel Schmittfull

Series: Cosmology & Gravitation

Date: March 23, 2021 - 11:00 AM

URL: http://pirsa.org/21030038

Abstract: Large surveys of the positions of galaxies in the Universe are becoming increasingly powerful to shed light on some of the unsolved problems of cosmology, including the question of what caused the early Universe to expand. The analysis of the data is challenging, however, because the signal is small, the data is difficult to model, and its probability distribution is not fully known. I will present some recent ideas to approach these challenges.
Precision Cosmology using the Clustering of Galaxies

Marcel Schmittfull
Perimeter Institute, March 2021
Nature of each building block is unknown

**Cosmic inflation**
What particle physics model led to the rapid expansion?
How did our Universe begin?

**Dark energy**
What causes this 2nd epoch of rapid expansion?
Is it a cosmological constant? Is General Relativity broken?

**Dark matter**
What particle(s) is it made of?

**Relativistic particles**
What is the mass (hierarchy) of neutrinos?
Are there additional light particles?
Non-Gaussian fluctuations from inflation

Cosmic inflation

Single field
Gaussian fluctuations
Skewness $f_{NL} \ll 1$

Multi-field
Non-Gaussian fluctuations
Skewness $f_{NL} \gg 1$

---

log(probability)

-4 -2 0 2 4
Fluctuation

-4 -2 0 2 4
Fluctuation

log(probability)

-2 -1 0 1

10^2 10^1 10^0

$10^{-2}$ $10^{-1}$ $10^0$

$10^{-2}$ $10^{-1}$ $10^0$
Tightest limits today

WMAP satellite: $f_{NL} = 37 \pm 19.9 \ (1\sigma)$

Planck satellite: $f_{NL} = -0.9 \pm 5.1 \ (1\sigma)$

Both consistent with zero (2$\sigma$)

WMAP Collaboration, Bennett et al. (2013)
Planck Collaboration, Akrami et al. (1905.05697)
SNR^2 \sim \text{Number of modes} \sim \text{Volume}

\Rightarrow \text{Use 3D distribution of galaxies to improve constraints}
The distribution of galaxies

**Single-field inflation**
Galaxies evolve from normally distributed initial conditions

\[ f_{NL} = 0 \]

**Multi-field inflation**
Galaxies evolve from non-Gaussian initial conditions with enhanced peaks
We are looking for a 5000x smaller signal

\[ f_{NL} = 5000 \]

Dalal et al. (2007)
How to measure this signal?

Count #peaks
Count #voids
Measure histogram of the galaxy number density
Measure skewness, kurtosis, etc
Measure skewness of all 3D Fourier modes
+ Many more ideas proposed in the literature

\[ f_{\text{NL}} = 5000 \]
Some challenges

Signal is tiny

Data is complicated, nonlinear function of initial conditions
⇒ Not easy to model the data

Data is not normally distributed
⇒ What is the optimal data analysis method?
Other questions suffer from similar challenges

**Cosmic inflation**
What particle physics model led to the rapid expansion?
How did our Universe begin?

**Dark energy**
What causes this 2nd epoch of rapid expansion?
Is it a cosmological constant? Is General Relativity broken?

**Dark matter**
What particle(s) is it made of?

**Relativistic particles**
What is the mass (hierarchy) of neutrinos?
Are there additional light particles?
Power spectrum as a summary statistic

Power spectrum \( P(k) = \frac{1}{N(k)} \sum_{k, |k|=k} |\delta(k)|^2 \)

Squared size of fluctuations

![Image of power spectrum](image.png)

Wavenumber \( k (\sim 1/\text{scale}) \)
Power spectrum as a summary statistic

Power spectrum \( P(k) = \frac{1}{N(k)} \sum_{|\mathbf{k}| = k} |\delta(\mathbf{k})|^2 \)

Squared size of fluctuations

Multi-field inflation couples peaks to grav. potential, enhancing power at low \( k \)
Divide by expectation for single-field inflation

Multi-field inflation  Single field inflation

Measured galaxy power (drawn from a normal distribution with mean given by dashed curve)
Sample variance (error on the mean)

Power spectrum / expectation

Wavenumber $k$ ($\sim 1$/scale)
Make a correlated measurement, insensitive to signal

Seljak (2009), McDonald & Seljak (2009), MS & Seljak (2018)
Make another, correlated measurement

Sample variance cancels in the ratio, so can detect multi-field inflation

Seljak (2009), McDonald & Seljak (2009), MS & Seljak (2018)
How did this work?

