Large scale structure analysis with galaxy redshift surveys

Florian Beutler

June, 2015

Lawrence Berkeley National Lab

Outline of the week

Tuesday:

- LSS and BAO, basics of likelihood analysis
- Hands-on session: Implement your own MCMC

Wednesday:

- Continue LSS and BAO... including basics of LSS analysis (power spectrum, correlation function)
- Hands-on session: implement a power spectrum estimator, estimate a covariance matrix

Thursday:

- LSS and RSD
- Hands-on session: Use MCMC, power spectrum and covariance matrix to make a likelihood analysis (if time include shot noise and galaxy bias)

Friday:

- Beyond BAO and RSD... neutrino mass and non-Gaussianity, Future outlook, Euclid and DESI
- Hands-on session: finish likelihood analysis

Outline of the week

Tuesday:

- LSS and BAO, basics of likelihood analysis.
- Hands-on session: Implement your own MCMC.

Wednesday:

- Continue LSS and BAO... including basics of LSS analysis (power spectrum, correlation function)
- Hands-on session: implement a power spectrum estimator, estimate a covariance matrix

Thursday:

- LSS and RSD
- Hands-on session: Use MCMC, power spectrum and covariance matrix to make a likelihood analysis (if time include shot noise and galaxy bias)

Friday:

- Beyond BAO and RSD... neutrino mass and non-Gaussianity, Future outlook, Euclid and DESI
- Hands-on session: finish likelihood analysis

What are Baryon Acoustic Oscillations?

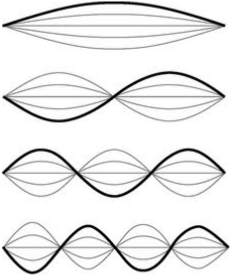
$$H^2(a) = H_0^2\left[\Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_K a^{-2} + \Omega_\Lambda\right]$$

The evolution eq. of baryon and photon perturbations in the radiation dominated era can be written as

$$\ddot{\delta} + 2H\dot{\delta} + \left(\frac{c_s^2 k^2}{a^2} - 4\pi G\overline{\rho}\right)\delta = F$$

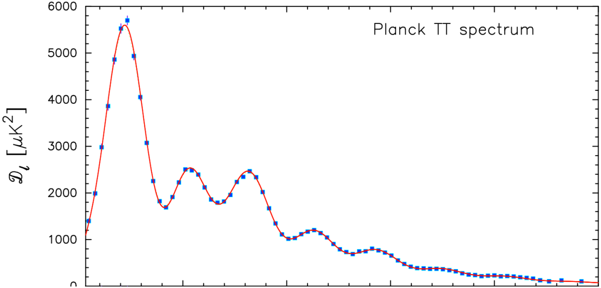
- Note that this is a forced and damped harmonic oscillator ($m\ddot{x} + b\dot{x} + kx = F$) with the plane wave solution $\delta \propto A\cos(\omega t - \phi)$, where $\omega^2 = c_s^2 k^2/a^2 - 4\pi G\overline{\rho}$.
- Because of the coupling between photons and baryons any perturbation in the baryon-photon plasma behaves like an acoustic wave thanks to the competing forces of gravity and radiation pressure.
  1. Gravity tries to compress the plasma where there is an over density...
  2. this increases the photon density...
  3. which leads to an increase of photon pressure...
  4. which causes the plasma to expand...and back to (1)

What are Baryon Acoustic Oscillations?

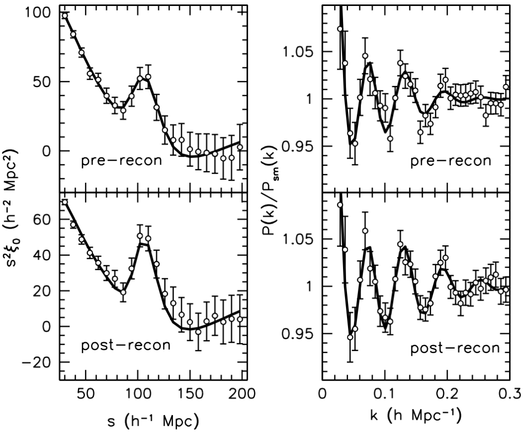


The oscillation depends on the speed of sound c_s and on the wavenumber k. At the freeze out, t_{ls}, some wave numbers will have $kc_st_{ls} = X\pi$ leading to maxima in the power spectrum, while $kc_st_{ls} = X\pi/2$ will yield minima.

These oscillations have been detected in the distribution of photons:



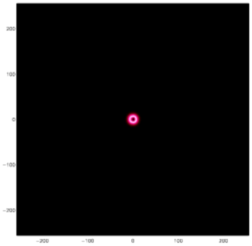
Since only 15% of the matter participated in the BAO physics, the signal in the matter distribution is suppressed compared to the photon distribution.

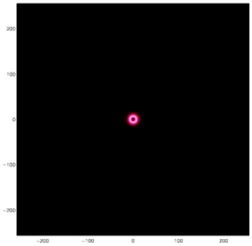


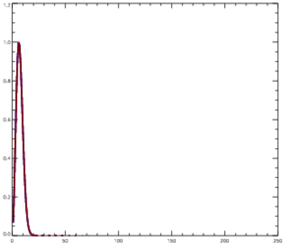
Anderson et al 2014

What are Baryon Acoustic Oscillations?

(1) We start with a point-like over density seeded by inflation. Directly after inflation we enter the radiation dominated regime. The over density sources a photon pressure which drives the baryons and photons outwards in a spherical wave. The dark matter continues to collapse into the original over density.

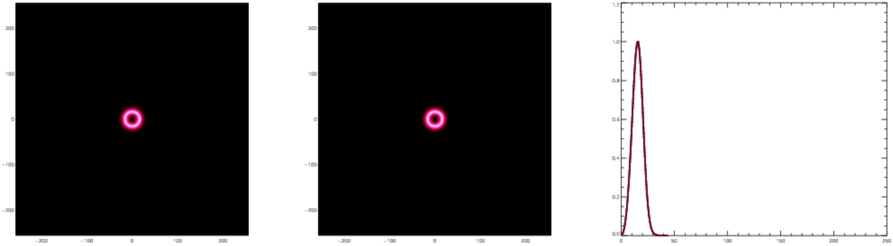






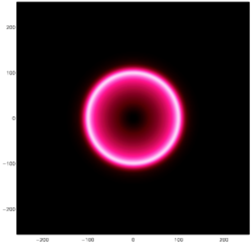
Martin White

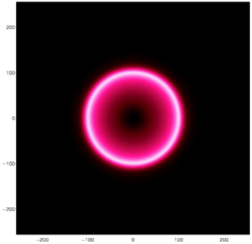
(2) The wave continues to move outwards with almost the speed of light.

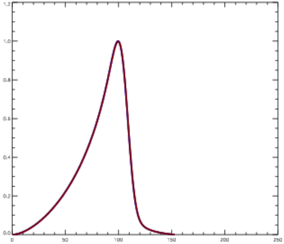


Martin White

(3) This expansion continues for 300 000 years.

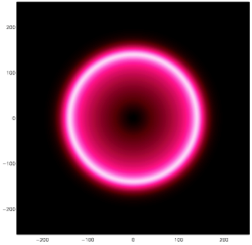


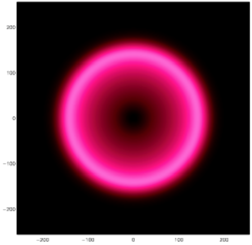


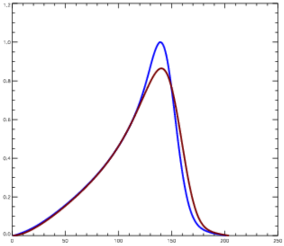


Martin White

(4) After 300 000 years the universe has cooled enough, so that protons capture the electrons to form neutral Hydrogen. This decouples the photons from the electrons (baryons). The photons quickly stream away, leaving the baryon peak stalled.

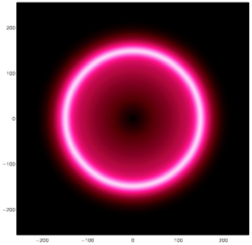


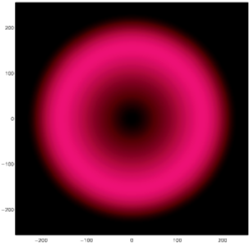


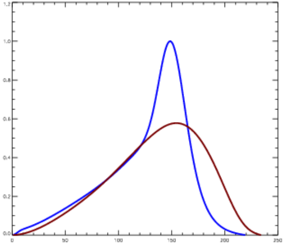


Martin White

(5) The mean free path of photons becomes very large and they basically do not interact (much) before reaching our CMB experiments.



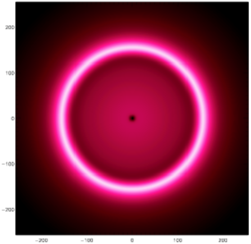


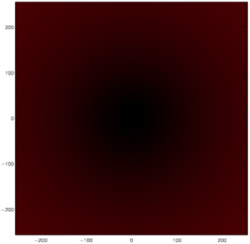


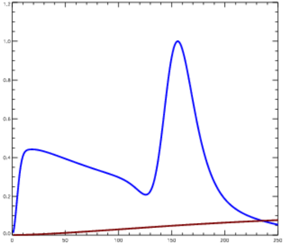
Martin White

What are Baryon Acoustic Oscillations?

(6) The baryons remain over-dense in a shell 100 Mpc/h in radius.



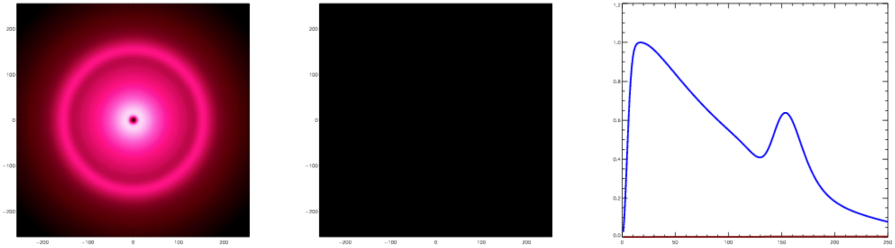




Martin White

What are Baryon Acoustic Oscillations?

(7) The final configuration is the original peak at the center and an echo in a shell roughly 100 Mpc/h in radius. The radius of this shell is known as the sound horizon. At later times galaxies will form. These galaxies will preferentially form in high-density regions, i.e. the shell over-density or the original over-density.



Martin White

Let's do a quick estimate

The challenge of the BAO method is primarily statistical: because this is a weak signal at a large scale, one needs to map enormous volumes of the universe to detect the BAO and obtain a precise distance measurement.

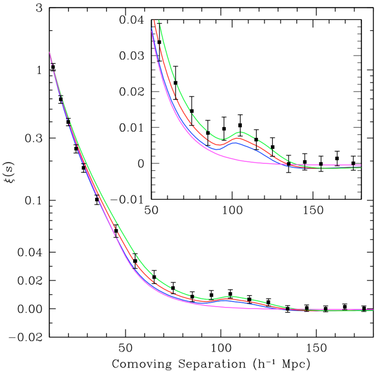
- Consider a typical Luminous Red Galaxy (LRG). How many LRGs do we expect in a BAO shell around it?
- Assume $n \sim 10^{-4}h^3/\text{Mpc}^{-3}$.
- With a shell width of $\Delta r = 20\,\text{Mpc}/h$ the volume of a shell around this LRG can be approximated as

$$V = \frac{4}{3}\pi(110^3 - 90^3) \approx 10^6(Mpc/h)^3$$

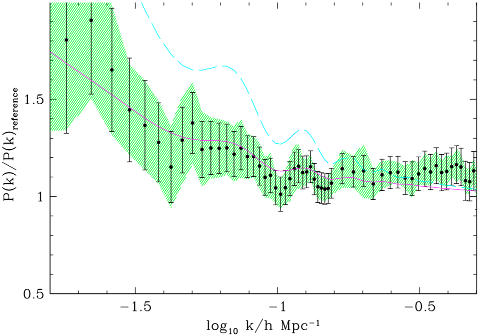
  which leads to about 100 LRGs in this volume.
- The correlation function predicts $\sim 1\%$ excess probability. So on average we have one extra galaxy in this volume... so we need a large volume to get good statistics.