Imagine you come up with a new image compression algorithm

Is it better than JPEG?
Method 1

a. Ask people to rate JPEG-compressed images

b. Also ask to rate *other* images compressed with new algorithm

c. Compare ratings to find winner

Subject to sample variance (error of the mean)
Method 2

- Ask people to rate same image compressed with JPEG & new algorithm
- Compare ratings 1-by-1 for each image

Less sample variance (can tell winner with 1 image)
How to measure the distribution of dark matter?

Measure gravitational lensing of Microwave Background radiation
Lensed Cosmic Microwave Background radiation

Local power spectrum

Local power spectrum

Estimate local magnification / demagnification
Measured magnification map

Planck Collaboration: Planck 2018 lensing
Forecasts for future experiments

~5x better than current constraint, factor ~2 from SVC

MS & Seljak (2018); arXiv: 1808.07445, 1907.08284; 1907.04473, 1908.01062; 1902.10541, 1908.07495
Challenge

Initial conditions

Data (galaxies)

Grav. evolution
Gas physics

Physical params.

Simple
Complicated, nonlinear
Observed
Standard approach

Initial conditions

Data (galaxies)

Grav. evolution
Gas physics

Nuisance params.

Physical params.

Summary statistic

Simple

Complicated, nonlinear

Observed
Beware of overfitting: 50 data points, 10 free parameters

Fitted parameters:

\[ b_1 = 0.98 \pm 0.01 \]
\[ b_2 = 0.01 \pm 2.73 \]
\[ b_3 = -0.62 \pm 1.43 \]
\[ b_4 = 0.58 \pm 2.33 \]
\[ \epsilon^{(c_0)}_{\text{c}4} = (5.3 \pm 4.7) \left( \frac{k_{\text{NL}}}{h\,\text{Mpc}^{-1}} \right)^2 \]
\[ \epsilon^{(c_1)}_{\text{c}4} = (-14 \pm 5) \left( \frac{k_{\text{M}}}{h\,\text{Mpc}^{-1}} \right)^2 \]
\[ \epsilon^{(c_2)}_{\text{c}4} = (-0.69 \pm 1.67) \left( \frac{k_{\text{M}}}{h\,\text{Mpc}^{-1}} \right)^2 \]
\[ \epsilon^{(c_3)}_{\text{c}4} = (0.70 \pm 14.74) \]
\[ \epsilon^{(c_4,2)}_{\text{c}4} = (8.9 \pm 3.4) \left( \frac{k_{\text{M}}}{h\,\text{Mpc}^{-1}} \right)^2 \]
\[ \epsilon^{(c_5,3)}_{\text{c}4} = (8.0 \pm 7.8) \left( \frac{k_{\text{M}}}{h\,\text{Mpc}^{-1}} \right)^2 . \]
Alternative: Predict 3D data given initial conditions

Initial conditions

Grav. evolution
Gas physics

Nuisance params.

Physical params.

Simple

Complicated, nonlinear

Data (galaxies)

Observed
Benefits of using 3D fields rather than summary statistics

+ No overfitting (6 parameters describe >1 million galaxy positions)
+ No sample variance, can use small volumes with high resolution
+ ‘All’ $n$-point functions measured simultaneously
+ Easy to isolate mistakes of the model
+ Useful for field-level likelihood and initial condition reconstruction

MS, Simonović, Assassi & Zaldarriaga (2019)
Setup

Initial conditions

Simulation

Regression

Synthetic galaxies

Model prediction

MS, Simonović, Assassi & Zaldarriaga (2019)
Simulation

$1536^3 = 3.6\text{B particles in a 3D cubic box}$

$3072^3 = 29\text{B grid points for long-range force computation}$

4000 time steps

5 realizations

1M CPU hours on a local cluster ($\sim2000$ CPUs) using MP-Gadget

N-body code
Comparison with linear regression model

Simulation, $\log M = 10.8 - 11.8$

Simulation

Linear Std. Eul. bias

Linear model $b_1 \delta_m(x)$

Reasonable prediction on large scales

Missing structure on small scales

MS, Simonović, Assassi & Zaldarriaga (2019)
How to improve?

Include all terms allowed by symmetries (effective field theory)

$$\delta_g(x) = b_1 \delta_m(x) + b_2 \delta_m^2(x) + \text{tidal term} + b_3 \delta_m^3(x) + \cdots$$

Desjacques, Jeong & Schmidt: Review of Large-Scale Galaxy Bias (2018)

Fit coefficients $b_i$ using least-squares regression

MS, Simonović, Assassi, Zaldarriaga (2019)

In practice:
- Run independent regression for each Fourier mode shell
- Fit resulting regression coefficients $b(k)$ with 6-parameter model
- Orthogonalize operators for robust numerics and interpretation
- Include large bulk flows nonperturbatively (see later)
Comparison with nonlinear model

Simulation, log$M = 10.8 - 11.8$

Simulation

Cubic bias

Nonlinear regression model

$\delta_g$

Much better agreement than linear model

MS, Simonović, Assassi, Zaldarriaga (2019)
How to improve?

Include all terms allowed by symmetries (effective field theory)

$$
\delta_g(x) = b_1 \delta_m(x) + b_2 \delta^2_m(x) + \text{tidal term} + b_3 \delta^3_m(x) + \cdots
$$

Desjacques, Jeong & Schmidt: Review of Large-Scale Galaxy Bias (2018)

Fit coefficients $b_i$ using least-squares regression

MS, Simonović, Assassi, Zaldarriaga (2019)

In practice:
- Run independent regression for each Fourier mode shell
- Fit resulting regression coefficients $b_i(k)$ with 6-parameter model
- Orthogonalize operators for robust numerics and interpretation
- Include large bulk flows nonperturbatively (see later)
Mean-squared model error

\[ P_{\text{err}} = \langle |\delta_{\text{truth}}^h - \delta_{\text{model}}^h|^2 \rangle \]

This white noise is crucial to avoid biasing physical parameters

MS, Simonović, Assassi, Zaldarriaga (2019)
Mean-squared model error

\[ P_{err} = \langle |\delta_{h}^{\text{truth}} - \delta_{h}^{\text{model}}|^2 \rangle \]

This white noise is crucial to avoid biasing physical parameters

MS, Simonović, Assassi, Zaldarriaga (2019)
How to improve?

Include all terms allowed by symmetries (effective field theory)

\[ \delta_g(x) = b_1 \delta_m(x) + b_2 \delta_m^2(x) + \text{tidal term} + b_3 \delta_m^3(x) + \cdots \]

Desjacques, Jeong & Schmidt: Review of Large-Scale Galaxy Bias (2018)

Fit coefficients \( b_i \) using least-squares regression

MS, Simonović, Assassi, Zaldarriaga (2019)

In practice:
- Run independent regression for each Fourier mode shell
- Fit resulting regression coefficients \( b_i(k) \) with 6-parameter model
- Orthogonalize operators for robust numerics and interpretation
- Include large bulk flows nonperturbatively (see later)
Bulk flows

Model with wrong bulk flows

Model with correct bulk flows

Correct bulk flows are crucial for small pixel-level residual

Cannot Taylor expand because bulk flows are large

$$\delta(x + \psi) \neq \delta(x) + \psi \nabla \delta(x)$$

MS, Simonović, Assassi, Zaldarriaga (2019)
Tried many other nonlinear models

3-6x larger model error

Our model allows using small-scale data (~10x larger volume)
Main reason: Bulk flows included nonperturbatively