First detection in 2005

SDSS-LRG



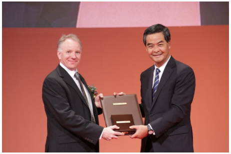
2dFGRS



Eisenstein et al. 2005

Cole et al. 2005

It’s worth doing



How can we measure the Baryon Acoustic Oscillation signal in the distribution of galaxies?

We have a 3D distribution of points (galaxies), which carries the BAO signal (and many other observables of interest). One possibility is to calculate the different moments of the point distribution. The general expression for the k-th order moment is

$$M_k = \int_{-\infty}^{+\infty} (x-\mu)^k f(x)dx.$$

The function $f(x)$ is the probability density function and the integral over $f(x)$ is 1. The first moment with $\mu = 0$ is known as the mean

$$\mu = \int_{-\infty}^{+\infty} xf(x)dx.$$

The second moment is the variance around the mean

$$\sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx.$$

The third moment is called the skewness and measures the asymmetry of the distribution...

For a distribution of discrete points the integrals turn into sums and we have e.g. for the variance

$$\sigma^2 = \frac{1}{N}\sum(x-\mu)^2$$

If the points are distributed as a Gaussian random field, we only need the first and second moment to describe the distribution.

A point distribution can be turned into an over-density field

$$\delta(\vec{x}) = \frac{\rho(\vec{x})}{\overline{\rho}} - 1,$$

where $\overline{\rho}$ is the mean density (of the Universe!).
We can easily Fourier transform such a density field

$$\delta(\vec{k}) = \sum \delta(\vec{x}) \exp(i\vec{k}\cdot\vec{x})$$

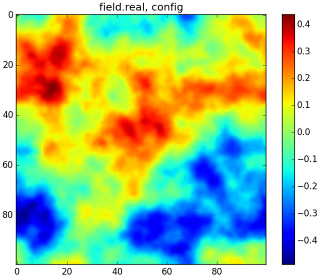
The power spectrum is then simply given by $P(\vec{k}) = \langle|\delta(\vec{k})|^2\rangle$.

In a Gaussian field the power spectrum carries all the information of the field because of Wick's theorem. Wick's theorem states that any ensemble average of products of variables can be obtained by a product of ensemble averages of pairs

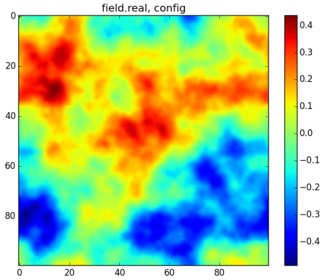
$$\langle\delta(k_1),...,\delta(k_{2p+1})\rangle = 0$$
$$\langle\delta(k_1),...,\delta(k_{2p})\rangle = \sum_{\text{all pair associations}} \prod_{\text{p pairs (i,j)}} \langle\delta(k_i)\delta(k_j)\rangle.$$

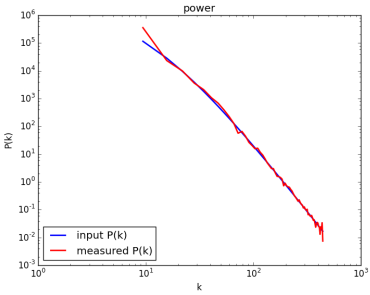
This means that the shape and normalization of $P(k)$ has all the information of the field $\delta(k)$. This is true shortly after inflation.

Does the power spectrum corresponding to this density field $\delta(\vec{x})$ have power on large scales or small scales?



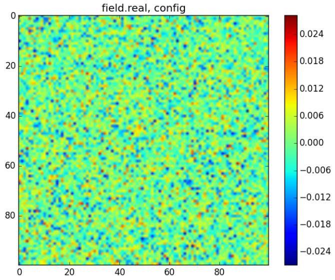
Does the power spectrum corresponding to this density field $\delta(\vec{x})$ have power on large scales or small scales?



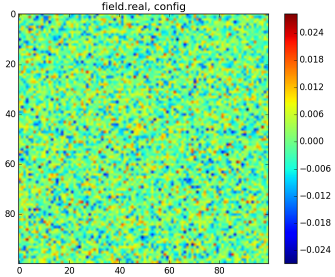


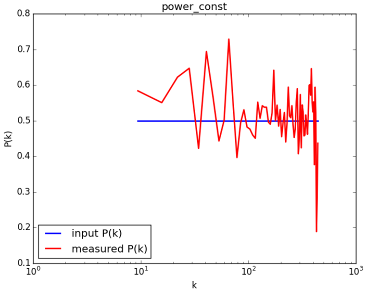
small k corresponds to large scale modes with roughly $k = 2\pi/L$.

Does the power spectrum corresponding to this density field $\delta(\vec{x})$ have power on large scales or small scales?



Does the power spectrum corresponding to this density field $\delta(\vec{x})$ have power on large scales or small scales?





Fourier space vs. configuration space

The correlation function and the power spectrum are just Fourier transforms of each other

$$P(\vec{k}) = \int_{-\infty}^{+\infty} \xi(\vec{r}) \exp(i\vec{k}\cdot\vec{r}) d\vec{r},$$

$$\xi(\vec{r}) = \frac{1}{(2\pi)^3} \int_{-\infty}^{+\infty} P(\vec{k}) \exp(-i\vec{k}\cdot\vec{r}) d\vec{k}.$$

Reference: Landy & Szalay (1993); Feldman, Kaiser & Peacock (1994)

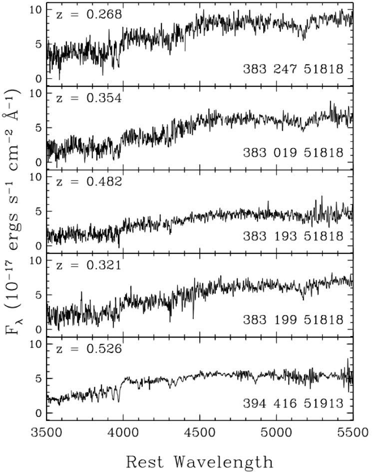
Photometric surveys vs. spectroscopic surveys

To get a redshift for each galaxy takes a lot of time. Do we need it?

- Using only the photometric data we can measure galaxy positions in two dimensions, but usually we have many more targets.
- The galaxy magnitude in each band can be used to get photometric redshifts (usually large errors and outliers, but often good enough for coarse separation into redshift bins).
- There are other interesting observables like galaxy shapes $\rightarrow$ weak lensing measurements.
- Multiple passes allow to search for transients (e.g. SN, gamma-ray bursts ...)
- The search for red-sequence galaxies allows the identification of clusters, see Rykoff et al. (2014)
- Surveys like DES and LSST will go down this route

Building a galaxy redshift survey

1. Assemble a homogeneous list of targets from an imaging survey
    - Magnitude cuts
    - Color cuts... useful to maximize volume
2. Take spectra to get the redshifts.
3.
4.
5.



Eisenstein et al. (2001)

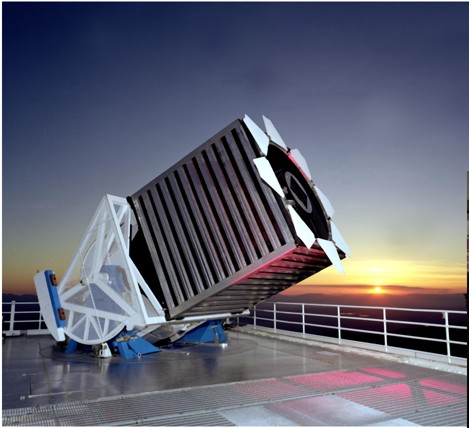
Building a galaxy redshift survey

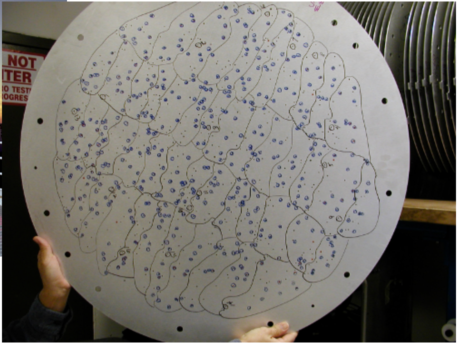
1. Assemble a homogeneous list of targets from an imaging survey
   - Magnitude cuts
   - Color cuts... useful to maximize volume
2. Take spectra to get the redshifts.
3.
4.
5.

Building a galaxy redshift survey

1. Assemble a homogeneous list of targets from an imaging survey
    - Magnitude cuts
    - Color cuts... useful to maximize volume
2. Take spectra to get the redshifts.
3. Make a three-dimensional galaxy density map, $\delta(\vec{x})$.
4. Estimate the correlation function/power spectrum.
5. Perform a likelihood analysis.

The BOSS survey

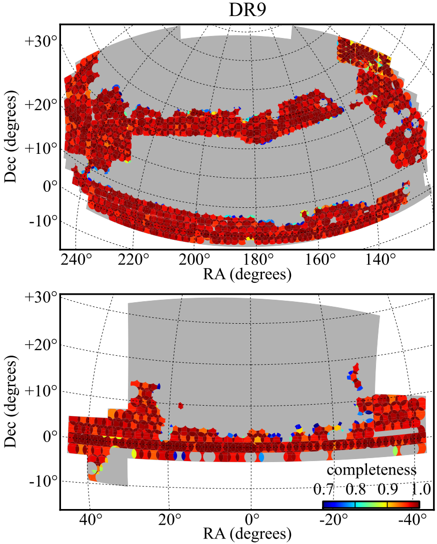




The BOSS survey

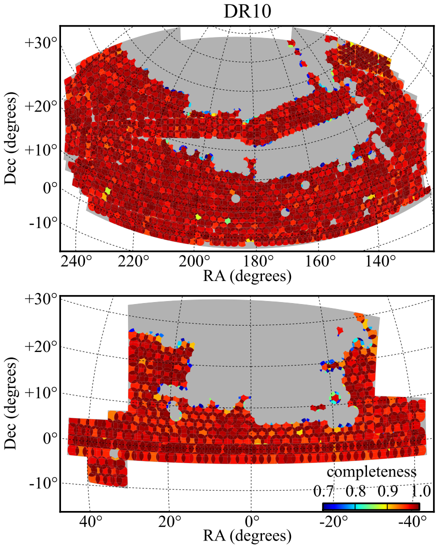
The BOSS survey

- Optimized for the measurement of Baryon Acoustic Oscillations.
- CMASS: $0.43 < z < 0.7$
- LOWz: < 0.43
- The effective volume is $6\,\mathrm{Gpc}^3$ for CMASS and $2.4\,\mathrm{Gpc}^3$ for LOWz.
- DR9: 3275 deg^2

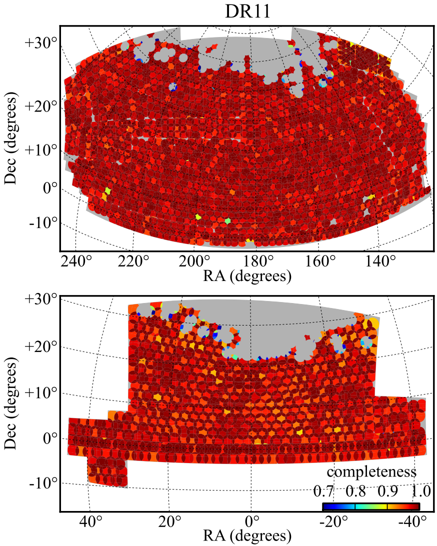


The BOSS survey

- Optimized for the measurement of Baryon Acoustic Oscillations.
- CMASS: $0.43 < z < 0.7$
- LOWz: < 0.43
- The effective volume is $6\,\mathrm{Gpc}^3$ for CMASS and $2.4\,\mathrm{Gpc}^3$ for LOWz.
- DR9: 3275 deg^2
- DR10: 6262.1 deg^2
-
-