MS, Simonović, Assassi, Zaldarriaga (2019)
Real data

Same model as recent EFT analyses of BOSS power spectrum

D’Amico, Gleyzes, Kokron et al. (1909.05271)
Ivanov, Simonović & Zaldarriaga (1909.05277)
Tröster, Sanchez, Asgari et al. (2020)

Model power spectrum is evaluated using FFTLog trick to speed up MCMC chains (reduce 2D loop integrals to 1D FFTs)

Hamilton (2000)
MS, Vlah & McDonald (2016)
McEwen, Fang, Hirata & Blazek (2016)
Cataneo, Foreman & Senatore (2017)
Simonović, Baldauf, Zaldarriaga et al. (2018)
Competitive for some parameters

Galaxies (SDSS)  
CMB (Planck 2018)

Similar precision as CMB

Amount of DM  Expansion rate  rms of fluctuations

Ivanov et al. (arXiv:1909.05277)
Redshift space

Distance to galaxy inferred from redshift = true distance + velocity

Model velocity field in Lagrangian perturbation theory to get a field level model for galaxies in redshift space

*Figure 1:* 2D slice of the linear density, Zel’dovich density, and $x$- and $y$-component of the continuous velocity field predicted by Eq. (2.12) for $n_{\text{max}} = 1$. The predicted velocity field is coherent over tens of Megaparsecs, with most regions flowing towards the cluster and filament in the center of the slice. To generate the Zel’dovich density and the velocity prediction, 1536$^2$ particles in a Lagrangian space box with $L = 500$ $h^{-1}$Mpc were shifted by the first-order displacement. All fields are evaluated at redshift $z = 0.6$.

Comparison with simulations

Model describes bulk flows in simulations well, but does not capture large velocities in highly clustered regions (satellites):

Simulation velocity  
Model  
Residual

Comparison with simulations

Model captures large-scale flows, but not Fingers of God
Comparison with simulations

Functional form of model error power spectrum:

\[ P_{\text{err}}(k, \mu) = \frac{1}{n_g} \left( c_{\epsilon,1} + c_{\epsilon,3} f \mu^2 \left( \frac{k}{k_M} \right)^2 \right) \]

\[ c_{\epsilon,1} = 0.599 \]

\[ c_{\epsilon,3} = 2.45 \left( \frac{k_M}{1 \text{hMpc}^{-1}} \right)^2 \]

Perko, Senatore, Jennings & Wechsler (arXiv:1610.09321)
https://github.com/mschmitt/full/perr
Additional thoughts from 3D tests of models

• Weighting galaxies by their mass can reduce model error 10x
• Removing 13% of galaxies gives 2x smaller rms RSD displacement, enabling higher model $k_{\text{max}}$
Additional thoughts from 3D tests of models

- Weighting galaxies by their mass can reduce model error 10x
- Removing 13% of galaxies gives 2x smaller rms RSD displacement, enabling higher model $k_{\text{max}}$

How to find weights/detect outliers?

Solve as optimization problem with custom-made objective (e.g. maximize quadrupole/monopole ratio at high $k$ to suppress FoG and shot noise)

Note: We can weight galaxies by any function of local observables and still use the same galaxy bias model