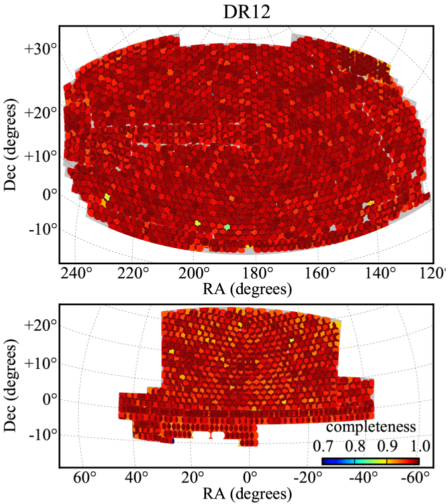


- Optimized for the measurement of Baryon Acoustic Oscillations.
- CMASS: $0.43 < z < 0.7$
- LOWz: < 0.43
- The effective volume is $6\,\mathrm{Gpc}^3$ for CMASS and $2.4\,\mathrm{Gpc}^3$ for LOWz.
- DR9: 3275 deg^2
- DR10: 6262.1 deg^2
- DR11: 8509.6 deg^2
-



The BOSS survey

- Optimized for the measurement of Baryon Acoustic Oscillations.
- CMASS: $0.43 < z < 0.7$
- LOWz: < 0.43
- The effective volume is $6\,\mathrm{Gpc}^3$ for CMASS and $2.4\,\mathrm{Gpc}^3$ for LOWz.
- DR9: 3275 deg^2
- DR10: 6262.1 deg^2
- DR11: 8509.6 deg^2
- DR12: 10000 deg^2



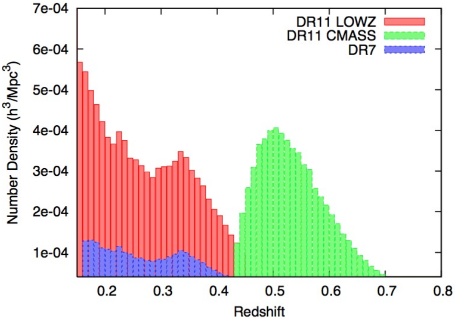
credit: Molly Swanson

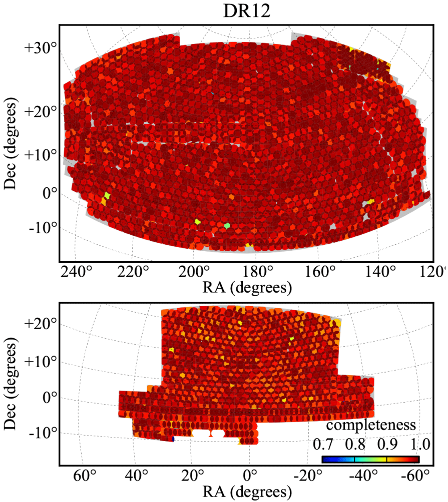
Building a galaxy redshift survey

1. Assemble a homogeneous list of targets from an imaging survey
   - Magnitude cuts
   - Color cuts... useful to maximize volume
2. Take spectra to get the redshifts.
3. Make a three-dimensional galaxy density map, $\delta(\vec{x})$.
4. Estimate the correlation function/power spectrum.
5. Perform a likelihood analysis.

Building a galaxy redshift survey – selection function

To get a density map we first need to define the probability of finding a galaxy as a function of sky position and redshift. This is called a survey selection function.



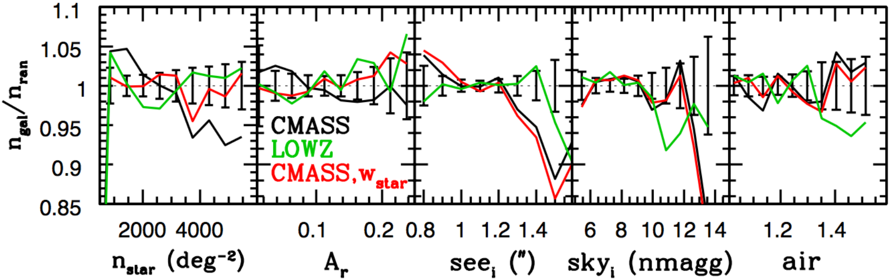


Building a galaxy redshift survey – selection function

We characterize the selection function with a set of random data points

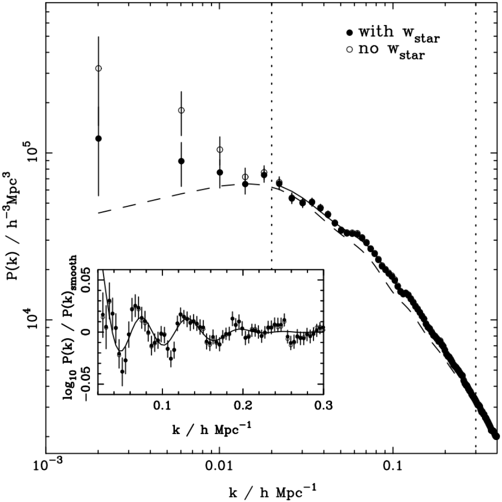
1. First we generate random positions on the sky (RA, DEC) with probability proportional to the angular completeness map.
2. We then assign redshifts from the actual catalogue randomly to these data points (by definition they have $\xi = 0$).

Understanding your survey selection can be very difficult...



Ross et al. (2012)

Building a galaxy redshift survey – selection function



Ross et al. (2012)

Clustering measurements – correlation function

The correlation function is defined via the excess probability of finding a galaxy pair at separation r

$$dP = \overline{n}^2 \left[1 + \xi(r)\right] dV_1 dV_2.$$

In practice we count the number of unique data-data, $DD(r)$, data-random, $DR(r)$, and random-random, $RR(r)$ pairs. The Landy-Szalay estimator is

$$\xi(r) = \frac{DD(r)N_{dd} - 2DR(r)N_{dr} + RR(r)}{RR(r)}$$

with $N_{dd} = n_r(n_r - 1)/(n_d(n_d - 1)$, $N_{dr} = n_r(n_r - 1)/(2n_d n_r)$.
References: Davis & Peebles (1983), Landy & Szalay (1993),

Since we have a direct estimate of the galaxy over-density at each point, we can compute the Fourier transform of the observed galaxy density field:

$$P^{\rm obs}(\vec{k}) = \frac{\int d\vec{r} \int d\vec{r}' w(\vec{r}')w(\vec{r})\langle[n_g(\vec{r}) - \alpha n_r(\vec{r})][n_g(\vec{r}') - \alpha n_r(\vec{r}')]\rangle e^{ik\cdot(\vec{r}-\vec{r}')}}{\int d\vec{r}\bar{n}^2(\vec{r})w^2(\vec{r})}$$

with $\alpha = N_{\rm gal}/N_{\rm ran}$. The observed power spectrum, $P^{\rm obs}(\vec{k})$, is related to the true galaxy power spectrum, $P^{\rm true}(\vec{k})$, as

$$P^{\rm obs}(\vec{k}) = \int \frac{d\vec{k}'}{(2\pi)^3} P^{\rm true}(\vec{k}')|W(\vec{k} - \vec{k}')|^2 + (1+\alpha)\frac{\int d\vec{r}\bar{n}(\vec{r})w^2(\vec{r})}{\int d\vec{r}\bar{n}^2(\vec{r})w^2(\vec{r})}$$
$$- \frac{|W(\vec{k})|^2}{|W(0)|^2} \int d\vec{k}' P^{\rm true}(\vec{k}')|W(\vec{k}')|^2$$

with W being the survey window function.

Since we have a direct estimate of the galaxy over-density at each point, we can compute the Fourier transform of the observed galaxy density field:

$$P^{\rm obs}(\vec{k}) = \frac{\int d\vec{r} \int d\vec{r}' w(\vec{r}')w(\vec{r})\langle [n_g(\vec{r}) - \alpha n_r(\vec{r})][n_g(\vec{r}') - \alpha n_r(\vec{r}')]\rangle e^{ik\cdot(\vec{r}-\vec{r}')}}{\int d\vec{r}\bar{n}^2(\vec{r})w^2(\vec{r})}$$

with $\alpha = N_{\rm gal}/N_{\rm ran}$. The observed power spectrum, $P^{\rm obs}(\vec{k})$, is related to the true galaxy power spectrum, $P^{\rm true}(\vec{k})$, as

$$P^{\rm obs}(\vec{k}) = \int \frac{d\vec{k}'}{(2\pi)^3} P^{\rm true}(\vec{k}')|W(\vec{k}-\vec{k}')|^2 + (1+\alpha)\frac{\int d\vec{r}\bar{n}(\vec{r})w^2(\vec{r})}{\int d\vec{r}\bar{n}^2(\vec{r})w^2(\vec{r})}$$
$$- \frac{|W(\vec{k})|^2}{|W(0)|^2} \int d\vec{k}' P^{\rm true}(\vec{k}')|W(\vec{k}')|^2$$

with W being the survey window function. To fully understand the measured power spectrum we have to account for the discrete sampling of the density field (shot noise), the window function and the integral constraint!

- Since the survey window multiplies the configuration density field, it results in a convolution in Fourier-space.
- In general the window function is given by the survey selection function.
- To account for the window function effect we can deconvolve the data or convolve the model.
- The convolution integral is

$$P^{\rm conv}(\vec{k}) = \int d\vec{k}' P^{\rm true}(\vec{k}')|W(\vec{k}-\vec{k}')|^2 - \frac{|W(\vec{k})|^2}{|W(0)|^2}\int d\vec{k}' P^{\rm true}(\vec{k}')|W(\vec{k}')|^2$$

- A straight forward calculation of this integral would have the complexity $\mathcal{O}(N^2_{\rm modes})$. But if the window function is compact, this can be simplified considerably.

What to measure? Your choice of data vector matters

Multipole decomposition:

$$P(k,\mu) = \sum_\ell P_\ell(k)\mathcal{L}_\ell(\mu)$$

$$\xi(s,\mu) = \sum_\ell \xi_\ell(s)\mathcal{L}_\ell(\mu)$$

And they are simply related by a Henkel transform

$$\xi_\ell(s) = i^\ell \int \frac{k^2 dk}{2\pi^2} P_\ell(k) j_\ell(ks)$$

Clustering measurements – power spectrum

- We can express the window function convolution in terms of the power spectrum multipoles as:

$$P_\ell^{\rm conv}(k) = 2\pi \int dk' k'^2 \sum_L P_L^{\rm true}(k')|W(k,k')|^2_{\ell L}$$
$$- 2\pi \frac{|W(k)|^2_\ell}{|W(0)|^2_0} \int dk' k'^2 \sum_L P_L^{\rm true}(k')|W(k')|^2_L \frac{2}{2L+1}$$

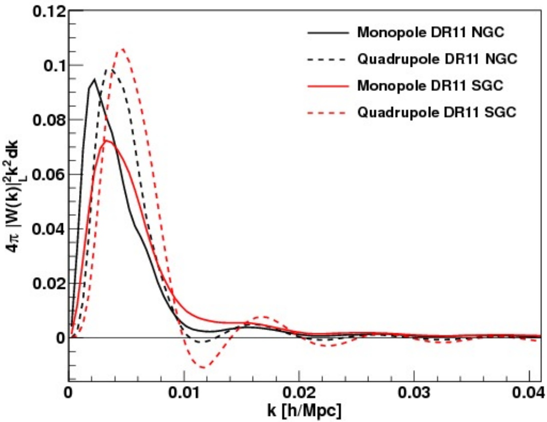
- The equation above only contains the mode amplitude $|\vec{k}|$ instead of the mode vector. For details check out our DR11 paper, Beutler et al. (2014).

Clustering measurements – power spectrum

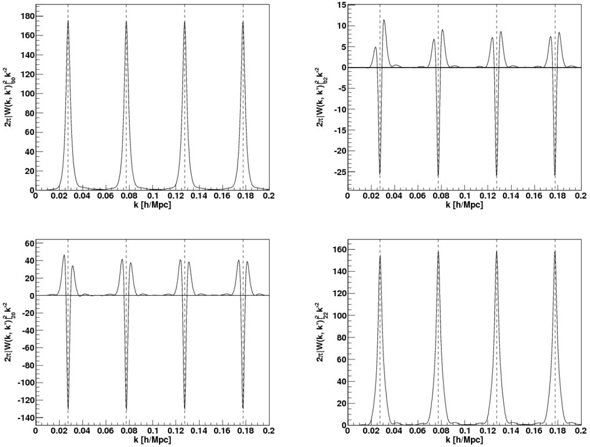
$$2\pi\frac{|W(k)|^2_\ell}{|W(0)|^2_0}\int dk'k'^2\sum_L P_L^{\rm true}(k')|W(k')|^2_L\frac{2}{2L+1}$$

- The integral constraint comes from the fact that we assumed that the mean density of the survey is equal to the mean density of the Universe. Sample variance tells us that this is wrong...
- Therefore our measured power spectrum has the condition $P(k\to 0)=0$ by design.
- We don't really care about the k=0 modes, since we usually don't include it in our analysis. However, the window function couples the $k=0$ mode with larger modes, depending on the width of the window function.
- Therefore the artificial $P(k\to 0)=0$ condition can impact our constraints.
- Reference: Peacock & Nicholson (1991)

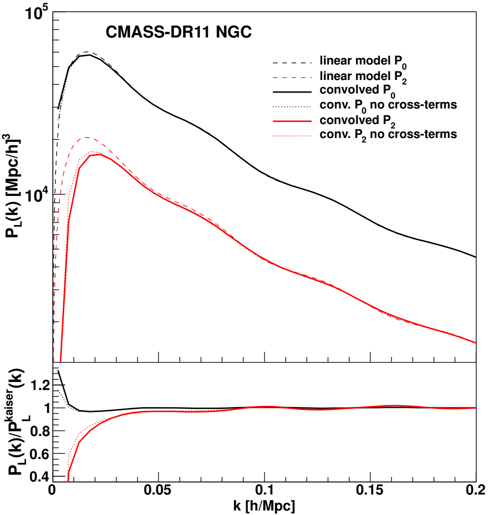
Clustering measurements – power spectrum



Beutler et al. (2014)

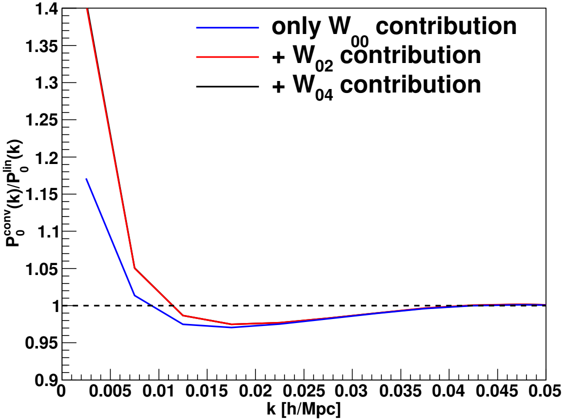


Beutler et al. (2014)



Beutler et al. (2014)

Ignoring the window function can lead to biased constraints



- The power spectrum multipoles can be calculated with FFT's. This allows a complexity of $\mathcal{O}(N \log N)$ instead of the naive $\mathcal{O}(N^2)$. Such a method does not exist for the configuration-space method.
- In principle FFTs cannot be applied to a line-of-sight dependent quantity, since the FFT assumes periodic boundary conditions. However one can separate the FFT into 1D FFTs when using the power spectrum multipoles.

$$P_0(k) = \frac{1}{2A}\int d\Omega_k \left[F_0(\vec{k})F_0^*(\vec{k}) - S\right]$$
$$P_2(k) = \frac{5}{4A}\int d\Omega_k F_0(\vec{k})\left[3F_2^*(\vec{k}) - F_0^*(\vec{k})\right]$$

with

$$F_n(\vec{k}) = \int d\vec{r}(\hat{k}\cdot\hat{r})^n F(\vec{r})e^{i\vec{k}\cdot r}$$

and with $\hat{k}\cdot\hat{r} = (k_x r_x + k_y r_y + k_z r_z)/kr$ we have for e.g. the quadrupole component

$$F_2(\vec{k}) = \frac{1}{k^2}\Bigg(k_x^2 B_{xx}(\vec{k}) + k_y^2 B_{yy}(\vec{k}) + k_z^2 B_{zz}(\vec{k})$$
$$+ 2\left[k_x k_y B_{xy}(\vec{k}) + k_x k_z B(\vec{k}) + k_y k_z B_{yz}(\vec{k})\right]\Bigg)$$

where $B_{ij}(\vec{k}) = \int d\vec{r}\frac{r_i r_j}{r^2}F(\vec{r})e^{i\vec{k}\cdot\vec{r}}$. Reference: Bianchi et al. (2015)

Is there an optimal weight for such a measurement?

$$\left(\frac{\sigma_{P(k)}}{P(k)}\right)^2 = \frac{1}{V_k}\int d\vec{k}' \left|Q(\vec{k}') + \frac{S(\vec{k}')}{P(\vec{k})}\right|^2$$

with S being the shot noise and

$$Q(\vec{k}) = \frac{1}{A}\int d\vec{x} n_g^2(\vec{x}) w_{\rm FKP}(\vec{x}) e^{-i\vec{k}\cdot\vec{x}}.$$

Minimizing the relative error leads to the optimal weight:

$$w_{\rm FKP}(\vec{x}) = \frac{1}{1+\overline{n}(\vec{x})P(k)}$$

- Regions with a small number of galaxies get a larger weight.
- Reference: Feldman, Kaiser & Peacock (1993)

BOSS data

- Now you can compute $\xi(r)$ and $P(k)$ yourself.
- You can download SDSS DR10 large scale structure catalogs here: `http://data.sdss3.org/sas/dr10/boss/lss/`
- And make your own subsamples following these instructions: `https://www.sdss3.org/dr10/tutorials/lss_galaxy.php`

What to measure? Your choice of data vector matters

- Several recent papers (e.g. Percival et al. 2014) pointed out that errors in the data covariance matrix propagate into errors on cosmological parameters; this can be mitigated by information compression i.e., choosing a shorter data vector.
- Percival et al. (2014) derives an optimal bin size for $P(k)$ or $\xi(s)$ for BAO and RSD measurements.
- From a theoretical point of view, your data vector should be restricted to scales where you have good reason to believe your model is sufficiently accurate
- Parameter constraints can be quite sensitive to this choice, since the number of available modes is $N_k \propto k^2\Delta k$!

What to measure? Your choice of data vector matters

Multipole decomposition:

$$P(k,\mu) = \sum_{\ell} P_\ell(k)\mathcal{L}_\ell(\mu)$$

$$\xi(s,\mu) = \sum_{\ell} \xi_\ell(s)\mathcal{L}_\ell(\mu)$$

And they are simply related by a Henkel transform

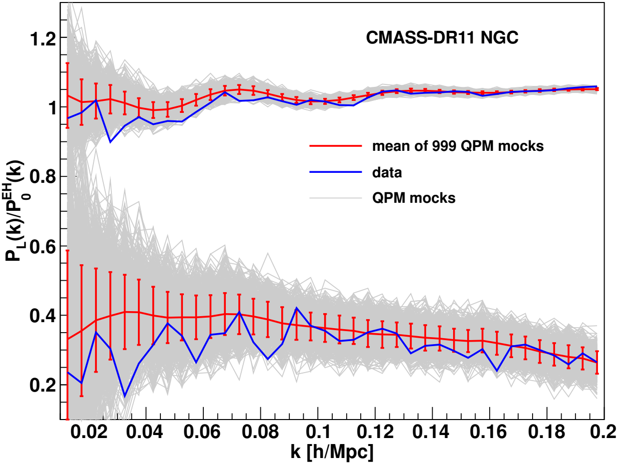
$$\xi_\ell(s) = i^\ell \int \frac{k^2 dk}{2\pi^2} P_\ell(k) j_\ell(ks)$$

Clustering measurements – Uncertainties

- For most practical purposes, there are no real measurement uncertainties in LSS; that is, the error in the angular position of the galaxy (RA, DEC) or redshift is much smaller than the scales on which you are interested in measuring correlations.
- Survey Geometry – the finite volume and complicated geometry of real surveys induce correlations between neighboring modes and change the effective number of independent modes contributing to bandpower $P(k_i)$.
- The galaxy density field is given by $\delta_g(k) = b\delta_m(k) + \epsilon$. Our density field estimate has an error contribution from sample variance and shot noise.
- The density field looks fairly Gaussian on large scales but has strong non-Gaussian features at small scales (introduced by gravitational evolution). So the assumption of Gaussian errors only works on very large scales.

Clustering measurements – Uncertainties

- A brute force solution is to generate hundreds or thousands of synthetic surveys (as realistic as possible), and measure the covariance matrix of your observable from this set.
- Approximations to real (expensive to compute) gravitational dynamics + galaxy prescription include Poisson sampling of lognormal matter density fields [easiest to generate], "PTHalos", "Addgals", "COLA", "QPM", "PATCHY", real N-body sims + halo model [currently infeasible], hydrodynamic simulations including galaxy formation [REALLY infeasible]. Reference: e.g. White, Tinker & McBride (2014)
- Then apply survey selection function to include effects of geometry, sampling variance (shot noise), veto masks, etc.
- Since we need to populate the mock survey with galaxies it needs to get the halos right. These simulations cannot just focus on BAO scales...



Beutler et al. (2014)

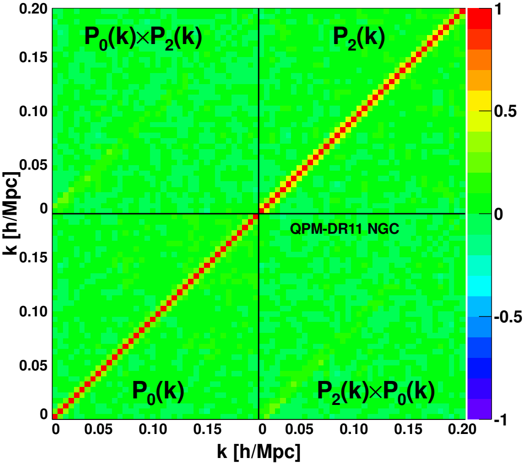
Now we can estimate the covariance matrix.

$$C_{ij} = \frac{1}{N-1}\sum_{n=1}^{N}\left[P_\ell(k_i) - \overline{P}_\ell(k_i)\right]\left[P_\ell(k_j) - \overline{P}_\ell(k_j)\right].$$

This matrix simplifies to the variance if the non-diagonal elements can be neglected. It is essential for any likelihood analysis

$$\chi^2 = \sum_{ij}(D_i - M_i)C_{ij}^{-1}(D_j - M_j)$$

and $\mathcal{L} \sim \exp(-\chi^2/2)$.



Beutler et al. (2014)

Summary so far

- Determine the survey selection function. If you don't understand the survey selection your data is useless.
- Decide whether you want to do the analysis in Fourier space or in configuration space.
- Carefully decide on your binning and the range of scales you want to include in the analysis.
- Determine your uncertainties (often the most difficult part).

Outline of the week

Tuesday:

- LSS and BAO, basics of likelihood analysis
- hands-on session: Implement your own MCMC

Wednesday:

- continue LSS and BAO... including basics of LSS analysis (power spectrum, correlation function)
- hands-on session: implement a power spectrum estimator, estimate a covariance matrix

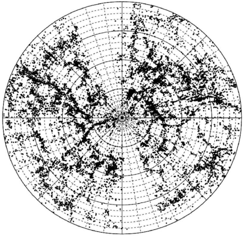
Thursday:

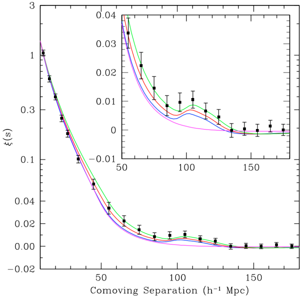
- LSS and RSD
- hands-on session: Use MCMC, power spectrum and covariance matrix to make a likelihood analysis (if time include shot noise and galaxy bias)

Friday:

- Beyond BAO and RSD... neutrino mass and non-Gaussianity, Future outlook, Euclid and DESI
- hands-on session: finish likelihood analysis

Geometric constraints from galaxy surveys





We have two choices:

1. Stick to statistics in observer coordinates (RA, DEC, z) – e.g., compute the angular correlation function/power spectra in MANY redshift bins.
2. Assume a fiducial cosmology to convert RA, DEC and z to co-moving coordinates; account for this choice in theory calculation.