Inspired by Obuljen, Percival & Dalal (2020)
Test on synthetic dark matter data

Galaxy skew-spectra in redshift space

Cross-spectra between 14 quadratic fields $S_n$ and galaxy density

\[
\begin{align*}
    b_1^3 : & \quad S_1 = F_2[\delta, \delta] \\
    b_1^2 b_2 : & \quad S_2 = \delta^2 \\
    b_1^2 b g_2 : & \quad S_3 = S^2[\delta, \delta] \\
    b_1^3 f : & \quad S_4 = \dot{\delta}_i \dot{\delta}_j \partial_i \left( \delta \frac{\partial_j}{\nabla^2} \delta \right) \\
    b_1^2 f : & \quad S_5 = 2F_2[\delta^\parallel, \delta] + G^\parallel_2[\delta, \delta] \\
    b_1 b_2 f : & \quad S_6 = \delta \delta^\parallel \\
    b_1 b g_2 f : & \quad S_7 = S^2[\delta, \delta^\parallel] \\
    b_1^2 f^2 : & \quad S_8 = \dot{\delta}_i \dot{\delta}_j \partial_i \left( \delta \frac{\partial_j}{\nabla^2} \delta^\parallel + 2\delta^\parallel \frac{\partial_j}{\nabla^2} \delta \right) \\
    b_1 f^2 : & \quad S_9 = F_2[\delta^\parallel, \delta^\parallel] + 2G^\parallel_2[\delta^\parallel, \delta] \\
    b_2 f^2 : & \quad S_{10} = (\delta^\parallel)^2 \\
    b g_2 f^2 : & \quad S_{11} = S^2(\delta^\parallel, \delta^\parallel) \\
    b_1 f^3 : & \quad S_{12} = \dot{\delta}_i \dot{\delta}_j \partial_i \left( \delta^\parallel \frac{\partial_j}{\nabla^2} \delta + 2\delta^\parallel \frac{\partial_j}{\nabla^2} \delta^\parallel \right) \\
    f^3 : & \quad S_{13} = G^\parallel_2[\delta^\parallel, \delta^\parallel] \\
    f^4 : & \quad S_{14} = \dot{\delta}_i \dot{\delta}_j \partial_i \left( \delta^\parallel \frac{\partial_j}{\nabla^2} \delta^\parallel \right).
\end{align*}
\]

MS & Moradinezhad Dizgah (2021)
Galaxy skew-spectra in redshift space

Cross-spectra between 14 quadratic fields $S_n$ and galaxy density

\[
\begin{aligned}
    b_1^3 : & \quad S_1 = F_2[\delta, \delta] \\
    b_1^2 b_2 : & \quad S_2 = \delta^2 \\
    b_1^2 b_2 : & \quad S_3 = S^2[\delta, \delta] \\
    b_1^3 f : & \quad S_4 = \hat{z}_i \hat{z}_j \partial_i \left( \delta \frac{\partial j}{\nabla^2} \delta \right) \\
    b_1^2 f : & \quad S_5 = 2F_2[\delta^\parallel, \delta] + G_2^\parallel[\delta, \delta] \\
    b_1 f : & \quad S_6 = \delta \delta^\parallel \\
    b_1 b_2 f : & \quad S_7 = S^2[\delta, \delta^\parallel] \\
    b_1 b_2^2 : & \quad S_8 = \hat{z}_i \hat{z}_j \partial_i \left( \delta \frac{\partial j}{\nabla^2} \delta^\parallel + 2 \delta^\parallel \frac{\partial j}{\nabla^2} \delta \right) \\
    b_1 f^2 : & \quad S_9 = F_2[\delta^\parallel, \delta^\parallel] + 2G_2^\parallel[\delta^\parallel, \delta] \\
    b_2 f^2 : & \quad S_{10} = (\delta^\parallel)^2 \\
    b_2^2 f^2 : & \quad S_{11} = S^2(\delta^\parallel, \delta^\parallel) \\
    b_1 f^3 : & \quad S_{12} = \hat{z}_i \hat{z}_j \partial_i \left( \delta^\parallel \frac{\partial j}{\nabla^2} \delta + 2 \delta^\parallel \frac{\partial j}{\nabla^2} \delta^\parallel \right) \\
    f^3 : & \quad S_{13} = G_2^\parallel[\delta^\parallel, \delta^\parallel] \\
    f^4 : & \quad S_{14} = \hat{z}_i \hat{z}_j \partial_i \left( \delta^\parallel \frac{\partial j}{\nabla^2} \delta^\parallel \right). 
\end{aligned}
\]

MS & Moradinezhad Dizgah (2021)
Test on synthetic galaxy data

Code & Simulations

- Field level model  [github.com/mschmittfull/perr]
- Skew-spectra    [github.com/mschmittfull/skewspec]
- Iterative reconstruction [github.com/mschmittfull/iterrec]

All based on nbodykit [github.com/bccp/nbodykit]

For questions email mschmittfull@gmail.com