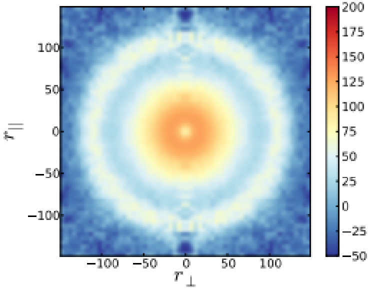
- Option 1 would require many z bins and therefore a large data vector to retain all the information.
- Therefore people usually choose option 2 (converting RA, DEC and z into co-moving coordinates using a fiducial cosmology).
- This makes our measurement dependent on the assumed fiducial cosmology!
- However, we can account for this in our theoretical modeling. The effect of the fiducial cosmology on the distance measure $\chi(z) = (1+z)D_A(z) = \int_0^z c\frac{dz}{H(z)}$ is smooth (they are integrals!), and so the resulting geometric distortions can be modeled very accurately.

Assume a survey in a very narrow redshift bin $z_s \pm dz/2$. Select a fiducial cosmology with which to measure ξ:

- The conversion between θ angular and transverse co-moving separation $x_\perp$ is $x_\perp \approx (1+z)D_{A,\mathrm{fid}}(z)\theta$
- The conversion between Δz and line-of-sight (LOS) co-moving separation $x_\parallel$ is $x_\parallel = \frac{c\Delta z}{H_{\mathrm{fid}}(z)}$
- The theoretical prediction based on a model with a different $D_A(z)$ and $H(z)$ for the observed ξ_{obs} [measured with $D_A(z)$ and $H(z)$] is

$$\xi^{obs}(s_\perp, s_\parallel) = \xi^{true}\left(s_\perp \frac{D_A(z)}{D_{A,\mathrm{fid}}(z)}, s_\parallel \frac{H_{\mathrm{fid}}(z)}{H(z)}\right)$$

- With a thicker redshift slice this is still a good approximation



- Let's assume our correlation function has a feature at a characteristic scale e.g. $s_{\rm BAO} = 150\,$Mpc.
- Suppose we observed this feature at $s_\perp$ and $s_\parallel$;

$$\xi^{\rm obs}(s_\perp, s_\parallel) = \xi^{\rm true}\left(s_{\rm BAO} = s_\perp\frac{D_A(z_{\rm eff})}{D_{A,fid}(z_{\rm eff})}, s_{\rm BAO} = s_\parallel\frac{H_{\rm fid}(z_{\rm eff})}{H(z_{\rm eff})}\right)$$

- So we actually measured $D_A(z_{\rm eff})/(s_{\rm BAO}D_{\rm A,fid}(z_{\rm eff}))$ and $H_{\rm fid}(z_{\rm eff})/(s_{\rm BAO}H(z_{\rm eff}))$.

Geometric constraints from galaxy surveys

- So the Alcock-Paczynski effect constrains the relative angular and line-of-sight separation, $F_{\mathrm{AP}}(z) = (1+z)D_A(z)H(z)/c$.
- The AP effect requires no standard ruler (just isotropy), so information from all scales is useful.
- Reference: Alcock & Paczynski (1979), Ballinger, Peacock & Heavens (1996), Matsubara & Suto (1996)
- In case of the BAO often people only quote the angular average quantity $D_V = \left[(1+z)^2 D_A^2(z)cz/H(z)\right]^{1/3}$. This means our BAO constraint is $D_V(z_{eff})/s_{\mathrm{BAO}}$.
- Reference: Eisenstein et al. (2005)

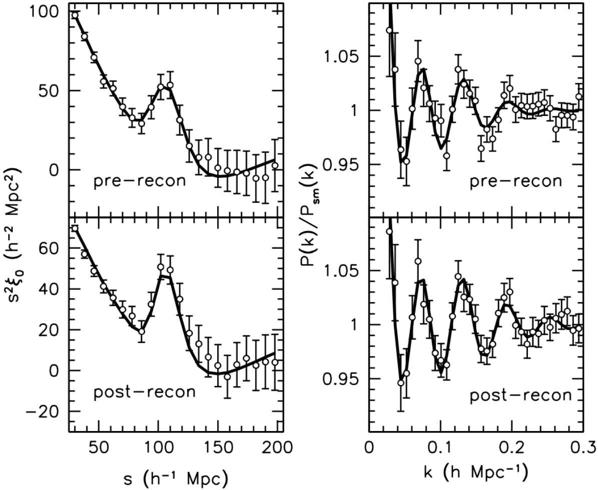
Summary Alcock-Paczynski effect

- The fact that we assume a fiducial cosmology to turn the observed RA, DEC and redshift into co-moving coordinates before the likelihood analysis means that our observed statistic, $\xi(r)$ or $P(k)$ depends on the fiducial cosmology.
- We can correct for that by scaling the measured statistic with the ratios $\alpha_{\perp} = D_A/D_{\mathrm{A,fid}}$ and $\alpha_{\parallel} = H_{\mathrm{fid}}/H$.
- The Alcock-Paczynski effect introduces an observable which is interesting just by itself, $F_{\mathrm{AP}}(z) = (1+z)D_A(z)H(z)/c$, but the constraints from this are far weaker than the BAO constraints.
- The AP effect comes from all scales, not just the BAO.
- Since the AP effect introduces anisotropies we have to worry about degeneracies with redshift-space distortions... more about this on Thursday.

Recap Baryon Acoustic Oscillations

- Baryon Acoustic Oscillations get imprinted in the matter distribution in the very early Universe.
- The process which produces BAO ends after $\sim 300\,000$ years.
- After that the BAO scale follows (mostly) linear evolution.
- Since galaxies trace the underlying density field we can use galaxy surveys to measure this scale.

Recap Baryon Acoustic Oscillations



Anderson et al. (2014)

Recap Baryon Acoustic Oscillations

So we look for:
$\rightarrow$ Oscillations in Fourier space and
$\rightarrow$ a peak in configuration-space

So we look for:
$\rightarrow$ Oscillations in Fourier space and
$\rightarrow$ a peak in configuration-space Questions:

- There is a special scale imprinted in the density field... so what?

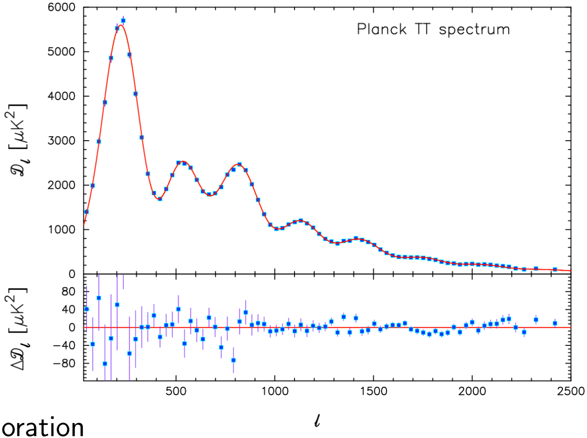
So we look for:
$\rightarrow$ Oscillations in Fourier space and
$\rightarrow$ a peak in configuration-space Questions:

- There is a special scale imprinted in the density field... so what?
- What would be the properties of a galaxy survey optimized for such a measurement?

So we look for:
$\rightarrow$ Oscillations in Fourier space and
$\rightarrow$ a peak in configuration-space Questions:

- There is a special scale imprinted in the density field... so what?
- What would be the properties of a galaxy survey optimized for such a measurement?
- How can we get a model for $\xi(r)/P(k)$ to extract the BAO scale?

BAO in the photon temperature power spectrum



Planck collaboration

$$r_s(z_*) = 144.75 \pm 0.66\,\mathrm{Mpc} \quad (0.46\%)$$
$$r_s(z_d) = 147.34 \pm 0.64\,\mathrm{Mpc} \quad (0.43\%)$$

Using BAO as a standard ruler

- So we know the absolute size of the sound horizon to high precision.
- Measuring its apparent size in the distribution of galaxies allows a distance measurement to the redshift of the galaxies.
- The large size of the acoustic scale protects this clustering feature from nonlinear structure formation in the low-redshift universe.
- We can measure its angular size $\Delta\theta$ which yields the angular diameter distance D_A at that redshift

$$D_A(z) \propto \int_0^z \frac{dz'}{H(z')}$$

- By measuring the redshift interval Δz we can measure the Hubble parameter $H(z)$.

$$c\Delta z = H(z) r_s$$

- While these are independent observables, they are of course connected and allow a valuable cross-check.

- The size of the sound horizon can easily be calculated

$$r_s = \int_0^{t_*} \frac{c_s(t)}{a(t)} dt = \int_{z_*}^{\infty} \frac{c_s(z)}{H(z)} dz$$

with

$$c_s(z) = \frac{c}{\sqrt{3\left(1 + \frac{3\rho_b}{4\rho_\gamma}\right)}} \propto \Omega_b h^2$$

$$H(z) \propto \Omega_m h^2$$

- While SN distance measurements are empirically calibrated and cannot be derived from theory, the BAO scale and its existence can be predicted directly by theory.

The fundamentals of a BAO survey

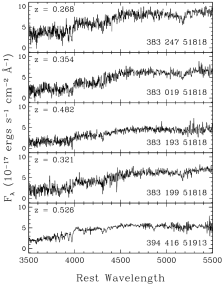
1. Select a tracer of the mass density field (high bias is usually good since the signal scales with the amplitude of the clustering... but see systematics discussion later).
2. Calculate the correlation function or power spectrum $\langle\delta^2\rangle$ of these galaxies and locate the sound horizon feature.
3. Measure the angular size $\Delta\theta$ and redshift width Δz related to the sound horizon at a variety of redshifts, z.
4. Compare to the sound horizon in the CMB to get $D_A(z)$ and $H(z)$.
5. Compare to theoretical predictions of $D_A(z)$ and $H(z)$ and test different cosmological models with e.g. varying dark energy equation of state $\Omega_\Lambda a^{-3(1+w)}$

What makes a good BAO galaxy survey

1. We have two error contributions, sample variance and shot noise.
2. The BAO signal and sample variance are proportional to the power spectrum amplitude, while the shot noise is constant.
3. Sample variance can be reduced by increasing the volume.
4. Shot noise can be reduced by increasing the density.
5. So there is an optimization problem:
    - Limited telescope time.
    - Different trace galaxies need different exposer time.
    - A highly biased tracer allows a lower density...

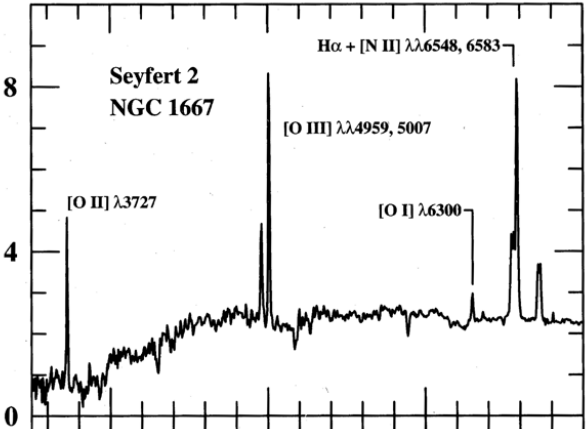
When designing an experiment, Fisher matrices are a quantitative way to answer such questions (see e.g. Font-Ribera 2014).
$\rightarrow$ LRGs seem to be close to optimal!

What makes a good BAO galaxy survey



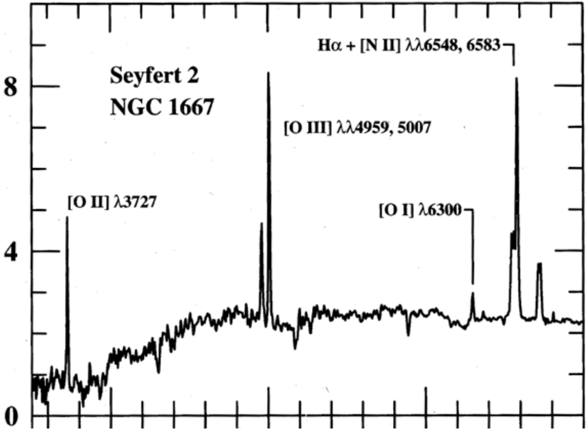
- Luminous Red Galaxies (LRGs) are dominated by red light (old stars). Not much star formation going on...
- Their spectra contains the prominent 4000 Angstrom break which makes obtaining redshifts quick.
- However, the density of LRGs gets quite low when going beyond redshift 0.7... so future galaxy surveys at redshifts higher than BOSS will target ELG (emission line galaxies) and quasars.

Emission Line Galaxies (ELGs): Young bright stars in star-forming galaxies photo-ionize surrounding gas.



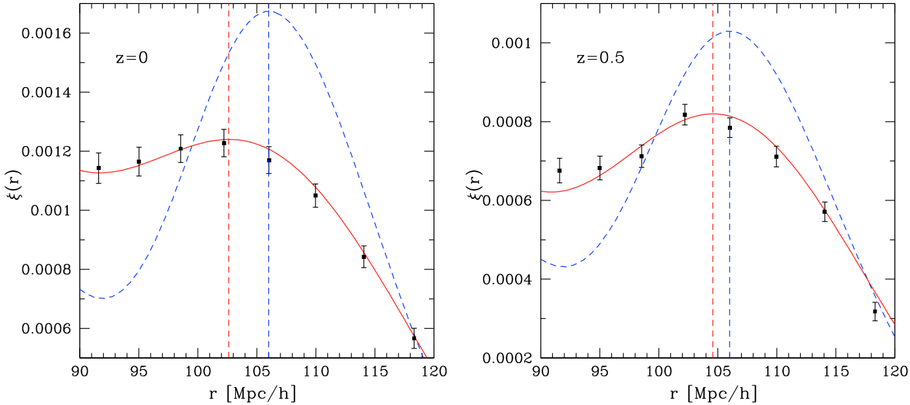
What makes a good BAO galaxy survey

WiggleZ, DESI (ground-based optical): [OII] doublet
Euclid/WFIRST (space-based infrared): Hα, [OIII]



Fitting the BAO – philosophy

- Lots of complicated physical effects alter the galaxy $\xi(r)/P(k)$ away from its linear theory behavior:
  - Non-linear gravitational evolution.
  - Non-linear biasing between tracers and matter field.
  - Non-linear redshift space distortions.
- At the same time, the BAO feature is located at very large scales and we might be able to get away with a linear model.
- We will exploit the fact that the messy non-linearities should generically be smooth, and marginalize over broad-band (smooth) contributions to the power spectrum/correlation function.
- We will use a smoothed template for the BAO feature.



$$\xi(r,z) = [e^{-r^2/\sigma^2} * \xi_0](r,z) + \xi_{\rm MC}(r,z)$$

Crocce & Scoccimarro 2008

- Start with linear $P(k)$ and separate the broadband shape, $P^{\rm sm}(k)$, and the BAO feature $O^{\rm lin}(k)$. Include a damping of the BAO feature:

$$P^{\rm sm,lin}(k) = P^{\rm sm}(k)\left[1 + (O^{\rm lin}(k/\alpha) - 1)e^{-k^2\Sigma^2_{\rm nl}/2}\right]$$

- add broadband nuisance terms

$$A(k) = a_1 k + a_2 + \frac{a_3}{k} + \frac{a_4}{k^2} + \frac{a_5}{k^3}$$

- Marginalize to get P(α)

$$P^{\rm fit}(k) = B^2 P^{\rm sm,lin}(k/\alpha) + A(k)$$

Fitting the BAO – test with mocks

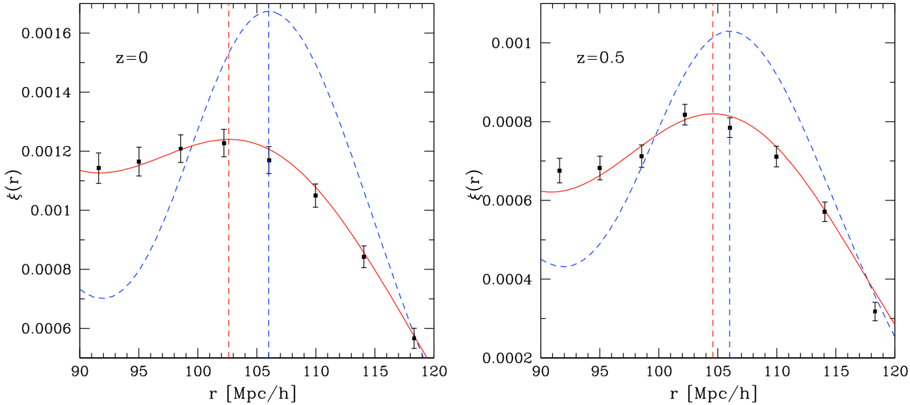
Estimator	$\langle\alpha\rangle$	S_α	$\langle\sigma\rangle$	$\langle\chi^2\rangle$/dof
DR11				
Consensus $P(k)$+$\xi(s)$	**1.0000**	**0.0090**	**0.0088**	
combined $P(k)$	1.0001	0.0092	0.0089	
combined $\xi(s)$	0.9999	0.0091	0.0090	
post-recon $P(k)$	1.0001	0.0093	0.0090	28.6/27
post-recon $\xi_0(s)$	0.9997	0.0095	0.0097	17.6/17
pre-recon $P(k)$	1.0037	0.0163	0.0151	27.7/27
pre-recon $\xi_0(s)$	1.0041	0.0157	0.0159	15.7/17
DR10				
post-recon $P(k)$	1.0006	0.0117	0.0116	28.4/27
post-recon $\xi_0(s)$	1.0014	0.0122	0.0126	17.2/17
pre-recon $P(k)$	1.0026	0.0187	0.0184	27.7/27
pre-recon $\xi_0(s)$	1.0038	0.0188	0.0194	15.8/17

Anderson et al. (2014)

The systematics in the BAO

- Any measurement which is based on galaxy clustering has to deal with systematic errors due to (1) non-linear structure formation, (2) redshift-space distortions, and (3) galaxy clustering bias.
- The galaxies might have formed following the density field set by the early Universe physics, but they do evolve with time. Gravitational interaction pulls galaxies away from their initial position. While these interactions are not significant at the sound horizon scale ($\sim 150\,$Mpc) they are important on the scale of the BAO width ($\sim 10\,$Mpc) and therefore can smear out the signal.
- This effect can be modeled by convolving the correlation function with a smoothing kernel $\exp(-r^2/\sigma^2)$, with σ depending on z.
- And there are further second order effects...

The systematics in the BAO



$$\xi(r, z) = [e^{-r^2/\sigma^2} * \xi_0](r, z) + \xi_{\rm MC}(r, z)$$

Crocce & Scoccimarro 2008

The systematics in the BAO

- Let's describe the over density field as a sum of small scale modes and large scale modes $\delta = \delta_L + \delta_s$.
- Regions where δ_L is high effectively live in a closed Universe, which undergoes larger growth. Regions where δ_L is low effectively live in an open Universe, which undergoes a smaller growth. In high density regions the BAO scale gets contracted, while in low density regions the BAO scale gets stretched.
- If both regions contributed similarly to the final correlation function, the two effects would cancel out. However, high density regions have a higher number of possible target galaxies and therefore will effectively contribute more. This leads to a shift in the correlation function to smaller scales.
- The effect is $\sim 0.3\% \, [D(z)/D(0)]^2$ (see e.g. Seo et al. 2010 or Metha et al. 2011)
- See Padmanabhan, White & Cohn 2009 for a proper PT derivation...

Density field reconstruction

- Smooth the density field to filter out high-k non-linearities.

$$\delta'(\vec{k}) \rightarrow e^{-\frac{k^2R^2}{4}}\delta(\vec{k})$$

- Solve the Poisson eq. to obtain the gravitational potential

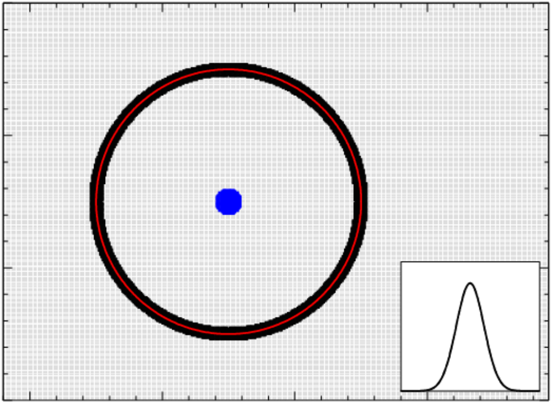
$$\nabla^2\phi = \delta$$

- The displacement (vector) field is given by

$$\Psi = \nabla\phi$$

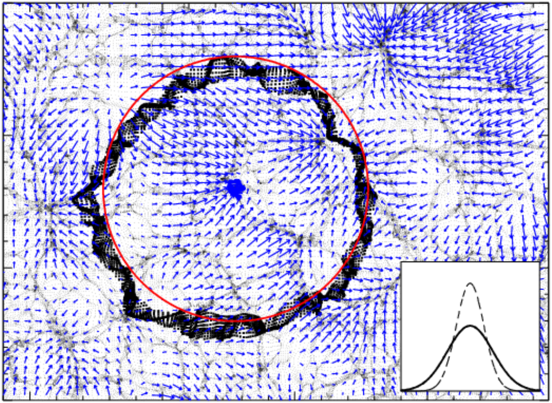
- Now we calculate the displaced density field by shifting the original particles.

Density field reconstruction



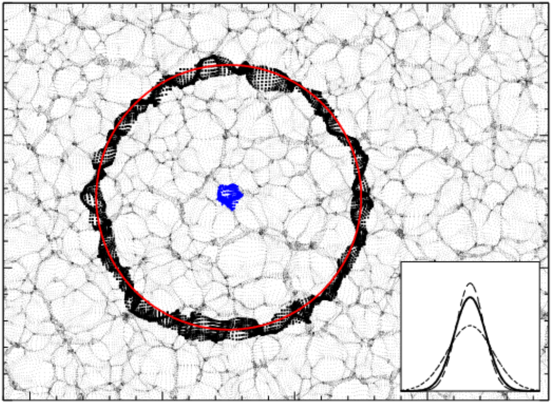
Padmanabhan et al. 2012

Density field reconstruction



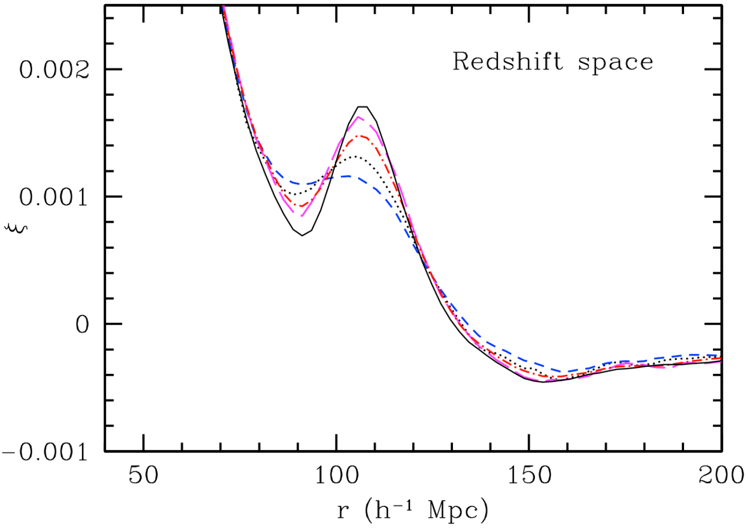
Padmanabhan et al. 2012

Density field reconstruction



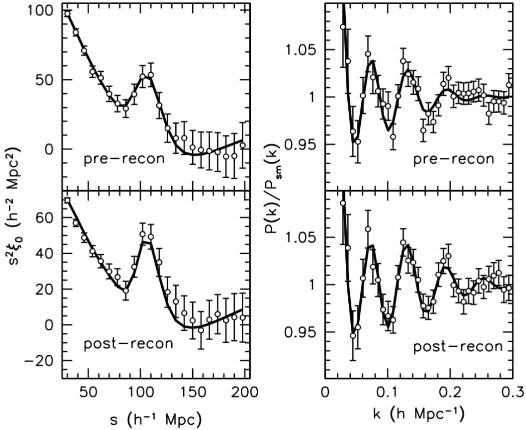
Padmanabhan et al. 2012

Density field reconstruction (it works in simulations)



Eisenstein et al. 2007

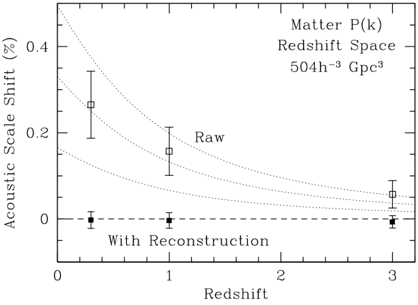
Density field reconstruction (it works in data as well!)

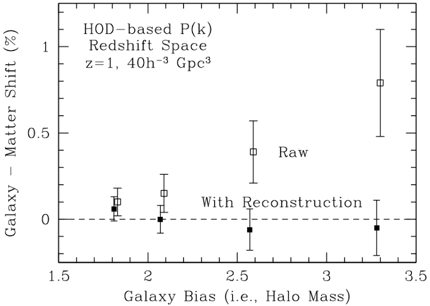


Anderson et al 2014

$\rightarrow > 7\sigma$ significance of the BAO detection

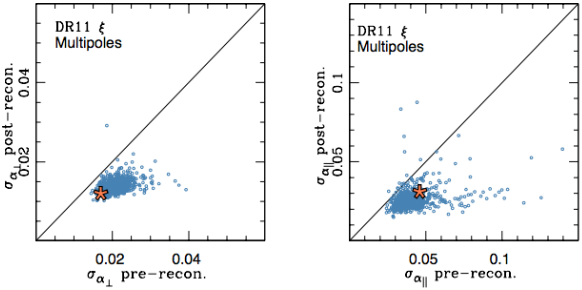
- Density field reconstruction increases the BAO signal and improves the distance constraint by a factor of ~ 2 which corresponds to a increase in survey volume by a factor of ~ 4 (Padmanabhan et al. 2012).
- One can show that reconstruction removes the shift due to the mode coupling term (see e.g. Sherwin et al. 2012).





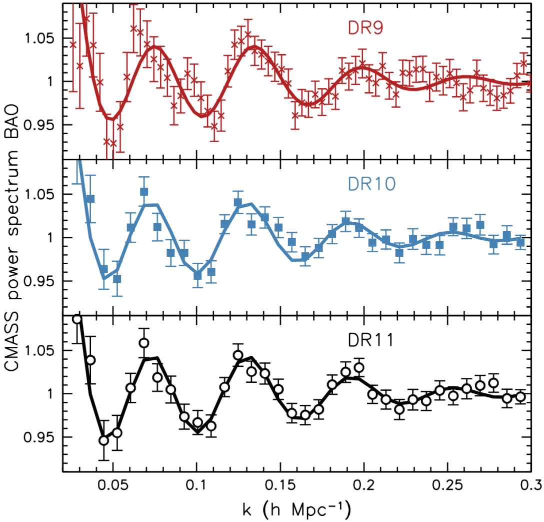
Seo et al. 2010 / Mehta et al 2011

- SDSS-II, LRGs ($z = 0.35$): 3.5% $\rightarrow$ 1.9%, Padmanabhan et al. (2012).
- SDSS-III, CMASS ($z = 0.57$): 1.4% $\rightarrow$ 0.9%, Anderson et al. (2014).
- WiggleZ ($z = 0.4 - 0.7$): about a factor of 1.5 improvement, Kazin et al. (2014).



Anderson et al. (2014)

from DR9 to DR11

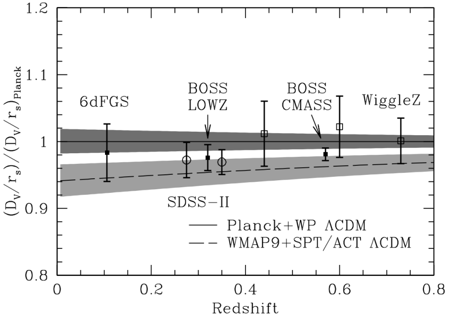


Anderson et al (2014)

Current Results and cosmological implications

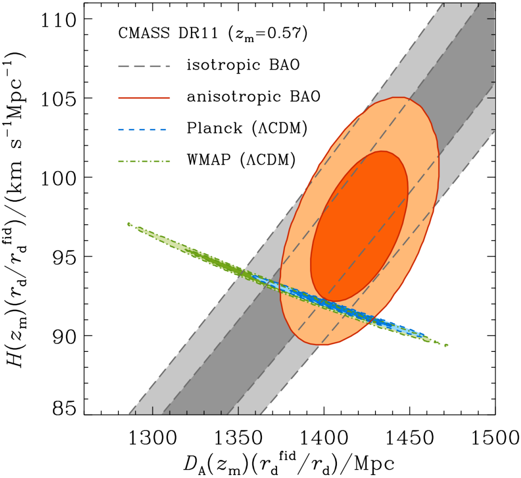
$$\text{Planck: } \Omega_m h^2 = 0.1427 \pm 0.0024$$

$$\text{eWMAP: } \Omega_m h^2 = 0.1353 \pm 0.0035$$

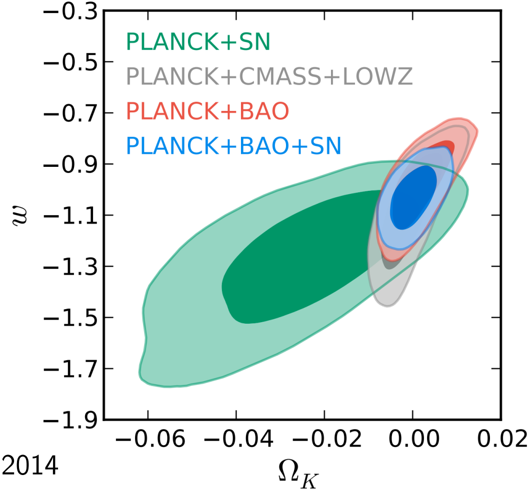


Anderson et al. 2013

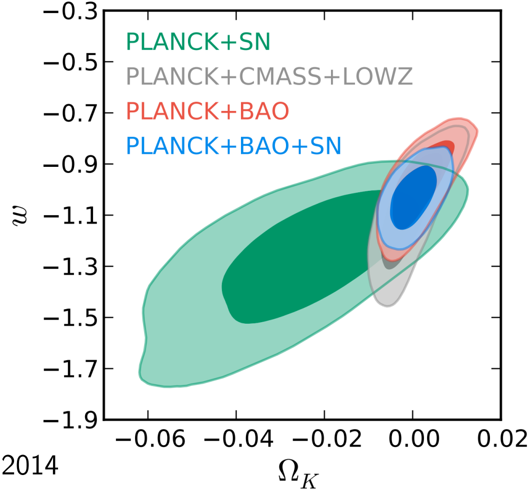
$$D_V(z) = \left[(1+z)^2 D_A^2(z) \frac{cz}{H(z)} \right]^{1/3}$$



Anderson et al 2014



Anderson et al 2014

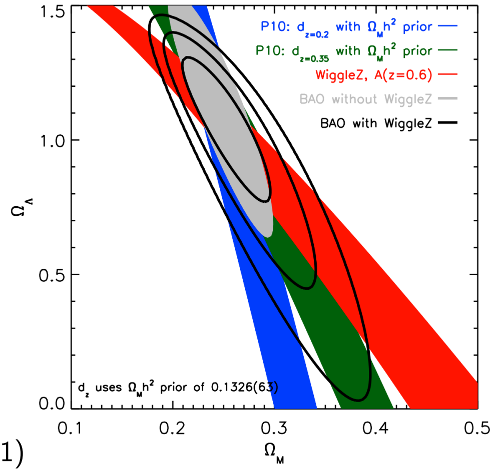


Anderson et al 2014

$$\Omega_k = 0.0002 \pm 0.0034$$

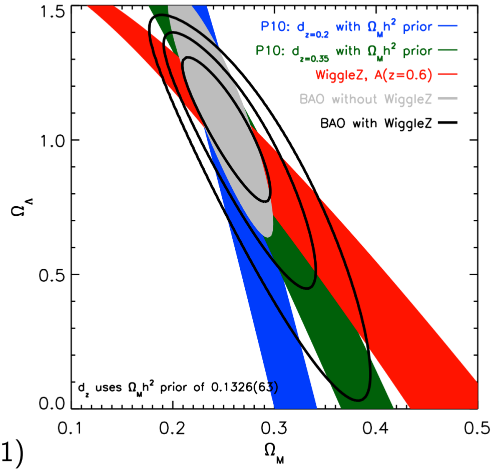
$$w = -1.03 \pm 0.07$$

Evidence for dark energy from BAO



Blake et al. (2011)

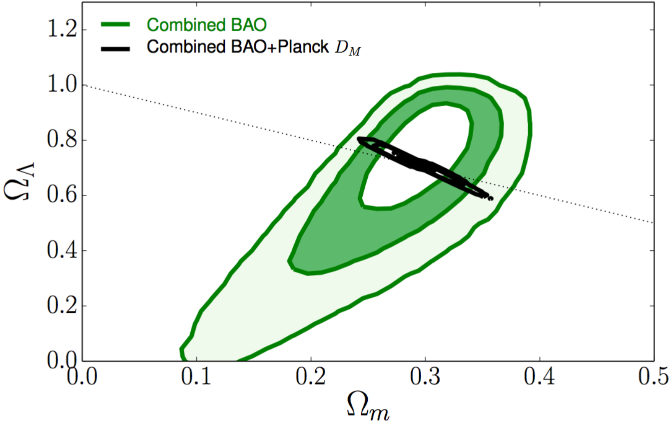
Evidence for dark energy from BAO



Blake et al. (2011)

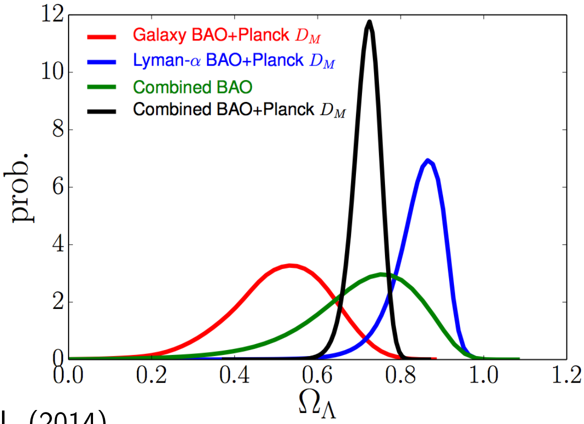
$$\Omega_\Lambda = 1.1^{+0.2}_{-0.4}$$

Evidence for dark energy from BAO



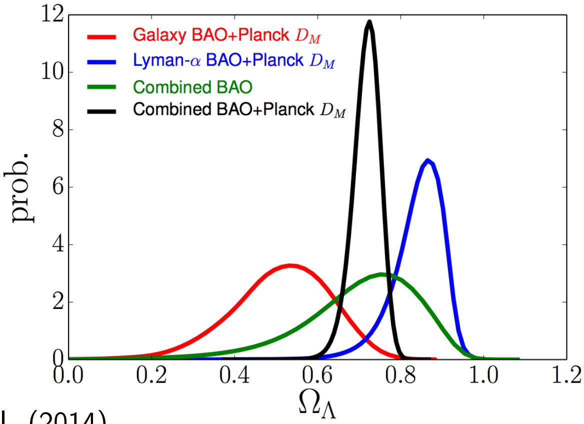
Aubourg et al. (2014)

Evidence for dark energy from BAO



Aubourg et al. (2014)

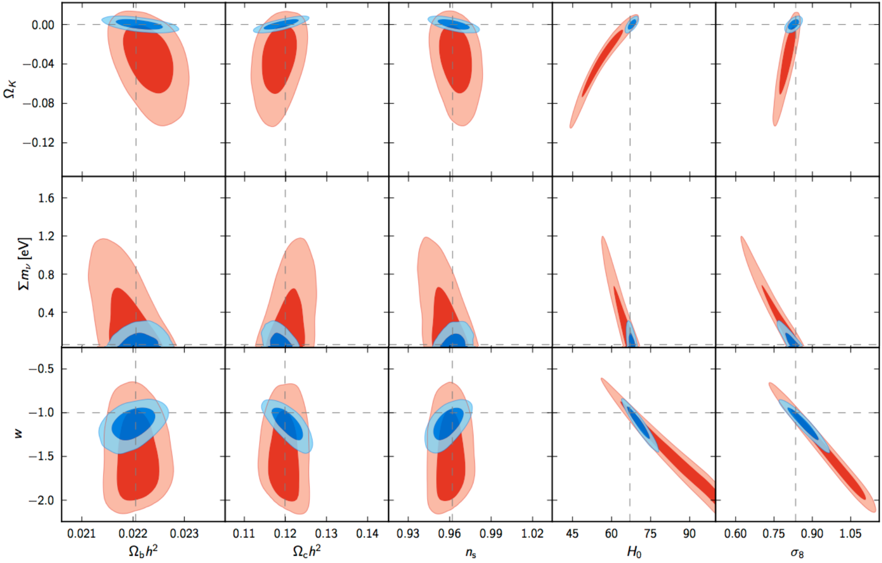
Evidence for dark energy from BAO



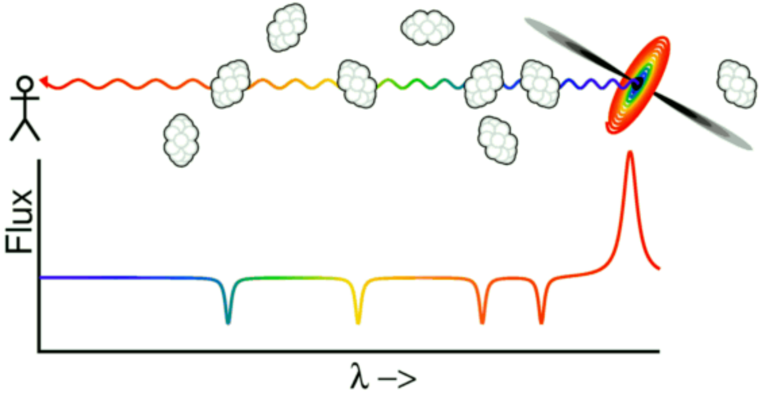
Aubourg et al. (2014)

$$\Omega_\Lambda = 0.73^{+0.25}_{-0.68} \qquad (99.7\% C.L.)$$

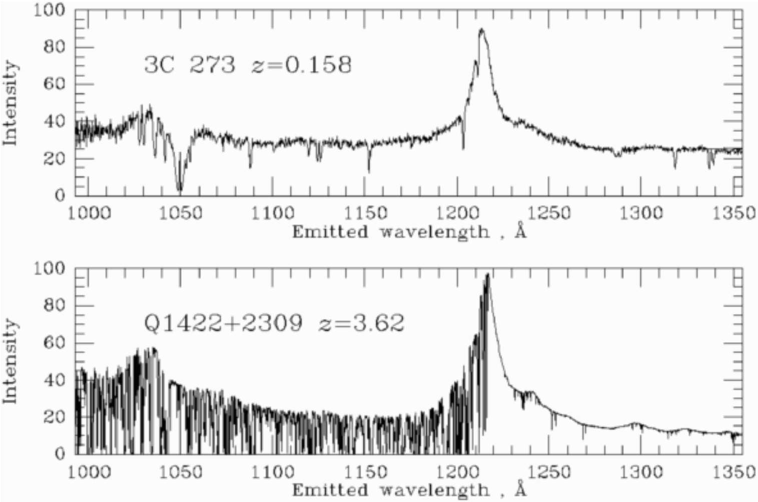
PLANCK, PLANCK+BAO

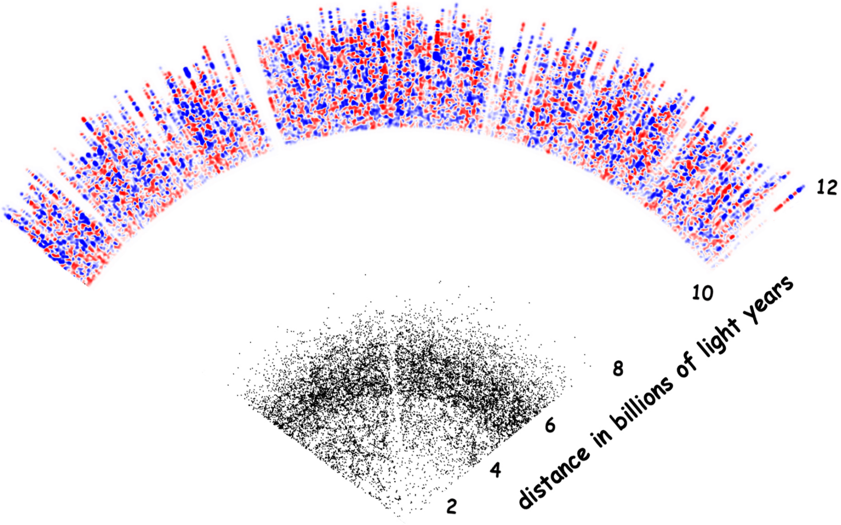


BAO in the Ly-α forrest

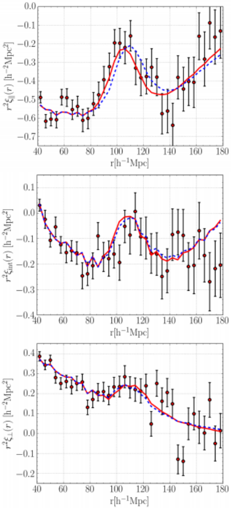


BAO in the Ly-α forrest



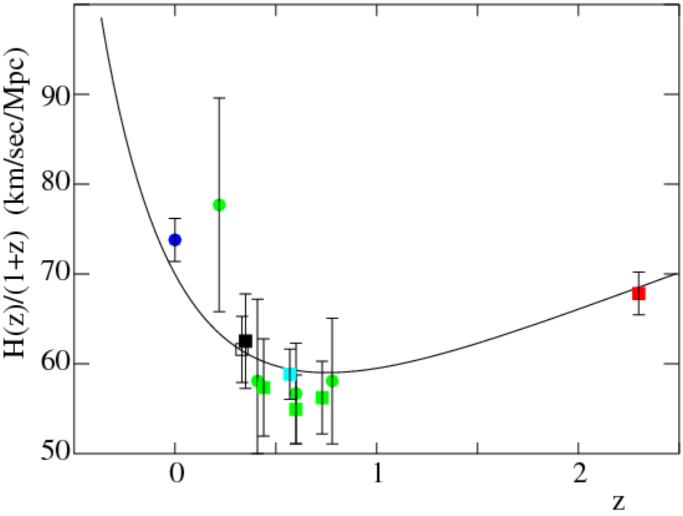

12
10
8
6
4
2
distance in billions of light years

- The absorption spectrum of quasars does carry information about the density field including BAO.
- First detected in BOSS DR9, Busca et al. (2012) and Slosar et al. (2012)
- The DR11 detection resulted in a 3% BAO constraint at $z = 2.5$, Delubac et al. (2014)
- We also detected the BAO signal in the cross-correlation between quasars and Ly-α forrest, Font-Ribera et al. (2013)



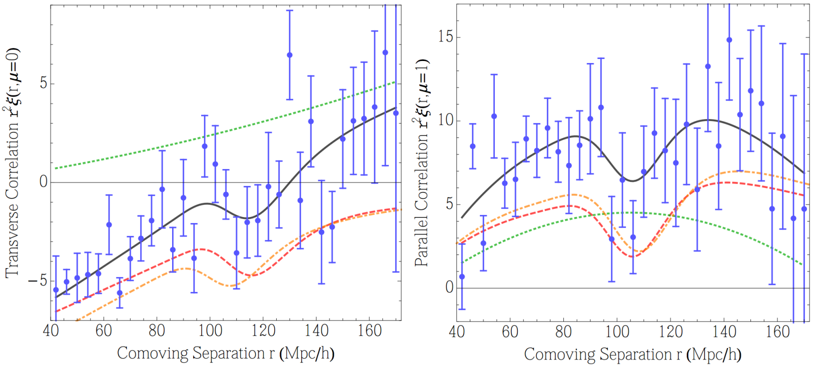
Delubac et al. (2014)

Extending the BAO Hubble diagram



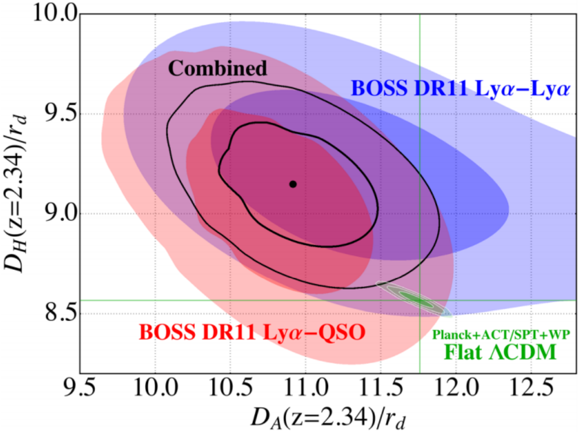
Busca et al. 2013

BAO in the cross-correlation of Quasars and Ly-α forrest



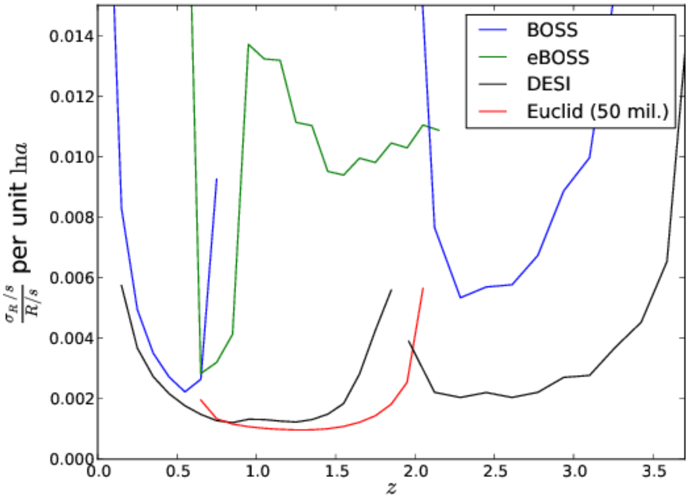
Font-Ribera et al. (2014)

We have a 2.5σ tension with ΛCDM.



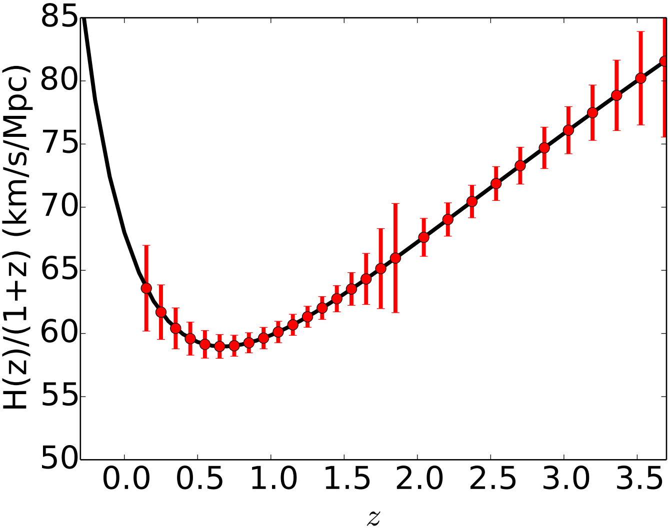
Delubac et al. (2014)

Looking into the future



Font-Ribera et al. 2014

Looking into the future



credit: Patrick McDonald

Conclusion

- Baryon Acoustic Oscillations are a firm prediction of CDM models.
- The acoustic signature has now been detected in many galaxy surveys (6dFGS, SDSS, 2dFGRS, WiggleZ, BOSS).
- The best current constraints (1%) come from the Baryon Oscillation Spectroscopic Survey (BOSS).
- Linear theory is (almost) all we need to measure the BAO signal.
- Density field reconstruction can improve the distance constraints by reducing the damping of the BAO peak. Reconstruction also removes second order systematic shifts in the BAO scale.
- The BAO signature can be used to make distance measurements up to very high redshift using Lyα absorption and maybe even higher using 21cm (?)
- Even future galaxy surveys like DESI will most likely be statistics limited so the future for BAO looks bright.